Received: 22 September 2019 | Revised: 20 December 2019 | Accepted: 24 December 2019

DOI: 10.1111/1755-0998.13132

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Chromosomal-level assembly of *Takifugu obscurus* (Abe, 1949) genome using third-generation DNA sequencing and Hi-C analysis

Seunghyun Kang[1] | Jin-Hyoung Kim[1] | Euna Jo[1,2] | Seung Jae Lee[2] | Jihye Jung[1] | Bo-Mi Kim[1] | Jun Hyuck Lee[1,3] | Tae-Jin Oh[4,5,6] | Seungshic Yum[7] | Jae-Sung Rhee[8,9] | Hyun Park[2]

[1]Unit of Polar Genomics, Korea Polar Research Institute (KOPRI), Incheon, Korea

[2]Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, Korea

[3]Polar Sciences, University of Science & Technology, Daejeon, Korea

[4]Department of Life Science and Biochemical Engineering, SunMoon University, Asan, Korea

[5]Genome-based BioIT Convergence Institute, Asan, Korea

[6]Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan, Korea

[7]Risk Assessment Research Center, Korea Institute of Ocean Science and Technology (KIOST), Geoje, Korea

[8]Department of Marine Science, College of Natural Sciences, Incheon National University, Incheon, Korea

[9]Research Institute of Basic Sciences, Incheon National University, Incheon, Korea

Correspondence
Jae-Sung Rhee, Department of Marine Science, College of Natural Sciences, Incheon National University, Incheon 22012, South Korea.
Email: jsrhee@inu.ac.kr

Hyun Park, Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Korea.
Email: hpark@korea.ac.kr

Funding information
Genome Program of the Korea Institute of Marine Science and Technology Promotion, Grant/Award Number: 20180430; Basic Science Research Program, Grant/Award Number: NRF-2017R1A6A1A06015181; Ministry of Oceans and Fisheries, Grant/Award Number: 20180430; National Research Foundation; Ministry of Education

Abstract

The Tetraodontidae family are known to have relatively small and compact genomes compared to other vertebrates. The obscure puffer fish *Takifugu obscurus* is an anadromous species that migrates to freshwater from the sea for spawning. Thus the euryhaline characteristics of *T. obscurus* have been investigated to gain understanding of their survival ability, osmoregulation, and other homeostatic mechanisms in both freshwater and seawater. In this study, a high quality chromosome-level reference genome for *T. obscurus* was constructed using long-read Pacific Biosciences (PacBio) Sequel sequencing and a Hi-C-based chromatin contact map platform. The final genome assembly of *T. obscurus* is 381 Mb, with a contig N50 length of 3,296 kb and longest length of 10.7 Mb, from a total of 62 Gb of raw reads generated using single-molecule real-time sequencing technology from a PacBio Sequel platform. The PacBio data were further clustered into chromosome-scale scaffolds using a Hi-C approach, resulting in a 373 Mb genome assembly with a contig N50 length of 15.2 Mb and and longest length of 28 Mb. When we directly compared the 22 longest scaffolds of *T. obscurus* to the 22 chromosomes of the tiger puffer *Takifugu rubripes*, a clear one-to-one orthologous relationship was observed between the two species, supporting the chromosome-level assembly of *T. obscurus*. This genome assembly can serve as a valuable genetic resource for exploring fugu-specific compact genome characteristics, and will provide essential genomic information for understanding

Seunghyun Kang and Jin-Hyoung Kim contributed equally to this work.

molecular adaptations to salinity fluctuations and the evolution of osmoregulatory mechanisms.

KEYWORDS
chromosome-level assembly, fugu genome, Hi-C assembly, long-read sequencing, *Takifugu obscurus*

1 | INTRODUCTION

The Tetraodontidae (simply called puffers) are biologically interesting fish due to their unique morphometric characteristics and defence systems (e.g., inflation, neurotoxin accumulation). Their ecological importance has been highlighted due to having a wide geographic distribution, being found from tropical and temperate seas to freshwater locales, and overall they exhibit enormous diversity in their shapes, sizes, and way of life (Stump, Ralph, Comeros-Raynal, Matsuura, & Carpenter, 2018). The order Tetraodontiformes comprises 412 extant species in 10 families, and 184 recognized species in 27 genera are contained in the family Tetraodontidae (Matsuura, 2014). The twin defence mechanisms of inflation and accumulation of potent neurotoxins (e.g., saxitoxin, tetrodotoxin) to discourage predatory activity are notable characteristics in puffers, and have rarely been observed in other fish (Miyazawa & Noguchi, 2001; Wainwright & Turingan, 1997). As an important aquaculture fish, several puffer species including *Takifugu obscurus* (Figure 1a) have been commercially cultured for culinary or medicinal preparations in East Asian countries such as China, Korea, and Japan.

The genus *Takifugu* belongs to the family Tetraodontidae and consists of over 20 species living in the Northwest Pacific Ocean (Kato, Doi, Nakada, Sakai, & Hirose, 2005). Extraordinary colouration, variety in body sizes, and thermal habitats have been highlighted as unique characteristics of *Takifugu* species compared to other pufferfish (Santini et al., 2013; Yamanoue et al., 2009). Of these, only two species, the obscure puffer *T. obscurus* and the ocellated puffer *T. ocellatus* are known to be uniquely anadromous (Wu, Jin, & Ni, 1978; Xu, 1990), with most puffers' habitats limited to shallow, warm, tropical and temperate seas (Matsuura, 2014; Stump et al., 2018). An earlier study on freshwater adaptability revealed that only *T. obscurus* could survive more than 10 days in freshwater while another five *Takifugu* species were unable to survive (Kato et al., 2005). The obscure puffers are distributed on the seabed of inshore and inland waters of the East China Sea, the South China Sea, and inland waters in China and the Korean Peninsula (Kato et al., 2005; Figure 1b). They normally live in seawater and when sexually mature, obscure puffers migrate to inland freshwaters to spawn, while fingerlings grow in these inland waters before returning to the sea. The mitochondrial genome of the obscure puffer shares a very high sequence identity with those of the genus *Takifugu* (Kato et al., 2005), while investigations of genomic similarity, unique genomic characteristics, and comparative genomics have rarely been conducted in these fish. Thus, to make full use of this anadromous model system and to understand the complex mechanisms behind the maintenance of body salt and water homeostasis as well as its molecular evolution, a high-quality reference genome is strongly required.

In the family Tetraodontidae, the genome size of *T. rubripes* was originally determined to be one-eighth the size of the human genome, but with a similar gene repertoire to that seen in humans (Brenner et al., 1993). Subsequently, the genome of *T. rubripes* was the first Tetraodontidae sequenced and was assembled as the second vertebrate genome in 2002, further establishing its compact genomic characteristics (Aparicio et al., 2002). However, large-scale genomic analysis at the chromosome level (e.g., chromosomal rearrangement) has not been well-characterized in *Takifugu* due to the fragmented nature of assemblies, although studies for increasing the contiguity of *T. rubripes* assembly to the chromosomal level have been conducted (Kai et al., 2011). Genomic information of the green spotted puffer *Tetraodon nigroviridis* was published, highlighting the smallest genomes among all vertebrates based on higher assembly contiguity compared to the *T. rubripes* genome (Jaillon et al., 2004). Recently, high quality genomes were published from *T. bimaculatus* (Zhou et al., 2019b) and *T. flavidus* (Zhou et al., 2019a). The relatively low content of repetitive sequences, higher gene density, and shorter introns and intergenic regions were shown to be remarkable characteristics in the smaller and compact genomes of puffers, thus establishing them as ideal models for comparative and molecular evolutionary research in vertebrates. In this study, a high-quality chromosome-level reference genome for the obscure puffer *T. obscurus* was established using a long-read PacBio strategy, followed by scaffolding with Hi-C-based chromatin contact maps. The genome of the obscure puffer was statistically compared to the available genomes of *T. rubripes*,*T. nigroviridis*, *T. bimaculatus* and *T. flavidus* for understanding puffer-specific genomic evolution. This assembly will be a useful resource for research on genetic and genomic studies as well as for studies on ecological and evolutionary aspects of puffers.

2 | MATERIALS AND METHODS

2.1 | Sample collection and DNA extraction

Obscure pufferfish, *T. obscurus* (≈30 cm in length) were collected from Imjin River in Korea in April 2016 using a net. A single collected specimen was used for both DNA isolation and total RNA preparation. No specific permission was required for fishing and the location is not privately-owned or protected. To obtain sufficient high-quality DNA molecules for long-read sequencing on the PacBio Sequel

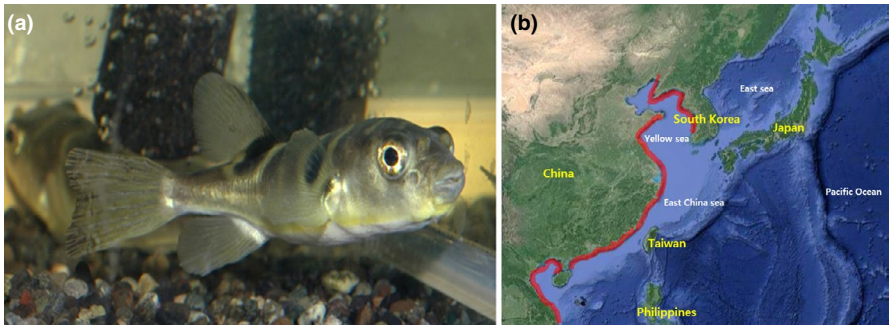

FIGURE 1 The obscure puffer *Takifugu obscurus* used in this study and their distribution. (a) *T. obscurus*. (b) The distribution is shown in dark red [Colour figure can be viewed at wileyonlinelibrary.com]

platform (Pacific Biosciences), one obscure pufferfish was dissected and fresh muscle tissues were used for DNA extraction using the phenol/chloroform extraction method. The quality of the DNA was checked using a 2100 Bioanalyzer (Agilent Technologies), and high integrity DNA molecules were measured using 1% agarose gel electrophoresis. The current study followed the guidelines of the Animal Welfare Ethical Committee and the Animal Experimental Ethics Committee of the Korea Polar Research Institute (KOPRI, South Korea).

2.2 | Library construction, sequencing, and assembly

Genomic DNA libraries were prepared according to the manufacturer's instructions. The single-molecule real-time sequencing (SMRT) bell (SMRT Bell) library was constructed using a PacBio DNA Template Prep Kit 1.0 (Pacific Biosciences). Quality and quantity of each library was checked using a 2100 Bioanalyzer (Agilent Technologies). The SMRT Bell-Polymerase complex was constructed using a PacBio Binding Kit 2.0 (Pacific Biosciences) based on the manufacturer's instructions. The complex was loaded onto SMRT cells (Pacific Biosciences, Sequel SMRT Cell 1M v2) and sequenced using Sequel Sequencing Kit 2.1 (Pacific Biosciences, Sequel SMRT Cell 1M v2). For each SMRT cell, 1 × 600 min movies were captured using the Sequel sequencing platform (Pacific Biosciences) at DNA Link (Seoul, South Korea). For de novo assembly, the FALCON-Unzip assembler (ver. 0.4, Falcon, Resource Identification Portal identification number (RRID):SCR_016089) was used (Chin et al., 2016) with parameters about length cutoff (length_cutoff = 12,000, length_cutoff_pr = 10,000) and the filtered subreads from SMRT Link (version 4.0.0) (Minimum subread length = 50). For improve the quality of genome assembly, the FALCON-Unzip assembler was polished using Arrow algorithm with unaligned BAM files as raw data.

Muscle samples from the same obscure puffer were used in constructing a Hi-C chromatin contact map to enable a chromosome-level assembly. Tissue fixation, chromatin isolation, and library construction for Hi-C analysis were performed according to the manufacturer's instructions (Belton et al., 2012). After checking the insert size, concentration, and effective concentration of the constructed libraries, final libraries were sequenced using the Illumina Novaseq platform with 150 bp paired-end strategy. A total of 171 million raw reads were generated from the Hi-C libraries and were mapped to the polished obscure puffer scaffolds using HiRise with default parameters (Putnam et al., 2016). Whole genome alignment was further conducted to compare consistency between the obscure puffer scaffolds and *T. rubripes* chromosomes using a pairwise aligner Large-Scale Genome Alignment Tool (LASTZ) version 1.10 (Harris, 2007). The k-mer-based genome size estimation was omitted, as a reference genome is available in the genus *Takifugu* (FUGU5). A total of 22 chromosome-level super-scaffolds were constructed for the *T. obscurus* genome.

2.3 | Assessment of the completeness of chromosome-level assembly

To evaluate the completeness of the obscure puffer chromosome-level assembly, the assembled scaffolds of *T. obscurus* were subjected to Benchmarking Universal Single-Copy Orthologs (BUSCO) ver. 3.0 (RRID:SCR_015008) with default parameters, by the conservation of a core set of genes of the Actinopterygii database (actinopterygii odb9 constructed from 20 fish species consisting of 4,584 orthologs; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015).

2.4 | Transcriptome sequencing

Total RNA was extracted from twelve tissues (blood, brain, eye, gall bladder, gill, gonad, heart, kidney, liver, muscle, skin, and stomach) of the same individual that was used for DNA extraction, using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. The quality and quantity of total RNA was evaluated using an Agilent Bioanalyzer (Agilent Technologies). Two micrograms of total RNA from each tissue was pooled for RNA sequencing. The pooled sample for the twelve organs were sequenced using one SMRT cell v3 based on P6-C4 chemistry after standard full-length cDNA (1–3 kb) library preparation, and a total of two SMRT cells were sequenced on a PacBio Sequel system (Pacific Biosciences). Filtering, quality control, clustering, and polishing of the Iso-Seq sequencing data were performed using SMRT Link (version 6.0.0).

2.5 | Genome annotation and repeat analysis

A de novo repeat library was constructed using REPEATMODELER version 1.0.3 (RRID:SCR_015027; Bao & Eddy, 2002), including RECON version 1.08 (Price, Jones, & Pevzner, 2005) and REPEATSCOUT version 1.0.5 (RRID:SCR_014653; Price et al., 2005), with default parameters. Tandem Repeats Finder version 4.09 was used to predict consensus sequences, classification information for each repeat, and tandem repeats including simple repeats, satellites, and low-complexity repeats (Benson, 1999). Kimura distances (Kimura, 1980) were calculated for all TE copies found in each family from the library to estimate the age of TEs.

Genome annotation was conducted using MAKER version 2.28 (MAKER, RRID:SCR_005309) with three rounds of reiterative training (Holt & Yandell, 2011). Ab initio gene prediction was performed using SNAP (SNAP, RRID:SCR_002127) and Augustus (Augustus: Gene Prediction, RRID:SCR_008417). MAKER was initially run in est2genome mode, which was based on Iso-Seq data including the 34,871 full-length transcripts. Additionally, protein evidence was provided from the genomes of twelve teleosts (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Dicentrarchus labrax*, *Astyanax mexicanus* and *Xiphophorus maculatus*). A next round of MAKER was run using retrained HMMs for gene prediction with all other settings that were the same as for the first run except est2genome and protein2genome settings. After one more retraining, MAKER selected and revised the final gene model considering all information produced.

The protein sequences were aligned to the nonredundant database ($E < 10^{-10}$) and the UniProtKB/Swissprot database with BLASTP version 2.2.29, and protein signatures were annotated using INTERPROSCAN 5 for further BLAST2GO-based gene ontology (GO) analysis (version 4.1.9; Gotz et al., 2008), and also aligned against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to retrieve KEGG relevant functional annotations. The INFERNAL software package version 1.1 (INFERNAL, RRID:SCR_011809; Nawrocki, Kolbe, & Eddy, 2009) and covariance models from the RFAM (RFAM, RRID:SCR_007891; Gardner et al., 2010) database were used to identify other non-coding RNAs in the *T. obscurus* scaffolds. Putative tRNA genes were identified using TRNASCAN-SE version 1.4 (TRNASCAN-SE, RRID:SCR_010835; Lowe & Eddy, 1997), which employs a covariance model that scores candidates based on sequence and predicted secondary structures.

2.6 | Gene family identification and phylogenetic analysis

Orthologous gene clusters were classified within 12 well-annotated and well-assembled fish species genomes (Table S5), along with *T. obscurus*, using both ORTHOMCL (Li, Stoeckert, & Roos, 2003) and the ORTHOMCL pipeline with application of the Markov Clustering Algorithm (MCL; Fischer et al., 2011) with default parameters (Table S1). We used coding sequences of genes annotated based on the MAKER annotation pipeline (Holt & Yandell, 2011). A phylogenetic tree was constructed based on one-to-one, single-copy orthologous gene clusters. The Probabilistic Alignment Kit (PRANK; Löytynoja & Goldman, 2005) was used to align protein coding genes with codon alignment option. The Gblocks (Castresana, 2000) was used to eliminate poorly aligned regions with gaps under a codon model (–t=c –e=-gb1 –b4=5 –d=y) for subsequent procedures. The MEGA 6 software (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) was used to calculate divergence time of the 13 fish species and a maximum-likelihood tree was built using RAXML (RRID:SCR_006086) (Stamatakis, 2014) under generalised time-reversible model with optimization of substitution rates and gamma model of rate heterogeneity with 1,000 bootstrap replicates. The divergence time calibration was performed with TIMETREE (minimum and maximum divergence time for *D. rerio* and *T. rubripes*: 204.5 and 255.3 million years ago; minimum and maximum divergence times for *G. aculeatus* and *T. rubripes* were: 141 and 170 million years ago) (Hedges, Dudley, & Kumar, 2006). CAFÉ 4.0 (Han, Thomas, Lugo-Martinez, & Hahn, 2013) was employed to identify the likelihood of gene family expansion and contraction with $p < .05$ and the separate birth (λ) and death (μ) rates were estimated with the same program using the lambda/mu command with –s and –t options.

2.7 | Positively selected genes analysis

Among the orthologous gene clusters classified in the previous analysis steps, orthologous analysis returned 911 one-to-one gene clusters between thirteen fish species, and we selected the sequences of each species in a cluster with the highest pairwise score among the other sequences of its cluster from many-to-many clusters. These selected gene clusters were used to identify positively selected genes (PSGs). The software PRANK was used to align the coding sequences and poorly aligned sequences with gaps were removed with a codon model using Gblocks. The aligned sequences with <50% identity and shorter than 150 bp were removed in subsequent analysis steps. The values of d*N*, d*S*, and ω were estimated with CODEML program implemented in the PAML package (RRID:SCR_014932) (Yang, 2007). The basic and branch-site models were tested to define PSGs, and genes under relaxation of selective pressure were eliminated by Likelihood Ratio Tests (LRTs). The functional categories and pathways enriched in the PSGs were identified by performing a BLAST2GO enrichment test (Conesa et al., 2005) with Fisher's exact test (cutoff: $p \leq .05$).

3 | RESULTS

3.1 | Genome sequencing and assembly

Information on genome diversity and evolution in puffer is incomplete, although valuable genomic information has been published for several puffers (*T. bimaculatus*, *T. flavidus*, *T. rubripes*, and *T. nigroviridis*).

TABLE 1 Sequencing data statistics generated for *Takifugu obscurus* genome assembly

Library type	Platform	Library size (bp)	Data size (Gb)	Coverage (×)	Application
Long reads	PacBio Sequel	20,000	30.8	81	Genome assembly
Hi-C	Novaseq	250	25.8	68	Chromosome construction
Isoseq	PacBio Sequel	20,000	3.0	–	Annotation

We therefore took advantage of long-read sequencing technology to perform chromosome-level assembly of *T. obscurus*. In total, 30.8 and 25.8 Gb of high-quality genomic data were produced by PacBio Sequel and Novaseq platforms, respectively (Table 1). The average coverage of SMRT sequences on the *T. obscurus* genome was 81 and 68-fold, respectively (Table 1). The N50 lengths of polymerase and subreads reached 15.2 kb and 12.0 kb, suggesting efficient production of ultra-long genomic sequences for the following assembly (Table S2). The genome was assembled with SMRT reads and yielded 380 Mb of reference genome size with an N50 length about 3.3 Mb and a longest length of 10.7 Mb (Table 2). In the case of the Hi-C super-scaffolding, the total size of the genome was 373 Mb with the longest contig being 28 Mb and the N50 value was 15 Mb in length (Table 2). The genome size of *T. obscurus* is quite comparable to those of *T. rubripes* (393 Mb) and *T. nigroviridis* (358 Mb). Overall, the N50 value of *T. obscurus* genome assemblies is comparable to the published fish genomes that were constructed by long-read PacBio technology. To the best of our knowledge in teleosts, only a few fish species including Asian seabass (1.19 Mb; Vij et al., 2016), Nile tilapia (3.09 Mb; Conte, Gammerdinger, Bartie, Penman, & Kocher, 2017), Tibetan endemic fish (Yang et al., 2019), orange clownfish (1.86 Mb; Lehmann et al., 2019), two-spotted puffer (1.31 Mb; Zhou et al., 2019b), and yellowbelly puffer (4.4 Mb; Zhou et al., 2019a) showed contig N50 longer than 1 Mb when their genomes were analyzed with long-read PacBio sequencing platforms.

Overall, we constructed a high quality chromosomal-level assembly of *T. obscurus* with 22 pseudo-chromosomes with lengths ranging from 9.84 to 28.07 Mb (Table 3). The total length of 22 pseudo-chromosomes consisted of 91% of all genomic sequences. The 22 pseudo-chromosomes assembled in *T. obscurus* were directly aligned to the 22 chromosomes of the *T. rubripes* genome and 22 pseudo-chromosomes of *T. bimaculatus* and *T. flavidus* (Figure 2a–d), suggesting that our assembly is highly contiguous compared to the other *Takifugu* genomes.

In addition to the validation of assembly contiguity, we assessed the completeness of the chromosome-level genome assemblies using BUSCO with the Actinopterygii lineage. Among 4,584 total BUSCO groups searched, 4,399 and 96 BUSCO core genes were completely and partially identified, respectively (Table 4), leading to a total of 98.0% of core eukaryotic genes being captured in the *T. obscurus* genome. Taken together, the use of both PacBio long reads and Hi-C-based chromatin contact maps successfully managed to increase contiguity and completeness in the chromosome-level assembly of the *T. obscurus* genome.

TABLE 2 *Takifugu obscurus* genome assembly statistics

Assembly	Falcon-unzip	Hi-C
Number of contigs (scaffolds)	1,133	586
Total size of contigs (bp)	380,655,037	373,462,439
Longest contig	10,685,970	28,067,044
N50 contig length (bp)	3,296,039	15,200,451
Number of scaffolds > 10 M	1	21
Gap (%)	0	0.01

3.2 | Gene annotation and comparison with puffer genomes

The *T. obscurus* genome contained 45.7% repetitive sequences, and 11.05% repetitive sequences were accounted for by transposable elements (TEs; Table S3). Among these, there are 2.57% of DNA transposons, 1.43% of long terminal repeats (LTRs), 3.75% of long interspersed elements (LINEs), and 0.20% of short interspersed elements (SINEs; Table S3). Overall, most crucial TE superfamilies were detected in the *T. obscurus* genome, suggesting that teleost-specific TE diversity is conserved in the genome, as has been shown in several teleost genomes (Chalopin, Naville, Plard, Galiana, & Volff, 2015). The amount of TEs in the *T. obscurus* genome is slightly higher than the proportions seen in *T. rubripes* and *T. nigroviridis*, while the amount of total interspersed repeat sequences in *T. obscurus* (11.05%) is quite a bit higher than in *T. rubripes* (7.53%) and *T. nigroviridis* (5.60%) (Table S3). These results suggest that TEs might have independently diversified and accumulated in the *T. obscurus* genome after divergence from the *Takifugu* genus lineage.

One of the fascinating characteristics of puffers is that their genomes, which are the smallest and most compact among all known vertebrates, are an excellent model for comparative gene network and evolutionary investigation (Aparicio et al., 2002; Jaillon et al., 2004; Venkatesh, Gilligan, & Brenner, 2000). It is believed that reduced intron and intergenic regions, higher gene density and chromosomal stability, and lower percentages of repetitive sequences could drive their genomes smaller (Jaillon et al., 2004), whereas the quantity of and length of genes are not different between puffers and teleosts. A positive correlation between genome size and TE content has been previously noted in teleosts, as the relatively smaller genome of *T. nigroviridis* (≈360 Mb) has 6% TEs, whereas zebrafish have a genome that is 55% TEs (≈1.4 Gb; Chalopin et al., 2015). This positive correlation also applies to *T. obscurus* (Figure S1). In the case of satellite sequences, both *T. rubripes* and *T. nigroviridis* have small amounts in their genome, but no sequence was detected in the *T. obscurus* genome. Since both TEs and satellite sequences have been suggested to be involved in genome

TABLE 3 Lengths of the 22 longest scaffolds assembled in *T. obscurus* genome

#Scaffold	Length (bp)
To_HiC1	28,067,044
To_HiC2	13,629,849
To_HiC3	17,107,491
To_HiC4	15,200,451
To_HiC5	13,660,805
To_HiC6	11,836,844
To_HiC7	16,025,588
To_HiC8	18,578,086
To_HiC9	14,710,971
To_HiC10	12,830,475
To_HiC11	15,217,180
To_HiC12	12,507,637
To_HiC13	18,477,698
To_HiC14	15,434,050
To_HiC15	13,980,399
To_HiC16	12,514,786
To_HiC17	15,055,448
To_HiC18	9,842,566
To_HiC19	16,267,691
To_HiC20	16,282,353
To_HiC21	17,438,062
To_HiC22	13,889,933
Total sequence clustered (%)	338,555,407 (90.65%)

size variation in insects (Bosco, Campbell, Leiva-Neto, & Markov, 2007; Kidwell, 2002), it may be concluded that satellite sequences are less relevant in contributing to genome size variation in teleosts. Kimura divergence was calculated for all TE copies of each element to estimate the age and transposition history of TEs. The *T. obscurus* genome underwent more recent TE amplification than *T. rubripes* and *T. nigroviridis*, as the genome is dominated by rather recent copies (*K*-values ≤5) and highly shaped by DNA transposons (Figure 3a). Overall, DNA transposons and LINEs have been active and strongly contributed during the evolution of the three puffers' genomes.

Following the characterization of repetitive sequences, a total of 22,105 *T. obscurus* genes were annotated using MAKER with generated Iso-Seq data. The annotated gene number is comparable to those of 12 fish species (Table S1) and to the average number of genes (23,475) analyzed in 22 fish species (Lehmann et al., 2019). The average length of *T. obscurus* genes was 8.7 kb, with an average intron length of 254 kb (9.8 exons per gene) (Table S4). Overall, 99.0% of the *T. obscurus* genes were annotated based on known proteins in public databases (Table S4), and 73.4% were similar to *T. rubripes* genes (Figure S2). The principal BLAST top hits indicated that about 16,223 T. *obscurus* contigs exhibited sequence similarities to transcripts of *T. rubripes*, and 1,364 exhibited similarities to transcripts of *T. nigroviridis* (Figure S2). Several KEGG pathways were enriched when we mapped *T. obscurus* genes to entire KEGG pathways (Figure S3). Gene family analysis identified a core set of 10,532 gene families that were shared among the five puffer genomes and of these 151 unique gene families containing 377 genes were *T. obscurus*-specific (Figure 3b). Further studies will focus on classifying and characterizing these unique gene families for understanding their fundamental role in *T. obscurus* physiology and adaptation.

3.3 | Genome evolution

The phylogenetic tree of *T. obscurus* and 12 other teleosts was constructed using single-copy orthologues (Figure 4; Table S5). The *T. obscurus* orthologues were closely clustered with those of *T. rubripe*, and further, formed a clade with *T. nigroviridis* in the family Tetraodontidae (Figure 4). The family Tetraodontidae diverged approximately 126 million years ago from its common ancestor and the genus *Takifugu* diverged about 53 million years ago from the genus *Tetraodon*. Genomic differences between *T. obscurus* and *T. rubripes* have accumulated for ~3 million years, resulting in good agreement with previous estimates, as *Takifugu* species have undergone explosive speciation with variations of behavior, morphology, and physiology in the last 2–5 Myr (Yamanoue et al., 2009). Thus, the phylogenetic relationship as evidenced by different TE compositions and unique orthologues, supports the independent genomic evolution of *T. obscurus* from the *Takifugu* lineage.

A comprehensive theory for the modulation of the unique compact genome of puffers is still lacking, and mechanisms for the replication and transcriptional regulation of such compact genomes are not, as yet, defined. In this study, it was observed that several pathways involved in chromatin regulation, chromosome maintenance, or histone modification are greatly expanded in the *T. obscurus* genome (Table S6). For example, the vast majority of expanded pathways (42%; 14 in 33 pathways) in the biological process category are involved in the modulation of DNA (GO0065004, *protein-DNA complex assembly*; GO0071103, *DNA conformation change*; GO0006323, *DNA packaging*; GO0071824, *protein-DNA complex subunit organization*), nucleosome (GO0043486, *histone exchange*; GO0034728; *nucleosome organization*; GO0006334, *nucleosome assembly*; GO0031063, *regulation of histone deacetylation*; GO0031056, *regulation of histone modification*) or chromatin/chromosome (GO1902275, *regulation of chromatin organization*; GO0031497, *chromatin assembly*; GO0006325, *chromatin organization*; GO0006333, *chromatin assembly or disassembly*; GO0043044, *ATP-dependent chromatin remodelling*). Similarly, many expanded gene families (75%; nine in 12 pathways) were located in DNA, nucleosome or chromatin/chromosomes (GO0032993, *protein-DNA complex*; GO0044815, *DNA packaging complex*, GO0005694, *chromosome*; GO0000788, *nuclear nucleosome*; GO0000786, *nucleosome*; GO0000785, *chromatin*; GO0000790, *nuclear chromatin*; GO0044454, *nuclear chromosome part*; GO0044427, *chromosomal part*) in the cellular component

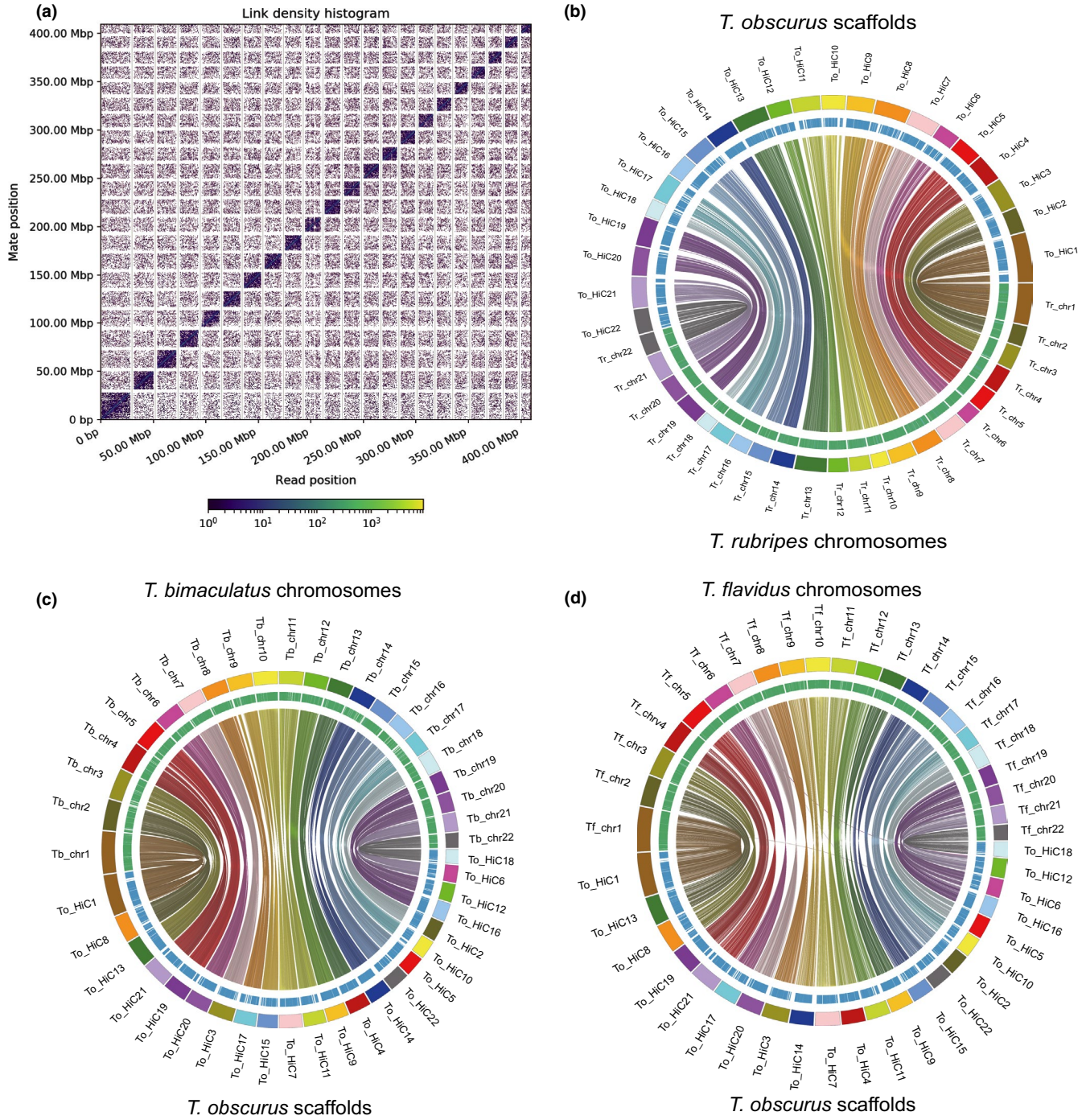

FIGURE 2 Hi-C interaction heat map and genome synteny. (a) Analysis of contiguity of *Takifugu obscurus* scaffolds constructed using PacBio Sequel platform and Hi-C-based chromatin contact maps. The blocks represent the contact between one location and the other locations. The colour reflects the intensity of each contact, for which the deeper colour represents the higher intensity. (b) Collinear blocks between *T. obscurus* and 22 chromosomes of *T. rubripes*. (c) Collinear blocks between *T. obscurus* and 22 chromosomes of *T. bimaculatus*. (d) Collinear blocks between *T. obscurus* and 22 chromosomes of *T. flavidus*. Each coloured arc represents a best match between the two species. To_HiC1–22 represents pseudo-chromosomes of the pufferfish genome and Tr_chr1–22, Tb_chr1–22 and Tf_chr1–22 represents chromosomes 1–22 of *T. rubripes*, *T. bimaculatus* and *T. flavidus*, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

category. In the molecular function category, *nucleosome binding* (GO0031491) and *histone deacetylase binding* (GO0042826) were identified to be relevant to the mechanism. Although the unique mechanism evolved in the *T. obscurus* compact genome is hard to explain based on just these expanded pathways, these components would provide an advantage in terms of access to a large concatenation of genes (e.g., replication, transcription) with reduced energy expenditures compared to much larger genomes

TABLE 4 Benchmarking Universal Single-Copy Orthologs (BUSCO) evaluated for the completeness of the *Takifugu obscurus* genome assembly

Actinopterygii_odb9	No.	%
Complete BUSCOs (C)	4,399	95.9
Complete and single-copy BUSCOs (S)	4,186	91.3
Complete and duplicated BUSCOs (D)	213	4.6
Fragmented BUSCOs (F)	96	2.1
Missing BUSCOs (M)	89	2
Total BUSCO groups searched	4,584	98

(Roest Crollius & Weissenbach, 2005). The molecular mechanisms of the puffer-specific compactness remain to be elucidated.

The obscure puffer possesses varying doses of the potent neurotoxin tetrodotoxin, which has been suggested to be produced by internal endosymbiotic bacteria introduced via the food chain, but studies on the exact mechanism of biosynthesis, transport, and accumulation of toxin are still being actively investigated (Bane, Lehane, Dikshit, O'Riordan, & Furey, 2014). Interestingly, toxin-related gene families were shown to be expanded and these included *toxin transport* (GO1901998) in the biological process category and *toxin transmembrane transporter activity* (GO0019534) in the molecular function category. The amount of neurotoxin present is differentially accumulated and is dependent on the species, maturation status, organs (mainly bioconcentrated in ovary, liver, intestine, and skin), and seasons. Thus, the toxin may be transported and stored through several steps consisting of cellular transporters, moving it from the primary source. Some previous evidence supports this idea, as tetrodotoxin was shown to be bioconcentrated by a carrier-mediated membrane transporter into *T. rubripes* liver tissue (Matsumoto et al., 2007).

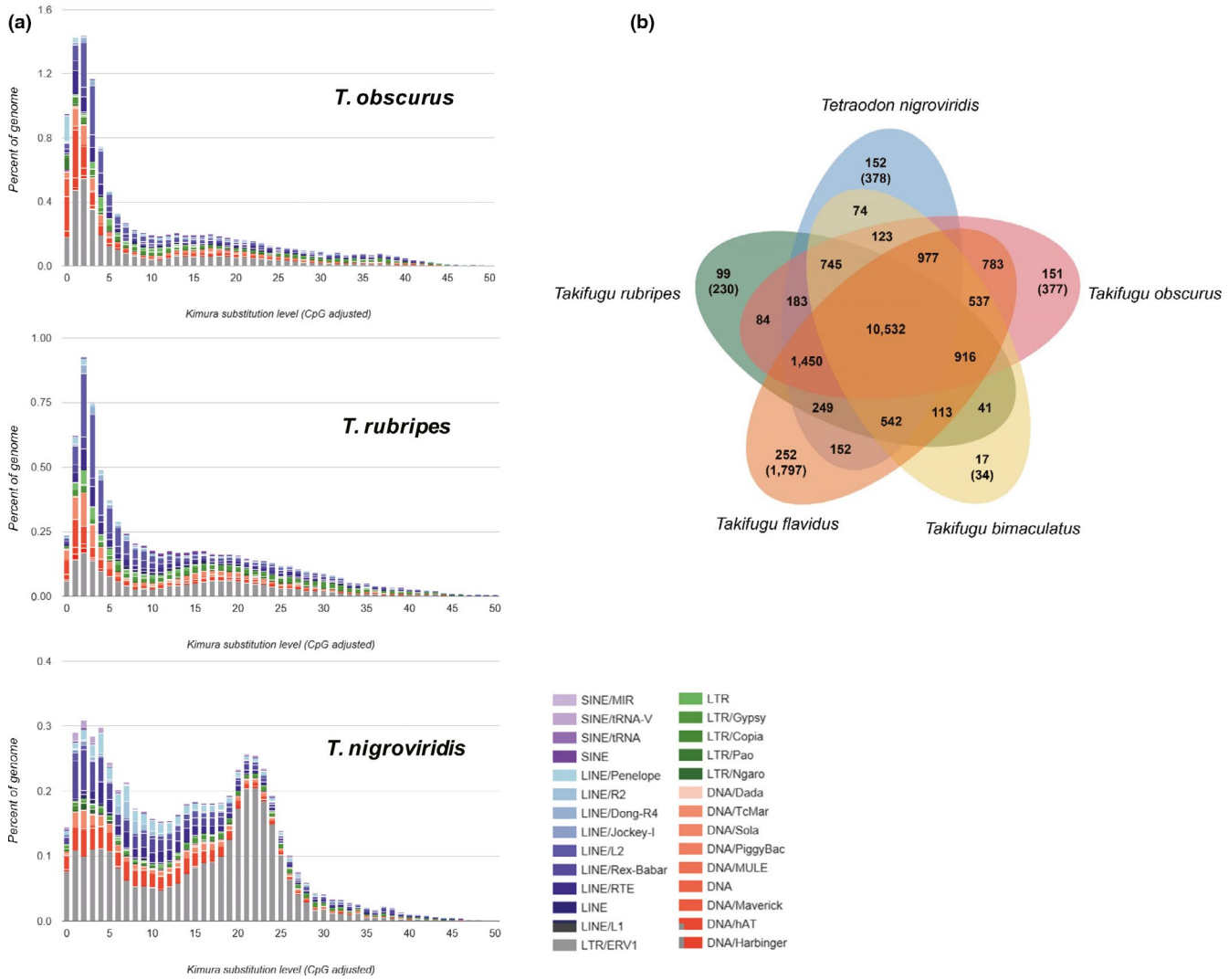

FIGURE 3 Comparison of repetitive components and orthologous in puffer genomes. (a) Kimura distance-based copy divergence analysis of transposable elements in three fugu genomes. Graphs represent genome coverage (*Y*-axis) for each type of TE (DNA transposons, SINE, LINE, and LTR retrotransposons) in the different genomes analyzed, clustered to their corresponding consensus sequence according to Kimura distances (*X*-axis, *K*-value from 0 to 50). (b) Venn diagram of orthologous gene families. Five fugu species (*T. obscurus*, *T. nigroviridis*, *T. rubripes*, *T. bimaculatus*, and *T. flavidus*) were used to generate the Venn diagram based on the gene family cluster analysis [Colour figure can be viewed at wileyonlinelibrary.com]

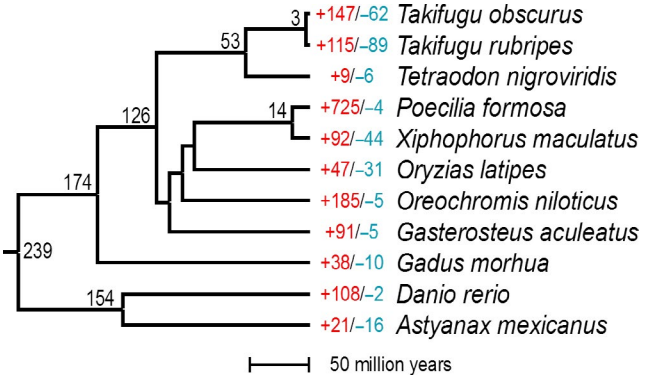

FIGURE 4 Phylogenetic analysis of *Takifugu obscurus* within the teleost lineage and gene family gain-and-loss analysis, including the number of gained gene families (+) and lost gene families (–). Each number on the branch site indicates divergence times between lineages and the 95% confidential intervals [Colour figure can be viewed at wileyonlinelibrary.com] [Colour figure can be viewed at wileyonlinelibrary.com]

The K/Cl cotransporters (KCCs) have important roles in epithelial ion transport and osmotic homeostasis in vertebrates by coupling electroneutral movement of K and Cl ions (Adragna, Di Fulvio, & Lauf, 2004). Although their molecular functions have been mainly investigated in mammals and in vitro models, several studies have suggested that KCCs in fish are ubiquitously present and involved in ionic/osmotic homeostasis and cell volume regulation for protection against cell swelling and shrinkage (Bogdanova & Nikinmaa, 2001; Guizouarn & Motais, 1999; Jensen, 1995; Lauf, 1982; Thomas & Egee, 1998). Thus, the expanded gene family *potassium:chloride symporter activity* (GO0015379) observed in the molecular function category could be associated with the unique anadromous characteristics of *T. obscurus*. In particular, fluctuations in salinity are a critical stressor for fish immune maintenance, as salinity challenge directly modulates parasite prevalence, host–parasite interactions, resistance against pathogens, and general innate immune functions (Uribe, Folch, Enriquez, & Moran, 2011). Several expanded immunity-related gene families were observed in the biological process component (GO0050851, *antigen receptor-mediated signalling pathway*; GO0050852, *T cell receptor signalling pathway*; GO0048002, *antigen processing and presentation of peptide antigen*; GO0002768, *immune response-regulating cell surface receptor signalling pathway*; GO0002474, *antigen processing and presentation of peptide antigen via MHC class I*; GO0002429, *immune response-activating cell surface receptor signalling pathway*; GO0002574, *thrombocyte differentiation*), the cellular component (GO0042612, *MHC class I protein complex*), and the molecular function category (GO0003823, *antigen binding*), are presumably associated with immune stability of *T. obscurus* upon salinity challenge during migration.

We found evidence of positive selection for 14 genes in *T. obscurus* (Table S7). Several genes identified in the *T. obscurus* genome have crucial functions in mammals, while very limited case reports have been published in teleosts. The major histocompatibility complex (MHC) is important in adaptive immunity in jawed vertebrates, by encoding several cell-surface glycoproteins for presenting peptides to T cells (Germain, 1994). Genomic structure and polymorphism of the two major subfamilies, MHC class I and class II, have been extensively investigated to understand genome evolution, phylogeny, immune response, and disease in teleosts (Uribe et al., 2011). In *T. obscurus* genome, MHC class I protein is positively selected, supporting the evidence that nonlinkage of MHC class I and class II loci would affect different patterns of MHC evolution in teleosts, as the genomic loci of class I gene are linked to class II gene loci in cartilaginous fish and tetrapods (Kelley, Walter, & Trowsdale, 2005; Ohta et al., 2000). Although consecutive sequence of whole MHC was firstly cloned in *F. rubripes* (Clark, Shaw, Kelly, Snell, & Elgar, 2001), there is no in-depth study on comparative genomic evolution of MHC in Tetraodontidae. Thus, this information would be helpful to understand fugu-specific MHC evolution with incorporation of additional MHC structures from fugu genomes that are sequenced to date. In case of other positively selected genes, very limited information on their roles has been suggested in teleost taxa. Although all discussion on these genes is presumption at this point in time, it will be useful for comparative genome analysis in Tetraodontidae or between teleosts.

In conclusion, we present a high quality, chromosome-level genome assembly of the obscure puffer, *T. obscurus*. A primary assembly was constructed with PacBio Sequel sequencing and Falcon_Unzip algorithm platform, and final assembly was established into chromosome-sized scaffolds including 22 longest scaffolds using Hi-C chromatin contact maps. The *T. obscurus* assemblies showed high contiguity and completeness. An effective gene model can be offered using the *T. obscurus* gene set, as 22,105 genes including 377 T. *obscurus*-specific orthologues were annotated. Since we know little about the biological aspects of *T. obscurus* in terms of development, physiology, reproduction, or anadromous life cycle, the high-quality *T. obscurus* reference genome assembly constructed in this study will facilitate the cross-disciplinary use of *T. obscurus* as a promising fish model to investigate evolutionary and ecological aspects of the puffer-specific small genome, and *T. obscurus*-specific anadromous characteristics.

ACKNOWLEDGEMENTS

This research was supported by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (MOF) (No. 20180430). This work was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2017R1A6A1A06015181).

AUTHOR CONTRIBUTIONS

H.P., J.K. and J.R. conceived the study. H.P., S.K., S.J.L., B.K., T.O., J.L. and S.Y. performed genome sequencing, assembly, and annotation. E.J. and J.J. performed experiments. J.R. and H.P. mainly wrote the paper.

All authors contributed to writing and editing the manuscript as well as providing supplementary information and producing the figures.

DATA AVAILABILITY STATEMENT

The obscure puffer genome project was deposited at NCBI under BioProject number PRJNA449558. The whole-genome sequence was deposited in the Sequence Read Archive (SRA) database under accession number SRR7081561 and SRR10127925.

ORCID

Jin-Hyoung Kim https://orcid.org/0000-0001-6746-7447
Hyun Park https://orcid.org/0000-0002-8055-2010

REFERENCES

Adragna, N. C., Di Fulvio, M., & Lauf, P. K. (2004). Regulation of K-Cl cotransport: From function to genes. *The Journal of Membrane Biology*, *201*, 109–137. https://doi.org/10.1007/s00232-004-0695-6

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., … Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, *297*, 1301–1310. https://doi.org/10.1126/science.1072104

Bane, V., Lehane, M., Dikshit, M., O'Riordan, A., & Furey, A. (2014). Tetrodotoxin: Chemistry, toxicity, source, distribution and detection. *Toxins*, *6*, 693–755. https://doi.org/10.3390/toxins6020693

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, *12*, 1269–1276. https://doi.org/10.1101/gr.88502

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*, 573–580. https://doi.org/10.1093/nar/27.2.573

Bogdanova, A. Y., & Nikinmaa, M. (2001). Reactive oxygen species regulate oxygen-sensitive potassium flux in rainbow trout erythrocytes. *The Journal of General Physiology*, *117*, 181–190. https://doi.org/10.1085/jgp.117.2.181

Bosco, G., Campbell, P., Leiva-Neto, J. T., & Markov, T. A. (2007). Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*, *177*, 1277–1290. https://doi.org/10.1534/genetics.107.075069

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., & Aparicio, S. (1993). Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature*, *366*, 265–268. https://doi.org/10.1038/366265a0

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

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. https://doi.org/10.1093/gbe/evv005

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., … Schatz, M. C. (2016). Phased diploid genome assembly with single molecule real-time sequencing. *Nature Methods*, *13*, 1050–1054. https://doi.org/10.1038/nmeth.4035

Clark, M. S., Shaw, L., Kelly, A., Snell, P., & Elgar, G. (2001). Characterization of the MHC class I region of the Japanese pufferfish (*Fugu rubripes*). *Immunogenetics*, *52*, 174–185. https://doi.org/10.1007/s002510000285

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). BLAST2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J., & Kocher, T. D. (2017). A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics*, *18*, 341. https://doi.org/10.1186/s12864-017-3723-5

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Stoeckert Jr, C. J. (2011). Using ORTHOMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, Chapter 6, Unit 6.12, 1–19. https://doi.org/10.1002/0471250953.bi0612s35

Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., … Bateman, A. (2010). RFAM: Wikipedia, clans and the "decimal" release. *Nucleic Acids Research*, *39*, D141–D145.

Germain, R. N. (1994). MHC-dependent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. *Cell*, *76*, 287–299. https://doi.org/10.1016/0092-8674(94)90336-0

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the BLAST2GO suite. *Nucleic Acids Research*, *36*, 3420–3435. https://doi.org/10.1093/nar/gkn176

Guizouarn, H., & Motais, R. (1999). Swelling activation of transport pathways in erythrocytes: Effects of Cl^-, ionic strength, and volume changes. *American Journal of Physiology*, *276*, C210–C220.

Han, M. V., Thomas, G. W., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, *30*, 1987–1997. https://doi.org/10.1093/molbev/mst100

Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA*. PhD thesis. State College, PA: Pennsylvania State University.

Hedges, S. B., Dudley, J., & Kumar, S. (2006). TIMETREE: A public knowledge-base of divergence times among organisms. *Bioinformatics*, *22*, 2971–2972. https://doi.org/10.1093/bioinformatics/btl505

Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*, 491. https://doi.org/10.1186/1471-2105-12-491

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., … Roest Crollius, H. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, *431*, 946–957. https://doi.org/10.1038/nature03025

Jensen, F. (1995). Regulatory volume decrease in carp red blood cells: Mechanisms and oxygenation-dependency of volume-activated potassium and amino acid transport. *Journal of Experimental Biology*, *198*, 155–165.

Kai, W., Kikuchi, K., Tohari, S., Chew, A. K., Tay, A., Fujiwara, A., … Venkatesh, B. (2011). Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in teleosts and mammals. *Genome Biology and Evolution*, *3*, 424–442. https://doi.org/10.1093/gbe/evr041

Kato, A., Doi, H., Nakada, T., Sakai, H., & Hirose, S. (2005). *Takifugu obscurus* is a euryhaline fugu species very close to *Takifugu rubripes* and suitable for studying osmoregulation. *BMC Physiology*, *5*, 18.

Kelley, J., Walter, L., & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, *56*, 683–695. https://doi.org/10.1007/s00251-004-0717-7

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*, 49–63.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111–120. https://doi.org/10.1007/BF01731581

Lauf, P. K. (1982). Evidence for chloride-dependent potassium and water transport induced by hyposmotic stress in erythrocytes of the

marine teleost, *Opsanus tau*. *Journal of Comparative Physiology*, *146*, 9–16. https://doi.org/10.1007/BF00688711

Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., … Ravasi, T. (2019). Finding Nemo's Genes: A chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Molecular Ecology Resources*, *19*, 570–585.

Li, L., Stoeckert, C. J. Jr, & Roos, D. S. (2003). ORTHOMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*, 2178–2189. https://doi.org/10.1101/gr.1224503

Lowe, T. M., & Eddy, S. R. (1997). TRNASCAN-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*, 955–964. https://doi.org/10.1093/nar/25.5.955

Löytynoja, A., & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences USA*, *102*, 10557–10562.

Matsumoto, T., Nagashima, Y., Kusuhara, H., Sugiyama, Y., Ishizaki, S., Shimakura, K., & Shiomi, K. (2007). Involvement of carrier-mediated transport system in uptake of tetrodotoxin into liver tissue slices of puffer fish *Takifugu rubripes*. *Toxicon*, *50*, 173–179. https://doi.org/10.1016/j.toxicon.2007.03.004

Matsuura, K. (2014). Taxonomy and systematics of tetraodontiform fishes: A review focusing primarily on progress in the period from 1980 to 2014. *Ichthyological Research*, *62*, 72–113. https://doi.org/10.1007/s10228-014-0444-5

Miyazawa, K., & Noguchi, T. (2001). Distribution and origin of tetrodotoxin. *Toxin Review*, *20*, 11–13. https://doi.org/10.1081/TXR-100103081

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). INFERNAL 1.0: Inference of RNA alignments. *Bioinformatics*, *25*, 1335–1337. https://doi.org/10.1093/bioinformatics/btp157

Ohta, Y., Okamura, K., McKinney, E. C., Bartl, S., Hashimoto, K., & Flajnik, M. F. (2000). Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. *Proceedings of the National Academy of Sciences USA*, *97*, 4712–4717. https://doi.org/10.1073/pnas.97.9.4712

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics*, *21*, i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., … Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*, 342–350. https://doi.org/10.1101/gr.193474.115

Roest Crollius, H., & Weissenbach, J. (2005). Fish genomics and biology. *Genome Research*, *15*, 1675–1682. https://doi.org/10.1101/gr.3735805

Santini, F., Nguyen, M. T. T., Sorenson, L., Waltzek, T. B., Lynch Alfaro, J. W., Eastman, J. M., & Alfaro, M. E. (2013). Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). *Journal of Evolutionary Biology*, *26*, 1003–1018. https://doi.org/10.1111/jeb.12112

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stump, E., Ralph, G. M., Comeros-Raynal, M. T., Matsuura, K., & Carpenter, K. E. (2018). Global conservation status of marine pufferfishes (Tetraodontiformes: Tetraodontidae). *Global Ecology and Conservation*, *14*, e00388. https://doi.org/10.1016/j.gecco.2018.e00388

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Thomas, S., & Egee, S. (1998). Fish red blood cells: Characteristics and physiological role of the membrane ion transporters. *Comparative Biochemistry and Physiology Part A*, *119*, 79–86. https://doi.org/10.1016/S1095-6433(97)00404-2

Uribe, C., Folch, H., Enriquez, R., & Moran, G. (2011). Innate and adaptive immunity in teleostfish: A review. *Veterinarni Medicina*, *56*, 486–503.

Venkatesh, B., Gilligan, P., & Brenner, S. (2000). Fugu: A compact vertebrate reference genome. *FEBS Letters*, *476*, 3–7. https://doi.org/10.1016/S0014-5793(00)01659-8

Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., … Orbán, L. (2016). Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLOS Genetics*, *12*, e1005954.

Wainwright, P. C., & Turingan, R. G. (1997). Evolution of pufferfish inflation behavior. *Evolution*, *51*, 506–518. https://doi.org/10.1111/j.1558-5646.1997.tb02438.x

Wu, H. L., Jin, X. B., & Ni, Y. (1978). *Toxic and pharmacological fish in China Shanghai*. Shanghai, China: Scientific & Technical Publishers.

Xu, C. (1990). *Takifugu obscurus* (Abe). In East China Sea Fisheries Institute, Chinese Academy of Fisheries Science (Ed.), *The fishes of Shanghai area* (pp. 377–378). Shanghai, China: Scientific & Technical Publishers.

Yamanoue, Y., Miya, M., Matsuura, K., Miyazawa, S., Tsukamoto, N., Doi, H., … Sakai, H. (2009). Explosive speciation of *Takifugu*: Another use of fugu as a model system for evolutionary biology. *Molecular Biology and Evolution*, *26*, 623–629. https://doi.org/10.1093/molbev/msn283

Yang, X., Liu, H., Ma, Z., Zou, Y., Zou, M., Mao, Y., … Yang, R. (2019). Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Molecular Ecology Resource*, *19*, 1027–1036.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Zhou, Y., Xiao, S., Lin, G., Chen, D., Cen, W., Xue, T., … Huang, Z. (2019a). Chromosome genome assembly and annotation of the yellowbelly pufferfish with PacBio and Hi-C sequencing data. *Scientific Data*, *6*, 267. https://doi.org/10.1038/s41597-019-0279-z

Zhou, Z., Liu, B. O., Chen, B., Shi, Y., Pu, F., Bai, H., … Xu, P. (2019b). The sequence and *de novo* assembly of *Takifugu bimaculatus* genome using PacBio and Hi-C technologies. *Scientific Data*, *6*, 187. https://doi.org/10.1038/s41597-019-0195-2

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Kang S, Kim J-H, Jo E, et al. Chromosomal-level assembly of *Takifugu obscurus* (Abe, 1949) genome using third-generation DNA sequencing and Hi-C analysis. *Mol Ecol Resour*. 2020;20:520–530. https://doi.org/10.1111/1755-0998.13132