



## Molecular phylogenetics of the Clupeiformes based on exon-capture data and a new classification of the order

Qian Wang<sup>a,b</sup>, Leyli Purrafee Dizaj<sup>c</sup>, Junman Huang<sup>b,d</sup>, Kishor Kumar Sarker<sup>b,d</sup>, Charalampos Kevrekidis<sup>e</sup>, Bettina Reichenbacher<sup>e,f</sup>, Hamid Reza Esmaeili<sup>c</sup>, Nicolas Straube<sup>g</sup>, Timo Moritz<sup>h,i,\*</sup>, Chenhong Li (李晨虹)<sup>b,d,\*</sup>

<sup>a</sup> East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China

<sup>b</sup> Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China

<sup>c</sup> Ichthyology and Molecular Systematics Research Laboratory, Zoology Section, Department of Biology, School of Science, Shiraz University, Shiraz, Iran

<sup>d</sup> Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, Shanghai 201306, China

<sup>e</sup> Ludwig-Maximilians-Universität München, Department für Geo- und Umweltwissenschaften, Paläontologie & Geobiologie, Richard-Wagner-Str. 10, 80333 Munich, Germany

<sup>f</sup> GeoBio-Center, Ludwig-Maximilians-Universität München, Munich, Germany

<sup>g</sup> University Museum, Department of Natural History, University of Bergen, Norway

<sup>h</sup> Deutsches Meeresmuseum, Katharinenberg 14-20, 18439 Stralsund, Germany

<sup>i</sup> Institute of Biological Sciences, University of Rostock, Albert-Einstein-Straße 3, 18059 Rostock, Germany

### ARTICLE INFO

#### Keywords:

Clupeiform fishes  
Clupeidae  
Target enrichment  
Nuclear genes  
Divergence time  
Taxonomic revision

### ABSTRACT

The Clupeiformes, including among others herrings, anchovies, shads and menhadens are ecologically and commercially important, yet their phylogenetic relationships are still controversial. Previous classification of Clupeiformes were based on morphological characters or lack of synapomorphic characters. More recent studies based on molecular data as well as new morphological evidence are keeping challenging their phylogenetic relations and there is still no consensus on many interrelationships within the Clupeiformes. In this study, we collected nuclear sequence data from 4,434 single-copy protein coding loci using a gene-capture method. We obtained a robust phylogeny based on 1,165 filtered loci with less than 30 % missing data. Our major findings include: 1) reconfirmation of monophyly of the Clupeiformes, that is, Denticipitidae is sister to all other clupeiforms; 2) the polyphyletic nature of dussumieriids and early branching of Spratelloididae from all other clupeoids were confirmed using datasets curated for less missing data and more balanced base composition in the respective taxa. The next branching clade is the monophyletic Engraulidae. Pristigasteridae also is monophyletic, but it was nested in the previously defined “Clupeidae”. Within Pristigasteridae there is no support for monophyletic Pelloninae. *Chirocentrus* is close to *Dussumieria* and not to engraulids. The miniaturized *Sundasalanx* is placed close to the ehiravine *Clupeonella*, however, with a relatively deep split. The genus *Clupea*, is not part of the diverse “Clupeidae”, but part of a clade containing additionally *Sprattus* and *Etrumeus*. Within the crown group clades, Alosidae and Dorosomatidae are retrieved as sister clades. Based on new fossil calibration points, we found that major lineages of the clupeiforms diverged in the late Cretaceous and early Paleogene. The extinction event at the end of the Cretaceous may have created ecological niches, which could have fueled the diversification of clupeiform fishes. Based on the strong evidence of the present study, we propose an updated classification of Clupeiformes consisting of ten families: Denticipitidae; Spratelloididae; Engraulidae (Engraulinae + Coiliinae); Clupeidae; Chirocentridae; Dussumieriidae; Pristigasteridae; Ehiravidae; Alosidae, Dorosomatidae.

\* Corresponding authors at: Shanghai Ocean University, Shanghai 201306, China (C. Li); Deutsches Meeresmuseum, Katharinenberg 14-20, 18439 Stralsund, Germany (T. Moritz).

E-mail addresses: [wangqian@ecs.ac.cn](mailto:wangqian@ecs.ac.cn) (Q. Wang), [lpurrafee@gmail.com](mailto:lpurrafee@gmail.com) (L. Purrafee Dizaj), [dreamenmurphy@hotmail.com](mailto:dreamenmurphy@hotmail.com) (J. Huang), [kishorssf91@gmail.com](mailto:kishorssf91@gmail.com) (K. Kumar Sarker), [xaralamboskevrekidis@yahoo.gr](mailto:xaralamboskevrekidis@yahoo.gr) (C. Kevrekidis), [b.reichenbacher@lrz.uni-muenchen.de](mailto:b.reichenbacher@lrz.uni-muenchen.de) (B. Reichenbacher), [hresmaeili22@gmail.com](mailto:hresmaeili22@gmail.com) (H. Reza Esmaeili), [nicolas.straube@uib.no](mailto:nicolas.straube@uib.no) (N. Straube), [timo.moritz@meeresmuseum.de](mailto:timo.moritz@meeresmuseum.de) (T. Moritz), [chli@shou.edu.cn](mailto:chli@shou.edu.cn) (C. Li).

<https://doi.org/10.1016/j.ympev.2022.107590>

Received 24 May 2022; Received in revised form 21 June 2022; Accepted 12 July 2022

Available online 16 July 2022

1055-7903/© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

Herrings and allies, i.e. Clupeomorpha sensu Greenwood et al. (1966) or Clupei sensu Wiley and Johnson (2010), presently comprise 435 extant species (Fricke et al., 2022). These herrings, anchovies, menhadens, shads and allies mostly inhabit coastal marine and euryhaline waters with half of the diversity occurring in the Indo-West Pacific. About 18 % of all species primarily inhabit freshwater habitats (Nelson et al., 2016) and about 7 % of the species show an anadromous lifestyle (Bloom et al., 2018). Many herring-like species have been of significant importance as food source for humans as they live in surface waters close to shore in large schools. Approximately 18 % of the global fishery catches, i.e. about 16 million tons of a total of 92.5 million tons are Clupei (FAO, 2021). For example, despite its small size the anchoveta (*Engraulis ringens*) is the most commonly caught species representing about 10 % (in weight) of the worldwide fisheries catches (FAO, 2020). Due to the high ecological and economic importance of these Clupei species, their stock assessment, life history and taxonomy have been extensively studied. Also, many attempts to resolve the phylogenetic intrarelationships of the Clupei have been undertaken. However, several contradictory hypotheses have been published in the last decades (Bloom et al., 2018; Bloom and Lovejoy, 2012, 2014; Di Dario, 2009; Grande, 1985; Lavoué et al., 2014, 2013, 2017; Li and Ortí, 2007; Miyashita, 2010; Nelson, 1967, 1970b; Whitehead, 1985; Wilson et al., 2008), and as of today the evolutionary relationships of Clupeiformes are not

resolved with confidence (Fig. 1).

Before the seminal work of Greenwood et al. (1966), herring-like fish were usually lumped together with many other ‘basal’ teleost taxa forming the Isospondyli (e.g., Cope, 1871; Regan, 1909). Although their position within the teleosts was vague over decades, synapomorphies such as the *recessus lateralis*, intracranial diverticula of the swim bladder or the fusion of the second hypural to the first ural centrum provided good evidence for their monophyly. Grande (1985) extensively investigated the intrarelationships of the Clupei based on morphology. Few studies added morphological details to this (Di Dario, 2002, 2009; Miyashita, 2010; Sato, 1994). These studies, as well as morphological studies on *Denticeps* (Greenwood and Chandler, 1960; Greenwood, 1968a, 1968b) all agreed that *Denticeps* represents the sister-group to all other recent herring-like fish. However, morphological investigations were inconsistent in their phylogenetic placement for some taxa, e.g. *Chirocentrus* or pristigasterids, or the coverage and intrarelationships of ‘true herrings’, the family Clupeidae sensu lato. In the advent of molecular phylogenetics, from 2007 onwards several molecular studies have been published testing phylogenetic hypotheses for the Clupei (Bloom et al., 2018; Bloom and Lovejoy, 2012, 2014; Lavoué et al., 2014, 2013, 2017; Li and Ortí, 2007). However, hitherto published molecular phylogenies are partially contradicting or contain polytomies. For example, the intrarelationships within the crown-group, the Clupeidae sensu lato, largely remained unresolved, and dussumieriids are paraphyletic containing taxa distributed across the entire clupeiform

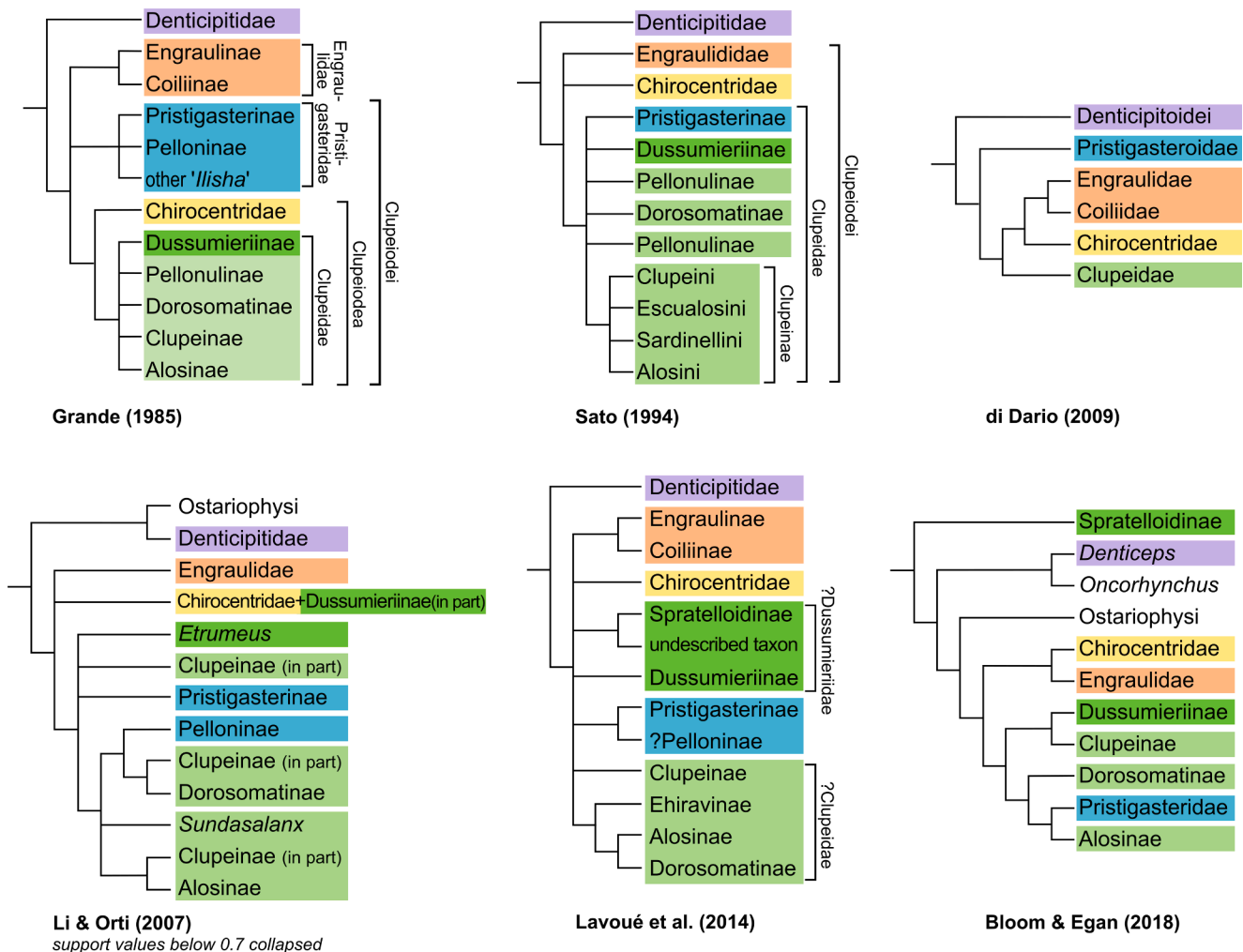


**Fig. 1.** Selected morphological (upper row from Grande, 1985; Sato, 1994; di Dario 2009) and molecular (lower row from Li & Ortí, 2007 based on 12 s, 16 s, rag1 and rag2 genes; Lavoué et al., 2014 based on mitochondrial genomes; Bloom and Egan, 2018 based on cytb, 16 s, rag1 and rag2 genes) hypotheses on the phylogeny of Clupeiformes.

phylogeny. In some studies Clupei do not form a monophyletic clade as *Denticeps* was recovered as sister to all other Clupei and Ostariophysi (Bloom and Lovejoy, 2014) or *Denticeps* was more closely related to Ostariophysi than to other Clupei (Fig. 1) (Li and Ortí, 2007).

Several studies have shown that base composition and GC biases are crucial metrics in phylogenetics and phylogenomics. Heterogeneity among loci of the analyzed taxa often provide misleading signal during phylogenetic tree reconstruction, i.e. detection of natural selection and estimation of codon usage certainly influences downstream molecular evolution analyses (Bossert et al., 2017; Romiguier and Roux, 2017). A reason for the difficulties in resolving the molecular phylogeny may be found in biased base composition of some taxa. Li and Ortí (2007) revealed that *Denticeps clupeioides* had the lowest GC content among all species of Clupei. The analysis on Li and Ortí (2007) on RY-coded data considering base compositional bias, rejected the close evolutionary relationship between *Denticeps* and Ostariophysi over *Denticeps* and Clupei which provided support for the morphology-based phylogenies. Nevertheless, a phylogeny with *Denticeps* as sister to the euteleost *Oncorhynchus*, Ostariophysi as sister to most Clupei, and Spratelloidinae as sister to all clupeocephalans was recently published (Fig. 1) (Bloom and Egan, 2018) and thus opened again the controversy of the association of Denticipitoidei with Clupeoidei.

A comprehensive investigation of the intrarelationships of Clupei should be based on both phylogenomic approach as well as a large taxon sampling. Advancing molecular techniques provide constantly improved methods to obtain not only larger but also tailored datasets to study evolutionary interrelationships. Remaining challenges of such datasets involve systematic errors and/or methodological artifacts. Systematic errors are for example incomplete lineage sorting, horizontal gene transfer, or uneven GC content across species (Li, 2007). Methodological artifacts include, e.g., improper phylogenetic approaches, inadequate character sampling, or incomplete taxon sampling (Li, 2007). The current trends in molecular phylogenetics are towards assembling large data matrices. Robust probabilistic models coupled with genome-scale datasets usually can accommodate heterogeneity among sites and sampling errors that aid to resolve systematic errors within a phylogenetic tree (Li, 2007). Meanwhile, high-throughput (next-generation) sequencing permits us to surmount limitations of data quantity for a large number of taxa in a rapid and cost-effective manner (Finstermeier et al., 2013; Oliveira et al., 2018; Prum et al., 2015).

Target gene capture, also known as target enrichment, is based on hybridizing RNA/DNA baits (probes) to DNA libraries of targeted species and fishing for sequences of thousands of loci for subsequent high-throughput sequencing. Target gene capture relies on selective enrichment of chosen genomic regions by hybridization to probes. These probes are short DNA or RNA molecules (~70–150 nucleotides) designed to capture usually exons or other conserved regions of the genome (Oliveira et al., 2018), which have been adopted for a variety of invertebrates and vertebrates (Crawford et al., 2012; Faircloth et al., 2012; Hugall et al., 2016; Jiang et al., 2019; Kuang et al., 2018; Lemmon et al., 2012; Li et al., 2013; Mayer et al., 2016; McCormack et al., 2012; Straube et al., 2015; Straube et al., 2018; Teasdale et al., 2016; Yuan et al., 2016). An adjusted gene capture method has been developed by Li et al. (2013) to target and isolate thousands of genes of interest across divergent taxa which is highly applicable for molecular systematics and evolutionary genetics (e.g. Jiang et al., 2019; Kuang et al., 2018; Straube et al., 2015; Yuan et al., 2016).

Here we applied target gene capture to sequence thousands of loci from samples representing the Clupei for retracing their phylogenetic relationships. The aims of this study are to: i) resolve controversial relationships among and within the families of the Clupei; ii) examine the effect of GC bias and missing data on the phylogenetic position of Spratelloidinae lineage; iii) provide a robust phylogeny of the Clupei allowing for fossil-calibrated divergence time estimation; and iv) propose a new classification for the Clupei.

## 2. Material and methods

### 2.1. Taxon sampling and DNA extraction

Over a time period of 12 years (2005–2018), specimens were collected and preserved from 14 countries covering major water bodies of clupeid distribution ranges. After initial morphological observations, fin or muscle tissues were directly fixed in 70 % to 100 % molecular grade ethanol. In total, tissues from 178 samples were obtained and preserved at 4 °C until DNA extraction. The sampled taxa represent 66 species (15 % of 436 species) and 39 genera (47 % of 83 genera) from all extant clupeiforms (Fricke et al., 2022; Lavoué et al., 2013). Three species were collected as out groups from the orders Gymnotiformes (*Apternotus albifrons*), Siluriformes (*Ictalurus punctatus*) and Characiformes (*Pygocentrus nattereri*). Besides these outgroups stated above, genomes of *Danio rerio* (GCA\_000002035) and *Lepisosteus oculatus* (GCA\_000242695) were retrieved from the NCBI repository to extract orthologous genes and to use as outgroup as well. Sample details are listed in Appendix A, Table S1. Total genomic DNA was extracted from tissues using Ezup Column Animal Genomic DNA Purification Kit (B518251; Sangon Biotech, Shanghai, China) following the manufacturer's instructions. The concentration of extracted DNA was measured using a NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

### 2.2. Cross species gene enrichment and sequencing

Illumina® DNA library preparation and target gene capture were carried out following Meyer and Kircher's (2010) protocol with some modifications (Li et al., 2013) to harbor target loci from divergent taxa. Prior to library preparation, the genomic DNA was fragmented using a M220 Focused-ultrasonicator (Covaris, Inc., MA, USA). A total of 500 ng of DNA was diluted in 270 µl and sonicated for each sample. DNA was sheared to pieces around 250 bp length. Length distribution of DNA fragments was checked visually using agarose gel electrophoresis. Illumina universal adapters were subsequently ligated to the sheared DNA fragments. The adapters contained 6 bp inline-indices after repairing the blunt ends that later aided to identify the cross contamination (Wang et al., 2022).

A suite of 4,434 single-copy nuclear exon markers for all ray-finned fishes was developed by Jiang et al. (2019). RNA-baits targeting these loci were designed based on sequences of *Denticeps clupeioides* and *Ilisha elongata* (both Clupeiformes) (#ClupeiformsV2, Daicel Arbor Biosciences, Ann Arbor, MI 48103, USA). The baits were used for hybridization following the MYbaits-manual-v3 protocol (Daicel Arbor Biosciences, Ann Arbor, MI 48103, USA) with some modifications described in Huang et al. (2021). Every sample was enriched twice during exon capture in order to increase the proportion of on-target reads as shown in e.g. Li et al. (2013). The amplified DNA library generated from the first enrichment experiment was used as basis for the 2nd round of capture. After the second enrichment, each sample was amplified with indexed primers so that all samples could be pooled together for sequencing and then the sequenced reads could be distinguished by their index. The concentration of twice enriched and amplified libraries was measured using a NanoDrop 3300. The samples then were pooled equimolarly (typically 12–16 samples per pool). Gel-based fragment size selection was conducted using a QIAquick Gel Extraction Kit (Cat #28706; QIAGEN China, Shanghai, China). DNA fragments larger than 1 kb were removed to insure high sequencing quality. The pooled samples were sequenced for 150 bp pair-end reads on an Illumina HiSeq 2500 sequencer by GENEWIZ (Suzhou, China).

### 2.3. Data preparation and read assembly

The raw sequencing reads were processed and assembled using the Assexon pipeline (Yuan et al., 2019). Assexon includes two major steps,

sequence quality control and assembly before sequence alignment. Raw files returned from the sequencer were demultiplexed according to the index on the P7 adapter and the inline index. Adapter sequences and low-quality sites ( $Q < 20$ ) were trimmed from the reads using cutadapt (v3.2) (Martin, 2011) included within the wrapper tool 'trim\_galore (v0.6.4)' ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). PCR duplicates were excluded by using the "bfastx\_uniques" command in USEARCH (v10.0.240) (Edgar, 2010). Unique reads were parsed to targeted loci using the UBLAST algorithm implemented in USEARCH (v10.0.240) which allows to accept low identity reads. Preliminary contigs were assembled by using SGA v0.10.15 (Simpson and Durbin, 2012), where the maximum number of edges was kept even for reads with low sequencing-depth. Exonerate v2.2.0 (Slater and Birney, 2005) was used to select edge and extract candidate exons. Reciprocal BLAST strategy was carried out to predict putative orthologs in candidate exons. Alignment length, GC content, percentage of missing data and average pairwise distance were calculated using a custom Perl script (statistics.pl; Yuan et al., 2019).

## 2.4. Phylogenetic tree reconstruction and taxon filtering

Sequences were aligned using MAFFT (v7.475) (Kato and Standley, 2013). We constructed three different datasets allowing for different levels of missing data for phylogenetic analysis. First, only loci with less than 30 % missing taxa in the single locus alignments were included for analyses (dataset 1). In the second dataset, we included loci with less than 70 % missing taxa (dataset 2) and in the third approach, we included all loci independent of the amount of missing data (dataset 3). The custom Perl script pick\_taxa.pl -min\_seq obtained from Yuan et al. (2019) was used for this step. An initial data partition scheme was created using another custom Perl script (Appendix A, extract\_DNA\_blocks.pl) based on gene and codon position. PartitionFinder2 (Lanfear et al., 2017) was subsequently used to search for the best partition scheme under the GTR + G model and the rclusterf searching algorithm. Across all partitions, branch lengths were set as linked, and model selection was based on AICc criteria. The maximum likelihood phylogeny based on the concatenated datasets were reconstructed using IQ-TREE v2.1.2 (Minh et al., 2020) (-m GTR + G -spp partition\_file -bb 1000) and RAXML v8.2.12 (Stamatakis, 2014) (-f a -p 12,345 -m GTRGAMMA -# 500 -x 12,345 -q partition\_file) respectively applying the GTR + G substitution model to each partition. Two differently partitioned analyses were performed for each of the three datasets. The first used a partition scheme of codon position, and the second applied the best partition scheme resulting from the PartitionFinder2 analysis.

As preliminary results suggested that the spratelloidines split early from the other Clupeoidei (Fig. 1), many loci showed missing data as well as biased GC content in this group (Appendix A, Fig. S1 and S2). We selected from dataset 1 for loci that had sequence information for all spratelloidines (*Jenkinsia lamprotaenia*, *Spratelloides delicatulus* and *Spratelloides gracilis*) to reconstruct the phylogeny to test whether the basal position of this group was an artifact due to missing data (pick\_taxa.pl; Yuan et al., 2019). Similarly, we filtered out loci which had more than 10 % difference in average GC content in the third codon position between the spratelloidines and the other clupeoids to test whether biased GC content falsely led to their basal position in the resulting tree (Appendix A, gccontent.pl, sift\_gc.pl). Phylogenetic reconstruction using the matrices with less missing data and balanced GC content of the Spratelloidinae followed the same procedure as above.

## 2.5. Fossil calibration

To select the 100 most clock-like loci for divergence time estimation from dataset 1, the global likelihood ratio of each gene was calculated by comparing a tree with a strict molecular clock enforced against a tree without constraints reconstructed for each gene. Under the null hypothesis, the phylogeny was rooted and the branch lengths were

enforced to be clock-like. Under the alternative hypothesis, each branch was allowed to vary independently. The likelihood of each gene tree with and without enforcing the molecular clock was estimated using PAUP\* 4.0a (Swofford, 2002). The probability of no difference between the alternative and null hypothesis was calculated based on the likelihoods from above using the custom Perl script clocklikeness\_test.pl from Yuan et al. (2019). Outgroups of the Clupeiformes were excluded from the analysis. BEAUti v2.6.6 was used to create the XML file as input file for BEAST v2.6.6 (Bouckaert et al., 2014). Data were partitioned by three codon positions. Site models were unlinked across partitions, while clock models and trees were linked. The substitution model was GTRGAMMA with a substitution rate of 1.0, 4 Gamma category count and 1.0 as shape parameter. A relaxed clock log normal model and a birth-death tree model was applied. The chronogram estimation based on the maximum likelihood tree derived from dataset 1 was calibrated with fossil node dating (listed in the following paragraph) using a penalized likelihood method provided in R package ape v5.5 with the function "chronopl" (-lambda 1) was used (Sanderson, 2002). The XML file was edited to select the chronogram estimation (ultrametric tree) as a starting tree instead of a random starting tree for the BEAST analysis to avoid potential failure of the BEAST run. The weight of the four parameters: Birth Death Subtree Slide, Narrow, Wide and Wilson-Balding were removed to fix the tree topology in the BEAST run. The BEAST analysis was run twice with 20,000,000 generations of MCMC, and parameter sampling every 1,000 generations. The MCMC log files were examined in Tracer v1.7.2 (Rambaut et al., 2018) to assess convergence and calculate effective sample sizes (ESS) of various parameters. Tree files from the two runs were combined with LogCombiner v2.6.6 with 20 % Burn-in (Drummond and Rambaut, 2007). TreeAnnotator v2.6.6 (Drummond and Rambaut, 2007) was used to make a maximum clade-credibility (MCC) tree with means and 95 % highest posterior density of divergence times. FigTree v1.4.4 (<https://tree.bio.ed.ac.uk/software/figtree/>) was used to view and edit the resulting trees.

Six calibration points were selected as priors for divergence time estimation. The fossil calibration nodes were used for setting:

1. A hard minimum age, 125 Ma and 95 % soft maximum age, 150.94 Ma (mean, 1.9; standard deviation, 0.8; offset 125) were applied to the ancestral node of the Denticipitidae and the rest families. Fossils used for calibration: Min. 125 Ma †*Pseudoellimma gallae* De Figueiredo, 2009 from Atol Quarry, São Miguel dos Campos, Alagoas, Brazil (De Figueiredo, 2009a); soft max 150.94 Ma †*Tischlingerichthys vohli* Arratia, 1997 from Mühlheim, Bavaria, Germany (Arratia, 1997, 2001; Benton et al., 2015).

2. A hard minimum age, 5.333 Ma and 95 % soft maximum age, 125 Ma (mean, 3.47; standard deviation, 0.8; offset, 5.333) were applied to the ancestral node of *Spratelloides gracilis* and *Spratelloides delicatulus*. Fossils used for calibration: Min 5.333 Ma, *Spratelloides gracilis* from Oran, Algeria (Arambourg, 1927; Sorbini, 1987); soft max 125 Ma †*Pseudoellimma gallae* De Figueiredo 2009.

3. A hard minimum age, 54.05 Ma and 95 % soft maximum age, 125 Ma (mean, 2.94; standard deviation, 0.8; offset 54.05) were applied to the ancestral node of the Clupeidae, Pristigasteridae, Engraulidae, genus *Dussumieria* and *Etrumeus teres*. Fossils used for calibration: Min 54.05 Ma †*Clupeopsis straeleni* Casier, 1946 from the Dubois clay pit, Chièvres, Belgium (Capobianco et al. 2020); soft max: 125 Ma †*Pseudoellimma gallae* De Figueiredo, 2009.

4. A hard minimum age, 48.5 Ma and 95 % soft maximum age, 54.05 Ma (mean, 1.7; standard deviation, 0.8; offset 48.5) were applied to the ancestral node of all taxa in the Engraulidae group. Fossils used for calibration: Min 48.5 Ma †*Eoengraulis fasoloi* Marramà and Carnevale, 2016 from the site Pesciara Cave, Monte Bolca locality, Italy (Marramà and Carnevale, 2016); soft max: 54.05 Ma †*Clupeopsis straeleni* Casier, 1946.

5. A hard minimum age, 5.333 Ma and 95 % soft maximum age, 54.05 Ma (mean, 2.57; standard deviation, 0.8; offset 5.333) were applied to the ancestral node of *Etrumeus teres* and all taxa in the

Clupeinae group. Fossils used for calibration: Min 5.333 Ma †*Etrumeus boulei* Arambourg, 1925 from Oran, Algeria (Arambourg, 1927; Grande, 1985); soft max: 54.05 Ma †*Clupeopsis straeleni* Casier, 1946.

6. A hard minimum age, 27.82 Ma and 95 % soft maximum age, 54.05 Ma (mean, 1.95; standard deviation, 0.8; offset, 27.82) were applied to the ancestral node of Alosinae and Dorosomatinae. Fossils used for calibration: Min 27.82 Ma †*Alosa aralensis* Khisarova, 1977 from the locality Sarybaur in the Northern Aral Sea, Agyspe, Kazakhstan (Khisarova, 1977); Soft max: 54.05 Ma †*Clupeopsis straeleni* Casier, 1946.

### 3. Results

#### 3.1. Sequencing results and assembly

Illumina sequencing generated 798,425,826 raw reads, i.e. an average of 11,406,083 reads per sample. After parsing reads, adapter trimming and excluding low quality reads, an average of 11,402,939 reads per sample were available for further processing. The smallest number of reads per sample was 1,180,452 (CL141, *Ictalurus punctatus*) and the largest was 78,276,866 (CL531, *Laeviscutella dekimpei*). In total 4,108 (92.6 %) loci were assembled, an average of 1,691 loci per sample. Samples with less than 900 loci (21 %) assembled were removed, except for the sample of *Dussumieria elopsoides* due to the importance of this sample for testing the hypothesis of monophyletic dussumieriids. The final dataset comprised 72 taxa in total after removing redundant samples of the same species except that two samples of *Sardinella longiceps* were kept. The average alignment length per locus was 253 bp, average GC content 55 %, average missing data in alignment sequence 3.8 %, and average pairwise distance 0.14 in 4,053 genes. All details about the raw reads are supplemented in Appendix A, Table S2 and Fig. S3.

#### 3.2. Phylogenetic relationships

Dataset 1 comprised 1,165 loci with a total alignment length of 206,130 bp. The maximum likelihood tree built with IQTREE based on dataset 1 is shown in Fig. 2. All internal nodes had a bootstrap support > 95 % except for a node in the genus *Sardinella* (78 %) and a node in *Anchoa* (93 %) (Fig. 2). The RAxML tree had the same topology with only small difference of bootstrap values on some nodes (Appendix A, Fig. S4). The Clupei are monophyletic, with Denticipitidae as the sister-group to all other Clupei. The dussumieriids, i.e. Dussumieriinae sensu Grande (1985), are not monophyletic, as *Jenkinsia* and *Spratelloides* are sister to other clupeoids, separated from *Etrumeus* and *Dussumieria*, which were not grouped together either (Fig. 2).

To test whether the basal placement of spratelloidines was an artifact due to missing data, we reconstructed the tree with 417 loci from dataset 1 filtered for missing data in the respective species. The results still supported the early branching of the spratelloidines within the Clupeoidei clade (Appendix A, Fig. S5). We also filtered out the loci that had more than 10 % difference in GC content between the spratelloidines and the other clupeoids. The phylogenetic analysis of these remaining 465 loci still supported the early splitting of the spratelloidines from the other clupeoids (Appendix A, Fig. S6). There were 438 loci left when we included only loci that had less than 10 % missing taxa. The resulting tree showed a similar topology compared to the tree built on 1,165 loci, except some divergent interrelationships within the Dorosomatinae (Appendix A, Fig. S7).

Pristigasterids are monophyletic, but within this taxon there is no support for monophyletic Pelloninae. *Chirocentrus* is close to *Dussumieria* and not to engraulids. The miniaturized *Sundasalanx* is placed close to the ehiravine *Clupeonella*, however, with a relatively deep split. The genus *Clupea*, is not part of the diverse crown-group, but part of a clade containing additionally *Sprattus* and *Etrumeus*. Within the crown group, clades are found which correspond to the subfamilies and tribes reported by Sato (1994), i.e. Dorosomatinae, Pellonulinae, Escualosini,

Sardinellini and Alosini. The genus *Sardinella*, however, apparently is not monophyletic.

#### 3.3. Divergence time analyses

Two independent BEAST v2.6.6 runs of the 100 clock-likeness genes reached stationarity after 2 million generations. Effective sample sizes for all parameters were greater than 200. The mean divergence time for each node and its 95 % higher posterior densities (HPDs) intervals are shown in Fig. 3. The origin of crown Clupei is estimated to be in the Early Cretaceous (130.82 Ma, CI: 125.5–138.9 Ma). The Spratelloididae (91.4 Ma, CI: 76.09–107.26 Ma) and possibly the Engraulidae (71.58 Ma, CI: 62.97–80.72 Ma) diverged in the Cretaceous, but the rest of the families diverged very close or after the end of the Cretaceous (66 Ma). The crown groups of the Engraulidae (50.28 Ma, CI: 48.64–53.11 Ma), Ehiravidae (45.87 Ma, CI: 37.86–53.87 Ma), as well as split between Alosidae and Dorosomatidae (52.92 Ma, CI: 44.77–60.95 Ma), appeared in the Paleogene to Eocene. On the contrary, crown Pristigasteridae (26.91, CI: 20.47–34.17), Dussumieriidae (11.94, CI: 6.67–17.64) and Alosidae (19.7 Ma, CI: 13.44–26.75 Ma) likely appeared during the Oligocene to Miocene.

### 4. Discussion

#### 4.1. Monophyly of the Clupeiformes and phylogenetic position of *Denticeps*

The extant *Denticeps* occurs in coastal freshwaters in the area of and around the Niger Delta in West Africa, predominantly from southeast Benin to northwest Cameroon while the fossil †*Paleodenticeps tangerikae* was recorded from the Eocene in East Africa (Clausen, 1959; Greenwood and Chandler, 1960). Prominent denticles on roofing bones of the skull and a complete lateral line are the characters that distinguish *Denticeps* from all other recent herring-like fishes (Clausen, 1959). Based on several additional morphological characters *Denticeps* has been unambiguously considered as the sister taxon to all other Clupeiformes (Grande, 1985; Greenwood, 1968b), which was also supported by a few molecular studies (e.g., Lavoué et al. 2007). Recently, Bloom and Egan (2018) placed *Denticeps clupeoides* outside the Clupeiformes and recovered it as sister to all other otomorphs based on recombination activating genes *rag1* and *rag2*. Bloom and Egan (2018) estimated *Denticeps* even more distant to other clupeiforms. In the present study, the phylogeny of Clupeiformes was reconstructed using over a thousand loci and 66 taxa from 46 genera with five outgroup species, including Lepisosteiformes, Cypriniformes, Gymnotiformes, Siluriformes, and Characiformes exceeding all hitherto analyzed molecular datasets in the number of loci analyzed. *Denticeps clupeoides*, however, has been retained unanimously as sister group to clupeoids. Based on the thousands of exons analyzed herein, our study reassures the monophyly of Clupeiformes (Fig. 2).

#### 4.2. Monophyly and phylogenetic position of *Dussumieriidae sensu lato*

Dussumieriidae (round herring), so far containing the genera *Dussumieria*, *Spratelloides*, *Jenkinsia*, and *Etrumeus*, were morphologically characterized by a 'W' shaped unkeeled pelvic scute while other scutes are absent (Nelson, 1970a; Whitehead, 1972). Based on such morphological similarities round herrings assigned to Spratelloidinae and Dussumieriinae were classified into the family Dussumieriidae. Yet, several molecular studies rejected this hypothesis and questioned the monophyly of Dussumieriidae. Li (2007) found that six nuclear loci highly support dussumieriids as the basal group to the rest of clupeoids with a bootstrap value of 99 % and a Bayesian posterior probability of 1.0. Lavoué et al. (2014) discussed that the phylogenetic position of *Dussumieria* and the non-monophyly of the Dussumieriidae have some consequences for the hypothesis of an early evolution and the biogeography

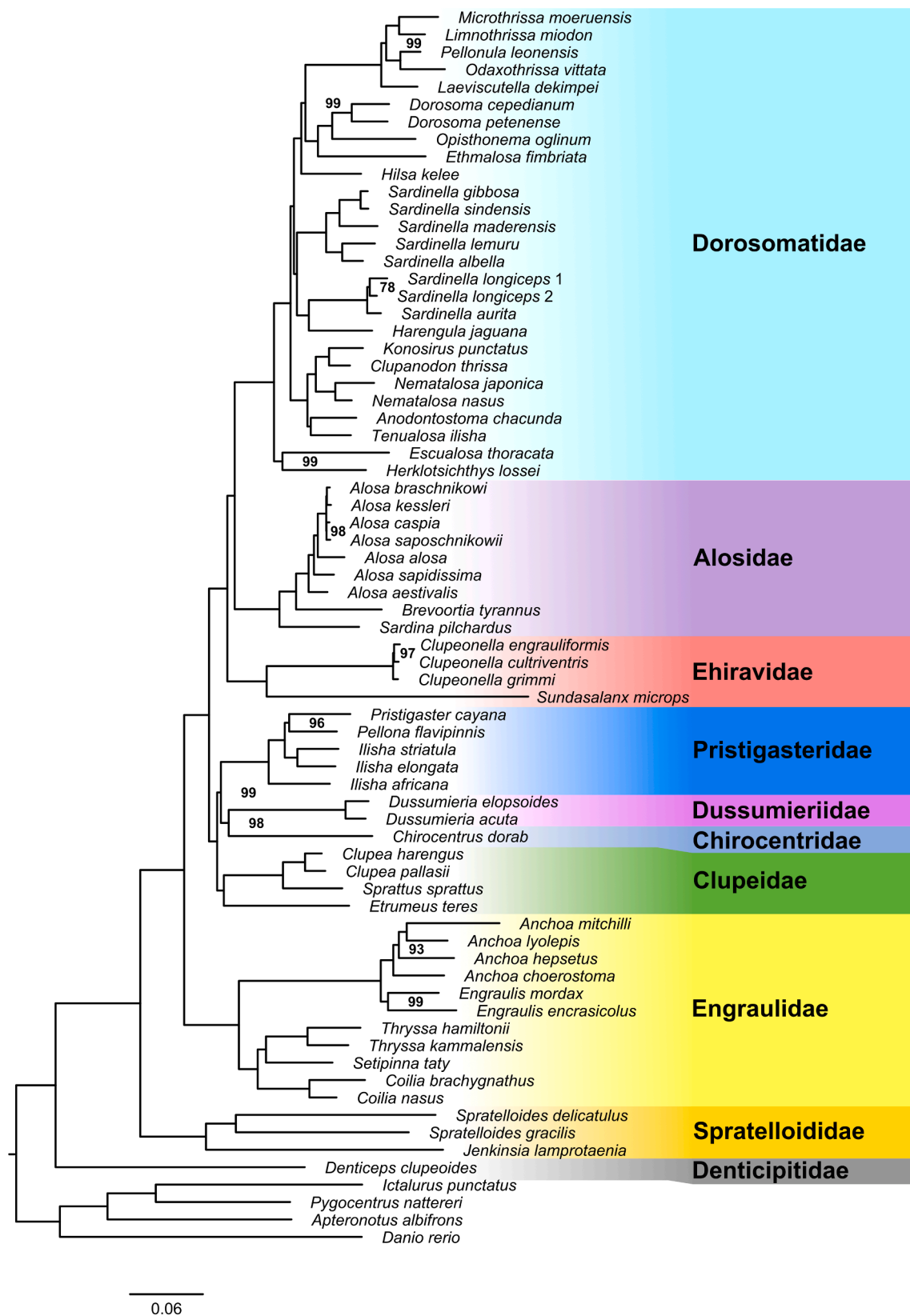
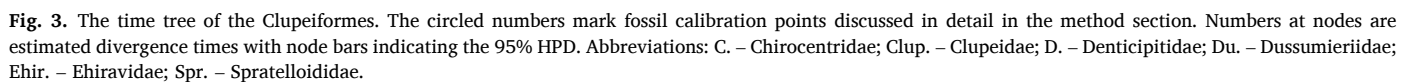


**Fig. 2.** Maximum likelihood tree of the Clupeiformes based on 1,165 loci reconstructed using IQtree with best partition scheme from PartitionFinder2. Only bootstrap supports not equal to 100% are shown. Tree was rooted with *Lepisosteus oculatus* (not shown).



of the Clupeoidei.

It is traditionally considered that the phylogenetic position of the family Dussumieriidae within the Clupeoidei is important to understand the evolution of morphology, because these fishes (round herrings) have a particular morphology, such as the presence of a W-shaped pelvic scute along with the absence or quasi absence of other abdominal scutes (Whitehead, 1962, 1963). Nelson (1967) and others considered the Dussumieriidae to be a “primitive” group within the large family Clupeidae and therefore suggested that the absence of an abdominal scute is the ancestral condition within the Clupeoidei. On the other hand, Grande (1985) considered the Dussumieriidae to be a derived group because of the absence of abdominal scutes, which he interpreted as a derived character. Lavoué et al. (2014) reconstructed the evolution of the scute morphology using a previous mitogenomic phylogenetic tree and they concluded that the most recent common ancestor of the Clupeoidei had a complete series of well-developed abdominal scutes and that the diminution in the number of scutes (along with the simplification of their morphology) evolved several times independently in the Clupeoidei.

Bloom and Lovejoy (2014) proposed a new position of Spratelloidinae (*Spratelloides* + *Jenkinsia*) separated from Dussumieriinae based on molecular phylogenetic analyses. Similarly, our study shows that both spratelloid species form a monophyletic group but phylogenetically rather distant to other dussumieriids (Fig. 2). The filtered datasets addressing the potential issue of missing data and biased GC content all confirmed the polyphyletic status of Dussumieriidae and early branching of Spratelloidinae from the other clupeoids. Hence, there is strong evidence for this result (Appendix A, Fig. S3 and S4). Despite of having similar molecular support, Lavoué et al. (2014), retained the family Dussumieriidae, but marked it with a question mark and the remark that it was placed temporarily on its initial position in absence of alternative hypothesis. In the present study, *Etrumeus teres* clustered with *Clupea* and *Sprattus* and not with *Dussumieria elopoides* and *D. acuta*. Considering these two conflicting hypotheses and in the light of our results, we suggest a specific review of both the morphological and molecular perspective of *Etrumeus* in the future.

#### 4.3. The phylogenetic position of *Sundasalanx*

The phylogenetic position of *Sundasalanx* (Sundasalangidae) is difficult to assess using morphology, as the species in the genus are highly pedomorphic (Roberts, 1981, 1984). All seven species of this genus are freshwater dwellers found in South-east Asia (Indonesia, Thailand, and Laos). The genus was initially described by Roberts (1981, 1984) based on its minute, transparent and pedomorphic feature and placed close to Salangidae within Osmeriformes. Later, Siebert (1997) identified *Sundasalanx* as clupeid based on non-pedomorphic features like an ossified prootic bulla and the *recessus lateralis* and further suggested *Sundasalanx* as a possible sister-group of a dussumieriid, namely the Caribbean endemic genus *Jenkinsia*. Britz and Kottelat (1999) reviewed the conclusion by Siebert's (1997) tentative placement of *Sundasalanx* where they found *Sundasalanx* shares synapomorphies with Clupeidae and advised for an extensive phylogenetic study. In subsequent molecular studies by Ishiguro et al. (2005) and Lavoué et al. (2007) *Sundasalanx* was confirmed as a member of the Clupeoidei but these studies also rejected a close relationship with *Jenkinsia*. Moreover, Li and Ortí (2007) placed *Sundasalanx* within “Clupeidae I” based on 12 s and 16 s ribosomal genes and suggested a further review to estimate the exact placement of *Sundasalanx*. A later study by Bloom and Lovejoy (2014), analyzed rag1, rag2, cytb and 16 s and identified *Sundasalanx mekongensis* as a member of the Ehiravini clade. Also Lavoué et al. (2013) found *Sundasalanx* sp. to be the sister-group of Ehiravinae and name these as “lineage 3” in their phylogenetic tree. Our results confirmed that *Sundasalanx* is a member of the Ehiravini group (Fig. 2; Fig. S4-5; Fig. S7).

There has been some controversy about the taxonomic status of the

genus *Clupeonella*. Nelson et al. (2016) kept this genus as a member of the subfamily Clupeinae, while it has been considered as a member of the subfamily Ehiravinae based on molecular results (Lavoué et al., 2017). The latter findings correspond to our results. However, the split between the ehiravini species of the genus *Clupeonella* with *Sundasalanx* is quite deep. A close relationship of *Clupeonella* with *Jenkinsia* can be discarded. Based on our findings aligning partially with previous studies, we erect the subfamily Ehiravinae to family status, i.e. **Ehiravidae**.

#### 4.4. Interrelationships among the major families of the Clupeoidei and resolving the Clupeidae, the ‘group of convenience’

Several subfamilies and/or tribes have been subsumed under the Clupeidae (*sensu lato*) in previous works. However, robust hypothesis of the relationships of these clupeid taxa were never defined. The amount of morphological diversity made it also difficult to define synapomorphies for the family. Recognizing the difficulties on defining reliable characters to distinguish clupeid species from other families, initially Nelson (1970b) excluded Pristigasterinae from Clupeidae. After dislocating Pristigasterinae Nelson (1970b) acknowledged that prevailing synapomorphies such as tendencies of missing teeth and proliferation of gill rakers did not provide strong support to define clupeids or the relationships of the subfamilies in it. Later, Grande (1985) introduced a single diagnostic character for the Clupeidae “the presence of two long, rod-like postcleithra” – a character that was proposed by Nelson (1970b) although this character is absent in several dussumieriids like the genera *Dussumieria*, *Jenkinsia* and *Spratelloides*. Grande (1985) was able to separate Pelloninae from Dussumieriinae but did not provided any diagnostic character for Alosinae, Dorosomatinae and Clupeinae. As a result, he hypothesized the Clupeidae as a “group of convenience” and explained that “the greatest remaining problem in Clupeid systematics is discovering how the members of the Dorosomatinae, Clupeinae and Alosinae are interrelated within the Clupeoidei” (Grande 1985, page 281). Moreover, several molecular studies on Clupeiformes did not agree on the phylogenetic position of the subfamilies of Clupeidae and therefore there is no molecular support for the monophyly of Clupeidae until now. The results recovered by Li and Ortí (2007) showed that *Chirocentrus* was nested within Clupeidae and as a result rejecting the monophyletic status of the Clupeidae. That study partitioned the Clupeidae into Clupeidae-I (Dorosomatinae + Alosinae including *Sundasalanx*) and Clupeidae-II (Dussumieriinae + Spratelloidinae + Clupeinae including *Chirocentrus*). In the current study using more than one thousand nuclear loci, we also found that the taxa formerly subsumed under Clupeidae are no longer existent as a monophyletic family but are segregated in several new clades. As a consequence we rank three monophyletic clades to family level, i.e. the Ehiravidae (see also above), the Alosidae, and the Dorosomatidae, with the latter two being retrieved as sister-groups. Finally, our analysis revealed that the former Clupeinae is the only remaining subfamily under Clupeidae which therefore resolves the “taxonomic basket” of the Clupeidae from “group of convenience” to monophylum based on molecular evidence.

Future studies should be carried out with a focus on resolving tribes within the diverse family Dorosomatidae using a denser taxon sampling. Misidentification of congener species in some genera such as the round-body *Sardinella* due to their overlapped similarities is one of serious issues in their taxonomy. Stern et al. (2018) revised the taxonomy of *Sardinella*, based on COI gene sequences and morphometric characters, suggesting the possible existence of only a single cosmopolitan species with two ecophenotypic variations. Our results show that there are at least two deeply diverged clades of *Sardinella*, with discernible species in each clade. The genus *Sardinella* is not monophyletic with *Harengula jaguana* grouped with the *Sardinella aurita* clade. This indicates that there may be a high cryptic diversity exists in *Sardinella* and a global taxon sampling is needed to test their taxonomy hypotheses using the gene-capture method.

#### 4.5. The estimated divergence time of major lineages of the Clupeiformes

The oldest †Ellimmichthyiformes, the taxon widely regarded as sister group to the Clupeiformes (Grande, 1985; Nelson et al., 2016), originates from the Lower Cretaceous of Brazil (ca. 140–129 Ma; Silva Santos and Silva, 1985; De Figueiredo and Ribeiro, 2017; De Figueiredo and Gallo, 2021). The oldest fossil record of the Clupeiformes (also from Brazil) is slightly younger (ca. 129–101 Ma, Maisey, 1993; De Figueiredo, 2009a; Malabarba and Di Dario, 2017), and this is consistent with our estimated age of crown Clupeiformes (130.82 Ma, CI: 125.56–138.83 Ma).

Crown Clupeoidei are estimated to have diverged during the latest Early Cretaceous or the beginning of the Late Cretaceous (91.4 Ma, CI: 76.09–107.26 Ma). There is at least one fossil species that indicates a late Early Cretaceous age for crown Clupeoidei, namely †*Nolfia riachuelensis* De Figueiredo 2009b from the Albian (ca. 100.5–113 Ma) of Brazil (De Figueiredo 2009b). Seven additional clupeoid species from the Upper Cretaceous of Southern Italy have been tentatively attributed to lineages such as “Dussumieriinae”, “Alosinae” or “Ehiravini” (Taverne 2002, 2004, 2007a, 2007b, 2011, 2014). These species are, however, in need of revision due to the fragmentary nature of the material or the obsolescence of these taxa, at least as they were traditionally defined (see Lavoué et al., 2014; Marramà and Carnevale, 2015; Argyriou and Davesne 2021).

The bulk of clupeoid families and major lineages diverged in the Late Cretaceous and early Paleogene, a period during which most major groups that dominate the actinopterygian diversity of today appeared (Near et al., 2012). The extinction event at the end of the Cretaceous may have created a vacancy of ecological niches previously held by non-clupeoid competitors, which could have resulted in new ecological opportunities for clupeoid fishes to diversify via adaptive radiation. Such a scenario is consistent with what has been observed in other teleost lineages as well (Alfaro et al., 2018; Ribeiro et al., 2018).

A common pattern throughout our divergence time estimates are the very large gaps between the estimated dates of divergence of a group and the first fossil representative of the same group (Fig. 3). This phenomenon is known as “ghost lineages” (Norell et al., 1993). For most family-level and other higher level major lineages in our tree, these gaps last from several million years up to several tens of millions of years. The oldest lineages in this estimate, namely the Denticipitidae, Spratelloididae and Engraulidae represent families whose origins are estimated in the Cretaceous, however, their oldest fossils are all documented from the Eocene, i.e., are about several millions of years younger than their estimated time of origin. Denticipitidae diverged between ca. 126–139 Ma (CI), but their first fossil representative is †*Palaeodenticiceps* from 45 to 46 million year old strata in Tanzania, East Africa (Greenwood and Chandler, 1960; Harrison et al., 2001). Spratelloididae diverged between 76 and 107 Ma (CI), whereas the first putative member of this group, †*Trollichthys bolcensis* from Monte Bolca, Italy is dated to ca. 48.5 Ma (Friedman and Carnevale, 2018; Marramà and Carnevale, 2015). A clear gap, albeit not so large, is also seen for the Engraulidae, whose divergence has been dated by 63–81 Ma, while their first fossil member, †*Clupeopsis straeleni* from Chièvres, Belgium, is about 54 Ma old (Capobianco et al., 2020).

Fossils of the Clupeiformes are more abundant in the Oligocene and Miocene, but many fossil species are not well preserved and their relationships to modern lineages are currently poorly known. Therefore, the evolutionary history of most lineages during the Oligocene and Miocene remains to be explored. However, based on our divergence time estimates, we can comment on the evolution and paleobiogeography of certain members of the Dorosomatidae, the second largest family of Clupeiformes after the Engraulidae. The family Dorosomatidae comprises, among other species, the lineage of the so-called West African freshwater herrings. These are represented in this study by the genera *Limnothrissa*, *Microthrissa*, *Pellonula*, *Odaxothrissa* and *Laeviscutella*. According to the molecular study of Wilson et al. (2008), this clade

transitioned from marine to freshwater ca. 37 Ma ago (CI: 25–53.3 Ma) and diversified in West Africa, then dispersed eastwards, with the Lake Tanganyika endemics *Stolothrissa tanganyicae* and *Limnothrissa miodon* splitting at 7.6 Ma (CI: 2.1–15.9 Ma). Because Lake Tanganyika is known to have formed between ca. 9–12 Ma (Cohen et al., 1993), Wilson et al. (2008) suggested that these freshwater herrings evolved *in situ*. Based on the upper 95 % CI of their estimate, however, it remains possible that these genera split before the formation of the lake and colonized it later, a scenario that also has been proposed for certain lineages of cichlid fishes in Lake Tanganyika (Altner et al., 2020; Altner et al., 2017). Our results estimate a younger divergence of the clade of the West African freshwater herrings, compared to Wilson et al.’s estimate (2008), that is the transition from marine to freshwater (29.04 Ma, CI: 24.14–34.13 Ma). Additionally, our estimates favor the scenario of the *in situ* evolution of the Tanganyika endemic *Limnothrissa*, as indicated by the age of its divergence from *Microthrissa moeruensis*, a native of the nearby Lake Mweru (5.75 Ma, CI: 3.57–8.04 Ma).

Another group which has attracted a lot of research interest with regard to its ecology and evolution and has a good fossil record is the family Alosidae (e.g., Bagils, 2001; Bagliniere and Élie, 2000; Ganias, 2014; Pozzobon et al., 2021). *Alosa* is suggested to have originated in the NW Atlantic, from where it later could have dispersed eastwards to the Mediterranean and eventually entered and diversified in the Black and Caspian Seas (Bloom and Lovejoy, 2014; Lavoué et al., 2013; Vernygora et al., 2018). The split between *Alosa* and its sister taxon *Brevoortia* has been dated to a period from the lower Oligocene to the upper Miocene, e.g., at 20.1 Ma (CI: ca. 17–30 Ma) in Vernygora et al. (2018) and 15.05 Ma (CI: 9.74–19.16 Ma) in Pozzobon et al. (2021). Vernygora et al. (2018) further estimated an upper Miocene to lower Pliocene age for the eastward expansion of *Alosa* in the Mediterranean and Ponto-Caspian regions. Our study supports a middle–upper Miocene age for the split between *Alosa–Brevoortia* (10.55 Ma, CI: 7.06–14.22 Ma). Accordingly, *Alosa* may have entered the Mediterranean not earlier than in the lower Pliocene (2.85 Ma, CI: 1.98–3.75 Ma) and the Black and Caspian Seas in the upper Pliocene to Pleistocene (2.21 Ma, CI: 1.49–2.97 Ma). However, several fossil species of *Alosa* have been described from the Oligocene and lower Miocene of the (peri-) Mediterranean region (Arambourg, 1927; Khisarova, 1977; Rückert-Ülkümen, 1965). This implies that either the systematic affinities of these fossil taxa need to be reconsidered, as already pointed out by Grande (1985), or that the split between *Alosa* and its sister taxon *Brevoortia* is older than currently estimated. A systematic revision of the clupeiform fossil record is necessary to allow us tracing the evolutionary history of these fishes in greater detail and also to further refine the dates estimated here.

#### 4.6. The evolution of decisive morphological characters of Clupeiformes

Clupeiformes are characterized by a high degree of homoplasy with regards to morphology. In the following we select three examples of morphological characters for discussion that were formerly regarded as synapomorphies putatively supporting major clades within Clupeiformes but are now shown to be homoplasies instead. Furthermore, we suggest ways in which the environment or biomechanical constraints, rather than phylogenetic conservatism, may have played a key role in the appearance or loss of these characters.

A striking apomorphy for Clupei is the presence of keeled scales along the ventral midline, the so-called scutes (Grande, 1982, 1985). The presence of a w-shaped pelvic scute and the loss of remaining ventral scutes is what defines the round herrings, *sensu* Grande (1985), i.e. the genera *Spratelloides*, *Jenkinsia*, *Dussumieria* and *Etrumeus*, as one of the easiest recognizable groups based on morphology (Grande, 1985; Whitehead, 1985; Nelson et al., 2016). In earlier work, the reduction of most of these scutes had led Whitehead (1963) to include in the dussumieriids, not only the round herrings (i.e. *Spratelloides*, *Jenkinsia*, *Dussumieria* and *Etrumeus*), but also the miniaturized clade of ehiravinids. Later Whitehead (1972) placed this latter group to the family

Pellonulinae, indicating without explicitly stating that the abdominal scute reduction must have happened convergently.

Lavoué et al. (2014) showed that reduction and simplification of ventral scutes happened several times within clupeoids. The primitive state seems to be a complete series of scutes as seen in *Denticeps*, and not the presence of only few or even a single scute as suggested by Whitehead (1963). According to the phylogeny presented herein, a partial reduction of pelvic scutes happened at least six times in the Clupeoidei: at the bases of Spratelloididae, Engraulinae, Dussumieriidae + *Chirocentrus*, and Ehiravidae, as well as in *Etrumeus*, and within the Pellonulini (part of Dorosomatidae). Further scute reduction likely happened within smaller clades or even within genera. For some, but not all those clades, miniaturization could provide an explanation for parallel loss of ventral scutes. Nevertheless, the w-shaped pelvic scute is an uncommon character that was supposed to unite the dussumieriids (sensu Grande, 1985), though it is apparent now that this is a case of homoplasy. This implies a mechanical constrain of the pelvic scute, so that this scute remains the last to be reduced and at the same time shows sometimes a morphology differing from pre- and post-pelvic scutes. This context, however, remains to be studied in detail in future.

The Pristigasteridae have a relatively low ratio of rib pairs compared to the total number of vertebrae, a character which they share with *Denticeps*, Engraulidae and †Elliimichthyiformes (Grande, 1985). The Pristigasteridae were therefore considered to be a more “primitive” group, i.e., not a part of the “Clupeoidea” which included “Clupeidae” + Chirocentridae (Di Dario, 2009; Grande, 1985). As, however, Pristigasteridae are deeply nested within what was formerly called “Clupeidae” and as *Spratelloides* + *Jenkinsia* is sister to the remaining Clupeoidei (Fig. 2), it appears that the evolution of this character has been much more complex than previously assumed. A low ribs-to-vertebrae ratio could be related to a reduced need for abdominal space, which accommodates the gut, and therefore reference to diet (Burns, 2021; Davis et al., 2013; Kapoor et al., 1976). Nevertheless, there does not seem to be an obvious connection between the trophic level (see Whitehead, 1985) and the ribs-to-vertebrae ratio in clupeiforms. An alternative explanation might be the swimming abilities, as we observe in Grande (1985, tables 1–10) that taxa with a low rib-to-vertebrae ratio also generally possess more anal fin rays.

Previous studies and classification schemes that were based on morphology have found it difficult to precisely define “Alosinae” relative to “Clupeinae” and “Dorosomatinae” (Grande, 1985) and have prioritized some characters that are now known to be homoplastic (Fig. 2), such as the structure of the gill arches (Nelson, 1967, 1970b) or the presence of a median notch in the upper jaw (Nelson et al., 2016; Regan, 1917; Whitehead, 1985). These characters may be related to feeding, particularly microphagy (Nelson, 1967), and the mechanics of the jaw occlusion. There is nonetheless a character that unequivocally unites the Alosidae genera as defined here, namely *Alosa*, *Brevoortia*, *Sardina* and *Sardinops*. This character is the presence of a striated operculum, and it is not found in any other clupeiform taxa (Vernygora, 2020). The branching sensory canals of the lateral line on the opercle may follow the striations on the opercle in Alosidae (personal observation, CK). Bone may form to enclose the sensory canals of the lateral line completely or partially in certain bones of the skull in teleosts (Lekander, 1949; Reno, 1969; Webb, 2013). Few studies have examined the distribution of cephalic sensory canals in Clupeiformes (Di Dario and de Pinna, 2006; Stephens, 1985; Sáez and Pequeño, 2018) and at present it remains unknown if there is a causal or even functional relationship between striation and lateral line canal branching in Alosidae.

In summary, the functional role of several distinct morphological characters within the Clupeiformes is insufficiently known or completely unknown. Our phylogenetic analyses, however, allow us to identify which of these characters were independently derived in multiple lineages and which are homologous. A detailed review of such characters is necessary in order to understand the evolutionary history of Clupeiformes and it is currently under preparation by CK. The

phylogeny presented in this paper paves the way for future studies to explore the functional role of convergent morphological characters in this group and the mechanisms by which they arise (Wake et al., 2011).

#### 4.7. An updated classification of the Clupeiformes

Based on the strong evidence of the present study, we propose an updated classification of Clupeiformes consisting of ten families: Denticipitidae; Spratelloididae; Engraulidae (including Engraulinae + Coiliinae); Clupeidae; Chirocentridae; Dussumieriidae; Pristigasteridae; Ehiravidae; Alosidae; Dorosomatidae. The classification of recent herring like fishes fitting in the classification proposed by Wiley and Johnson (2010), updates the revised classification of Lavoué et al. (2014):

Subcohort **Clupei** (sensu Wiley and Johnson, 2010; corresponding to Clupeomorpha Greenwood et al., 1966).

##### Order Clupeiformes

###### Suborder Denticipitoidei

###### Family Denticipitidae Clausen 1959

Type genus: *Denticeps* Clausen 1959

###### *Denticeps*

###### Suborder Clupeoidei

###### Family Spratelloididae Jordan 1925

Type genus: *Spratelloides* Bleeker 1851

###### *Jenkinsia*, *Spratelloides*

###### Family Engraulidae Gill 1861

Type genus: *Engraulis* Cuvier 1816

###### Subfamily Coiliinae Bleeker 1872

Type genus: *Coilia* Gray 1830

###### *Coilia*, *Lycotrissa*, *Papuengraulis*, *Pseudosetipinna*, *Setipinna*, *Thyrssa*

###### Subfamily Engraulinae Gill 1861

Type genus: *Engraulis* Cuvier 1816

###### *Amazonsprattus*, *Anchoa*, *Anchovia*, *Anchoviella*, *Cetengraulis*, *Encrasicholina*,

###### *Engraulis*, *Jurengraulis*, *Lycengraulis*, *Pterengraulis*, *Stolephorus*

###### Family Clupeidae Cuvier 1816

Type genus: *Clupea* Linnaeus 1758

###### *Clupea*, *Ethmidium*, *Hyperlophus*, *Potamalosa*, *Ramnogaster*, *Sprattus*, *Strangomera* incertae sedis: *Etrumeus* Bleeker 1853

###### Family Dussumieriidae Gill 1816

Type genus: *Dussumieria* Valenciennes 1847

###### *Dussumieria*

###### Family Chirocentridae Bleeker 1872

Type genus: *Chirocentrus* Cuvier 1816

###### *Chirocentrus*

###### Family Pristigasteridae Bleeker 1872

Type genus: *Pristigaster* Cuvier 1816

###### *Chirocentrodon*, “*Ilisa*”, *Neopisthopterus*, *Odontognathus*, *Opisthopterus*, *Pellona*, *Pliosteostoma*, *Pristigaster*, *Raconda*

###### Family Ehiravidae Deraniyagala, 1929

Type genus: *Ehirava* Deraniyagala 1929

###### *Clupeichthys*, *Clupeoides*, *Clupeonella*, *Corica*, *Dayella*, *Ehirava*, *Gilchristella*, *Minyclupeoides*, *Sauvagella*, *Spratellomorpha*, *Sundasanax*

###### Family Alosidae Svetovidov 1953

Type genus: *Alosa* Linck 1790

###### *Alosa*, *Brevoortia*, *Sardina*, *Sardinops*

###### Family Dorosomatidae Bleeker, 1872

Type genus: *Dorosoma* Rafinesque 1820

###### *Amblygaster*, *Anodontostoma*, *Clupanodon*, *Congothrissa*, *Dorosoma*, *Escualosa*, *Ethmalosa*, *Gonialosa*, *Gudusia*, *Harengula*, *Herklotsichthys*, *Hilsa*, *Konosirus*, *Laeviscutella*, *Lile*, *Limnothrissa*, *Microthrissa*, *Nannothrissa*, *Nematalosa*, *Odaxothrissa*, *Opisthonema*, *Pellonula*, *Platanichthys*, *Potamothrissa*, *Rhinosardinia*, *Sardinella*, *Sierrathrissa*, *Stolothrissa*, *Tenuialosa*

## 5. Conclusions

We reconstructed a robust phylogeny of the Clupeiformes based on thousands of nuclear coding sequences obtained through target enrichment and curated for phylogenetic informativeness. We confirmed that the Clupeiformes is monophyletic, and *Denticeps* is sister to the Clupeoidei. Dussumieriids in their former understanding are polyphyletic, as Spratelloididae being sister to all other clupeoids. *Sundasanax* belongs to Ehiravidae, but with deep split to other ehiravine

species. A time tree of the Clupeiformes was calibrated with new fossil data and an updated classification of the Clupeiformes is provided.

## Author contributions

C Li, N Straube, HR Esmaeili and T Moritz conceived the research ideas. Q Wang, LP Dizaj, and N Straube collected the data, Q Wang, J Huang and LP Dizaj analyzed the data. C Kevrekidis and B Reichenbacher curated fossil data for divergence time estimations. C Kevrekidis, B Reichenbacher and T Moritz analyzed the results and discussed the evolution of morphological characters. Q Wang and KK Sarker wrote the first draft. All authors revised and approved the final version of the manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We thank S. Lavoué, E. Wiley, X. Gong, L. Ma, and J. Zhao for providing fish material for molecular work.

## Funding

This work was supported by the Science and Technology Commission of Shanghai Municipality (19410740500; 19050501900) to CL, the Deutsche Forschungsgemeinschaft (Grant No. RE 1113/24-1) to CK and BR, and VolkswagenStiftung within the focal point program "Forschung in Museen" (AZ-88 353) to TM and NS.

## Data accessibility

Research data associated with this study is available from the Mendeley Data platform (<https://data.mendeley.com/drafts/3g4rf47yfb>). Raw sequence reads are available from the NCBI Sequence Read Archive, under BioProject PRJNA823840.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2022.107590>.

## References

- Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E., Oliveros, C.H., Cerny, D., Near, T.J., 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nat. Ecol. Evol.* 2 (4), 688–696. <https://doi.org/10.1038/s41559-018-0494-6>.
- Altnier, M., Schliwien, U.K., Penk, S.B.R., Reichenbacher, B., 2017. *Tugenchromis pickfordi*, gen. et sp. nov., from the Upper Miocene—A Stem-Group Cichlid of the 'East African Radiation'. *Palaeodiversity* 37 (2). <https://doi.org/10.1080/02724634.2017.1297819>.
- Altnier, M., Ruthensteiner, B., Reichenbacher, B., 2020. New haplochromine cichlid from the upper Miocene (9–10 MYA) of Central Kenya. *BMC Evol. Biol.* 20 (1), 65. <https://doi.org/10.1186/s12862-020-01602-x>.
- Arambourg, C., 1927. Les Poissons fossiles d'Oran. Bastide-Jourdan, Algiers.
- Argyriou, T., Davesne, D., 2021. Offshore marine actinopterygian assemblages from the Maastrichtian-Paleogene of the Pindos Unit in Eurytania. Greece. *PeerJ* 9.
- Arratia, G., 1997. Basal teleosts and teleostean phylogeny. *Palaeo Ichthyologica*. 75–168.
- Arratia, G., 2001. The monophyly of Teleostei and stem-group teleosts. Consensus and disagreements. In: Arratia, G., Schultze, H.-P. (Eds.), *Mesozoic Fishes 2—Systematics and Fossil Record*. Verlag Dr. Friedrich Pfeil, Munich, pp. 265–334.
- Bagils, R.Z., 2001. Les *Alosa* fossiles et la notion de genre en systématique phylogénétique. *Bull. Fr. Pêche Piscic.* 1059–1066. <https://doi.org/10.1051/kmae:2001036>.
- Bagliniere, J.L., Élie, P., 2000. Les aloses (*Alosa alosa* et *Alosa fallax* spp.). *Écobiologie Et Variabilité Des Populations*. Louis-Jean, Paris.
- Benton, M.J., Donoghue, P.C.J., Asher, R.J., Friedman, M., Near, T.J., Vinther, J., 2015. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electronica* 18, 1–107. <https://doi.org/10.26879/424>.
- Bloom, D.D., Egan, J., 2018. Systematics of Clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. *Neotrop. Ichthyol.* 16 (3). <https://doi.org/10.1590/1982-0224-20180095>.
- Bloom, D.D., Lovejoy, N.R., 2014. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proc. R. Soc. B* 281 (1778), 20132081. <https://doi.org/10.1098/rspb.2013.2081>.
- Bloom, D.D., Burns, M., Schriever, T., 2018. Evolution of body size and trophic position in migratory fishes: A phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). *Biol. J. Linn. Soc. Lond.* 125 (2), 302–314. <https://doi.org/10.1093/BIOLINNEAN/BLY106>.
- Bloom, D.D., Lovejoy, N.R., 2012. Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *J. Evol. Biol.* 25 (4), 701–715. <https://doi.org/10.1111/j.1420-9101.2012.02464.x>.
- Bossert, S., Murray, E.A., Blaimer, B.B., Danforth, B.N., 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. *Mol. Phylogenet. Evol.* 111, 149–157. <https://doi.org/10.1016/j.ympev.2017.03.022>.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10 (4). <https://doi.org/10.1371/journal.pcbi.1003537>.
- Britz, R., Kottelat, M., 1999. *Sundasalanx mekongensis*, a new species of clupeiform from the Mekong basin (Teleostei: Sundasalangidae). *Ichthyol. Explor. Freshwaters* 10 (4), 337–344.
- Burns, M.D., 2021. Adaptation to herbivory and detritivory drives the convergent evolution of large abdominal cavities in a diverse freshwater fish radiation (Otophysi: Characiformes). *Evolution* 75 (3), 688–705. <https://doi.org/10.1111/evo.14178>.
- Capobianco, A., Beckett, H.T., Steurbaut, E., Gingerich, P.D., Carnevale, G., Friedman, M., 2020. Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. *R. Soc. Open Sci.* 7 (5). <https://doi.org/10.1098/rsos.192260>.
- Clausen, H.S., 1959. Denticipitidae, a new family of primitive isospondylous teleosts from West African fresh-water. *Videnskabelige Meddelelser fra Dansk Naturhistorisk Forening, Kjøbenhavn* 121, 141–151.
- Cohen, A.S., Bills, R., Cocquyt, C., Caljon, A.G., 1993. The impact of sediment pollution on biodiversity in Lake Tanganyika. *Conserv. Biol.* 7 (3), 667–677. <https://doi.org/10.1046/j.1523-1739.1993.07030667.x>.
- Cope, E.D., 1871. Contribution to the Ichthyology of the Lesser Antilles. *Trans. Am. Philos. Soc.* 14 (3), 445. <https://doi.org/10.2307/1005256>.
- Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., Glenn, T.C., 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett.* 8 (5), 783–786. <https://doi.org/10.1098/rsbl.2012.0331>.
- Davis, A.M., Unmack, P.J., Pusey, B.J., Pearson, R.G., Morgan, D.L., 2013. Ontogenetic development of intestinal length and relationships to diet in an Australasian fish family (Terapontidae). *BMC Evol. Biol.* 13, 1–16. <https://doi.org/10.1186/1471-2148-13-53>.
- De Figueiredo, F.J., 2009a. A new clupeiform fish from the Lower Cretaceous (Barremian) of Sergipe-Alagoas Basin, northeastern Brazil. *J. Vertebr. Paleontol.* 29 (4), 993–1005. <https://doi.org/10.1671/039.029.0402>.
- De Figueiredo, F.J., 2009b. A new marine clupeoid fish from the Lower Cretaceous of the Sergipe-Alagoas Basin, northeastern Brazil. *Zootaxa* 2164 (1), 21–32.
- De Figueiredo, F.J., Gallo, V., 2021. Revision of *Ellimmichthys longicostatus* (Clupeomorpha: *Ellimmichthyiformes*) from the Lower Cretaceous of Brazil with comments on the taxonomy of related species. *J. S. Am. Earth Sci.* 105.
- De Figueiredo, F.J., Ribeiro, D.R., 2017. Reassessment and Relationships of *Scutatuspinosus itapagipensis* (Teleostei, Clupeomorpha, *Ellimmichthyiformes*) from the Neocomian of Recôncavo Basin, Northeastern Brazil. *Anais Da Academia Brasileira De Ciencias* 89, 799–823.
- Di Dario, F., 2002. Evidence supporting a sister-group relationship between Clupeoidea and Engrauloidea (Clupeomorpha). *Copeia* 102 (2), 496–503. [https://doi.org/10.1643/0045-8511\(2002\)002\[0496:ESASGRJ2.0.CO;2](https://doi.org/10.1643/0045-8511(2002)002[0496:ESASGRJ2.0.CO;2).
- Di Dario, F., 2009. Chirocentrids as engrauloids: Evidence from suspensorium, branchial arches, and infraorbital bones (Clupeomorpha, Teleostei). *Zool. J. Linn. Soc.* 156 (2), 363–383. <https://doi.org/10.1111/j.1096-3642.2008.00472.x>.
- Di Dario, F., de Pinna, M.C.C., 2006. The supratemporal system and the pattern of ramification of cephalic sensory canals in *Denticeps clupeoides* (Denticipitoidei, Teleostei): additional evidence for monophyly of Clupeiformes and Clupeoidei. *Pap. Avulsos Zool.* 46 (10), 107–123. <https://doi.org/10.1590/S0031-10492006001000001>.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214. <https://doi.org/10.1186/1471-2148-7-214>.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26 (19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61 (5), 717–726. <https://doi.org/10.1093/sysbio/sys004>.
- FAO, 2020. *FAO Yearbook. Fishery and Aquaculture Statistics 2018/FAO annuaire. Statistiques des pêches et de l'aquaculture 2018/FAO anuario. Estadísticas de pesca y acuicultura 2018*. FAO, Rome.

- FAO, 2021. FAO Yearbook. Fishery and Aquaculture Statistics 2019/FAO annuaire. Statistiques des pêches et de l'aquaculture 2019/FAO anuario. Estadísticas de pesca y acuicultura 2019. FAO, Rome.
- Finsternberger, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., Roos, C., 2013. A mitogenomic phylogeny of living primates. *PLoS ONE* 8 (7). <https://doi.org/10.1371/journal.pone.0069504>.
- Fricke, R., Eshmeier, W.N., Fong, J.D., 2022. Eshmeier's catalog of fishes: genera/species by family/subfamily, in: Fricke, R., Eshmeier, W. N. & R. van der Laan (Eds.), Eshmeier's catalog of fishes: genera, species, references. <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp>. Electronic version accessed 6 June 2022.
- Friedman, M., Carnevale, G., 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. *J. Geol. Soc.* 175 (4), 569–579. <https://doi.org/10.1144/jgs2017-164>.
- Ganias, K., 2014. Biology and Ecology of Sardines and Anchovies. CRC Press, Boca Raton.
- Grande, L., 1982. A revision of the fossil genus *Diplomystus* : with comments on the interrelationships of clupeomorph fishes. *Am. Mus. Novit.* 2728, 1–34.
- Grande, L., 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bull. Am. Mus. Nat. Hist.* 181 (2), 231–372.
- Greenwood, P.H., 1968a. Notes on the visceral anatomy of *Denticipes clupeoides* Clausen, 1959, a West African clupeomorph fish. *Rev. Zool. Bot. Afr.* 77, 1–10.
- Greenwood, P.H., 1968b. The osteology and relationships of the Denticipitidae, a family of clupeomorph fishes. *Bull. Br. Mus. Nat. Hist. Zool.* 16, 215–273.
- Greenwood, P.H., Chandler, M.E.J., 1960. Fossil denticipitid fishes from east Africa. *Bull. Brit. Mus. (Nat. Hist.) Geol.* 5, 1–11.
- Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Myers, G.S., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bull. Am. Mus. Nat. Hist.* 131 (4), 339–455.
- Harrison, T., Msuya, C.P., Murray, A.M., Jacobs, B.F., Báez, A.M., Mundil, R., Ludwig, K. R., 2001. Paleontological investigations at the Eocene locality of Mahenge in north-central Tanzania, East Africa. In: Gunnell, G.F. (Ed.), *Eocene Biodiversity: Unusual Occurrences and Rarely Sampled Habitats*. Springer, US, Boston, pp. 39–74.
- Huang, J., Yuan, H., Li, C., 2021. Protocol for Cross-species Target-gene Enrichment. *Bio-protoc.* e1010606 <https://doi.org/10.21769/BioProtoc.1010606>.
- Hugall, A.F., O'Hara, T.D., Hunjan, S., Nilsen, R., Moussalli, A., 2016. An Exon-Capture System for the Entire Class Ophiuroidea. *Mol. Biol. Evol.* 33, 281–294. <https://doi.org/10.1093/molbev/msv216>.
- Ishiguro, N.B., Miya, M., Inoue, J., Nishida, M., 2005. *Sundasalang* (Sundasalangidae) is a progenetic clupeiform, not a closely-related group of salangids (Osmereformes): mitogenomic evidence. *J. Fish Biol.* 67 (2), 561–569. <https://doi.org/10.1111/j.0022-1112.2005.00746.x>.
- Jiang, J., Yuan, H., Zheng, X., Wang, Q., Kuang, T., Li, J., Liu, J., Song, S., Wang, W., Cheng, F., Li, H., Huang, J., Li, C., 2019. Gene markers for exon capture and phylogenomics in ray-finned fishes. *Ecol. Evol.* 9 (7), 3973–3983. <https://doi.org/10.1002/eece.3.5026>.
- Kapoor, B.G., Smit, H.J.M., Verighina, I.A., 1976. The Alimentary Canal and Digestion in Teleosts. *Adv. Mar. Biol.* 13, 109–239. [https://doi.org/10.1016/S0065-2881\(08\)60281-3](https://doi.org/10.1016/S0065-2881(08)60281-3).
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Khisarova, G.D., 1977. Novyy vid *Alosa* (Clupeidae) iz oligotsena (Clupeidae) iz oligotsena Severnogo Priaral'ya [A new species of *Alosa* (Clupeidae) from the Oligocene of the northern Peri-Aral area]. In: Kojamkulova, B.S., Tleuberdina, P.A., Savinof, P.F., Orlofskaja, E.R. (Eds.), *Mezokazyozyskaya Fauna I Flora Severo-Zapadnogo Kazakhstana, Materlaly Po Istorii Fauny I flory Kazahstana*. Nauka, Kazakh SSR, pp. 5–9.
- Kuang, T., Tornabene, L., Li, J., Jiang, J., Chakrabarty, P., Sparks, J.S., Naylor, G.J., Li, C., 2018. Phylogenomic analysis on the exceptionally diverse fish clade Gobioidae (Actinopterygii: Gobiiformes) and data-filtering based on molecular clocklikeness. *Mol. Phylogenet. Evol.* 128, 192–202. <https://doi.org/10.1016/j.ympev.2018.07.018>.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* 34 (3), 772–773. <https://doi.org/10.1093/molbev/msw260>.
- Lavoué, S., Miya, M., Saitoh, K., Ishiguro, N.B., Nishida, M., 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Mol. Phylogenet. Evol.* 43 (3), 1096–1105. <https://doi.org/10.1016/j.ympev.2006.09.018>.
- Lavoué, S., Miya, M., Musikasinthorn, P., Chen, W.J., Nishida, M., 2013. Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). *PLoS ONE* 8 (2). <https://doi.org/10.1371/journal.pone.0056485>.
- Lavoué, S., Konstantinidis, P., Chen, W.J., 2014. Progress in clupeiform systematics. In: Ganias, K. (Ed.), *Biology and ecology of sardines and anchovies*. CRC Press, pp. 3–42.
- Lavoué, S., Bertrand, J.A.M., Chen, W.J., Ho, H.C., Motomura, H., Sado, T., Miya, M., 2017. Phylogenetic position of the rainbow sardine *Dussumieria* (Dussumeriidae) and its bearing on the early evolution of the Clupeoidei. *Gene* 623, 41–47. <https://doi.org/10.1016/j.gene.2017.04.032>.
- Lekander, B., 1949. The sensory line system and the canal bones in the head of some Ostariophysi. *Acta Zool.* 30, 1–131.
- Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61 (5), 727–744. <https://doi.org/10.1093/sysbio/sys049>.
- Li, C., 2007. A Genome-scale approach to phylogeny of ray-finned fish (Actinopterygii) and molecular systematics of Clupeiformes. The University of Nebraska-Lincoln. Doctoral Thesis.
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J., 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques* 54 (6), 321–326. <https://doi.org/10.2144/000114039>.
- Li, C., Ortí, G., 2007. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 44, 386–398. <https://doi.org/10.1016/j.ympev.2006.10.030>.
- Maisey, J.G., 1993. A new clupeomorph fish from the Santana Formation (Albian) of NE Brazil. *Am. Mus. Novit.* no. 3076.
- Malabarba, M.C., Di Dario, F., 2017. A new predatory herring-like fish (Teleostei: Clupeiformes) from the early Cretaceous of Brazil, and implications for relationships in the Clupeoidei. *Zool. J. Linn. Soc.* 180 (1), 175–194.
- Marramà, G., Carnevale, G., 2015. Eocene Round Herring from Monte Bolca. Italy. *Acta Palaeontol. Pol.* 60 (3), 701–700. <https://doi.org/10.4202/app.00057.2014>.
- Marramà, G., Carnevale, G., 2016. An Eocene anchovy from Monte Bolca, Italy: The earliest known record for the family Engraulidae. *Geol. Mag.* 153, 84–94. <https://doi.org/10.1017/S0016756815000278>.
- Martin, M., 2011. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17, 10–12. <https://doi.org/10.14806/ebj.17.1.200>.
- Mayer, C., Sann, M., Donath, A., Meixner, M., Podsiadlowski, L., Peters, R.S., Petersen, M., Meusemann, K., Liere, K., Wagele, J.W., Misof, B., Bleidorn, C., Ohl, M., Niehuis, O., 2016. BaitFisher: A Software Package for Multispecies Target DNA Enrichment Probe Design. *Mol. Biol. Evol.* 33 (7), 1875–1886. <https://doi.org/10.1093/molbev/msw056>.
- McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* 22 (4), 746–754. <https://doi.org/10.1101/gr.125864.111>.
- Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010 (6), 1–10. <https://doi.org/10.1101/pdb.prot5448>.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37 (5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Miyashita, T., 2010. Unique occipital articulation with the first vertebra found in pristigasterids, chirocentrids, and clupeids (Teleostei: Clupeiformes: Clupeoidei). *Ichthyol. Res.* 57 (2), 121–132. <https://doi.org/10.1007/s10228-009-0132-z>.
- Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P. C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. U.S.A.* 109 (34), 13698–13703. <https://doi.org/10.1073/pnas.1206625109>.
- Nelson, G.J., 1967. Gill Arches of Teleostean Fishes of the Family Clupeidae. *Copeia* 1967 (2), 389–399. <https://doi.org/10.2307/1442129>.
- Nelson, G.J., 1970a. Dorsal Scutes in the Chinese Gizzard Shad *Clupanodon thrissa* (Linnaeus). *Japanese Journal of Ichthyology* 17, 131–134.
- Nelson, G.J., 1970b. The hyobranchial apparatus of teleostean fishes of the families Engraulidae and Chirocentridae. *Am. Mus. Novit.* 1–30.
- Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. *Fishes of the World*, fifth ed. John Wiley & Sons.
- Norell, M.A., Dodson, P., Gingerich, P.D., 1993. Tree-based approaches to understanding history: comments on ranks, rules and the quality of the fossil record. *Am. J. Sci.* 293-A, 407–417. <https://doi.org/10.2475/ajs.293.A.407>.
- Oliveira, M.C., Repetti, S.I., Iha, C., Jackson, C.J., Díaz-Tapia, P., Lubiana, K.M.F., Cassano, V., Costa, J.F., Cremen, M.C.M., Marcelino, V.R., Verbruggen, H., 2018. High-throughput sequencing for algal systematics. *Eur. J. Phycol.* 53, 256–272. <https://doi.org/10.1080/09670262.2018.1441446>.
- Pozzobon, A.P.B., Gonçalves, P.R., Anderson, J.D., Rocha, L.A., Diaz de Astarloa, J.M., Di Dario, F., 2021. Phylogenetic relationships, genetic diversity and biogeography of menhaden, genus *Brevoortia* (Clupeiformes, Clupeidae). *Mol. Phylogenet. Evol.* 160 <https://doi.org/10.1016/j.ympev.2021.107108>.
- Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526 (7574), 569–573. <https://doi.org/10.1038/nature15697>.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67 (5), 901–904. <https://doi.org/10.1093/sysbio/syy032>.
- Regan, C.T., 1909. XII.—The classification of Teleostean fishes. *Ann. Mag. Nat. Hist.* 3 (8), 75–86. <https://doi.org/10.1080/00222930908692548>.
- Regan, C.T., 1917. XXVI.—A revision of the Clupeoid fishes of the genera *Pomolobus*, *Brevoortia* and *Dorosoma*, and their Allies. *Ann. Mag. Nat. Hist.* 19, 297–316.
- Reno, H.W., 1969. Cephalic lateral-line systems of the cyprinid genus *Hybopsis*. *Copeia* 1969, 736–773.
- Ribeiro, E., Davis, A.M., Rivero-Vega, R.A., Ortí, G., Betancur, R.R., 2018. Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. *Proc. Biol. Sci.* 285 (1893), 20182010. <https://doi.org/10.1098/rspb.2018.2010>.
- Roberts, T.R., 1981. Sundasalangidae, a new family of minute freshwater salmoniform fishes from southeast Asia. *Proc. California Acad. Sci.* 42 (9), 295–302.
- Roberts, T.R., 1984. Skeletal anatomy and classification of the neotenic Asian salmoniform superfamily Salangoidea (icefishes or noodlefishes). *Proc. California Acad. Sci.* 43 (13), 179–220.
- Romiguier, J., Roux, C., 2017. Analytical Biases Associated with GC-Content in Molecular Evolution. *Front. Genet.* 8, 16. <https://doi.org/10.3389/fgene.2017.00016>.
- Rückert-Ulkümen, N., 1965. Fossile Fische aus dem Sarmat von Pınarhisar (Türkisch-Thrakien). *Senckenbergiana Lethaea* 46a, 315–361.

- Sáez, S., Pequeño, G., 2018. Gross morphology of the cephalic sensory canals of *Strangomera bentincki* (Norman, 1936) (Teleostei: Clupeidae). *Rev. Biol. Mar. Oceanogr.* 53, 1–5. [10.22370/rbmo.2018.53.0.1248](https://doi.org/10.22370/rbmo.2018.53.0.1248).
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>.
- Sato, Y., 1994. Cladistic-phenetic relationships and the evolutionary classification of the suborder Clupeoidei (Pisces, Teleostei). Tokyo Metropolitan University. Doctoral Thesis.
- Siebert, D.J., 1997. Notes on the anatomy and relationships of *Sundasalanx* Roberts (Teleostei, Clupeidae), with descriptions of four new species from Borneo. *Bull. Nat. Hist. Mus. Lond. (Zool.)* 63, 13–26.
- Silva Santos, R.D., Silva, Corrêa V., 1985. Contribuição ao conhecimento da paleoictiofauna do Cretáceo no Brasil. *Coletânea de Trabalhos Paleontológicos Série Geologia* 27 (2), 169–174.
- Simpson, J.T., Durbin, R., 2012. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* 22 (3), 549–556. <https://doi.org/10.1101/gr.126953.111>.
- Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* 6, 31. <https://doi.org/10.1186/1471-2105-6-31>.
- Sorbin, L., 1987. Biogeography and climatology of Pliocene and Messinian fossil fish of Eastern-Central Italy. *Boll. Mus. civ. stor. nat. Verona* 14, 1–85.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Stephens, R.R., 1985. The lateral line system of the gizzard shad, *Dorosoma cepedianum* Lesueur (Pisces : Clupeidae). *Copeia* 1985, 540–556. <https://doi.org/10.2307/1444742>.
- Stern, N., Douek, J., Goren, M., Rinkevich, B., 2018. With no gap to mind: a shallow genealogy within the world's most widespread small pelagic fish. *Ecography* 41, 491–504. <https://doi.org/10.1111/ecog.02755>.
- Straube, N., Li, C., Claes, J.M., Corrigan, S., Naylor, G.J., 2015. Molecular phylogeny of Squaliformes and first occurrence of bioluminescence in sharks. *BMC Evol. Biol.* 15, 162. <https://doi.org/10.1186/s12862-015-0446-6>.
- Straube, N., Li, C., Mertz, M., Yuan, H., Moritz, T., 2018. A phylogenomic approach to reconstruct interrelationships of main clupeocephalan lineages with a critical discussion of morphological apomorphies. *BMC Evol. Biol.* 18, 158. <https://doi.org/10.1186/s12862-018-1267-1>.
- Swofford, D., 2002. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.
- Taverne, L., 2002. Les poissons crétacés de Nardò. 12°. *Nardoclupea grandei* gen. et sp. nov. (Teleostei, Clupeiformes, Dussumieriinae). *Bollettino del Museo Civico di Storia Naturale di Verona* 26, 3–23.
- Taverne, L., 2004. Les poissons crétacés de Nardò. 18°. *Pugliacupea nolardi* gen. et sp. nov. (Teleostei, Clupeiformes, Clupeidae). *Bollettino del Museo Civico di Storia Naturale di Verona* 28, 17–28.
- Taverne, L., 2007a. Les poissons crétacés de Nardò. 25°. *Italoclupea nolfi* gen. et sp. nov. (Teleostei, Clupeiformes, Clupeidae). *Bollettino del Museo Civico di Storia Naturale di Verona* 31, 21–35.
- Taverne, L., 2007b. Les poissons crétacés de Nardò. 26°. Un second Dussumieriinae: *Portoselvaggioclupea whiteheadi* gen. et sp. nov. et complément à l'étude de *Pugliacupea nolardi* (Teleostei, Clupeidae). *Bollettino del Museo Civico di Storia Naturale di Verona* 31, 37–42.
- Taverne, L., 2011. Les poissons crétacés de Nardò. 33°. *Lecceclupea ehiravaensis* gen. et sp. nov. (Teleostei, Clupeidae). *Bollettino del Museo Civico di Storia Naturale di Verona* 35, 3–17.
- Taverne, L., 2014. Les poissons du Santonien (Crétacé supérieur) d'Apricena (Italie du Sud). 7 *Garganoclupea svetovidovi* gen. et sp. nov. et *Apricencalupea ridewoodi* gen. et sp. nov. (Teleostei, Clupeiformes). *Bollettino del Museo Civico di Storia Naturale di Verona, Geologia Paleontologia Preistoria* 38, 27–49.
- Teasdale, L.C., Kohler, F., Murray, K.D., O'Hara, T., Moussalli, A., 2016. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Mol. Ecol. Resour.* 16 (5), 1107–1123. <https://doi.org/10.1111/1755-0998.12552>.
- Vernygora, O.V., 2020. Systematics of Clupeomorpha (Osteichthyes:Teleostei) with methodological considerations for morphological phylogenetics. University of Alberta. Doctoral Thesis.
- Vernygora, O.V., Davis, C.S., Murray, A.M., Sperling, F.A.H., 2018. Delimitation of *Alosa* species (Teleostei: Clupeiformes) from the Sea of Azov: integrating morphological and molecular approaches. *J. Fish Biol.* 93 (6), 1216–1228. <https://doi.org/10.1111/jfb.13847>.
- Wake, D.B., Wake, M.H., Specht, C.D., 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science* 331 (6020), 1032–1035. <https://doi.org/10.1126/science.1188545>.
- Wang, Y., Yuan, H., Huang, J., Li, C., 2022. Inline index helped in cleaning up data contamination generated during library preparation and the subsequent steps. *Mol. Biol. Rep.* 49, 385–392. <https://doi.org/10.1007/s11033-021-06884-y>.
- Webb, J.F., 2013. Morphological Diversity, Development, and Evolution of the Mechanosensory Lateral Line System. In: Coombs, S., Bleckmann, H., Fay, R.R., Popper, A.N. (Eds.), *The Lateral Line System*. Springer, New York, New York, pp. 17–72.
- Whitehead, P.J.P., 1962. A review of the Indo-Pacific gizzard shad genera *Nematalosa*, *Clupanodon* and *Konosirus* (Pisces: Dorosomatidae). *Bull. Br. Mus. Nat. Hist.* 9, 87–102. <https://doi.org/10.5962/bhl.part.16338>.
- Whitehead, P.J.P., 1963. A revision of the recent round herrings (Pisces : Dussumieriidae). *Bull. Br. Mus. Nat. Hist.* 10, 305–380. <https://doi.org/10.5962/bhl.part.20529>.
- Whitehead, P.J.P., 1972. A synopsis of the clupeoid fishes of India. *J. Mar. Biol. Ass. India* 141, 160–256.
- Whitehead, P.J.P., 1985. FAO species catalogue. Vol. 7. Clupeoid fishes of the world (suborder Clupeoidei): an annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies, and wolffherrings. Part 1. Chirocentridae, Clupeidae and Pristigasteridae. British Museum (Natural History), Rome.
- Wiley, E.O., Johnson, J.G., 2010. A teleost classification based on monophyletic groups. In: Nelson, J.S., Schultze, H.-P., Wilson, M.V.H. (Eds.), *Origin and Phylogenetic Interrelationships of Teleosts, Honoring Gloria Arratia*. Verlag Dr. Friedrich Pfeil, München, pp. 123–182.
- Wilson, A.B., Teugels, G.G., Meyer, A., 2008. Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into west Africa. *PLoS One* 3 (4). <https://doi.org/10.1371/journal.pone.0001979>.
- Yuan, H., Jiang, J., Jiménez, F.A., Hoberg, E.P., Cook, J.A., Galbreath, K.E., Li, C., 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. *Mol. Ecol. Resour.* 16 (5), 1095–1106. <https://doi.org/10.1111/1755-0998.12532>.
- Yuan, H., Atta, C., Tornabene, L., Li, C., 2019. Assexon: Assembling Exon Using Gene Capture Data, 1176934319874792 *Evol. Bioinform.* 15. <https://doi.org/10.1177/1176934319874792>.