



# Genome Analysis of *Lagocephalus sceleratus*: Unraveling the Genomic Landscape of a Successful Invader

Theodoros Danis<sup>1,2</sup>, Vasileios Papadogiannis<sup>2</sup>, Alexandros Tsakogiannis<sup>2</sup>, Jon B. Kristoffersen<sup>2</sup>, Daniel Golani<sup>3</sup>, Dimitris Tsaparis<sup>2</sup>, Aspasia Sterioti<sup>2</sup>, Panagiotis Kasapidis<sup>2</sup>, Georgios Kotoulas<sup>2</sup>, Antonios Magoulas<sup>2</sup>, Costas S. Tsigenopoulos<sup>2</sup> and Tereza Manousaki<sup>2\*</sup>

<sup>1</sup>School of Medicine, University of Crete, Heraklion, Greece, <sup>2</sup>Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Greece, <sup>3</sup>Department of Ecology, Evolution and Behavior and the National Natural History Collections, The Hebrew University, Jerusalem, Israel

## OPEN ACCESS

### Edited by:

Norman A. Johnson,  
University of Massachusetts Amherst,  
United States

### Reviewed by:

George Minos,  
International Hellenic University,  
Greece

Marco Gerdol,  
University of Trieste, Italy

### \*Correspondence:

Tereza Manousaki  
tereza@hcmr.gr

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

Received: 07 October 2021

Accepted: 16 November 2021

Published: 08 December 2021

### Citation:

Danis T, Papadogiannis V,  
Tsakogiannis A, Kristoffersen JB,  
Golani D, Tsaparis D, Sterioti A,  
Kasapidis P, Kotoulas G, Magoulas A,  
Tsigenopoulos CS and Manousaki T  
(2021) Genome Analysis of  
*Lagocephalus sceleratus*: Unraveling  
the Genomic Landscape of a  
Successful Invader.  
Front. Genet. 12:790850.  
doi: 10.3389/fgene.2021.790850

The Tetraodontidae family encompasses several species which attract scientific interest in terms of their ecology and evolution. The silver-cheeked toadfish (*Lagocephalus sceleratus*) is a well-known “invasive sprinter” that has invaded and spread, in less than a decade, throughout the Eastern and part of the Western Mediterranean Sea from the Red Sea through the Suez Canal. In this study, we built and analysed the first near-chromosome level genome assembly of *L. sceleratus* and explored its evolutionary landscape. Through a phylogenomic analysis, we positioned *L. sceleratus* closer to *T. nigroviridis*, compared to other members of the family, while gene family evolution analysis revealed that genes associated with the immune response have experienced rapid expansion, providing a genetic basis for studying how *L. sceleratus* is able to achieve highly successful colonisation. Moreover, we found that voltage-gated sodium channel (NaV 1.4) mutations previously connected to tetrodotoxin resistance in other pufferfishes are not found in *L. sceleratus*, highlighting the complex evolution of this trait. The high-quality genome assembly built here is expected to set the ground for future studies on the species biology.

**Keywords:** silver-cheeked toadfish, synteny, phylogenomic analyses, gene families, tetrodotoxin, genome evolution

## INTRODUCTION

The opening of the Suez Canal in 1869 initiated a process of biological invasions from the Red Sea into the Mediterranean, an event commonly known as Lessepsian migration (Por, 1971; Golani and Appelbaum-Golani, 1977). This influx of marine organisms has greatly impacted the local communities in ecological, evolutionary (Sax et al., 2007), and economical terms (Arim et al., 2006). Lessepsian fishes comprise a significant percentage of all recorded invasive species in the Mediterranean Sea (Zenetos et al., 2012) and may be causing several indigenous species displacements (Golani and Appelbaum-Golani, 2010). Lessepsian migration, having both direct and indirect human-driven origins, is a phenomenon that offers a unique opportunity for studying fast evolutionary change (Palumbi, 2001).

One of the most successful Lessepsian migrant species is the silver-cheeked toadfish, *Lagocephalus sceleratus* (Gmelin 1789), a member of the Tetraodontidae family (called puffers), widely distributed throughout the Indian and Pacific Oceans (Akyol et al., 2005). The first record of *L. sceleratus*

invasion in the Mediterranean Sea was reported in 2003 in the Gökova Bay, in the south-eastern Aegean Sea coast of Turkey (Filiz and Er, 2004) and 2 years later in the North Cretan Sea (Kasapidis et al., 2007). Since then, it has spread rapidly throughout the entire Levant, Aegean and Ionian Seas (Kalogirou, 2013; Akyol and Ünal, 2017).

The survival and dispersal rate of alien species into novel habitats are dependent on multiple factors including encountering previously unknown pathogens. Thus, the invasion success may be affected by the effectiveness of the immune response (Kolar and Lodge, 2001; Lee and Klasing, 2004) rendering the gene families involved in the adaptive immune response potentially under positive selection.

*Lagocephalus sceleratus* is a highly toxic species, which along with other pufferfishes, contains in different tissues, a toxic organic compound, tetrodotoxin (TTX) (Chua and Chew, 2009; Kosker et al., 2016). TTX is accumulated in high concentration in ovary, liver and intestine of *L. sceleratus* individuals, while lower amounts have been detected in skin and muscle tissues (Akboru et al., 2020). TTX is the metabolic product of TTX-producing bacteria and is retained in the fish tissues. However, the mechanisms of absorption and the differential concentration in the tissues and the organs of pufferfishes remain unclear (Nagashima and Arakawa, 2016). TTX toxicity is due to its binding to the outer pore of the NaV1.4 (SCN4) channel, blocking the transport of sodium ions across the pore (Hanifin, 2010). Multiple pufferfishes have been shown to deploy TTX resistance through mutations in specific domains of NaV1.4, (Venkatesh et al., 2005; Soong and Venkatesh 2006; Jost et al., 2008).

In this study, we provide and analyse the first high-quality genome assembly of *L. sceleratus*. To that end, we combine Illumina and Oxford Nanopore Technology (ONT) reads to construct a highly contiguous assembly, which allows us to explore the genomic landscape of this successful invader and study the genetic background of TTX-resistance in the species. This valuable and robust genome resource will facilitate future studies on the ecology, evolution and potential exploitation of this invasive species.

## MATERIALS AND METHODS

### Sample Collection and Sequencing

Animal care and handling were carried out following well established guidelines [Guidelines for the treatment of animals in behavioral research and teaching. *Anim. Behav.* 53, 229–234 (1997)].

One female fish (58 cm in length) was caught alive in Agios Georgios (Hersonissos), Crete, Greece (35°20′07.50″N 25°23′11.30″E) at the pre-spawning/spawning stage (as determined by stereoscopic investigation of the oocytes) and was anesthetized using clove oil. In total, 10 ml of blood was collected using a sterilized syringe and stored in tubes that contained ~1/10 of volume heparin for subsequent DNA extraction.

DNA extraction for the purpose of ONT sequencing was conducted on the day of sampling, from 2 µl of the freshly

collected blood, using Qiagen Genomic tip (20 G) and following ONT's protocol for DNA extraction from chicken blood. The final elution was made with 50 µl AE buffer providing 90.4 ng/µl (Qubit measurement) of high purity, high molecular weight DNA (Nanodrop ratios: 260/280 = 1.87 and 260/230 = 2.12). DNA integrity was assessed by electrophoresis in 0.4% w/v Bio-Rad Megabase agarose gel. We constructed four ligation libraries (SQK-LSK109) following the manufacturer's instructions. Approximately 1.2 µg of unsharded DNA was used for each library. Two of the prepared libraries were divided into two aliquots. Each library was run for approximately 24 h on a MinION sequencer at HCMR, after which the ONT nuclease flush protocol was performed, and a fresh library or library aliquot was loaded onto the same R9.4.1 flow cell. The total run time was ~130 h. Basecalling was done with Guppy v3.2.4 (<https://community.nanoporetech.com/posts/guppy-3-2-4-release>) in High Accuracy Mode with minimum quality score 7.

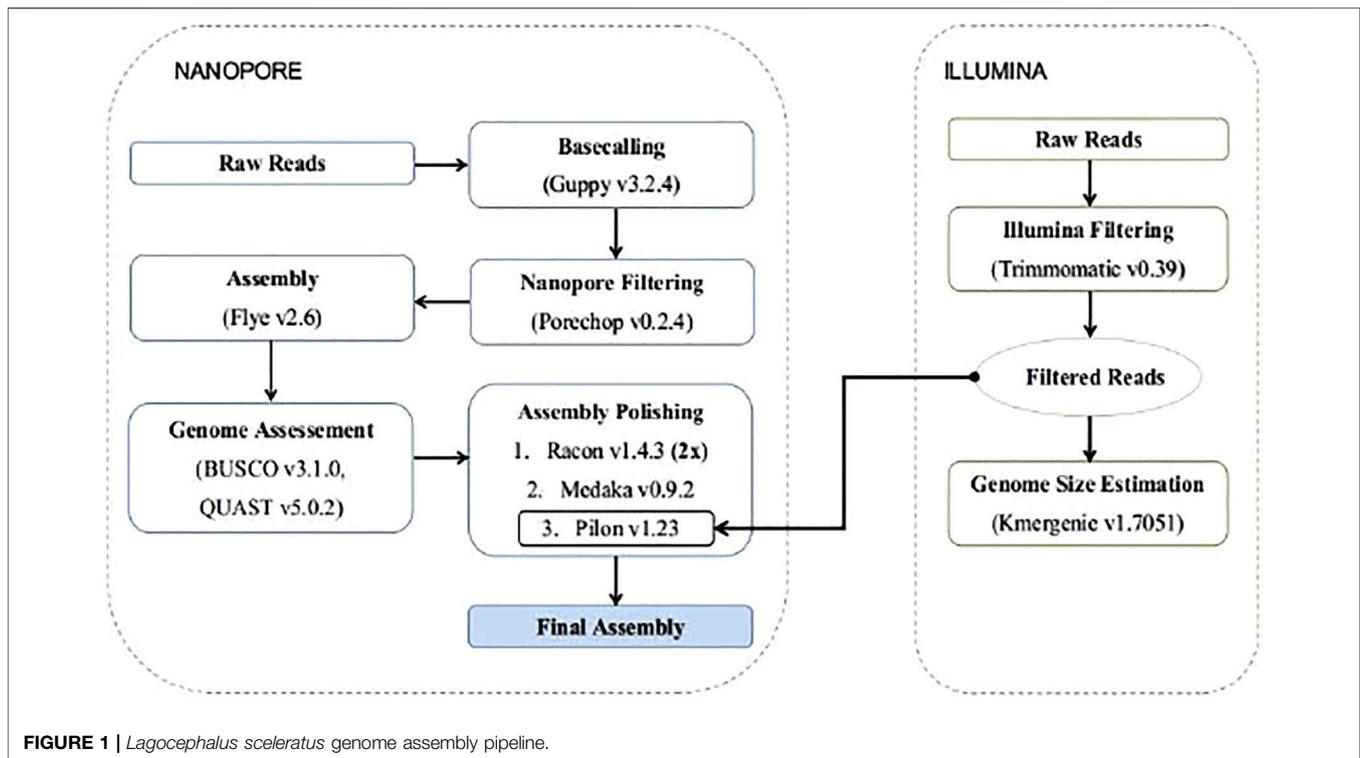
For Illumina sequencing, we proceeded with 2-days old, refrigerated blood samples using the same DNA extraction protocol as above. We used 4 µl of blood eluted in 100 µl AE buffer which resulted in 79.2 ng/µl (Qubit measurement) of pure DNA (260/280 = 1.85 and 260/230 = 2.21).

Template DNA for Illumina sequencing was sheared by ultrasonication in a Covaris S220 instrument. A PCR-free library was prepared with the Kapa Hyper Prep DNA kit with TruSeq Unique Dual Indexing. Paired end 2 × 150 bp sequencing was performed on an Illumina HiSeq4000 platform.

Tissue samples from brain, gonad, skin, liver, spleen and muscle of the same female individual, were grounded and powdered using pestle and mortar under liquid nitrogen, and except the spleen tissue were all homogenized in TRIzol<sup>®</sup> reagent (Invitrogen) assisting the homogenization procedure using needle and syringe. Total RNA was extracted from the TRIzol<sup>®</sup> homogenate according to the manufacturer's instructions. In the case of spleen, due to its nature and the inability to clean properly the RNA from "dirt" even after five washes with ethanol, the spleen RNA was finally extracted using the commercial kit NucleoSpin<sup>®</sup> RNA. The quantity of the isolated RNA was measured spectrophotometrically with NanoDrop<sup>®</sup> ND-1000 (Thermo Scientific), while its quality was tested on an agarose gel (electrophoresis in 1.5% w/v). Finally, all samples were used for mRNA paired-end library construction with the Illumina TruSeq<sup>™</sup> RNA Sample Preparation Kits v2 following the manufacturer's protocol (Poly-A mRNA isolation with oligo-dT beads, mRNA fragmentation, followed by transcription into first-strand cDNA using reverse transcriptase and random hexamer primers) and sequenced as 150 bp paired reads in half a lane of a HiSeq4000<sup>®</sup> following the protocols of Illumina Inc. (San Diego, CA).

### Raw Data Pre-Processing and Genome Size Estimation

Quality assessment of the raw DNA Illumina sequence data was performed with FastQC v0.11.8 (Andrews, 2010). Low quality



reads and adapters were removed using Trimmomatic v0.39 (Bolger et al., 2014). The reads were scanned by a 4-based sliding window with an average cutting threshold lower than 15 Phred score. Leading and trailing bases with quality scores less than 10 were also filtered out. Reads with total length shorter than 75 bp and average score below 30 were omitted. The same process was applied to the RNASeq reads.

Adapter trimming and length filtering of basecalled ONT data was done using Porechop v0.2.4 (<https://github.com/rrwick/Porechop>) with default parameters and the option `--discard_middle` to discard reads with internal adapters.

The genome size was estimated using the k-mer histogram method with Kmergenic v1.7051 (Chikhi and Medvedev, 2014) from the Illumina genomic sequencing data.

## De Novo Genome Assembly

To build the genome assembly the long ONT reads were used for the construction of an initial *de novo* assembly, and then the Illumina reads were used for the polishing stages. (Figure 1). To construct the initial assembly, we used the v. Flye v2.6 (Kolmogorov et al., 2019) algorithm, a repeat graph assembler. The assembly was evaluated by assessing: 1) the N50 sizes of contigs, using QUILT v5.0.2 (Gurevich et al., 2013), and 2) a gene completeness score using BUSCO v3.1.0 (Simão et al., 2015) against the Actinopterygii ortholog dataset v10, with default parameters. In addition Genomescope2 was used to for kmer frequency profiling (Ranallo-Benavidez et al., 2020).

The produced assembly was polished with two rounds of Racon v1.4.3 (Vaser et al., 2017; Vaser and Šikić, 2020), using the preprocessed long reads mapped against the assembly with

Minimap2 v2.17 (Li, 2018). Further polishing was performed with Medaka v0.9.2 (<https://github.com/nanoporetech/medaka>) and the final polishing was completed using Pilon v1.23 (Walker et al., 2014) after mapping the Illumina reads against the partially polished assembly with Minimap2 v2.17.

## Genome Annotation

### Repeat Elements Annotation

To identify the repeat elements' sequences, we constructed *de novo* repeat library using RepeatModeler2 (Flynn et al., 2020), including RECON v1.08 (Bao and Eddy, 2002), RepeatScout v1.0.6 (Price et al., 2005), LtrHarvest (Ellinghaus et al., 2008), which is incorporated in GenomeTools v1.5.9, Ltr\_retriever v2.7 (Ou and Jiang, 2018), assuming default parameters and the extra LTRStruct pipeline which includes Mafft v7.453 (Kato and Standley, 2013), CD-HIT v4.8.1 (Li and Godzik, 2006) and Ninja v0.95 (Wheeler, 2009). Thereafter, sequences that were obtained by RepeatModeler, were combined with Repbase v17.01 and a custom database constructed with the entries of *Takifugu rubripes*, *Takifugu flavidus*, and *Tetraodon nigroviridis* of the FishTEDB (Shao et al., 2018). Finally, RepeatMasker v4.1.0 (Tarailo-Graovac and Chen, 2009) was used to annotate repeat elements based on the above-described database.

### Gene Prediction and Functional Annotation

After repeat masking, gene prediction was conducted using MAKER2 pipeline v2.31.10 (Holt and Yandell, 2011) with two iterative rounds. We used a combined strategy of *ab initio*, homology-based and transcriptome-based methods. In the first round, for homology annotation, MAKER2 was initially run in

**TABLE 1** | Summary of sequencing results.

Sequencing technology	Raw reads	Quality-controlled reads	Coverage
Illumina	57,303,140	44,475,382	38 x
MinION	552,476	484,152	20 x

**TABLE 2** | Polished genome assembly statistics and completeness.

Total contigs	235
Total contig sequence	373,851,781 bp
GC (%)	46.7
Contig N50	11,297,640 bp
Contig N75	6,386,954 bp
Longest contig	17,085,954 bp
Contig L50	14
Contig L75	25
BUSCO completeness score	
Complete	96.60%
Single	95.70%
Duplicated	0.90%
Fragmented	1.00%
Missing	2.40%
Total number of Actinopterygii orthologs	3,640 (97.60%)

protein2genome mode, while SWISS-PROT (www.uniprot.org) was used for protein sequences extraction of three closely related species, *Mola mola*, *Tetraodon nigroviridis*, and *Takifugu rubripes*. For annotation using the RNA-Seq data, est2genome mode was enabled, which is based on transcriptome evidence. Transcriptomic reads from all sequenced tissues were mapped and assembled through the genome-guide approach, using HISAT2 v2.2.0 (Kim et al., 2015) and StringTie v2.1.1 (Pertea et al., 2015). *Ab initio* prediction was performed with SNAP (Korf, 2004) (<http://korflab.ucdavis.edu>), which was independently trained on *L. sceleratus* genome with default parameters and AUGUSTUS v3.3.3 (Stanke et al., 2006) previously trained through BUSCO v3.1.0 (Simão et al., 2015) with the extra parameter “-long.” The second round of MAKER2 was run using the previously trained models with the same settings as round one, except est2genome and protein2genome modes. The previous custom repeat library and MAKER2 repeat library that used for genome masking, remained for both rounds. The completeness of putative genes was assessed using BUSCO v4.0.5 (Simão et al., 2015) against the Actinopterygii odb10 database.

The functional annotation of the predicted genes of *L. sceleratus* was performed by similarity search against the UniprotKB/Swissprot database (release-2020\_03) with BLASTP v.2.9.0+ (e-value 1e-6, -max\_target\_seqs = 10) (Altschul et al., 1990). InterProScan v5 (Jones et al., 2014) was used to search motifs and domains against all default databases and the extra of SignalP\_EUK and TMHMM. Functional annotation results were also retrieved using eggNOG-mapper (Huerta-Cepas et al., 2017) based on fast orthology assignments using precomputed eggNOG v5.0 (Huerta-Cepas et al., 2019) clusters and phylogenies.

## Gene Ontology Mapping

Gene ontology analysis was carried out using a custom python script (gene\_ontology\_mapping.py). Gene ontology terms were retrieved through the Uniprot API service ([https://www.uniprot.org/help/programmatic\\_access](https://www.uniprot.org/help/programmatic_access)) and as queries we chose the best blast hits that we extracted after the functional annotation step against UniProtKB/Swiss-Prot.

## Phylogenomic Analysis Orthology Assignment

To identify paralogous and orthologous genes, we compared 30 whole-genome protein-coding gene sets from teleost fish (**Supplementary Table S3**) (27 species from our previous dataset of Natsidis et al. (2019) in addition with two extra Tetraodontidae species i.e., *Takifugu bimaculatus*, *T. flavidus*, and *L. sceleratus*) with Orthofinder v2.3.12 (Emms and Kelly, 2015) using default parameters. Firstly, the longest isoform of each gene was kept using the *primary\_transcript.py* script provided by Orthofinder suite. For *L. sceleratus*, only the longest isoforms over 30 amino acids were extracted with a custom script (longestIsoforms.py) and used in the analysis.

## Species Tree Reconstruction

For the phylogenomic analysis the orthogroups produced by orthofinder were filtered, keeping those containing a single gene per species to avoid inclusion of paralogs. Then, we kept those with representation from at least 26 out of the 30 taxa analysed in total, using a custom python script (filtered\_orthogroups.py) The amino-acid sequences encoded by selected genes of each orthogroup were aligned using MAFFT v7.453 (Katoh and Standley, 2013), with the -auto mode. The aligned orthogroups were then concatenated using a python script by P. Natsidis ([https://github.com/pnatsi/Sparidae\\_2019/blob/master/concatenate.py](https://github.com/pnatsi/Sparidae_2019/blob/master/concatenate.py)). The resulted alignments were filtered with Gblocks v0.91b (Castresana, 2000) to exclude poorly aligned regions with the following parameters: “Allowed Gap Positions” was set to half, “Minimum Length of a Block” was set to 8, “Minimum Number of Sequences for a Flanking Position” was set to 20, and “Minimum Number of Sequences for a Conserved Position” was set to 18.

Then, we ran RAXML-NG v0.9.0 (Kozlov et al., 2019) for phylogenetic tree reconstruction, and in order to select the best model, we used ModelTest-NG v0.1.6 (Darriba et al., 2020) specifying the --topology type parameter to maximum likelihood (ml) mode. The phylogenomic inference was run using the selected model, JTT + I + G4 + F. To assess the branch confidence, we ran 100 bootstrap replicates. The final tree was visualized using R/RStudio RStudio Team (2021) with a custom script using *Lepisosteus oculatus* as outgroup (phylo\_tree\_plot.r).

**TABLE 3** | Repeat elements annotation statistics.

Repetitive elements	Number of elements	Length occupied (bp)	Percentage of sequence (%)
Retroelements	204,205	29,198,482	7.81
SINEs	5,447	743,745	0.20
Penelope	3,627	2,572,446	0.69
LINEs	171,108	20,706,770	5.54
CRE/SLACS	0	0	0.00
L2/CR1/Rex	45,521	7,208,493	1.93
R1/LOA/Jockey	573	179,409	0.05
R2/R4/NeSL	48,675	3,708,881	0.99
RTE/Bov-B	47,510	3,856,263	1.03
L1/CIN4	16,080	2,144,919	0.57
LTR elements	27,650	7,747,967	2.07
BEL/Pao	410	259,661	0.07
Ty1/Copia	235	96,868	0.03
Gypsy/DIRS1	13,595	3,917,10	1.05
Retroviral	4,582	1,298,868	0.35
DNA transposons	59,679	8,587,997	2.30
hobo-Activator	25,182	2,911,788	0.78
Tc1-IS630-Pogo	14,227	2,885,165	0.77
En-Spm	0	0	0.00
MuDR-IS905	0	0	0.00
PiggyBac	644	131,052	0.04
Tourist/Harbinger	2,399	392,537	0.10
Other (Mirage, P-element, Transib)	99	5,804	0.00
Unclassified	113,777	1,126,996	5.97
Small RNA	0	0	0.00
Satellites	20	85,955	0.02
Simple repeats	9,305	1,126,996	0.30

**TABLE 4** | Summary statistics of functional annotated protein-coding genes.

Type	Number	Percent (%)
Blast	18,805	88
InterProScan	20,347	96
EggNog-Mapper	17,849	84
Predicted genes	20,578	97
Total Genes	21,251	

## Gene Family Expansion and Contraction

The expansion and contraction of gene families were analysed using CAFE v4.2.1 (De Bie et al., 2006). The sequences were first clustered using MCL (Enright et al., 2002) following the CAFE developers' instructions (<https://iu.app.box.com/v/cafetutorial-files>) and filtering the gene families using a custom python script (`cafe_filterin_blast_dump.py`) to exclude gene families that contain at least one or more species with  $\geq 100$  genes. An ultrametric tree was produced with r8s v1.81 (Sanderson, 2003) using the phylogenetic tree produced in our phylogenomic analysis and the divergence time for *Thunnus thynnus* and *Oreochromis niloticus* taken from TIMETREE (<http://www.timetree.org/>). Finally, CAFE was run with conditional *p*-values, for each gene family below 0.01.

## Synteny Analysis

Synteny analyses were performed on two tiers, on whole-genome sequence comparison and at the gene level.

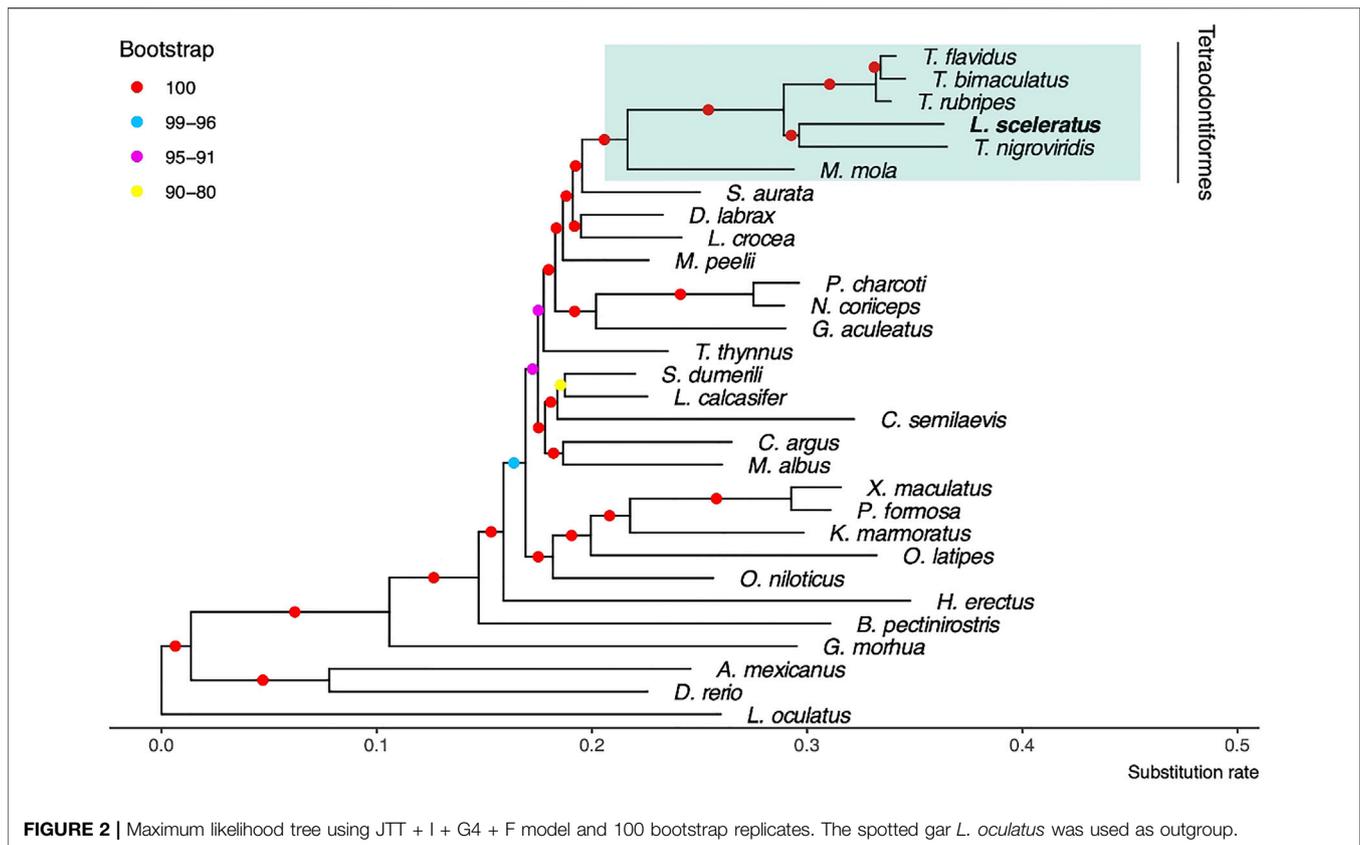
For the whole-genome comparison, the *L. sceleratus* genome assembly was compared with all available genomes of

Tetraodontiformes to date (*T. nigroviridis*, *T. rubripes*, *T. flavidus* and *T. bimaculatus*). Furthermore, we performed similar comparisons aligning the genomes of *T. nigroviridis* and *T. rubripes* against the other species. For the alignment, we used LAST v1145 (Kielbasa et al., 2011), implementing the sensitive alignment protocol as described for Human-mouse whole-genome project comparison (<https://github.com/mcfrith/last-genome-alignments>) with *e*-value cut-off 0.001. On the gene level, we used the one-to-one orthologues outputs of Orthofinder v2.3.12 analysis for comparing the same set of species.

We selected for visualization the 41 largest contigs of the *L. sceleratus* assembly representing  $\sim 91\%$  of the genome (**Supplementary Figure S1**), for both whole-genome and gene-based synteny results. In each set of fish, the whole-genome pairwise alignments were plotted by custom python scripts (`synteny_plot.py`) (**Supplementary Figures S3–S8**), while the one-to-one orthologs relationships of all the above-mentioned comparisons, were visualized through Circos (`circos_plot.py`) (Krzywinski et al., 2009) (**Supplementary Figures S9–S13**).

## Alignment of Voltage-Gated Sodium Channel Alpha Subunit 4 (SCN4A)

Mutations in SCN4A proteins have been previously associated with pufferfish TTX resistance (Venkatesh et al., 2005). To investigate these mutations further, sequences of the SCN4 alpha subunit were identified *via* BLAST (reciprocal best hit approach) against *D. rerio*, *L. sceleratus* and other fishes from the species tree (**Supplementary Table S3**), as well as *Takifugu pardalis* (Genbank: BAA90398.1),



*Latimeria chalumnae* (NCBI Reference: XP\_006003324.1), *Callorhynchus milii* (NCBI Reference: XP\_007907351.1), using the *D. rerio* SCN4AA and SCN4AB sequences as queries. The retrieved protein sequences were aligned using MAFFT v7.453 (Katoh and Standley, 2013) with the -auto mode. The alignment was manually curated using Jalview v2.11.1.4 (Waterhouse et al., 2005).

## RESULTS

### Genome Size and Assembly Completeness

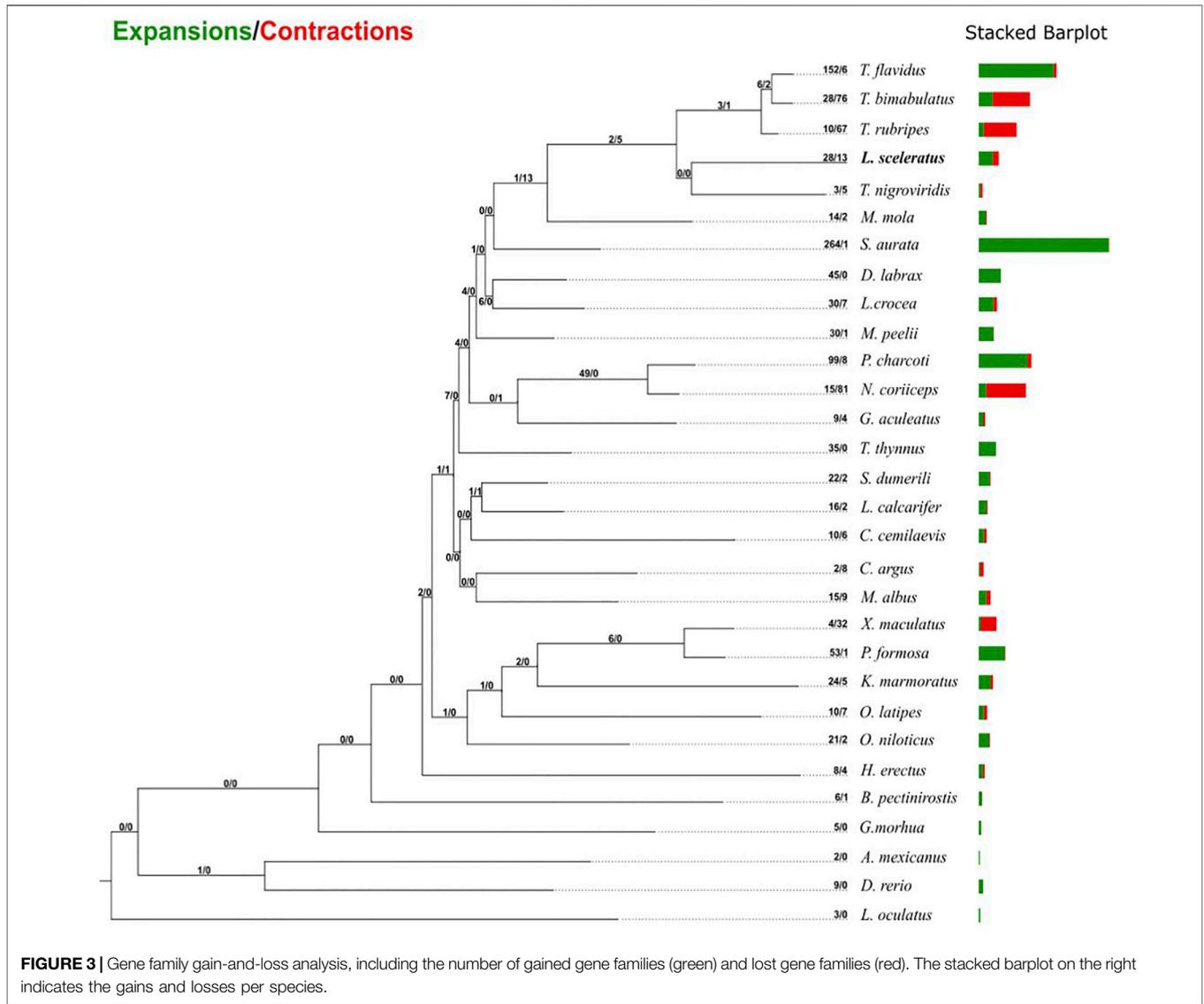
Sequencing yielded 57.30 Gb of raw Illumina reads and 9.68 Gb of ONT reads above Q7, with N50 of 48.85 Kb. The estimated genome size was calculated *via* kmergenie (Supplementary Figure S1) and found to be ~360 Mb and the best predicted  $k = 81$ . After quality trimming and filtering, we retained 44.45 Gb of Illumina data for genome polishing and 9.67 Gb of ONT data (Table 1) for the genome assembly. The final assembled and polished genome contained 235 contigs with total length of ~373 Mb, with 41 contigs representing ~91% of the genome and the largest contig sizing 17 Mb and N50 of 11 Mb (Table 2). In addition, kmer frequency profiling showed low levels of heterozygosity (0.521983%), as calculated using Genomescope2 (Supplementary Figure S2). Regarding genome completeness, we found 98% (4,513 out of 4,584) of the genes included in the BUSCO Actinopterygian gene set. Of those, 96.20% (4,410) were complete (Table 2), suggesting a high level of completeness and contiguity in the built assembly.

### Repeat Annotation, Gene Prediction and Functional Annotation

A total of 61.9 Mb of repeat sequences that accounted for 16.55% of the genome assembly were masked in *L. sceleratus*. The class of Retroelements makes up 7.81% of the total assembly and LINES are the most abundant of this class, with 5.54%. LTR elements sequences (2.07%) is the second most abundant group in the Retroelements class, and the results also indicated that 2.30% of the genome assembly consists of sequences of the class DNA transposons (Table 3).

### Gene Prediction and Functional Annotation

The combination of *ab initio*-based, homologue-based and RNASeq-based methods resulted in 32,451 putative protein-coding genes. After removing putative genes with Annotation Edit Distance (AED) (Eilbeck et al., 2009) score below one ( $AED < 1$ ), with custom script (longestIsoforms.py), we ended up with 21,251 genes of average gene length and exon size 587,43 and 249,62 bp, respectively. A total of 20,578 genes were successfully annotated, accounting for 97% of the predicted gene set (Table 4). Using BLAST to search the longest proteins encoded by each locus, a total of 17,706 *L. sceleratus* genes had a match in *T. rubripes*, and 16,706 in *T. nigroviridis*. Of those matches, we found 14,271 reciprocal best hits in *T. rubripes* and 13,425 reciprocal best hits in *T. nigroviridis*.



The completeness of the gene set was assessed using BUSCO v4.0.5 (Simão et al., 2015). From a core set of 3,640 single-copy ortholog genes from the *Actinopterygii* (odb 10) lineage, 92.2% were complete (70.5% as single-copy, 21.7% as duplicates), 1.7% were fragmented and 6.1% were not found.

### Orthology Assignment and Phylogenomic Analysis

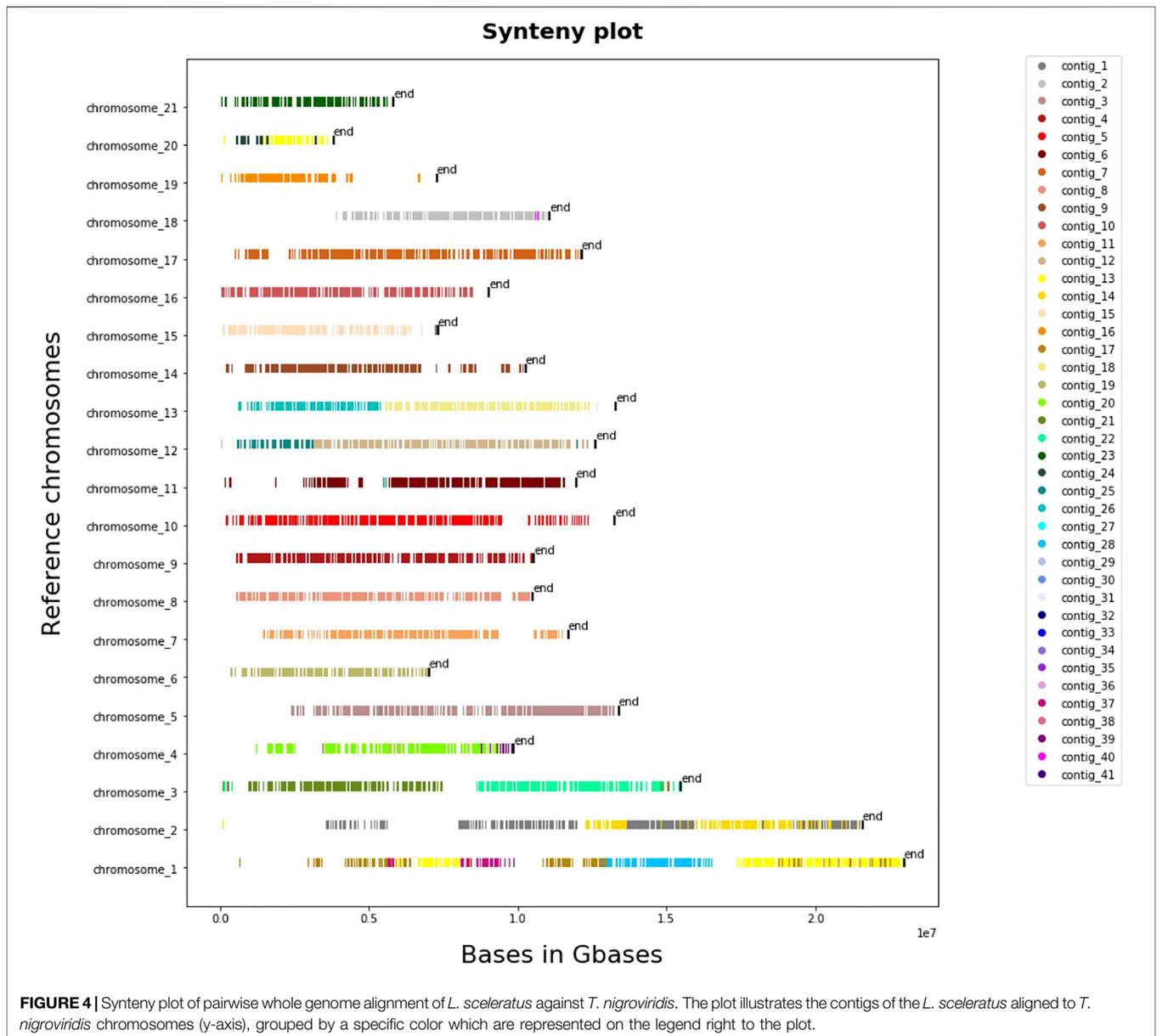
The total number of genes from all 30 fish proteomes (Supplementary Table S3) analysed by Orthofinder was 731,383 while 21,897 orthogroups were identified. After filtering, we chose 731 one-to-one orthogroups to construct the super-alignment. The initial matrix consisted of 494,732 amino acid positions. The Gblocks-filtered matrix contained 252,477 positions (51% of the initial), which were used for the phylogenomic analysis.

We identified JTT + I + G4 + F as the best model which was used for the estimation of the phylogenetic tree (Figure 2). At

the resulted phylogeny, almost all branches were supported with 100 bootstraps. The recovered phylogenetic position of *L. sceleratus* is within Tetraodontidae clade and is placed closer to *T. nigroviridis*.

### Gene Family Evolution

Gene family evolutionary analysis revealed multiple predicted rapidly expanded and contracted gene families of *L. sceleratus* (Figure 3; Supplementary Tables S4, S5), respectively. Gene families found only in one or more pufferfish species and not in any other taxon were considered as puffer-specific. Among rapidly evolving families, four gene families were identified as *L. sceleratus* specific, two as rapidly expanding (Supplementary Table S1) and two gene families as rapidly contracting. Finally, 43 *T. bimaculatus* and 45 *T. rubripes* families were found to be rapidly contracted (Supplementary Figures S1–S2). The gene family analysis did not reveal a core set of rapidly evolving families at the base of all five puffer species (Supplementary Figures S2–S3).



## Synteny Analysis

### Whole Genome-Based Synteny

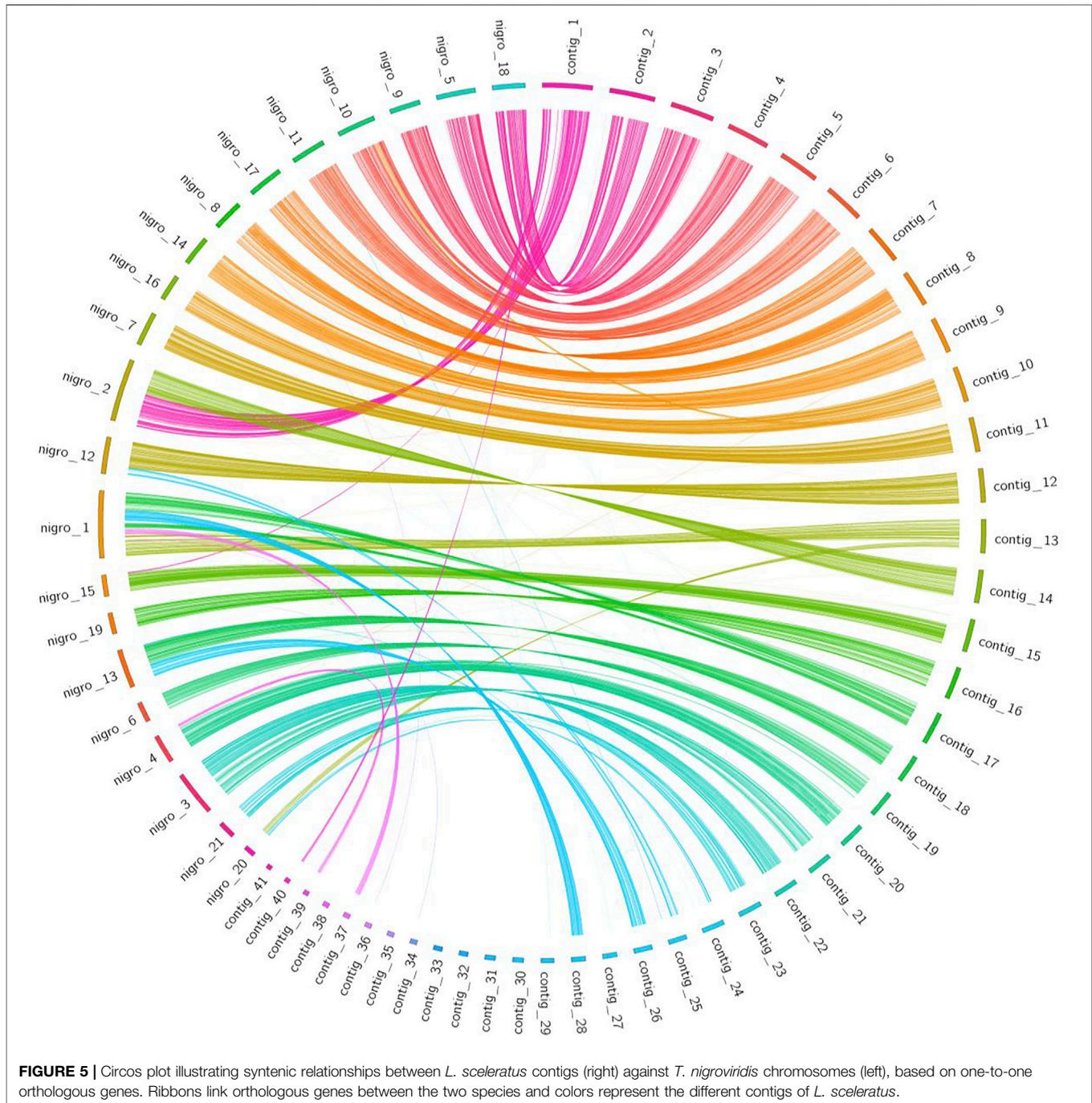
The results that we obtained from whole-genome pairwise alignment of *L. sceleratus* against the other four puffers (*T. nigroviridis*, *T. rubripes*, *T. flavidus* and *T. bimaculatus*) are summarised in **Supplementary Table S2**. Comparisons of *L. sceleratus* against *T. nigroviridis* (**Figure 4**), *T. rubripes*, *T. flavidus* and *T. bimaculatus* (**Supplementary Figures S9–S11**) indicated high collinearity of the studied species' genome against the rest. In particular, we aligned the 41 largest contigs of *L. sceleratus* against the chromosomes of *T. flavidus* (~231.3 aligned Mb), *T. bimaculatus* (~224 aligned Mb), *T. rubripes* (~227 aligned Mb) and *T. nigroviridis* (~154.7 aligned Mb).

Whole genome alignments were also carried out between *T. nigroviridis* and *T. flavidus*, as well as *T. bimaculatus* and *T.*

*rubripes* (**Supplementary Figures S4–S6**) revealing highly contiguous matches across all species, with higher collinearity observed between *T. nigroviridis* and *T. bimaculatus* compared to *T. nigroviridis* and *T. rubripes*.

### Gene-Based Synteny Analysis

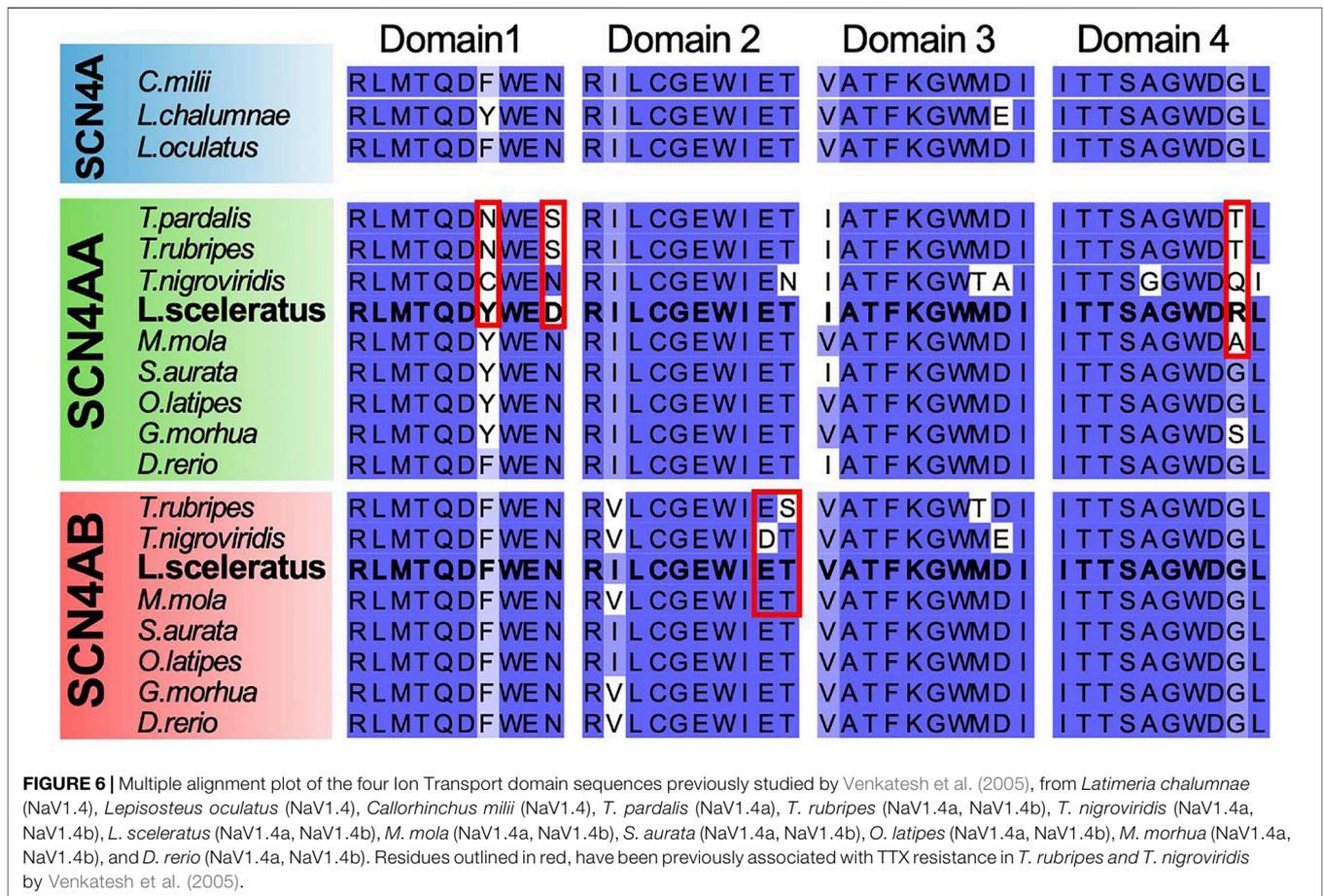
Synteny information obtained from a gene-based analysis revealed higher conserved synteny between *L. sceleratus* and *T. nigroviridis* (**Figure 5**) than between *L. sceleratus* and *T. rubripes* or *T. bimaculatus* (**Supplementary Figures S14–S15**). *L. sceleratus*' contigs that are unaligned to *T. nigroviridis* chromosomes in **Figure 6**, were aligned to unplaced regions of the *T. nigroviridis* assembly (**Supplementary Figure S6**).



## *L. sceleratus* Lacks SCN4A Mutations Previously Associated With Pufferfish TTX Resistance

The SCN4A protein has been previously associated with pufferfish TTX resistance, showing that mutations in the core of its ion transport domain led to changes in TTX binding affinity. We performed multiple sequence alignment

of the ion transport domain sequences from different Tetraodontiformes species and selected outgroups, including teleosts (SCN4AA and SCN4AB paralogs) and other vertebrates. We found that *L. sceleratus* and *M. mola* lack the previously characterized mutations in domains 1 and 2 in residues outlined in red. In contrast, all Tetraodontiformes show changes in the second to last residue of the domain 4 sequence outlined in red in **Figure 6**.



## DISCUSSION

### Genome Size and Assembly Completeness

In this study, a high-quality pufferfish genome assembly of high contiguity was reconstructed, from data obtained from a single MinION flow cell and half a lane of Illumina HiSeq. To our knowledge, only one other highly contiguous reference Tetraodontiformes genome assembly has previously been constructed using the same strategy, for *Thamnaconus septentrionalis* (Bian et al., 2019). The assembly of *L. sceleratus* (360 Mb) is consistent with the predicted genome size from kmergenie and is comparable in size with that of other puffers, such as *Fugu rubripes* (~365 Mb; Aparicio et al., 2002), *T. flavidus* (~377 Mb; Zhou Y. et al., 2019), *T. bimaculatus* (~393.15 Mb; Zhou Z. et al., 2019), *T. obscurus* (~373 Mb; Kang et al., 2020) and *T. nigroviridis* (340 Mb, Jaillon et al., 2004). The contig N50 value (~11 Mb) of the *L. sceleratus* assembly is considerably greater than that reported for the genomes of *T. bimaculatus* (1.31 Mb; Zhou Z. et al., 2019) and *T. flavidus* (4.4 Mb; Zhou Y. et al., 2019). Similarly, our assembly appears of equivalent levels of completeness to other Tetraodontidae genomes, based on BUSCO scores (e.g., *T. obscurus* (Kang et al., 2020) and *T. flavidus* (Zhou et al., 2019a)].

### Repeat Content, Gene Prediction and Functional Annotation

The percentage of transposable elements (TEs) found in the *L. sceleratus* genome (16.55% of the assembled genome) is marginally higher than the one found in *T. septentrionalis* (14.2%) (Bian et al., 2019), *T. obscurus* (11.05%) (Kang et al., 2020), and *M. mola* (11%) (Pan et al., 2016). Moreover, it is almost twofold higher than in *T. rubripes* (7.53%) and threefold higher than *T. nigroviridis* (5.60%) and *T. flavidus* (6.87%) (Gao et al., 2014). *T. rubripes* contain more copies of transposable elements than *T. nigroviridis*, which have been proposed to contribute to its marginally larger genome size (365–370 Mb) (Jatllon et al., 2004). Although the *L. sceleratus* genome has a comparable size to reported *Takifugu* genomes, it harbors much higher repeat content. Moreover, *D. holocanthus* genome of the Diodontidae family contains 36.35% repetitive sequences, almost double the repeat content of *L. sceleratus*. These findings imply that TEs might follow an independent pathway of accumulation and diversification across Tetraodontiformes species. In the case of *L. sceleratus*, such differential repeat expansion may have taken place after the divergence of the *Takifugu* and Tetraodon genera.

Despite such TE content variation across closely related taxa, positive correlation of genome size and TE repeat content has

been documented across a larger evolutionary scale in teleosts (Shao et al., 2019). For example, the relatively smaller genome of *T. nigroviridis* (~360 Mb) contains 5.6% TEs, in contrast to the zebrafish genome (~1.4 Gb) which is composed of 55% repetitive sequences (Shao et al., 2019). This positive correlation is also reflected in the small size and relatively low repeat content of the *L. sceleratus* genome, regardless of differences with other pufferfish. However, it would be interesting to further explore these differences, as they may be informative for genome evolution. As an interesting example, LINE elements are the most abundant in the *L. sceleratus* genome, with ~170,000 copies, as compared to the ~12,300 copies of the *T. rubripes* genome. This finding indicates dynamic genome evolution in the two species. Previous studies have shown a correlation between genome TEs and species adaptations to new environments, suggesting they may be associated to invasiveness (Stapley et al., 2015; Yuan et al., 2018). Thus, the repeat content of *L. sceleratus* may play a role in its fast adaptation to novel environments and should be investigated further.

## Species Tree Reconstruction

Although the order Tetraodontiformes is a cosmopolitan taxonomic group that includes multiple families, large parts of their phylogenetic relationships remain unexplored. In this study, we presented the first phylogenetic tree based on whole genome data including the invasive “sprinter” *L. sceleratus*. The recovered phylogenetic position of *L. sceleratus* is within Tetraodontidae and is placed closer to *T. nigroviridis*, while the long branch length of the Tetraodontidae clade possibly suggests a faster evolutionary rate. Regarding relationships within the pufferfish group (*T. nigroviridis*, *T. rubripes*, *T. flavidus*, *T. bimaculatus* and *L. sceleratus*), the resulting topology agrees with previous studies (Yamanoue et al., 2011; Meynard et al., 2012; Hughes et al., 2018; Hughes et al., 2021). Moreover, the Tetraodontidae group was recovered confidently as monophyletic in accordance with Yamanoue et al. (2011). Our results suggest that Tetraodontiformes are the closest group to Sparidae and corroborates the results of Natsidis et al. (2019) and of others (Kawahara et al., 2008; Meynard et al., 2012), based both on six mitochondrial and two nuclear genes.

## Synteny Analysis

All pairwise comparisons of the whole-genome alignment analysis of *L. sceleratus* against the four other Tetraodontidae species (Figure 5) (Supplementary Figures S9–S11), showed highly conserved synteny. The genome that exhibited the highest synteny conservation with the *L. sceleratus* genome was that of *T. nigroviridis*, in accordance with our reconstructed phylogeny which places the two species as more closely related to each other compared to the rest.

The synteny between *L. sceleratus* and the three species of the genus *Takifugu* (*T. rubripes*, *T. bimaculatus* and *T. flavidus*) was less conserved, especially between *L. sceleratus* and *T. bimaculatus*.

To sum up, the higher synteny between *L. sceleratus* and *T. nigroviridis* corroborates their closer phylogenetic position compared to the three *Takifugu* species.

## Gene Family Evolution and Adaptation

Adapting to a new habitat is a challenging task for a species, requiring a certain degree of physiological plasticity. To achieve establishment in a new niche, an invader must face environmental challenges that involve both biotic and abiotic factors (Crowl et al., 2008). Invasive species are facing novel pathogens during the colonisation of new environments and the ability to deal with these new immune challenges is key to their invasive success (Lee and Klasing, 2004). Interestingly, we found several expanded immune related families, including *immunoglobulins* (*C-Type* and *V-Type*), *Ig heavy chain Mem5-like*, *B-cell receptors* and the *Fish-specific NACHT associated domain*, which are related to the innate immunity (Stein et al., 2007).

In addition, we also detected major histocompatibility complex (*MHC class I*) genes in the expanded gene families. MHC genes are crucial for the immune response, involved in pathogen recognition by T cells (Germain, 1994), thus initiating the adaptive immune response. The expanded repertoire of *L. sceleratus* immune response associated genes might be related to its survival in novel habitats, through the detection and inhibition of a wide range pathogens. Therefore, in this context, we suggest further research to explore the role of the expanded genes related to immune response.

Another interesting finding was the expansion of the fucosyltransferase (FUT) gene family. In particular, we detected 24 FUT9 [alpha (1.3) fucosyltransferase 9] genes. Glycosylation is one of the most frequent post-translational modifications of a protein. Many proteins involved in the immune response are glycosylated, extending their diversity and functionality (Bednarska et al., 2017). Fucosylation, a type of glycosylation, plays an essential role in cell proliferation, metastasis and immune escape (Jia et al., 2018). In mice, FucTC has been shown to regulate leukocyte trafficking between blood and the lymphatic system, after its engagement in selectin ligand biosynthesis (Maly et al., 1996).

Overall, based on our results, we believe that the expanded innate immune system gene families identified herein deserve further study as to whether they have contributed to the ability of *L. sceleratus* to spread rapidly (Kalogirou, 2013).

## TTX Resistance of SCNA4 Sodium Channel

Sodium channels (NaV) are formed by an  $\alpha$  subunit consisting of four domains (I-IV) and an optional  $\beta$  subunit that may alter its activity (Yu et al., 2005). Furthermore, the third whole-genome duplication of teleosts has expanded the  $\alpha$  subunit family to eight members (Zakon et al., 2011). NaV channels are the target of several neurotoxins (Daly, 1995; Fry et al., 2009) and TTX is known to bind to the outer pore of the NaV1.4 (SCN4) channel, blocking the transport of sodium ions across the pore (Hanifin, 2010). As different pufferfishes have been shown to acquire TTX resistance through mutations in specific domains of NaV1.4 (Venkatesh et al., 2005; Soong and Venkatesh, 2006; Jost et al., 2008), we checked if this is also the case for *L. sceleratus*. For this purpose, we investigated whether the sequences of voltage-gated sodium channels (SCN4AA, SCN4AB) in *L. sceleratus* carry the previously reported mutations associated with TTX resistance in

other pufferfish. According to these studies (Venkatesh et al., 2005; Soong and Venkatesh, 2006), the residues that are associated with TTX resistance are located in the same position in NaV1.4a Domain I and were mutated to Cys and Asn, in *T. nigroviridis* and *T. rubripes*, respectively (Figure 6). Surprisingly, despite the fact that the Domain I mutations lead to extensive decrease in TTX binding (Satin et al., 1992; Kaneko et al., 1997; Yotsu-Yamashita et al., 2000; Venkatesh et al., 2005) and have been associated with TTX resistance in pufferfishes and in one *C. pyrrhogaster* newt (Kaneko et al., 1997), we did not observe any of these reported mutations in the SCN4A\_A gene of *L. sceleratus* (Figure 6). Similarly, no changes from the ancestral sequence of the same paralog were found in *M. mola*, which is also a toxin resistant species (Halstead, 1988). The absence of mutations in *L. sceleratus* and *M. mola*, in conjunction with the updated Tetraodontiformes relationships presented in Figure 3, suggest that the changes seen in *T. rubripes* and *T. nigroviridis* have arisen via convergent evolution, a hypothesis further supported by the different amino acid replacement observed in these two species. This could also suggest that the position of the replacement is more important than the amino acid change *per se*. Our results are also congruent with previous findings for *Haplochlaena lunulata* (Geffeney et al., 2019) and *Haplochlaena maculosa* (Whitelaw et al., 2020), as NaV1.4a Domain I is highly conserved in both octopod species. Furthermore, unlike the pufferfishes *A. nigropunctatus* and *T. nigroviridis* (Jost et al., 2008), the garter snake *Thamnophis couchii* (Feldman et al., 2012) and the octopuses *Haplochlaena lunulata* (Geffeney et al., 2019) and *Haplochlaena maculosa* (Whitelaw et al., 2020), replacements were also not observed in the NaV1.4a Domain III of *L. sceleratus*. Similar to *L. sceleratus*, mutated residues are not found in *T. rubripes* either. However, in NaV1.4a Domain IV all the Tetraodontiformes species have a replacement in the same position (outlined in red in Figure 7). Interestingly, a different mutation is observed in each species, except for *T. rubripes* and *T. pardalis*, which both share a change to Thr. Similar to what is discussed earlier for the changes observed in domain I, the position of the replacement may be more crucial than the specific changes observed. This could hint to a decrease in TTX binding by disturbing the conserved ancestral binding interface, as previously shown in domain I mutations in *T. rubripes* and *T. nigroviridis* (Venkatesh et al., 2005; Soong and Venkatesh, 2006). Previously, these mutations for *T. rubripes* and *T. nigroviridis* had not been associated with TTX resistance (Venkatesh et al., 2005; Soong and Venkatesh, 2006). At the last positions of Domain IV, there are also two replacements to His and Ser in *H. lunulata* (Geffeney et al., 2019) and *H. maculosa* (Whitelaw et al., 2020), which may inhibit TTX binding.

The same studies (Venkatesh et al., 2005; Soong and Venkatesh, 2006; Jost et al., 2008) also showed that an Asp mutation in Domain II of *T. nigroviridis* NaV1.4b is highly associated with TTX resistance. A similar substitution has been reported in the soft-shelled toxin-resistant clam *M. arenaria* (Bricelj et al., 2005). Strikingly, the *L. sceleratus* and *M. mola* NaV1.4b sequences also lack mutations in Domain II, comparable to NaV1.4a domain I (Figure 6).

While the origin of TTX resistance remains elusive, several studies have provided insight into resistance mechanisms. Toxin tolerance may appear through mutations in sodium channels or toxin binding proteins (Ho et al., 1994; Zou, 2020). A wide range of different organisms have taken advantage of mutations which confer TTX tolerance either independently or complementary. The lack of previously characterised pufferfish mutations associated with TTX resistance in *L. sceleratus* raises several questions about how TTX resistance has evolved. Nevertheless, the results of our analysis imply that combined effects of complex polygenic adaptations working redundantly have played a role in the evolution of this complex trait, while similar genetic changes have arisen convergently multiple times.

## CONCLUSION

Invader fishes, such as *L. sceleratus*, often thrive in novel environments. Our analysis provides the first high-quality genome assembly built from sequencing a single female individual and a comprehensive evolutionary genomic analysis of the species. We uncovered a close phylogenetic position of *L. sceleratus* with *T. nigroviridis*, untangling relationships within the pufferfish group, that were not clearly resolved in previous studies. The study gives insights into a variety of genomic signatures that may be associated with *L. sceleratus* invasion and colonisation effectiveness. Surprisingly, examination of voltage-gated sodium channels (NaV1.4) revealed a lack of TTX resistance associated mutations found in other pufferfishes, highlighting the complex evolution of the trait. Overall, the *L. sceleratus* genome will be an invaluable resource for additional studies on immune response in novel environments, osmoregulation, reconstruction of ancient chromosome rearrangements, investigation of complex TTX resistance mechanisms and population genomics and adaptation. Such studies are expected to elucidate the mechanisms behind the high invasiveness of *L. sceleratus* and assist the management of this invasive sprinter in the Mediterranean.

## DATA AVAILABILITY STATEMENT

Genomic and transcriptomic Illumina and Oxford Nanopore raw data can be accessed in ENA with the IDs ERX6020380, ERS6677531 and ERR6391424 respectively. The genome assembly has been uploaded with the ENA ID CAJVRN01000000.

## ETHICS STATEMENT

Ethical review and approval was not required for the animal study because this study does not include experimentation on laboratory animals for which a protocol approval is required by the law. Further, as stated within the manuscript we followed guidelines for animal care and handling [Guidelines for the

treatment of animals in behavioral research and teaching. *Anim. Behav.* 53, 229–234 (1997)].

## AUTHOR CONTRIBUTIONS

TM, CT, VP, and TD designed research; TD, VP, AT, JBC, and AS performed research and/or analysed data; TD, VP, and TM wrote the paper. All authors reviewed the manuscript.

## ACKNOWLEDGMENTS

This project received internal funding from the HCMR programme SUPP-GEN: “Research development of environmental genetics and biotechnology (SR: AM).

## REFERENCES

- Akbora, H. D., Kunter, İ., Erçetin, T., Elagöz, A. M., and Çiçek, B. A. (2020). Determination of Tetrodotoxin (TTX) Levels in Various Tissues of the Silver Cheeked Puffer Fish (*Lagocephalus Sceleratus* (Gmelin, 1789)) in Northern Cyprus Sea (Eastern Mediterranean). *Toxicol.* 175, 1–6. FebEpub 2019 Dec 4. PMID: 31833474. doi:10.1016/j.toxicol.2019.12.002
- Akyol, O., Unal, V., Ceyhan, T., and Bilecenoglu, M. (2005). First Confirmed Record of *Lagocephalus sceleratus*(Gmelin, 1789) in the Mediterranean Sea. *J. Fish Biol.* 66, 1183–1186. doi:10.1111/j.0022-1112.2005.00667.x
- Akyol, O., and Ünal, V. (2017). Long Journey of *Lagocephalus Sceleratus* (Gmelin, 1789) throughout the Mediterranean Sea. *Nat. Eng. Sci.* 2 (3), 41–47. doi:10.28978/nesciences.369534
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215 (3), 403–410. Oct 5PMID: 2231712. doi:10.1016/S0022-2836(05)80360-2
- Andrews, S. (2010). *FASTQC. A Quality Control Tool for High Throughput Sequence Data.*
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-m., Dehal, P., et al. (2002). Whole-genome Shotgun Assembly and Analysis of the Genome of *Fugu Rubripes*. *Science* 297, 1301–1310. doi:10.1126/science.1072104
- Arim, M., Abades, S. R., Neill, P. E., Lima, M., and Marquet, P. A. (2006). Spread Dynamics of Invasive Species. *Proc. Natl. Acad. Sci.* 103, 374–378. doi:10.1073/pnas.0504272102
- Bao, Z., and Eddy, S. R. (2002). Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Res.* 12, 1269–1276. doi:10.1101/gr.88502
- Bednarska, N. G., Wren, B. W., and Willcocks, S. J. (2017). The Importance of the Glycosylation of Antimicrobial Peptides: Natural and Synthetic Approaches. In *Drug Discovery Today. Drug Discov. Today* 22 (Issue 6), 919–926. doi:10.1016/j.drudis.2017.02.001
- Bian, L., Li, F., Ge, J., Wang, P., Chang, Q., Zhang, S., et al. (2019). Chromosome-level Genome Assembly of the Greenfin Horse-Faced Filefish (*Thamnaconus Septentrionalis*) Using Oxford Nanopore PromethION Sequencing and Hi-C Technology. *bioRxiv Genomics*, 1–25. doi:10.1101/798744
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Bricelj, V. M., Connell, L., Konoki, K., MacQuarrie, S. P., Scheuer, T., Catterall, W. A., et al. (2005). Sodium Channel Mutation Leading to Saxitoxin Resistance in Clams Increases Risk of PSP. *Nature* 434, 763–767. doi:10.1038/nature034110.1038/nature03415
- Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17, 540–552. doi:10.1093/oxfordjournals.molbev.a026334
- Further, this research was supported through computational resources provided by IMBBC (Institute of Marine Biology, Biotechnology and Aquaculture) of the HCMR (Hellenic Centre for Marine Research)–Zorbas HPC infrastructure (Zafeiropoulos et al., 2021). Funding for establishing the IMBBC HPC has been received by the MARBIGEN (EU Regpot) project, LifeWatchGreece RI, and the CMBR (Centre for the study and sustainable exploitation of Marine Biological Resources).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.790850/full#supplementary-material>

- Chikhi, R., and Medvedev, P. (2014). Informed and Automated K-Mer Size Selection for Genome Assembly. *Bioinformatics* 30, 31–37. doi:10.1093/bioinformatics/btt310
- Chua, H. H., and Chew, L. P. (2009). Puffer Fish Poisoning: a Family Affair. *Med. J. Malaysia* 64 (2), 181–182. JunPMID: 20058587.
- Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., and Lugo, A. E. (2008). The Spread of Invasive Species and Infectious Disease as Drivers of Ecosystem Change. *Front. Ecol. Environ.* 6, 238–246. doi:10.1890/070151
- Daly, J. W. (1995). The Chemistry of Poisons in Amphibian Skin. *Proc. Natl. Acad. Sci. U S A.* 92, 9–13. doi:10.1073/pnas.92.1.9
- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., Flouri, T., et al. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* 37 (Issue 1), 291–294. January. doi:10.1093/molbev/msz189
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a Computational Tool for the Study of Gene Family Evolution. *Bioinformatics* 22, 1269–1271. doi:10.1093/bioinformatics/btl097
- Eilbeck, K., Moore, B., Holt, C., and Yandell, M. (2009). Quantitative Measures for the Management and Comparison of Annotated Genomes. *BMC Bioinformatics*, 10, 67. doi:10.1186/1471-2105-10-67
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an Efficient and Flexible Software for De Novo Detection of LTR Retrotransposons. *BMC Bioinformatics* 9, 18. doi:10.1186/1471-2105-9-18
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy. *Genome Biol.* 16, 157. doi:10.1186/s13059-015-0721-Enright.AJ
- Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An Efficient Algorithm For Large-Scale Detection Of Protein Families. *Nucleic Acids Res.* 30 (7), 1575–1584. doi:10.1093/nar/30.7.1575
- Feldman, C. R., Brodie, E. D., Jr, Brodie, E. D., 3rd, and Pfrender, M. E. (2012). Constraint Shapes Convergence in Tetrodotoxin-Resistant Sodium Channels of Snakes. *Proc. Natl. Acad. Sci.* 109 (12), 4556–4561. Mar 20. doi:10.1073/pnas.1113468109
- Filiz, H., and Er, M. (2004). *Akdenizin Yeni Misafiri (New Guests in the Mediterranean Sea).*
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc. Natl. Acad. Sci. USA* 117, 9451–9457. doi:10.1073/pnas.1921046117
- Fry, B. G., Roelants, K., Champagne, D. E., Scheib, H., Tyndall, J. D. A., King, G. F., et al. (2009). The Toxicogenomic Multiverse: Convergent Recruitment of Proteins into Animal Venoms. *Annu. Rev. Genom. Hum. Genet.* 10, 483–511. doi:10.1146/annurev.genom.9.081307.164356
- Gao, Y., Gao, Q., Zhang, H., Wang, L., Zhang, F., Yang, C., et al. (2014). Draft Sequencing and Analysis of the Genome of Pufferfish Takifugu Flavidus. *DNA Res.* 21 (6), 627–637. doi:10.1093/dnares/dsu025

- Geffeney, S. L., Williams, B. L., Rosenthal, J. J. C., Birk, M. A., Felkins, J., Wisell, C. M., et al. (2019). Convergent and Parallel Evolution in a Voltage-Gated Sodium Channel Underlies TTX-Resistance in the Greater Blue-Ringed octopus: *Hapalochlaena lunulata*. *Toxicon* 170, 77–84. doi:10.1016/j.toxicon.2019.09.013
- Germain, R. N. (1994). MHC-dependent Antigen Processing and Peptide Presentation: Providing Ligands for T Lymphocyte Activation. *CellElsevier* 76 (2), 287–299. doi:10.1016/0092-8674(94)90336-0
- Golani, D., and Appelbaum-Golani, B. (2010). FISH INVASIONS OF THE MEDITERRANEAN SEA: *Change and Renewal*. Sofia Bulgaria: PENSOFT Publishers.
- Golani, D., and Appelbaum-Golani, B. (1977). Guidelines for the Treatment of Animals in Behavioral Research and Teaching. *Anim. Behav.* 53, 229–234.
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* 29, 1072–1075. doi:10.1093/bioinformatics/btt086
- Halstead, B. W. (1988). *Poisonous and Venomous Marine Animals of the World: Vertebrates*.
- Hanifin, C. T. (2010). The Chemical and Evolutionary Ecology of Tetrodotoxin (TTX) Toxicity in Terrestrial Vertebrates. *Mar. Drugs* 8, 577–593. doi:10.3390/md8030577
- Ho, B., Yeo, D. S. A., and Ding, J. L. (1994). A Tetrodotoxin Neutralizing System in the Haemolymph of the Horseshoe Crab, *Carciniscorpius rotundicauda*. *Toxicon* 32 (7), 755–762. Jul. doi:10.1016/0041-0101(94)90001-9
- Holt, C., and Yandell, M. (2011). MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects. *BMC Bioinformatics* 12. doi:10.1186/1471-2105-12-491
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast Genome-wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi:10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* 47, D309–D314. doi:10.1093/nar/gky1085
- Hughes, L. C., Orti, G., Huang, Y., Sun, Y., Baldwin, C. C., Thompson, A. W., et al. (2018). Comprehensive Phylogeny of ray-finned Fishes (Actinopterygii) Based on Transcriptomic and Genomic Data. *Proc. Natl. Acad. Sci. USA* 115, 6249–6254. doi:10.1073/pnas.1719358115
- Hughes, L. C., Orti, G., Saad, H., Li, C., White, W. T., Baldwin, C. C., et al. (2021). Exon Probe Sets and Bioinformatics Pipelines for All Levels of Fish Phylogenomics. *Mol. Ecol. Resour.* 21, 816–833. doi:10.1111/1755-0998.13287
- Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., et al. (2004). Genome Duplication in the Teleost Fish *Tetraodon nigroviridis* Reveals the Early Vertebrate Proto-Karyotype. *Nature* 431, 946–957. doi:10.1038/nature03025
- Jia, L., Zhang, J., Ma, T., Guo, Y., Yu, Y., and Cui, J. (2018). The Function of Fucosylation in Progression of Lung Cancer. *Front. Oncol.* 8, 565. doi:10.3389/fonc.2018.00565
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* 30, 1236–1240. doi:10.1093/bioinformatics/btu031
- Jost-Hillis, M. C., Hillis, D. M., Lu, Y., Kyle, J. W., Fozzard, H. A., and Zakon, H. H. (2008). Toxin-Resistant Sodium Channels: Parallel Adaptive Evolution across a Complete Gene Family. *Mol. Biol. Evol.* 25 (Issue 6), 1016–1024. June. doi:10.1093/molbev/msn025
- Kalogirou, S. (2013). Ecological Characteristics of the Invasive Pufferfish *Lagocephalus sceleratus* (Gmelin, 1789) in the Eastern Mediterranean Sea - a Case Study from Rhodes. *Medit. Mar. Sci.*, 14(2), 251–260. doi:10.12681/mms.364
- Kaneko, Y., Matsumoto, G., and Hanyu, Y. (1997). TTX Resistivity of Na<sup>+</sup>-Channel in Newt Retinal Neuron. *Biochem. Biophysical Res. Commun.* 240 (3), 651–656. doi:10.1006/bbrc.1997.7696
- Kang, S., Kim, J. H., Jo, E., Lee, S. J., Jung, J., Kim, B. M., et al. (2020). Chromosomal-level Assembly of Takifugu Obscurus (Abe, 1949) Genome Using Third-generation DNA Sequencing and Hi-C Analysis. *Mol. Ecol. Resour.* 20, 520–530. doi:10.1111/1755-0998.13132
- Kasapidis, P., Peristeraki, P., Tserpes, G., and Magoulas, A. (2007). First Record of the Lessepsian Migrant *Lagocephalus Sceleratus* (Gmelin 1789) (Osteichthyes: Tetraodontidae) in the Cretan Sea (Aegean, Greece). *Ai 2*, 71–73. doi:10.3391/ai.2007.2.1.9
- Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010
- Kawahara, R., Miya, M., Mabuchi, K., Lavoué, S., Inoue, J. G., Satoh, T. P., et al. (2008). Interrelationships of the 11 Gasterosteiform Families (Sticklebacks, Pipefishes, and Their Relatives): A New Perspective Based on Whole Mitogenome Sequences from 75 Higher Teleosts. *Mol. Phylogenet. Evol.* 46, 224–236. doi:10.1016/j.ympev.2007.07.009
- Kielbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011). Adaptive Seeds Same Genomic Sequence Comparison. *Genome Res.* 21 (3), 487PMC3044862–493. MarEpub 2011 Jan 5. PMID: 21209072; PMCID. doi:10.1101/gr.113985.110
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317
- Kolar, C. S., and Lodge, D. M. (2001). *Progress in Invasion Biology*.
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* 37, 540–546. doi:10.1038/s41587-019-0072-8
- Korf, I. (2004). Gene Finding in Novel Genomes. *BMC Bioinformatics* 5, 59. doi:10.1186/1471-2105-5-59
- Kosker, A. R., Özogul, F., Durmus, M., Ucar, Y., Ayas, D., Regenstein, J. M., et al. (2016). Tetrodotoxin Levels in Pufferfish (*Lagocephalus Sceleratus*) Caught in the Northeastern Mediterranean Sea. *Food Chem. 210*, 332–337. Nov 1Epub 2016 Apr 27. PMID: 27211655. doi:10.1016/j.foodchem.2016.04.122
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., and RAxML-, N. G. (2019). RAxML-NG: a Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* 35 (Issue 21), 4453–4455. 1 November. doi:10.1093/bioinformatics/btz305
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an Information Aesthetic for Comparative Genomics. *Genome Res.* 19 (9), 1639–1645. SepEpub 2009 Jun 18. PMID: 19541911; PMCID: PMC275213. doi:10.1101/gr.092759.109
- Lee, K. A., and Klasing, K. C. (2004). A Role for Immunology in Invasion Biology. *Trends Ecol. Evol.* 19, 523–529. doi:10.1016/j.tree.2004.07.012
- Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191
- Li, W., and Godzik, A. (2006). Cd-hit: a Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158
- Malý, P., Thall, A. D., Petryniak, B., Rogers, C. E., Smith, P. L., Marks, R. M., et al. (1996). The  $\alpha(1,3)$ Fucosyltransferase Fuc-TVII Controls Leukocyte Trafficking through an Essential Role in L-, E-, and P-Selectin Ligand Biosynthesis. *Cell* 86, 643–653. doi:10.1016/s0092-8674(00)80137-3
- Meynard, C. N., Mouillot, D., Mouquet, N., and Douzery, E. J. P. (2012). A Phylogenetic Perspective on the Evolution of Mediterranean Teleost Fishes. *PLoS ONE* 7, e36443. doi:10.1371/journal.pone.0036443
- Nagashima, Y., and Arakawa, O. (2016). “Pufferfish Poisoning and Tetrodotoxin,” in *Marine and Freshwater Toxins*. Editors P. Gopalakrishnakone, V. Haddad Jr., A. Tubaro, E. Kim, and W. Kem, and, 259–284. doi:10.1007/978-94-007-6419-410.1007/978-94-007-6419-4\_12
- Natsidis, P., Tsakogiannis, A., Pavlidis, P., Tsigenopoulos, C. S., and Manousaki, T. (2019). Phylogenomics Investigation of Sparids (Teleostei: Spariformes) Using High-Quality Proteomes Highlights the Importance of Taxon Sampling. *Commun. Biol.* 2, 1–10. doi:10.1038/s42003-019-0654-5
- Ou, S., and Jiang, N. (2018). LTR\_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol.* 176, 1410–1422. doi:10.1104/pp.17.01310
- Palumbi, S. R. (2001). *The Evolution Explosion: How Humans Cause Rapid Evolutionary Change*. New York: W. W. Norton.
- Pan, H., Yu, H., Ravi, V., Li, C., Lee, A. P., Lian, M. M., et al. (2016). The Genome of the Largest Bony Fish, Ocean Sunfish (*Mola mola*), Provides Insights into its Fast Growth Rate. *GigaSci* 5, 36. doi:10.1186/s13742-016-0144-3

- Perteua, M., Perteua, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122
- Por, F. D. (1971). One Hundred Years of Suez Canal-A Century of Lessepsian Migration: Retrospect and Viewpoints. *Syst. Zool.* 20, 138–159. doi:10.2307/2412054
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De Novo identification of Repeat Families in Large Genomes. *Bioinformatics* 21, i351–i358. doi:10.1093/bioinformatics/bti1018
- Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for Reference-free Profiling of Polyploid Genomes. *Nat. Commun.* 11, 1432. doi:10.1038/s41467-020-14998-3
- Sanderson, M. J. (2003). r8s: Inferring Absolute Rates of Molecular Evolution and Divergence Times in the Absence of a Molecular Clock. *Bioinformatics* 19 (Issue 2), 301–302. 22 January. doi:10.1093/bioinformatics/19.2.301
- Satin, J., Kyle, J. W., Chen, M., Bell, P., Cribbs, L. L., Fozzard, H. A., et al. (1992). A Mutant of TTX-Resistant Cardiac Sodium Channels with TTX-Sensitive Properties. *Science* 256, 1202–1205. doi:10.1126/science.256.5060.1202
- Sax, D., Stachowicz, J., Brown, J., Bruno, J., Dawson, M., Gaines, S., et al. (2007). Ecological and Evolutionary Insights from Species Invasions. *Trends Ecol. Evol.* 22, 465–471. doi:10.1016/j.tree.2007.06.009
- Shao, F., Han, M., and Peng, Z. (2019). Evolution and Diversity of Transposable Elements in Fish Genomes. *Sci. Rep.* 9. doi:10.1038/s41598-019-51888-1
- Shao, F., Wang, J., Xu, H., and Peng, Z. (2018). FishTEDB: a Collective Database of Transposable Elements Identified in the Complete Genomes of Fish. *Database* 2018 (2018), 106. doi:10.1093/database/bax106
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351
- Soong, T., and Venkatesh, B. (2006). Adaptive Evolution of Tetrodotoxin Resistance in Animals. *Trends Genet.* 22 (11), 621–626. NovEpub 2006 Sep 7. doi:10.1016/j.tig.2006.08.010
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* 34, W435–W439. Jul 1 (Web Server issue) PMID: PMC1538822.4. doi:10.1093/nar/gkl200.PMID:16845043
- Stapley, J., Santure, A. W., and Dennis, S. R. (2015). Transposable Elements as Agents of Rapid Adaptation May Explain the Genetic Paradox of Invasive Species. *Mol. Ecol.* 24, 2241–2252. doi:10.1111/mec.13089
- Stein, C., Caccamo, M., Laird, G., and Leptin, M. (2007). Conservation and Divergence of Gene Families Encoding Components of Innate Immune Response Systems in Zebrafish. *Genome Biol.* 8, R251. doi:10.1186/gb-2007-8-11-r251
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinformatics* 25, 2009 4 (Issue SUPPL. 25). doi:10.1002/0471250953.bi0410s25
- Vaser, R., and Šikić, M. (2020). Raven: A De Novo Genome Assembler for Long Reads. *bioRxiv*, 2020. doi:10.1101/2020.08.07.242461
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and Accurate De Novo Genome Assembly from Long Uncorrected Reads. *Genome Res.* 27, 737–746. doi:10.1101/gr.214270.116
- Venkatesh, B., Lu, S. Q., Dandona, N., See, S. L., Brenner, S., and Soong, T. W. (2005). Genetic Basis of Tetrodotoxin Resistance in Pufferfishes. *Curr. Biol.* 15, 2069–2072. doi:10.1016/j.cub.2005.10.068
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* 9, e112963. doi:10.1371/journal.pone.0112963
- Waterhouse, A., Procter, J., Martin, D. A., and Barton, G. J. (2005). Jalview: Visualization and Analysis of Molecular Sequences, Alignments, and Structures. *BMC Bioinformatics* 6, P28. doi:10.1186/1471-2105-6-S3-P28
- Wheeler, T. J. (2009). Large-scale Neighbor-Joining with NINJA. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes. *Bioinformatics* 5724 LNBI, 375–389. doi:10.1007/978-3-642-04241-6\_31
- Whitelaw, B. L., CookeCooke, I. R., Finn, J., da Fonseca, R. R., Ritschard, E. A., Gilbert, M. T. P., et al. (2020). Adaptive Venom Evolution and Toxicity in Octopods Is Driven by Extensive Novel Gene Formation, Expansion, and Loss. *GigaScience* 9 (Issue), 11, November. doi:10.1093/gigascience/giaa120
- Yamanoue, Y., Miya, M., Doi, H., Mabuchi, K., Sakai, H., and Nishida, M. (2011). Multiple Invasions into Freshwater by Pufferfishes (Teleostei: Tetraodontidae): A Mitogenomic Perspective. *PLoS ONE* 6, e17410. doi:10.1371/journal.pone.0017410
- Yotsu-Yamashita, M., Nishimori, K., Nitana, Y., Isemura, M., Sugimoto, A., and Yasumoto, T. (2000). Binding Properties of 3H-PbTx-3 and 3H-Saxitoxin to Brain Membranes and to Skeletal Muscle Membranes of Puffer Fish Fugu Pardalis and the Primary Structure of a Voltage-Gated Na<sup>+</sup> Channel  $\alpha$ -Subunit (fMNa1) from Skeletal Muscle of *F. pardalis*. *Biochem. Biophysical Res. Commun.* 267 (1), 403–412. Jan 7. doi:10.1006/bbrc.1999.1974
- Yu, F. H., Yarov-Yarovoy, V., Gutman, G. A., and Catterall, W. A. (2005). Overview of Molecular Relationships in the Voltage-Gated Ion Channel Superfamily. *Pharmacol. Rev.* 57, 387–395. doi:10.1124/pr.57.4.13
- Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., et al. (2018). Comparative Genome Analysis of 52 Fish Species Suggests Differential Associations of Repetitive Elements with Their Living Aquatic Environments. *BMC Genomics* 19. doi:10.1186/s12864-018-4516-1
- Zenetos, A., Gofas, S., Morri, C., Rosso, A., Violanti, D., Garcia Raso, J. E., et al. (2012). Alien Species in the Mediterranean Sea by 2012. A Contribution to the Application of European Union's Marine Strategy Framework Directive (MSFD). Part 2. Introduction Trends and Pathways. *Medit. Mar. Sci.* 13, 328. doi:10.12681/mms.327
- Zafeiropoulos, H., Gioti, A., Ninidakis, S., PotirakisParagkamian, A., Paragkamian, S., Angelova, N., et al. (2021). 0s and 1s in marine Molecular Research: a Regional HPC Perspective. *GigaScience* 10 (Issue 8), giab053, 2021. August. doi:10.1093/gigascience/giab053
- Zakon, H. H., Jost, M. C., and Lu, Y. (2011). Expansion of Voltage-dependent Na<sup>+</sup> Channel Gene Family in Early Tetrapods Coincided with the Emergence of Terrestriality and Increased Brain Complexity. *Mol. Biol. Evol.* 28, 1415–1424. doi:10.1093/molbev/msq325
- Zhou, Y., Xiao, S., Lin, G., Chen, D., Cen, W., Xue, T., et al. (2019a). Chromosome Genome Assembly and Annotation of the Yellowbelly Pufferfish with PacBio and Hi-C Sequencing Data. *Sci. Data* 6, 267. doi:10.1038/s41597-019-0279-z
- Zhou, Z., Liu, B., Chen, B., Shi, Y., Pu, F., Bai, H., et al. (2019b). The Sequence and De Novo Assembly of *Takifugu bimaculatus* Genome Using PacBio and Hi-C Technologies. *Sci. Data* 6, 187. doi:10.1038/s41597-019-0195-2
- Zou, S. (2020). Genetic Patterns for Adaptive Evolution of TTX in *Nassarius*. *Authorea* August 12, 2020. doi:10.22541/au.159724546.6127140

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Danis, Papadogiannis, Tsakogiannis, Kristoffersen, Golani, Tsaparis, Steriotti, Kasapidis, Kotoulas, Magoulas, Tsigenopoulos and Manousaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.