Supplementary Materials for

Genome structures resolve the early diversification of teleost fishes

Elise Parey, Alexandra Louis, Jerome Montfort, Olivier Bouchez, Céline Roques, Carole Iampietro, Jerome Lluch, Adrien Castinel, Cécile Donnadieu, Thomas Desvignes, Christabel Floi Bucao, Elodie Jouanno, Ming Wen, Sahar Mejri, Ron Dirks, Hans Jansen, Christiaan Henkel, Wei-Jen Chen, Margot Zahm, Cédric Cabau, Christophe Klopp, Andrew W. Thompson, Marc Robinson-Rechavi, Ingo Braasch, Guillaume Lecointre, Julien Bobe, John H. Postlethwait, Camille Berthelot, Hugues Roest Crollius, Yann Guiguen.

Correspondence to: camille.berthelot@pasteur.fr, hrc@bio.ens.psl.eu, yann.guiguen@inrae.fr

This PDF file includes:

Materials and Methods

Fish sampling

Information on the Elopomorpha specimens that were collected for genome sequencing is provided in table S2.

High Molecular Weight (HMW) genomic DNA (gDNA) extraction

For all Elopomorpha species, HMW gDNA for long-read sequencing was extracted from blood (0.5 ml) sampled with a 10% EDTA (Ethylenediaminetetraacetic acid) coated syringe and directly stored in 25 ml of a TNES-Urea lysis buffer (TNES-Urea: 4 M urea; 10 mM Tris-HCl, pH 7.5; 125 mM NaCl; 10 mM EDTA; 1% SDS). HMW gDNA was then extracted from each blood sample in TNES-urea using a modified phenol/chloroform protocol as previously described (*28*).

Genome-wide Chromatin Conformation Capture (Hi-C) sample collection

For all Elopomorpha species except *Aldrovandia affinis*, 1.5 ml of blood was sampled with a 10% EDTA (Ethylenediaminetetraacetic acid) coated syringe and slowly cryopreserved with 15 % dimethyl sulfoxide (DMSO) for the construction chromosome contact map (Hi-C) libraries.

RNA-sequencing (RNA-Seq) sample collection and RNA extraction

For all Elopomorpha species, except *A. anguilla* in which previous RNA-Seq were publicly available (*29*), some of the following organs or tissues: kidney, brain, gills, gonads, liver, fins and skin, were sampled on freshly euthanized individuals and stored in RNAlater solution or directly snap-frozen in liquid nitrogen. All samples were kept at -80°c until RNA extraction. Total RNA was extracted using a modified phenol / chloroform extraction method with Tri Reagent (Euromedex). Frozen tissues or organs were grinded in 1 ml of Tri Reagent with a Precellys grinder (Precellys Evolution Bertin) in a 2 ml tube containing 28 mm ceramic balls (Ozyme) with the following grinder parameters: 5,800 rpm, 2x30’’cycles with a 30’’pause. Chloroform (0.2 ml) was then added and mixed with each

grinded sample in Tri Reagent solution and centrifuged (12,000 g, 30 min, 4°C) in order to separate the organic and aqueous phases. The aqueous phase in the supernatant was then recovered and RNAs were precipitated by adding 0.5 ml of cold isopropanol (-20°C). After 2 hours at -20°C solutions were centrifuged (12,000 g, 45 min, 4°C) and the supernatant removed. The RNA pellets were washed twice with 1 ml of 75% ethanol at -20°C followed by centrifugation (12,000 g,15 min, 4°C). The final RNA pellets were then dried at room temperature for 5-10 min and resuspended in 10 to 50 µl nuclease-free water depending on the size of the pellet. RNAs were then stored at -80°C until library construction.

Genome sequencing, assembly and annotation

Long-read sequencing was carried out in Elopomorpha species using the Oxford Nanopore **(**ONT) sequencing technology with the single exception of the Kaup's arrowtooth eel, *Synaphobranchus kaupii* that was sequenced using the PacBio Hifi sequencing technology. HMW gDNA quality and purity was assessed using spectrophotometry, fluorometry and capillary electrophoresis. Additional purification steps were performed using AMPure XP beads (Beckman Coulter).

Oxford Nanopore (ONT) long-read sequencing

All library preparation and sequencing were performed using the Oxford Nanopore Ligation Sequencing Kit SQK-LSK108 (*A. anguilla*), and SQK-LSK109 according to the manufacturer's instructions. For each library, 10 µg of DNA was purified then sheared between 20kb to 35kb using the megaruptor1 system (Diagenode). For *A. anguilla*, DNA was sheared with a G-tube (Covaris), or not. For some samples a size selection step using the Short Read Eliminator XS Kit (Circulomics) was performed. Then a one-step DNA-damage repair + END-repair-dA-tail procedure was performed on 2 µg of DNA. Adapters were ligated to DNAs in the library. Libraries were loaded onto R9.4 (*A. anguilla*), and R9.4.1 flowcells and sequenced on either MinION, GridION or PromethION instruments for 24h to 72h (see table S3 for details of long read sequencing outputs).

PacBio Hifi long-read sequencing

For *S. kaupii*, libraries preparation and sequencing were performed according to the manufacturer's instructions "Procedure & Checklist Preparing HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0". Fifteen µg of gDNA was purified then sheared at 20 kb using the megaruptor3 system (Diagenode). Then a single strand overhangs removal and a DNA and END damage repair step were performed on 10 µg of the sample. Blunt hairpin adapters were ligated to the library. The library was treated with an exonuclease cocktail to digest unligated DNA fragments. A size selection step using a 11 kb cutoff was performed on the BluePippin Size Selection system (Sage Science) with "0.75% DF Marker S1 High Pass 15-20 kb" protocol. Using Binding kit 2.0 kit and sequencing kit 2.0, the primer V2 annealed and polymerase 2.0 bounded library was sequenced by diffusion loading onto 2 SMRTcells on Sequel2 instrument at 50pM and 60 pM with a 2 hours pre-extension and a 30 hours movie (see table S3 for details of long read sequencing outputs).

10x Genomics© Linked-Reads

HMW gDNA quality and purity was assessed using spectrophotometry, fluorometry and capillary electrophoresis. 10X Chromium Library was prepared according to 10X Genomics protocols from 1.25 ng (*S. kaupii*), 0.625 ng (*Conger conger*, *A. affinis*, *Albula goreensis*, *Megalops atlanticus*) and 0.98 ng (*Gymnothorax javanicus*) of gDNA using the Genome Reagent Kits v2. Library quality was assessed using capillary electrophoresis and quantified by QPCR using the Kapa Library Quantification Kit. All species were sequenced on an Illumina HiSeq3000 using a 2x150 pb paired-end read length except *S. kaupii* that was sequenced on an Illumina HiSeq4000 using a 2x150 pb paired-end read length.

Genome-wide Chromatin Conformation Capture (Hi-C) libraries

For *G. javanicus*, and *A. Anguilla*, Hi-C libraries were prepared according to a protocol adapted from Foissac and collaborators (*30*). Briefly, the cryopreserved blood sample was spun down, and the cell pellet was resuspended and fixed in 1% formaldehyde. Five million cells were processed through overnight digestion with *HindIII* (NEB), and DNA ends were labeled with Biotin-14-DCTP (Invitrogen) using the klenow (NEB) and religated. DNA (1.4 g) was sheared to an average size of 550 bp (Covaris). Biotinylated DNA fragments were pulled down using M280 Streptavidin

Dynabeads (Invitrogen) and ligated to PE adaptors (Illumina). Hi-C libraries were then amplified using PE primers (Illumina) with 10 PCR amplification cycles and each library was sequenced on either on an Illumina NovaSeq6000 platform for *G. javanicus* or an Illumina HiSeq platform for *A. anguilla* (see table S4 for details of short read sequencing outputs). For *C. conger*, *M. atlanticus*, *A. goreensis*, and *S. kaupii*, Hi-C libraries were processed using the Arima-HiC Kit (San Diego, CA) according to the manufacturer's instructions and the resulting Hi-C libraries were sequenced on an Illumina NovaSeq6000 platform. No Hi-C library has been constructed in *A. affinis*, due to the lack of a cryopreserved blood sample in this deep-sea species.

RNA-sequencing (RNA-Seq) for genome annotation

RNA-Seq libraries have been prepared according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. Briefly, mRNA was selected using poly-T beads, fragmented to generate double stranded cDNA and adaptors were ligated to be sequenced. 11 cycles of PCR were applied to amplify libraries. Library quality was assessed using a Fragment Analyser and libraries were quantified by QPCR using the Kapa Library Quantification Kit. RNA-Seq libraries were sequenced on an Illumina NovaSeq 6000 using a paired-end read length of 2x150 pb with the Illumina NovaSeq 6000 sequencing kits.

Genome assembly

All genomes except those of *A. affinis* and *S. kaupii*, were assembled using the same assembly procedure described hereafter. Nanopore reads were assembled with wtdbg2 (*31*) version 2.3 using standard parameters. Contigs were polished with one round of racon (v.1.3.1) (*32*), using long reads aligned with minimap2 (v.2.7) (*33*) and one round of pilon (v.1.22) (*34*), using 10X Illumina reads. These reads were aligned with bwa mem (v.0.7.12-r1039) (*35*) with standard parameters and the alignments were compressed, sorted and indexed with samtools (*36*) view, sort and index v.1.3.1, using standard parameters. The polished contigs were then scaffolded using Hi-C and 10X as sources of linking information. 10X reads were aligned using Long Ranger v2.1.1 (10x Genomics) (*37*). Hi-C reads were aligned to the draft genome using Juicer (*38*) with default parameters. A candidate assembly was then generated with the 3D de novo assembly (3D-DNA) pipeline (*39*) with the -r 0

and --polisher-input-size 100000 parameters. Finally, the candidate assembly was manually reviewed using the Juicebox assembly tools (*40*). Because no Hi-C data was available for *A. affinis*, the contigs were scaffolded using 10X Illumina reads with arcs (v.1.1.1) (*41*). As *S. kaupii* was sequenced with HiFi, reads on the PacBio Sequel II were assembled with hifiasm (v.0.9) (*42*) and the contigs purged with purge_dups (v1964aaa). The contigs were not polished before scaffolding using the procedure described previously.

Genome annotation

All genomes have been annotated using the same procedure. The first annotation step was to identify repetitive content using RepeatMasker v4.0.7 (*43*), Dust v1.0.0 (*44*) and TRF v4.09 (*45*). From each genome, a species-specific de novo repeat library was built with RepeatModeler v1.0.11 and repeated regions were located using RepeatMasker with the *de novo* and *Danio rerio* libraries. Bedtools v2.26.0 (*46*) was used to aggregate repeated regions identified with the three tools and to soft-mask the genome. The MAKER3 genome annotation pipeline (*47*) v3.01.02-beta combined annotations and evidence from three approaches: similarity with fish proteins, assembled transcripts (see below), and de novo gene predictions. Protein sequences from 11 fish species (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Xiphophorus maculatus*) found in Ensembl were aligned to the masked genome using Exonerate v2.4 (*48*) with the alignment model protein2genome that allows translated alignments with modeling of introns. RNA-Seq were mapped to the genome using STAR v2.5.1b (*49*) with outWigType and outWigStrand options to output signal wiggle files. Cufflinks v2.2.1 (*50*) was used to assemble the transcripts which were used as RNA-seq evidence. Braker v2.0.4 (*51*) provided de novo gene models from wiggle files provided by STAR as hint files for GeneMark (*51*) and Augustus (*52*) training. The best supported transcript for each gene was chosen using the quality metric called Annotation Edit Distance (AED) (*53*). Finally, genome annotation gene completeness was assessed by BUSCO (*54*) based on orthologs derived from the Actinopterygii lineage.

Genome-wide phylogenetic gene trees

In order to extract single-copy orthologous genes to use as phylogenetic markers, we reconstructed a genome-wide set of gene trees including genes of all 25 selected genomes (see table S4 for species and genome assembly references). To do so, we used a pipeline similar to the one employed by the Ensembl Compara database (*55*). We started by performing an all-against-all BLASTP+ on the set of proteins derived from the longest transcripts of each genome (*56*), using the following parameters '-seg no -max_hsps 1 -use_sw_tback -evalue 1e-5'. From the blast results, we defined gene families with the clustering algorithm hcluster_sg, using parameters '-m 750 -w 0 -s 0.34 -O'. For each gene family, we next built a protein multiple alignment using T-Coffee (*57*), with the command 't-coffee -type=PROTEIN -method mafftgins_msa, muscle_msa,kalign_msa'. We reconstructed gene trees with TreeBeST, using default parameters (*55*). Since TreeBeST requires a species tree to guide gene tree inference and perform reconciliation, we used the previously reported consensus molecular phylogeny (i.e., Osteoglossocephala scenario, Figure 2B), in an effort to not bias the inferences towards the Eloposteoglossocephala scenario. We however note that while the chosen species phylogeny impacts TreeBeST gene tree topologies, it does not alter the orthology inferences and orthologous gene sets that we leverage in all downstream analyses. Finally, since whole genome duplications (WGD) are known as a prominent source of errors in gene trees, we ran SCORPiOs (version 2.0.0) to account for the teleost WGD and correct gene trees accordingly (*58*). We ran SCORPiOs for five rounds of iterative gene tree correction, using the bowfin and spotted gar genomes as outgroups to the WGD and the same species phylogeny as for the TreeBeST gene tree inferences. Similarly, SCORPiOs only uses the species phylogeny to guide gene tree topologies, thus it does not impact the definition of orthologous genes sets.

Selection of orthologous marker genes

We exploit the high-confidence orthology relationships inferred in the SCORPiOs gene trees to derive sets of orthologous marker genes. For molecular phylogenies reconstruction, we extracted the set of all 955 strictly one-to-one orthologous gene families. The average size of aligned sequences for these 955 families is of 2,438 nucleotides, with the smallest alignment comprising 321 nucleotides and a total combined alignment size of 2,328,657 nucleotides, which is significantly more than all previous studies (Fig. S1). For the microsynteny-based phylogeny, we completed the set of 955 high-confidence one-to-one markers with 2,086 additional genes, thus obtaining a total of 3,041 markers.

The additional 2,086 markers comprise genes that exist in exactly one copy in non-duplicated outgroup genomes (chicken, western clawed frog, spotted gar and bowfin) and in either one or two copies in teleosts as a result of the teleost WGD. These additional markers were included in order to cover a larger proportion of the genomes and leverage the post-WGD rediploidisation history (shared gene copy losses) for phylogeny reconstruction (see Microsynteny phylogeny below).

Molecular phylogeny from gene tree collections

To reconstruct molecular phylogenies for teleost genomes, we took advantage of ASTRAL-III (*23*), a summary method to infer species trees from collections of gene trees. ASTRAL accounts for discordance among gene trees and species trees by explicitly modeling incomplete lineage sorting under the multi-species coalescent model. Here, we first built individual gene trees for each of the 955 gene families using RAxML 8.2.12 (*59*), under the GTRGAMMA model of sequence evolution (codon alignments) or PROTGAMMAJTT (protein alignments). The Maximum Likelihood (ML) searches were conducted from 10 starting trees (-N 10). Codon alignments were analyzed as two partitions: one partition for first and second codon positions and one partition for the third codon position. Species phylogenies were then computed from the set of 955 estimated ML trees using ASTRAL-III version 5.7.3 (*23*), with default parameters.

Molecular phylogeny from concatenated sequences

We used RAxML 8.2.12 (*59*) to infer phylogenies from the nucleic and protein concatenated sequences of the 955 orthologous marker genes. We conducted Maximum Likelihood searches from 10 starting trees and generated 100 bootstrap replicates. Phylogenies were inferred under the GTRGAMMA model of sequence evolution for codon alignments, partitioned by each codon position and the PROTGAMMAJTT model for protein alignments.

Gene genealogy interrogation

We conducted a gene genealogy interrogation analysis (*60*) to evaluate the support given by each of the 955 gene alignments to the three phylogenetic hypotheses (Fig. 3B). We first performed three

additional ML tree reconstructions for each of the 955 alignments, where we constrained the tree topologies to follow each of the three branching hypotheses presented in Fig. 1. To do so, we used constrained RAxML searches with the same models and parameters as detailed previously (see Molecular phylogeny from gene tree collections). We next used CONSEL (*61*) to rank gene trees according to their likelihood and perform likelihood AU-tests (*62*). Note that we included the unconstrained ML tree in the set of trees considered by the AU-tests, as it mitigates cases were none of the 3 hypotheses are supported by the data (due to undue paralogs inclusion, strong incomplete lineage sorting, introgression or other evolutionary events). One gene tree was considered significantly better than the others when both alternative topologies were rejected by AU-tests at α=0.05.

Microsynteny phylogeny

We built a microsynteny-based phylogeny using an approach previously successful in resolving the bowfin position in the fish tree of life (*21*). Here, we reduced the 25 studied genomes to the set of 3,041 ordered marker genes (see Selection of orthologous marker genes). The average genomic distance between these markers is 91 kb, with an end-to-end coverage encompassing 86% of selected genomes (Fig. S5). We leveraged adjacency conservation between these markers to reconstruct the teleost phylogeny. In this setting, gene adjacencies are broken along teleosts evolution either by small-scale rearrangements or loss of duplicated gene copies. We first estimated pairwise evolutionary distances between genomes using a normalized breakpoint distance, simply computed as: (1-SHARED_ADJ) / min (ADJ1, ADJ2), where SHARED_ADJ is the number of shared gene adjacencies, and ADJ1 and ADJ2 the total number of adjacencies in genome 1 and genome 2, respectively. Based on these computed distances, we constructed a rearrangement-based neighbor-joining (NJ) tree. Bootstrap supports were obtained from 100 random re-sampling of the columns of the complete adjacency absence/presence matrix, with replacement.

Macrosynteny phylogeny

We used PhyChro (*22*) to reconstruct the teleost phylogeny from the evolutionary signal contained in synteny block breakpoints. We first computed all pairwise synteny block breakpoints for the 25

studied genomes using SynChro (*22*), with the parameter --delta 3. Across comparisons, we recovered an average of 1,642 synteny blocks, an average block size of 191 kb and an average genomic coverage of 63% (Fig. S5). As such, the macrosyntenic approach employed here indeed considers larger genomic scale rearrangement events than the gene adjacency approach (Fig. S5). We then directly invoked PhyChro on the set of computed synteny comparisons, with default parameters and --delta 3, to reconstruct the teleost phylogeny. Note that synteny blocks inferred by SynChro do not make use of our pre-defined orthologous marker genes, but infers orthologous genes for each pairwise comparison, using both sequence-based and synteny-based orthology criteria. From the complete set of synteny breakpoints revealed by SynChro, PhyChro assembles a collection of phylogenetically informative breakpoints, i.e., breakpoints that allow grouping genomes into two disjoint sets. PhyChro then computes a distance between all genome pairs, based on the number of times the two genomes are found into different groupings of such sets. Finally, PhyChro reconstructs the phylogeny with the fewest breakpoint inconsistencies, by iteratively grouping genomes with the smallest distance. Confidence scores range between 0 and 1 and correspond to the proportion of informative breakpoints supporting that node, reflecting the phylogenetic signal supporting internal nodes. The PhyChro strategy has been shown to accurately reconstruct branches even with a very low phylogenetic signal, such as the position of the fast-evolving Rodentia clade (PhyChro confidence score = 0.03) (*22*).

Fusions of ancestral chromosomes in modern teleost karyotypes

We identified pre-duplication ancestral chromosomes based on a previous reconstruction of the ancestral teleost karyotype (*63*). We next applied an approach similar to (*64*) to identify ‘a’ and ‘b’ ancestrally-duplicated chromosome copies across teleost genomes. Here, we leveraged paralogy and orthology relationships contained in the full set of SCORPiOs phylogenetic gene trees (see “Genome-wide phylogenetic gene trees"). Using the goldeye genome as a reference, we use (i) paralogous genes to identify ‘a’ and ‘b’ chromosome copies within the goldeye genome and (ii) orthologous genes to propagate these ‘a’ and ‘b’ annotations across species. We note that while for some genes, errors might remain in the orthology assignments, these errors are unlikely to affect whole chromosomes. Indeed, gene orthologies are inferred from independent gene trees. Finally, we used RIdeogram (*65*) to plot ancestral chromosomes on modern teleost karyotypes. We used plotting parameters aimed at

reducing the noise induced by small-scale rearrangements and fragmented genome assemblies. We show only chromosome/scaffolds with over 50 genes and at least 5% of genes assigned to ancestral chromosomes under study (1a, 1b, 2a, 2b). Similarly, an ancestral chromosome color is painted on a modern chromosome/scaffold if it contains at least 5% of these ancestral chromosome genes. The two fusion events that we present (Fig. 4, Fig. S6) are the only chromosome-scale shared rearrangement events with one fusion only shared in Elopomorpha and Osteoglossomorpha, and the second only amongst Clupeocephala.

Data and scripts availability

All input data (sets of orthologous marker genes, CDS codons alignments, gene coordinates files) and the generated reconstructed species phylogenies have been deposited in Zenodo (doi: 10.5281/zenodo.6414307), along with all scripts and environments to reproduce the analyses. In particular, the ASTRAL-III and gene genealogy interrogation analyses can be reproduced in a single command from the deposited snakemake pipeline. The microsynteny phylogeny can be reproduced by running the provided Jupyter Notebook. Finally, we provide instructions and commands to run PhyChro and reproduce the macrosynteny phylogeny.

Supplementary Text

Section 1: Rationale behind the naming of the Eloposteoglossocephala clade

Eloposteoglossocephala, the name given to the clade containing the Elopomorpha and the Osteoglossomorpha, has been made from the fusion of the prefix *elopo*- referring to the Elopomorpha and the prefix *osteoglosso*- referring to the Osteoglossomorpha. The suffix -*cephala* is proposed to rank the clade as a supercohort, as close as possible to the ranking proposed by Betancur et al. (2017) (*12*). As the tree gains an element of symmetry, a rank disappears and the "-*cephalai*" megacohort rank of Betancur et al. (2017) (*12*) is no longer necessary as in the new topology the two sister groups have the same rank:

Infraclass: Teleostei (as in (*12*))

Supercohort: Eloposteoglossocephala (this study)

Cohort: Elopomorpha (as in (*12*))

Cohort: Osteoglossomorpha (proposed as a cohort in this study)

Supercohort: Clupeocephala (supercohort as in (*12*))

Cohort: Otomorpha (cohort, as in (*12*))

Cohort: Euteleosteomorpha (as in (*12*): the rest of teleosts, i.e., Euteleostei).

Section 2: Review of the anatomical literature and search for potential Eloposteoglossocephala synapomorphies.

From anatomical and morphological characters of fossil and extant ray-finned fish taxa, Patterson and Rosen (*6*) found a sister group relationship between Osteoglossomorpha and the rest of teleosteans, calling this new group Elopocephala. Elopocephalan synapomorphies were (a) only two uroneurals extending beyond the second ural centrum and (b) intermuscular epipleural bones present

345 in the caudal region. Arratia (*7*), however, found these characters as homoplastic when adding new
346 fossils to an anatomical matrix of fossil and extant taxa, and concluded on a sister-group relationship
347 between Elopomorpha and the rest of teleosteans, calling this new group Osteoglossocephala.
348 Osteoglossocephalan synapomorphies were (a) the antorbital branch of the infraorbital sensory canal
349 not enclosed by the antorbital bone, (b) the posterior opening of the mandibular sensory canal laterally
350 placed on the angular portion of the lower jaw, (c) seven or fewer hypurals, (d) absence of urodermal
351 (at least primitively), (e) long dorsal segmented procurrent rays in the caudal fin, (f) no dorsal
352 processes in the innermost principal caudal rays of the upper lobe of the caudal fin, (g) principal
353 caudal rays with straight segmentation. However, figure 106 in (*7*) shows that all these characters are
354 homoplastic, and the Osteoglossocephala group does not seem to be better supported than Patterson
355 and Rosen's Elopocephala group. A close examination of the matrix in (*7*) shows that there are no
356 character states exclusively shared by *Elops*, †*Anaethalion*, i.e., the two sampled Elopomorpha, and
357 *Hiodon* and †*Lycoptera*, i.e., the two Osteoglossomorpha sampled. Actually, none of the anatomical
358 studies having sampled both Elopomorpha and Osteoglossomorpha for a cladistic analysis concluded
359 in favor of a sister-group relationship of the two groups. Unfortunately, these studies did not exhibit
360 their homoplastic characters onto their trees (excepted (*7*)), a practice that would allow to identify a
361 possible weak anatomical signal, i.e., characters shared by Osteoglossomorpha and Elopomorpha but
362 appearing twice in a tree where other -more consistent- characters separate the two groups. The
363 careful examination of the character matrices published to date did not allow the detection of
364 characters exclusively shared by the two groups.

365 One character though, Char. 247 of Diogo et al. (2008) (*26*) is shared by Elopomorpha and the
366 osteoglossomorphans *Mormyrus*. The shared derived state of this character is the "absence of a
367 retroarticular as an independent ossification". Fusion of the retroarticular with the angular (*6*, *66*)
368 and/or the articular (*26*) has been previously considered a synapomorphy of the Elopomorpha (Arratia
369 coded this derived character as fused angular and retroarticular, with the articular bone partially fusing
370 with the retroarticular late in ontogeny for *Elops* and †*Anaethalion* (*7*)). According to Diogo et al.
371 (2008) (*26*) this character is ambiguous in *Hiodon* because some authors have not observed
372 independent retroarticulars, which were instead fused with the angulars (*27*, *66*), while other authors
373 have observed retroarticulars independent from the angulars (*7*, *67*). This derived character is absent
374 in the two other Osteoglossomorpha of Diogo et al.'s sample, *Xenomystus* and *Pantodon*. This derived
375 character is, however, also present in siluriformes. Thus, this derived state could represent a

synapomorphy of the clade uniting Elopomorpha and Osteoglossomorpha, that would have been secondarily lost in *Xenomystus* and *Pantodon*, or alternatively could represent a homoplastic character independently acquired in Elopomorpha and *Mormyrus*. The former hypothesis may be favored because, following Nelson (1973) (*66*), Hilton (2003) (*27*) considered that several patterns of bone fusion observed in Elopomorpha are also found in Osteoglossomorpha.

The fusion of retroarticular and angular was shown to occur during ontogeny in *Hiodon* (*68*). Among other osteoglossomorpha, the angulars are fused with the articulars in some osteoglossids (*Pantodon*, *Scleropages*, *Osteoglossum*), and notopterids (*Chitala*, *Xenomystus, Papyrocranus*) (*27*). Retroarticulars, articulars, and angulars are not fused in the osteoglossids *Arapaima* and *Heterotis*, while fused in mormyrids (*Petrocephalus, Gnathonemus, Campylomormyrus*) (*27*). The basal fossils Osteoglossomorpha †*Lycoptera* and †*Ostariostoma* have unclear conditions. Because *Hiodon*, which is the most early diverging Osteoglossomorpha, its fossil sister group †*Eohiodon* (*27*), and mormyrids exhibit an angular-retroarticular fusion, we provisionally consider that the tendency of such fusion could be the primitive condition in Osteoglossomorpha, thus a possible synapomorphy of the Eloposteoglossocephala group uniting Elopomorpha and Osteoglossomorpha, even though with some homoplastic changes (i.e., reversal to ancestral unfused character state) within Osteoglossomorpha. Further studies are necessary to better grasp the extent of angular-retroarticular bone fusion across the early diverging (i.e., non acanthomorph) teleostean diversity.

References

28. Q. Pan, R. Feron, A. Yano, R. Guyomard, E. Jouanno, E. Vigouroux, M. Wen, J.-M. Busnel, J. Bobe, J.-P. Concordet, H. Parrinello, L. Journot, C. Klopp, J. Lluch, C. Roques, J. Postlethwait, M. Schartl, A. Herpin, Y. Guiguen, Identification of the master sex determining gene in Northern pike (Esox lucius) reveals restricted sex chromosome differentiation. *PLOS Genetics*. **15**, e1008013 (2019).

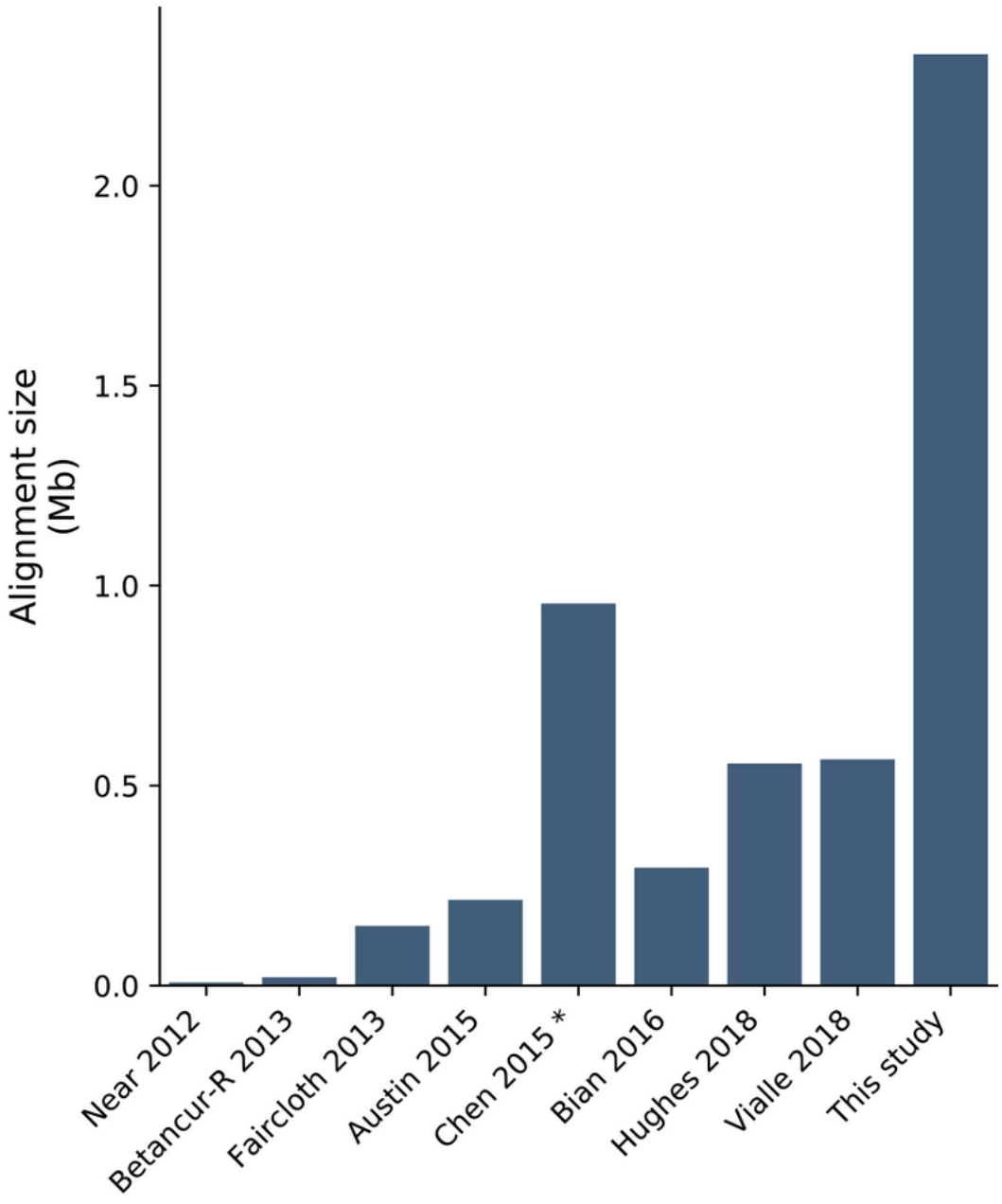
29. J. Pasquier, C. Cabau, T. Nguyen, E. Jouanno, D. Severac, I. Braasch, L. Journot, P. Pontarotti, C. Klopp, J. H. Postlethwait, Y. Guiguen, J. Bobe, Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics*. **17**, 368 (2016).

30. S. Foissac, S. Djebali, K. Munyard, N. Vialaneix, A. Rau, K. Muret, D. Esquerré, M. Zytnicki, T. Derrien, P. Bardou, F. Blanc, C. Cabau, E. Crisci, S. Dhorne-Pollet, F. Drouet, T. Faraut, I. Gonzalez, A. Goubil, S. Lacroix-Lamandé, F. Laurent, S. Marthey, M. Marti-Marimon, R. Momal-Leisenring, F. Mompart, P. Quéré, D. Robelin, M. S. Cristobal, G. Tosser-Klopp, S. Vincent-Naulleau, S. Fabre, M.-H. Pinard-Van der Laan, C. Klopp, M. Tixier-Boichard, H. Acloque, S. Lagarrigue, E. Giuffra, Multi-species annotation of transcriptome and chromatin structure in

410 domesticated animals. *BMC Biology*. **17**, 108 (2019).
411 31. J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. **17**, 155–158
412 (2020).
413 32. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from
414 long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
415 33. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–
416 3100 (2018).
417 34. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q.
418 Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: an integrated tool for comprehensive microbial
419 variant detection and genome assembly improvement. *PLoS ONE*. **9**, e112963 (2014).
420 35. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform.
421 *Bioinformatics*. **25**, 1754–1760 (2009).
422 36. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
423 Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and
424 SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
425 37. P. Marks, S. Garcia, A. M. Barrio, K. Belhocine, J. Bernate, R. Bharadwaj, K. Bjornson, C.
426 Catalanotti, J. Delaney, A. Fehr, I. T. Fiddes, B. Galvin, H. Heaton, J. Herschleb, C. Hindson, E. Holt,
427 C. B. Jabara, S. Jett, N. Keivanfar, S. Kyriazopoulou-Panagiotopoulou, M. Lek, B. Lin, A. Lowe, S.
428 Mahamdallie, S. Maheshwari, T. Makarewicz, J. Marshall, F. Meschi, C. J. O'Keefe, H. Ordonez, P.
429 Patel, A. Price, A. Royall, E. Ruark, S. Seal, M. Schnall-Levin, P. Shah, D. Stafford, S. Williams, I.
430 Wu, A. W. Xu, N. Rahman, D. MacArthur, D. M. Church, Resolving the full spectrum of human
431 genome variation using Linked-Reads. *Genome Res.* (2019), doi:10.1101/gr.234443.118.
432 38. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L.
433 Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell*
434 *Syst*. **3**, 95–98 (2016).
435 39. O. Dudchenko, S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S.
436 Shamim, I. Machol, E. S. Lander, A. P. Aiden, E. L. Aiden, De novo assembly of the Aedes aegypti
437 genome using Hi-C yields chromosome-length scaffolds. *Science*. **356**, 92–95 (2017).
438 40. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L.
439 Aiden, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *cels*.
440 **3**, 99–101 (2016).
441 41. S. Yeo, L. Coombe, R. L. Warren, J. Chu, I. Birol, ARCS: scaffolding genome drafts with
442 linked reads. *Bioinformatics*. **34**, 725–731 (2018).
443 42. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly
444 using phased assembly graphs with hifiasm. *Nat Methods*. **18**, 170–175 (2021).
445 43. M. Tarailo-Graovac, N. Chen, *Current Protocols in Bioinformatics*, in press,
446 doi:10.1002/0471250953.bi0410s25.
447 44. A. Morgulis, E. M. Gertz, A. A. Schäffer, R. Agarwala, A fast and symmetric DUST
448 implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
449 45. G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*
450 **27**, 573–580 (1999).
451 46. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
452 features. *Bioinformatics*. **26**, 841–842 (2010).
453 47. C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management
454 tool for second-generation genome projects. *BMC Bioinformatics*. **12**, 491 (2011).
455 48. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence

456 comparison. *BMC Bioinformatics*. **6**, 31 (2005).
457 49. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson,
458 T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
459 50. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L.
460 Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals
461 unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–
462 515 (2010).
463 51. K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised
464 RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. **32**,
465 767–769 (2016).
466 52. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: ab initio
467 prediction of alternative transcripts. *Nucleic Acids Res*. **34**, W435-439 (2006).
468 53. K. Eilbeck, B. Moore, C. Holt, M. Yandell, Quantitative measures for the management and
469 comparison of annotated genomes. *BMC Bioinformatics*. **10**, 67 (2009).
470 54. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO Update: Novel
471 and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of
472 Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*. **38**, 4647–4654
473 (2021).
474 55. A. J. Vilella, J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, E. Birney, EnsemblCompara
475 GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. **19**, 327–
476 335 (2009).
477 56. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search
478 tool. *J Mol Biol*. **215**, 403–410 (1990).
479 57. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: combining multiple
480 sequence alignment methods with T-Coffee. *Nucleic Acids Res*. **34**, 1692–1699 (2006).
481 58. E. Parey, A. Louis, C. Cabau, Y. Guiguen, H. R. Crollius, C. Berthelot, Synteny-guided
482 resolution of gene trees clarifies the functional impact of whole genome duplications. *Mol. Biol. Evol.*
483 (2020), doi:10.1093/molbev/msaa149.
484 59. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
485 phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
486 60. D. Arcila, G. Ortí, R. Vari, J. W. Armbruster, M. L. J. Stiassny, K. D. Ko, M. H. Sabaj, J.
487 Lundberg, L. J. Revell, R. Betancur-R., Genome-wide interrogation advances resolution of
488 recalcitrant groups in the tree of life. *Nat Ecol Evol*. **1**, 1–10 (2017).
489 61. H. Shimodaira, M. Hasegawa, CONSEL: for assessing the confidence of phylogenetic tree
490 selection. *Bioinformatics*. **17**, 1246–1247 (2001).
491 62. H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. **51**,
492 492–508 (2002).
493 63. Y. Nakatani, A. McLysaght, Genomes as documents of evolutionary history: a probabilistic
494 macrosynteny model for the reconstruction of ancestral genomes. *Bioinformatics*. **33**, i369–i378
495 (2017).
496 64. E. Parey, A. Louis, J. Monfort, Y. Guiguen, H. R. Crollius, C. Berthelot, A high-resolution
497 comparative atlas across 74 fish genomes illuminates teleost evolution after whole-genome
498 duplication (2022), p. 2022.01.13.476171, , doi:10.1101/2022.01.13.476171.
499 65. Z. Hao, D. Lv, Y. Ge, J. Shi, D. Weijers, G. Yu, J. Chen, RIdeogram: drawing SVG graphics
500 to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci*. **6**, e251 (2020).
501 66. G. Nelson, Relationships of clupeomorphs, with remarks on the structure of the lower jaw in

fishes. *Interrelationships of fishes*, 333–349 (1973).
67. L. Taverne, *Ostéologie, phylogénèse et systématique des Téléostéens fossiles et actuels du super-ordre des Ostéoglossomorphes: Ostéologie des genres Hiodon, Eohiodon, Lycoptera, Osteoglossum, Scleropages, Heterotis et Arapaima* (Palais des académies, 1977; https://books.google.fr/books?id=AhwMAQAAIAAJ), vol. 42 of *Académie Royale de Belgique, Mémoires de la Classe des Sciences*.
68. E. J. Hilton, *Osteology of the extant North American fishes of the genus Hiodon Lesueur, 1818 (Teleostei: Osteoglossomorpha: Hiodontiformes)* (Field Museum of Natural History, 2002; https://www.biodiversitylibrary.org/bibliography/2666), *Fieldiana (Zoology) New Series*.
69. B. C. Faircloth, L. Sorenson, F. Santini, M. E. Alfaro, A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS One*. **8**, e65923 (2013).
70. C. M. Austin, M. H. Tan, L. J. Croft, M. P. Hammer, H. M. Gan, Whole Genome Sequencing of the Asian Arowana (Scleropages formosus) Provides Insights into the Evolution of Ray-Finned Fishes. *Genome Biol Evol*. **7**, 2885–2895 (2015).



519

520 **<u>Fig. S1.</u> Alignments used in previous teleost molecular phylogenies.** Total size of nucleotide
521 alignments leveraged for previous phylogeny inference in teleosts (*10–12*, *16–18*, *69*, *70*), compared
522 with our dataset. Alignment sizes are provided as reported by the authors of the respective studies,
523 with the exception of Chen et al 2015 (*16*), as the authors did not only focus on teleost fishes but
524 inferred a complete vertebrate phylogeny. We thus report here the subset of the alignment that is
525 informative for teleost fishes (i.e., with a low proportion of gaps and missing data in teleosts), as
526 identified by Takezaki 2021 (*9*).

527

528

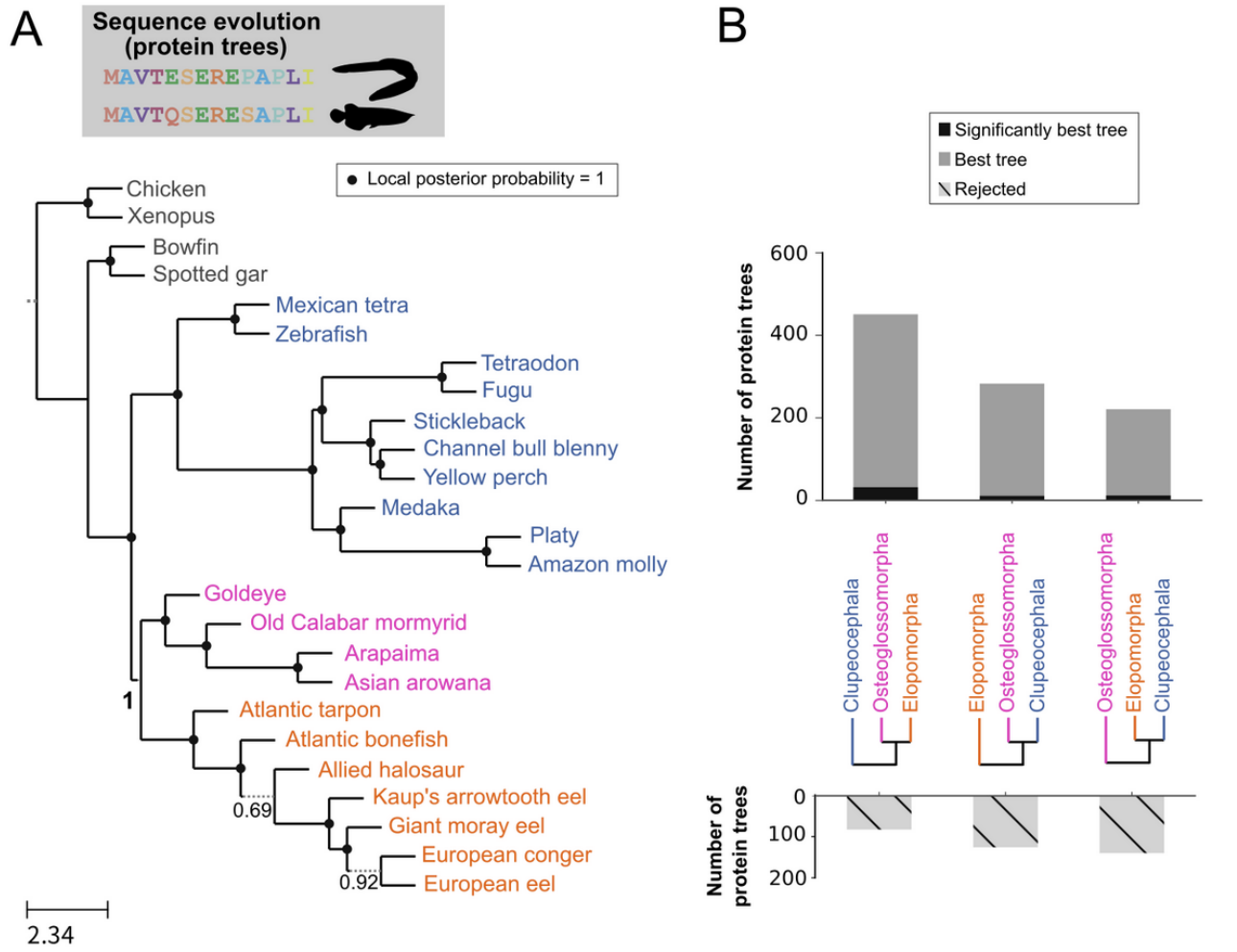
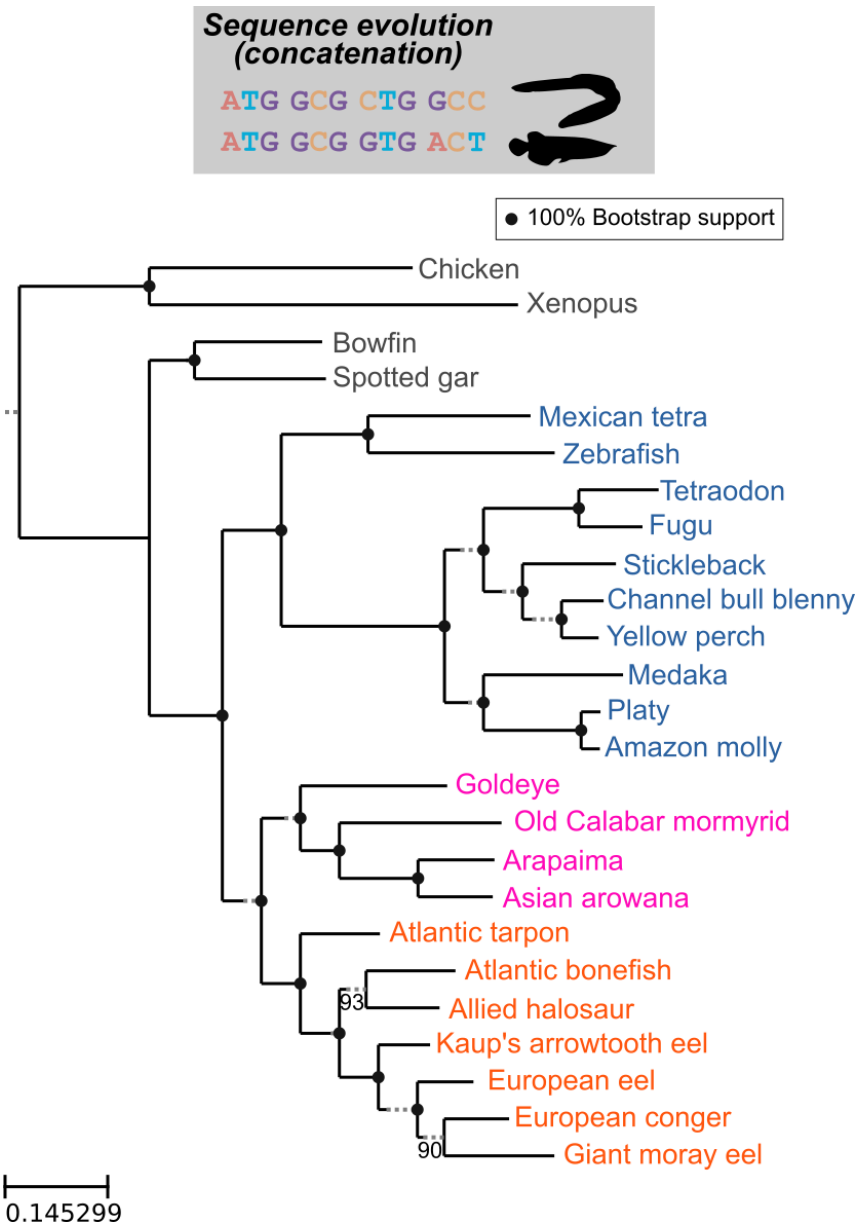
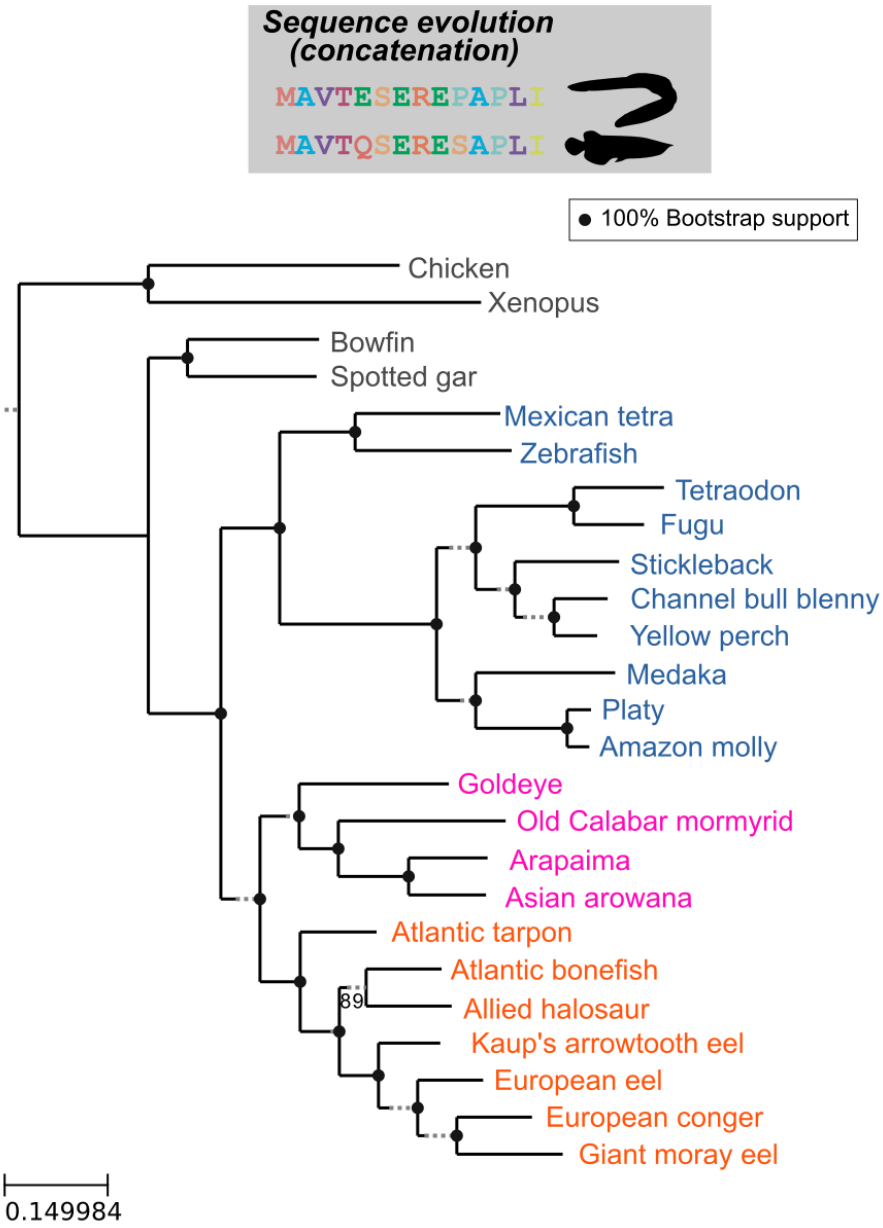


Fig. S2. Molecular phylogeny inferred from protein trees. A. Species tree inferred with ASTRAL-III from 955 single-copy protein trees. **B.** Gene genealogy interrogation: number of protein trees supporting each hypothesis (top, gray bars), significantly supporting each hypothesis (top, black bars), and number of significantly rejected gene trees (bottom, dashed bar).



<u>Fig. S3.</u> Molecular phylogeny inferred from the concatenation of nucleotide coding sequences. RAxML tree inferred from the concatenation of the 955 nucleotide coding sequences.



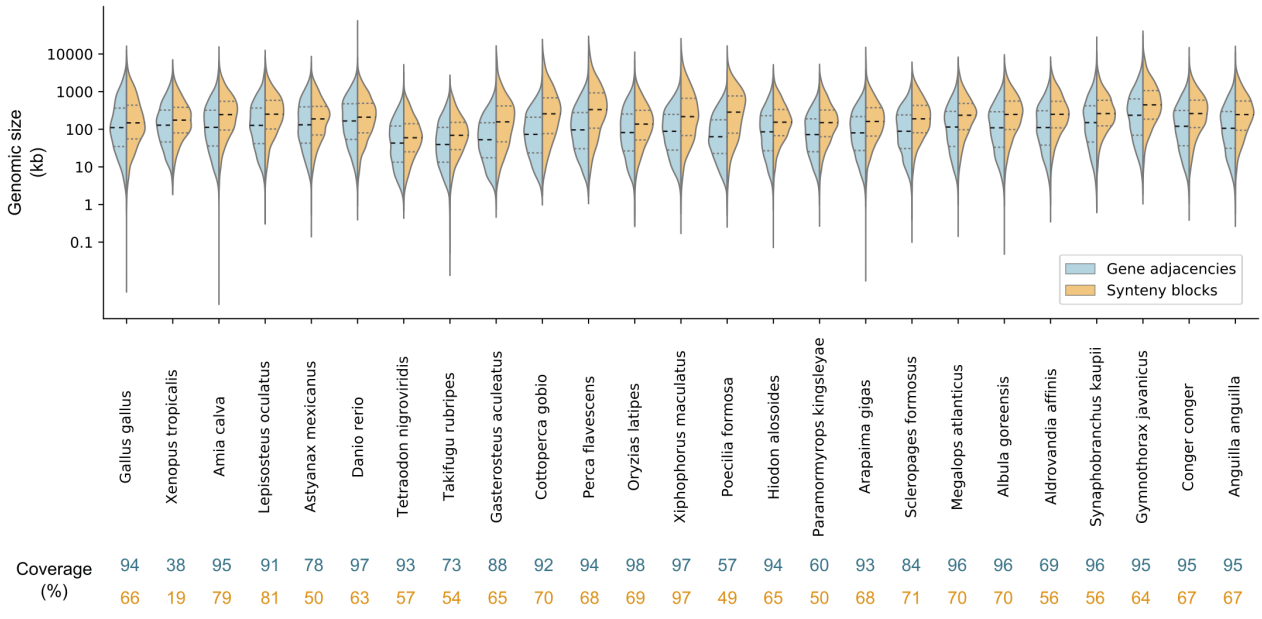
543

544

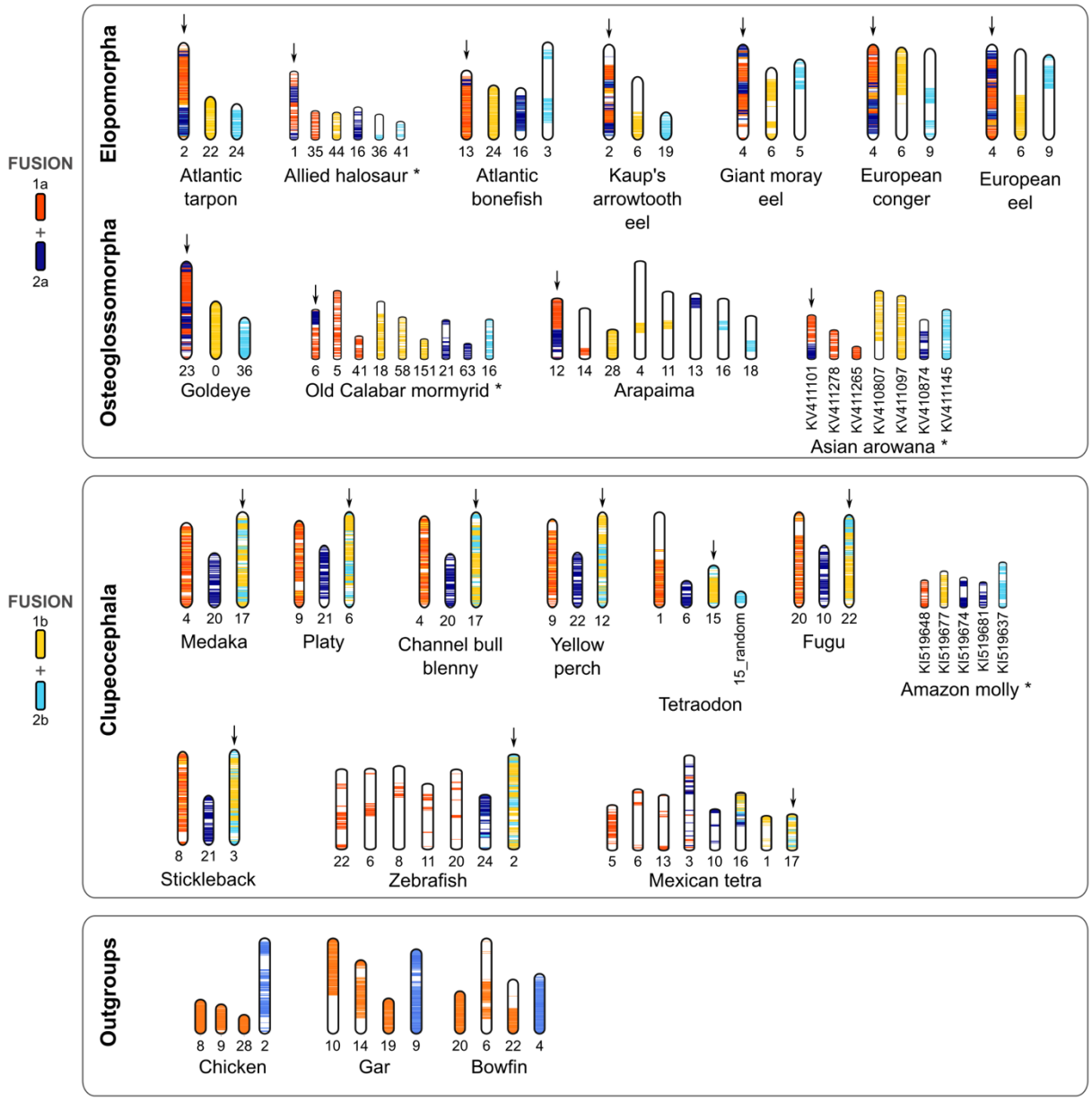
545 **<u>Fig. S4.</u> Molecular phylogeny inferred from the concatenation of protein sequences.** RAxML
546 tree inferred from the concatenation of the 955 protein sequences.

547

548



<u>Fig. S5.</u> Genomic size of the gene adjacencies and synteny blocks leveraged for phylogeny inference. For each genome, we show the distribution of distances between adjacent marker genes (microsynteny analysis) and the sizes of synteny blocks (macrosynteny analysis). We performed a total of 24 pairwise synteny inferences for each genome, but selected here the comparison with the most representative mean, for visualization purposes. Genomic coverage is indicated below the plot: end-to-end genomic coverage of adjacent marker genes (shown in blue, first line) and total coverage of synteny blocks (shown in orange, second line).



<u>Fig. S6.</u> Fusions of ancestral chromosomes across all teleosts and outgroup genomes considered in this study. Chromosomes are colored as in Figure 4, modern chromosomes descending from the fusion of ancestral chromosomes are indicated by an arrow. Stars at the end of species names indicate genomes with a scaffold-level assembly. The fusion of ancestral chromosomes 1a and 1b is identifiable in all Osteoglossomorpha and Elopomorpha genomes. Note that a subsequent fission seems to have occurred in the allied halosaur. The fusion of ancestral chromosomes 2a and 2b is identifiable in all Clupeocephala genomes with the exception of the Amazon molly, which has a highly fragmented assembly.

Table S1. **Elopomorpha genome assembly metrics**

Scientific name	Genome size	Contig N	Contig N50	Contig L50	Chr N	% Chr	Scaffold N50	Scaffold L50	Buscos scores (C ; S ; D ; F ; M)
Megalops atlanticus	0.99	920	17.7	71	25	98.2	39.4	10	96.3 ; 86.2 ; 10.1 ; 0.6 ; 3.1
Albula goreensis	0.93	993	17.5	90	25	99.3	39.1	11	95.4 ; 87.9 ; 7.5 ; 0.4 ; 4.2
Aldrovandia affinis	1.23	3957	2	169	NA	NA	3.8	88	89.2 ; 82.4 ; 6.8 ; 2.8 ; 8.0
Synaphobranchus kaupii	1.5	5597	0.5	3068	24	97.9	76.8	7	91.0 ; 84.1 ; 6.9 ; 2.2 ; 6.8
Gymnothorax javanicus	2.1	2280	8.5	59	20	98.9	133.6	7	92.8 ; 88.3 ; 4.5 ; 1.5 ; 5.7
Anguilla anguilla	1	2729	4.3	49	19	97.7	55.5	8	93.5 ; 86.6 ; 6.9 ; 1.8 ; 4.
Conger conger	1	3867	1.4	194	19	96.8	55.2	8	91.4 ; 85.9 ; 5.5 ; 1.4 ; 7.2

Genome assembly size (Gb), Contig N = number of contigs, Contig or Scaffold N50 (Mb), Chr N = number of chromosomes, % Chr = percentage of the assembly anchored in chromosomes, Buscos (V4, in genome mode with actinopterygii lineage) scores in percentage (C = Complete, S = Single copy, D = Duplicated, F = Fragmented, M = Missing). N.A = Not Applicable as the *Aldrovandia affinis* genome assembly was not anchored on chromosomes.

Table S2. Information on the Elopomorpha specimens collected for genome sequencing

Species name	Sequencing	Specimen origin	Sample collectors	Sampling date	Sex	Developmental stage
Megalops atlanticus	ONT, 10X, RNA-Seq	Aquarium trade, Nigeria coast according to provider	Yann Guiguen, Julien Bobe, Ming Wen INRAE LPGP, France	March 15, 2019	unknown	Juvenile
Albula goreensis	ONT, 10X, HIC, RNA-Seq	Florida sea	Sahar Mejri, Aaron J. Adams Florida University, USA	July 30, 2019	unknown	adult
Aldrovandia affinis	ONT, 10X, RNA-Seq	New Caledonia sea (at a depth around 1080 m)	Wei-Jen Chen National Taiwan University, Taiwan	September 29, 2019	female	adult
Synaphobranchus kaupii	ONT, 10X, RNA-Seq	Taiwan sea (station: CP4170, 22°17.466N 119°59.652 E, depth 967 m)	Wei-Jen Chen, Janette Chen National Taiwan University, Taiwan	November 01, 2017	unknown	adult
	HIC	Taiwan sea (CP4209; ~1000m, off Kaohsiung)	Wei-Jen Chen National Taiwan University, Taiwan	November 03, 2019	female	adult
Gymnothorax javanicus	ONT, 10X, HIC, RNA-Seq	Taiwan sea	Wei-Jen Chen, Michelle Lin National Taiwan University.	September 07, 2018	unknown	adult
Anguilla anguilla	ONT, Illumina	Netherlands, Lake Veere	Future Genomics Technologies The Netherlands	2014	female	adult
	HIC	France, Grand-Lieu lake	Yann Guiguen, Ming Wen INRAE LPGP, France	December 23, 2018	unknown	adult
Conger conger	ONT, 10X, HIC, RNA-Seq	France, South Ouest coast near Lorient	Yoann Guilloux, Fabien Quendo, Ming Wen (INRAE LPGP) France	July 23, 2019	unknown	adult

Table S3. Elopomopha long read sequencing. NR = not recorded

Species name	Shearing (Kb)	Optionnal Sizing	Loading	FlowCell	Output (Gb)	Total output (Gb)
Megalops atlanticus	20	no	25 fM	MinION 48H	7	50
	20	no	20 fM	PromethION 64H	43	
Albula goreensis	20	Circulomics XS	22 fM	GridION 72H	5	44
	20	Circulomics XS	22 fM	PromethION 72H	39	
Aldrovandia affinis	20	Circulomics XS	25 fM	GridION 72H	7	47
	20	Circulomics XS	24 fM	PromethION 72H	17	
	20	Circulomics XS	25 fM	PromethION 72H	23	
Synaphobranchus kaupii	20	no	50 pM	Sequel II HiFi SMRTcell 30 H	12 (HiFi reads)	20 (HiFi reads)
	20	no	60 pM	Sequel II HiFi SMRTcell 30 H	8 (HiFi reads)	
Gymnothorax javanicus	20	no	30 fM	GridION 48H	15	105
	25	no	30 fM	PromethION 64H	35	
	25	no	25 fM	PromethION 64H	55	
Anguilla anguilla	NR	no	NR	GridION 48H	11	35
	NR	no	NR	PromethION 64H	24	
Conger conger	35	Circulomics XS	16 fM	GridION 48H	3	39
	30	Circulomics XS	8 fM	Flongle 24H	0.09	
	35	Circulomics XS	16 fM	GridION 48H	5	
	20	Circulomics XS	24 fM	PromethION 72H	31	

Table S4. Information on the species and the genome assembly resources used in this study

Common name	Scientific name	class (subclass)	infraclass	Cohort	Order	Family	Genome assembly ID
Chicken (1)	*Gallus gallus*	Aves	NA	NA	Galliformes	Phasianidae	GCA_000002315.5
Xenopus (2)	*Xenopus tropicalis*	Amphibia	NA	NA	Anura	Pipidae	GCA_000004195.1
Bowfin	*Amia calva*	Actinopteri (Neopterygii)	Holostei	NA	Amiiformes	Amiidae	GCA_017591415.1
Spotted gar	*Lepisosteus oculatus*	Actinopteri (Neopterygii)	Holostei	NA	Lepisosteiformes	lepisosteidae	GCA_000242695.1
Atlantic tarpon	*Megalops atlanticus*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Elopiformes	Megalopidae	GCA_019176425.1
Atlantic bonefish	*Albula goreensis*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Albuliformes	Albulidae	Submitted & Available Dataverse
Allied Halosaur	*Aldrovandia affinis*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Notacanthiformes	Halosauridae	Submitted & Available Dataverse
Kaup's arrowtooth eel	*Synaphobranchus kaupii*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Anguilliformes	Protanguilloidae	Submitted & Available Dataverse
Moray eel	*Gymnothorax javanicus*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Anguilliformes	Muraenidae	Submitted & Available Dataverse
European eel	*Anguilla anguilla*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Anguilliformes	Anguillidae	GCA_018320845.1
European conger	*Conger conger*	Actinopteri (Neopterygii)	Teleostei	Elopomorpha	Anguilliformes	Congridae	Submitted & Available Dataverse
Goldeye	*Hiodon alosoides*	Actinopteri (Neopterygii)	Teleostei	Osteoglossomorpha (3)	Hiodontiformes	Hiodontidae	Available Dataverse
Old Calabar mormyrid	*Paramormyrops kingsleyae*	Actinopteri (Neopterygii)	Teleostei	Osteoglossomorpha (3)	Osteoglossiformes	Mormyridae	GCA_002872115.1
Arapaima	*Arapaima gigas*	Actinopteri (Neopterygii)	Teleostei	Osteoglossomorpha (3)	Osteoglossiformes	Osteoglossidae	GCA_007844225.1
Asian arowana	*Scleropages formosus*	Actinopteri	Teleostei	Osteoglossomorpha	Osteoglossiformes	Osteoglossidae	GCA_001624265.1

		(Neopterygii)		(3)			
Mexican tetra	*Astyanax mexicanus*	Actinopteri (Neopterygii)	Teleostei	Otomorpha	Characiformes	Characidae	GCA_000372685.2
Zebrafish	*Danio rerio*	Actinopteri (Neopterygii)	Teleostei	Otomorpha	Cypriniformes	Danionidae	GCA_000002035.4
Tetraodon	*Tetraodon nigroviridis*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Tetraodontiformes	Tetraodontidae	Ensembl TETRAODON 8.0
Fugu	*Takifugu rubripes*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Tetraodontiformes	Tetraodontidae	GCA_000180615.2
Stickleback	*Gasterosteus aculeatus*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Perciformes	Gasterosteidae	Ensembl broad S1
Channel bull blenny	*Cottoperca gobio*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Perciformes	Bovichtidae	GCF_900634415.1
Yellow perch	*Perca flavescens*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Perciformes	Percidae	GCF_004354835.1
Medaka	*Oryzias latipes*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Atheriniformes	Adrianichthyidae	GCA_002234675.1
Platy	*Xiphophorus maculatus*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Cyprinodontiformes	Poeciliidae	GCA_002775205.2
Amazon molly	*Poecilia formosa*	Actinopteri (Neopterygii)	Teleostei	Euteleosteomorpha	Cyprinodontiformes	Poeciliidae	GCA_000485575.1

(1) & (2) not ranked according to Betancur et al., 2017 (*12*); (3) = Supercohort according to Betancur et al., 2017 (*12*); NA = Not applicable; Available dataverse = https://doi.org/10.15454/GWL0GP