

Genome analysis of *Lagocephalus sceleratus*: unraveling the genomic landscape of a successful invader

Theodoros Danis¹, Vasilis Papadogiannis¹, Alexandros Tsakogiannis¹, Jon Kristoffersen¹, Daniel Golani², Dimitris Tsaparis¹, Aspasia Sterioti¹, Panagiotis Kasapidis¹, Georgios Kotoulas¹, Antonios Magoulas¹, Costas Tsigenopoulos¹, and Tereza Manousaki³

¹Hellenic Centre for Marine Research

²The Hebrew University of Jerusalem

³Hellenic Center for Marine Research

April 05, 2024

Abstract

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.

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 2010). 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 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 two 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 (Akyol & Ünal, 2017; Kalogirou, 2013).

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 (Lee and Klasing, 2004; Kolar and Logde, 2001) 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., 2019). 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 & METHODS

Sample collection & 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 (20G) and following ONTs 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 & 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 unsheared DNA was used for each library. Two of the prepared libraries were divided into two aliquots. Each library was run for approximately 24 hours 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 hours. 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 two-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 & 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 2x150 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® reagent (Invitrogen) assisting the homogenization procedure using needle and syringe. Total RNA was extracted from the TRIzol® homogenate according to the manufacturer's instructions. In the case of spleen, due 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® RNA. The

quantity of the isolated RNA was measured spectrophotometrically with NanoDrop® 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™ 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® 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 et al., 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 Kmergenie 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 QUAST 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 v9, with default parameters.

The produced assembly was polished with two rounds of Racon v1.4.3 (Vaser et al. 2017), using the pre-possessed 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 & 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 (Katoh and Standley 2013), CD-HIT v4.8.1 (Li & 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 & 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 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, S.F 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>) 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 3) (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 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). The resulted alignments were filtered with Gblocks v0.91b (Castrecana 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. 2019) 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).

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 [?] 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 ~91% of the genome (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) (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) (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 3), as well as *Takifugu pardalis* (Genbank: BAA90398.1), *Latimeria chalumnae* (NCBI Reference: XP_006003324.1), *Callorhinchus 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., 2009).

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 ~360 Mb and 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). 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 bp and 249,62 bp, respectively). A total of 20,578 genes were successfully annotated, accounting for 97% of the predicted gene set (Table 4). The BLAST top hits indicated that *L. sceleratus* genes exhibited among others, 6,704 sequence similarities to *T. rubripes*, and 4,057 to *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 3) 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 evolution analysis revealed multiple predicted rapidly expanded and contracted gene families of *L. sceleratus* (Figure 3; Supplementary Table 4; Supplementary Table 5), 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, 4 gene families were identified as *L. sceleratus*-specific, two as rapidly expanding (Supplementary Table 1) and two gene families as rapidly contracting. Finally, 43 *T. bimaculatus* and 45 *T. rubripes* families were found to be rapidly contracted (Figure S1-S2). The gene family analysis did not reveal a core set of rapidly evolving families at the base of all five puffer species (Figure 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 2. Comparisons of *L. sceleratus* against *T. nigroviridis* (Figure 4), *T. rubripes*, *T. flavidus* and *T. bimaculatus* (Figure 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* (Figure 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* (Figure S14-S15). *L. sceleratus*' contigs that are unaligned to *T. nigroviridis* chromosomes in Fig.6, were aligned to unplaced regions of the *T. nigroviridis* assembly (Fig. 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 Fig.7.

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 comparable in size with that of other puffers, such as *Fugu rubripes* (~365 Mb; Aparicio et al. 2002), *T. flavidus* (~377 Mb; Zhou et al. 2019a), *T. bimaculatus* (~393.15 Mb; Zhou et al. 2019b), *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 et al. 2019b) and *T. flavidus* (4.4 Mb; Zhou et al. 2019a). 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) (Jaillon 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 (Yuan et al. 2018, Stapley et al., 2015). 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 (Hughes et al., 2020, Hughes et al., 2018, Meynard et al., 2012, Yamanoue et al., 2009). 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 (Figures 5) (Figure 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 may hypothesize that the rapidly expanded innate immune system gene families identified play a role in the ability of *L. sceleratus* spread rapidly throughout the Eastern Mediterranean (Kalogirou 2011).

TTX resistance of *SCNA4* sodium channel

Sodium channels (NaV) are formed by an α subunit consisting of four domains (I-IV) and an optional β subunit that may alter its activity (Yu et al., 2005). Furthermore, the third whole-genome duplication of teleosts has expanded the α 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 Fig.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 *Haplochromis lunulata* (Geffeney et al., 2019) and *Haplochromis 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 *Haplochromis lunulata* (Geffeney et al., 2019) and *Haplochromis 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 Fig.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.

CONCLUSIONS

Invasive fishes, such as *L. sceleratus*, often thrive in novel environments. Our analysis provides the first high-quality genome assembly 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.

ACKNOWLEDGMENTS

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

REFERENCES

- Akbora HD, Kunter İ, Erçeti N T, Elagöz AM, Çi Çek BA. Determination of tetrodotoxin (TTX) levels in various tissues of the silver cheeked puffer fish (*Lagocephalus sceleratus* (Gmelin, 1789)) in Northern Cyprus Sea (Eastern Mediterranean). *Toxicon*. 2020 Feb; 175:1-6. doi: 10.1016/j.toxicon.2019.12.002. Epub 2019 Dec 4. PMID: 31833474.
- Akyol O, Unal V, Ceyhan T and Bilecenoglu M (2005) First confirmed record of *Lagocephalus sceleratus* (Gmelin, 1789) in the Mediterranean. *Journal of Fish Biology* 66: 1183-1186
- Akyol, O., and Ünal, V. (2017). Long journey of *Lagocephalus sceleratus* (Gmelin, 1789) throughout the

- Mediterranean Sea. *Natural and Engineering Sciences* 2(3): 41-47. doi: 10.28978/nesciences.369534
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.
- Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. ming, 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. U.S.A.* 103, 374–378. doi:10.1073/pnas.0504272102
- Bakiu, P.A and Durmisshaj (2019). First record of the silver-cheeked toadfish *Lagocephalus sceleratus* (Gmelin, 1789) in Albanian waters. Pp. 237-238. In: Kousteni, V., Bakiu, R. A., Benhmida, A., Crocetta, F., Martino, V. Di, Dogrammatzi, A., et al. New Mediterranean Biodiversity Records (April 2019). *Med. Mar. Sci.* 20(1), 230–247
- Bao, Z. and 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>
- Bednarska, N. G., Wren, B. W., & Willcocks, S. J. (2017). The importance of the glycosylation of antimicrobial peptides: natural and synthetic approaches. In *Drug Discovery Today* (Vol. 22, Issue 6, pp. 919–926). Elsevier Ltd. <https://doi.org/10.1016/j.drudis.2017.02.001>
- Bian, L., Li, F., Wang, P., Zhang, S., Liu, K., Liu, X., 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
- Bickmeyer, U., Rössler, W., & Wiegand, H. (1994). Omega AGA toxin IVA blocks high-voltage-activated calcium channel currents in cultured pars intercerebralis neurosecretory cells of adult locusta migratoria. *Neuroscience Letters* , 181 , 113–116. [https://doi.org/10.1016/0304-3940\(94\)90572-X](https://doi.org/10.1016/0304-3940(94)90572-X)
- 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., Connell, L., Konoki, K. *et al.* Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. *Nature* 434, 763–767 (2005). <https://doi.org/10.1038/nature0341>
- Brooke L Whitelaw, Ira R Cooke, Julian Finn, Rute R da Fonseca, Elena A Ritschard, M T P Gilbert et al. Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion, and loss, *GigaScience* , Volume 9, Issue 11, November 2020, g1aa120, <https://doi.org/10.1093/gigascience/g1aa120>.

Bruce W. Halstead: Poisonous and Venomous Marine Animals of the World: Vertebrates

- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics* , 46 , 253–260. <https://doi.org/10.1038/ng.2890>
- 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 HH, Chew LP. Puffer fish poisoning: a family affair. *Med J Malaysia*. 2009 Jun;64(2):181-2. PMID: 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. *Frontiers in Ecology and the Environment* , 6 , 238–246. <https://doi.org/10.1890/070151>

Daly JW. The chemistry of poisons in amphibian skin, *Proc Natl Acad Sci U S A.*, 1995, vol. 92 1(pg. 9-13)

Darriba, D, Posada, D, Kozlov, MA, Stamatakis, A, Morel, B, Flouri, T, ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models, *Molecular Biology and Evolution*, Volume 37, Issue 1, January 2020, Pages 291–294, <https://doi.org/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. <https://doi.org/10.1093/bioinformatics/btl097>

Demuth, J. P., and Hahn, M. W. (2009). The life and death of gene families. *BioEssays* , 31 (1), 29–39. <https://doi.org/10.1002/bies.080085>

Deniz Magazin (Istanbul) 68, 52-54. [in Turkish]

Eilbeck K, Moore B, Holt C, Yandell M. Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics*. 2009 Feb 23, 10:67. doi: 10.1186/1471-2105-10-67. PMID: 19236712; PMCID: PMC2653490.

Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9, 18 (2008). <https://doi.org/10.1186/1471-2105-9-18>

Emms, D.M., Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16, 157 (2015). <https://doi.org/10.1186/s13059-015-0721->

Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002 Apr 1;30(7):1575-84. doi: 10.1093/nar/30.7.1575. PMID: 11917018; PMCID: PMC101833.

Feldman CR, Brodie ED Jr, Brodie ED 3rd, Pfrender ME. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc Natl Acad Sci U S A*. 2012 Mar 20;109(12):4556-61. doi: 10.1073/pnas.1113468109.

Filiz, H. and Er, M. (2004). Akdenizin yeni misafiri (New guests in the Mediterranean Sea).

Flajnik, M. F. (2018). A cold-blooded view of adaptive immunity. In *Nature Reviews Immunology* (Vol. 18, Issue 7, pp. 438–453). Nature Publishing Group. <https://doi.org/10.1038/s41577-018-0003-9>

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C. and Smit, A. F. (2020). Repeat-Modeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* , 117 , 9451–9457. <https://doi.org/10.1073/pnas.1921046117>

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms, *Annu Rev Genomics Hum Genet.*, 2009, vol. 10 (pg. 483-511)

Gao, Y., Gao, Q., Zhang, H., Wang, L., Zhang, F., Yang, C., & Song, L. (2014). Draft Sequencing and Analysis of the Genome of Pufferfish *Takifugu flavidus*. *DNA Research*, 21(6), 627–637. <https://doi.org/10.1093/dnares/dsu025>

Ge, H., Lin, K., Shen, M., Wu, S., Wang, Y., Zhang, Z., et al. (2019). De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epimetheus akaara*) using nanopore sequencing and Hi-C. *Mol. Ecol. Resour.* 19, 1461–1469. doi:10.1111/1755- 0998.13064

- Geffeney SL, Williams BL, Rosenthal JJC, et al. Convergent and parallel evolution in a voltage-gated sodium channel underlies TTX-resistance in the greater blue-ringed octopus: *Haplochromis lunulata*. *Toxicon*. 2019; 170:77–84.
- Germain, R. N. (1994). MHC-dependent antigen processing and peptide presentation: Providing ligands for T lymphocyte activation. In *Cell* (Vol. 76, Issue 2, pp. 287–299). Elsevier. [https://doi.org/10.1016/0092-8674\(94\)90336-0](https://doi.org/10.1016/0092-8674(94)90336-0)
- Golani, D and Appelbaum-Golani, B. (2010). *FISH INVASIONS of the MEDITERRANEAN SEA* : Change and Renewal, PENSOFT Publishers
- Guidelines for the treatment of animals in behavioral research and teaching. (1977). *Anim. Behav.* 53, 229–234
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi:10.1093/bioinformatics/btt086
- Hanifin CT. The chemical and evolutionary ecology of tetrodotoxin (TTX) toxicity in terrestrial vertebrates, *Mar Drugs*, 2010, vol. 8 3(pg. 577-593)
- Ho B, Yeo DS, Ding JL. A tetrodotoxin neutralizing system in the haemolymph of the horseshoe crab, *Carcinoscorpius rotundicauda*. *Toxicon*. 1994 Jul;32(7):755-62. 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 . <https://doi.org/10.1186/1471-2105-12-491>
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* , 496 , 498–503. <https://doi.org/10.1038/nature12111>
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C. and Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution* , 34 , 2115–2122. <https://doi.org/10.1093/molbev/msx148>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., et al. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* , 47 , D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Hughes, L. C., Orti, G., Huang, Y., Sun, Y., Baldwin, C. C., Thompson, A. W., Arcila, D., et al. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences of the United States of America* , 115 , 6249–6254. <https://doi.org/10.1073/pnas.1719358115>
- Hughes, L. C., Orti, G., Saad, H., Li, C., White, W. T., Baldwin, C. C., Crandall, K. A., et al. (2021). Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Molecular Ecology Resources* , 21 , 816–833. <https://doi.org/10.1111/1755-0998.13287>
- Jatllon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Maucell, 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., & Cui, J. (2018). The Function of Fucosylation in Progression of Lung Cancer. *Frontiers in oncology*, 8, 565. <https://doi.org/10.3389/fonc.2018.00565>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* , 30 , 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>

- Kalogirou, S. (2013). Ecological characteristics of the invasive pufferfish *Lagocephalus sceleratus* (Gmelin, 1789) in Rhodes, Eastern Mediterranean Sea. A case study. *Med. Mar. Sci.* 14, 251–260. doi:10.12681/mms.364
- Kalogirou, S. (2013). Ecological characteristics of the invasive pufferfish *Lagocephalus sceleratus* (Gmelin, 1789) in the eastern Mediterranean Sea – a case study from Rhodes. *Mediterranean Marine Science*, 14(2), 251-260. doi: <https://doi.org/10.12681/mms.364>
- Kaneko Y, Matsumoto G, Hanyu Y. TTX resistivity of Na⁺ channel in newt retinal neuron. *Biochem Biophys Res Commun.* 1997 Nov 26;240(3):651-6. doi: 10.1006/bbrc.1997.7696.
- Kang, S., Kim, J., Jo, E., Lee, S. J., Jung, J., Kim, B., et al. (2020). Chromosomal-level assembly of *Takifugu obscurus* (Abe, 1949) genome using third-generation DNA sequencing and Hi-C analysis. *Mol. Ecol. Resour.* 00, 1–11. doi:10.1111/1755- 0998.13132
- Kang, S., Kim, J., Jo, E., Lee, S. J., Jung, J., Kim, B., Lee, J. H., Oh, T., Yum, S., Rhee, J., & Park, H. (2020). Chromosomal-level assembly of *Takifugu obscurus* (Abe, 1949) genome using third-generation DNA sequencing and Hi-C analysis. *Molecular Ecology Resources*, 1755-0998.13132. <https://doi.org/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). *Aquat. Invasions* 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. *Molecular Biology and Evolution* , 30 , 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kawahara, R., Miya, M., Mabuchi, K., Lavoue, S., Inoue, J. G., Satoh, T. P., Kawaguchi, A. and Nishida, M. (2008). Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): A new perspective based on whole mitogenome sequences from 75 higher teleosts. *Molecular Phylogenetics and Evolution* ,46 , 224–236. <https://doi.org/10.1016/j.ympev.2007.07.009>
- Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res.* 2011 Mar;21(3):487-93. doi: 10.1101/gr.113985.110. Epub 2011 Jan 5. PMID: 21209072; PMCID: PMC3044862.
- Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods* , 12 , 357–360. <https://doi.org/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. Gene finding in novel genomes. *BMC Bioinformatics* 5, 59 (2004). <https://doi.org/10.1186/1471-2105-5-59>
- Kosker AR, Özogul F, Durmus M, Ucar Y, Ayas D, Regenstein JM, Özogul Y. Tetrodotoxin levels in pufferfish (*Lagocephalus sceleratus*) caught in the Northeastern Mediterranean Sea. *Food Chem.* 2016 Nov 1; 210:332-7. doi: 10.1016/j.foodchem.2016.04.122. Epub 2016 Apr 27. PMID: 27211655.
- Kozlov, MA, Darriba, D, Flouri, T, Morel, B, Stamatakis, A, RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference, *Bioinformatics*, Volume 35, Issue 21, 1 November 2019, Pages 4453–4455, <https://doi.org/10.1093/bioinformatics/btz305>
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009 Sep;19(9):1639-45. doi: 10.1101/gr.092759.109. Epub 2009 Jun 18. PMID: 19541911; PMCID: PMC275213

Lee KA, Klasing KC. A role for immunology in invasion biology. *Trends Ecol Evol.* 2004 Oct;19(10):523-9. doi: 10.1016/j.tree.2004.07.012. PMID: 16701317.

Lee, K. A., and Klasing, K. C. (2004). A role for immunology in invasion biology. *Trends in Ecology and Evolution* , 19 , 523–529. <https://doi.org/10.1016/j.tree.2004.07.012>

Lewis RJ, Nielsen KJ, Craik DJ, Loughnan ML, Adams DA, Sharpe IA, Luchian T, Adams DJ, Bond T, Thomas L, Jones A, Matheson JL, Drinkwater R, Andrews PR, Alewood PF. Novel omega-conotoxins from *Conus catus* discriminate among neuronal calcium channel subtypes. *J Biol Chem.* 2000 Nov 10;275(45):35335-44. doi: 10.1074/jbc.M002252200. PMID: 10938268.

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. <https://doi.org/10.1093/bioinformatics/btl158>

Maly, P. et al. 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 (1996).

Manda Clair Jost, David M. Hillis, Ying Lu, John W. Kyle, Harry A. Fozzard, Harold H. Zakon, Toxin-Resistant Sodium Channels: Parallel Adaptive Evolution across a Complete Gene Family, *Molecular Biology and Evolution* , Volume 25, Issue 6, June 2008, Pages 1016–1024, <https://doi.org/10.1093/molbev/msn025>.

McGaugh, S. E., Gross, J. B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., Hinaux, H., Jeffery, et al. (2014). The cavefish genome reveals candidate genes for eye loss. *Nature Communications* , 5 , 1–10. <https://doi.org/10.1038/ncomms6307>

Meng, Y., & Yang, R. L. (2019). [Comparative analysis of gene family size provides insight into the adaptive evolution of vertebrates]. *Yi Chuan = Hereditas* , 41 , 158–174. <https://doi.org/10.16288/j.ycz.18-225>

Meynard CN, Mouillot D, Mouquet N, Douzery EJP (2012) A Phylogenetic Perspective on the Evolution of Mediterranean Teleost Fishes. *PLoS ONE* 7: e36443. <https://doi.org/10.1371/journal.pone.0036443>

Michael J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock, *Bioinformatics*, Volume 19, Issue 2, 22 January 2003, Pages 301–302, <https://doi.org/10.1093/bioinformatics/19.2.301>

Nagashima, Y. and Arakawa, O. (2016). Pufferfish Poisoning and Tetrodotoxin. In Gopalakrishnakone, P., Haddad, V., Jr., Tubaro, A., Kim, E., and Kem, W., R (Eds). *Marine and Freshwater Toxins* (pp. 259-284). doi:10.1007/978-94-007-6419-4

Narahashi, t., Moore, j. W., & Scott, w. R. (1964). Tetrodotoxin Blockage of Sodium Conductance Increase in Lobster Giant Axons. *The Journal of General Physiology* , 47 , 965–974. <https://doi.org/10.1085/jgp.47.5.965>

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. *Communications Biology* , 2 , 1–10. <https://doi.org/10.1038/s42003-019-0654-5>

Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., Wainwright, P. C, et al. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America* , 109 , 13698–13703. <https://doi.org/10.1073/pnas.1206625109>

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. In *Annual Review of Genetics* 39, pp. 121–152). NIH Public Access. <https://doi.org/10.1146/annurev.genet.39.073003.112240>

- Nishimura, O., Hara, Y., and Kuraku, S. (2017). GVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 33, 3635–3637. doi:10.1093/bioinformatics/btx445
- Nishio, H., Kumagaye, K. Y., Kubo, S., Chen, Y. N., Momiyama, A., Takahashi, T., Kimura, T., & Sakakibara, S. (1993). Synthesis of ω -agatoxin IVA and its related peptides. *Biochemical and Biophysical Research Communications* , 196 , 1447–1453. <https://doi.org/10.1006/bbrc.1993.2354>
- Olyarnik, S. V., Bracken, M. E. S., Byrnes, J. E., Hughes, A. R., Hultgren, K. M., and Stachowicz, J. J. (2009). *Ecological Factors Affecting Community Invasibility* (pp. 215–238). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-79236-9_12
- Ou, S. and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology* , 176 , 1410–1422. <https://doi.org/10.1104/pp.17.01310>
- Palumbi, S. R. (2001). *The Evolution Explosion: how Humans Cause Rapid Evolutionary Change* . New York, W.W. Norton & Co
- Pan, H., Yu, H., Ravi, V. et al. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *GigaSci* 5, 36 (2016). <https://doi.org/10.1186/s13742-016-0144-3>
- Pertea, M., Pertea, 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. *Nature Biotechnology* , 33 , 290–295. <https://doi.org/10.1038/nbt.3122>
- Pinto, A. (2014). Secure because math: A deep-dive on machine learning-based monitoring. *Black Hat Briefings* 25, 1–11. doi: 10.1101/gr.215087.116.Freely
- Por, F.D. 1971. One hundred years of Suez Canal – A century of Lessepsian migration: retrospect and viewpoints. *Systematic Zoology* 20: 138-159
- Gerlach, J. D. (2001). Predicting invaders. *Trends in Ecology & Evolution*, 16(10), 545.
- Price, A. L., Jones, N. C. and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* , 21 , i351–i358. <https://doi.org/10.1093/bioinformatics/bti1018>
- Renwick JH (1971) The mapping of human chromosomes. *Annual Reviews in Genetics* 5: 81–120
- Sathyajith, C., Yamanoue, Y., Yokobori, S. I., Thampy, S., & Vattiringal Jayadrathan, R. K (2019). Mitogenome analysis of dwarf pufferfish (*Carinotetraodon travancoricus*) endemic to southwest India and its implications in the phylogeny of Tetraodontidae. *Journal of Genetics* , 98 (5), 1–11. <https://doi.org/10.1007/s12041-019-1151-9>
- Satin J, Kyle JW, Chen M, Bell P, Cribbs LL, Fozzard HA, Rogart RB. A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties, *Science* , 1992, vol. 256 (pg. 1202-1205)
- Sax, D. F., Stachowicz, J. J., Brown, J. H., Bruno, J. F., Dawson, M. N., Gaines, S. D., et al. (2007). Ecological and evolutionary insights from species invasions. *Trends Ecol. Evol.* 22, 465–471. doi: 10.1016/j.j.tree.2007.06.009
- Shanmei Zou. Genetic patterns for adaptive evolution of TTX in *Nassarius*. Authorea. August 12, 2020. DOI: 10.22541/au.159724546.6127140
- Shao, F., Han, M. and Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports* , 9 . <https://doi.org/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. <https://doi.org/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.
- Simmons, M. A. (2007). Tetrodotoxin. In *xPharm: The Comprehensive Pharmacology Reference* (pp. 1–4). Elsevier Inc. <https://doi.org/10.1016/B978-008055232-3.62740-0>
- Soong TW, Venkatesh B. Adaptive evolution of tetrodotoxin resistance in animals. *Trends Genet.* 2006 Nov;22(11):621-6. doi: 10.1016/j.tig.2006.08.010. Epub 2006 Sep 7
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W435-9. doi: 10.1093/nar/gkl200. PMID: 16845043; PMCID: PMC1538822.4
- Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology* , 24 , 2 235–2252. <https://doi.org/10.1111/mec.13089>
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., Rounge, T. B., et al. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* , 477 , 207–210. <https://doi.org/10.1038/nature10342>
- Stein, C., Caccamo, M., Laird, G. et al. Conservation and divergence of gene families encoding components of innate immune response systems in zebrafish. *Genome Biol* 8, R251 (2007). <https://doi.org/10.1186/gb-2007-8-11-r251>
- Streftaris, N. and Zenetos, A. (2006). Alien marine species in the Mediterranean - the 100 “worst invasives” and their impact. *Med. Mar. Sci.* 7, 87–118. doi:10.12681/mms.180
- Tarailo-Graovac, M. and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. In *Current Protocols in Bioinformatics: Vol. Chapter 4* (Issue SUPPL. 25). Curr Protoc Bioinformatics. <https://doi.org/10.1002/0471250953.bi0410s25>
- Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., et al. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications* , 5 , 5770. <https://doi.org/10.1038/ncomms6770>
- Vaser, R. and Šikić, M. (2020). Raven: A de novo genome assembler for long reads. In *bioRxiv* (p.2020.08.07.242461). bioRxiv. <https://doi.org/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., & Soong, T. W. (2005). Genetic basis of tetrodotoxin resistance in pufferfishes. *Current Biology* , 15 , 2069–2072. <https://doi.org/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. doi: 10.1371/journal.pone.0112963
- Waterhouse, A., Procter, J., Martin, D.A. et al. Jalview: Visualization and Analysis of Molecular Sequences, Alignments, and Structures. *BMC Bioinformatics* 6, P28 (2005). <https://doi.org/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 in Bioinformatics)* ,5724 LNBI , 375–389. https://doi.org/10.1007/978-3-642-04235-6_31

- Xia, X. (2013). What is comparative Genomics? *Comparative Genomics* (Springer), pp. 1-20.
- Xiao, C. Le, Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., Luo, F. and Xie, Z. (2017). MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nature Methods* , 14 , 1072–1074. <https://doi.org/10.1038/nmeth.4432>
- Xu, J., Bian, C., Chen, K., Liu, G., Jiang, Y., Luo, Q., You, X., et al. (2017). Draft genome of the Northern snakehead, *Channa argus*. In *GigaScience* (Vol. 6, Issue 4, pp. 1–5). Oxford University Press. <https://doi.org/10.1093/gigascience/gix011>
- Yamanoue Y, Miya M, Doi H, Mabuchi K, Sakai H, Nishida M (2011) Multiple Invasions into Freshwater by Pufferfishes (Teleostei: Tetraodontidae): A Mitogenomic Perspective. *PLoS ONE* 6: e17410. <https://doi.org/10.1371/journal.pone.0017410>
- Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* **39**, 306–314 (1994). <https://doi.org/10.1007/BF00160154>
- Yotsu-Yamashita M, Nishimori K, Nitanaï Y, Isemura M, Sugimoto A, Yasumoto T. Binding properties of (3)H-PbTx-3 and (3)H-saxitoxin to brain membranes and to skeletal muscle membranes of puffer fish *Fugu pardalis* and the primary structure of a voltage-gated Na(+) channel alpha-subunit (fMNa1) from skeletal muscle of *F. pardalis*. *Biochem Biophys Res Commun.* 2000 Jan 7;267(1):403-12. doi: 10.1006/bbrc.1999.1974.
- You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., Wang, J., et al. (2014). Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nature Communications* ,5 (1), 1–8. <https://doi.org/10.1038/ncomms6594>
- Yu FH, Yarov-Yarovoy V, Gutman GA, Catterall WA. Overview of molecular relationships in the voltage-gated ion channel superfamily, *Pharmacol Rev*, 2005, vol. 57 (pg. 387-395)
- Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., & Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics* , 19 . <https://doi.org/10.1186/s12864-018-4516-1>
- Zafeiropoulos, H., Gioti, G., Ninidakis, S., Potirakis, Paragkamian, S., Angelova, N., Antoniou, A., et al. 0s and 1s in marine molecular research: a regional HPC perspective, *GigaScience* , Volume 10, Issue 8, August 2021, giab053, <https://doi.org/10.1093/gigascience/giab053>
- Zakon HH, Jost MC, Lu Y. Expansion of voltage-dependent Na⁺ channel gene family in early tetrapods coincided with the emergence of terrestriality and increased brain complexity, *Mol Biol Evol.*, 2011, vol. 28 (pg. 1415-1424)
- 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. *Mediterr. Mar. Sci.* 13, 328. doi:10.12681/mms.327
- Zhao, T., and Eric Schranz, M. (2019). Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proceedings of the National Academy of Sciences of the United States of America* , 116 , 2165–2174. <https://doi.org/10.1073/pnas.1801757116>
- 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

DATA AVAILABILITY

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 GCA_911728415.

CODE AVAILABILITY

The custom scripts that have been used in this study are available at this github repository (https://github.com/genomenerds/Lagocephalus_scleratus) and (https://github.com/Tdanis/Lagocephalus_genome_analysis)

AUTHOR CONTRIBUTIONS

T.M., C.S.T., V.P., T.D. designed research; T.D., V.P., A.T., J.C., A.S. performed research and/or analysed data; T.D., V.P., T.M. wrote the paper. All authors reviewed the manuscript.

FIGURES

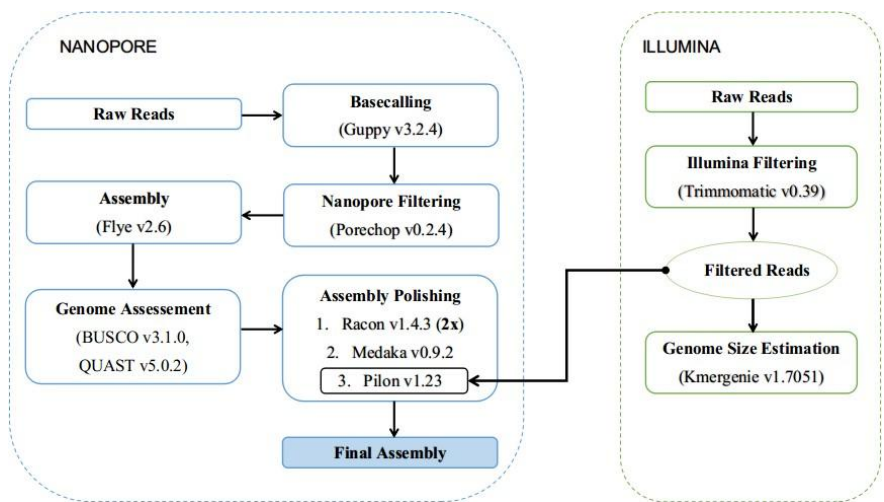


Figure 1. *Lagocephalus scleratus* genome assembly pipeline.

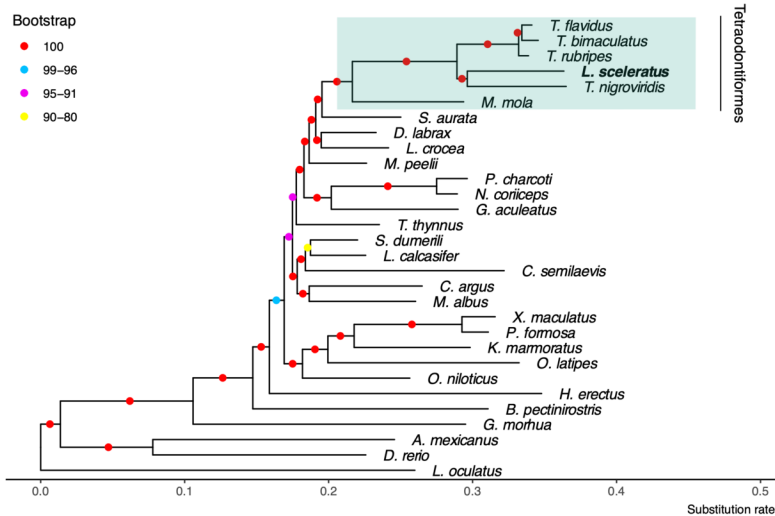


Figure 2. Maximum likelihood tree using JTT +I+ G4+F model and 100 bootstrap replicates. The spotted gar *L. oculatus* was used as outgroup.

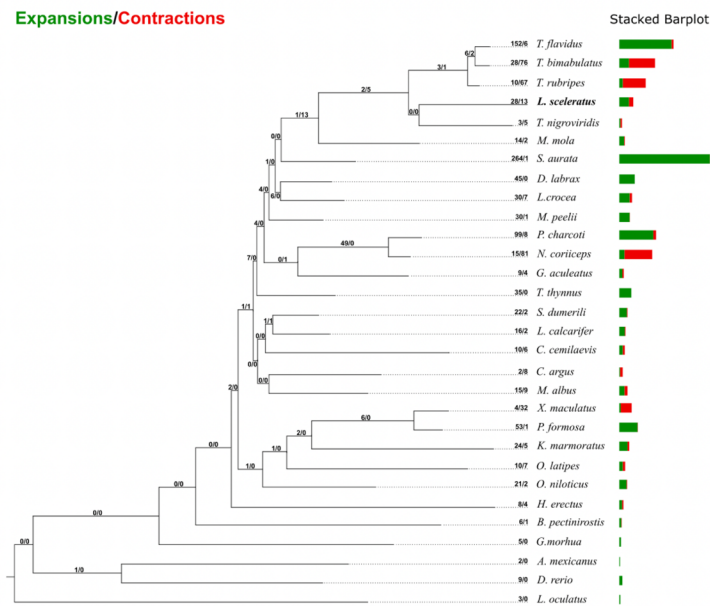


Figure 3. Gene family gain-and-loss analysis, including the number of gained gene families (green) and lost gene families (red). The stacked barplot on the right indicates the gains and losses per species.

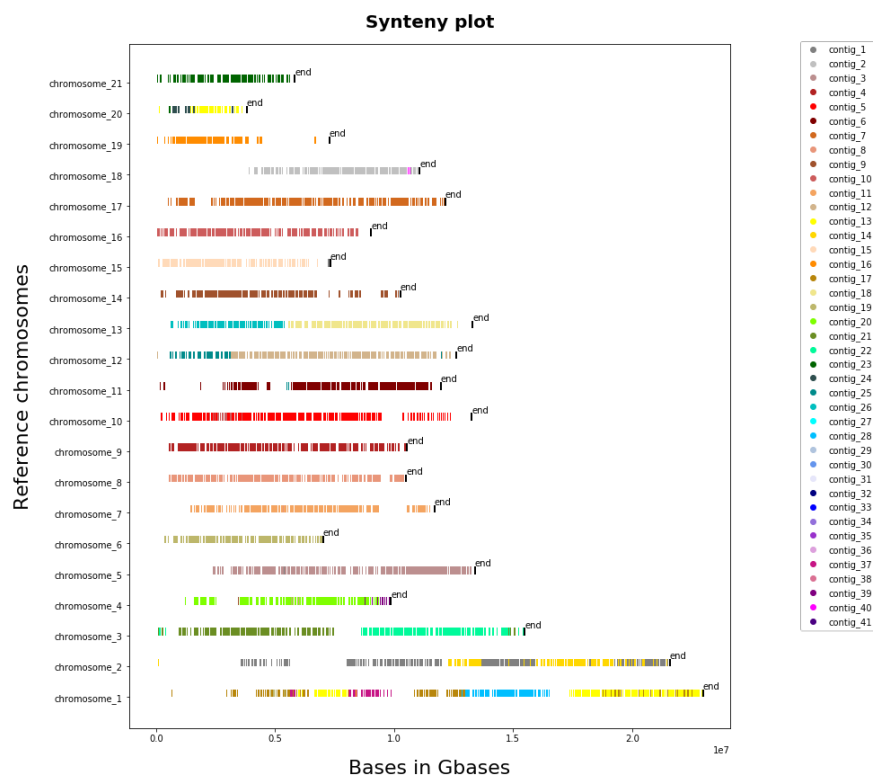


Figure 4. Synteny plot of pairwise whole genome alignment of *L. scleratus* against *T. nigroviridis*. The plot illustrates the contigs of the *L. scleratus* aligned to *T. nigroviridis* chromosomes (y-axis), grouped by a specific color which are represented on the legend right to the plot.

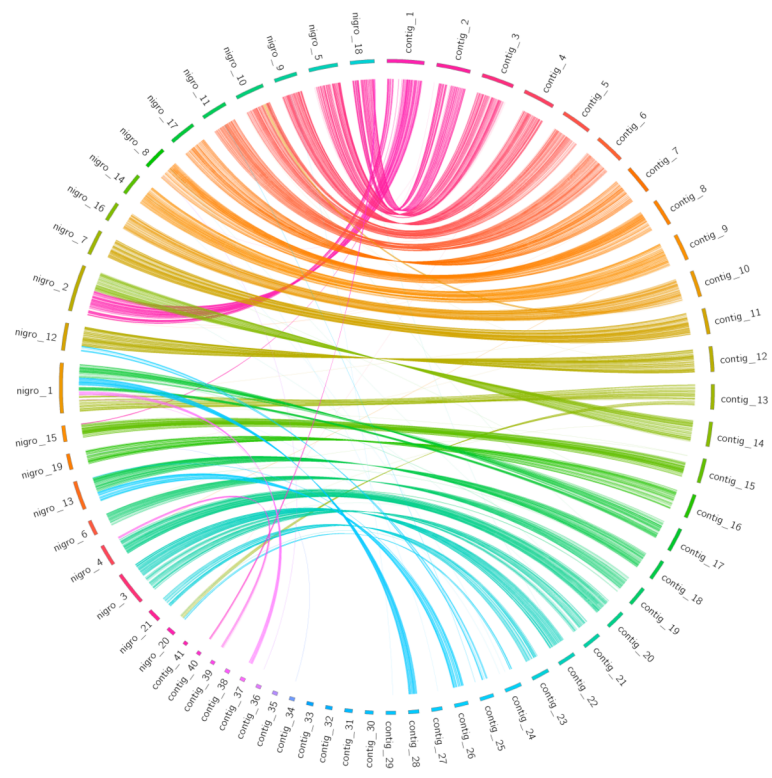


Figure 5. Circos plot illustrating syntenic relationships between *L. scleratus* contigs (right) against *T. nigroviridis* chromosomes (left), based on one-to-one orthologous genes. Ribbons link orthologous genes between the two species and colors represent the different contigs of *L. scleratus* .

		Domain1	Domain 2	Domain 3	Domain 4
SCN4A	<i>C.millii</i>	RLMTQDFWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
	<i>L.chalumnae</i>	RLMTQDYWEN	RILCGEWIET	VATFKGWMEI	ITTSAGWDGL
	<i>L.oculatus</i>	RLMTQDFWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
SCN4AA	<i>T.pardalis</i>	RLMTQDNWES	RILCGEWIET	IATFKGWMDI	ITTSAGWDTL
	<i>T.rubripes</i>	RLMTQDNWES	RILCGEWIET	IATFKGWMDI	ITTSAGWDTL
	<i>T.nigroviridis</i>	RLMTQDCWEN	RILCGEWIEN	IATFKGWTAI	ITTSAGWDQI
	<i>L.scleratus</i>	RLMTQDYWED	RILCGEWIET	IATFKGWMDI	ITTSAGWDR
	<i>M.mola</i>	RLMTQDYWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDAL
	<i>S.aurata</i>	RLMTQDYWEN	RILCGEWIET	IATFKGWMDI	ITTSAGWDGL
	<i>O.latipes</i>	RLMTQDYWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
	<i>G.morhua</i>	RLMTQDYWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDSL
	<i>D.rerio</i>	RLMTQDFWEN	RILCGEWIET	IATFKGWMDI	ITTSAGWDGL
SCN4AB	<i>T.rubripes</i>	RLMTQDFWEN	RVLCGEWIES	VATFKGWTDI	ITTSAGWDGL
	<i>T.nigroviridis</i>	RLMTQDFWEN	RVLCGEWIDT	VATFKGWMEI	ITTSAGWDGL
	<i>L.scleratus</i>	RLMTQDFWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
	<i>M.mola</i>	RLMTQDFWEN	RVLCGEWIDT	VATFKGWMDI	ITTSAGWDGL
	<i>S.aurata</i>	RLMTQDFWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
	<i>O.latipes</i>	RLMTQDFWEN	RILCGEWIET	VATFKGWMDI	ITTSAGWDGL
	<i>G.morhua</i>	RLMTQDFWEN	RVLCGEWIDT	VATFKGWMDI	ITTSAGWDGL
	<i>D.rerio</i>	RLMTQDFWEN	RVLCGEWIDT	VATFKGWMDI	ITTSAGWDGL

Figure 6. Multiple alignment plot of the four Ion Transport domain sequences previously studied by

Venkatesh et al., 2005, from *Latimeria chalumnae* (NaV1.4), *Lepisosteus oculatus* (NaV1.4), *Callorhinchus milii* (NaV1.4), *T. pardalis* (NaV1.4a), *T. rubripes* (NaV1.4a, NaV1.4b), *T. nigroviridis* (NaV1.4a, NaV1.4b), *L. scleratus* (NaV1.4a, NaV1.4b), *M. mola* (NaV1.4a, NaV1.4b), *S. aurata* (NaV1.4a, NaV1.4b), *O. latipes* (NaV1.4a, NaV1.4b), *M. morhua* (NaV1.4a, NaV1.4b) and *D. rerio* (NaV1.4a, NaV1.4b). Residues outlined in red, have been previously associated with TTX resistance in *T. rubripes* and *T. nigroviridis* by Venkatesh et al., 2005.

TABLES

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
BUSCO completeness score	
Complete	96.20 %
Single	94.40 %
Duplicated	2.20 %
Fragmented	1.40 %
Missing	2.00 %
Total number of Actinopterygii orthologs	4,492 (98%)

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 %
<i>SINEs</i> :	5,447	743,745	0.20 %
Penelope	3,627	2,572,446	0.69 %
<i>LINEs</i> :	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 %

Repetitive Elements	Number of elements	Length occupied (bp)	Percentage of sequence
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	Per cent (%)
Blast	18,805	88%
InterProScan	20,347	96%
EggNog-Mapper	17,849	84%
Predicted genes	20,578	97%
Total Genes	21,251	