Comparative analysis of repetitive elements in fish genomes

by

Zihao Yuan

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 5, 2018

Keywords: Fish, Repetitive elements, Genome, Expression, Evolution

Approved by

Zhanjiang Liu, Chair, Professor of School of Fisheries, Aquaculture and Aquatic Sciences
Rex Dunham, Co-chair, Professor of School of Fisheries, Aquaculture and Aquatic Sciences
Charles Y. Chen, Professor of Department of Crop, Soil and Environmental Sciences
Bill Daniels, Associate Professor of School of Fisheries, Aquaculture and Aquatic Sciences

Abstract

Repetitive elements make up significant proportions of genomes. However, their roles in evolution remain largely unknown. To provide insights into the roles of repetitive elements in fish genomes, this dissertation work focused on comparative analysis of repetitive elements presented in three chapters: general overview and literature review (Chapter 1), analysis and annotation of the repetitive elements in the channel catfish genome (Chapter 2), and comparative analysis of repetitive elements from a large number (52) of the fish genomes (Chapter 3)

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 the industry. 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 transposable elements are the most abundant type, making up ~20% of the total repetitive

elements, followed by the microsatellites (14%). The prevalent 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.

The content of the repetitive elements varies significantly among the fish genomes. To provide insights into the roles of repetitive elements in fish genomes, a comparative analysis of repetitive elements from all sequenced teleost fish genomes of 52 species in 22 orders was conducted. The proportions of repetitive elements in various genomes were found to be positively correlated with genome sizes, with a few exceptions. More importantly, specific enrichment appeared between some repetitive element categories with species habitats. Specifically, class II transposons appear to be more abundant in freshwater bony fishes than in marine bony fishes when the phylogeny is not considered. In contrast, marine bony fishes harbor more tandem repeats than freshwater species.

In addition, class I transposable elements appear to be more abundant in primitive species, such as cartilaginous fish and lamprey than in bony fishes. The enriched association of specific categories of repetitive elements with fish habitats suggests the importance of repetitive elements in genome evolution and their potential roles in fish adaptation to their living environments.

Acknowledgments

I would like to express my sincere thanks to my major professor, Dr. Zhanjiang Liu, for his guidance of my research in the entire course of my graduate education. Without his guidance and support, none of my research in the past four years would be possible. I am also grateful to my committee co-chair Dr. Rex Dunham for his guidance and support. I would like to thank my committee member Dr. Charles Chen and Dr. Bill Daniels and my University reader Dr. Luxin Wang for their time, patience and advice related to my research work. I would also like to thank Dongya Gao for her constant guidance, professional advice and support that makes me feel like being at home 8,000 miles from my hometown.

My graduate education in the United States was made possible with a scholarship from the China Scholarship Council, and I am extremely grateful for this financial support. I would like to thank all my lab mates including Dr. Shikai Liu, Ludmilla Kaltenboeck, Dr. Jiaren Zhang, Dr. Luyang Sun, Dr. Ailu Chen, Dr. Chen Jiang, Dr. Yun Li, Dr. Jun Yao, Dr. Xin Geng, Dr. Lisui Bao,

Dr. Qifan Zeng, Dr. Sen Gao, Dr. Ning Li, Dr. Tao Zhou, Dr. Xiaozhu Wang, Dr. Chenglong Wu, Dr. Xiaoyan Xu, Dr. Jianhua Ming, Dr. Lei Gao, Dr. Xiaoxiao Zhong, Dr. Wansheng Jiang, Dr. Donghong Niu, Dr. Changxu Tian, Mr. Wendy Prabowo, Dr. Ribamar Nunes, Mr. Erlong Wang, Ms. Yujia Yang, Mr. Suxu Tan, Ms. Yulin Jin, Ms. Huitong Shi and Ms. Wenwen Wang for their help and support. The research environment in Dr. Liu's laboratory is incredibly powerful for my professional training. I appreciate the help and assistance from Dr. Wei Huang from University of Edinburgh.

Finally, but not leastly, I would like to express my deepest thanks to my beloved parents, my family and all my friends for their unconditional support.

Table of Contents

List of Tables

List of Illustrations

Chapter 1. Introduction and Literature Review

1.1 Overview

1.1.1 The study of repetitive elements in aquatic species

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 (Maumus and Quesneville, 2014). The non-protein coding repetitive sequences were once thought to be junk DNA (Ohno, 1972), but recent studies have indicated that they play important roles in propelling genome evolution and adaptation to environments, and those repetitive elements are of vital importance for the organism to adapt to the changing environment. (De Boer et al., 2007; Lynch and Conery, 2003; Meagher and Vassiliadis 2005; Schmidt and Anderson 2006; Schrader et al., 2014; Sun et al., 2015; Thomas 1971; Thornburg et al., 2006; Tucker et al., 2011; Wang et al., 2014; Wang et al., 2015; Yi et al., 2005; Yu et al.,2016). Besides facilitating the genome evolution, the insertion of repetitive elements can also support genome structure and provide alternative promoters, exons, terminators and splice junctions to protein-coding loci (Babushok et al, 2007; Chung et al., 2011; Day et al. 2010; Häsler et al, 2007; Jurka, 2004; Kazazian, 2004; Peaston et al, 2004; Rosenfeld et al., 2009; Speek, 2001). For instance, retrotransposon insertions can disrupt gene expression (Han et al, 2004; Ustyugova et al, 2006) and cause numerous diseases (Batzer and Deininger, 2002; Yang and Kazazian, 2006). DNA and RNA transposable elements are also transcribed and accumulate in various conditions such as cancer (Criscione et al., 2014; Lamprecht et al., 2010; Lee et al., 2012; Sciamanna et al., 2013; Sciamanna et al., 2014; Shukla et al., 2013; Tubio et al., 2014), eurodegenerative diseases (Bundo et al., 2014; Reilly et al., 2013), embryogenesis (Fadloun et al., 2013; Macia et al., 2011), neural development (Bodega et al, 2011; Coufal et al., 2009; Coufal et al., 2011; Faulkner et al., 2009; Muotri et al., 2005; Perrat et al., 2013; Thomas et al., 2012) as well as aging (De Cecco et al., 2013; Li et al., 2013; Sedivy et al., 2013). However, to date, most of the studies of the repeatomes are focused in the plants and higher vertebrates, especially those of mammals, the studies of repetitive elements are still limited for aquatic species especially in fish.

First appeared 530 million years ago, the fish had colonized on the earth and had undergone huge evolution selections. As many phylogenetic classifications suggest, the three main existing clades of the fish are: Osteichthyes (bony fish), Chondrichthyes (cartilious fish) and Petromyzontiformes (lamprey) (Nelson, 2006). The fish consist the largest and the most diverse vertebrate group on earth, compromising over half of all vertebrates (Hurley et al., 2007; Nelson 2006), and possess both the smallest and largest genomes in vertebrates (Hardie & Hebert, 2004), the sequenced genomes of the fish varies dramatically from the smallest *Tetraodon nigroviridis* of 342.4 Mb (Jaillon et al, 2004) to the largest *Salmon salar* of 2,966.89 MB, the research indicated that the whole genome duplications are known to be a major force of variations in genome sizes (Allendorf et al., 1984; Meyer and Peer, 2005; Xu et al., 2014). Besides propelling the whole genome duplications, the previous researches also indicated that, like in other species, the expansion of repetitive elements can also act as a significant propelling factor to the variations of genome sizes in different genomes (Bengtén et al., 2006; Boulesteix et al., 2005; Holland et al., 2001; Ingle et al., 1975; Liu et al., 2003; Organ et al., 2007; Rubin and Levy 1997; Sen et al., 2006; Ungerer et al., 2006). Thus the characterization and analysis of repetitive elements landscape in fish can not only help to understand the “dark matter” in the fish genome, but also provide information for the study of the genome evolution of life from a broad aspect. In addition, the study of how the repetitive elements are evolved in the fish are of vital importance to the understanding of the evolution dynamics and the environmental influences on the Eukaryotes genome structures.

Catfish (order Siluriformes) is one of the largest orders of teleost containing ~4,100 species, representing ~12% of all teleost and ~6.3% of all vertebrates (Eschmeyer and Fong, 2014). The catfish is a benthic feeding species discovered in all continents (Bruton, 1996, Grande and Eastman, 1986, Lundberg and Friel, 2003), making it important not only for evolution but also for the biogeographical studies (Sullivan et al., 2006). The channel catfish (*Ictalurus punctatus*) is a freshwater benthic species distributed from lower Canada to southern North America. Its high tolerance and adaptability to harsh environments such as tolerance to low oxygen, resistance to disease (eg. Columnaris disease or Enteric Septicemia of Catfish), as well as high feed conversion ration up to 2.4, (Robinson et al., 1998) making it one of the most popular aquaculture species in the world especially in the south-east United States (FAO, 2015). The channel catfish along with its F1 hybrid with blue catfish (*Ictalurus furcatus*) is the primary aquaculture species, accounting for over 60% of all U.S. aquaculture production (Liu, 2011;

USDA-NASS, 2006; United States Department of Agriculture, 2014). Its world-wide importance is constantly rising as its market share is constantly increasing in some Asian countries such as China and Vietnam (Liu, 2008).

The research on the channel catfish genome was initiated in early 1990s, (Liu et al., 1992; Liu et al., 1999; Liu 2003), aiming to dissect the entire genome of channel catfish, after almost 30 years of research, great efforts have been done to the exploration of the basic facts of channel catfish genome such as the development of genome markers, analyzing the genome structure as well as sequencing its whole genome sequences. Like most other teleost, the channel catfish had also undergone the teleost-specific genome duplication (TSGD) in its evolutionary histories (Kasahara et al, 2007; Meyer and Van de Peer, 2005). Channel catfish genome contains 29 pairs of chromosome, rich in A/T and its sizes is approximately 1.0 GB (LeGrande et al, 1984; Tiersch and Goudie, 1993; Xu et al., 2006). The channel catfish reference genome sequence with annotation of protein coding genes was published in 2016, as a milestone for the channel catfish genome project (Liu et al., 2016). Previous works reported the presence of an A/T-rich tandem repeats named Xba on the centromere regions (Liu et al., 1998; Quiniou et al., 2005), and the presence of dispersed short interspersed nuclear elements (Kim et al., 2000) or DNA transposable elements (Liu et al., 1999). But the repetitive elements which make up more than 40% of the channel catfish genome was not fully characterized. Genomic sequencing surveys can provide additional information about other repetitive sequences such as microsatellites and transposable elements (Chen et al., 2016; Liu et al., 2009; Nandi et al., 2007; Serapion et al., 2004; Xu et al., 2006, Xu et al., 2007), making the comprehensive annotation and characterization of the repeatome in channel catfish genome possible. The repetitive sequences in channel catfish can be generally divided into three major categories: the dispersed repeats such as transposable elements or transposons, tandem repeats, and high copy number genes (Maumus and Quesneville, 2014; Nagarajan et al., 2008).

Transposable elements were first discovered in 1940s to 1950s, the concept was applied to explain the genes that are responsible for controlling the physical characteristics in maize (*Zea mays*). The discoverer of the transposable elements, Barbara McClintock (1902-1992) regarded the transposable elements as “normal components of the chromosome responsible for controlling, differentially, the time and type of activity of individual genes” (McClintock, 1956).

In spite there are skepticism at the moment, the important role of the transposable elements in the genome was understood and reevaluated in 1960s and 1970s, bringing insights into the understanding of the complexity of the repetitive elements in the genomes. (Chuong et al, 2017). Transposable elements are commonly found in the eukaryotes genomes (Dombroski et al, 1991), and they are capable of moving along the genome, and therefore, they are believed to be a major driving force for genomic evolutions and genome expansions (Hurst and Schilthuizen 1998; Hurst and Werren 2001; Kazazian 1999; Kazazian 2004). Transposable elements are dispersed across genomes and their proportion are highly variable among genomes, ranging from 3% to 85% in terms of the sequence length (SanMiguel et al., 1996; Wicker et al., 2007). For instance, the genome of *Utricularia gibba* contains only 3% of repetitive sequences (Ibarra-Laclette et al., 2013; Lee et al., 2014), while 85% of the maize genome are composed of transposable elements (Schnable et al., 2009; Vicient, 2010).

Based on the proliferation mechanism of the transposable elements, in 1989, Finnegan proposed the first classification system for the transposable elements, which distinguished two classes of transposable elements by their transposition intermediate: RNA transposable elements

(Class I) or DNA transposable elements (Class II). The transposition mechanism of Class I transposable element is commonly called “copy-and-paste”, and that of Class II transposable element is known as “cut-and-paste” (Finnegan,1989). Class I transposable elements contain three main subclasses: transposable elements with long terminal repeats (LTRs); long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). The detailed descriptions of the respective transposable elements categories are followed.

1.1.2 The Class I transposable elements and its potential role

1.1.2.1 Long terminal repeats (LTRs)

Long terminal repeats (LTRs) are functional DNA sequences found at each end of open reading frames (ORFs) in the integrated retrovirus. The long terminal repeats flanking the ORF regions of the retrovirus is enabled with the promoter capabilities. The flanked ORFs codes for the essential viral proteins such as capsid protein (GAG), aspartic proteinase (AP), reverse transcriptase (RT), and RNase H (RH). First discovered in 1977 (Schwartz et al, 1977), it was characterized that the LTR regions can recognize host cell DNA sequences, and facilitate the integration of the retrovirus genome into the host cell chromosomes. Previous analysis had indicated that at least 20,000-40,000 copies of solitary LTRs in human genome are considered to be relics of ancient proviruses insertions (Schön et al., 2001; Wilkinson et al., 1994). Thus the study of LTR sequences can help the understanding the origination of external DNAs in Eukaryote genomes.

The insertion of LTRs into the chromosome had greatly influenced the human evolution (Katoh and Kurata, 2013). The human gene and the external LTR sequences had co-evolved and it was estimated that in human genome, over 400,000 copies of LTR transposable elements had provided enhancer-promoter sequences and initiation sites for neighboring human genes (Griffiths, 2001). LTRs are being suppressed in normal biological conditions, however, the incorrect LTR activation can trigger diseases, such as rheumatoid arthritis (Seidl et al., 1999; Pascual et al., 2001), type I diabetes (Pascual et al., 2001), and schizophrenia (Hegyi, 2013).

1.1.2.2 Long interspersed nuclear elements (LINEs)

Long interspersed nuclear elements (LINEs) are a group of ancient non-LTR retrotransposons which are widespread in the genome of many Eukaryotes (Fanning and Singer, 1987; Hutchison et al, 1989; Malik et al., 1999) and are propagated by a unique mechanism:

target-DNA primed reverse transcription (TPRT) (Luan et al., 1993). The structure of the LINEs contains two ORFs. The function of ORF1 is not clear, the ORF2 contains endonuclease (EN), reverse transcriptase (RVT) domain as well as a cysteine-rich domain (C-rich). The ORF2 is responsible for integration of the LTR retrotransposons into the genomic DNA.

The LINEs are important in the evolution of the primate genome, in human genome, the LINEs compromise about 21% of the genome, being one of the most abundant and active repetitive elements in human genome. Besides being important in the expansion of the human genome (Smit., 1996), some full-length LINEs elements retain the ability to retrotranpose in the genome and cause human disease and genetic disorders such as systemic autoimmune disease (Burwinkel and Kilimann., 1998; Crow, 2010; Mavragani et al., 2016; Segal et al., 1999). Besides, since the some of the activated LINEs elements may skip their polyadenylation signals and use a second downstream polyadenylation site, thus those LINEs are equipped with the ability to transduce the DNA flanking their 3′ ends to new genomic locations, called the 3’tranduction. (Goodier et al., 2000), this kind of transduction was mainly observed in human genome (Holmes et al., 1994; McNaughton et al., 1997; Miki et al., 1992; Rozmahel et al., 1997).

1.1.2.3 Short interspersed nuclear elements (SINEs)

The 100bp - 400bp long SINEs are characterized by an internal polymerase III promoter and a left and a right dimer. The promoter is presented in a tRNA-derived region in all SINEs on the left dimer region (Kachroo et al., 1995; Okada, 1991; Yoshioka et al., 1993). The left dimer also contains functional, but weak A and B boxes of the RNA polymerase III promoter. The left dimer is complementary to the right dimer, which can sometimes form a helix with the left dimer. (Häsler and Strub, 2006).

The SINEs elements are also abundant in human as well as other primate linkages, covering a large number of copies (300,000 to 500,000 copies in human) (Bailey et al., 2003; Pace and Feschotte, 2007; Rogers, 1985). Researches indicated that the SINE elements are actively involved in a number of biological activities, for example, the study of SINE/B2 element had indicated that the transcribed RNAs from SINE elements can regulated the mRNA productions (Endoh et al., 1990; Ponicsan et al., 2010). In addition, the SINEs elements are also involved in regulation of the gene expressions (Li et al., 1999), the researches in mouse also suggests that the SINEs elements can increase the integration efficiency of the external genes into the genome (Kang et al., 1999).

1.1.3. The Class II DNA transposable elements and its potential role

The Class II DNA transposable elements are also called the RNA-independent Class II DNA transposable elements. DNA transposable elements are generally moved by a “cut-and-paste” mechanism in which the transposable element is excised from one location and reintegrated elsewhere on the genome. Most DNA transposable elements are moved and duplicated through a non-replicative mechanism. A typical structure of DNA transposable elements is consisted of a transposase gene that is flanked by two terminal inverted repeats (TIRs). The transposase is capable of recognizing the TIRs sites and excise the main transposable element DNA body, then subsequently insert it into a new genomic location. Once inserted, target site DNA is duplicated, resulting the target site duplications (TSDs), which is a unique hallmark for each DNA transposable element (Muñoz-López and García-Pérez, 2010).

DNA transposable elements significantly contributed to the genome evolution, in terms of size as well as functions (Gao et al., 2016; Petrov, 2001; Piegu et al., 2006). Besides, since DNA transposable elements possess the mighty force in changing the genome, the modified DNA transposable elements can be applied as effective transgenic tools (Ivics et al., 1997; Mátés et al., 2009). For example, in normal biological conditions, the repetitive Tc1/mariner-type elements were inactive and silent (Izsvák et al., 1995). But in certain conditions, they can be re-activated by artificial intervention. Researchers had resurrected the inactive Tc1/mariner-like transposons in salmonids and reformed it into a trans-genetic tool named Sleeping Beauty system (Ivics et al., 1996; Ivics et al., 1997). The Sleeping Beauty system replace the transpose gene in the transposable element with an expression cassette. By inserting the expression cassette into the target site on the genome, the system can transfer the external DNA sequence into the target genome (Geurts et al., 2003). After first invented in 1997, the sleeping beauty system had been widely applied as a gene-transfer tool for transgenic studies and has huge potential in the field of human gene therapies and mutagenesis screens. (Aronovich et al., 2011; Eisenstein, 2005; Geurts, 2003; Muñoz-López and García-Pérez, 2010).

In general, transposable elements constitute the largest component of mammalian genomes and because their unique ability to transpose and are frequently amplified, the accumulated transposable elements have become a major determinants of genome size (Petrov, 2001; Piegu et al., 2006), and with huge impacts on shaping the structure and evolution of eukaryotic nulear genomes.

1.1.4 The tandem repeats and its potential role

Tandem repeats are individual repeats of DNA located adjacent to one another in the genome (Benson, 1999; Kubis et al., 1998; Tóth et al., 2000; Ugarković and Plohl 2002). 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 (Hacch and Mazrimas 1974; Petitpierre et al., 1995). The amplification and/or mutations of tandem repeats may also affect the genome by changing the genome structures or genome sizes (Charlesworth et al., 1994; Lindahl 1994; Strand et al., 1993), thereby affecting recombination of genomes, gene expressions, gene conversions, and chromosomal organizations (Balaresque et al., 2014; Martin et al., 2005; Moxon et al., 1994; Pardue et al., 1987; Richard et al., 2000).

There are different classifications systems of the tandem repeats (Beckmann and Weber, 1992; Gordenin et al., 1997; Litt and Luty, 1989; Nakamura,1987; Tautz,1993; Weber and

May,1989), causing confusions. One of the recommended nomenclature for the tandem repeats are shown in table 1-1 (Chambers and MacAvoy, 2000).

Table1-1. A classification of microsatellites, minisatellites and satellite sequences.

	Unit	Location	Examples
Microsatellites	2-6	Euchromatic	Genomic markers
Minisatellites	10-100	Euchromatic	DNA fingerprints
Satellites	>100	heterochromatic	Telomere DNA

Tandem repeats, especially the microsatellites are widely applied as the second generation genomic markers, although there is a trend to be replaced by third generation Single Nucleotide Polymorphism (SNP) markers and whole genome sequencing technology. The function of the tandem repeats is hardly seen and are still under studies. However, the accumulation of the simple nucleotide units in the genome can imply an increased risk of homologous recombination between chromosomes and resulting in chromosomal deletions, translocations as well as inversions (Flavell, 1985). From a evolutionary point, the accumulation or loss of microsatellites sequences on the chromosome may lead to the chances of chromosome rearrangements (Huang et al., 2015), and subsequently reduce the fertility of heterozygous hybrids and cause speciations (Rieseberg., 2001; White et al., 1978).

1.2 Goals and specific objectives

1.2.1 Long-term goal

The long-term goal of this study was to characterize, annotate and analyze the repetitive elements and their evolutionary dynamics in the channel catfish. In addition, through the comparative analysis of the repetitive elements across the fish phylogeny, it is hoped to contribute to the understanding the evolution dynamics of the repetitive elements and how the changing environment shaped the genome architecture. To reach these long-term goals, the dissertation project will accomplish the following specific objectives.

1.2.2 Specific objectives:

1. Characterization of repetitive elements in channel catfish genome;
2. Analyze and compare the evolutionary dynamics of different categories of repetitive elements in channel catfish and blue catfish;
3. Comparison of repetitive elements covering the existing fish clades: Osteichthyes (bony fish),

Chondrichthyes (cartilaginous fish) and Petromyzontiformes (lamprey)

References

Allendorf, F.W., Thorgaard, G.H., 1984. Tetraploidy and the evolution of salmonid fishes, Evolutionary genetics of fishes. Springer, pp. 1-53.

Aronovich, E.L., McIvor, R.S., Hackett, P.B., 2011. The Sleeping Beauty transposon system: a non-viral vector for gene therapy. Human molecular genetics. 20, R14-R20.

Babushok, D., Ostertag, E., Kazazian, H., 2007. Current topics in genome evolution: molecular mechanisms of new gene formation. Cellular and molecular life sciences. 64, 542-554.

Bailey, J.A., Liu, G., Eichler, E.E., 2003. An Alu transposition model for the origin and expansion of human segmental duplications. The American Journal of Human Genetics. 73, 823-834.

Balaresque, P., King, T.E., Parkin, E.J., Heyer, E., Carvalho- Silva, D., Kraaijenbrink, T., Knijff, P., Tyler- Smith, C., Jobling, M.A., 2014. Gene Conversion Violates the Stepwise Mutation Model for Microsatellites in Y- Chromosomal Palindromic Repeats. Human mutation. 35, 609-617.

Batzer, M.A., Deininger, P.L., 2002. Alu repeats and human genomic diversity. Nature reviews genetics. 3, 370-379.

Beckmann, J.S., Weber, J.L., 1992. Survey of human and rat microsatellites. Genomics. 12, 627-631.

Bengtén, E., Clem, L.W., Miller, N.W., Warr, G.W., Wilson, M., 2006. Channel catfish immunoglobulins: repertoire and expression. Developmental & Comparative Immunology. 30, 77-92.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research. 27, 573.

Bodega, B., Valle, D., Marasca, F., Carninci, P., Faulkner, G., Orlando, V., 2011. Repetitive elements transcription and mobilization contribute to human skeletal muscle differentiation and Duchenne muscular dystrophy progression. The Febs Journal. 278, 75-76.

Boulesteix, M., Weiss, M., Biémont, C., 2005. Differences in genome size between closely related species: the Drosophila melanogaster species subgroup. Molecular biology and evolution. 23, 162-167.

Bruton, M.N., 1996. Alternative life-history strategies of catfishes. Aquatic Living Resources. 9, 35-41.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A., 2014. Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron. 81, 306-313.

Burwinkel, B., Kilimann, M.W., 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. Journal of Molecular Biology. 277, 513-517.

Chambers, G.K., MacAvoy, E.S., 2000. Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 126, 455-476.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 371, 215-220.

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., Liu, S., Lu, H., Tian, S., Li, R., 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies Apis mellifera sinisxinyuan n. ssp. Molecular biology and evolution. 33, 1337-1348.

Chen, X., Zhong, L., Bian, C., Xu, P., Qiu, Y., You, X., Zhang, S., Huang, Y., Li, J., Wang, M., 2016. High-quality genome assembly of channel catfish, Ictalurus punctatus. GigaScience. 5, 39.

Chung, D., Kuan, P.F., Li, B., Sanalkumar, R., Liang, K., Bresnick, E.H., Dewey, C., Keleş, S., 2011. Discovering transcription factor binding sites in highly repetitive regions of genomes with multi-read analysis of ChIP-Seq data. PLoS computational biology. 7, e1002111.

Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nature Reviews Genetics. 18, 71.

Criscione, S.W., Zhang, Y., Thompson, W., Sedivy, J.M., Neretti, N., 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC genomics. 15, 583.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., Gage, F.H., 2009. L1 retrotransposition in human neural progenitor cells. Nature. 460, 1127-1131.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V., Gage, F.H., 2011. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. Proceedings of the National Academy of Sciences. 108, 20382-20387.

Crow, M.K., 2010. Long interspersed nuclear elements (LINE-1): potential triggers of systemic autoimmune disease. Autoimmunity. 43, 7-16.

Day, D.S., Luquette, L.J., Park, P.J., Kharchenko, P.V., 2010. Estimating enrichment of repetitive elements from high-throughput sequence data. Genome biology. 11, R69.

De Cecco, M., Criscione, S.W., Peterson, A.L., Neretti, N., Sedivy, J.M., Kreiling, J.A., 2013. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging (Albany NY). 5, 867.

De Boer, J.G., Yazawa, R., Davidson, W.S., Koop, B.F., 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC genomics. 8, 422.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., Kazazian Jr, H.H., 1991. Isolation of an active human transposable element. Science. 254, 1805-1808.

Eisenstein, M., 2005. Wake-up call for Sleeping Beauty. Nature methods. 2, 637-637.

Endoh, H., Nagahashi, S., Okada, N., 1990. A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. The FEBS Journal. 189, 25-31.

Eschmeyer, W., Fong, J., 2014. Catalog of fishes: species by family/subfamily (http://research. calacademy. org/research/ichthyology/catalog/SpeciesByFamily. asp). Accessed.

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., Torres-Padilla, M.-E., 2013. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nature structural & molecular biology. 20, 332-338.

Fanning, T., Singer, M., 1987. LINE-1: a mammalian transposable element. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression. 910, 203-212.

FAO 2015. *Ictalurus punctatus*. Avalilable at http://www.fao.org/fishery/culturedspecies/Ictalurus_punctatus/en.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., 2009. The regulated retrotransposon transcriptome of mammalian cells. Nature genetics. 41, 563-571.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. Trends in genetics. 5, 103-107.

Flavell, R., 1985. Repeated sequences and genome change. Genetic Flux in Plants, 139-156.

Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., Song, C., 2016. The contribution of transposable elements to size variations between four teleost genomes. Mobile DNA. 7, 4.

Geurts, A.M., Yang, Y., Clark, K.J., Liu, G., Cui, Z., Dupuy, A.J., Bell, J.B., Largaespada, D.A., Hackett, P.B., 2003. Gene transfer into genomes of human cells by the sleeping beauty transposon system. Molecular Therapy. 8, 108-117.

Goodier, J.L., Ostertag, E.M., Kazazian Jr, H.H., 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Human molecular genetics. 9, 653-657.

Gordenin, D.A., Kunkel, T.A., Resnick, M.A., 1997. Repeat expansion—all in flap? Nature genetics. 16, 116-118.

Grande, L., Eastman, J.T., 1986. A review of Antarctic ichthyofaunas in the light of new fossil discoveries. Palaeontology. 29, 113-137.

Griffiths, D.J., 2001. Endogenous retroviruses in the human genome sequence. Genome biology.

2, reviews1017. 1011.

Hacch, F., Mazrimas, J., 1974. Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys ordii*). Nucleic acids research. 1, 559-576.

Han, J.S., Szak, S.T., Boeke, J.D., 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature. 429, 268-274.

Hardie, D.C., Hebert, P.D., 2004. Genome-size evolution in fishes. Canadian Journal of Fisheries and Aquatic Sciences. 61, 1636-1646.

Häsler, J., Samuelsson, T., Strub, K., 2007. Useful'junk': Alu RNAs in the human transcriptome. Cellular and molecular life sciences. 64, 1793.

Häsler, J., Strub, K., 2006. Alu elements as regulators of gene expression. Nucleic Acids Research. 34, 5491-5497.

Hegyi, H., 2013. GABBR1 has a HERV-W LTR in its regulatory region–a possible implication for schizophrenia. Biology direct. 8, 5.

Holland, J., Helland, S., Sharopova, N., Rhyne, D., 2001. Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. Genome. 44, 1065-1076.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., Kazazian, H.H., 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nature genetics. 7, 143-148.

Huang, J., Zhao, Y., Bai, D., Shiraigol, W., Li, B., Yang, L., Wu, J., Bao, W., Ren, X., Jin, B., 2015. Donkey genome and insight into the imprinting of fast karyotype evolution. Scientific reports. 5, 14106.

Hurley, I.A., Mueller, R.L., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V.E., Yang, Z., Thomas, M.G., Coates, M.I., 2007. A new time-scale for ray-finned fish evolution. Proceedings of the Royal Society of London B: Biological Sciences. 274, 489-498.

Hurst, G.D., Schilthuizen, M., 1998. Selfish genetic elements and speciation. Heredity. 80, 2-8.

Hurst, G.D., Werren, J.H., 2001. The role of selfish genetic elements in eukaryotic evolution. Nature Reviews Genetics. 2, 597-606.

Hutchison, C., 1989. LINEs and related retroposons: long interspersed repeated sequences in the eukayotic genome. Mobile DNA, 593-617.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.-H., Lan, T., Welch, A.J., Juárez, M.J.A., Simpson, J., 2013. Architecture and evolution of a minute plant genome. Nature. 498, 94-98.

Ingle, J., Timmis, J.N., Sinclair, J., 1975. The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. Plant physiology. 55, 496-501.

Ivics, Z., Hackett, P.B., Plasterk, R.H., Izsvák, 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., Izsvák, Z., Minter, A., Hackett, P.B., 1996. Identification of functional domains and evolution of Tc1-like transposable elements. Proceedings of the National Academy of Sciences. 93, 5008-5013.

Izsvák, Z., Ivics, Z., Hackett, P.B., 1995. Characterization of a Tc1-like transposable element in zebrafish (*Danio rerio*). Molecular and General Genetics MGG. 247, 312-322.

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431, 946-957.

Jurka, J., 2004. Evolutionary impact of human Alu repetitive elements. Current opinion in genetics & development. 14, 603-608.

Kachroo, P., Leong, S.A., Chattoo, B.B., 1995. Mg-SINE: a short interspersed nuclear element from the rice blast fungus, Magnaporthe grisea. Proceedings of the National Academy of Sciences. 92, 11125-11129.

Kang, Y.-K., Park, J.S., Lee, C.-S., Yeom, Y.I., Chung, A.-S., Lee, K.-K., 1999. Efficient integration of short interspersed element-flanked foreign DNA via homologous recombination. Journal of Biological Chemistry. 274, 36585-36591.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature. 447, 714-719.

Katoh, I., Kurata, S.-i., 2013. Association of endogenous retroviruses and long terminal repeats with human disorders. Frontiers in oncology. 3.

Kazazian, H.H., 1999. An estimated frequency of endogenous insertional mutations in humans. Nature genetics. 22, 130-130.

Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. science. 303, 1626-1632.

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

Kubis, S., Schmidt, T., Heslop-Harrison, J.S., 1998. Repetitive DNA elements as a major component of plant genomes. Annals of Botany. 82, 45-55.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M.A., Richter, J., Soler, E., 2010. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nature medicine. 16, 571-579.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., 2012. Landscape of somatic retrotransposition in human cancers. Science. 337, 967-971.

Lee, S.-I., Kim, N.-S., 2014. Transposable elements and genome size variations in plants. Genomics & informatics. 12, 87-97.

LeGrande, W.H., Dunham, R.A., Smitherman, R., 1984. Karyology of three species of catfishes (Ictaluridae: Ictalurus) and four hybrid combinations. Copeia, 873-878.

Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., Dubnau, J., 2013. Activation of transposable elements during aging and neuronal decline in Drosophila. Nature neuroscience. 16, 529-531.

Li, T.-H., Spearow, J., Rubin, C.M., Schmid, C.W., 1999. Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. Gene. 239, 367-372.

Lindahl, T., 1994. DNA Repair: DNA surveillance defect in cancer cells. Current Biology. 4, 249-251.

Litt, M., Luty, J.A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American journal of human genetics. 44, 397.

Liu, G., Zhao, S., Bailey, J.A., Sahinalp, S.C., Alkan, C., Tuzun, E., Green, E.D., Eichler, E.E., Program, N.C.S., 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome research. 13, 358-368.

Liu, Q., Goudie, C., Simco, B., Davis, K., Morizot, D., 1992. Gene-centromere mapping of six enzyme loci in gynogenetic channel catfish. Journal of Heredity. 83, 245-248.

Liu, H., Jiang, Y., Wang, S., Ninwichian, P., Somridhivej, B., Xu, P., Abernathy, J., Kucuktas, H., Liu, Z., 2009. Comparative analysis of catfish BAC end sequences with the zebrafish genome. BMC genomics. 10, 592.

Liu, Z., 2003. A review of catfish genomics: progress and perspectives. Comparative and Functional Genomics. 4, 259-265.

Liu Z. 2008. Catfish. In: Kocher, T.D., Kole, C., 2008. Genome mapping and genomics in fishes and aquatic animals Volume 2. Springer Science & Business Media. p. 85-100.

Liu, Z., 2011. Development of genomic resources in support of sequencing, assembly, and annotation of the catfish genome. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics. 6, 11-17.

Liu, Z., Li, P., Dunham, R.A., 1998. Characterization of an A/T-rich family of sequences from channel catfish (*Ictalurus punctatus*). Molecular marine biology and biotechnology. 7, 232-239.

Liu, Z., Li, P., Argue, B., Dunham, R., 1999. Random amplified polymorphic DNA markers: usefulness for gene mapping and analysis of genetic variation of catfish. Aquaculture. 174, 59-68.

Liu, Z., Li, P., Kucuktas, H., Dunham, R., 1999. Characterization of nonautonomous Tc1-like transposable elements of channel catfish (Ictalurus punctatus). Fish Physiology and Biochemistry. 21, 65-72.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., Jiang, C., Sun, L., Wang, R., Zhang, Y., 2016. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nature communications. 7.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 72, 595-605.

Lungberg, J., Friel, J., 2003. Siluriformes. Tree of Life Web Project. Retrieved (2012): 06-16.

Lynch, M., Conery, J.S., 2003. The origins of genome complexity. science. 302, 1401-1404.

Macia, A., Muñoz-Lopez, M., Cortes, J.L., Hastings, R.K., Morell, S., Lucena-Aguilar, G., Marchal, J.A., Badge, R.M., Garcia-Perez, J.L., 2011. Epigenetic control of retrotransposon expression in human embryonic stem cells. Molecular and cellular biology. 31, 300-316.

Malik, H.S., Burke, W.D., Eickbush, T.H., 1999. The age and evolution of non-LTR retrotransposable elements. Molecular biology and evolution. 16, 793-805.

Martin, P., Makepeace, K., Hill, S.A., Hood, D.W., Moxon, E.R., 2005. Microsatellite instability regulates transcription factor binding and gene expression. Proceedings of the National Academy of Sciences of the United States of America. 102, 3800-3804.

Mátés, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nature genetics. 41, 753-761.

Maumus, F., Quesneville, H., 2014. Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS One. 9, e94101.

Mavragani, C.P., Sagalovskiy, I., Guo, Q., Nezos, A., Kapsogeorgou, E.K., Lu, P., Zhou, J.L., Kirou, K.A., Seshan, S.V., Moutsopoulos, H.M., 2016. Long interspersed nuclear element-1 retroelements are expressed in patients with systemic autoimmune disease and induce type I interferon. Arthritis & rheumatology (Hoboken, NJ). 68, 2686.

McClintock, B., 1956. Controlling elements and the gene, Cold Spring Harbor symposia on quantitative biology. Cold Spring Harbor Laboratory Press, pp. 197-216.

McNaughton, J.C., Hughes, G., Jones, W.A., Stockwell, P.A., Klamut, H.J., Petersen, G.B., 1997. The evolution of an intron: analysis of a long, deletion-prone intron in the human dystrophin gene. Genomics. 40, 294-304.

Meagher, T.R., Vassiliadis, C., 2005. Phenotypic impacts of repetitive DNA in flowering plants. New Phytologist. 168, 71-80.

Meyer, A., Van de Peer, Y., 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 27, 937-945.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., Nakamura, Y., 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer research. 52, 643-645.

Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current biology. 4, 24-33.

Muñoz-López, M., García-Pérez, J.L., 2010. DNA transposons: nature and applications in genomics. Current genomics. 11, 115-128.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., Gage, F.H., 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature. 435, 903-910.

Nagarajan, N., Navajas-Pérez, R., Pop, M., Alam, M., Ming, R., Paterson, A.H., Salzberg, S.L., 2008. Genome-wide analysis of repetitive elements in papaya. Tropical Plant Biology. 1, 191-201.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science. 235, 1616-1623.

Nandi, S., Peatman, E., Xu, P., Wang, S., Li, P., Liu, Z., 2007. Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (*Ictalurus punctatus*). Genetica. 131, 81-90.

Nelson J. Fishes of the world 4th edition. Hoboken: John Wiley & Sons, Inc; 2006. p 334; p456.

Ohno, S., 1972. So much" junk" DNA in our genome, Brookhaven symposia in biology, pp. 366-370.

Okada, N., 1991. SINEs. Current opinion in genetics & development. 1, 498-504.

Organ, C.L., Shedlock, A.M., Meade, A., Pagel, M., Edwards, S.V., 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature. 446, 180-184.

Quiniou, S., Wolters, W., Waldbieser, G., 2005. Localization of Xba repetitive elements to channel catfish (*Ictalurus punctatus*) centromeres via fluorescence in situ hybridization. Animal genetics. 36, 353-354.

Pace, J.K., Feschotte, C., 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome research. 17, 422-432.

Pardue, M., Lowenhaupt, K., Rich, A., Nordheim, A., 1987. (dC-dA) n.(dG-dT) n sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. The EMBO journal. 6, 1781.

Pascual, M., Martin, J., Nieto, A., Giphart, M.J., van der Slik, A.R., de Vries, R.R., Zanelli, E., 2001. Distribution of HERV-LTR elements in the 5′-flanking region of HLA-DQB1 and association with autoimmunity. Immunogenetics. 53, 114-118.

Peaston, A.E., Evsikov, A.V., Graber, J.H., De Vries, W.N., Holbrook, A.E., Solter, D., Knowles, B.B., 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Developmental cell. 7, 597-606.

Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., Waddell, S., 2013. Transposition-driven genomic heterogeneity in the Drosophila brain. Science. 340, 91-95.

Petitpierre, E., Juan, C., Pons, J., Plohl, M., Ugarkovic, D., 1995. Satellite DNA and constitutive heterochromatin in tenebrionid beetles, Kew Chromosome Conference IV. Royal Botanic Gardens, London, pp. 351-362.

Petrov, D.A., 2001. Evolution of genome size: new approaches to an old problem. TRENDS in Genetics. 17, 23-28.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D.S., Jackson, S., Wing, R.A., 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome research. 16, 1262-1269.

Ponicsan, S.L., Kugel, J.F., Goodrich, J.A., 2010. Genomic gems: SINE RNAs regulate mRNA production. Current opinion in genetics & development. 20, 149-155.

Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I., Gage, F.H., 2013. The role of transposable elements in health and diseases of the central nervous system. Journal of Neuroscience. 33, 17577-17586.

Richard, G.F., Pâques, F., 2000. Mini-and microsatellite expansions: the recombination connection. EMBO reports. 1, 122-126.

Rieseberg, L.H., 2001. Chromosomal rearrangements and speciation. Trends in ecology & evolution. 16, 351-358

Robinson, E.H., Li, M.H., Brunson, M.W., 1998. Feeding catfish in commercial ponds. Southern Regional Aquaculture Center.

Rogers, J.H., 1985. The origin and evolution of retroposons. International review of cytology. 93, 187-279.

Rosenfeld, J.A., Xuan, Z., DeSalle, R., 2009. Investigating repetitively matching short sequencing reads: the enigmatic nature of H3K9me3. Epigenetics. 4, 476-486.

Rozmahel, R., Heng, H., Duncan, A., Shi, X.-M., Rommens, J., Tsui, L.-C., 1997. Amplification of CFTR exon 9 sequences to multiple locations in the human genome. Genomics. 45, 554-561.

Rubin, E., Levy, A.A., 1997. Abortive gap repair: underlying mechanism for Ds element formation. Molecular and Cellular Biology. 17, 6294-6302.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science. 274, 765-768.

Schmidt, A.L., Anderson, L.M., 2006. Repetitive DNA elements as mediators of genomic change in response to environmental cues. Biological Reviews. 81, 531-543.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., 2009. The B73 maize genome: complexity, diversity, and dynamics. science. 326, 1112-1115.

Schön, U., Seifarth, W., Baust, C., Hohenadl, C., Erfle, V., Leib-Mösch, C., 2001. Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. Virology. 279, 280-291.

Schrader, L., Kim, J.W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., 2014. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nature communications. 5, 5495.

Schwartz, D.E., Zamecnik, P.C., Weith, H.L., 1977. Rous sarcoma virus genome is terminally redundant: the 3'sequence. Proceedings of the National Academy of Sciences. 74, 994-998.

Sciamanna, I., Gualtieri, A., Piazza, P.V., Spadafora, C., 2014. Regulatory roles of LINE-1-encoded reverse transcriptase in cancer onset and progression. Oncotarget. 5, 8039.

Sciamanna, I., Gualtieri, A., Cossetti, C., Osimo, E.F., Ferracin, M., Macchia, G., Aricò, E., Prosseda, G., Vitullo, P., Misteli, T., 2013. A tumor-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in human transformed cell lines.

Oncotarget. 4, 2271.

Sedivy, J.M., Kreiling, J.A., Neretti, N., Cecco, M.D., Criscione, S.W., Hofmann, J.W., Zhao, X., Ito, T., Peterson, A.L., 2013. Death by transposition–the enemy within? Bioessays. 35, 1035-1043.

Seidl, C., Donner, H., Petershofen, E., Usadel, K.H., Seifried, E., Kaltwasser, J.P., Badenhoop, K., 1999. An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis. Human immunology. 60, 63-68.

Segal, Y., Peissel, B., Renieri, A., de Marchi, M., Ballabio, A., Pei, Y., Zhou, J., 1999. LINE-1 elements at the sites of molecular rearrangements in Alport syndrome–diffuse leiomyomatosis. The American Journal of Human Genetics. 64, 62-69.

Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., Dyer, M., Cordaux, R., Liang, P., Batzer, M.A., 2006. Human genomic deletions mediated by recombination between Alu elements. The American Journal of Human Genetics. 79, 41-53.

Serapion, J., Kucuktas, H., Feng, J., Liu, Z., 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (Ictalurus punctatus). Marine biotechnology. 6, 364-377.

Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., 2013. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 153, 101-111.

Smit, A.F., 1996. The origin of interspersed repeats in the human genome. Current opinion in genetics & development. 6, 743-748.

Speek, M., 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Molecular and cellular biology. 21, 1973-1985.

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D., 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 365, 274-276.

Sullivan, J.P., Lundberg, J.G., Hardman, M., 2006. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. Molecular phylogenetics and evolution. 41, 636-662.

Sun, Y.B., Xiong, Z.J., Xiang, X.Y., Liu, S.P., Zhou, W.W., Tu, X.L., Zhong, L., Wang, L., Wu, D.D., Zhang, B.L., 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences. 112, E1257-E1262.

Tautz, D., 1993. Notes on the definition and nomenclature of tandemly repetitive DNA sequences, DNA fingerprinting: State of the science. Springer, pp. 21-28.

Thomas Jr, C.A., 1971. The genetic organization of chromosomes. Annual review of genetics. 5, 237-256.

Thomas, C.A., Paquola, A.C., Muotri, A.R., 2012. LINE-1 retrotransposition in the nervous system. Annual review of cell and developmental biology. 28, 555-573.

Thornburg, B.G., Gotea, V., Makałowski, W., 2006. Transposable elements as a significant source of transcription regulating signals. Gene. 365, 104-110.

Tiersch, T., Goudie, C., 1993. Inheritance and variation of genome size in half-sib families of hybrid catfishes. Journal of Heredity. 84, 122-125.

Tóth, G., Gáspári, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome research. 10, 967-981.

Tubio, J.M., Li, Y., Ju, Y.S., Martincorena, I., Cooke, S.L., Tojo, M., Gundem, G., Pipinikas, C.P., Zamora, J., Raine, K., 2014. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science. 345, 1251343.

Tucker, B.A., Scheetz, T.E., Mullins, R.F., DeLuca, A.P., Hoffmann, J.M., Johnston, R.M., Jacobson, S.G., Sheffield, V.C., Stone, E.M., 2011. Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (MAK) as a cause of retinitis pigmentosa. Proceedings of the National Academy of Sciences. 108, E569-E576.

Ugarković, Đ., Plohl, M., 2002. Variation in satellite DNA profiles—causes and effects. The EMBO journal. 21, 5955-5959.

United States Department of Agriculture, 2014. Census of Aquaculture (2012), Volume 1, Geographic Area Series. Part 51. Washington, DC: United States Department of Agriculture.

Ungerer, M.C., Strakosh, S.C., Zhen, Y., 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Current Biology. 16, R872-R873.

USDA-NASS (United States Department of Agriculture, National Agriculture Statistic Service), 2006. Census of Aquaculture (2005), Volume 3, Special Studies, Part 2. Publication AC-02-SP-2. Washington, DC: United States Department of Agriculture

Ustyugova, S.V., Lebedev, Y.B., Sverdlov, E.D., 2006. Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. Genetica. 128, 261-272.

Vicient, C.M., 2010. Transcriptional activity of transposable elements in maize. BMC genomics. 11, 601.

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., 2014. The locust genome provides insight into swarm formation and long-distance flight. Nature communications. 5, 2957.

Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., Tu, X.-L., Dong, Y., Zhu, C.-L., Wang, L., 2015. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Molecular biology and evolution. 32, 1880-1889.

Weber, J.L., May, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American journal of human genetics. 44, 388.

White, M., Davern, C., 1978. Chromosomal models of speciation. Modes of speciation, 169-276.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics. 8, 973-982.

Wilkinson, D.A., Mager, D.L., Leong, J.-A.C., 1994. Endogenous human retroviruses, In: The retroviridae. Springer, pp. 465-535.

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. Animal genetics. 37, 321-326.

Xu, P., Wang, S., Liu, L., Thorsen, J., Kucuktas, H., Liu, Z., 2007. A BAC-based physical map of the channel catfish genome. Genomics. 90, 380-388.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., 2014. Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nature genetics. 46, 1212-1219.

Yang, N., Kazazian, H.H., 2006. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nature structural & molecular biology. 13, 763-771.

Yi, S., Streelman, J.T., 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends in Genetics. 21, 643-646.

Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., Machida, Y., 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proceedings of the National Academy of Sciences. 90, 6562-6566.

Yu, L., Wang, G.-D., Ruan, J., Chen, Y.-B., Yang, C.-P., Cao, X., Wu, H., Liu, Y.-H., Du, Z.-L., Wang, X.-P., 2016. Genomic analysis of snub-nosed monkeys (Rhinopithecus) identifies genes and processes related to high-altitude adaptation. Nature genetics. 48, 947-952.

Chapter 2. The Annotation of Repetitive Elements in Channel Catfish

2.1 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 transposable elements 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.

2.2. Background

Eukaryotic genomes contain significant amounts of repetitive DNA sequences, and the collective of the repeated sequences in an organism is known as the repeatome of the organism (Maumus and Quesneville, 2014). Such non-protein coding repetitive sequences were once thought to be junk DNA (Ohno 1972), but recent studies have indicated that they play important roles in propelling genome evolution and adaptation to environments (Thomas 1971; Meagher and Vassiliadis 2005; Schmidt and Anderson 2006; Thornburg et al., 2006; Sun et al., 2015; Wang et al., 2014; Yu et al., 2016). 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 (Maumus and Quesneville, 2014). Transposable elements are dispersed across genomes and their proportion are highly variable among genomes, ranging from 3% to 85% in terms of physical size (SanMiguel et al., 1996; Wicker et al., 2007). For instance, the genome of

Utricularia gibba contains only 3% of repetitive sequences (Ibarra-Laclette et al., 2013; Lee et al., 2014), while the genome of maize contains over 85% transposable elements (Schnable et al., 2009; Vicient, 2010). Based on their mechanisms of proliferation, transposable elements can be further classified into RNA-mediated Class I transposable elements and RNA-independent Class II DNA transposable elements. Class I transposable elements contain three main subclasses: short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and transposable elements with long terminal repeats (LTRs). Transposable elements are capable of moving in the genome, and, therefore, they are believed to be a major driving force for genome evolution (Hurst and Schilthuizen 1998; Hurst and Werren 2001; Kazazian 1999; Kazazian 2004).

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 (Kubis et al., 1998; Tóth et al., 2000; Ugarković and Plohl 2002). 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 (Hacch and Mazrimas

1974; Petitpierre et al., 1995). The amplification and/or mutations of tandem repeats may also affect the genome by changing the genome structures or genome sizes (Charlesworth et al., 1994; Lindahl 1994; Strand et al., 1993), thereby affecting recombination of genomes, gene expression, gene conversion, and chromosomal organization (Balaresque et al., 2014; Martin et al., 2005; Moxon et al., 1994; Pardue et al., 1987; Richard et al., 2000).

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*) (Ingle et al., 1975). In *Saccharomyces cerevisiae*, a single cluster of rRNA can cover about 60% of the chromosome XII (Kobayashi et al., 1998). It has been considered as the "king of the housekeeping genes" in terms of function and quantity (Kobayashi, 2011). 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 (Bengtén et al., 2006; Maumus and Quesneville, 2014; Nagarajan et al., 2008).

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 (USDA-NASS, 2006, United States Department of Agriculture, 2014). Its reference genome sequence with annotation of protein coding genes was published (Liu et al., 2016), but its repeatome was not fully characterized. Previous works reported the presence of an A/T-rich tandem Xba elements on the centromere regions (Liu et al., 1998, Quiniou, 2005), the presence of dispersed SINE elements (Kim et al., 2000) and DNA transposable elements (Liu et al., 1999). Genomic sequencing surveys provided additional information for the discovery of other repetitive sequences types such as microsatellites and transposable elements (Chen et al., 2016; Liu et al., 2009; Nandi et al., 2007; Serapion et al., 2004; Xu et al., 2006; Xu et al., 2007). Here, the repeatome of the channel catfish genome has been annotated and characterized from the sequences generated for whole genome sequencing

2.3. Materials and Methods

2.3.1 Annotation of repetitive elements in channel catfish genome

The identification and annotation of the repetitive elements in the channel catfish genome were conducted using the RepeatModeler 1.0.8 package (http://www.repeatmasker.org/RepeatModeler.html) containing RECON (Bao and Eddy 2002) and RepeatScout (Price et al., 2005). The identified channel catfish repetitive sequences were searched against curated libraries and repetitive DNA sequence database such as Repbase (Bao et al., 2015) and Dfam (Wheeler et al., 2013) derived from RepeatMasker package (http://www.repeatmasker.org/). To further determine characters of the repetitive elements classified as "Unknown" by the RepeatModeler, they were first clustered by self-alignments via BLASTCLUST and CD-HIT (Huang et al, 2010; Li and Godzik, 2006). Then, all the "Unknown" sequences were searched against the NCBI Nucleotide collection database (nt) using blastn: 2.2.28+ with a relatively relaxed E-value ($<10^{-5}$).

2.3.2 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 (Deng et al., 2014).

2.3.3 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% (Yang et al., 2012). In addition to channel catfish and blue catfish, the sequences of blind cave fish (*Astyanax mexicanus*), common carp (*Cyprinus carpio*), and zebrafish (*Danio rerio*) were also used for phylogenetic analysis. The analysis was performed using the BEAST v.1.8.0 package

(Drummond et al., 2012). Two independent runs were performed with 1,000 generations sampled from every 10 million generations for each dataset using MCMC chains (Drummond et al., 2002). The input files were constructed in BEAUTi, and the best substitution model was selected by Prottest 3.2.1 according to the alignments (Darriba et al., 2011). Model parameters consisted of a GTR+I+G model with a log normal relaxed clock (Drummond et al., 2006; Drummond and Suchard, 2010), the speciation birth-death process, and random starting tree that was also applied in the phylogenetic analysis.

For the files of the resulting trees, the TreeAnnotator v1.8.0 was used to discard 10% of samples as burn-in and the information of the remaining samples of trees summarized onto a maximum clade credibility chronogram. 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, three teleost fossil records were selected for calibration, and the following node ages were set using lognormal 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 (Blanton et al., 2013; Hardman and Hardman, 2008).

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 (Wang et al., 2014; Werner et al., 1994).

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 (Benton, 1993).

2.3.4 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 (Chinwalla et al., 2002), 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 $r=K/(2T)$ (Li and

Graur; 1991), 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.

2.4. Results

2.4.1 Annotation of repetitive elements in the channel catfish genome

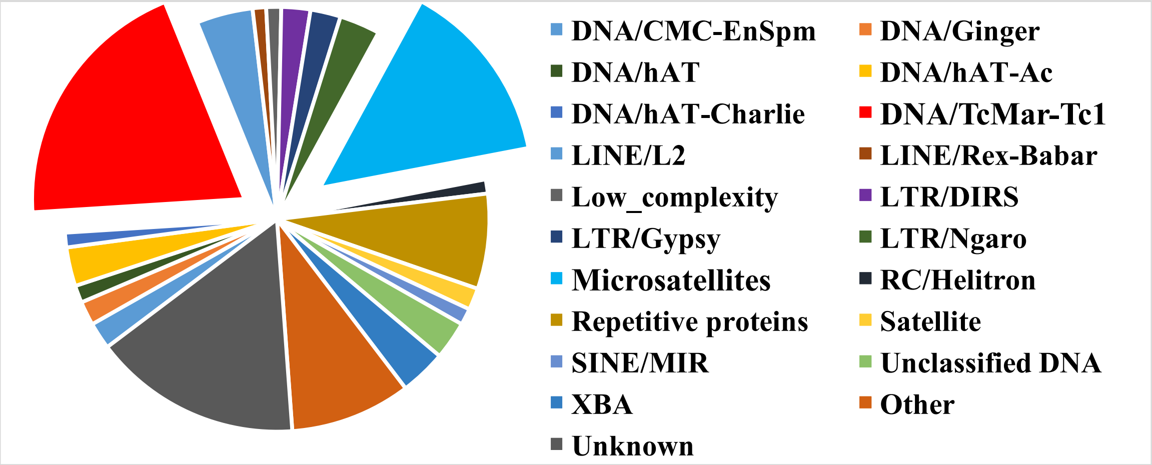
The major categories of the repetitive elements in the channel catfish genome are shown in Figure 2-1 and detailed in Table 2-1. The channel catfish genome harbored a total of 417.8 Mb of repetitive elements, accounting for 44% of the catfish genome. Of all the repetitive elements, 84.1% were annotated 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 transposable elements accounting for the largest percentage (19.9%), followed by microsatellites (14.1%), repetitive proteins (7.2%), LINE/L2 (4.3%), Xba elements (3.6%), LTR/Nagro (3.1%), hAT/Ac (3.0%), unclassified DNA transposable elements (2.9%), LTR/Gypsy (2.3%), LTR/DIRS (2.2%), CMC-EnSpm (2.1%), Ginger (1.8%), satellite (1.7%), hAT (1.3%), SINE/MIR (1.3%), low complexity elements (1.1%), DNA/hAT-Charlie (1.1%), RC/Helitron (1.1%), and LINE/Rex-Babar (1.0%). All the remaining categories represented less than 1% each of the repetitive elements (Table 2-1).

Table 2-1. A list of the major categories of repetitive elements in channel catfish and their percentage in the total repeatome.

Categories	Length	Percentage
DNA/CMC-EnSpm	8,588,035	2.06%
DNA/Crypton	242,175	0.06%
DNA/Crypton-V	78,594	0.02%
DNA/Ginger	7,637,015	1.83%
DNA/hAT	5,487,708	1.31%
DNA/hAT-Ac	12,360,979	2.96%
DNA/hAT-Blackjack	897,878	0.21%
DNA/hAT-Charlie	4,668,464	1.12%
DNA/hAT-hAT5	312,477	0.07%
DNA/hAT-hobo	171,935	0.04%
DNA/hAT-Tip100	872,668	0.21%
DNA/hAT-Tol2	1,158,482	0.28%
DNA/IS3EU	827,786	0.20%
DNA/Kolobok	245,734	0.06%
DNA/Kolobok-T2	825,932	0.20%
DNA/Maverick	2,710,651	0.65%
DNA/MULE-MuDR	346,485	0.08%
DNA/MuLE-NOF	71,642	0.02%
DNA/Novosib	133,898	0.03%
DNA/P	267,664	0.06%
DNA/PIF-Harbinger	3,848,539	0.92%
DNA/PIF-ISL2EU	358,050	0.09%
DNA/PiggyBac	639,874	0.15%
DNA/Sola	91,729	0.02%
DNA/TcMar	535,998	0.13%
DNA/TcMar-ISRm11	85,918	0.02%

DNA/TcMar-Mariner	90,702	0.02%
DNA/TcMar-Tc1	83,037,626	19.88%
DNA/TcMar-Tc2	149,381	0.04%
DNA/TcMar-Tigger	950,714	0.23%
DNA/Zisupton	238,823	0.06%
UnclassifiedDNA	12,145,751	2.91%
LINE/I	281,092	0.07%
LINE/I-Nimb	355,326	0.09%
LINE/L1	1,937,508	0.46%
LINE/L1-Tx1	2,595,401	0.62%
LINE/L2	17,933,406	4.29%
LINE/Penelope	422,511	0.10%
LINE/R1	1,362,877	0.33%
LINE/R2-Hero	131,887	0.03%
LINE/Rex-Babar	4,283,938	1.03%
LINE/RTE-BovB	785,854	0.19%
LINE/RTE-X	127,422	0.03%
UnclassifiedLINE	111,578	0.03%
Low_complexity	4,746,556	1.14%
LTR/Copia	1,425,541	0.34%
LTR/DIRS	9,233,420	2.21%
LTR/ERV	224,039	0.05%
LTR/ERV1	3,025,892	0.72%
LTR/ERVK	65,915	0.02%
LTR/Gypsy	9,627,395	2.30%
LTR/Ngaro	12,888,999	3.09%
LTR/Pao	51,682	0.01%
LTR/Viper	125,356	0.03%
UnclassifiedLTR	508,483	0.12%
RC/Helitron	4,491,267	1.07%
Retroposon	230,172	0.06%
rRNA	1,469,993	0.35%
Satellite	6,945,285	1.66%
Microsatellites	58,745,860	14.06%
SINE/5S	1,193,623	0.29%
SINE/5S-Deu-L2	1,004,623	0.24%
SINE/Alu	495,527	0.12%
SINE/MIR	5,245,061	1.26%
SINE/tRNA	125,260	0.03%
SINE/tRNA-L2	194,202	0.05%
SINE/tRNA-V	1,944,266	0.47%
SINE/tRNA-V-CR1	431,196	0.10%
SINE/U	133,552	0.03%
SINE?	1,427,942	0.34%
snRNA	81,984	0.02%
XBA	14,700,253	3.52%
Repetitivegenes	30,274,984	7.25%
Unknown	66,328,071	15.88%

Figure 2-1. The proportion of major categories (>1% of the total repeatome) of repetitive elements in the channel catfish genome.



2.4.2 The distribution of repeats cross genome

The Tc1/Mariner transposable elements are distributed cross the whole genome, with no major differences among chromosomes or among chromosomal regions within chromosomes (Figure 2-2). 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 (Figure 2-3.), 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 tandem repeats (Kamnert et al., 1997, Melters et al., 2013; Witzany, G., 2008; Wicky et al., 1996).

Figure 2-2. The distribution of Tc1/Mariner transposable elements 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 transposable elements in the chromosomal regions. Each color bar represented a physical distance of 1 Mb DNA.

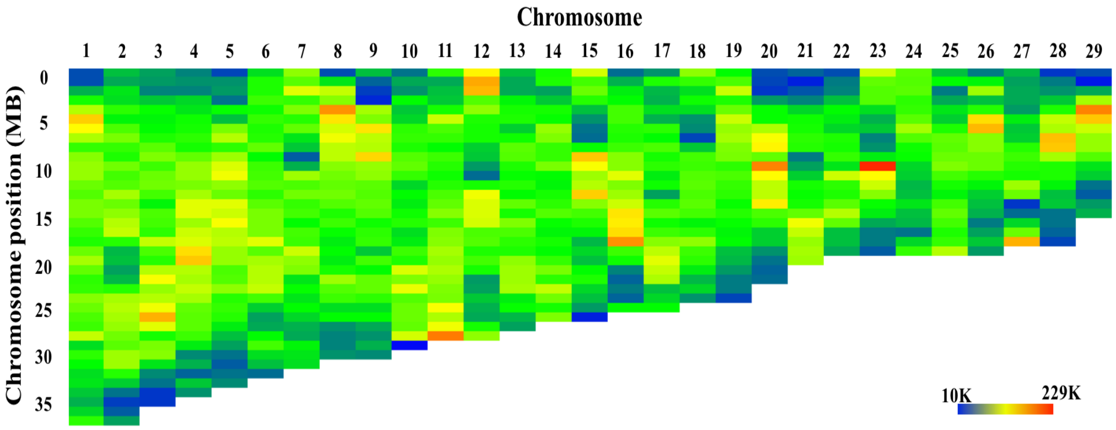
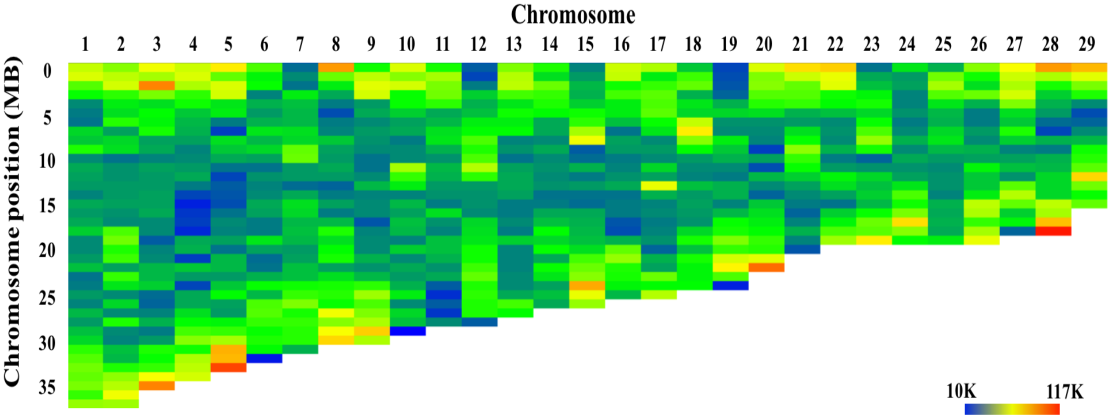


Figure 2-3. The distribution of microsatellites and satellites along the chromosomes of the channel catfish genome. Color key is indicated at the lower right of the figure, with blue color to indicate low and red color to indicate high levels of the transposable elements in the chromosomal regions. Each color bar represented a physical distance of 1 Mb DNA.



2.4.3 Substitution rates

The analysis of evolutionary rate of the unique Xba elements within catfish and the most abundant DNA/TcMar-Tc1 transposable elements is useful to assess their limitations of evolution, providing assessment of their potential functions. The divergence analysis indicated that the Xba elements have a low average number of substitutions per site of 3.53, lower than the average number of substitutions per site of 13.34 of the channel catfish reaptome. Meanwhile, compared with Xba elements, the substitution distribution of the catfish DNA/TcMar-Tc1 transposable elements, 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 average divergence rate of approximately 12% (Figure 2-4). This indicated a long history of evolution as well as a more active evolutionary dynamics during the evolution of DNA/TcMar-Tc1 transposable elements 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 (Figure 2-5). 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 (Blanton et al., 2013). 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 (Chen et al., 2010), the rate of nucleotide substitution of Xba elements are higher than the sequences 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}).

Figure 2-4. The divergence distribution of channel catfish Xba elements (blue) and DNA/TcMar-Tc1 transposable elements (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 (%).

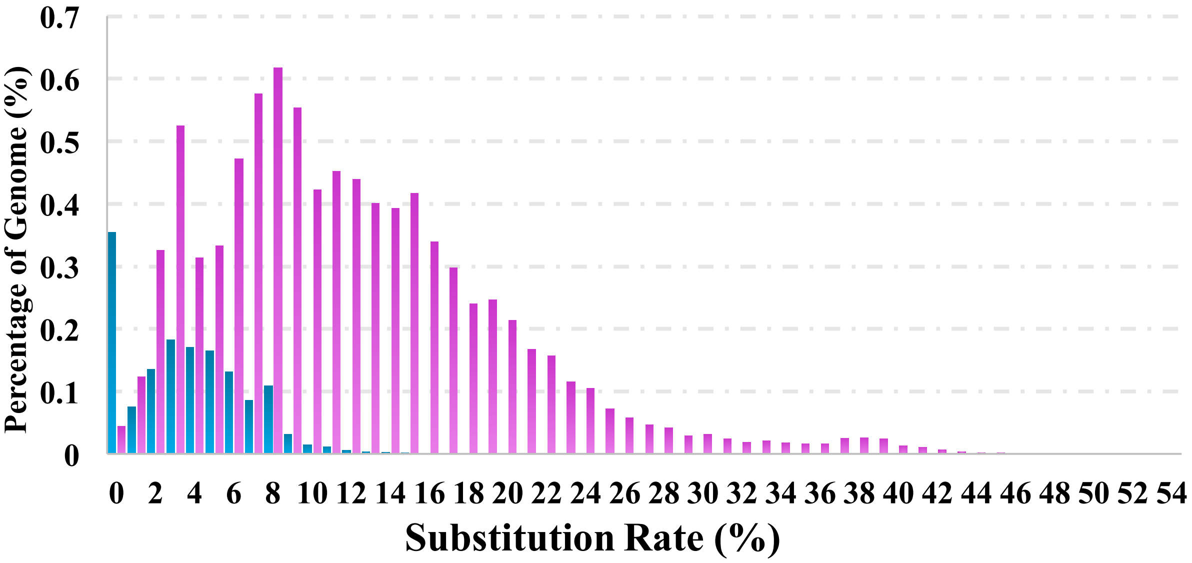
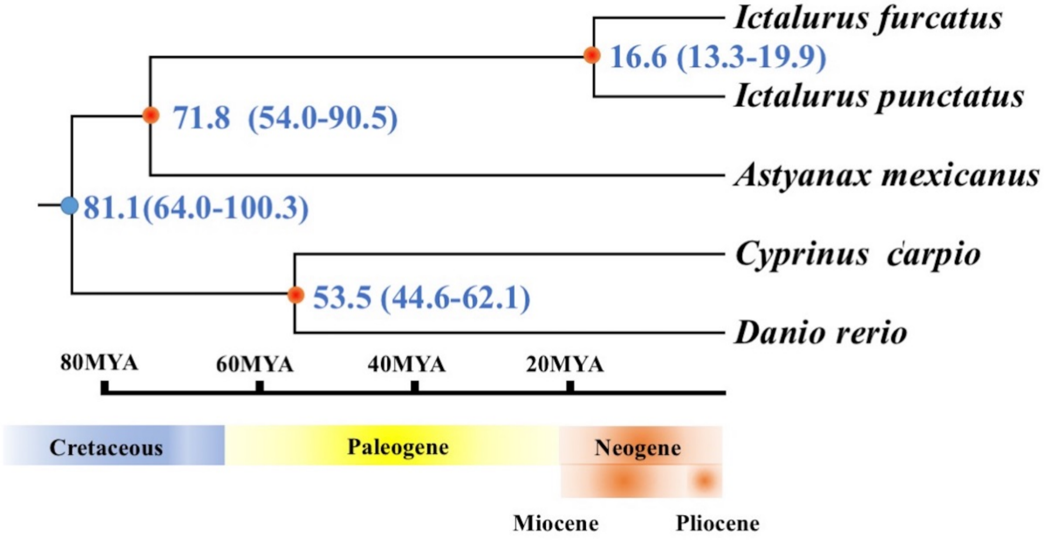


Figure 2-5. The divergence time of the channel catfish and blue catfish with fossil calibrations (orange nodes) based on the mitochondrial cytochrome b sequences.



2.4.4. 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 (Table 2-2). The top categories with over 500 Kb in length and their representative sequences on the genome are listed in Table 2-3. 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. This work provides a brief classification of those repetitive elements. 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.

Table 2-2. A list of the clustering of novel repetitive elements in channel catfish, ranked by the number of contained sequences

Number of Sequences	Representative Sequence		
	scaffold	**scf_start**	**scf_end**
15,687	IpCoco_scf00610	3,078	6,525
10,668	IpCoco_scf00474	141,399	143,674
9,962	IpCoco_scf00517	9,296,133	9,298,172
9,372	IpCoco_scf00172	16,372,640	16,373,425
9,256	IpCoco_scf00369	2,537,431	2,538,007
7,789	IpCoco_scf00203	3,637,116	3,638,889
7,711	IpCoco_scf00563_1077_652_655	1,427,889	1,431,117
7,314	IpCoco_scf00540	509	4,548
6,816	IpCoco_scf00419	7,439,807	7,440,803
6,593	IpCoco_scf00021	1,307,023	1,308,497
6,040	IpCoco_scf00789	5,607	11,243
5,982	IpCoco_scf00570_502_500	2,136,706	2,137,523
5,665	IpCoco_scf05654_567	71,525	74,021
5,432	IpCoco_scf00563_1077_652_655	1,372,133	1,373,248
5,414	IpCoco_scf00077_78	260,739	262,006
5,385	IpCoco_scf00396	143,979	146,295
5,263	IpCoco_scf00077_78	1,199,900	1,201,056
5,180	IpCoco_scf00799	9,869	11,749
4,997	IpCoco_scf00354_353	643,250	644,319
4,764	IpCoco_scf00504	608,431	609,640
4,736	IpCoco_scf00002_4_6	2,140,474	2,141,692
4,735	IpCoco_scf00161_160	5,433,996	5,434,743
4,538	IpCoco_scf00136_7095_137_139	9,991,070	9,991,859

4,535	IpCoco_scf00141_142_968_5056_143	1,351,661	1,352,594
4,510	IpCoco_scf00204	1,383,666	1,384,283
4,477	IpCoco_scf00096	575,684	576,725
4,446	IpCoco_scf00172	2,251,598	2,252,801
4,355	IpCoco_scf00249	2,654,914	2,655,810
4,279	IpCoco_scf00340_341_342	19,062,478	19,063,434
4,138	IpCoco_scf00172	8,598,037	8,598,997
4,102	IpCoco_scf05231	1	1,093
4,079	IpCoco_scf00496_498_497	7,933,106	7,934,163
4,054	IpCoco_scf02085	433	1,678
3,968	IpCoco_scf00369	2,792,428	2,793,279
3,927	IpCoco_scf00077_78	1,077,726	1,078,788
3,895	IpCoco_scf00019_718	6,408,693	6,409,508
3,885	IpCoco_scf00366	478,563	479,717
3,881	IpCoco_scf00340_341_342	10,530,823	10,531,924
3,825	IpCoco_scf00040	8,026,450	8,027,394
3,771	IpCoco_scf00489_490_491	4,089,377	4,090,058
3,759	IpCoco_scf00198_201	2,500,027	2,501,092
3,709	IpCoco_scf00161_160	2,491,966	2,492,910
3,667	IpCoco_scf00391_390	3,109,908	3,110,627
3,614	IpCoco_scf00010	4,363,749	4,364,706
3,606	IpCoco_scf00010	2,599,200	2,600,332
3,562	IpCoco_scf00274_275	7,889,580	7,890,361
3,554	IpCoco_scf00448	89,120	90,036
3,482	IpCoco_scf00141_142_968_5056_143	11,727,384	11,728,311
3,474	IpCoco_scf00096	309,930	310,766
3,472	IpCoco_scf00136_7095_137_139	3,075,579	3,076,208
3,446	IpCoco_scf00635_446	238,998	239,646
3,435	IpCoco_scf00327_328	13,320,235	13,321,149
3,435	IpCoco_scf00340_341_342	15,848,812	15,849,806
3,414	IpCoco_scf00334	1,635,586	1,636,149
3,413	IpCoco_scf00204	185,132	185,729
3,391	IpCoco_scf00099_100	346,026	346,988
3,389	IpCoco_scf00489_490_491	7,584,469	7,585,338
3,370	IpCoco_scf00198_201	3,312,261	3,313,178
3,370	IpCoco_scf00458_460	9,455,461	9,456,157
3,369	IpCoco_scf00008	446,573	447,360
3,332	IpCoco_scf00021	4,198,429	4,199,326
3,322	IpCoco_scf00086	143,561	144,604
3,319	IpCoco_scf00562_252	1,374,434	1,375,410
3,289	IpCoco_scf00317	360,705	361,648
3,266	IpCoco_scf00393_392	3,100,399	3,101,253
3,262	IpCoco_scf00633	18,326	18,860
3,259	IpCoco_scf00274_275	2,867,477	2,868,704
3,219	IpCoco_scf00503	210,495	211,229
3,213	IpCoco_scf00071	1,633,412	1,634,325
3,197	IpCoco_scf00389	3,414,455	3,415,293
3,170	IpCoco_scf00278	422,055	422,964
3,157	IpCoco_scf04438	40	822
3,133	IpCoco_scf00516	1,491,132	1,491,961

3,099	IpCoco_scf00036	2,572	3,537
3,032	IpCoco_scf00148	1,685,841	1,686,671
3,022	IpCoco_scf00496_498_497	2,415,958	2,416,937
3,022	IpCoco_scf00041	1,528,807	1,530,113
3,010	IpCoco_scf00136_7095_137_139	7,830,924	7,831,811
3,008	IpCoco_scf00525	3,198,193	3,199,630
2,951	IpCoco_scf00525	6,200,193	6,200,986
2,950	IpCoco_scf00327_328	2,324,626	2,325,657
2,919	IpCoco_scf00042	1,116,339	1,117,093
2,917	IpCoco_scf00286_7269	2,926,236	2,927,230
2,868	IpCoco_scf00327_328	11,461,354	11,462,034
2,859	IpCoco_scf00389	8,469,223	8,469,861
2,856	IpCoco_scf00437	978,114	978,867
2,849	IpCoco_scf04837	1	755
2,795	IpCoco_scf00009	4,749,269	4,749,978
2,760	IpCoco_scf00114_115	11,405,239	11,405,911
2,751	IpCoco_scf00550_548	1,101,252	1,102,226
2,733	IpCoco_scf00036	568,560	569,426
2,699	IpCoco_scf00011	9,958,281	9,959,135
2,677	IpCoco_scf00389	6,727,127	6,727,877
2,673	IpCoco_scf00036	6,166	7,059
2,670	IpCoco_scf00585	161,644	162,548
2,641	IpCoco_scf00396	630,241	631,059
2,603	IpCoco_scf00489_490_491	186,285	186,949
2,596	IpCoco_scf00108_980	109,118	110,342
2,585	IpCoco_scf00399	1,527,144	1,528,051
2,583	IpCoco_scf05905	1	699
2,574	IpCoco_scf00517	7,105,932	7,106,758
2,568	IpCoco_scf00300_298_299	9,453,069	9,453,877
2,561	IpCoco_scf00437	4,041,640	4,042,395
2,554	IpCoco_scf00389	8,199,498	8,200,369
2,547	IpCoco_scf00249	9,209,951	9,210,757
2,546	IpCoco_scf00574	143,785	144,426
2,538	IpCoco_scf00246	477,926	478,409
2,517	IpCoco_scf00534	1,361,886	1,363,113
2,500	IpCoco_scf00220	75,142	76,079
2,491	IpCoco_scf00581_386	4,887,047	4,887,716
2,482	IpCoco_scf00068_10963_69	2,015,920	2,016,548
2,468	IpCoco_scf00232	8,205,457	8,206,098
2,462	IpCoco_scf00936	9,673	10,464
2,410	IpCoco_scf00092	2,673,153	2,674,146
2,406	IpCoco_scf00021	5,350,196	5,351,010
2,387	IpCoco_scf03832	621	1,104
2,344	IpCoco_scf00550_548	1,159,155	1,159,846
2,293	IpCoco_scf00496_498_497	10,977,345	10,977,986
2,280	IpCoco_scf00136_7095_137_139	3,486,353	3,487,182
2,255	IpCoco_scf00040	5,274,383	5,275,071
2,249	IpCoco_scf00246	2,486,986	2,487,786
2,247	IpCoco_scf00375	469,803	470,567
2,239	IpCoco_scf00369	1,440,700	1,441,442

2,217	IpCoco_scf00200	6,979,352	6,980,065
2,205	IpCoco_scf00136_7095_137_139	5,679,024	5,680,155
2,187	IpCoco_scf00428	7,162	7,938
2,179	IpCoco_scf00609	147,467	148,139
2,137	IpCoco_scf00389	3,057,922	3,058,552
2,034	IpCoco_scf00391_390	3,173,448	3,173,715
2,017	IpCoco_scf00366	1,994,548	1,995,380
2,010	IpCoco_scf00200	11,537,699	11,538,622
2,006	IpCoco_scf00369	9,305,743	9,306,545
1,998	IpCoco_scf00315	1,172,781	1,173,466
1,989	IpCoco_scf00171	1,308,407	1,309,107
1,970	IpCoco_scf00141_142_968_5056_143	9,809,096	9,809,691
1,963	IpCoco_scf00313_314	714,174	714,810
1,962	IpCoco_scf00161_160	9,603,209	9,603,910
1,956	IpCoco_scf00001	2,158,589	2,159,301
1,955	IpCoco_scf00449	83,930	84,750
1,949	IpCoco_scf00322	3,909,356	3,910,382
1,922	IpCoco_scf04734	463	1,128
1,909	IpCoco_scf00311_312	2,885,636	2,886,341
1,889	IpCoco_scf00327_328	8,102,843	8,103,477
1,884	IpCoco_scf00181_186	40,421	41,176
1,861	IpCoco_scf00530_1311	162,525	163,199
1,852	IpCoco_scf00437	6,867,161	6,867,864
1,814	IpCoco_scf00327_328	13,947,769	13,948,538
1,811	IpCoco_scf00550_548	2,870,570	2,871,331
1,791	IpCoco_scf00340_341_342	15,931,004	15,931,648
1,790	IpCoco_scf00525	11,756,306	11,756,809
1,787	IpCoco_scf00171	1,219,854	1,220,499
1,780	IpCoco_scf00771	25,166	25,765
1,762	IpCoco_scf00200	14,190,348	14,191,003
1,751	IpCoco_scf00096	814,481	814,993
1,720	IpCoco_scf00136_7095_137_139	1,965,895	1,966,445
1,678	IpCoco_scf00246	1,935,174	1,935,902
1,670	IpCoco_scf00256	62,349	62,902
1,662	IpCoco_scf00172	14,988,719	14,989,469
1,649	IpCoco_scf00300_298_299	5,560,993	5,561,596
1,645	IpCoco_scf00517	10,128,725	10,129,417
1,640	IpCoco_scf00019_718	2,159,479	2,160,116
1,634	IpCoco_scf00525	8,578,223	8,578,983
1,626	IpCoco_scf00300_298_299	8,566,043	8,566,671
1,561	IpCoco_scf00562_252	400,566	401,138
1,556	IpCoco_scf00019_718	1,917,470	1,918,033
1,526	IpCoco_scf00428	437,421	438,095
1,526	IpCoco_scf00041	1,051,141	1,051,745
1,467	IpCoco_scf00300_298_299	2,631,644	2,632,265
1,467	IpCoco_scf00286_7269	2,067,689	2,068,092
1,454	IpCoco_scf00352	1,929,317	1,929,879
1,452	IpCoco_scf00274_275	12,438,671	12,439,359
1,451	IpCoco_scf00419	5,751,012	5,751,689
1,440	IpCoco_scf00420	12,159	12,798

1,405	IpCoco_scf00581_386	211,372	212,113
1,389	IpCoco_scf00496_498_497	6,569,791	6,570,391
1,387	IpCoco_scf00114_115	11,583,522	11,584,112
1,384	IpCoco_scf00455	3,789,094	3,789,670
1,372	IpCoco_scf00011	28,281	28,982
1,368	IpCoco_scf00393_392	2,768,677	2,769,258
1,364	IpCoco_scf00151_2157_666_736	243,129	243,558
1,344	IpCoco_scf00385	4,986,958	4,987,744
1,335	IpCoco_scf00053	190,276	191,076
1,335	IpCoco_scf00249	15,180,700	15,181,426
1,326	IpCoco_scf00055	1,653,385	1,653,957
1,302	IpCoco_scf00279	274,418	274,926
1,248	IpCoco_scf00558	1,555,028	1,555,659
1,239	IpCoco_scf00327_328	8,413,457	8,414,076
1,219	IpCoco_scf00489_490_491	15,768,967	15,769,447
1,219	IpCoco_scf00340_341_342	20,303,697	20,304,321
1,216	IpCoco_scf04073	744	1,236
1,200	IpCoco_scf00169	21,562	22,023
1,192	IpCoco_scf00076	613,682	614,296
1,182	IpCoco_scf00396	1,112,297	1,112,962
1,181	IpCoco_scf00159	3,809,170	3,809,865
1,172	IpCoco_scf00340_341_342	16,633,255	16,633,750
1,130	IpCoco_scf00249	15,626,023	15,626,577
1,104	IpCoco_scf00136_7095_137_139	695,027	697,319
1,097	IpCoco_scf00385	3,752,059	3,752,742
1,067	IpCoco_scf00169	44,619	45,147
1,053	IpCoco_scf00517	4,189,110	4,189,611
1,031	IpCoco_scf00114_115	8,376,101	8,376,678
997	IpCoco_scf00008	315,955	316,492
981	IpCoco_scf00826	15,287	15,773
918	IpCoco_scf00203	2,217,944	2,218,386
914	IpCoco_scf00369	7,486,442	7,486,954
885	IpCoco_scf00279	699,557	700,099
875	IpCoco_scf01014	1	479
848	IpCoco_scf00114_115	8,292,985	8,293,540
812	IpCoco_scf00011	7,410,043	7,410,515
779	IpCoco_scf00102	20,267	20,625
767	IpCoco_scf00233_234_235	293,553	293,962
747	IpCoco_scf00389	1,861,045	1,861,731
684	IpCoco_scf00204	317,790	318,339
645	IpCoco_scf00370	3,751,623	3,752,048
574	IpCoco_scf00711	8,014	8,428
136	IpCoco_scf00009	3,058,070	3,058,197

Table 2-3. The major novel repetitive elements and their characteristics in the channel catfish repeatome.

Size	AT content	Representative Sequence		
		scaffold	scf _start	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

2.5. Discussion

2.5.1 Repetitive elements in channel catfish

Using repetitive element libraries combined with the nucleotide (nt) database, I 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 transposable elements 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 the previous observations through genome sequence surveys (Liu et al., 2009; Xu et al., 2006), but the data were analyzed from the whole genome and therefore is more complete.

The DNA/TcMar-Tc1 transposable element sequences make up the highest percentage among Class II transposable elements in channel catfish genome, accounting for about ~20% of the total repetitive elements and interspersed on the genome. The DNA-TcMAr/Tc1 is a typical "cut-paste" transposable element (Finnegan 1985), which is prevalent in nature and can be transferred not only vertically but also horizontally cross species during evolution (Aziz et al., 2010). It is this character that allows DNA-TcMAr/Tc1 transposable elements to escape from the vertical extinction and being so abundant in nature (Lawrence and Hartl 1992; Lohe et al., 1995; Maruyama and Hartl 1991). 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 transposable element footprints in channel catfish genome may indicate an external origin of DNA/TcMar-Tc1 invasion to the genome during evolution. As "parasitic" mobile elements, DNA transposable elements are known to be potent sources of mutation, and the long-time shrinking of effective population sizes can contribute to the evolution of more complex genomes such as more mobile elements or larger genome sizes (Lynch and Conery, 2003, Yi and Streelman, 2005). It is believed that the large amount of mobile transposable elements 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 (González et al., 2008; González et al., 2009). Previous studies indicated that the transposition by a member of the Tc1/mariner family of transposable elements appears to have integrated in the duplicated Cμ region of the immunoglobulin (Ventura-Holman and Lobb, 2002). Channel catfish is a quite hardy fish species that can survive in a wide range of environmental conditions (Wellborn 1988). It is also possible that the prevalent of DNA/TcMar-Tc1 sequences, as well as other transposable elements in channel catfish genomes, play important roles in their adaptation to environments. There are no significant differences on the density of DNA/TcMar-Tc1 on each individual chromosome, with no hot-spot of transposition.

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 and ubiquitous in nearly all eukaryotic genomes (Tautz and Renz 1984; Tautz et al., 1985; Weber 1990), the expansion of microsatellites is disputable but it is generally considered to be expanded through DNA polymerase slippage (Levinson and Gutman 1987a; Levinson and Gutman 1987b; Schlötterer and Tautz 1992). High content of microsatellites in catfish genomes indicates a high level of DNA polymerase slippage, probably related to the high magnesium concentration (meq/L) in the channel catfish tissue (Chen et al., 2003; Wedemeyer, 1996). It was speculated that the magnesium concentration can contribute to DNA polymerase slippage by stabilizing the hairpin structure (Castillo-Lizardo et al., 2014). 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 a number of proteins involved in various DNA replication processes (Bachtrog et al., 1999; Castoe et al., 2011; Cordaux and Batzer. 2009; Glenn et al., 1996; Grady et al., 1992; Janes et al., 2010; Mellon et al., 1996; Melters et al., 2013; Primmer et al., 1997). 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 (Chang et al., 2001; Rocha et al., 2002). This analysis of the distribution of the microsatellites indicates that they are were mostly present on the telomere and the centromere regions of the chromosome, consistent with the previous analysis (Areshchenkova and Ganal., 1999; Vidaurreta et al., 2007).

The catfish genome also contains a large fraction of repetitive proteins in the reaptome. The main types of repetitive proteins are related to the adaptive immunology and metabolism as previous analysis indicated (Bengtén et al., 2006). 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 immunoglobulin in the channel catfish genome (Table 2-4). This may suggest that the expansion of the immunoglobulin family in the channel catfish genome can be one of the mechanisms of its defense system against various pathogens.

Table 2-4. The major repetitive protein domains characterized from the channel catfish repeatome.

Rank	Length	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

2.5.2 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 (*Ameiurus catus*) and flathead catfish (*Pylodictus olivaris*) (Liu et al., 1999; Quiniou et al., 2005). 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 (Liu et al., 1998; Quiniou et al., 2005). 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 (Mehta et al., 2010; Wang et al., 2009). The large amount of conservative Xba elements on centromere suggests those sequences may be involved in centromere functions, such as kinetochore assembly and chromosome segregation during mitosis or meiosis (Pidoux and Allshire., 2005; Westhorpe and Straight., 2013), or even some epigenetic regulations (Karpen and Allshire., 1997).

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, the rate of nucleotide substitutions of Xba elements is higher than that for the sequences in the open reading frames, but lower than that of those in untranslated regions. Slower rates of evolution suggest functional constraints

(Grauer & Li, 2000). The relatively slow evolutionary rate of Xba elements may indicate their potential functions, although unknown at present.

2.6. Conclusion

In this study, I 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 transposable elements, 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, and it should facilitate the annotation of various teleost genomes.

References

Areshchenkova, T., Ganal, M.W., 1999. Long tomato microsatellites are predominantly associated with centromeric regions. Genome. 42, 536-544.

Aziz, R.K., Breitbart, M., Edwards, R.A., 2010. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic acids research. 38, 4207-4217.

Bachtrog, D., Weiss, S., Zangerl, B., Brem, G., Schlötterer, C., 1999. Distribution of dinucleotide microsatellites in the Drosophila melanogaster genome. Molecular Biology and Evolution. 16, 602-610.

Balaresque, P., King, T.E., Parkin, E.J., Heyer, E., Carvalho- Silva, D., Kraaijenbrink, T., Knijff, P., Tyler- Smith, C., Jobling, M.A., 2014. Gene Conversion Violates the Stepwise Mutation Model for Microsatellites in Y- Chromosomal Palindromic Repeats. Human mutation. 35, 609-617.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome research. 12, 1269-1276.

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 6, 11.

Bengtén, E., Clem, L.W., Miller, N.W., Warr, G.W., Wilson, M., 2006. Channel catfish immunoglobulins: repertoire and expression. Developmental & Comparative Immunology. 30, 77-92.

Benton M.J., 1993. The fossil record 2: London. Chapman and Hallp. 629.

Blanton, R.E., Page, L.M., Hilber, S.A., 2013. Timing of clade divergence and discordant estimates of genetic and morphological diversity in the slender madtom, Noturus exilis (Ictaluridae). Molecular phylogenetics and evolution. 66, 679-693.

Castillo-Lizardo, M., Henneke, G., Viguera, E., 2014. Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota Pyrococcus abyssi. Frontiers in microbiology. 5.

Castoe, T.A., Hall, K.T., Mboulas, G., Marcel, L., Gu, W., de Koning, A., Fox, S.E., Poole, A.W., Vemulapalli, V., Daza, J.M., 2011. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. Genome biology and evolution. 3, 641-653.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 371, 215-220.

Chen, F., Lee, Y., Jiang, Y., Wang, S., Peatman, E., Abernathy, J., Liu, H., Liu, S., Kucuktas, H., Ke, C., 2010. Identification and characterization of full-length cDNAs in channel catfish (Ictalurus punctatus) and blue catfish (*Ictalurus furcatus*). PLoS One. 5, e11546.

Chen, X., Zhong, L., Bian, C., Xu, P., Qiu, Y., You, X., Zhang, S., Huang, Y., Li, J., Wang, M., 2016. High-quality genome assembly of channel catfish, Ictalurus punctatus. GigaScience. 5, 39.

Chang, D.K., Metzgar, D., Wills, C., Boland, C.R., 2001. Microsatellites in the eukaryotic DNA mismatch repair genes as modulators of evolutionary mutation rate. Genome Research. 11, 1145-1146.

Chen, C.-Y., Wooster, G.A., Getchell, R.G., Bowser, P.R., Timmons, M.B., 2003. Blood chemistry of healthy, nephrocalcinosis-affected and ozone-treated tilapia in a recirculation system, with application of discriminant analysis. Aquaculture. 218, 89-102.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420, 520-562.

Cordaux, R., Batzer, M.A., 2009. The impact of retrotransposons on human genome evolution. Nature Reviews Genetics. 10, 691-703.

Danilova, N., Bussmann, J., Jekosch, K., Steiner, L.A., 2005. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nature immunology. 6, 295-302.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 27, 1164-1165.

Deng, W., Wang, Y., Liu, Z., Cheng, H., Xue, Y., 2014. HemI: a toolkit for illustrating heatmaps. PloS one. 9, e111988.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS biology. 4, e88.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics. 161, 1307-1320.

Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. BMC biology. 8, 114.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 29, 1969-1973.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nature reviews genetics. 5, 435-445.

Finnegan, D., 1985. Transposable elements in eukaryotes. International review of cytology. 93, 281-326.

Glenn, T.C., Stephan, W., Dessauer, H.C., Braun, M.J., 1996. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. Molecular Biology and Evolution. 13, 1151-1154.

González, J., Lenkov, K., Lipatov, M., Macpherson, J.M., Petrov, D.A., 2008. High rate of recent transposable element–induced adaptation in Drosophila melanogaster. PLoS biology. 6, e251.

González, J., Macpherson, J.M., Petrov, D.A., 2009. A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila melanogaster. Molecular biology and evolution. 26, 1949-1961.

Grady, D.L., Ratliff, R.L., Robinson, D.L., McCanlies, E.C., Meyne, J., Moyzis, R.K., 1992. Highly conserved repetitive DNA sequences are present at human centromeres. Proceedings of the National Academy of Sciences. 89, 1695-1699.

Li, W., Graur, D., Sunderland, M., 1992. Fundamental of Molecular Evolution. Sunderland: Sinauer Associates, Inc. 67-98.

Gao, S., Liu, S., Yao, J., Li, N., Yuan, Z., Zhou, T., Li, Q., Liu, Z., 2017. Genomic organization and evolution of olfactory receptors and trace amine-associated receptors in channel catfish, Ictalurus punctatus. Biochimica et Biophysica Acta (BBA)-General Subjects. 1861, 644-651.

Hacch, F., Mazrimas, J., 1974. Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys ordii*). Nucleic acids research. 1, 559-576.

Hardman, M., Hardman, L.M., 2008. The relative importance of body size and paleoclimatic change as explanatory variables influencing lineage diversification rate: an evolutionary analysis of bullhead catfishes (Siluriformes: Ictaluridae). Systematic Biology. 57, 116-130.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 26, 680-682.

Hurst, G.D., Schilthuizen, M., 1998. Selfish genetic elements and speciation. Heredity. 80, 2-8.

Hurst, G.D., Werren, J.H., 2001. The role of selfish genetic elements in eukaryotic evolution. Nature Reviews Genetics. 2, 597-606.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.-H., Lan, T., Welch, A.J., Juárez, M.J.A., Simpson, J., 2013. Architecture and evolution of a minute plant genome. Nature. 498, 94-98.

Ingle, J., Timmis, J.N., Sinclair, J., 1975. The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. Plant physiology. 55, 496-501.

Kazazian, H.H., 1999. An estimated frequency of endogenous insertional mutations in humans.

Nature genetics. 22, 130-130.

Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. science. 303, 1626-1632.

Kobayashi, T., Heck, D.J., Nomura, M., Horiuchi, T., 1998. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. Genes & development. 12, 3821-3830.

Kobayashi, T., 2011. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. Cellular and molecular life sciences. 68, 1395-1403.

Janes, D.E., Organ, C.L., Fujita, M.K., Shedlock, A.M., Edwards, S.V., 2010. Genome evolution in Reptilia, the sister group of mammals. Annual review of genomics and human genetics. 11, 239-264.

Kamnert, I., López, C.C., Rosén, M., Edström, J.E., 1997. Telomeres terminating with long complex tandem repeats. Hereditas. 127, 175-180.

Karpen, G.H., Allshire, R.C., 1997. The case for epigenetic effects on centromere identity and function. Trends in Genetics. 13, 489-496.

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

Kubis, S., Schmidt, T., Heslop-Harrison, J.S., 1998. Repetitive DNA elements as a major component of plant genomes. Annals of Botany. 82, 45-55.

Lawrence, J.G., Hartl, D., 1992. Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. Genetics. 131, 753-760.

Lee, S.-I., Kim, N.-S., 2014. Transposable elements and genome size variations in plants. Genomics & informatics. 12, 87-97.

Levinson, G., Gutman, G.A., 1987. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Research. 15, 5323-5338.

Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular biology and evolution. 4, 203-221.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22, 1658-1659.

Lindahl, T., 1994. DNA Repair: DNA surveillance defect in cancer cells. Current Biology. 4, 249-251.

Liu, Z., Li, P., Dunham, R.A., 1998. Characterization of an A/T-rich family of sequences from channel catfish (Ictalurus punctatus). Molecular marine biology and biotechnology. 7, 232-239.

Liu, Z., Li, P., Kucuktas, H., Dunham, R., 1999. Characterization of nonautonomous Tc1-like transposable elements of channel catfish (Ictalurus punctatus). Fish Physiology and Biochemistry. 21, 65-72.

Liu, H., Jiang, Y., Wang, S., Ninwichian, P., Somridhivej, B., Xu, P., Abernathy, J., Kucuktas, H., Liu, Z., 2009. Comparative analysis of catfish BAC end sequences with the zebrafish genome. BMC genomics. 10, 592.

Liu, Z., 2008. Catfish. In: Thomas, K.D., Chittaranjan, K. (Eds.), Genome Mapping and Genomics in Fishes and Aquatic Animals Volume 2. Berlin Heidelberg. 85–100.

Liu, Z., 2011. Development of genomic resources in support of sequencing, assembly, and annotation of the catfish genome. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics. 6, 11-17.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., Jiang, C., Sun, L., Wang, R., Zhang, Y., 2016. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nature communications. 7: 11757.

Lohe, A.R., Moriyama, E.N., Lidholm, D.-A., Hartl, D.L., 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Molecular biology and evolution. 12, 62-72.

Lynch, M., Conery, J.S., 2003. The origins of genome complexity. science. 302, 1401-1404.

Martin, P., Makepeace, K., Hill, S.A., Hood, D.W., Moxon, E.R., 2005. Microsatellite instability regulates transcription factor binding and gene expression. Proceedings of the National Academy of Sciences of the United States of America. 102, 3800-3804.

Maruyama, K., Hartl, D.L., 1991. Evidence for interspecific transfer of the transposable element mariner betweenDrosophila andZaprionus. Journal of molecular evolution. 33, 514-524.

Maumus, F., Quesneville, H., 2014. Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS One. 9, e94101.

Meagher, T.R., Vassiliadis, C., 2005. Phenotypic impacts of repetitive DNA in flowering plants. New Phytologist. 168, 71-80.

Mehta, G.D., Agarwal, M.P., Ghosh, S.K., 2010. Centromere identity: a challenge to be faced. Molecular genetics and genomics. 284, 75-94.

Mellon, I., Rajpal, D.K., Koi, M., Boland, C.R., Champe, G.N., 1996. Transcription-coupled repair deficiency and mutations in human mismatch repair genes. SCIENCE-NEW YORK THEN WASHINGTON-, 557-559.

Melters, D.P., Bradnam, K.R., Young, H.A., Telis, N., May, M.R., Ruby, J.G., Sebra, R., Peluso, P., Eid, J., Rank, D., 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome biology. 14, R10.

Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current biology. 4, 24-33.

Nagarajan, N., Navajas-Pérez, R., Pop, M., Alam, M., Ming, R., Paterson, A.H., Salzberg, S.L.,

2008. Genome-wide analysis of repetitive elements in papaya. Tropical Plant Biology. 1, 191-201.

Nandi, S., Peatman, E., Xu, P., Wang, S., Li, P., Liu, Z., 2007. Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (Ictalurus punctatus). Genetica. 131, 81-90.

Ohno, S., 1972. So much" junk" DNA in our genome, Brookhaven symposia in biology, pp. 366-370.

Pardue, M., Lowenhaupt, K., Rich, A., Nordheim, A., 1987. (dC-dA) n.(dG-dT) n sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. The EMBO journal. 6, 1781-1789.

Petitpierre, E., Juan, C., Pons, J., Plohl, M., Ugarkovic, D., 1995. Satellite DNA and constitutive heterochromatin in tenebrionid beetles, Kew Chromosome Conference IV. Royal Botanic Gardens, London, pp. 351-362.

Pidoux, A.L., Allshire, R.C., 2005. The role of heterochromatin in centromere function. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 360, 569-579.

Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. Bioinformatics. 21, i351-i358.

Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Møller, A.P., Ellegren, H., 1997. Low frequency of microsatellites in the avian genome. Genome Research. 7, 471-482.

Quiniou, S., Wolters, W., Waldbieser, G., 2005. Localization of Xba repetitive elements to channel catfish (Ictalurus punctatus) centromeres via fluorescence in situ hybridization. Animal genetics. 36, 353-354.

Richard, G.F., Pâques, F., 2000. Mini- and microsatellite expansions: the recombination connection. EMBO reports. 1, 122-126.

Rocha, E.P., Matic, I., Taddei, F., 2002. Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? Nucleic Acids Research. 30, 1886-1894.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science. 274, 765-768.

Schlötterer, C., Tautz, D., 1992. Slippage synthesis of simple sequence DNA. Nucleic acids research. 20, 211-215.

Schmidt, A.L., Anderson, L.M., 2006. Repetitive DNA elements as mediators of genomic change in response to environmental cues. Biological Reviews. 81, 531-543.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., 2009. The B73 maize genome: complexity, diversity, and dynamics. science. 326, 1112-1115.

Serapion, J., Kucuktas, H., Feng, J., Liu, Z., 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (Ictalurus punctatus). Marine biotechnology. 6, 364-377.

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D., 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 365: 274-276.

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L., Wang, L., Wu, D.-D., Zhang, B.-L., 2015. Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences. 112, E1257-E1262.

Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic acids research. 12, 4127-4138.

Tautz, D., Trick, M., Dover, G.A., 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature. 322, 652-656.

Thomas Jr, C.A., 1971. The genetic organization of chromosomes. Annual review of genetics. 5, 237-256.

Thornburg, B.G., Gotea, V., Makałowski, W., 2006. Transposable elements as a significant source of transcription regulating signals. Gene. 365, 104-110.

Tóth, G., Gáspári, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome research. 10, 967-981.

United States Department of Agriculture, 2014. Census of Aquaculture (2012), Volume 1, Geographic Area Series. Part 51. Washington, DC: United States Department of

Agriculture.

Ugarković, Đ., Plohl, M., 2002. Variation in satellite DNA profiles—causes and effects. The EMBO journal. 21, 5955-5959.

USDA-NASS (United States Department of Agriculture, National Agriculture Statistic Service), 2006. Census of Aquaculture (2005), Volume 3, Special Studies, Part 2. Publication AC-02-SP-2. Washington, DC: United States Department of Agriculture

Ventura-Holman, T., Lobb, C.J., 2002. Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters. Molecular immunology. 38, 557-564.

Vicient, C.M., 2010. Transcriptional activity of transposable elements in maize. BMC genomics. 11, 601.

Vidaurreta, M., Maestro, M., Rafael, S., Veganzones, S., Sanz-Casla, M., Cerdán, J., Arroyo, M., 2007. Telomerase activity in colorectal cancer, prognostic factor and implications in the microsatellite instability pathway. World Journal of Gastroenterology: WJG. 13, 3868.

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., 2014.

The locust genome provides insight into swarm formation and long-distance flight. Nature communications. 5, 2957.

Wang, G., Zhang, X., Jin, W., 2009. An overview of plant centromeres. Journal of Genetics and Genomics. 36, 529-537.

Wang, Q., Xu, C., Xu, C., Wang, R., 2014. Complete mitochondrial genome of the Southern catfish (Silurus meridionalis Chen) and Chinese catfish (S. asotus Linnaeus): Structure, phylogeny, and intraspecific variation. Genetics and Molecular Research. 14, 18198-18209.

Wedemeyer, G., 1996. Physiology of fish in intensive culture systems. Springer Science & Business Media. p. 54

Weber, J.L., 1990. Informativeness of human (dC-dA) n·(dG-dT) n polymorphisms. Genomics. 7, 524-530.

Wellborn, T.L., 1988. Channel catfish: life history and biology. SRAC publication (USA). No.180.

Werner, C., 1994. Die kontinentale Wirbeltierfauna aus der unteren Oberkreide des Sudan (Wadi

Milk Formation). Berliner Geowissenschaftliche Abhandlungen E (B. Krebs-Festschrift). 13, 221-249.

Westhorpe, F.G., Straight, A.F., 2013. Functions of the centromere and kinetochore in chromosome segregation. Current opinion in cell biology. 25, 334-340.

Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F., Finn, R.D., 2012. Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic acids research. 41, D70-D82.

Wicky, C., Villeneuve, A.M., Lauper, N., Codourey, L., Tobler, H., Müller, F., 1996. Telomeric repeats (TTAGGC) n are sufficient for chromosome capping function in Caenorhabditis elegans. Proceedings of the National Academy of Sciences. 93, 8983-8988.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics. 8, 973-982.

Witzany, G., 2008. The viral origins of telomeres and telomerases and their important role in eukaryogenesis and genome maintenance. Biosemiotics. 1, 191-206.

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. Animal genetics. 37, 321-326.

Xu, P., Wang, S., Liu, L., Thorsen, J., Kucuktas, H., Liu, Z., 2007. A BAC-based physical map of the channel catfish genome. Genomics. 90, 380-388.

Yang, J.-Q., Tang, W.-Q., Liao, T.-Y., Sun, Y., Zhou, Z.-C., Han, C.-C., Liu, D., Lin, H.-D., 2012. Phylogeographical analysis on Squalidus argentatus recapitulates historical landscapes and drainage evolution on the island of Taiwan and mainland China. International journal of molecular sciences. 13, 1405-1425.

Yi, S., Streelman, J.T., 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends in Genetics. 21, 643-646.

Yu, L., Wang, G.-D., Ruan, J., Chen, Y.-B., Yang, C.-P., Cao, X., Wu, H., Liu, Y.-H., Du, Z.-L., Wang, X.-P., 2016. Genomic analysis of snub-nosed monkeys (Rhinopithecus) identifies genes and processes related to high-altitude adaptation. Nature genetics. 48, 947-952.

Chapter 3. The Comparative Analysis of Repetitive Elements Through the Phylogentic Clades

3.1 Abstract:

Repetitive elements make up significant proportions of genomes. However, their roles in evolution remain largely unknown. To provide insights into the roles of repetitive elements in fish genomes, I conducted a comparative analysis of repetitive elements of 52 fish species in 22 orders. The proportions of repetitive elements in various genomes are found to be positively correlated with genome sizes, with a few exceptions. More importantly, there appears to be specific enrichment between some repetitive element categories with species habitats. Specifically, class II transposons appears to be more abundant in freshwater bony fish than in marine bony fish when the phylogeny is not considered. In contrast, marine bony fish harbor more tandem repeats than freshwater species. In addition, class I transposable elements appear to be more abundant in primitive species such as cartilaginous fish and lamprey than in bony fish. The enriched association of specific categories of repetitive elements with fish habitats suggests the importance of repetitive elements in genome evolution and their potential roles in fish

adaptation to their living environments. However, due to the restriction of the limited sequenced species, further analysis need to be done to alleviate the phylogeny biases.

3.2 Background

The majority of eukaryotic genomes contain a large proportion of repetitive elements. Based on their arrangements in the genome, repetitive elements can be divided into two major categories: the transposable elements (transposable elements) and the tandem repeats. Transposable elements can be divided into RNA-mediated class I transposable elements, which include transposable elements with long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs); and RNA-independent class II DNA transposable elements. Tandem repeats are copies of DNA repeats located adjacent to one other (Kubis et al., 1998; Tóth et al., 2000; Ugarković and Ploh, 2002). Tandem repeats themselves can be dispersed across the whole genome such as the case of microsatellites, and they can be clustered in the highly repetitive genome regions such as centromeric, telomeric and subtelomeric regions (Hacch et al., 1974; Petitpierre et al., 1995).

Although repetitive elements were considered to be junk DNA (Ohno, 1972), recent studies suggested that they are functional in regulating gene expressions and contribute to genome evolution (Meagher and Vassiliadis , 2005; Schmidt and Anderson, 2006; Sun et al., 2015; Thornburg et al., 2006; Wang et al., 2014). Transposable elements are considered to be drivers of genetic diversification because of their ability to co-opt into genetic processes such as restructuring the chromosomes or providing genetic material on which natural selection can act on (Hurst and Werren, 2001; Kazazian, 1999; Kazazian, 2004), and thus can be the major reason for species difference in genome size (Lee and Kim, 2014; SanMiguel et al., 1996; Wicker et al., 2007). Similarly, expansion or contraction of tandem repeats can also affect genome size (Charlesworth et al., 1994; Lindahl, 1994; Strand et al., 1993), and consequently affect recombination, gene expressions, conversions and chromosomal organizations (Balaresque et al., 2014; Hancock, 1996; Martin et al., 2005; Moxon et al., 1994; Pardue et al., 1987; Richard and Pâques et al., 2000)

Fish comprise a large and highly diverse group of vertebrates inhabiting a range of different aquatic environments (Volff et al., 2005). Sequenced fish genomes vary in size from 342 Mb of *Tetraodon nigroviridis* to 2,967 Mb of *Salmo salar*. Some studies have been conducted on the diversity of repetitive elements in fish (Chalopin et al., 2015; Chalopin et al., 2015; Gao et al., 2016), but a systematic comparative analysis has been hindered by the lack of whole genome sequences from a large number of species. Recent availability of a large number of fish genome sequences made it possible to determine the repetitive element profiles of fish species from a broad taxonomic spectrum. In this study, I annotated the repetitive elements of 52 fish genomes from 22 orders, and determined their distribution in relationship with environmental adaptations. Based on the phylogeny tree, the effects of phylogeny on the differences between freshwater or marine bony fish were evaluated with the phylogenetically independent contrasts. Here, I observed a correlation between high numbers of DNA transposable elements, especially the Tc1 transposable elements, with freshwater bony fish, high level of microsatellites with marine bony fish, and high numbers of class I transposable elements with cartilaginous fish and lamprey.

3.3 Materials and Methods

3.3.1 Annotation of repetitive elements in fish genome assemblies

The channel catfish genome was assembled by our group (Liu et al., 2016), the genome sequences of other 51 species were retrieved from NCBI or Ensembl databases (AlMomin et al., 2015; Amemiya et al., 2013; Aparicio et al., 2002; Barrio et al., 2016; Braasch et al., 2016; Brawand et al., 2014; Burns et al., 2016; Chen et al., 2014; Conte and Kocher, 2015; Fraser et al., 2015; Gao et al., 2014; Jones et al., 2012; Kasahara et al., 2007; Lien et al., 2016; McGaugh et al., 2014; Nakamura et al., 2013; Read et al., 2015; Rondeau et al., 2014; Schartl et al., 2013; Shin et al., 2014; Smith et al., 2013; Smolka et al., 2015; Star et al., 2011; Tine et al., 2014; Venkatesh et al., 2014; Watson et al., 2009; Wu et al., 2014; Xu et al., 2014; Xu et al., 2016; Yang et al., 2016) (Table 3-1). The repetitive elements were identified using RepeatModeler 1.0.8 containing RECON (Bao and Eddy, 2002) and RepeatScout with default parameters (Price et al., 2005). The derived repetitive sequences were searched against Dfam (Wheeler et al., 2013) and Repbase (Bao et al., 2015). If the sequence is classified as "Unknown", they were further searched against the NCBI-nt database using blastn 2.2.28+.

Table 3-1. Fish genomes used for analysis, with their habitats indicated by color, the pink indicates freshwater habitats, the blue indicates marine habitats and the yellow indicates the diadromous habitats.

Order	Species	Habitats	Reference OR Accession
Cyprinodontiformes	*Poecilia Formosa*	Freshwater	GCA_000485575.1
Cyprinodontiformes	*Poecilia Mexicana*	Freshwater	GCA_001443325.1
Cyprinodontiformes	*Poecilia latipinna*	Freshwater	GCA_001443285.1
Cyprinodontiformes	*Poecilia reticulata*	Freshwater	GCA_000633615.2 (Fraser et al., 2015)
Cyprinodontiformes	*Xiphophorus couchianus*	Freshwater	GCA_001444195.1 (Schartl et al., 2013)
Cyprinodontiformes	*Xiphophorus hellerii*	Freshwater	GCA_001443345.1 (Schartl et al., 2013)
Cyprinodontiformes	*Xiphophorus maculatus*	Freshwater	GCA_000241075.1 (Schartl et al., 2013)
Cyprinodontiformes	*Fundulus heteroclitus*	Freshwater	GCA_000826765.1
Cyprinodontiformes	*Cyprinodon variegatus*	Freshwater	GCA_000732505.1
Cyprinodontiformes	*Cyprinodon nevadensis*	Freshwater	GCA_000776015.1
Beloniformes	*Oryzias latipes*	Freshwater	MEDAKA1 (Ugarković and Plohl., 2002)
Perciformes	*Amphilophus citrinellus*	Freshwater	GCA_000751415.1
Perciformes	*Neolamprologus brichardi*	Freshwater	GCA_000239395.1 (Brawand et al., 2014)
Perciformes	*Pundamilia nyererei*	Freshwater	GCA_000239375.1 (Brawand et al., 2014)
Perciformes	*Haplochromis burtoni*	Freshwater	GCA_000239415.1 (Brawand et al., 2014)
Perciformes	*Oreochromis niloticus*	Freshwater	GCA_000188235.1 (Brawand et al., 2014)
Perciformes	*Maylandia zebra*	Freshwater	GCA_000238955.3 (Conte and Kocher, 2015)
Siluriformes	*Ictalurus punctatus*	Freshwater	(Liu et al., 2016)
Characiformes	*Astyanax mexicanus*	Freshwater	GCA_000372685.1 (McGaugh et al., 2014)
Clupeiformes	*Clupea harengus*	Marine	GCA_000966335.1 (Barrio et al., 2016)
Perciformes	*Notothenia coriiceps*	Marine	GCA_000735185.1 (Shin et al., 2014)
Perciformes	*Dicentrarchus labrax*	Marine	GCA_000689215.1 (Tine et al., 2014)
Gasterosteiformes	*Gasterosteus aculeatus*	Marine	BROAD S1 (Jones et al., 2012)
Scorpaeniformes	*Cottus rhenanus*	Freshwater	GCA_001455555.1 (Smolka et al., 2015)
Scorpaeniformes	*Sebastes nigrocinctus*	Marine	GCA_000475235.1
Scorpaeniformes	*Sebastes rubrivinctus*	Marine	GCA_000475215.1
Perciformes	*Pampus argenteus*	Marine	GCA_000697985.1 (AlMomin et al., 2015)
Perciformes	*Thunnus orientalis*	Marine	GCA_000418415.1 (Nakamura et al., 2013)
Perciformes	*Larimichthys crocea*	Marine	GCA_000742935.1 (Wu et al., 2014)
Perciformes	*Miichthys miiuy*	Marine	GCA_001593715.1 (Xu et al., 2016)
Pleuronectiformes	*Cynoglossus semilaevis*	Marine	GCA_000523025.1 (Chen et al., 2014)
Tetraodontiformes	*Tetraodon nigroviridis*	Freshwater	TETRAODON 8.0 (Jaillon et al., 2004)
Tetraodontiformes	*Takifugu rubripes*	Marine	FUGU 4.0 (Aparicio et al., 2002)
Tetraodontiformes	*Takifugu flavidus*	Marine	GCA_000400755.1 (Gao et al., 2014)
Salmoniformes	*Salmo salar*	Diadromous	GCA_000233375.4 (Lien et al., 2016)
Esociformes	*Esox lucius*	Freshwater	GCA_000721915.2 (Rondeau et al., 2014)
Cypriniformes	*Pimephales promelas*	Freshwater	GCA_000700825.1 (Burns et al., 2016)
Cypriniformes	*Sinocyclocheilus rhinocerous*	Freshwater	GCA_001515625.1 (Yang et al., 2016)
Cypriniformes	*Sinocyclocheilus grahami*	Freshwater	GCA_001515645.1 (Yang et al., 2016)
Cypriniformes	*Sinocyclocheilus anshuiensis*	Freshwater	GCA_001515605.1 (Yang et al., 2016)
Cypriniformes	*Cyprinus carpio*	Freshwater	GCA_000951615.1 (Xu et al., 2014)
Cypriniformes	*Danio rerio*	Freshwater	GRCz10
Gadiformes	*Gadus morhua*	Marine	GCA_000231765.1 (Star et al., 2011)
Anguilliformes	*Anguilla anguilla*	Diadromous	GCA_000695075.1
Anguilliformes	*Anguilla rostrata*	Diadromous	GCA_001606085.1
Lepisosteiformes	*Lepisosteus oculatus*	Freshwater	GCA_000242695.1 (Braasch et al., 2016)
Coelacanthiformes	*Latimeria chalumnae*	Marine	GCA_000225785.1 (Amemiya et al., 2013)
Rajiformes	*Leucoraja erinacea*	Marine	GCA_000238235.1
Orectolobiformes	*Rhincodon typus*	Marine	GCA_001642345.1 (Read et al., 2015)
Chimaeriformes	*Callorhinchus milii*	Marine	GCA_000165045.2 (Venkatesh et al., 2014)
Petromyzontiforme	*Lethenteron camtschaticum*	Freshwater	GCA_000466285.1
Petromyzontiformes	*Petromyzon marinus*	Freshwater	Pmarinus_7.0 (Smith et al., 2013)

3.3.2 Phylogenetic analysis

The phylogenetic analysis was based on the divergence of cytochrome b (Castresana, 2001). Multiple alignments were conducted by MAFFT (Katoh and Standley, 2013). The best substitution model was selected by Prottest 3.2.1 (Darriba et al., 2011). The phylogenetic tree was constructed using MEGA7 with the maximum likelihood method (Kumar et al., 2016), using JTT with Freqs. (+ F) model, and gaps were removed by partial deletion. The topological stability was evaluated with 1,000 bootstraps.

3.3.3 Divergence distribution of DNA/TcMar-Tc1

The average number of substitutions per sites (K) for each DNA/TcMar-Tc1 fragment was subtotaled. The K was calculated based on the Jukes-Cantor formula: $K=-300/4\times \mathrm{Ln}(1-D\times 4/300)$, the D represents the proportion of each DNA/TcMar-Tc1 fragment differ from the consensus sequences (Chinwalla et al., 2002).

3.3.4 Statistics and plotting

The statistical analyses for the significance of differences between different groups and the habitats were performed by Wilcoxon rank test function in R language package because the data are not normally distributed (R Core Team, 2003). The Pearson correlation analysis in Excel was applied for the correlation between genome size and the content of repetitive elements. The heat map was plotted using the Heml1.0 (Deng et al., 2014).

3.4 Results

3.4.1 Contents of repetitive elements in various fish genomes

A total of 128 categories of repetitive elements are identified from the 52 fish species (Appendices and Supplementary tables). The contents of repetitive elements in fish are correlated with their genome sizes, independent of the phylogeny (Figure 3-1, Pearson correlation r=0.6, P-value=1.45e-06, Table 3-2), ranging from less than 10% in pufferfish to over 58% in zebrafish. However, several exceptions existed. For instance, the whale shark genome is 2.57 Gb, but contains only 26.2% of repetitive elements; in contrast, the mid-sized zebrafish genome is ~1.5 Gb in size, but contains over 58% of repetitive elements.

Figure 3-1. Correlation between genome sizes and contents of repetitive elements. Genome sizes against the percentages of repetitive elements to the whole genome are plotted for 52 species of species for which genome sequences are available. The major orders are plotted in different colors and shapes: Yellow circle: Tetraodontiformes; Orange circle: Perciformes circle; Green circle: Scorpaeniformes; Brown circle: Cypriniformes; Red circle: Cyclostomata; Purple circle: Cyprinodontiformes; Blue triangle: Chondrichthyes; Blue circle: Other species.

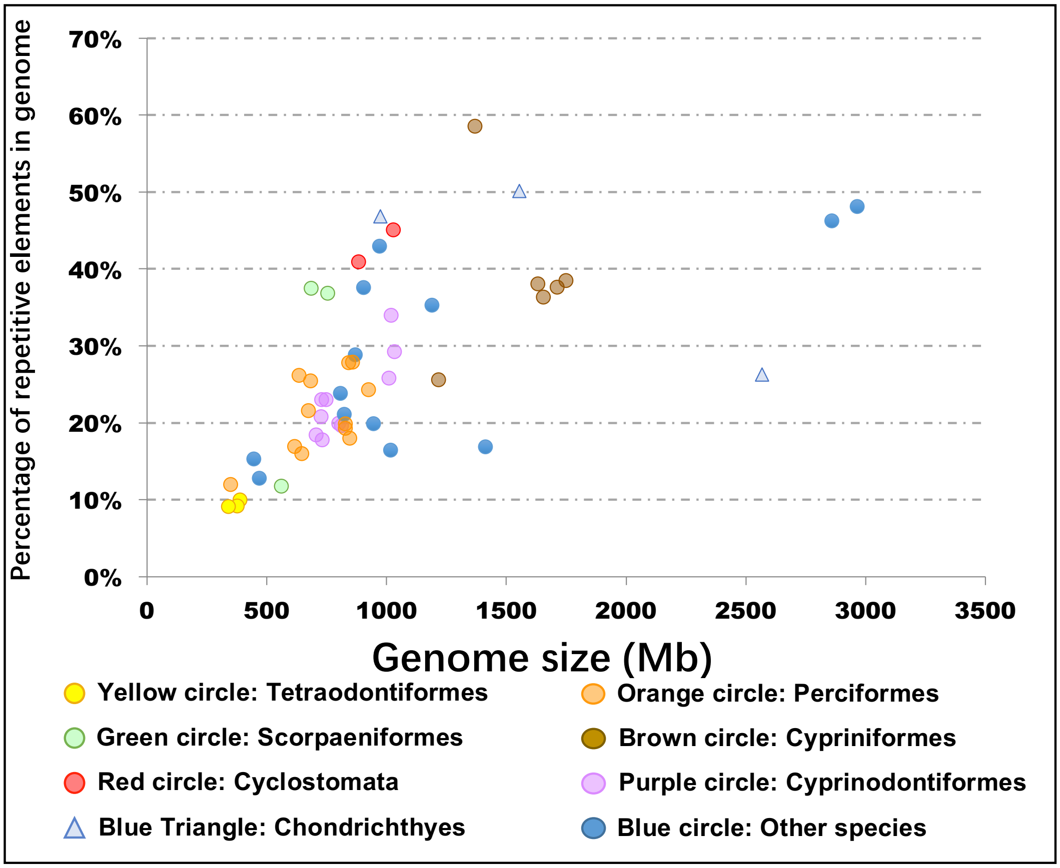


Table 3-2. The relationship between fish genome sizes and the content of repeatomes.

Common Names	Latin names	Genome Size (MB)	Repeatome content (%)
Amazon molly	*Poecilia formosa*	748.9	23.0
Shortfin molly	*Poecilia mexicana*	801.7	19.9
Sailfin molly	*Poecilia latipinna*	815.1	19.5
Guppy	*Poecilia reticulata*	731.6	22.9
Monterrey platyfish	*Xiphophorus couchianus*	708.4	18.4
Green swordtail	*Xiphophorus hellerii*	733.8	17.7
Southern platyfish	*Xiphophorus maculatus*	729.7	20.7
Mummichog	*Fundulus heteroclitus*	1021.9	33.9
Sheepshead minnow	*Cyprinodon variegatus*	1035.2	29.2
Amargosa pupfish	*Cyprinodon nevadensis*	1011.9	25.8
Japanese rice fish	*Oryzias latipes*	869.8	28.8
Midas cichlid	*Amphilophus citrinellus*	844.9	27.8
Princess cichlid	*Neolamprologus brichardi*	847.9	18.0
Flame back cichlid	*Pundamilia nyererei*	830.1	19.8
Burton's mouthbrooder	*Haplochromis burtoni*	831.4	19.2
Nile tilapia	*Oreochromis niloticus*	927.7	24.3
Zebra mbuna	*Maylandia zebra*	859.8	27.8
channel catfish	*Ictalurus punctatus*	942.2	44.2

Mexican tetra	*Astyanax mexicanus*	1191.2	35.2
Atlantic herring	*Clupea harengus*	807.7	23.7
Black rockcod	*Notothenia coriiceps*	636.6	26.1
European seabass	*Dicentrarchus labrax*	675.9	21.5
Three-spined stickleback	*Gasterosteus aculeatus*	446.6	15.3
Chabot de Rhénanie	*Cottus rhenanus*	563.6	11.7
Tiger rockfish	*Sebastes nigrocinctus*	687.6	37.4
Flag rockfish	*Sebastes rubrivinctus*	756.3	36.8
Silver pomfret	*Pampus argenteus*	350.4	11.9
Pacific bluefin tuna	*Thunnus orientalis*	684.5	25.4
Large yellow croaker	*Larimichthys crocea*	648.4	15.9
Mi-iuy croaker	*Miichthys miiuy*	619.3	16.9
Tongue sole	*Cynoglossus semilaevis*	470.2	12.8
Spotted green pufferfish	*Tetraodon nigroviridis*	342.4	9.0
Japanese pufferfish	*Takifugu rubripes*	391.5	9.9
Yellowbelly pufferfish	*Takifugu flavidus*	378.0	9.1
Atlantic salmon	*Salmo salar*	2966.9	47.0
Northern pike	*Esox lucius*	904.5	37.5
Fathead minnow	*Pimephales promelas*	1219.3	25.5
horned Golden-line barbell	*Sinocyclocheilus rhinocerous*	1655.8	36.3
golden-line barbel	*Sinocyclocheilus grahami*	1750.3	38.4
Golden-line barbell (Anshui)	*Sinocyclocheilus anshuiensis*	1632.7	38.0
Common carp	*Cyprinus carpio*	1713.6	37.6
zebrafish	*Danio rerio*	1371.7	58.5
Atlantic cod	*Gadus morhua*	824.3	21.0
European eel	*Anguilla anguilla*	1018.7	16.4
American eel	*Anguilla rostrata*	1413.1	16.8
Spotted gar	*Lepisosteus oculatus*	945.9	19.8
Coelacanth	*Latimeria chalumnae*	2860.6	46.2
Little skate	*Leucoraja erinacea*	1555.5	50.1
Whale shark	*Rhincodon typus*	2567.6	26.2
Ghost shark	*Callorhinchus milii*	974.5	46.8
Arctic lamprey	*Lethenteron camtschaticum*	1030.7	45.0
Sea lamprey	*Petromyzon marinus*	885.5	40.9

In addition to the proportion, the number of repetitive element categories is also correlated with genome sizes. For example, the largest 26 genomes harbor more categories of repetitive elements than the smallest 26 genomes (Wilcoxon rank test, P-value=1.79e-02). Of all these species, DNA transposable elements are the predominant type of repetitive elements, followed by microsatellites.

3.4.2 Differential associations of repetitive elements across species

I investigated the possible correlation between repeat elements and aquatic environment. Comparison of diversity and abundance of repetitive elements across the 52 fish genomes revealed significant differences among species (Figure 3-2 and Table 3-3). Class I transposable elements are more prevalent in cartilaginous fish and lampreys than bony fish species (Wilcoxon rank test, P-value=1.41e-04). For example, class I transposable elements represent 76.6% of repetitive elements in elephant shark, but the bony fish genomes are more abundant with class II transposable elements and tandem repeats.

Figure 3-2. Classification and distribution of 128 repetitive elements in 52 species. The total number of each category of repeats to the all repeats are displayed in columns while different species are displayed in rows. The pink shade represents the freshwater living bony fish, the blue represents the marine living bony fish and yellow represents the diadromous species.

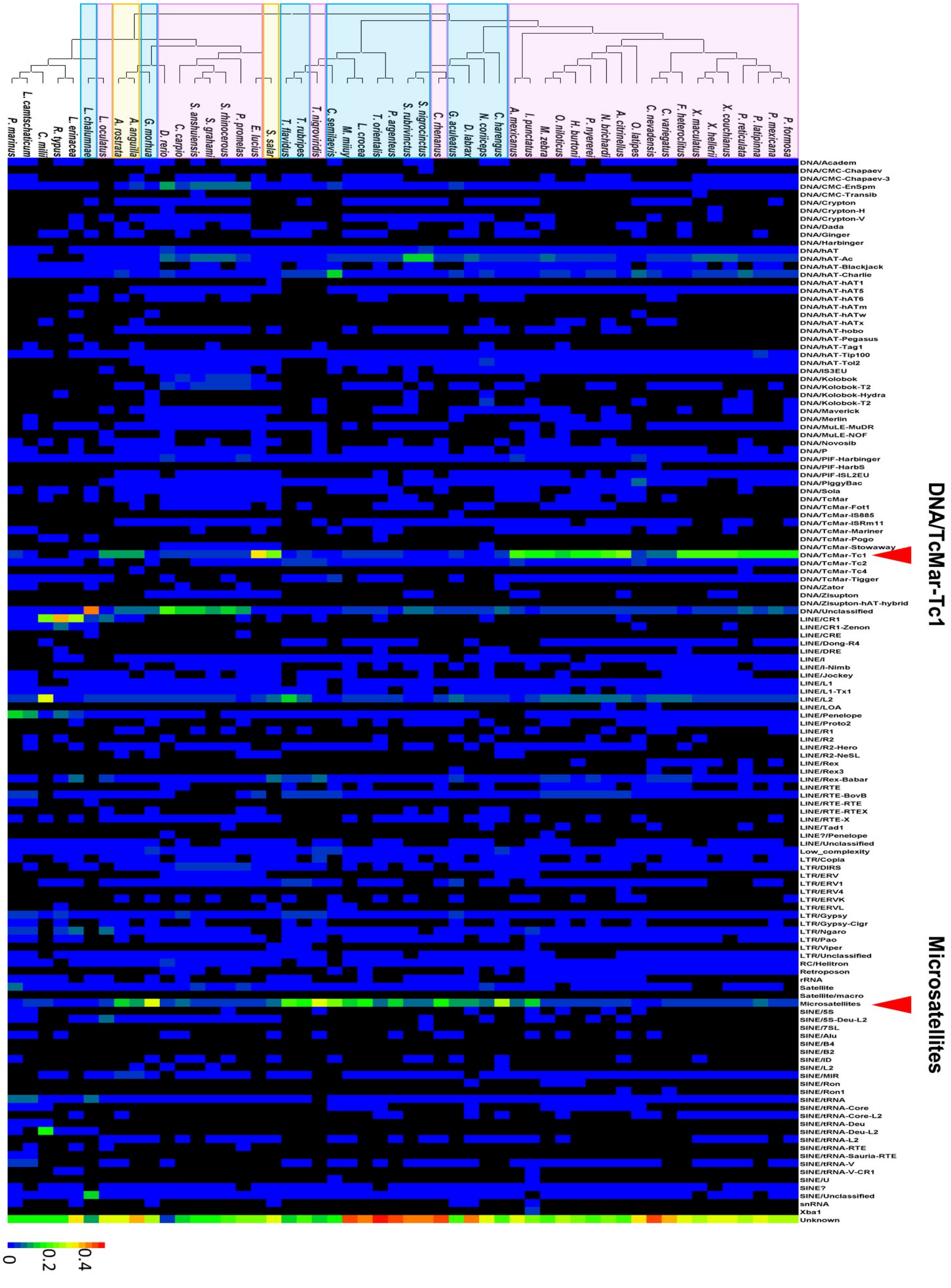

P. marinus
L. camtschaticum
C. milii
R. typus
L. erinacea
L. chalumnae
L. oculatus
A. rostrata
A. anguilla
G. morhua
D. rerio
C. carpio
S. anshuiensis
S. grahami
S. rhinocerous
P. promelas
E. lucius
S. salar
T. flavidus
T. rubripes
T. nigroviridis
C. semilaevis
M. miiuy
L. crocea
T. orientalis
P. argenteus
S. rubrivinctus
S. nigrocinctus
C. rhenanus
G. aculeatus
D. labrax
N. coriiceps
C. harengus
A. mexicanus
I. punctatus
M. zebra
O. niloticus
H. burtoni
P. nyererei
N. brichardi
A. citrinellus
O. latipes
C. nevadensis
C. variegatus
F. heteroclitus
X. maculatus
X. hellerii
X. couchianus
P. reticulata
P. latipinna
P. mexicana
P. formosa
DNA/Academ
DNA/CMC-Chapaev
DNA/CMC-Chapaev-3
DNA/CMC-EnSpm
DNA/CMC-Transib
DNA/Crypton
DNA/Crypton-H
DNA/Crypton-V
DNA/Dada
DNA/Ginger
DNA/Harbinger
DNA/hAT
DNA/hAT-Ac
DNA/hAT-Blackjack
DNA/hAT-Charlie
DNA/hAT-hAT1
DNA/hAT-hAT5
DNA/hAT-hAT6
DNA/hAT-hATm
DNA/hAT-hATw
DNA/hAT-hATx
DNA/hAT-hobo
DNA/hAT-Pegasus
DNA/hAT-Tag1
DNA/hAT-Tip100
DNA/hAT-Tol2
DNA/IS3EU
DNA/Kolobok
DNA/Kolobok-T2
DNA/Kolobok-Hydra
DNA/Kolobok-T2
DNA/Maverick
DNA/Merlin
DNA/MuLE-MuDR
DNA/MuLE-NOF
DNA/Novosib
DNA/P
DNA/PIF-Harbinger
DNA/PIF-HarbS
DNA/PIF-ISL2EU
DNA/PiggyBac
DNA/Sola
DNA/TcMar
DNA/TcMar-Fot1
DNA/TcMar-IS885
DNA/TcMar-ISRm11
DNA/TcMar-Mariner
DNA/TcMar-Pogo
DNA/TcMar-Stowaway
DNA/TcMar-Tc1
DNA/TcMar-Tc2
DNA/TcMar-Tc4
DNA/TcMar-Tigger
DNA/Zator
DNA/Zisupton
DNA/Zisupton-hAT-hybrid
DNA/Unclassified
LINE/CR1
LINE/CR1-Zenon
LINE/CRE
LINE/Dong-R4
LINE/DRE
LINE/I
LINE/I-Nimb
LINE/Jockey
LINE/L1
LINE/L1-Tx1
LINE/L2
LINE/LOA
LINE/Penelope
LINE/Proto2
LINE/R1
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/Rex
LINE/Rex3
LINE/Rex-Babar
LINE/RTE
LINE/RTE-BovB
LINE/RTE-RTE
LINE/RTE-RTEX
LINE/RTE-X
LINE/Tad1
LINE?/Penelope
LINE/Unclassified
Low_complexity
LTR/Copia
LTR/DIRS
LTR/ERV
LTR/ERV1
LTR/ERV4
LTR/ERVK
LTR/ERVL
LTR/Gypsy
LTR/Gypsy-Cigr
LTR/Ngaro
LTR/Pao
LTR/Viper
LTR/Unclassified
RC/Helitron
Retroposon
rRNA
Satellite
Satellite/macro
Microsatellites
SINE/5S
SINE/5S-Deu-L2
SINE/7SL
SINE/Alu
SINE/B4
SINE/B2
SINE/ID
SINE/L2
SINE/MIR
SINE/Ron
SINE/Ron1
SINE/tRNA
SINE/tRNA-Core
SINE/tRNA-Core-L2
SINE/tRNA-Deu
SINE/tRNA-Deu-L2
SINE/tRNA-L2
SINE/tRNA-RTE
SINE/tRNA-Sauria-RTE
SINE/tRNA-V
SINE/tRNA-V-CR1
SINE/U
SINE?
SINE/Unclassified
snRNA
Xba1
Unknown
DNA/TcMar-Tc1
Microsatellites
0
0.2
0.4

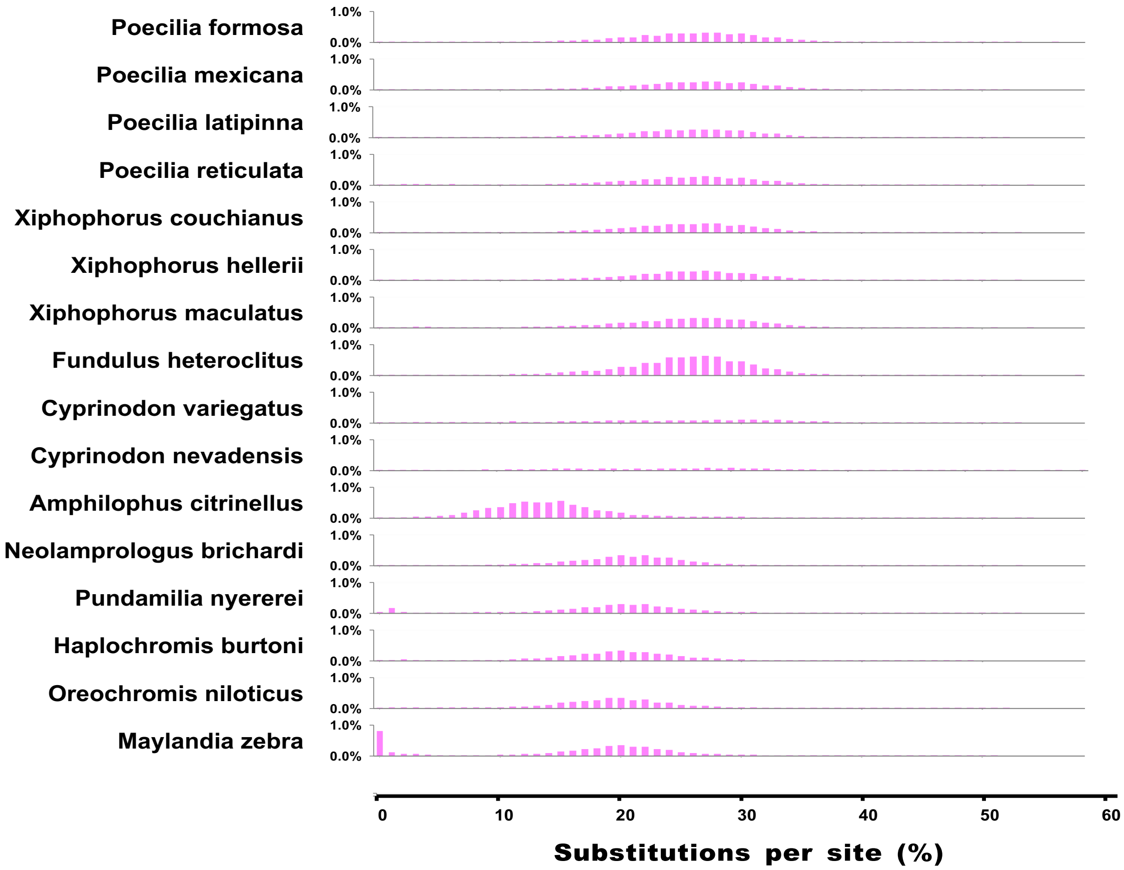
Table 3-3. Proportion of DNA/TcMar-Tc1, microsatellites contents out of all repeats in freshwater, marine and diadromous teleost species.

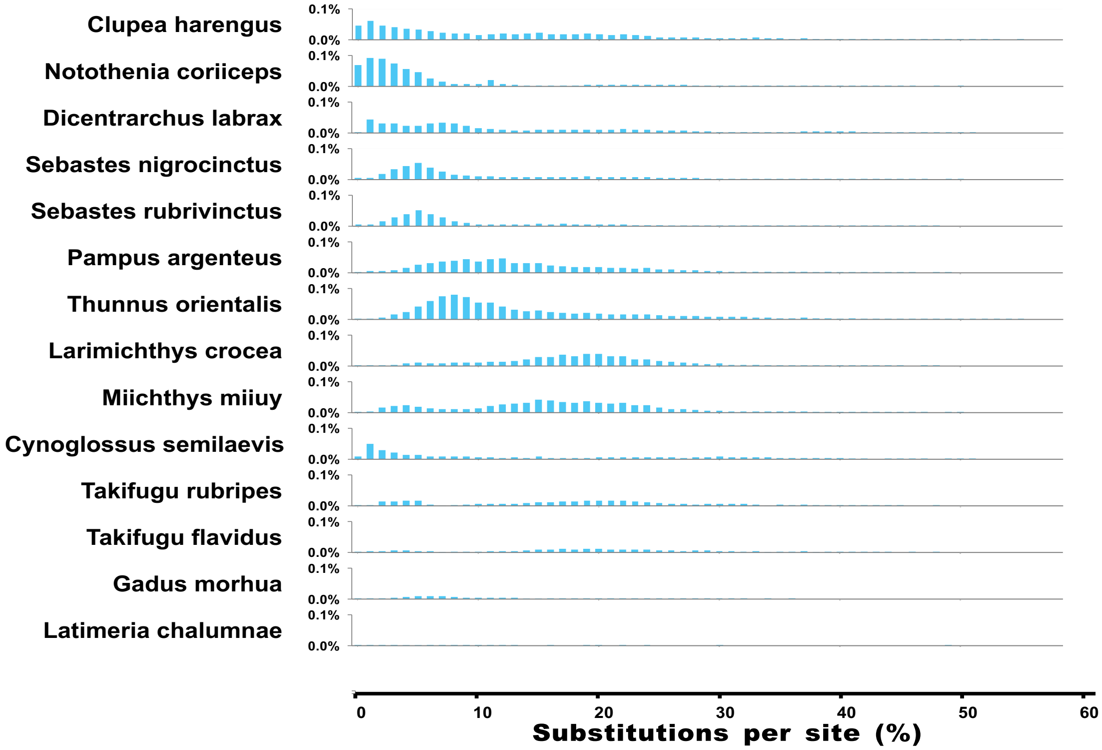
	Species	Order	DNA/TcMar-Tc1	Microsatellites
Freshwater species	*Esox lucius*	Esociformes	35.9%	4.8%
	Fundulus heteroclitus	Cyprinodontiformes	22.6%	4.6%
	Xiphophorus hellerii	Cyprinodontiformes	22.7%	5.1%
	Xiphophorus couchianus	Cyprinodontiformes	22.3%	5.1%
	Amphilophus citrinellus	Perciformes	23.0%	6.2%
	Xiphophorus maculatus	Cyprinodontiformes	22.0%	6.2%
	Lepisosteus oculatus	Lepisosteiformes	10.5%	3.0%
	Pundamilia nyererei	Perciformes	18.7%	5.8%
	Haplochromis burtoni	Perciformes	19.6%	6.1%
	Maylandia zebra	Perciformes	16.8%	5.3%
	Neolamprologus brichardi	Perciformes	20.9%	6.7%
	Cyprinodon nevadensis	Cyprinodontiformes	7.6%	2.5%
	Poecilia formosa	Cyprinodontiformes	19.3%	6.5%
	Oreochromis niloticus	Perciformes	15.9%	5.4%
	Poecilia reticulata	Cyprinodontiformes	17.8%	6.1%
	Poecilia mexicana	Cyprinodontiformes	18.9%	6.7%
	Astyanax mexicanus	Characiformes	21.8%	8.0%
	Poecilia latipinna	Cyprinodontiformes	19.5%	7.4%
	Cyprinodon variegatus	Cyprinodontiformes	8.2%	3.4%
	Oryzias latipes	Beloniformes	5.0%	2.6%
	Ictalurus punctatus	Siluriformes	19.9%	14.1%
	Danio rerio	Cypriniformes	6.1%	5.9%
	Cyprinus carpio	Cypriniformes	6.4%	7.1%
	Sinocyclocheilus grahami	Cypriniformes	4.7%	5.7%
	Sinocyclocheilus rhinocerous	Cypriniformes	3.4%	6.0%
	Sinocyclocheilus anshuiensis	Cypriniformes	3.2%	6.2%
	Pimephales promelas	Cypriniformes	3.1%	6.7%
	Cottus rhenanus	Scorpaeniformes	0.9%	17.8%
	Tetraodon nigroviridis	Tetraodontiformes	1.3%	31.1%
Diadromous species	*Salmo salar*	Salmoniformes	23.6%	7.5%
	Anguilla anguilla	Anguilliformes	11.9%	11.4%
	Anguilla rostrata	Anguilliformes	11.8%	13.9%
Marine species	*Thunnus orientalis*	Perciformes	3.6%	9.3%
	Pampus argenteus	Perciformes	5.6%	15.2%
	Gasterosteus aculeatus	Gasterosteiformes	4.1%	12.8%
	Miichthys miiuy	Perciformes	4.0%	14.7%
	Notothenia coriiceps	Perciformes	2.3%	9.5%
	Dicentrarchus labrax	Perciformes	2.4%	11.9%
	Larimichthys crocea	Perciformes	3.4%	17.7%
	Takifugu rubripes	Tetraodontiformes	3.3%	19.9%
	Sebastes nigrocinctus	Scorpaeniformes	1.1%	8.9%
	Cynoglossus semilaevis	Pleuronectiformes	2.7%	23.1%
	Takifugu flavidus	Tetraodontiformes	2.4%	21.8%
	Sebastes rubrivinctus	Scorpaeniformes	1.0%	9.1%
	Clupea harengus	Clupeiformes	3.0%	29.8%
	Latimeria chalumnae	Coelacanthiformes	0.0%	1.7%
	Gadus morhua	Gadiformes	0.4%	31.4%

Of the bony fish genomes, the most striking discovery is that freshwater bony fish contained a greater proportion of Tc1/mariner transposable elements than marine species (Figure 3-2, PIC p- value: 0.117, Wilcoxon rank test, P-value= 8.23e-06). However, the results were not significant when the phylogeny was taken into consideration. In contrast, the marine bony fish contain a greater proportion of microsatellites (PIC p-value: 3.12e-02, Wilcoxon rank test, P-value=3.72e-05) than the freshwater species, independent of the phylogeny. Interestingly, the diadromous species such as *Anguilla rostrata*, *Anguilla anguilla*, and *S. salar* contain high proportions of both the Tc1/mariner transposable elements and microsatellites (Table 3-3).

Analysis of the sequence divergence rates suggest that Tc1 transposable elements have been present in the genomes of freshwater species for much a longer period of time or are more active than in marine species (Figure 3-3). The Tc1 transposable elements in freshwater species are not only more abundant, but also exhibited a higher average K (average number of substitutions per site) (PIC p-value: 2.10e-02, Wilcoxon rank test, P-value=5.39e-03) than those in marine species. This is particularly notable in Cyprinodontiformes and Labroidei in Perciformes, where Tc1 transposable elements appeared to have the strongest activity over a long history, as reflected by the broad distribution and sharp peaks with higher substitution rates per site (Figure 3-3). The long history and high transposition activities in freshwater fish accounted, at least in part, for the high proportion of Tc1 transposable elements in the genomes of freshwater species.

Figure 3-3. Divergence distribution analysis of DNA/TcMar-Tc1 transposable elements in the representative fish genomes. The Cyprinodontiformes, Labroidei species (red) and marine bony fish (blue) are displayed. The y-axis represents the percentage of the genome comprised of repeat classes (%) and the x-axis represents the substitution rate from consensus sequences (%). Please note that not all y-axis scales are the same, particularly in marine species which are 10 times smaller.


Poecilia formosa
Poecilia mexicana
Poecilia latipinna
Poecilia reticulata
Xiphophorus couchianus
Xiphophorus hellerii
Xiphophorus maculatus
Fundulus heteroclitus
Cyprinodon variegatus
Cyprinodon nevadensis
Amphilophus citrinellus
Neolamprologus brichardi
Pundamilia nyererei
Haplochromis burtoni
Oreochromis niloticus
Maylandia zebra
1.0%
0.0%
0
10
20
30
40
50
60
Substitutions per site (%)


Clupea harengus
Notothenia coriiceps
Dicentrarchus labrax
Sebastes nigrocinctus
Sebastes rubrivinctus
Pampus argenteus
Thunnus orientalis
Larimichthys crocea
Miichthys miiuy
Cynoglossus semilaevis
Takifugu rubripes
Takifugu flavidus
Gadus morhua
Latimeria chalumnae
0.1%
0.0%
0
10
20
30
40
50
60
Substitutions per site (%)

3.5 Discussion

3.5.1. Accumulation of repetitive elements in fish genomes

In this work, I determined the correlation between the categories, proportions of repetitive elements and the living environments of various fish species. Based on the results, it is observed that class II transposons appeared to be more abundantly associated with freshwater bony fish than with marine bony fish, when phylogeny was not considered. In contrast, microsatellites are more abundantly associated with marine bony fish than with freshwater bony fish, independent of phylogeny. In addition, class I transposable elements are more abundant in primitive species such as cartilaginous fish and lamprey than in bony fish. Such findings suggest that the repetitive elements could potentially be related to the adaptability of fish to their living environment, although it is unknown at present if the differential categories and proportions of repetitive elements led to the adaptation to their living environments (the cause) or the living environments led to the accumulation of different repetitive elements (the consequences).

With teleost fish, the genome sizes are greatly affected by the teleost-specific round of whole genome duplication (Allendorf and Thorgaard, 1984; Meyer et al., 2005; Xu et al., 2014). However, whole genome duplication did not dramatically change the proportion of the repetitive elements in the genomes. In contrast, the expansion of repetitive elements may have contributed to the expansion of fish genome sizes because in this analysis, the fish genome sizes, with exceptions, are well correlated with their contents of repetitive elements. High contents of repetitive elements in the genome can accelerate the generation of novel genes for adaptations, but their overburden can also cause abnormal recombination and splicing, resulting in unstable genomes (Jiang et al., 1997). Therefore, the content of the repetitive elements cannot grow unlimited with the genome size; it must be limited to certain levels and shaped under specific natural selection by the environment.

It is worthwhile noting that the quality of the genome assembly varied greatly. As one would expect, many of the repetitive elements may have not been assembled into the reference genome sequences, especially with those of lower assembly qualities. This may have affected the assessment of the proportions of the repetitive elements in the genomes. However, most of the genomes sequencing methods are overall similar via next generation sequencing especially Illumina sequencing, thus the systematic biases related to repeat resolution can be disregarded. Besides, if the unassembled repetitive elements are more or less random, the quality of the genome assemblies should not have systematically affected the enrichment of specific categories of repetitive elements with habitats. In addition, as the total number of genomes used in the study is relatively large (52), the impact of sequence assembly quality should have been minimized.

3.5.2 Comparison of the repetitive elements among species

The distributions of repetitive elements are significantly associated with various clades during evolution. For example, class I transposable elements are more prevalent in cartilaginous fish and lampreys than in bony fish species. However, the cartilaginous fish and lamprey lack the class II transposable elements. Although there were no unifying explanations for this difference, it is speculated from the fact that active transposable elements types in mammals are also RNA transposable elements, that this is probably because internal fertilization of cartilaginous fish, which may have minimized the exposure of gametes and embryos from horizontal transfer of Class II transposable elements (Compagno, 1990; Gao et al., 2016; Huang et al., 2012). For lamprey, since it is still unclear how it fertilizes and develops in the wild (Clemens et al., 2010; Siwicke and Seitz, 2015), its accumulation of class I transposable elements deserve further investigation. As class I transposable elements are involved in various biological processes such as regulation of gene expression (Brosius, 1999; Brosius, 2000), the ancient accumulation of class I transposable elements in cartilaginous fish and lamprey are probably related with their evolutionary adaptations (Gess et al, 2006). The contents of class I transposable elements are low in bony fish; the exact reasons are unknown, but could involve putative mechanisms that counteract the invasiveness of RNAs on their genomes. Although there is a much larger number of bony fish genomes are used in this study than those from cartilaginous fish and lampreys, but this is dictated by the availability of genome sequences. However, if the repetitive elements are more conserved in their categories and proportions of the genome among most closely related species, such bias in the number of genomes used in the analysis should not significantly change the results.

Most freshwater bony fish are dominated by DNA transposable elements except *C. rhenanus* and *T. Nigroviridis* which contain high levels of microsatellites. Although *T. Nigroviridis* is a freshwater species, vast majority (497 out of 509) of species in Tetraodontidae family are marine species (Jaillon et al., 2004; Nelson , 2006; Watson et al., 2009). Thus it is likely that *T. Nigroviridis* had a marine origin. Similarly, *C. rhenanus* is a freshwater species, but most species of the Cottidae family are marine species (Nelson, 2006). In addition, the biology of *C. rhenanus* is largely unknown (Ovidio et al., 2009; Xiang-Yi et al., 2012), the origin of *C. rhenanus* as a freshwater species remains unexplained.

Uncovering the route of class II transposable elements expansion is difficult, because they can be transferred both vertically and horizontally (Abrusán and Krambeck, 2006; Huang et al., 2012; Zhang et al., 2014). However, when phylogenic relationships were not considered, the observed the prevalent class II transposable element in freshwater species may indicate that the freshwater environments are more favorable for proliferation and spreading of DNA transposable elements. Besides, like in other species, the frequent stress such as droughts and floods in the freshwater ecosystem can accelerate transpositions, which facilitate the host adaptions to the environment by generating new genetic variants (Schrader et al., 2014). Previous studies showed that freshwater ray-finned fish have smaller effective population sizes and larger genome sizes than marine species (Yi et al., 2005). These results lend additional support to the idea that shrinking effective population sizes may have underlined the evolution of more complex genomes (Howe et al., 2013; Lynch and Conery, 2003). Although the significance for the more prevalence of Tc1 transposon in freshwater species was reduced when the phylogenetic independent analysis was counted, which indicates the taxa in our data set for the analysis are not statistically independent because of shared evolutionary history. However, due to the dictation of the limited and uneven sequenced species available so far, it is inevitably to introduce the phylogenetic bias into the analysis. For example, a large number of the sequenced fish species are concentrated in the family of Cichlidae (6) or Cyprinidae (6). However, there is only one genome available (*Ictalurus punctatus*) from the order of Siluriformes, which comprise 12% of all fish species (Liu et al., 2016; Sullivan et al., 2006). Analyses of independent contrasts are robust to random species sampling (Ackerly and Reich, 1999), thus, further analysis should be conducted with a broader scope with more sequenced fish species, to complement the broader comparative studies.

Although the *Gasterosteus aculeatus* is collected from freshwater, studies indicated that limnetic *G. aculeatus* are formed as a result of marine populations trapped in freshwater recently (McPhail, 1993; McPhail, 1994; Jones et al., 2012). Thus the *G. aculeatus* is still classified as marine species. The population of marine species tend to be more stable than those in freshwater. Besides, the marine teleost species tend to have a higher osmotic pressure of body fluid (Parry, 1966; Yancey et al., 1982), thus, the high salinity environment may be prone to DNA polymerase slippage while not favorable for proliferation and spreading of transposable elements, since previous studies indicated that the higher salt concentration might stabilize the hairpin structure during the DNA polymerase slippage (Canceill and Ehrlich, 1996). Future research covering a broader scope of sequenced fish linages will address whether passive increases in genome size have in fact been co-opted for the adaptive evolution of complexity in fish as well as other lineages.

3.6 Conclusions

In this study, I investigated the diversity, abundance, and distribution of repetitive elements among 52 fish species in 22 orders, with the correction of the phylogenetic independent contrasts analysis. Differential associations of repetitive elements are found correlated with various clades as well as with their living environments. Class I transposable elements are abundant in lamprey and cartilaginous fish, but less so in bony fish. Tc1/mariner transposable elements are more abundant in freshwater bony fish than in marine fish when phylogeny was not taken into consideration, while microsatellites are more abundant in marine species than those in freshwater species independent of phylogeny. The average number of substitutions per sites of Tc1 among bony fish species suggested their longer and more active of expansion in freshwater species than in marine species, suggesting that freshwater environment is more favorable for the proliferations of Tc1 transposable elements. The analysis of the number of repeats within each microsatellite locus suggested that DNA polymerases are more prone to slippage during replication in marine environments than in freshwater environments. These observations support the notion that repetitive elements have roles for environmental adaptations during evolution. However, whether that is the cause or the consequences requires future studies with more comprehensive sequenced genomes.

References

Abrusán, G., Krambeck, H.-J., 2006. Competition may determine the diversity of transposable elements. Theoretical population biology. 70, 364-375.

Ackerly, D.D., Reich, P.B., 1999. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. American Journal of Botany. 86, 1272-1281.

Allendorf, F.W., Thorgaard, G.H., 1984. Tetraploidy and the evolution of salmonid fishes, Evolutionary genetics of fishes. Springer, pp. 1-53.

AlMomin, S., Kumar, V., Al-Amad, S., Al-Hussaini, M., Dashti, T., Al-Enezi, K., Akbar, A., 2015. Draft genome sequence of the silver pomfret fish, Pampus argenteus. Genome. 59, 51-58.

Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., MacCallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature. 496, 311-316.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-m., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science. 297, 1301-1310.

Balaresque, P., King, T.E., Parkin, E.J., Heyer, E., Carvalho- Silva, D., Kraaijenbrink, T., Knijff, P., Tyler- Smith, C., Jobling, M.A., 2014. Gene Conversion Violates the Stepwise Mutation Model for Microsatellites in Y- Chromosomal Palindromic Repeats. Human mutation. 35, 609-617.

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 6, 11.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome research. 12, 1269-1276.

Barrio, A.M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., Dainat, J., Ekman, D., Höppner, M., Jern, P., 2016. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. Elife. 5, e12081.

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., 2015. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nature genetics. 47, 427-437.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513, 375-381.

Brosius, J., 1999. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica. 107, 209-238.

Brosius, J., 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene. 238, 115-134.

Burns, F.R., Cogburn, A.L., Ankley, G.T., Villeneuve, D.L., Waits, E., Chang, Y.J., Llaca, V., Deschamps, S.D., Jackson, R.E., Hoke, R.A., 2016. Sequencing and de novo draft assemblies of a fathead minnow (Pimephales promelas) reference genome. Environmental toxicology and chemistry. 35, 212-217.

Canceill, D., Ehrlich, S.D., 1996. Copy-choice recombination mediated by DNA polymerase III holoenzyme from Escherichia coli. Proceedings of the National Academy of Sciences. 93, 6647-6652.

Castresana, J., 2001. Cytochrome b phylogeny and the taxonomy of great apes and mammals.

Molecular Biology and Evolution. 18, 465-471.

Chalopin, D., Naville, M., Plard, F., Galiana, D., Volff, J.-N., 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome biology and evolution. 7, 567-580.

Chalopin, D., Volff, J.-N., Galiana, D., Anderson, J.L., Schartl, M., 2015. Transposable elements and early evolution of sex chromosomes in fish. Chromosome research. 23, 545-560.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 371, 215-220.

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J.-N., 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature genetics. 46, 253-260.

Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420, 520-562.

Clemens, B.J., Binder, T.R., Docker, M.F., Moser, M.L., Sower, S.A., 2010. Similarities, differences, and unknowns in biology and management of three parasitic lampreys of North America. Fisheries. 35, 580-594.

Compagno, L.J., 1990. Alternative life-history styles of cartilaginous fishes in time and space, Alternative life-history styles of fishes. Springer, pp. 33-75.

Conte, M.A., Kocher, T.D., 2015. An improved genome reference for the African cichlid, Metriaclima zebra. BMC genomics. 16, 724.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 27, 1164-1165.

Deng, W., Wang, Y., Liu, Z., Cheng, H., Xue, Y., 2014. HemI: a toolkit for illustrating heatmaps. PloS one. 9, e111988.

Fraser, B.A., Künstner, A., Reznick, D.N., Dreyer, C., Weigel, D., 2015. Population genomics of natural and experimental populations of guppies (Poecilia reticulata). Molecular Ecology. 24, 389-408.

Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., Song, C., 2016. The contribution of transposable elements to size variations between four teleost genomes. Mobile DNA. 7, 4.

Gao, Y., Gao, Q., Zhang, H., Wang, L., Zhang, F., Yang, C., Song, L., 2014. Draft sequencing and analysis of the genome of pufferfish Takifugu flavidus. DNA research. 21, 627-637.

Gess, R.W., Coates, M.I., Rubidge, B.S., 2006. A lamprey from the Devonian period of South Africa. Nature. 443, 981-984.

Hacch, F., Mazrimas, J., 1974. Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys ordii*). Nucleic acids research. 1, 559-576.

Hancock, J.M., 1996. Simple sequences and the expanding genome. Bioessays. 18, 421-425.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 496, 498-503.

Huang, C.R.L., Burns, K.H., Boeke, J.D., 2012. Active transposition in genomes. Annual review of genetics. 46, 651-675.

Huang, C.R.L., Burns, K.H., Boeke, J.D., 2012. Active transposition in genomes. Annual review of genetics. 46, 651-675.

Hurst, G.D., Werren, J.H., 2001. The role of selfish genetic elements in eukaryotic evolution.

Nature Reviews Genetics. 2, 597-606.

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431, 946-957.

Jiang, H., 1998. The distribution trends in simple repetitive stretches of DNA. Chinese journal of biochemistry and molecular biology. 14, 65-70.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484, 55-61.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature. 447, 714-719.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 30, 772-780.

Kazazian, H.H., 1999. An estimated frequency of endogenous insertional mutations in humans.

Nature genetics. 22, 130-130.

Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. science. 303, 1626-1632.

Kubis, S., Schmidt, T., Heslop-Harrison, J.S., 1998. Repetitive DNA elements as a major component of plant genomes. Annals of Botany. 82, 45-55.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution. 33, 1870-1874.

Lee, S.-I., Kim, N.-S., 2014. Transposable elements and genome size variations in plants. Genomics & informatics. 12, 87-97.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 533, 200-205.

Lindahl, T., 1994. DNA Repair: DNA surveillance defect in cancer cells. Current Biology. 4, 249-251.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., Li, Y., Jiang, C., Sun, L., Wang, R., Zhang, Y., 2016. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nature communications. 7:11757.

Lynch, M., Conery, J.S., 2003. The origins of genome complexity. science. 302, 1401-1404.

Martin, P., Makepeace, K., Hill, S.A., Hood, D.W., Moxon, E.R., 2005. Microsatellite instability regulates transcription factor binding and gene expression. Proceedings of the National Academy of Sciences of the United States of America. 102, 3800-3804.

McGaugh, S.E., Gross, J.B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, W.R., Keene, A., Ma, L., 2014. The cavefish genome reveals candidate genes for eye loss. Nature communications. 5:5307.

McPhail, J., 1993. Ecology and evolution of sympatric sticklebacks (Gasterosteus): origin of the species pairs. Canadian Journal of Zoology. 71, 515-523.

McPhail, J., 1993. Speciation and the evolution of reproductive isolation in the sticklebacks Gasterosteus) of southwestern British Columbia. The evolutionary biology of the threespine stickleback. p. 399-437.

Meagher, T.R., Vassiliadis, C., 2005. Phenotypic impacts of repetitive DNA in flowering plants. New Phytologist. 168, 71-80.

Meyer, A., Van de Peer, Y., 2005. From 2R to 3R: evidence for a fish- specific genome duplication (FSGD). Bioessays. 27, 937-945.

Moxon, E.R., Rainey, P.B., Nowak, M.A., Lenski, R.E., 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current biology. 4, 24-33.

Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., Shigenobu, Y., Ojima, N., Muta, S., Fujiwara, A., 2013. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proceedings of the National Academy of Sciences. 110, 11061-11066.

Nelson J. Fishes of the world 4th edition. Hoboken: John Wiley & Sons, Inc; 2006. p 334; p456.

Ohno, S., 1972. So much" junk" DNA in our genome, Brookhaven symposia in biology, pp. 366-370.

Ovidio, M., Detaille, A., Bontinck, C., Philippart, J.-C., 2009. Movement behaviour of the small benthic Rhine sculpin Cottus rhenanus (Freyhof, Kottelat & Nolte, 2005) as revealed by radio-telemetry and pit-tagging. Hydrobiologia. 636, 119-128.

Pardue, M., Lowenhaupt, K., Rich, A., Nordheim, A., 1987. (dC-dA) n.(dG-dT) n sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. The EMBO journal. 6, 1781-1789.

Parry, G., 1966. Osmotic adaptation in fishes. Biological Reviews. 41, 392-440.

Petitpierre, E., Juan, C., Pons, J., Plohl, M., Ugarkovic, D., 1995. Satellite DNA and constitutive heterochromatin in tenebrionid beetles, Kew Chromosome Conference IV. Royal Botanic Gardens, London, pp. 351-362.

Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. Bioinformatics. 21, i351-i358.

R Core Team. 2013. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/

Read, T.D., Petit, R.A., Joseph, S.J., Alam, M.T., Weil, M.R., Ahmad, M., Bhimani, R., Vuong, J.S., Haase, C.P., Webb, D.H., 2017. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: Rhincodon typus Smith 1828. BMC genomics. 18, 532.

Richard, G.F., Pâques, F., 2000. Mini- and microsatellite expansions: the recombination connection. EMBO reports. 1, 122-126.

Rondeau, E.B., Minkley, D.R., Leong, J.S., Messmer, A.M., Jantzen, J.R., von Schalburg, K.R., Lemon, C., Bird, N.H., Koop, B.F., 2014. The genome and linkage map of the northern pike (Esox lucius): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One. 9, e102089.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science. 274, 765-768.

Schartl, M., Walter, R.B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J.-N., Lesch, K.-P., 2013. The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nature genetics. 45, 567-572.

Schmidt, A.L., Anderson, L.M., 2006. Repetitive DNA elements as mediators of genomic change in response to environmental cues. Biological Reviews. 81, 531-543.

Schrader, L., Kim, J.W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., 2014. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nature communications. 5, 5495.

Shin, S.C., Ahn, D.H., Kim, S.J., Pyo, C.W., Lee, H., Kim, M.-K., Lee, J., Lee, J.E., Detrich, H.W., Postlethwait, J.H., 2014. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. Genome biology. 15, 468.

Siwicke, K.A., Seitz, A.C., 2015. Interpreting Lamprey Attacks on Pacific Cod in the Eastern Bering Sea. Transactions of the American Fisheries Society. 144, 1249-1262.

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature genetics. 45, 415-421.

Smolka, M., Rescheneder, P., Schatz, M.C., von Haeseler, A., Sedlazeck, F.J., 2015. Teaser: Individualized benchmarking and optimization of read mapping results for NGS data. Genome biology. 16, 235.

Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T.F., Rounge, T.B., Paulsen, J., Solbakken, M.H., Sharma, A., 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 477, 207-210.

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D., 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 365, 274-276.

Sullivan, J.P., Lundberg, J.G., Hardman, M., 2006. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. Molecular phylogenetics and evolution. 41, 636-662.

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L., Wang, L., Wu, D.-D., Zhang, B.-L., 2015. Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. Proceedings of the National Academy of Sciences. 112, E1257-E1262.

Thornburg, B.G., Gotea, V., Makałowski, W., 2006. Transposable elements as a significant source of transcription regulating signals. Gene. 365, 104-110.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S., Hecht, J., Knaust, F., Belkhir, K., Klages, S., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nature communications. 5, 5770.

Tóth, G., Gáspári, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome research. 10, 967-981.

Ugarković, Đ., Plohl, M., 2002. Variation in satellite DNA profiles—causes and effects. The EMBO journal. 21, 5955-5959.

Venkatesh, B., Lee, A.P., Ravi, V., Maurya, A.K., Lian, M.M., Swann, J.B., Ohta, Y., Flajnik, M.F., Sutoh, Y., Kasahara, M., 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 505, 174-179.

Volff, J., 2005. Genome evolution and biodiversity in teleost fish. Heredity. 94, 280-294.

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., 2014. The locust genome provides insight into swarm formation and long-distance flight. Nature communications. 5, 2957.

Watson, C.A., Hill, J.E., Graves, J.S., Wood, A.L., Kilgore, K.H., 2009. Use of a novel induced spawning technique for the first reported captive spawning of Tetraodon nigroviridis. Marine genomics. 2, 143-146.

Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F., Finn, R.D., 2012. Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic acids research. 41, D70-D82.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics. 8, 973-982.

Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y., Zhou, D., Zhang, J., Zhang, Z., Xu, M., 2014. The draft genome of the large yellow croaker reveals well-developed innate immunity. Nature communications. 5.

Xiang-Yi L., Nolte AW., Vincx M., Sedlazek F., Konrad K., 2012. Genome evolution following admixture in invasive sculpins; Master Thesis, Max-Planck-Institute für Evolutionsbiologie Plön.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., 2014. Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nature genetics. 46, 1212-1219.

Xu, T., Xu, G., Che, R., Wang, R., Wang, Y., Li, J., Wang, S., Shu, C., Sun, Y., Liu, T., 2016. The genome of the miiuy croaker reveals well-developed innate immune and sensory systems.

Scientific reports. 6, 21902.

Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D., Somero, G.N., 1982. Living with water stress: evolution of osmolyte systems. Science. 217, 1214-1222.

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., 2016. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. BMC biology. 14, 1.

Yi, S., Streelman, J.T., 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends in Genetics. 21, 643-646.

Zhang, H.-H., Feschotte, C., Han, M.-J., Zhang, Z., 2014. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. Genome biology and evolution. 6, 1375-1386.

Appendices and Supplementary Tables

Table S-1 **The content of repetitive elements in the order Anguilliformes including *Anguilla Anguilla; Anguilla rostrate.***

Types	*Anguilla anguilla*	*Anguilla rostrata*
DNA/Academ	0.1%	0.1%
DNA/CMC-Chapaev-3	0.0%	/
DNA/CMC-EnSpm	3.0%	2.1%
DNA/Crypton	0.2%	0.1%
DNA/Crypton-V	0.1%	/
DNA/Dada	/	0.0%
DNA/Ginger	1.0%	0.9%
DNA/hAT	1.1%	1.4%
DNA/hAT-Ac	1.2%	1.0%
DNA/hAT-Blackjack	0.0%	0.6%
DNA/hAT-Charlie	2.3%	3.0%
DNA/hAT-hAT5	0.0%	/
DNA/hAT-hAT6	/	0.0%
DNA/hAT-hATx	0.0%	/
DNA/hAT-hobo	0.2%	0.2%
DNA/hAT-Tag1	0.0%	/
DNA/hAT-Tip100	0.3%	0.4%
DNA/hAT-Tol2	0.0%	/
DNA/Kolobok-T2	0.1%	0.2%
DNA/Maverick	0.0%	/
DNA/MuLE-MuDR	0.1%	0.1%
DNA/Novosib	0.5%	0.6%
DNA/P	0.0%	0.1%
DNA/PIF-Harbinger	1.1%	1.3%
DNA/PIF-ISL2EU	0.1%	0.3%
DNA/PiggyBac	0.1%	0.3%
DNA/Sola	0.1%	0.1%
DNA/TcMar	0.3%	0.3%
DNA/TcMar-Fot1	0.2%	/
DNA/TcMar-ISRm11	0.0%	0.1%
DNA/TcMar-Tc1	11.9%	11.8%
DNA/TcMar-Tc4	0.1%	/
DNA/Unclassified	7.5%	8.8%
LINE/CR1	1.0%	1.1%
LINE/I	0.2%	0.4%
LINE/I-Nimb	0.0%	/

LINE/Jockey	/	0.0%
LINE/L1	0.1%	0.2%
LINE/L1-Tx1	0.0%	0.1%
LINE/L2	4.7%	5.3%
LINE/Penelope	0.0%	0.0%
LINE/R1	0.1%	0.1%
LINE/R2-Hero	0.1%	0.0%
LINE/R2-NeSL	/	0.0%
LINE/Rex-Babar	0.2%	0.2%
LINE/RTE-BovB	0.2%	0.2%
LINE/RTE-X	0.5%	0.4%
LINE/Unclassified	0.0%	0.0%
Low_complexity	2.2%	2.2%
LTR/Copia	0.0%	0.1%
LTR/DIRS	0.2%	0.1%
LTR/ERV1	0.2%	1.6%
LTR/ERVK	/	0.0%
LTR/Gypsy	0.8%	0.7%
LTR/Gypsy-Cigr	/	0.0%
LTR/Ngaro	1.4%	0.4%
LTR/Pao	/	0.1%
LTR/Unclassified	0.0%	0.1%
RC/Helitron	1.0%	0.8%
Satellite	0.3%	0.3%
Microsatellites	11.4%	13.9%
SINE/5S-Deu-L2	0.8%	1.0%
SINE/L2	0.0%	/
SINE/MIR	3.8%	4.0%
SINE/tRNA	0.1%	0.1%
SINE/tRNA-Core-L2	/	0.5%
SINE/tRNA-L2	0.2%	0.1%
SINE/tRNA-V	/	0.1%
SINE?	0.1%	0.0%
SINE/Unclassified	0.2%	0.1%
Unknown	38.3%	32.1%

Table S-2 **The content of repetitive elements in the order Perciformes including: *Amphilophus citrinellus*; *Neolamprologus brichardi*; *Pundamilia nyererei*; *Haplochromis burtoni.***

Types	***Amphilophus citrinellus***	***Neolamprologus brichardi***	***Pundamilia nyererei***	***Haplochromis burtoni***
DNA/Academ	0.1%	0.0%	/	0.0%
DNA/CMC-Chapaev-3	0.1%	0.0%	0.0%	/
DNA/CMC-EnSpm	0.6%	0.5%	0.4%	0.5%
DNA/Dada	0.1%	/	/	/
DNA/hAT	1.3%	2.1%	1.9%	1.9%
DNA/hAT-Ac	2.7%	5.0%	5.2%	5.3%
DNA/hAT-Blackjack	0.2%	0.1%	0.0%	0.0%
DNA/hAT-Charlie	2.0%	3.6%	3.3%	3.0%
DNA/hAT-hAT5	0.1%	0.1%	0.1%	0.1%
DNA/hAT-hATx	/	0.3%	/	0.0%
DNA/hAT-hobo	0.1%	0.0%	/	/
DNA/hAT-Tip100	0.6%	0.7%	1.1%	0.4%
DNA/hAT-Tol2	1.1%	0.6%	0.7%	0.8%
DNA/IS3EU	0.4%	0.1%	0.2%	0.2%
DNA/Kolobok-T2	/	0.0%	0.0%	/
DNA/Maverick	/	0.1%	/	/
DNA/Merlin	0.1%	0.1%	/	0.2%
DNA/MuLE-MuDR	0.0%	0.1%	0.1%	0.0%
DNA/Novosib	/	0.1%	/	/
DNA/P	0.1%	0.1%	0.1%	0.2%
DNA/PIF-Harbinger	1.7%	1.0%	1.8%	0.8%
DNA/PIF-ISL2EU	0.1%	/	0.1%	/
DNA/PiggyBac	0.2%	1.1%	1.1%	1.5%
DNA/Sola	0.0%	0.1%	0.0%	0.1%
DNA/TcMar	0.0%	/	0.1%	0.0%
DNA/TcMar-Fot1	/	0.1%	0.0%	0.0%
DNA/TcMar-ISRm11	0.1%	0.1%	0.0%	/
DNA/TcMar-Mariner	0.1%	0.1%	0.1%	0.2%
DNA/TcMar-Tc1	23.0%	20.9%	18.7%	19.6%
DNA/TcMar-Tc2	0.4%	1.0%	1.5%	1.4%
DNA/TcMar-Tigger	0.1%	0.4%	0.2%	0.1%
DNA/Zisupton	0.0%	/	/	/
DNA/Unclassified	4.7%	3.1%	5.7%	4.9%
LINE/CR1	0.0%	/	/	/
LINE/Dong-R4	1.7%	1.2%	1.1%	1.2%
LINE/I	0.1%	0.1%	0.0%	/
LINE/I-Nimb	/	/	0.1%	0.1%

LINE/L1	1.1%	1.2%	1.3%	1.6%
LINE/L1-Tx1	1.0%	/	0.0%	/
LINE/L2	9.2%	8.1%	7.7%	8.4%
LINE/LOA	/	0.0%	/	/
LINE/Penelope	0.8%	2.2%	2.2%	2.3%
LINE/Proto2	0.0%	/	/	/
LINE/R1	/	0.1%	/	/
LINE/R2	/	/	0.1%	/
LINE/R2-Hero	0.1%	0.2%	0.1%	0.1%
LINE/Rex-Babar	2.1%	2.4%	3.2%	2.8%
LINE/RTE	/	0.1%	0.9%	0.1%
LINE/RTE-BovB	6.2%	3.9%	3.3%	3.9%
LINE/RTE-X	/	/	0.0%	/
LINE/Unclassified	0.1%		/	/
Low_complexity	0.8%	1.2%	1.0%	1.1%
LTR/Copia	0.1%	0.1%	/	0.0%
LTR/DIRS	0.1%	0.1%	0.1%	0.1%
LTR/ERV1	0.4%	0.0%	0.2%	0.1%
LTR/ERV4	0.4%	/	/	/
LTR/ERVK	0.0%	/	/	0.0%
LTR/Gypsy	1.1%	1.1%	1.2%	1.2%
LTR/Gypsy-Cigr	/	/	0.0%	/
LTR/Ngaro	1.7%	1.1%	1.8%	2.1%
LTR/Pao	0.3%	0.0%	0.0%	0.0%
LTR/Unclassified	0.1%	/	/	/
RC/Helitron	1.2%	0.1%	0.1%	0.3%
Retroposon	0.6%	0.5%	0.5%	0.5%
rRNA	0.1%	0.0%	0.0%	0.0%
Satellite	0.3%	0.3%	0.5%	0.0%
Microsatellites	6.2%	6.7%	5.8%	6.1%
SINE/5S	/	0.0%	/	/
SINE/5S-Deu-L2	0.0%	/	0.0%	/
SINE/Alu	0.0%	/	/	/
SINE/ID	0.0%	/	/	/
SINE/MIR	0.1%	1.5%	1.5%	1.5%
SINE/Ron	/	/	/	0.0%
SINE/Ron1	0.0%	/	/	/
SINE/tRNA	0.2%	0.1%	0.1%	0.1%
SINE/tRNA-Core	1.4%	0.8%	1.1%	1.1%
SINE/tRNA-V	0.1%	1.0%	0.3%	0.2%
SINE?	0.0%	0.3%	0.0%	0.1%
SINE/Unclassified	0.1%	0.1%	0.2%	0.2%
Unknown	22.0%	24.0%	22.9%	23.4%

Table S-3 **The content of repetitive elements in the order Perciformes including: *Oreochromis niloticus*; *Maylandia zebra*; *Notothenia coriiceps*; *Dicentrarchus labrax.***

Types	*Oreochromis niloticus*	*Maylandia zebra*	*Notothenia coriiceps*	*Dicentrarchus labrax*
DNA/Academ	/	/	0.1%	/
DNA/CMC-Chapaev-3	/	0.0%	/	/
DNA/CMC-EnSpm	0.8%	0.4%	1.5%	1.5%
DNA/Crypton	/	/	0.6%	0.4%
DNA/Crypton-V	/	0.1%	0.1%	0.1%
DNA/Dada	/	0.0%	/	0.1%
DNA/Ginger	/	/	/	0.0%
DNA/hAT	2.5%	2.1%	1.6%	1.6%
DNA/hAT-Ac	4.9%	8.1%	6.4%	7.2%
DNA/hAT-Blackjack	0.1%	0.0%	0.3%	0.2%
DNA/hAT-Charlie	3.1%	3.2%	5.2%	2.8%
DNA/hAT-hAT5	0.1%	0.1%	0.7%	0.6%
DNA/hAT-hAT6	/	/	0.3%	/
DNA/hAT-hATx	0.0%	/	0.0%	/
DNA/hAT-hobo	0.1%	0.0%	0.6%	0.1%
DNA/hAT-Tag1	0.0%	/	0.1%	/
DNA/hAT-Tip100	1.2%	0.7%	1.7%	0.7%
DNA/hAT-Tol2	0.8%	0.5%	3.1%	1.3%
DNA/IS3EU	0.1%	0.1%	0.4%	0.2%
DNA/Kolobok-T2	/	/	/	0.2%
DNA/Kolobok-Hydra	/	/	0.1%	/
DNA/Kolobok-T2	/	/	3.6%	/
DNA/Maverick	0.2%	0.1%	0.0%	0.0%
DNA/Merlin	0.3%	0.1%	/	/
DNA/MuLE-MuDR	0.0%	/	/	/
DNA/MuLE-NOF	0.0%	/	/	/
DNA/Novosib	0.2%	0.1%	/	/
DNA/P	0.3%	0.2%	0.1%	0.3%
DNA/PIF-Harbinger	1.1%	1.7%	2.1%	0.9%
DNA/PIF-ISL2EU	0.0%	0.0%	0.0%	0.1%
DNA/PiggyBac	0.9%	0.8%	0.7%	0.1%
DNA/TcMar	0.0%	/	/	/
DNA/TcMar-Fot1	0.1%	0.1%	0.1%	0.0%
DNA/TcMar-ISRm11	0.1%	0.3%	0.2%	0.1%
DNA/TcMar-Mariner	0.1%	0.1%	/	/
DNA/TcMar-Stowaway	/	/	0.1%	/
DNA/TcMar-Tc1	15.9%	16.8%	2.3%	2.4%
DNA/TcMar-Tc2	0.9%	1.7%	0.4%	0.0%
DNA/TcMar-Tigger	0.4%	0.2%	/	/
DNA/Unclassified	4.7%	3.9%	5.1%	3.5%

LINE/CR1	/	/	0.2%	/
LINE/Dong-R4	1.4%	1.6%	/	0.1%
LINE/DRE	/	/	0.0%	/
LINE/I	0.0%	0.1%	1.3%	0.8%
LINE/I-Nimb	/	/	0.1%	/
LINE/Jockey	0.1%	/	0.2%	/
LINE/L1	1.5%	3.0%	0.6%	0.1%
LINE/L1-Tx1	0.1%	0.0%	0.9%	0.0%
LINE/L2	8.1%	8.4%	6.7%	6.8%
LINE/Penelope	1.6%	2.1%	/	0.1%
LINE/Proto2	0.0%	0.0%	0.0%	/
LINE/R2	/	/	0.2%	/
LINE/R2-Hero	0.2%	0.1%	/	0.0%
LINE/Rex-Babar	3.9%	5.8%	2.4%	2.9%
LINE/RTE	0.0%	0.6%	0.4%	/
LINE/RTE-BovB	3.4%	3.0%	0.3%	0.4%
LINE/RTE-RTEX	/	/	0.0%	/
LINE/RTE-X	/	/	0.0%	0.4%
LINE/Tad1	0.0%	/	0.1%	/
LINE?/Penelope	/	0.0%	/	/
LINE/Unclassified	0.1%	0.5%	0.0%	0.1%
Low_complexity	0.9%	0.8%	0.7%	1.4%
LTR/Copia	/	/	0.0%	/
LTR/DIRS	0.3%	0.1%	0.8%	0.0%
LTR/ERV1	0.8%	0.1%	0.2%	/
LTR/Gypsy	1.8%	1.6%	1.0%	0.4%
LTR/Gypsy-Cigr	0.1%	/	/	/
LTR/Ngaro	1.0%	1.3%	0.0%	0.3%
LTR/Pao	0.2%	0.3%	0.2%	/
LTR/Unclassified	/	0.0%	0.0%	0.2%
RC/Helitron	0.6%	0.2%	0.7%	2.7%
Retroposon	0.4%	0.5%	/	0.0%
rRNA	0.0%	0.0%	/	1.1%
Satellite	/	0.2%	4.0%	0.3%
Microsatellites	5.4%	5.3%	9.5%	11.9%
SINE/5S	0.0%	/	/	0.1%
SINE/5S-Deu-L2	/	0.0%	0.1%	/
SINE/Alu	0.1%	/	/	/
SINE/MIR	0.7%	1.3%	1.0%	/
SINE/Ron	0.0%	0.0%	/	0.0%
SINE/Ron1	/	/	/	
SINE/tRNA	0.1%	0.0%	0.1%	1.0%
SINE/tRNA-Core	0.9%	0.7%	/	0.5%
SINE/tRNA-Core-L2	0.1%	/	/	0.2%
SINE/tRNA-L2	0.0%	/	0.0%	0.0%
SINE/tRNA-V	0.7%	0.3%	/	0.1%

SINE?	0.1%	0.0%	0.0%	0.0%
SINE/Unclassified	0.1%	0.2%	0.5%	0.8%
snRNA	/	/	/	0.0%
Unknown	26.5%	20.5%	30.3%	42.7%

Table S-4* The content of repetitive elements in the order Perciformes including: *Pampus argenteus*; *Thunnus orientalis*; *Larimichthys crocea*; *Miichthys miiuy.

Types	*Pampus argenteus*	*Thunnus orientalis*	*Larimichthys crocea*	*Miichthys miiuy*
DNA/CMC-Chapaev-3	0.1%	0.0%	0.2%	0.1%
DNA/CMC-EnSpm	0.8%	1.3%	0.8%	0.8%
DNA/Crypton	0.3%	0.1%	0.2%	0.4%
DNA/Crypton-V	0.0%	0.0%	0.0%	0.0%
DNA/Dada	0.6%	/	/	/
DNA/Ginger	/	0.1%	0.1%	0.2%
DNA/hAT	0.1%	1.9%	1.3%	1.1%
DNA/hAT-Ac	5.7%	5.8%	3.8%	5.2%
DNA/hAT-Blackjack	/	0.1%	/	/
DNA/hAT-Charlie	1.2%	1.0%	2.1%	2.0%
DNA/hAT-hAT5	0.2%	0.1%	0.2%	0.1%
DNA/hAT-hAT6	/	0.0%	0.0%	/
DNA/hAT-hobo	0.1%	0.3%	/	0.0%
DNA/hAT-Tip100	0.2%	0.7%	0.4%	0.4%
DNA/hAT-Tol2	2.6%	2.0%	0.6%	1.3%
DNA/IS3EU	/	0.0%	0.0%	/
DNA/Kolobok-T2	0.2%	0.3%	0.0%	0.1%
DNA/Maverick	0.0%	0.0%	/	
DNA/MuLE-MuDR	0.1%	0.0%	/	0.0%
DNA/Novosib	0.2%	/	/	/
DNA/P	0.6%	0.2%	0.1%	0.1%
DNA/PIF-Harbinger	1.2%	2.0%	0.6%	0.4%
DNA/PIF-ISL2EU	/	0.0%	/	/
DNA/PiggyBac	/	0.1%	0.0%	0.1%
DNA/Sola	0.0%	0.3%	0.0%	0.1%
DNA/TcMar	0.0%	/	/	/
DNA/TcMar-Fot1	0.0%	0.1%	/	/
DNA/TcMar-ISRm11	0.1%	0.2%	0.1%	0.1%
DNA/TcMar-Mariner	0.0%	0.0%	/	/
DNA/TcMar-Pogo	/	0.0%	/	/
DNA/TcMar-Tc1	5.7%	3.6%	3.4%	4.0%
DNA/TcMar-Tc2	0.6%	0.3%	0.1%	0.1%
DNA/TcMar-Tigger	0.1%	0.0%	0.5%	/
DNA/Unclassified	2.0%	4.2%	6.1%	6.5%
LINE/CR1	/	0.0%	/	/
LINE/Dong-R4	0.1%	0.1%	0.0%	0.0%
LINE/DRE	/	0.0%	/	/
LINE/I	0.1%	0.6%	0.1%	0.0%
LINE/I-Nimb	/	0.0%	/	0.1%
LINE/Jockey	/	/	/	0.1%

LINE/L1	0.1%	0.2%	0.2%	0.4%
LINE/L1-Tx1	/	0.1%	0.0%	0.1%
LINE/L2	4.5%	6.4%	4.8%	5.3%
LINE/Penelope	0.6%	0.3%	0.3%	0.3%
LINE/Proto2	/	0.0%	/	0.0%
LINE/R2	/	0.1%	/	/
LINE/R2-Hero	/	0.1%	0.0%	0.0%
LINE/Rex-Babar	2.0%	1.3%	2.6%	2.6%
LINE/RTE	/	/	0.0%	/
LINE/RTE-BovB	1.6%	0.4%	1.3%	1.4%
LINE/RTE-X	0.3%	0.4%	/	0.1%
LINE/Unclassified	0.1%	0.1%	/	/
Low_complexity	2.4%	1.3%	2.0%	1.9%
LTR/Copia	0.2%	0.1%	0.7%	0.1%
LTR/DIRS	/	0.4%	/	0.1%
LTR/ERV1	0.0%	0.1%	0.1%	/
LTR/ERVK	/	0.1%	/	/
LTR/ERVL	/	/	/	0.1%
LTR/Gypsy	2.1%	1.0%	1.2%	0.6%
LTR/Ngaro	0.5%	0.2%	1.4%	0.5%
LTR/Pao	0.1%	0.0%	0.0%	0.0%
LTR/Unclassified	0.2%	0.1%	0.3%	0.3%
RC/Helitron	0.3%	1.4%	0.8%	0.4%
Retroposon	/	/	0.0%	/
rRNA	/	/	0.2%	/
Satellite	0.1%	0.1%	0.6%	0.5%
Microsatellites	15.2%	9.3%	17.7%	14.7%
SINE/Alu	0.1%	/	/	/
SINE/MIR	0.3%	0.1%	0.9%	1.0%
SINE/tRNA	0.3%	0.3%	1.1%	1.0%
SINE/tRNA-Core	/	0.1%	0.3%	0.1%
SINE/tRNA-Core-L2	0.1%	0.1%	/	/
SINE/tRNA-L2	0.0%	/	/	0.0%
SINE/tRNA-V	/	0.1%	/	1.1%
SINE?	0.0%	/	0.0%	0.0%
SINE/Unclassified	0.4%	0.8%	0.8%	0.6%
Unknown	45.7%	49.4%	41.7%	43.2%

Table S-5 **The content of repetitive elements in the order Cyprinodontiformes including: *Poecilia Formosa; Poecilia Mexicana; Poecilia; latipinna; Poecilia reticulata Xiphophorus couchianus.***

Types	*Poecilia formosa*	*Poecilia mexicana*	*Poecilia latipinna*	*Poecilia reticulata*	*Xiphophorus couchianus*
DNA/Academ	0.1%	0.1%	0.0%	0.1%	0.1%
DNA/CMC-Chapaev-3	0.1%	0.1%	0.1%	0.1%	0.1%
DNA/CMC-EnSpm	1.3%	0.9%	0.9%	1.2%	0.4%
DNA/Crypton	/	0.0%	/	/	/
DNA/Ginger	/	/	0.1%	/	/
DNA/hAT	1.7%	1.5%	1.5%	1.5%	1.7%
DNA/hAT-Ac	6.2%	6.1%	5.9%	5.6%	7.8%
DNA/hAT-Blackjack	/	0.0%	0.0%	/	0.0%
DNA/hAT-Charlie	4.2%	4.6%	3.9%	7.9%	4.0%
DNA/hAT-hAT5	0.7%	1.1%	0.8%	0.8%	0.9%
DNA/hAT-Tip100	2.4%	1.3%	3.3%	1.8%	1.4%
DNA/hAT-Tol2	1.2%	1.2%	1.4%	2.1%	1.3%
DNA/IS3EU	0.5%	0.2%	0.2%	0.2%	0.3%
DNA/Kolobok-T2	/	1.2%	/	/	/
DNA/Kolobok-Hydra	/	/	/	0.0%	/
DNA/Kolobok-T2	0.7%	/	/	0.7%	0.2%
DNA/Maverick	0.1%	/	/	0.1%	0.1%
DNA/Merlin	0.1%	/	/	/	/
DNA/MuLE-MuDR	0.1%	0.1%	0.1%	0.1%	0.5%
DNA/Novosib	/	/	/	0.0%	/
DNA/P	0.2%	/	0.2%	0.1%	/
DNA/PIF-Harbinger	2.4%	2.4%	2.2%	2.0%	2.4%
DNA/PIF-ISL2EU	0.3%	0.2%	0.3%	0.6%	0.2%
DNA/PiggyBac	0.1%	0.1%	0.1%	0.0%	/
DNA/Sola	/	/	/	0.1%	/
DNA/TcMar-ISRm11	0.1%	/	/	0.0%	0.1%
DNA/TcMar-Mariner	/	/	/	/	0.1%
DNA/TcMar-Tc1	19.3%	18.9%	19.5%	17.8%	22.3%
DNA/TcMar-Tc2	0.3%	0.5%	0.6%	0.7%	0.4%
DNA/TcMar-Tc4	/	/	/	0.2%	/
DNA/TcMar-Tigger	0.6%	0.7%	0.4%	0.8%	0.8%
DNA/Zisupton	/	/	/	/	0.0%
DNA/Unclassified	5.9%	7.2%	5.8%	7.2%	3.8%
LINE/CR1-Zenon	/	/	/	/	0.0%
LINE/Dong-R4	0.3%	0.4%	0.4%	/	0.1%
LINE/I	0.3%	0.5%	0.4%	0.3%	0.3%
LINE/I-Nimb	0.2%	0.1%	0.3%	/	0.1%
LINE/Jockey	/	/	/	0.1%	0.1%

LINE/L1	0.3%	0.2%	0.2%	0.3%	0.2%
LINE/L1-Tx1	0.1%	0.1%	0.2%	0.1%	0.0%
LINE/L2	4.4%	3.8%	3.8%	4.0%	4.2%
LINE/Penelope	0.1%	0.2%	0.1%	0.1%	0.2%
LINE/Proto2	0.2%	0.2%	0.1%	0.0%	0.0%
LINE/R1	0.0%	0.0%	/	/	/
LINE/Rex	/	0.0%	/	0.1%	/
LINE/Rex3	/	/	/	0.1%	/
LINE/Rex-Babar	1.8%	1.8%	2.0%	1.9%	2.5%
LINE/RTE	/	0.2%	/	0.1%	/
LINE/RTE-BovB	1.3%	1.0%	1.2%	1.1%	1.6%
LINE/RTE-X	/	0.0%	/	/	/
LINE/Tad1	/	/	/	0.3%	/
LINE/Unclassified	/	0.2%	0.1%	0.2%	0.1%
Low_complexity	1.1%	1.1%	1.1%	0.9%	0.9%
LTR/Copia	0.2%	0.2%	/	0.1%	0.0%
LTR/DIRS	0.1%	0.1%	0.0%	0.0%	0.1%
LTR/ERV1	/	0.0%	0.1%	/	/
LTR/ERVK	0.0%	/	/	/	/
LTR/Gypsy	1.0%	0.6%	0.5%	0.4%	1.0%
LTR/Gypsy-Cigr	/	0.1%	/	/	/
LTR/Ngaro	0.8%	1.1%	1.4%	1.4%	0.4%
LTR/Pao	0.1%	0.5%	0.3%	0.0%	/
LTR/Unclassified	0.4%	0.3%	0.5%	0.3%	0.3%
RC/Helitron	1.2%	1.2%	0.9%	1.4%	1.2%
Retroposon	0.3%	0.2%	0.4%	0.2%	0.1%
Satellite	/	0.0%	0.0%	0.0%	0.1%
Microsatellites	6.5%	6.7%	7.4%	6.1%	5.1%
SINE/MIR	0.2%	0.2%	0.4%	0.3%	0.4%
SINE/tRNA	0.5%	0.3%	0.5%	0.4%	0.5%
SINE/tRNA-Core-L2	0.2%	0.2%	0.0%	0.2%	0.1%
SINE/tRNA-L2	0.0%	/	0.0%	0.0%	0.0%
SINE/tRNA-V	0.4%	0.2%	/	1.0%	1.1%
SINE?	0.3%	0.3%	0.1%	0.0%	0.0%
SINE/Unclassified	0.0%	0.1%	/	0.5%	0.0%
Unknown	28.8%	29.3%	30.3%	26.3%	30.6%

Table S-6 **The content of repetitive elements in the order Cyprinodontiformes including: *Xiphophorus hellerii; Xiphophorus maculatus; Fundulus heteroclitus; Cyprinodon variegatus; Cyprinodon nevadensis.***

Types	***Xiphophorus hellerii***	***Xiphophorus maculatus***	***Fundulus heteroclitus***	***Cyprinodon variegatus***	***Cyprinodon nevadensis***
DNA/Academ	0.0%	/	/	0.4%	0.1%
DNA/CMC-Chapaev-3	0.1%	0.1%	/	/	/
DNA/CMC-EnSpm	0.7%	0.4%	0.6%	0.5%	0.3%
DNA/CMC-Transib	/	0.0%	/	/	/
DNA/Crypton	/	/	0.1%	0.0%	0.1%
DNA/Crypton-H	0.0%	/	/	/	/
DNA/Crypton-V	0.3%	/	/	/	0.0%
DNA/Dada	/	/	0.0%	/	/
DNA/Ginger	/	/	0.1%	/	/
DNA/hAT	1.7%	1.6%	0.8%	1.5%	1.0%
DNA/hAT-Ac	7.8%	8.1%	4.0%	5.3%	5.0%
DNA/hAT-Blackjack	/	0.1%	0.1%	0.9%	0.1%
DNA/hAT-Charlie	3.5%	3.9%	2.8%	3.0%	3.1%
DNA/hAT-hAT5	0.9%	1.0%	0.3%	0.4%	0.3%
DNA/hAT-hAT6	/	/	/	0.1%	0.2%
DNA/hAT-hATx	/	/	/	0.2%	0.0%
DNA/hAT-Tip100	1.2%	0.7%	0.6%	1.5%	1.6%
DNA/hAT-Tol2	1.3%	1.4%	0.5%	1.5%	1.5%
DNA/IS3EU	0.3%	0.3%	0.0%	0.7%	0.6%
DNA/Kolobok-T2	0.9%	1.0%	0.9%	0.2%	0.4%
DNA/Maverick	/	/	0.1%	/	/
DNA/MuLE-MuDR	0.1%	0.1%	0.3%	0.0%	0.0%
DNA/P	/	0.2%	0.0%	0.6%	0.1%
DNA/PIF-Harbinger	2.0%	2.3%	1.4%	1.4%	1.6%
DNA/PIF-HarbS	/	/	/	/	0.0%
DNA/PIF-ISL2EU	0.3%	0.3%	0.4%	0.1%	0.0%
DNA/PiggyBac	/	0.0%	0.0%	0.1%	0.2%
DNA/TcMar	/	/	/	0.0%	/
DNA/TcMar-ISRm11	/	/	0.1%	0.0%	0.0%
DNA/TcMar-Mariner	/	/	/	0.0%	0.0%
DNA/TcMar-Tc1	22.7%	22.0%	22.6%	8.2%	7.6%
DNA/TcMar-Tc2	0.4%	0.3%	1.1%	0.4%	0.3%
DNA/TcMar-Tigger	0.6%	0.3%	0.0%	0.0%	0.0%
DNA/Zisupton	0.0%	/	/	/	/
DNA/Unclassified	4.0%	6.9%	2.0%	4.6%	5.1%
LINE/Dong-R4	0.0%	0.0%	0.2%	0.4%	0.4%
LINE/DRE	/	/	0.0%	/	/

LINE/I	0.1%	0.1%	0.1%	0.2%	0.6%
LINE/I-Nimb	0.2%	0.0%	0.1%	0.2%	0.2%
LINE/Jockey	0.1%	0.2%	0.2%	0.0%	/
LINE/L1	0.2%	0.3%	1.8%	1.1%	0.9%
LINE/L1-Tx1	0.1%	0.0%	0.1%	0.1%	0.2%
LINE/L2	4.4%	4.0%	8.7%	8.5%	8.1%
LINE/LOA	/	/	/	/	0.0%
LINE/Penelope	0.1%	0.1%	1.2%	2.4%	2.1%
LINE/Proto2	0.0%	0.0%	0.0%	0.0%	0.0%
LINE/R1	/	/	/	0.0%	0.0%
LINE/R2	/	0.2%	/	/	/
LINE/R2-Hero	/	/	0.0%	/	/
LINE/Rex	0.1%	0.1%	0.0%	0.0%	0.0%
LINE/Rex3	/	0.0%	/	/	0.0%
LINE/Rex-Babar	2.5%	2.9%	4.6%	5.7%	4.7%
LINE/RTE	/	/	0.5%	0.2%	0.2%
LINE/RTE-BovB	1.5%	1.2%	1.6%	1.6%	0.6%
LINE/Tad1	/	/	/	/	0.1%
LINE/Unclassified	0.1%	/	0.2%	0.1%	0.1%
Low_complexity	0.9%	0.9%	0.5%	0.6%	0.5%
LTR/Copia	/	0.1%	0.1%	0.1%	/
LTR/DIRS	0.1%	0.0%	0.1%	0.1%	0.0%
LTR/ERV1	/	0.0%	0.4%	0.1%	0.1%
LTR/Gypsy	0.4%	0.4%	0.8%	1.8%	1.1%
LTR/Ngaro	1.4%	0.4%	1.6%	1.5%	1.4%
LTR/Pao	0.1%	0.0%	0.0%	0.0%	0.1%
LTR/Unclassified	0.3%	0.3%	0.1%	0.7%	0.0%
RC/Helitron	1.4%	0.8%	0.5%	0.1%	0.2%
Retroposon	0.1%	0.1%	0.3%	0.1%	0.2%
Satellite	/	/	0.6%	0.1%	0.3%
Microsatellites	5.1%	6.2%	4.6%	3.4%	2.5%
SINE/Alu	/	/	/	/	0.0%
SINE/ID	/	0.0%	/	/	/
SINE/MIR	0.3%	0.4%	0.7%	1.6%	1.3%
SINE/Ron1	/	/	/	0.0%	/
SINE/tRNA	0.7%	0.8%	0.2%	0.2%	0.0%
SINE/tRNA-Core-L2	0.3%	0.1%	0.2%	/	0.1%
SINE/tRNA-L2	0.0%	0.0%	/	/	/
SINE/tRNA-V	0.1%	0.2%	/	/	0.1%
SINE?	/	0.0%	0.0%	0.0%	0.0%
SINE/Unclassified	0.0%	/	0.5%	0.5%	0.5%
Unknown	30.7%	28.7%	30.3%	37.0%	43.8%

Table S-7 **The content of repetitive elements in the order Cypriniformes including: *Pimephales promelas; Sinocyclocheilus rhinocerous; Sinocyclocheilus grahami.***

Types	***Pimephales promelas***	***Sinocyclocheilus rhinocerous***	***Sinocyclocheilus grahami***
DNA/Academ	0.4%	0.0%	/
DNA/CMC-Chapaev-3	/	0.0%	0.1%
DNA/CMC-EnSpm	7.1%	7.0%	7.0%
DNA/Crypton	0.3%	0.2%	0.1%
DNA/Crypton-V	0.1%	0.2%	0.2%
DNA/Dada	0.0%	0.0%	/
DNA/hAT	1.7%	1.9%	1.8%
DNA/hAT-Ac	6.2%	7.9%	9.3%
DNA/hAT-Blackjack	0.1%	/	0.2%
DNA/hAT-Charlie	0.8%	2.3%	1.4%
DNA/hAT-hAT5	0.7%	0.3%	1.0%
DNA/hAT-hAT6	0.4%	0.1%	/
DNA/hAT-hobo	0.1%	0.0%	0.1%
DNA/hAT-Tag1	/	/	0.0%
DNA/hAT-Tip100	0.6%	0.4%	0.7%
DNA/hAT-Tol2	0.5%	0.4%	0.7%
DNA/IS3EU	1.5%	1.7%	1.0%
DNA/Kolobok	5.2%	4.4%	3.5%
DNA/Kolobok-T2	3.9%	4.7%	6.4%
DNA/Maverick	0.1%	0.2%	0.1%
DNA/Merlin	0.4%	0.2%	0.4%
DNA/MuLE-MuDR	0.1%	0.0%	0.0%
DNA/Novosib	/	/	0.1%
DNA/P	0.2%	0.1%	0.1%
DNA/PIF-Harbinger	3.7%	2.3%	1.8%
DNA/PIF-ISL2EU	0.3%	0.2%	0.2%
DNA/PiggyBac	0.9%	0.8%	0.7%
DNA/Sola	0.0%	0.0%	0.1%
DNA/TcMar	0.4%	0.2%	0.2%
DNA/TcMar-Fot1	0.5%	0.0%	0.0%
DNA/TcMar-ISRm11	0.2%	0.3%	0.2%
DNA/TcMar-Stowaway	0.1%	0.0%	0.0%
DNA/TcMar-Tc1	3.1%	3.4%	4.7%
DNA/TcMar-Tc2	0.2%	0.2%	0.1%
DNA/TcMar-Tigger	/	0.0%	/
DNA/Zator	0.0%	/	/
DNA/Zisupton	0.3%	0.3%	0.0%
DNA/Unclassified	10.5%	15.6%	12.4%
LINE/CR1	0.0%	/	/
LINE/CRE	0.0%	/	/

LINE/I	0.1%	0.2%	0.6%
LINE/I-Nimb	0.1%	0.2%	0.1%
LINE/Jockey	0.0%	/	0.1%
LINE/L1	0.1%	0.0%	0.1%
LINE/L1-Tx1	0.2%	0.2%	0.1%
LINE/L2	2.4%	3.1%	3.6%
LINE/Penelope	0.0%	0.0%	/
LINE/Proto2	0.0%	/	/
LINE/R1	/	0.0%	/
LINE/R2-Hero	0.3%	0.0%	0.0%
LINE/R2-NeSL	0.0%	/	/
LINE/Rex-Babar	0.2%	0.7%	0.9%
LINE/RTE-BovB	0.2%	/	/
LINE/RTE-RTEX	0.0%	/	/
LINE/RTE-X	0.0%	0.0%	0.1%
LINE/Tad1	/	0.0%	/
LINE/Unclassified	0.0%	0.0%	/
Low_complexity	0.7%	0.8%	0.6%
LTR/Copia	/	0.0%	0.1%
LTR/DIRS	3.1%	3.3%	4.3%
LTR/ERV	0.1%	0.1%	0.1%
LTR/ERV1	0.2%	0.1%	0.1%
LTR/ERVK	/	/	0.0%
LTR/Gypsy	1.3%	2.2%	2.8%
LTR/Ngaro	0.2%	0.2%	0.2%
LTR/Pao	0.1%	0.2%	0.1%
LTR/Unclassified	0.5%	0.7%	1.1%
RC/Helitron	1.5%	1.4%	1.4%
Retroposon	0.0%	0.0%	/
Satellite	5.2%	1.4%	2.9%
Microsatellites	6.7%	6.0%	5.7%
SINE/5S	/	0.0%	/
SINE/5S-Deu-L2	0.2%	0.1%	0.2%
SINE/ID	/	0.0%	/
SINE/L2	/	/	0.0%
SINE/MIR	0.4%	0.0%	0.1%
SINE/tRNA	0.0%	0.1%	0.1%
SINE/tRNA-L2	0.1%	0.2%	0.2%
SINE/tRNA-RTE	0.0%	/	/
SINE?	/	0.0%	0.0%
SINE/Unclassified	0.1%	0.4%	0.8%
Unknown	25.1%	22.6%	18.9%

Table S-8 **The content of repetitive elements in the order Cypriniformes including: *Sinocyclocheilus anshuiensis; Cyprinus carpio; Danio rerio.***

Types	***Sinocyclocheilus anshuiensis***	***Cyprinus carpio***	***Danio rerio***
DNA/Academ	0.1%	/	0.1%
DNA/CMC-Chapaev-3	0.0%	0.0%	/
DNA/CMC-EnSpm	8.8%	5.9%	11.1%
DNA/CMC-Transib	0.0%	/	/
DNA/Crypton	0.2%	0.3%	0.2%
DNA/Crypton-H	/	/	0.0%
DNA/Crypton-V	0.3%	0.1%	0.1%
DNA/Dada	0.0%	0.3%	0.0%
DNA/Ginger	/	0.0%	/
DNA/hAT	1.9%	1.0%	4.0%
DNA/hAT-Ac	8.3%	6.1%	7.3%
DNA/hAT-Blackjack	0.3%	/	/
DNA/hAT-Charlie	1.5%	1.6%	3.0%
DNA/hAT-hAT5	0.7%	0.5%	0.9%
DNA/hAT-hAT6	0.1%	0.3%	0.2%
DNA/hAT-hobo	0.1%	0.0%	0.0%
DNA/hAT-Tip100	0.5%	0.7%	0.7%
DNA/hAT-Tol2	0.6%	0.5%	0.2%
DNA/IS3EU	1.5%	1.3%	0.6%
DNA/Kolobok	2.5%	3.3%	0.4%
DNA/Kolobok-T2	5.3%	2.8%	5.0%
DNA/Maverick	0.1%	0.0%	0.1%
DNA/Merlin	0.2%	0.1%	0.8%
DNA/MuLE-MuDR	0.0%	0.0%	0.3%
DNA/MuLE-NOF	/	/	0.0%
DNA/Novosib	/	/	0.2%
DNA/P	0.1%	0.0%	0.3%
DNA/PIF-Harbinger	2.0%	2.1%	5.2%
DNA/PIF-ISL2EU	0.2%	0.1%	/
DNA/PiggyBac	0.8%	1.5%	0.7%
DNA/Sola	0.1%	0.1%	0.2%
DNA/TcMar	0.3%	0.8%	0.6%
DNA/TcMar-Fot1	0.0%	0.2%	0.1%
DNA/TcMar-ISRm11	0.3%	0.2%	0.2%
DNA/TcMar-Mariner	0.1%	/	/
DNA/TcMar-Stowaway	0.0%	0.0%	0.1%
DNA/TcMar-Tc1	3.2%	6.4%	6.1%
DNA/TcMar-Tc2	0.1%	/	/
DNA/TcMar-Tigger	/	/	0.1%
DNA/Zator	/	0.1%	/

DNA/Zisupton	0.2%	0.1%	0.1%
DNA/Unclassified	14.7%	14.6%	20.2%
LINE/I	0.1%	0.0%	0.1%
LINE/I-Nimb	0.3%	0.2%	/
LINE/Jockey	/	/	0.1%
LINE/L1	0.1%	0.3%	0.4%
LINE/L1-Tx1	0.2%	0.2%	0.4%
LINE/L2	3.2%	6.6%	3.2%
LINE/Penelope	0.0%	0.0%	0.1%
LINE/R1	0.0%	/	/
LINE/R2	/	/	0.2%
LINE/R2-Hero	0.0%	0.0%	/
LINE/Rex-Babar	0.8%	2.2%	0.6%
LINE/RTE	/	0.0%	0.0%
LINE/RTE-BovB	/	/	0.4%
LINE/RTE-X	0.1%	0.1%	0.0%
LINE?/Penelope	/	/	0.0%
LINE/Unclassified	/	0.0%	/
Low_complexity	0.8%	1.0%	0.5%
LTR/Copia	0.0%	0.0%	/
LTR/DIRS	3.2%	3.3%	1.9%
LTR/ERV	0.1%	0.2%	0.3%
LTR/ERV1	0.1%	0.1%	0.5%
LTR/ERVK	/	0.0%	/
LTR/Gypsy	2.6%	4.0%	2.3%
LTR/Gypsy-Cigr	0.0%	/	/
LTR/Ngaro	0.1%	0.3%	2.4%
LTR/Pao	0.2%	0.2%	0.1%
LTR/Unclassified	0.9%	0.7%	0.8%
RC/Helitron	1.4%	2.3%	3.5%
Retroposon	0.0%	0.1%	0.1%
rRNA	/	0.0%	/
Satellite	3.8%	2.9%	2.5%
Microsatellites	6.2%	7.1%	5.9%
SINE/5S-Deu-L2	0.2%	0.1%	0.8%
SINE/Alu	/	0.0%	0.1%
SINE/ID	0.2%	0.2%	/
SINE/L2	/	/	0.0%
SINE/MIR	0.1%	0.0%	/
SINE/tRNA	0.2%	0.1%	0.2%
SINE/tRNA-L2	/	0.1%	/
SINE/tRNA-V	/	/	0.1%
SINE?	0.0%	0.0%	0.1%
SINE/Unclassified	0.5%	0.9%	0.0%
Unknown	19.6%	15.2%	3.1%

Table S-9 **The content of repetitive elements in the order Scorpaeniformes including: *Cottus rhenanus; Sebastes nigrocinctus; Sebastes rubrivinctus.***

Types	***Cottus rhenanus***	***Sebastes nigrocinctus***	***Sebastes rubrivinctus***
DNA/Academ	/	0.1%	0.4%
DNA/CMC-Chapaev	/	0.0%	/
DNA/CMC-EnSpm	1.0%	1.1%	1.4%
DNA/Crypton	0.0%	0.8%	0.7%
DNA/Crypton-V	0.0%	0.0%	0.0%
DNA/Dada	/	/	0.0%
DNA/hAT	0.8%	3.7%	2.0%
DNA/hAT-Ac	3.0%	15.5%	15.6%
DNA/hAT-Blackjack	/	0.3%	0.0%
DNA/hAT-Charlie	2.4%	2.4%	2.3%
DNA/hAT-hAT5	0.4%	0.1%	0.0%
DNA/hAT-hobo	0.3%	0.1%	0.1%
DNA/hAT-Tip100	0.6%	0.2%	0.7%
DNA/hAT-Tol2	1.6%	1.6%	1.9%
DNA/IS3EU	0.1%	/	0.0%
DNA/Kolobok	/	/	0.1%
DNA/Kolobok-T2	/	0.4%	/
DNA/Kolobok-Hydra	/	/	0.0%
DNA/Kolobok-T2	/	/	0.1%
DNA/Maverick	/	/	0.3%
DNA/MuLE-MuDR	0.0%	0.3%	0.3%
DNA/Novosib	0.3%	/	/
DNA/P	0.0%	0.7%	0.7%
DNA/PIF-Harbinger	1.0%	2.2%	2.0%
DNA/PIF-ISL2EU	/	0.3%	0.1%
DNA/PiggyBac	0.2%	0.2%	0.1%
DNA/Sola	0.0%	/	/
DNA/TcMar	/	/	0.0%
DNA/TcMar-Fot1	/	0.0%	0.0%
DNA/TcMar-ISRm11	0.0%	0.1%	0.0%
DNA/TcMar-Tc1	0.9%	1.1%	1.0%
DNA/TcMar-Tc2	0.0%	0.0%	0.0%
DNA/Zisupton	/	0.0%	0.0%
DNA/Unclassified	5.3%	7.4%	8.1%
LINE/Dong-R4	/	/	0.0%
LINE/I	0.0%	0.1%	0.1%
LINE/I-Nimb	/	0.0%	0.0%
LINE/Jockey	/	0.0%	/
LINE/L1	0.5%	0.2%	0.5%
LINE/L1-Tx1	/	0.5%	0.2%

LINE/L2	3.9%	1.9%	1.8%
LINE/Penelope	0.1%	0.1%	0.1%
LINE/Proto2	0.0%	0.0%	0.0%
LINE/R2-Hero	/	0.0%	0.0%
LINE/Rex-Babar	2.7%	1.3%	1.4%
LINE/RTE	1.1%	/	/
LINE/RTE-BovB	1.2%	0.6%	0.5%
LINE/RTE-RTEX	0.0%	/	0.0%
LINE/RTE-X	0.0%	0.1%	0.1%
Low_complexity	2.2%	0.8%	0.9%
LTR/Copia	0.4%	/	0.2%
LTR/DIRS	/	0.0%	0.1%
LTR/ERV1	0.1%	/	/
LTR/ERVK	/	/	0.0%
LTR/Gypsy	1.7%	0.6%	0.6%
LTR/Ngaro	0.5%	/	0.4%
LTR/Pao	0.2%	0.0%	0.1%
LTR/Unclassified	0.1%	0.0%	/
RC/Helitron	1.0%	1.0%	0.9%
rRNA	0.1%	0.4%	0.5%
Satellite	0.0%	0.9%	0.7%
Microsatellites	17.8%	8.9%	9.1%
SINE/5S-Deu-L2	/	0.0%	/
SINE/7SL	/	0.0%	/
SINE/Alu	0.1%	/	/
SINE/ID	/	0.0%	0.0%
SINE/MIR	2.5%	0.4%	0.4%
SINE/tRNA	0.1%	0.2%	0.2%
SINE/tRNA-L2	0.3%	0.0%	/
SINE/tRNA-V	0.1%	0.0%	0.0%
SINE?	0.0%	0.0%	0.0%
SINE/Unclassified	0.7%	0.2%	0.1%
Unknown	44.5%	42.6%	42.7%

Table S-10 **The content of repetitive elements in the order Tetraodontiformes, contains *Tetraodon nigroviridis; Takifugu rubripes; Takifugu flavidus.***

Types	***Tetraodon nigroviridis***	***Takifugu rubripes***	***Takifugu flavidus***
DNA/CMC-Chapaev-3	/	0.3%	0.4%
DNA/CMC-EnSpm	0.4%	0.6%	0.4%
DNA/Crypton	/	0.0%	/
DNA/Dada	0.5%	0.2%	0.2%
DNA/Ginger	/	0.2%	0.2%
DNA/hAT	0.1%	0.3%	0.1%
DNA/hAT-Ac	1.5%	0.6%	0.7%
DNA/hAT-Blackjack	0.1%	0.1%	/
DNA/hAT-Charlie	5.1%	3.4%	3.3%
DNA/hAT-Tag1	0.1%	/	/
DNA/hAT-Tip100	1.0%	0.3%	0.4%
DNA/Kolobok-T2	0.4%	/	/
DNA/Maverick	0.1%	/	/
DNA/MuLE-MuDR	0.1%	0.2%	0.2%
DNA/MuLE-NOF	0.0%	/	/
DNA/Novosib	0.1%	/	/
DNA/P	0.1%	/	/
DNA/PIF-Harbinger	0.9%	2.0%	2.3%
DNA/PiggyBac	0.1%	0.2%	0.3%
DNA/TcMar-Fot1	/	0.1%	/
DNA/TcMar-ISRm11	0.1%	/	/
DNA/TcMar-Tc1	1.3%	3.3%	2.4%
DNA/TcMar-Tc2	3.0%	3.8%	3.7%
DNA/TcMar-Tigger	0.2%	0.7%	0.8%
DNA/Unclassified	0.7%	0.9%	1.2%
LINE/CR1	0.3%	/	/
LINE/Dong-R4	0.1%	1.7%	1.7%
LINE/I	0.9%	0.1%	0.1%
LINE/Jockey	0.2%	0.5%	/
LINE/L1	0.1%	0.6%	/
LINE/L1-Tx1	0.8%	1.9%	2.1%
LINE/L2	1.9%	8.6%	13.8%
LINE/Penelope	2.8%	2.0%	2.2%
LINE/R1	/	0.0%	/
LINE/R2-Hero	2.2%	0.3%	0.3%
LINE/Rex-Babar	7.7%	4.4%	4.8%
LINE/RTE	/	/	1.4%
LINE/RTE-BovB	3.7%	5.7%	4.2%
LINE/Unclassified	0.3%	/	0.1%
Low_complexity	3.1%	2.1%	2.0%

LTR/Copia	6.8%	/	0.2%
LTR/DIRS	0.5%	0.2%	0.2%
LTR/ERV1	0.9%	3.5%	4.2%
LTR/ERVK	/	/	0.2%
LTR/Gypsy	4.5%	6.7%	6.2%
LTR/Gypsy-Cigr	0.1%	/	0.1%
LTR/Ngaro	1.5%	0.3%	0.3%
LTR/Pao	0.0%	0.1%	0.1%
LTR/Viper	/	0.1%	0.1%
LTR/Unclassified	0.3%	0.5%	0.3%
RC/Helitron	0.2%	0.2%	/
rRNA	/	0.2%	0.2%
Satellite	0.6%	0.0%	0.3%
Microsatellites	31.1%	19.9%	21.8%
SINE/5S	0.1%	/	/
SINE/Alu	0.0%	/	/
SINE/MIR	/	0.0%	0.1%
SINE/tRNA	/	0.0%	0.1%
SINE/tRNA-V	/	0.7%	0.2%
SINE?	0.1%	0.1%	0.0%
SINE/Unclassified	0.1%	1.7%	1.8%
Unknown	13.4%	21.0%	14.5%

Table S-11 **The content of repetitive elements in the order Petromyzontiforme, contains *Lethenteron camtschaticum; Petromyzon marinus.***

Types	***Lethenteron camtschaticum***	***Petromyzon marinus***
DNA/Academ	0.1%	0.1%
DNA/CMC-Chapaev-3	1.0%	0.8%
DNA/CMC-EnSpm	0.6%	0.2%
DNA/hAT	0.2%	0.1%
DNA/hAT-Ac	0.4%	0.1%
DNA/hAT-Blackjack	0.1%	0.0%
DNA/hAT-Charlie	1.0%	1.6%
DNA/hAT-Tag1	0.0%	/
DNA/hAT-Tip100	0.8%	0.9%
DNA/Merlin	0.1%	/
DNA/MuLE-MuDR	0.0%	/
DNA/Novosib	/	0.1%
DNA/P	0.0%	0.0%
DNA/PIF-Harbinger	0.2%	0.1%
DNA/Sola	/	0.1%
DNA/TcMar-Mariner	0.0%	0.0%
DNA/TcMar-Tc1	0.6%	3.6%
DNA/TcMar-Tc2	0.1%	0.2%
DNA/TcMar-Tigger	1.6%	2.2%
DNA/Zator	0.9%	1.1%
DNA/Unclassified	2.6%	0.8%
LINE/CR1	0.0%	0.0%
LINE/CR1-Zenon	0.3%	0.4%
LINE/Jockey	1.7%	1.7%
LINE/L1	0.0%	0.1%
LINE/L1-Tx1	0.4%	0.3%
LINE/L2	5.0%	5.5%
LINE/Penelope	12.7%	14.3%
LINE/R2	0.0%	/
LINE/R2-Hero	0.1%	/
LINE/Rex-Babar	0.1%	0.3%
LINE/RTE-BovB	6.4%	4.9%
LINE/RTE-RTE	2.1%	2.2%
LINE/RTE-X	/	0.0%
LINE/Unclassified	0.7%	2.3%
Low_complexity	0.4%	0.2%
LTR/Copia	/	0.0%
LTR/DIRS	0.0%	/
LTR/Gypsy	5.6%	3.8%
LTR/Gypsy-Cigr	1.7%	1.5%

LTR/Ngaro	4.4%	3.7%
LTR/Pao	0.2%	0.2%
LTR/Unclassified	0.5%	1.4%
RC/Helitron	2.5%	2.2%
rRNA	0.1%	0.7%
Satellite	2.0%	7.0%
Microsatellites	5.4%	2.1%
SINE/5S	0.6%	0.2%
SINE/5S-Deu-L2	0.9%	0.3%
SINE/Alu	0.0%	0.2%
SINE/ID	/	0.1%
SINE/tRNA	8.2%	7.7%
SINE/tRNA-Core	2.0%	1.4%
SINE/tRNA-Deu	1.5%	1.7%
SINE/tRNA-Deu-L2	/	0.0%
SINE/tRNA-RTE	0.0%	/
SINE/tRNA-Sauria-RTE	0.3%	0.4%
SINE/tRNA-V	4.3%	3.9%
SINE?	/	0.0%
SINE/Unclassified	1.3%	0.4%
snRNA	/	0.0%
Unknown	18.0%	17.1%

Table S-12 **The content of repetitive elements in the *Oryzias latipes*, order Beloniformes; *Astyanax mexicanus*, order Characiformes and *Esox Lucius*, order Esociformes.**

Types	***Oryzias latipes***	***Astyanax mexicanus***	***Esox lucius***
DNA/Academ	0.1%	/	0.1%
DNA/CMC-Chapaev-3	/	/	0.0%
DNA/CMC-EnSpm	0.8%	1.3%	0.3%
DNA/Crypton	0.7%	0.1%	/
DNA/Dada	0.1%	0.1%	/
DNA/Ginger	0.6%	0.2%	0.0%
DNA/hAT	1.0%	1.2%	0.7%
DNA/hAT-Ac	2.7%	6.7%	1.8%
DNA/hAT-Blackjack	/	0.2%	0.1%
DNA/hAT-Charlie	7.0%	1.4%	1.2%
DNA/hAT-hAT5	0.1%	0.5%	0.1%
DNA/hAT-hAT6	/	0.0%	/
DNA/hAT-hATw	0.0%	/	/
DNA/hAT-Tip100	1.1%	0.4%	0.9%
DNA/hAT-Tol2	0.1%	0.9%	0.0%
DNA/IS3EU	0.3%	0.0%	/
DNA/Kolobok	/	/	0.2%
DNA/Kolobok-T2	/	0.4%	0.0%
DNA/Kolobok-T2	0.1%	/	/
DNA/Maverick	/	0.1%	1.5%
DNA/Merlin	/	/	0.3%
DNA/MuLE-MuDR	/	0.0%	0.0%
DNA/MuLE-NOF	/	/	0.1%
DNA/Novosib	/	/	0.2%
DNA/P	0.0%	0.0%	0.3%
DNA/PIF-Harbinger	4.3%	4.1%	1.2%
DNA/PIF-ISL2EU	0.1%	0.1%	0.3%
DNA/PiggyBac	7.0%	0.3%	0.1%
DNA/Sola	0.0%	0.0%	0.1%
DNA/TcMar	0.0%	0.2%	0.0%
DNA/TcMar-ISRm11	0.3%	0.1%	0.7%
DNA/TcMar-Mariner	/	0.7%	0.1%
DNA/TcMar-Stowaway	/	/	0.0%
DNA/TcMar-Tc1	5.0%	21.8%	35.9%
DNA/TcMar-Tc2	3.1%	3.4%	0.0%
DNA/TcMar-Tigger	1.0%	0.1%	/
DNA/Zisupton	/	0.0%	0.1%
DNA/Unclassified	3.8%	6.4%	2.3%
LINE/CR1-Zenon	0.5%	/	/
LINE/Dong-R4	0.7%	/	/

LINE/I	0.1%	0.0%	0.0%
LINE/I-Nimb	/	/	0.1%
LINE/Jockey	/	/	0.0%
LINE/L1	1.4%	0.0%	0.7%
LINE/L1-Tx1	0.5%	0.0%	0.1%
LINE/L2	5.9%	1.6%	6.9%
LINE/Penelope	0.3%	/	0.2%
LINE/Proto2	0.1%	/	/
LINE/R2	0.2%	/	/
LINE/R2-Hero	0.0%	0.1%	/
LINE/R2-NeSL	/	0.0%	0.1%
LINE/Rex-Babar	1.7%	0.9%	2.6%
LINE/RTE	0.3%	/	/
LINE/RTE-BovB	2.4%	0.0%	4.8%
LINE/RTE-RTEX	/	/	0.0%
LINE/RTE-X	0.0%	/	0.4%
LINE/Unclassified	0.0%	/	0.5%
Low_complexity	0.4%	1.1%	0.9%
LTR/Copia	0.1%	0.1%	0.3%
LTR/DIRS	/	0.0%	0.4%
LTR/ERV	/	0.0%	0.1%
LTR/ERV1	0.6%	/	0.5%
LTR/ERVK	0.1%	0.0%	0.0%
LTR/Gypsy	2.4%	1.2%	2.0%
LTR/Gypsy-Cigr	0.0%	/	/
LTR/Ngaro	2.4%	0.4%	/
LTR/Pao	0.0%	/	0.1%
LTR/Unclassified	0.0%	0.2%	0.0%
RC/Helitron	0.6%	1.2%	0.5%
Retroposon	0.1%	/	/
rRNA	0.1%	/	/
Satellite	0.0%	0.4%	1.0%
Microsatellites	2.6%	8.0%	4.8%
SINE/5S	0.0%	/	/
SINE/5S-Deu-L2	/	/	0.2%
SINE/MIR	1.1%	0.1%	/
SINE/tRNA	0.3%	2.5%	0.3%
SINE/tRNA-Core	0.1%	/	/
SINE/tRNA-Core-L2	0.2%	/	/
SINE/tRNA-L2	/	0.2%	0.5%
SINE/tRNA-V	0.2%	0.0%	/
SINE?	0.0%	0.0%	0.1%
SINE/Unclassified	0.2%	0.1%	0.0%
Unknown	34.7%	30.6%	23.0%

Table S-13 **The content of repetitive elements in the *Salmo salar*, order Salmoniformes; *Ictalurus punctatus*, order Siluriformes; *Gadus morhua*, order Gadiformes; *Lepisosteus oculatus* order Lepisosteiformes.**

Types	***Salmo salar***	***Ictalurus punctatus***	***Gadus morhua***	***Lepisosteus oculatus***
DNA/Academ	/	/	0.0%	/
DNA/CMC-Chapaev	/	/	0.0%	/
DNA/CMC-EnSpm	0.2%	2.1%	1.2%	/
DNA/Crypton	/	0.1%	0.8%	/
DNA/Crypton-V	/	0.0%	0.8%	0.0%
DNA/Dada	/	/	0.0%	/
DNA/Ginger	0.0%	1.8%	/	/
DNA/hAT	0.6%	1.3%	0.7%	0.3%
DNA/hAT-Ac	1.2%	3.0%	2.8%	0.3%
DNA/hAT-Blackjack	0.1%	0.2%	0.2%	0.1%
DNA/hAT-Charlie	1.2%	1.1%	1.4%	1.9%
DNA/hAT-hAT1	0.0%	/	/	/
DNA/hAT-hAT5	0.0%	0.1%	0.1%	/
DNA/hAT-hAT6	/	/	0.2%	/
DNA/hAT-hATw	/	/	0.1%	/
DNA/hAT-hobo	0.1%	0.0%	0.0%	/
DNA/hAT-Tip100	0.1%	0.2%	1.0%	0.3%
DNA/hAT-Tol2	0.5%	0.3%	0.1%	/
DNA/IS3EU	0.0%	0.2%	0.2%	0.1%
DNA/Kolobok	/	0.1%	/	/
DNA/Kolobok-T2	0.0%	/	/	/
DNA/Kolobok-Hydra	/	/	0.1%	/
DNA/Kolobok-T2	/	0.2%	0.5%	/
DNA/Maverick	0.0%	0.7%	0.0%	0.4%
DNA/Merlin	/	/	0.1%	/
DNA/MuLE-MuDR	0.1%	0.1%	0.3%	/
DNA/MuLE-NOF	/	0.0%	/	0.0%
DNA/Novosib	0.2%	0.0%	0.4%	0.0%
DNA/P	/	0.1%	0.2%	/
DNA/PIF-Harbinger	0.2%	0.9%	2.8%	0.1%
DNA/PIF-ISL2EU	0.0%	0.1%	0.1%	/
DNA/PiggyBac	0.6%	0.2%	0.0%	0.1%
DNA/Sola	0.1%	0.0%	1.1%	/
DNA/TcMar	0.0%	0.1%	/	/
DNA/TcMar-Fot1	0.1%	/	/	/
DNA/TcMar-ISRm11	0.2%	0.0%	0.2%	/
DNA/TcMar-Mariner	/	0.0%	/	/
DNA/TcMar-Stowaway	0.0%	/	/	/

DNA/TcMar-Tc1	23.6%	19.9%	0.4%	10.5%
DNA/TcMar-Tc2	0.2%	0.0%	0.2%	0.0%
DNA/TcMar-Tigger	/	0.2%	/	1.0%
DNA/Zisupton	0.1%	0.1%	0.0%	/
DNA/Unclassified	2.9%	2.9%	7.0%	1.3%
LINE/CR1	0.1%	/	/	9.4%
LINE/I	0.1%	0.1%	0.0%	/
LINE/I-Nimb	0.2%	0.1%	/	/
LINE/Jockey	0.2%	/	0.1%	0.0%
LINE/L1	0.5%	0.5%	0.1%	1.0%
LINE/L1-Tx1	0.7%	0.6%	0.1%	0.3%
LINE/L2	8.2%	4.3%	3.7%	5.4%
LINE/Penelope	0.2%	0.1%	0.1%	1.4%
LINE/Proto2	/	/	0.1%	/
LINE/R1	0.1%	0.3%	/	/
LINE/R2-Hero	/	0.0%	0.0%	0.0%
LINE/R2-NeSL	0.0%	/	0.1%	/
LINE/Rex-Babar	8.0%	1.0%	0.9%	4.4%
LINE/RTE-BovB	/	0.2%	0.8%	0.6%
LINE/RTE-RTEX	/	/	0.0%	0.0%
LINE/RTE-X	1.3%	0.0%	0.1%	1.0%
LINE/Unclassified	0.1%	0.0%	0.0%	0.2%
Low_complexity	0.8%	1.1%	3.0%	0.6%
LTR/Copia	0.0%	0.3%	0.2%	0.1%
LTR/DIRS	0.1%	2.2%	0.9%	/
LTR/ERV	/	0.1%	/	/
LTR/ERV1	0.1%	0.7%	0.2%	0.2%
LTR/ERVK	/	0.0%	0.0%	/
LTR/ERVL	0.0%	/	/	/
LTR/Gypsy	2.6%	2.3%	1.5%	2.3%
LTR/Gypsy-Cigr	/	/	0.3%	0.3%
LTR/Ngaro	/	3.1%	0.1%	8.0%
LTR/Pao	/	0.0%	0.2%	/
LTR/Viper	0.0%	0.0%	/	/
LTR/Unclassified	0.3%	0.1%	0.0%	0.2%
RC/Helitron	0.1%	1.1%	0.7%	0.1%
Retroposon	/	0.1%	/	/
rRNA	0.8%	0.3%	0.0%	0.0%
Satellite	1.8%	1.7%	0.9%	0.1%
Microsatellites	7.5%	14.1%	31.4%	3.0%
SINE/5S	/	0.3%	/	/
SINE/5S-Deu-L2	0.3%	0.2%	0.3%	8.8%
SINE/Alu	0.0%	0.1%	/	/
SINE/ID	0.1%	/	0.0%	/
SINE/MIR	0.1%	1.3%	1.1%	1.0%
SINE/tRNA	0.1%	0.0%	0.1%	0.5%

SINE/tRNA-L2	2.1%	0.0%	0.3%	0.3%
SINE/tRNA-RTE	/	/	/	0.0%
SINE/tRNA-V	/	0.5%	/	0.6%
SINE/tRNA-V-CR1	/	0.1%	/	/
SINE/U	/	0.0%	/	/
SINE?	1.0%	0.3%	0.2%	0.2%
SINE/Unclassified	0.0%	/	0.0%	0.5%
snRNA	0.0%	0.0%	0.0%	/
Xba1	/	3.2%	/	/
Unknown	30.3%	23.2%	29.2%	32.9%

Table S-14 **The content of repetitive elements in the *Clupea harengus*, order Clupeiformes; *Latimeria chalumnae*, order Coelacanthiformes; *Cynoglossus semilaevis*, order Pleuronectiformes; *Gasterosteus aculeatus* order Gasterosteiformes.**

Types	***Clupea harengus***	***Latimeria chalumnae***	***Cynoglossus semilaevis***	***Gasterosteus aculeatus***
DNA/Academ	0.6%	/	0.3%	/
DNA/CMC-EnSpm	2.3%	0.0%	3.4%	3.1%
DNA/Crypton	0.5%	0.0%	/	/
DNA/Crypton-V	/	/	/	0.6%
DNA/Dada	0.1%	/	/	/
DNA/Ginger	0.5%	0.0%	/	0.1%
DNA/hAT	0.9%	0.0%	1.3%	1.1%
DNA/hAT-Ac	3.5%	0.5%	6.7%	4.8%
DNA/hAT-Blackjack	0.4%	0.2%	0.1%	/
DNA/hAT-Charlie	1.3%	0.3%	15.1%	4.2%
DNA/hAT-hAT5	0.1%	0.1%	0.0%	0.3%
DNA/hAT-hAT6	0.0%	0.0%	/	0.2%
DNA/hAT-hobo	0.1%	/	/	/
DNA/hAT-Tag1	/	0.0%	/	/
DNA/hAT-Tip100	1.8%	0.5%	0.6%	1.2%
DNA/hAT-Tol2	0.5%	/	0.3%	0.6%
DNA/IS3EU	0.0%	/	0.0%	0.3%
DNA/Kolobok-T2	0.4%	/	/	0.1%
DNA/Maverick	0.1%	/	/	0.4%
DNA/Merlin	/	/	/	0.3%
DNA/MuLE-MuDR	/	0.0%	/	/
DNA/Novosib	1.3%	/	/	/
DNA/PIF-Harbinger	1.2%	0.6%	0.3%	2.8%
DNA/PIF-ISL2EU	0.1%	/	/	0.1%
DNA/PiggyBac	0.1%	/	0.4%	0.1%
DNA/Sola	2.6%	0.0%	0.1%	/
DNA/TcMar	0.0%	0.0%	/	/
DNA/TcMar-IS885	/	/	/	0.1%
DNA/TcMar-ISRm11	0.5%	/	/	0.3%
DNA/TcMar-Mariner	0.0%	/	0.5%	/
DNA/TcMar-Pogo	/	0.0%	/	/
DNA/TcMar-Tc1	3.0%	0.0%	2.7%	4.1%
DNA/TcMar-Tc2	0.4%	/	2.5%	0.9%
DNA/TcMar-Tigger	0.0%	0.0%	3.9%	0.0%
DNA/Zator	0.0%	0.0%	/	/
DNA/Unclassified	8.5%	40.8%	1.0%	5.7%
LINE/CR1	/	5.9%	/	/
LINE/CR1-Zenon	/	0.2%	/	/

LINE/CRE	/	0.0%	/	/
LINE/DRE	0.3%	/	/	/
LINE/I	0.0%	/	0.6%	/
LINE/I-Nimb	/	/	/	0.2%
LINE/Jockey	/	0.1%	0.6%	/
LINE/L1	0.1%	2.6%	/	0.1%
LINE/L1-Tx1	0.2%	0.5%	/	0.6%
LINE/L2	5.0%	3.8%	1.1%	8.4%
LINE/Penelope	/	0.4%	1.3%	0.4%
LINE/Proto2	/	/	/	0.1%
LINE/R1	0.1%	/	/	/
LINE/R2	/	/	/	0.0%
LINE/R2-Hero	0.7%	/	0.3%	/
LINE/R2-NeSL	0.1%	/	/	/
LINE/Rex3	/	/	/	0.0%
LINE/Rex-Babar	2.0%	/	1.4%	5.8%
LINE/RTE	/	/	1.9%	/
LINE/RTE-BovB	0.1%	0.8%	4.4%	1.8%
LINE/RTE-RTE	/	0.3%	/	/
LINE/RTE-RTEX	0.0%	/	/	/
LINE/RTE-X	0.4%	/	/	/
LINE/Unclassified	/	1.8%	0.0%	0.1%
Low_complexity	3.8%	0.2%	3.3%	1.5%
LTR/Copia	0.1%	/	/	0.1%
LTR/DIRS	0.7%	1.3%	/	0.2%
LTR/ERV	0.2%	/	/	/
LTR/ERV1	0.1%	/	0.4%	3.3%
LTR/ERVK	/	/	0.0%	0.0%
LTR/ERVL	/	/	0.0%	/
LTR/Gypsy	0.7%	0.3%	0.4%	6.1%
LTR/Gypsy-Cigr	0.1%	/	/	0.1%
LTR/Ngaro	0.3%	/	0.3%	0.5%
LTR/Pao	0.2%	/	/	1.4%
LTR/Unclassified	0.1%	2.1%	/	0.2%
RC/Helitron	1.0%	0.1%	/	0.5%
Retroposon	0.0%	/	0.1%	/
rRNA	/	0.0%	/	0.4%
Satellite	0.5%	0.1%	0.8%	0.3%
Microsatellites	29.8%	1.7%	23.1%	12.8%
SINE/5S-Deu-L2	/	1.8%	/	/
SINE/Alu	/	/	/	0.0%
SINE/ID	0.0%	/	0.0%	0.0%
SINE/L2	0.4%	/	/	/
SINE/MIR	0.0%	0.1%	/	0.8%
SINE/tRNA	0.2%	7.5%	0.1%	0.0%
SINE/tRNA-Core	/	/	0.1%	/

SINE/tRNA-Core-L2	0.1%	/	/	/
SINE/tRNA-Deu-L2	/	0.2%	/	/
SINE/tRNA-L2	0.4%	/	0.2%	/
SINE/tRNA-V	/	/	0.5%	0.7%
SINE/U	/	/	0.0%	/
SINE?	0.1%	/	0.8%	0.6%
SINE/Unclassified	0.0%	13.2%	2.4%	0.0%
snRNA	/	/	/	0.2%
Unknown	21.5%	11.4%	16.7%	21.5%

Table S-15 **The content of repetitive elements in the *Rhincodon typus*, order Orectolobiformes; *Callorhinchus milii*, order Chimaeriformes; *Leucoraja erinacea*, order Rajiformes.**

Types	***Rhincodon typus***	***Callorhinc hus milii***	***Leucoraja erinacea***
DNA/Academ	0.0%	/	0.0%
DNA/CMC-EnSpm	0.6%	0.2%	0.5%
DNA/Crypton-V	/	/	0.0%
DNA/Dada	/	/	0.0%
DNA/Ginger	0.8%	0.3%	/
DNA/hAT	/	0.0%	0.1%
DNA/hAT-Blackjack	0.0%	0.0%	0.1%
DNA/hAT-Charlie	0.1%	0.1%	0.3%
DNA/hAT-hAT6	/	/	0.1%
DNA/hAT-hATw	/	/	0.0%
DNA/hAT-hATx	/	0.0%	/
DNA/hAT-Pegasus	/	/	0.0%
DNA/hAT-Tip100	/	0.0%	0.2%
DNA/Kolobok-Hydra	0.5%	/	/
DNA/MuLE-MuDR	0.2%	/	0.0%
DNA/Novosib	/	0.0%	0.0%
DNA/P	/	/	0.0%
DNA/PIF-Harbinger	0.5%	0.0%	0.1%
DNA/Sola	0.5%	0.2%	/
DNA/TcMar	/	0.0%	/
DNA/TcMar-Pogo	0.0%	0.0%	/
DNA/TcMar-Tc1	/	/	0.7%
DNA/TcMar-Tc2	0.1%	0.0%	0.0%
DNA/TcMar-Tigger	0.3%	0.0%	0.4%
DNA/Unclassified	0.9%	1.0%	0.4%
LINE/CR1	39.4%	23.0%	28.4%
LINE/CR1-Zenon	9.2%	0.1%	0.3%
LINE/Dong-R4	/	0.1%	/
LINE/DRE	0.0%	/	/
LINE/I	0.0%	/	0.0%
LINE/Jockey	0.1%	/	0.4%
LINE/L1	/	/	0.9%
LINE/L1-Tx1	0.6%	0.1%	0.5%
LINE/L2	1.5%	32.2%	0.5%
LINE/Penelope	8.1%	0.1%	4.3%
LINE/Proto2	/	/	0.1%
LINE/R2	0.0%	/	/
LINE/R2-Hero	/	/	0.2%
LINE/Rex-Babar	1.7%	/	7.1%
LINE/RTE-BovB	0.7%	/	1.0%

LINE/RTE-X	/	0.0%	0.0%
LINE/Unclassified	0.2%	0.6%	0.2%
Low_complexity	0.3%	0.4%	0.4%
LTR/Copia	0.4%	/	0.0%
LTR/DIRS	/	/	1.7%
LTR/ERV1	0.6%	0.1%	0.2%
LTR/ERVL	0.4%	/	/
LTR/Gypsy	3.6%	0.1%	1.9%
LTR/Gypsy-Cigr	0.0%	/	/
LTR/Ngaro	4.0%	/	10.0%
LTR/Pao	1.1%	/	0.3%
LTR/Unclassified	0.0%	/	0.0%
RC/Helitron	/	0.0%	0.3%
rRNA	0.0%	0.0%	/
Satellite	0.1%	0.2%	0.1%
Microsatellites	1.4%	3.5%	1.4%
SINE/5S-Deu-L2	/	/	0.0%
SINE/Alu	0.0%	/	/
SINE/MIR	0.0%	0.1%	/
SINE/tRNA	0.0%	0.3%	0.0%
SINE/tRNA-Core	/	0.0%	/
SINE/tRNA-Core-L2	/	/	0.0%
SINE/tRNA-Deu	/	0.0%	/
SINE/tRNA-Deu-L2	2.0%	18.7%	0.2%
SINE/tRNA-RTE	/	/	0.0%
SINE/tRNA-V-CR1	0.1%	/	2.7%
SINE?	0.2%	0.2%	0.0%
SINE/Unclassified	0.5%	0.9%	0.0%
snRNA	/	/	0.0%
Unknown	18.9%	17.2%	33.7%