BMC Genomics

RESEARCH ARTICLE

Open Access

CrossMark

Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments

Zihao Yuan[1], Shikai Liu[1], Tao Zhou[1], Changxu Tian[1], Lisui Bao[1], Rex Dunham[1] and Zhanjiang Liu[2*]

Abstract

Background: 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, we conducted a comparative analysis of repetitive elements of 52 fish species in 22 orders in relation to their living aquatic environments.

Results: The proportions of repetitive elements in various genomes were found to be positively correlated with genome sizes, with a few exceptions. More importantly, there appeared to be specific enrichment between some repetitive element categories with species habitat. Specifically, class II transposons appear to be more abundant in freshwater bony fish than in marine bony fish when phylogenetic relationship is not considered. In contrast, marine bony fish harbor more tandem repeats than freshwater species. In addition, class I transposons appear to be more abundant in primitive species such as cartilaginous fish and lamprey than in bony fish.

Conclusions: 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 needs to be done to alleviate the phylogenetic biases.

Keywords: Fish, Evolution, Repeat, Transposon, Microsatellite, Habitat

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 or transposons and the tandem repeats. Transposons can be divided into RNA-mediated class I transposons, which include transposons with long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs); and RNA-independent class II DNA transposons. Tandem repeats are copies of DNA repeats located adjacent to one other [1–3]. 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 [4, 5].

Although repetitive elements were considered to be junk DNA [6], recent studies suggested that they are functional in regulating gene expression and contribute to genome evolution [7–11]. Transposons 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 [12–14], and thus can be the major reason for species difference in genome size [15–17]. Similarly, expansion or contraction of tandem repeats can also affect genome size [18–20], and consequently affect recombination, gene expression, and conversion and chromosomal organization [21–26].

---

* Correspondence: johnliu@syr.edu
[2]Department of Biology, College of Art and Sciences, Syracuse University, Syracuse, NY 13244, USA
Full list of author information is available at the end of the article

Yuan *et al. BMC Genomics* (2018) 19:141

Page 2 of 10

Fish comprise a large and highly diverse group of vertebrates inhabiting a wide range of different aquatic environments [27]. Sequenced fish genomes vary in size from 342 Mb of *Tetraodon nigroviridis* to 2967 Mb of *Salmo salar*. Some studies have been conducted on the diversity of repetitive elements in fish [28–30], but systematic comparative studies have 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, we annotated the repetitive elements of 52 fish genomes from 22 orders, and determined their distribution in relationship with environmental adaptations. Here, we observed the correlation between high numbers of DNA transposons, especially the Tc1 transposons, with freshwater bony fish, high level of microsatellites with marine bony fish, and high numbers of class I transposons with cartilaginous fish and lamprey. 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 (PIC).
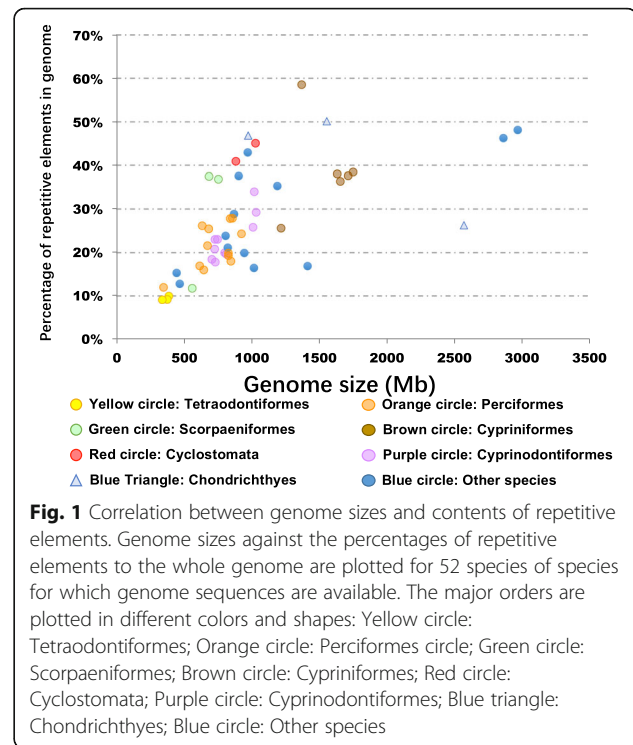
Results

Contents of repetitive elements in various fish genomes

A total of 128 categories of repetitive elements are identified from the 52 fish species (Additional file 1: Table S1). We found overall positive correlation between contents of repetitive elements in fish and their genome sizes. This correlation, was still significant when implementing phylogenetically independent contrasts (Fig. 1, PIC *p*-value: 1.88e-03, Pearson correlation r = 0.6, *p*-value = 1.45e-06). 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.
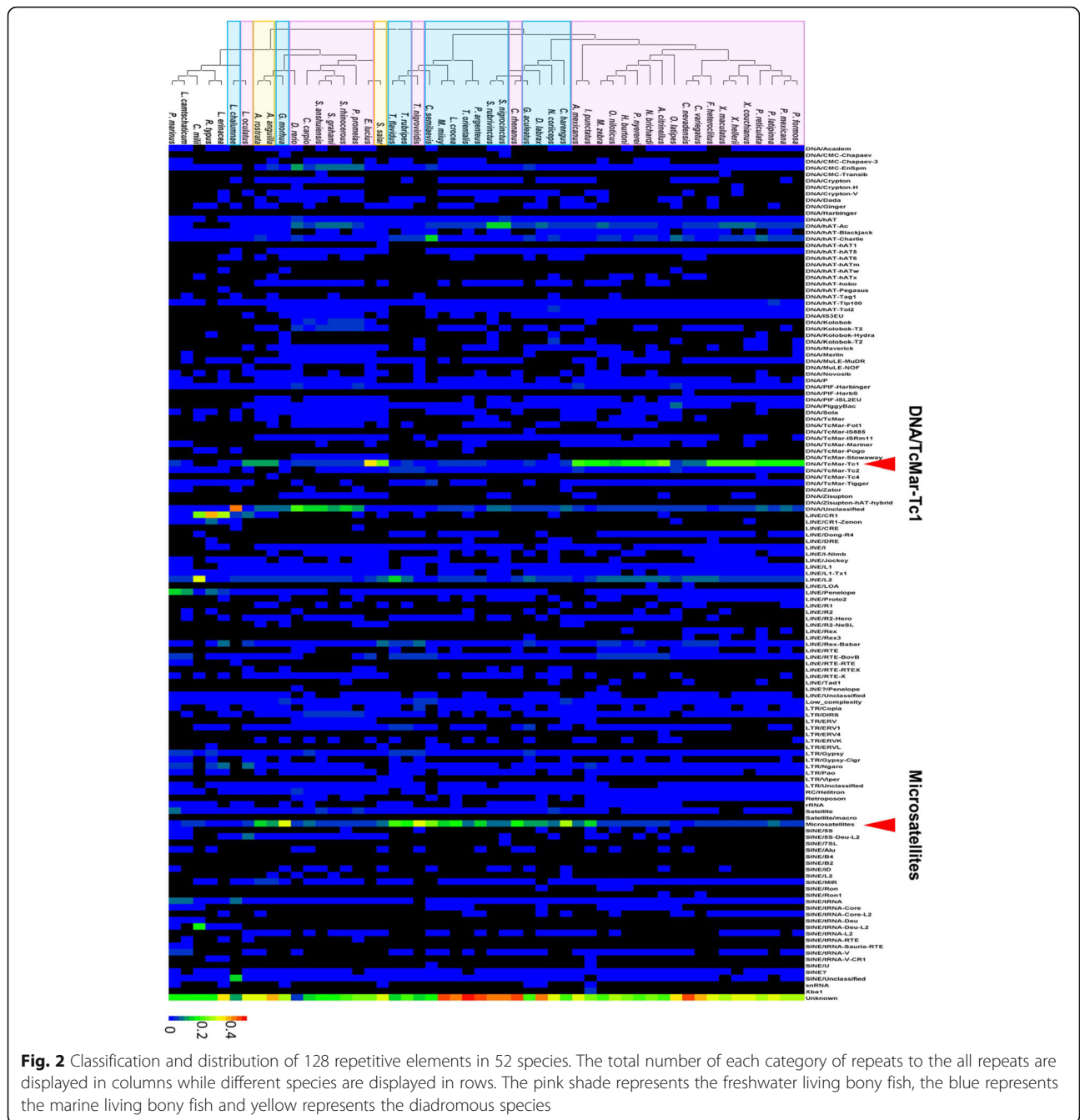
Differential associations of repetitive elements across species

We investigated the possible association between repetitive elements and aquatic environment. Comparison of diversity and abundance of repetitive elements across the 52 fish genomes revealed significant differences among species (Fig. 2 and Additional file 2: Table S2). Class I transposons are more prevalent in cartilaginous fish and lampreys than bony fish species (Wilcoxon rank test, *p*-value = 1.41e-04). For example, class I transposons represent 76.6% of repetitive elements in elephant shark, but the bony fish genomes are more abundant with class II transposons and tandem repeats.

Fig. 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

Of the bony fish genomes, the freshwater bony fish contained a greater proportion of Tc1/mariner transposons than marine species (Fig. 2, Wilcoxon rank test, *p*-value = 8.23e-06). However, the results were not significant when the phylogeny was taken into consideration (PIC p- value: 0.117). 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 transposons and microsatellites (Table 1).

Analysis of the sequence divergence rates suggest that Tc1 transposons have been present in the genomes of freshwater species for much a longer period of time or are more active than in marine species (Fig. 3). The Tc1 transposons 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 transposons 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 (Fig. 3). The long history and high transposition activities in freshwater fish accounted, at least in part, for the high proportion of Tc1 transposons in the genomes of freshwater species.

Yuan *et al. BMC Genomics* (2018) 19:141

Page 3 of 10

Fig. 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

Discussion

Accumulation of repetitive elements in fish genomes

In this work, we determined the correlation between the categories and proportions of repetitive elements and the living environments of various fish species. We found 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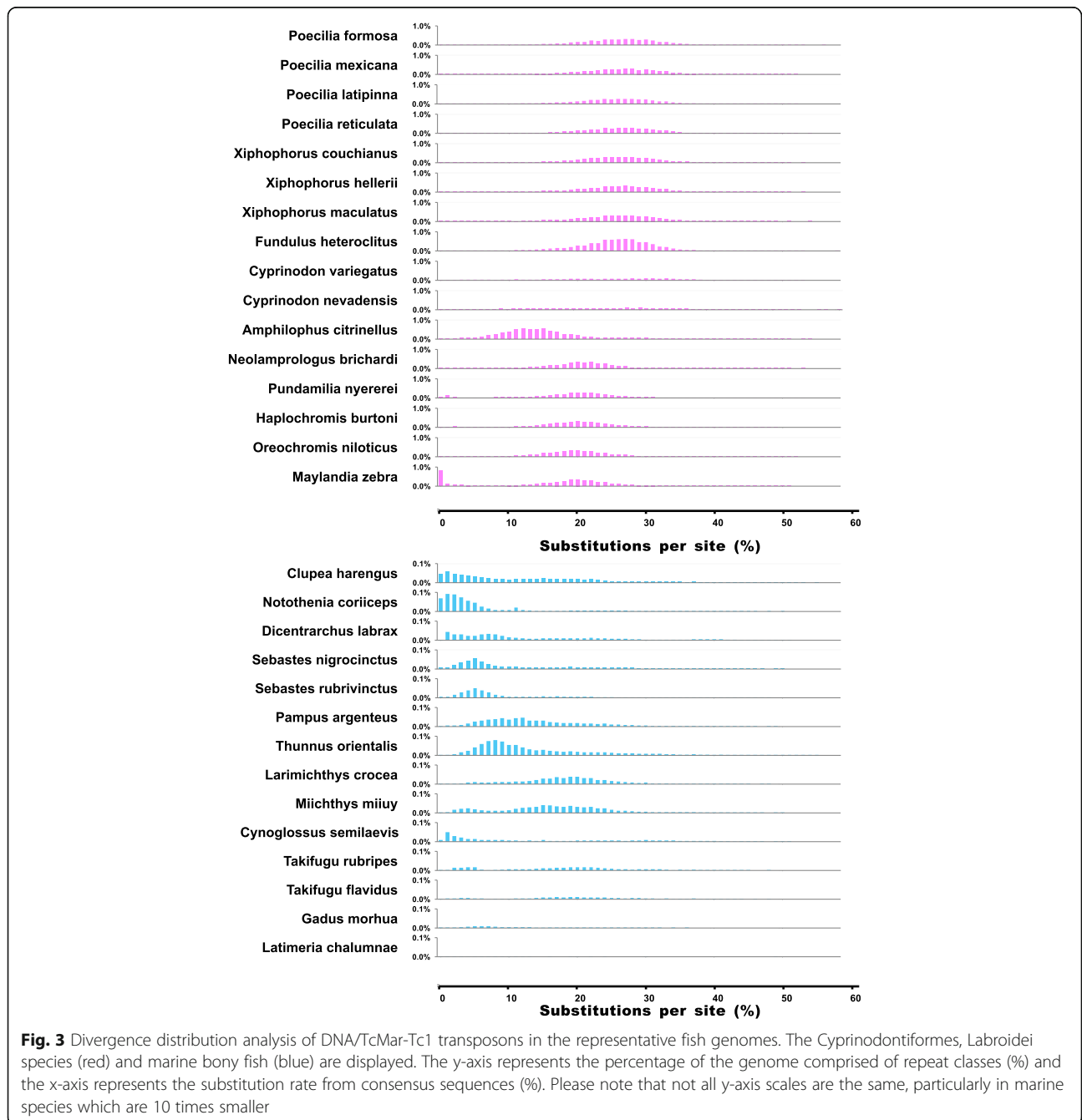 freshwater bony fish, independent of phylogenetic relationship. In addition, class I transposons are more abundant in primitive species such as cartilaginous fish and lamprey than in bony fish. Such findings suggest that these repetitive elements are related to the adaptability of fish to their living environments, 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).

Yuan *et al. BMC Genomics* (2018) 19:141

Page 4 of 10

Table 1 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%

Yuan *et al. BMC Genomics* (2018) 19:141

Page 5 of 10

Table 1 Proportion of DNA/TcMar-Tc1, microsatellites contents out of all repeats in freshwater, marine and diadromous teleost species *(Continued)*

Species	Order	DNA/TcMar-Tc1	Microsatellites
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%

Fig. 3 Divergence distribution analysis of DNA/TcMar-Tc1 transposons 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

Yuan *et al. BMC Genomics* (2018) 19:141

Page 6 of 10

With teleost fish, the genome sizes are greatly affected by the teleost-specific round of whole genome duplication [31–33]. 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 as observed in our analysis, fish genome sizes, with exceptions, were found to be 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 [34]. 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 should be small. In addition, 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. The total number of genomes used in the study is relatively large (52), the impact of sequence assembly quality should have been minimized.

Comparison of the repetitive elements among species

The distributions of repetitive elements are significantly associated with various clades during evolution. For example, class I transposons are more prevalent in cartilaginous fish and lampreys than in bony fish species. However, the cartilaginous fish and lamprey lack the class II transposons. Although there were no unifying explanations for this difference, it is speculated that it may be related to the internal fertilization of cartilaginous fish, which may have minimized the exposure of gametes and embryos from horizontal transfer of Class II transposons [30, 35, 36]. Interestingly, active transposable elements in mammals are also RNA transposons. For lamprey, since it is still unclear how it fertilizes and develops in the wild [37, 38], its accumulation of class I transposons deserve further investigation. As class I transposons are involved in various biological processes such as regulation of gene expression [39, 40], the ancient accumulation of class I transposons in cartilaginous fish and lamprey are probably related to their evolutionary adaptations [41]. The contents of class I transposons are low in bony fish; the exact reasons are unknown, but could involve putative mechanisms that counteract the invasiveness of RNAs on their genomes. We realized that a much larger number of bony fish genomes are used in this study than those from cartilaginous fish and lamprey, 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.

Repetitive elements of most freshwater bony fish are dominated by DNA transposons except *C. rhenanus* and *T. nigroviridis* which contain high levels of microsatellites. Although *T. nigroviridis* is a freshwater species, the vast majority (497 out of 509) of species in Tetraodontidae family are marine species [42–44]. 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 [43]. In addition, the biology of *C. rhenanus* is largely unknown [45, 46], and the origin of *C. rhenanus* as a freshwater species remains unexplained.

Uncovering the route of class II transposons expansion is difficult, because they can be transferred both vertically and horizontally [47–49]. However, when phylogenic relationships were not considered, the observed prevalent class II transposon in freshwater species may indicate that the freshwater environments are more favorable for proliferation and spreading of DNA transposons. In addition, as found 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 [50]. Previous studies showed that freshwater ray-finned fish have smaller effective population sizes and larger genome sizes than marine species [51]. Our results lend additional support to the idea that shrinking effective population sizes may have underlined the evolution of more complex genomes [52, 53]. The significance for more prevalence of Tc1 transposon in freshwater species was reduced when accounting for phylogenetic relationship, which indicates the taxa in our data set for 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 will inevitably introduce phylogenetic bias into the analysis. For example, a large number of the sequenced fish species belong to 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 [54, 55]. Considering the fact that the phylogenetic independent contrasts analysis is robust to random species sampling [56], thus, further analysis should be conducted

Yuan *et al. BMC Genomics* (2018) 19:141

Page 7 of 10

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 [57–59]. Thus we still classify the *G. aculeatus* as marine species. Because 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 [60, 61], thus, the high salinity environment may be prone to DNA polymerase slippage while not favorable for proliferation and spreading of transposons, since previous studies indicated that the higher salt concentration might stabilize the hairpin structure during the DNA polymerase slippage [62]. 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.

Conclusions
In this study, we investigated the diversity, abundance, and distribution of repetitive elements among 52 fish species in 22 orders. Differential associations of repetitive elements were found from various clades and their living environments. Class I transposons are abundant in lamprey and cartilaginous fish, but less so in bony fish. Tc1/mariner transposons 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 transposons. 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.

Methods
Annotation of repetitive elements in fish genome assemblies
The channel catfish genome was assembled by our group [54], the genome sequences of other 51 species were retrieved from NCBI or Ensembl databases [33, 42, 56, 63–89] (Additional file 1: Table S1). The repetitive elements were identified using RepeatModeler 1.0.8 containing RECON [90] and RepeatScout with default parameters [91]. The derived repetitive sequences were searched against Dfam [92] and Repbase [93]. If the sequence is classified as "Unknown", they were further searched against the NCBI-nt database using blastn 2.2.28 + .

Phylogenetic analysis
The phylogenetic analysis was based on the cytochrome b [94]. Multiple alignments were conducted by MAFFT [95]. The best substitution model was selected by Prottest 3.2.1 [96]. The phylogenetic tree was constructed using MEGA7 with the maximum likelihood method [97], using JTT with Freqs. (+ F) model, and gaps were removed by partial deletion. The topological stability was evaluated with 1000 bootstraps.

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: $K = -300/4 \times Ln(1-D \times 4/300)$, the D represents the proportion of each DNA/TcMar-Tc1 fragment differ from the consensus sequences [98].

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 [99]. The Pearson correlation analysis in Excel was applied for the correlation between genome size and the content of repetitive elements. Based on the phylogeny tree of the species generated in the previous method, the phylogenetically independent contrasts between the environments and different characters was conducted to evaluate the bias of the phylogeny. The freshwater and sea water was represented by their respective salinities (0.5 for freshwater and 35 for seawater) [100]. The phylogenetically independent contrast test was conducted via the "drop.tip ()" and "pic ()" function in ape package provided by R [101]. The heat map was plotted using the Heml1.0 [102].

Additional files

Additional file 1: Table S1. Fish genomes used for analysis. (DOCX 33 kb)

Additional file 2: Table S2. Distribution of repetitive elements among species. (XLS 96 kb)

Abbreviations
DIAG: Data Intensive Academic Grid; ENA: EMBL Nucleotide Sequence Database; K: Average number of substitutions per sites; LINEs: Long interspersed nuclear elements; LTRs: Long terminal repeats; PIC: Phylogenetically independent contrasts; SINEs: Short interspersed nuclear elements

Yuan *et al. BMC Genomics* (2018) 19:141

Page 8 of 10

Availability of data and materials

The datasets analyzed during the current study are available in the Genbank, https://www.ncbi.nlm.nih.gov/genbank/; EMBL Nucleotide Sequence Database (ENA), http://www.ebi.ac.uk/ena/, all genome accessions are included in this published article (Additional file 1: Table S1).

Authors' contributions

ZY performed the major part of data analysis of this work and drafted the manuscript. SL, TZ, CT and LB contributed to the data analysis and manuscript preparation. RD and ZL supervised the whole study and revised the manuscript. All authors have read and approved the manuscript for submission.

Ethics approval and consent to participate

This study is a retrospective analysis of the public available data and therefore no ethics approval was needed. The Genome sequences are downloaded and cited from Genbank, https://www.ncbi.nlm.nih.gov/genbank/; EMBL Nucleotide Sequence Database (ENA), http://www.ebi.ac.uk/ena/ as outlined in the additional file (Additional file 1: Table S1).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests exist.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]The Fish Molecular Genetics and Biotechnology Laboratory, Aquatic Genomics Unit, School of Fisheries, Aquaculture and Aquatic Sciences and Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, USA. [2]Department of Biology, College of Art and Sciences, Syracuse University, Syracuse, NY 13244, USA.

References

1. Kubis S, Schmidt T, Heslop-Harrison JSP. Repetitive DNA elements as a major component of plant genomes. Ann Bot. 1998;82:45–55.
2. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 2000;10:967–81.
3. Ugarković Đ, Plohl M. Variation in satellite DNA profiles—causes and effects. EMBO J. 2002;21:5955–9.
4. Hacch F, Mazrimas J. Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys Ordii*). Nucleic Acids Res. 1974;1:559–76.
5. Petitpierre E, Juan C, Pons J, Plohl M, Ugarkovic D. Satellite DNA and constitutive heterochromatin in tenebrionid beetles. In: Kew chromosome conference IV: Royal Botanic Gardens; London. 1995. p. 351-62.
6. Ohno S. So much "junk" DNA in our genome. In: Brookhaven symposia in biology; 1972. p. 366–70.
7. Meagher TR, Vassiliadis C. Phenotypic impacts of repetitive DNA in flowering plants. New Phytol. 2005;168:71–80.
8. Schmidt AL, Anderson LM. Repetitive DNA elements as mediators of genomic change in response to environmental cues. Biol Rev. 2006;81:531–43.
9. 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. Whole-genome sequence of the Tibetan frog Nanorana Parkeri and the comparative evolution of tetrapod genomes. Proc Natl Acad Sci. 2015;112:E1257–62.
10. Thornburg BG, Gotea V, Makałowski W. Transposable elements as a significant source of transcription regulating signals. Gene. 2006;365:104–10.
11. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C. The locust genome provides insight into swarm formation and long-distance flight. Nat Commun. 2014;5:2957.
12. Hurst GD, Werren JH. The role of selfish genetic elements in eukaryotic evolution. Nat Rev Genet. 2001;2:597–606.
13. Kazazian HH. An estimated frequency of endogenous insertional mutations in humans. Nat Genet. 1999;22:130.
14. Kazazian HH. Mobile elements: drivers of genome evolution. Science. 2004; 303:1626–32.
15. Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. Genomics Inform. 2014;12:87–97.
16. SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N. Nested retrotransposons in the intergenic regions of the maize genome. Science. 1996;274:765–8.
17. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
18. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 1994;371:215–20.
19. Lindahl T. DNA repair: DNA surveillance defect in cancer cells. Curr Biol. 1994;4:249–51.
20. Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 1993;365:274–6.
21. Balaresque P, King TE, Parkin EJ, Heyer E, Carvalho-Silva D, Kraaijenbrink T, Knijff P, Tyler-Smith C, Jobling MA. Gene conversion violates the stepwise mutation model for microsatellites in Y-chromosomal palindromic repeats. Hum Mutat. 2014;35:609–17.
22. Hancock JM. Simple sequences and the expanding genome. BioEssays. 1996;18:421–5.
23. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER. Microsatellite instability regulates transcription factor binding and gene expression. Proc Natl Acad Sci. 2005;102:3800–4.
24. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol. 1994;4:24–33.
25. Pardue M, Lowenhaupt K, Rich A, Nordheim A. (dC-dA) n.(dG-dT) n sequences have evolutionarily conserved chromosomal locations in drosophila with implications for roles in chromosome structure and function. EMBO J. 1987;6:1781–9.
26. Richard GF, Pâques F. Mini-and microsatellite expansions: the recombination connection. EMBO Rep. 2000;1:122–6.
27. Volff J. Genome evolution and biodiversity in teleost fish. Heredity. 2005;94: 280–94.
28. Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 2015;7(2):567–80.
29. Chalopin D, Volff J-N, Galiana D, Anderson JL, Schartl M. Transposable elements and early evolution of sex chromosomes in fish. Chromosom Res. 2015;23:545–60.
30. Gao B, Shen D, Xue S, Chen C, Cui H, Song C. The contribution of transposable elements to size variations between four teleost genomes. Mob DNA. 2016;7:4.
31. Allendorf FW, Thorgaard GH. Tetraploidy and the evolution of salmonid fishes. In: Evolutionary genetics of fishes: Springer;Boston.1984. p. 1-53.
32. Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays. 2005;27:937–45.
33. Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G. Genome sequence and genetic diversity of the common carp, Cyprinus Carpio. Nat Genet. 2014;46:1212–9.
34. Jiang H. The distribution trends in simple repetitive stretches of DNA. Chinese J Biochem Mol. 1997;14:65–70.
35. Compagno LJ. Alternative life-history styles of cartilaginous fishes in time and space. Environ Biol Fishes. 1990;28:33-75.
36. Huang CRL, Burns KH, Boeke JD. Active transposition in genomes. Annu Rev Genet. 2012;46:651–75.
37. Siwicke KA, Seitz AC. Interpreting lamprey attacks on Pacific cod in the eastern Bering Sea. T Am Fish Soc. 2015;144:1249–62.

Yuan *et al. BMC Genomics* (2018) 19:141

Page 9 of 10

38. Clemens BJ, Binder TR, Docker MF, Moser ML, Sower SA. Similarities, differences, and unknowns in biology and management of three parasitic lampreys of North America. Fisheries. 2010;35:580–94.

39. Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene. 1999;238:115–34.

40. Brosius J. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica. 1999;107:209-38.

41. Gess RW, Coates MI, Rubidge BS. A lamprey from the Devonian period of South Africa. Nature. 2006;443:981–4.

42. Watson CA, Hill JE, Graves JS, Wood AL, Kilgore KH. Use of a novel induced spawning technique for the first reported captive spawning of Tetraodon Nigroviridis. Mar Genomics. 2009;2:143–6.

43. Nelson J. Fishes of the world 4th edition. Hoboken: John Wiley & Sons, Inc; 2006. p.334–456.

44. Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature. 2004;431:946–57.

45. Ovidio M, Detaille A, Bontinck C, Philippart J-C. Movement behaviour of the small benthic Rhine sculpin *Cottus rhenanus* (Freyhof, Kottelat & Nolte, 2005) as revealed by radio-telemetry and pit-tagging. Hydrobiologia. 2009;636:119–28.

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

47. Abrusán G, Krambeck H-J. Competition may determine the diversity of transposable elements. Theor Popul Biol. 2006;70:364–75.

48. McDonald JF. Evolution and consequences of transposable elements. Curr Opin Genet Dev. 1993;3:855-64.

49. Zhang H-H, Feschotte C, Han M-J, Zhang Z. Recurrent horizontal transfers of Chapaev transposons in diverse invertebrate and vertebrate animals. Genome Biol Evol. 2014;6:1375–86.

50. Schrader L, Kim JW, Ence D, Zimin A, Klein A, Wyschetzki K, Weichselgartner T, Kemena C, Stökl J, Schultner E. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun. 2014;5:5495.

51. Yi S, Streelman JT. Genome size is negatively correlated with effective population size in ray-finned fish. Trends Genet. 2005;21:643–6.

52. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013;496:498–503.

53. Lynch M, Conery JS. The origins of genome complexity. Science. 2003;302: 1401–4.

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

55. Sullivan JP, Lundberg JG, Hardman M. A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences [J]. Mol Phylogenet Evol. 2006;41:636–62.

56. Ackerly DD, Reich PB. Convergence and correlations among leaf size and function in seed plants: a comparative test using independent contrasts. Am J Bot. 1999;86:1272–81.

57. McPhail J. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): origin of the species pairs. Can J Zool. 1993;71:515–23.

58. McPhail J. Speciation and the evolution of reproductive isolation in the sticklebacks (*Gasterosteus*) of south-western British Columbia. The evolutionary biology of the threespine stickleback; 1994. p. 399–437.

59. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484(7392):55–61.

60. Parry G. Osmotic adaptation in fishes. Biol Rev. 1966;41(3):392–440.

61. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water stress: evolution of osmolyte systems. Science. 1982;217:1214–22.

62. Canceill D, Ehrlich SD. Copy-choice recombination mediated by DNA polymerase III holoenzyme from *Escherichia coli*. Proc Natl Acad Sci. 1996;93(13):6647–52.

63. Fraser BA, Künstner A, Reznick DN, Dreyer C, Weigel D. Population genomics of natural and experimental populations of guppies (Poecilia Reticulata). Mol Ecol. 2015;24:389–408.

64. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff J-N, Lesch K-P. The genome of the platyfish, *Xiphophorus*

maculatus, provides insights into evolutionary adaptation and several complex traits. Nat Genet. 2013;45:567–72.

65. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y. The medaka draft genome and insights into vertebrate genome evolution. Nature. 2007;447:714–9.

66. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 2014;513:375–81.

67. Conte MA, Kocher TD. An improved genome reference for the African cichlid, Metriaclima Zebra. BMC Genomics. 2015;16:724.

68. McGaugh SE, Gross JB, Aken B, Blin M, Borowsky R, Chalopin D, Hinaux H, Jeffery WR, Keene A, Ma L. The cavefish genome reveals candidate genes for eye loss. Nat Commun. 2014;5:5307.

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

70. Shin SC, Ahn DH, Kim SJ, Pyo CW, Lee H, Kim M-K, Lee J, Lee JE, Detrich HW, Postlethwait JH. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. Genome Biol. 2014;15:468.

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

72. Smolka M, Rescheneder P, Schatz MC, von Haeseler A, Sedlazeck FJ. Teaser: individualized benchmarking and optimization of read mapping results for NGS data. Genome Biol. 2015;16:235.

73. AlMomin S, Kumar V, Al-Amad S, Al-Hussaini M, Dashti T, Al-Enezi K, Akbar A. Draft genome sequence of the silver pomfret fish, Pampus Argenteus. Genome. 2015;59:51–8.

74. Nakamura Y, Mori K, Saitoh K, Oshima K, Mekuchi M, Sugaya T, Shigenobu Y, Ojima N, Muta S, Fujiwara A. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proc Natl Acad Sci. 2013;110:11061–6.

75. Wu C, Zhang D, Kan M, Lv Z, Zhu A, Su Y, Zhou D, Zhang J, Zhang Z, Xu M. The draft genome of the large yellow croaker reveals well-developed innate immunity. Nat Commun. 2014;5:5227.

76. Xu T, Xu G, Che R, Wang R, Wang Y, Li J, Wang S, Shu C, Sun Y, Liu T. The genome of the miiuy croaker reveals well-developed innate immune and sensory systems. Sci Rep. 2016;6:21902.

77. Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff J-N. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet. 2014;46:253–60.

78. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A. Whole-genome shotgun assembly and analysis of the genome of Fugu Rubripes. Science. 2002;297:1301–10.

79. Gao Y, Gao Q, Zhang H, Wang L, Zhang F, Yang C, Song L. Draft sequencing and analysis of the genome of pufferfish Takifugu Flavidus. DNA Res. 2014;21:627–37.

80. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016;533:200–5.

81. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, Lemon C, Bird NH, Koop BF. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One. 2014;9:e102089.

82. Burns FR, Cogburn AL, Ankley GT, Villeneuve DL, Waits E, Chang YJ, Llaca V, Deschamps SD, Jackson RE, Hoke RA. Sequencing and de novo draft assemblies of a fathead minnow (Pimephales Promelas) reference genome. Environ Toxicol Chem. 2016;35:212–7.

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

84. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 2011;477:207–10.

85. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48:427–37.

86. Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, MacCallum I, Braasch I, Manousaki T, Schneider I, Rohner N. The African coelacanth genome provides insights into tetrapod evolution. Nature. 2013;496:311–6.

Yuan *et al. BMC Genomics* (2018) 19:141

Page 10 of 10

87. Read TD, Petit RA III, Joseph SJ, Alam MT, Weil R, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb H. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* smith 1828. Peer J Pre Prints. 2015;14:837v1.

88. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 2014;505:174–9.

89. Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet. 2013;45:415–21.

90. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 2002;12:1269–76.

91. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21:i351–i8.

92. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, Smit AF, Finn RD. Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. 2013;41:D70–82.

93. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.

94. Castresana J. Cytochrome b phylogeny and the taxonomy of great apes and mammals. Mol Biol Evol. 2001;8(4):465–71.

95. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

96. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27:1164–5.

97. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4. https://doi.org/10.1093/molbev/msw054.

98. Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, Graves TA, Hillier LW, Mardis ER, McPherson JD. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420:520–62.

99. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. 2003. http://www.R-project.org/.

100. Fofonoff NP. Physical properties of seawater: a new salinity scale and equation of state for seawater. J Geophys Res-Oceans. 1985;90:3332–42.

101. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language[J]. Bioinformatics. 2004;20(2):289–90.

102. Deng W, Wang Y, Liu Z, Cheng H, Xue Y. Heml: a toolkit for illustrating heatmaps. PLoS One. 2014;9:e111988.