Genomics 114 (2022) 110297

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Mitogenomics of the suborder Cottoidei (Teleostei: Perciformes): Improved assemblies, mitogenome features, phylogeny, and ecological implications

Simo N. Maduna [a,*], Adam Vivian-Smith [b], Ólöf Dóra Bartels Jónsdóttir [c], Albert K.D. Imsland [c,d], Cornelya F.C. Klütsch [a], Tommi Nyman [a], Hans Geir Eiken [a], Snorre B. Hagen [a,*]

[a] Norwegian Institute of Bioeconomy Research (NIBIO), Division of Environment and Natural Resources, Svanhovd, N-9925 Svanvik, Norway
[b] Norwegian Institute of Bioeconomy Research (NIBIO), Division of Forestry and Forest Resources, P.O. Box 115, N-1431 Ås, Norway
[c] Akvaplan-niva, Iceland Office, Akralind 4, 201 Kópavogur, Iceland
[d] Department of Biosciences, University of Bergen, 5020 Bergen, Norway

ARTICLE INFO

Keywords:
Control region
Cyclopterus lumpus
Cyclopteridae
Intergenic regions
Liparidae
Positive selection

ABSTRACT

We determined the mitogenome of *Cyclopterus lumpus* using a hybrid sequencing approach, and another four closely related species in the Liparidae based on available next-generation sequence data. We found that the mitogenome of *C. lumpus* was 17,266 bp in length, where the length and organisation were comparable to those reported for cottoids. However, we found a GC-homopolymer region in the intergenic space between $tRNA^{Leu2}$ and *ND1* in liparids and cyclopterids. Phylogenetic reconstruction confirmed the monophyly of infraorders and firmly supported a sister-group relationship between Cyclopteridae and Liparidae. Purifying selection was the predominant force in the evolution of cottoid mitogenomes. There was significant evidence of relaxed selective pressures along the lineage of deep-sea fish, while selection was intensified in the freshwater lineage. Overall, our analysis provides a necessary expansion in the availability of mitogenomic sequences and sheds light on mitogenomic adaptation in Cottoidei fish inhabiting different aquatic environments.

1. Introduction

The vertebrate mitochondrial genome (hereafter mitogenome) is a compact double-stranded and closed circular molecule that is typically maternally inherited and small, being 15–20 kb in size and scarcely undergoing recombination [15,79]. With few exceptions, the mitogenome consists of 37 genes: 13 protein-coding genes (PCGs) encoding sub-units of protein complexes [Complex I, III and IV, and adenosine triphosphate (ATP) synthase] directly involved in the oxidative phosphorylation pathway, two ribosomal (r)RNAs (small *12S* and large *16S rRNA*) and 22 transfer (t)RNAs [10,15]. The PCGs are transcribed by the two mitogenomic rRNA subunits and translated by the tRNA genes (two tRNA isoacceptors for the amino acids serine ($tRNA^{Ser}$) and leucine ($tRNA^{Leu}$) and one for each of the remaining twenty tRNAs) in the mitoplast system [10,15]. The gene content and organisation of the mitogenome in fishes is conserved, although, gene rearrangements (shuffling and translocations) are present in some genera [90].

The mitogenome also contains diverse non-coding regions; (i) the control region (*CR*) or *D*-loop (for displacement loop) that encompasses the sites of initiation of heavy(H)-strand replication and both H- and low (L)-strand transcription, (ii) L-strand replication origin (O_L), and (iii) intergenic spacers (IGSs) containing transcription breakpoints [15,79]. The *CR*, located between the genes for proline tRNA ($tRNA^{Pro}$) and phenylalanine tRNA ($tRNA^{Phe}$), is the main non-coding region within the mitogenome [4,18,44,87,90]. Within and between species, the *CR* varies in sequence length owing to the presence of three unique features with different rates of evolution and genetic variation: (i) a domain associated with the termination-associated sequences (TAS or ETAS), a conserved central domain (CCD), and the conserved sequences in blocks (CSB) domain [4,18,90,103].

The mitochondrial DNA (mtDNA) sequence properties, including relatively fast evolutionary rates, presumed selective neutrality, usual lack of recombination, and high copy number of the mitogenome has made mtDNA a very powerful molecular tool and marker of choice in phylogeography, species identification, wildlife forensics, molecular phylogenetics and studies based on environmental DNA [3,8,26,30,35,85,110, 112]. Interestingly, several studies in fishes have found evidence of strong negative (purifying) and positive (diversifying)

* Corresponding authors.
E-mail addresses: simo.maduna@nibio.no (S.N. Maduna), snorre.hagen@nibio.no (S.B. Hagen).

https://doi.org/10.1016/j.ygeno.2022.110297
Received 14 May 2021; Received in revised form 5 January 2022; Accepted 1 February 2022
Available online 5 February 2022

selection acting on the mitogenome [23,43,92,93,98]. Adaptive evolution driven by positive selection has been attributed to thermal adaption [24,57], protein function and respiration [32], latitudinal gradients [23], hydrogen sulphide-rich springs [101] and pressure and depth [92,93]. Moreover, mounting evidence suggests that the mitogenome has co-evolved with the nuclear genome and/or that changes in the mtDNA can have selection effects on the nuclear genome [36,99]. Mitogenomic research has become increasingly prevalent, and there is a growing need for the assembly of complete mitogenomes for studies on genome evolution [90] and for producing mtDNA reference databases for biodiversity assessment and monitoring [64,91].

The original Sanger sequencing method [88] has served as the workhorse technology for DNA sequencing since its invention, but it has proved impractical in genome sequencing [6]. The methodological limitations of Sanger sequencing have affected many mitogenome sequencing studies in fishes, where homopolymer regions and the *CR* have been largely missing or misassembled (e.g., [68,93]). The emergence of cost-effective short- and long-read high-throughput sequencing technologies and new genome assembly methods have opened novel avenues for generating accurately assembled reference-quality mitogenomes that allow studying gene structure, function and evolution [25,55,65,89,100].

Fish in the suborder Cottoidei (sculpins, sandfishes, and snailfishes) have adapted to and colonized extreme underwater bathymetric ranges, with the deepest observation being 8178 m [74]. Collectively, this monophyletic group of fish possesses considerable ecological and morphological diversity. Previously classified as the order Scorpaeniformes [13,14,105], cottoidean fish species were recently revised into the teleostean order Perciformes based on molecular-genetic data from 20 nuclear genes and one mitochondrial DNA marker. The revised classification of Cottoidei now comprises 29 families organised into six infraorders (previously considered as suborders): Anoplopomatales (formerly Anoplopomatoidei), Cottales (former Cottoidei), Gasterosteales (akin to Gasterosteoidei, but without Indostomidae), Hexagrammales (formerly Hexagrammoidei), Zaniolepidoales (formerly Zaniolepidoidei) and Zoarcales (formerly Zoarcoidei) [14,104,106]. Despite these reclassifications, the taxonomy and systematics of the Cottoidei remains problematic, particularly for several presumably polyphyletic families such as Bathymasteridae and Stichaeidae [14,103]. Moreover, support for inter- and intra-relationships within the valid infraorders is sparse since phylogenetic analyses implementing dense taxonomic sampling are not yet available (sensu [92,93,103]). Furthermore, molecular dating estimates required for reconstructing biogeographic history and determining the role of past climate change on the diversification of clades are largely missing. Apart from understanding the molecular systematics of cottoid fishes, their high biological, morphological, behavioral and ecological diversity [2,19,69], offers opportunities for studying the type of selection pressures operating on the protein-coding regions of their mitogenome.

The family Cyclopteridae (lumpfishes or lumpsuckers) currently comprises approx. 30 small to moderately large fish (< 70 cm total length) abundant in temperate to frigid coastal habitats in the northern Atlantic, northern Pacific and Arctic Oceans [73,104,106]. Members of Cyclopteridae are characterised by a globose body, skin that is typically covered with tubercles, the pelvic fins modified to form a sucking disc, and the usual presence of two short dorsal fins [75,107]. To date, only a few genera representing the family Cyclopteridae and Liparidae have been sampled in molecular-phylogenetic investigations based on the mitogenome [68]. For instance, the incomplete mitogenome of the smooth lumpfish *Aptocyclus ventricosus* (Pallas 1769) is the only representative of the family Cyclopteridae in molecular-phylogenetic investigations [68]. Given the paucity of cyclopterid mitogenomic sequences, we here report the complete mitogenome of lumpfish (or lumpsucker) *C. lumpus* (Linnaeus 1758), a member of Cyclopteridae. Lumpfish is a semi-pelagic cold-water fish distributed across the North-Atlantic Ocean, and is highly prized for its roe (ripe egg masses) in commercial fisheries. Moreover, the lumpfish is utilized for biologically control of sea louse infestations in Atlantic salmon mariculture farms due to its delousing performance under low temperatures [39–41]. Additionally, we assembled the mitogenomes of another four species based on available data from GenBank: the common seasnail *Liparis liparis* (Linnaeus 1766), Tanaka's snailfish *Liparis tanakae* (Gilbert and Burke 1912), the hadal snailfish *Pseudoliparis swirei* [33,108], and the Mariana snailfish (*Pseudoliparis* sp. from the Yap Trench, [60,74]). We also reassembled the mitogenome of the shorthorn sculpin *Myoxocephalus scorpius* (Linnaeus 1758) based on GenBank data.

In the present study our goals were four-fold. Firstly, we conduct a detailed analysis and comparisons of the basic structure, sequence properties and arrangement of genes in mitogenomes of the revised Cottoidei, including the description of a new mitogenome (i.e., *C. lumpus*). Secondly, we use mitogenomic data to reconstruct the phylogeny of the revised Cottoidei, with denser taxonomic sampling than in previous studies. We include several recently completed mitogenomes from the Liparidae, among another set from five infraorders and 14 families. We hypothesise that this analysis confirms the sister-group relationship of the Cyclopteridae to the Liparidae, and that the cyclopterid subfamily Cyclopterinae (*C. lumpus*) will show a sister-group relationship with Liparopsinae/Aptocyclinae (*A. ventricosus*). Thirdly, we estimate divergence times for the infraorders and families. Finally, we investigate mitogenomic selection pressures in members of Cottoidei that are characterised by distinct phenotypes inhabiting ecologically diverse and extreme niches (freshwater, brackish, marine and the hadal-abyssal environment).

2. Results

2.1. Genome structure, organisation, and composition

The mitogenome of *C. lumpus* is a 17,266 bp long circular molecule, with a gene order and content similar to other related teleostean species. Like in *A. ventricosus* [68] and *Pseudoliparis swirei* (Liparidae sp. 1 YYS-2017 MT-2016 in [93]), Illumina and Sanger sequencing methods could not provide a complete mitochondrial genome due to the presence of GC-rich regions, repeat arrays, and complex secondary structures (see **Supplementary Information S1**). Even despite a mean Illumina sequencing coverage of 196.7× and the use of different approaches to close the sequence, only Nanopore sequencing provided a successful means for gap closure (**Fig. S1**). The mitogenome of *C. lumpus* encodes 37 genes including 13 protein-coding genes (PCGs), two rRNA genes, and 22 tRNA genes (duplication of two tRNAs: $tRNA^{Leu}$ and $tRNA^{Ser}$) on both strands. Nine of the genes (1 PCG and 8 tRNAs) are encoded on the L-strand, while the other 28 are encoded on the H-strand (Table 1 and Fig. 1).

In *C. lumpus*, we found a total of three non-coding regions, which include the 1131 bp control region, 39 bp of the L-strand replication origin (O_L) and 484 bp of an intergenic spacer between $tRNA^{Leu2}$ and *ND1* (Fig. 1 and Table 2). A relatively large intergenic spacer between $tRNA^{Leu2}$ and *ND1* has not been reported in fish species, therefore, we evaluated this intergenic spacer in other members of Cyclopteridae and closely related families, i.e., Liparidae and Trichodontidae. We found that the spacer was present also in *A. ventricosus* (208 bp) and the Okhotsk snailfish *Liparis ochotensis* (Schmidt 1904; Liparidae, 410 bp; Table 1), but was absent in the sailfin (Japanese) sandfish *Arctoscopus japonicas* (Steindachner 1881; Trichodontida). All *Liparis* species had a characteristic G-tract homopolymer of 13–22 bp embedded in the intergenic spacer **Table S3**), whereas *C. lumpus* had both an imperfect GC-rich tract of 25 bp and a 19 bp C-tract homopolymer which was found 72 bp upstream of the GC-rich tract. The C-tract clearly inhibited both Illumina and Sanger sequencing (see **Supplementary Information S1**). The G-tract homopolymer between $tRNA^{Leu2}$ and *ND1* appears to be absent from *Pseudoliparis* species (**Data File S1**). The presence of inverted G- and C-homopolymer tracts between liparid and cyclopterid

Table 1
General information and nucleotide composition for 68 representative mitochondrial genomes of the suborder Cottioidei (Perciformes) used in this study.

Infraorder	Family	Species	Depth zone	Accession number	Size (bp)	Whole genome composition						
						A%	G%	T%	C%	A + T%	AT skew	GC skew
Anoplopomatales	Anoplopomatidae	*Anoplopoma fimbria*	175–2740	NC_018119	16,507	26.1	18.3	26.0	29.6	52.1	0.001	−0.235
		Erilepis zonifer	0–680	NC_026889	16,500	26.7	17.8	26.5	29.0	53.2	0.004	−0.24
Cottales	Agonidae	*Aspidophoroides olrikii* (then *Ulcina olrikii*)[a]	0–632	NC_027600	17,200	27.3	16.9	28.5	27.3	55.8	−0.021	−0.235
	Cottidae	*Clinocottus analis*[b]	0–18	NC_013828	18,374	25.4	17.0	24.1	33.5	49.5	0.025	−0.327
		Comephorus baikalensis	0–1600	NC_036148	16,530	26.7	17.2	26.1	30.0	52.8	0.012	−0.27
		Comephorus dybowskii	0–1600	NC_036149	16,527	26.7	17.2	26.2	29.9	52.9	0.01	−0.27
		Cottus amblystomopsis	0–10	NC_035002	16,528	25.9	17.8	26.1	30.2	52	−0.003	−0.258
		Cottus asper	NA	NC_036145	16,511	27.2	16.9	26.3	29.6	53.5	0.017	−0.275
		Cottus bairdii	NA – 16	NC_028277	16,529	27.4	16.7	26.1	29.9	53.4	0.024	−0.284
		Cottus czerskii	0–15	KY783660	16,560	26.2	18.0	26.2	29.6	52.4	−0.001	−0.243
		Cottus dzungaricus (then *Cottus sibiricus altaicus*)	NA	NC_024739	16,527	26.9	17.1	26.3	29.7	53.2	0.012	−0.27
		Cottus hangiongensis	NA	NC_014851	16,598	25.5	18.2	25.9	30.4	51.4	−0.008	−0.251
		Cottus perifretum	NA	NC_036146	16,523	27.0	17.1	26.1	29.8	53.1	0.016	−0.272
		Cottus poecilopus	0–15	NC_014849	16,560	25.7	18.2	25.7	30.4	51.4	−0.001	−0.251
		Cottus reinii	NA	NC_004404	16,561	26.3	17.6	25.8	30.3	52.1	0.01	−0.264
		Cottus rhenanus	NA	NC_036147	16,522	27.1	17.0	26.2	29.7	53.3	0.017	−0.274
		Cottus szanaga	NA	NC_032039	16,518	26.5	17.4	26.2	29.9	52.7	0.007	−0.265
		Cottus volki	NA	NC_035001	16,536	27.2	16.8	26.3	29.7	53.5	0.018	−0.278
		Enophrys diceraus	0–380	NC_022147	16,976	27.5	16.6	27.2	28.6	54.7	0.006	−0.265
		Gymnocanthus herzensteini	0–300	NC_034651	16,691	26.6	17.5	25.9	30.0	52.5	0.012	−0.264
		Gymnocanthus intermedius	15–256	NC_034650	16,639	26.4	17.6	25.5	30.4	51.9	0.017	−0.266
		Hemilepidotus gilberti	0–604	NC_034653	16,907	26.8	17.0	25.8	30.4	52.6	0.018	−0.282
		Icelus spatula	12–930	NC_027587	16,384	26.5	17.4	26.0	30.1	52.5	0.008	−0.266
		Icelus toyamensis (*Lycodes toyamensis?*)[c]	NA	NC_004409	16,697	25.4	18.9	25.4	30.3	50.8	0	−0.233
		Mesocottus haitej	NA	NC_022181	16,527	26.6	17.4	26.1	29.9	52.8	0.01	−0.265
		Myoxocephalus scorpius†	0–451	This study	16,607	27.2	16.8	26.7	29.3	53.9	0.009	−0.271
		Trachidermus fasciatus	NA	NC_018770	16,536	26.3	18.1	25.5	30.1	51.8	0.015	−0.249
	Cyclopteridae	*Aptocyclus ventricosus*	612–1700	NC_008129	15,974	28.6	15.5	25.8	30.2	54.4	0.051	−0.322
		Cyclopterus lumpus	0–868	This study	17,266	26.5	17.6	26.8	29.0	53.3	−0.006	−0.245
	Liparidae	*Liparis agassizii*	0–100	KX156765	17,896	29.5	15.2	28.0	27.4	57.5	0.026	−0.286
		Liparis ochotensis	0–761	MG718032	17,522	28.8	15.5	28.6	27.4	57.4	0.003	−0.277
		Liparis tanakae (*L. ochotensis*)[d†]	10–1100	This study	17,036	28.8	15.3	28.8	27.1	57.6	0.000	−0.278
		Pseudoliparis swirei[e]	6198–8078	KY242356	16,772	31.0	13.8	29.1	26.1	60.1	0.032	−0.308
		Pseudoliparis swirei (*P. amblystomopsis*)[f†]	7210–7230	This study	20,607	34.7	13.6	27.9	23.8	62.6	0.109	−0.273
		Pseudoliparis sp. Yap Trench (*P. yapensis*)†	6898–7966	This study	20,560	34.7	12.7	28.8	23.9	63.5	0.093	−0.306
	Trichodontidae	*Arctoscopus japonicus*	0–550	NC_002812	16,577	26.2	17.9	27.6	28.3	53.7	−0.026	−0.225
	Aulorhynchidae	*Aulorhynchus flavidus*	0–30	NC_010268	16,894	29.8	14.9	30.1	25.2	59.9	−0.006	−0.259
		Aulichthys japonicus	NA	AB445127	16,594	28.2	17.2	29.1	25.5	57.3	−0.015	−0.196
Gasterosteales	Gasterosteidae	*Apeltes quadracus*	0–3	NC_011580	16,472	27.6	16.7	27.5	28.1	55.1	0.001	−0.254
		Culaea inconstans	0–55	NC_011577	16,465	28.8	16.3	28.3	26.6	57.1	0.008	−0.242
		Gasterosteus aculeatus	0–100	AP002944	15,742	27.1	17.0	28.2	27.6	55.3	−0.021	−0.238
		Gasterosteus wheatlandi	NA	NC_011570	16,538	27.7	16.9	28.7	26.7	56.4	−0.018	−0.226
		Pungitius hellenicus	NA	NC_029471	16,713	27.4	17.5	26.4	28.6	53.8	0.019	−0.24
		Pungitius kaibarae	NA	NC_014893	16,505	27.5	17.2	26.5	28.8	54	0.019	−0.252
		Pungitius laevis	NA	NC_029473	16,576	27.6	17.3	27.0	28.0	54.7	0.011	−0.237
		Pungitius platygaster	NA	NC_029474	16,570	27.8	17.2	26.6	28.4	54.4	0.023	−0.245
		Pungitius pungitius	0–110	NC_011571	16,388	27.5	17.3	26.8	28.4	54.3	0.013	−0.243
		Pungitius sinensis	NA	NC_014889	16,581	27.5	17.3	26.9	28.3	54.4	0.011	−0.24
		Pungitius tymensis	NA	NC_029472	16,481	27.2	17.6	26.4	28.8	53.6	0.015	−0.241
		Spinachia spinachia	NA	NC_011582	16,359	29.2	15.4	31.0	24.4	60.2	−0.03	−0.227
	Hypoptychidae	*Hypoptychus dybowskii*	NA	NC_004400	16,479	24.6	18.6	28.0	28.8	52.6	−0.065	−0.214
Hexagrammales	Hexagrammidae	*Hexagrammos agrammus*	NA	NC_021459	16,514	26.9	17.2	26.2	29.7	53	0.014	−0.266
		Hexagrammos lagocephalus	0–596	NC_026888	16,505	27.0	17.3	26.3	29.5	53.3	0.013	−0.262
		Hexagrammos otakii	139–155	NC_028630	16,513	26.9	17.3	25.9	29.9	52.8	0.019	−0.266
		Ophiodon elongatus	0–475	NC_026887	16,528	26.7	17.5	25.6	30.2	52.3	0.02	−0.265
		Pleurogrammus azonus	0–240	NC_023129	16,591	26.9	17.2	27.1	28.8	54	−0.002	−0.251
		Pleurogrammus monopterygius	0–720	NC_023475	16,575	27.1	17.1	27.1	28.7	54.2	−0.001	−0.252
Zoarcales	Anarhichadidae	*Anarhichas denticulatus*	60–1700	NC_037606	16,519	26.7	17.8	27.3	28.3	54	−0.012	−0.228
		Anarhichas lupus	1–600	NC_009773	16,516	26.7	17.8	27.4	28.1	54.1	−0.014	−0.225
		Anarhichas minor	25–600	NC_037609	16,519	26.7	17.8	27.2	28.3	53.9	−0.009	−0.229
	Pholidae	*Pholis crassispina*	0–5	NC_004410	16,522	25.6	18.7	27.4	28.3	53	−0.034	−0.204
		Pholis fangi	NA	NC_029842	16,523	25.4	18.8	27.1	28.7	52.4	−0.032	−0.208
		Pholis nebulosa	0–200	NC_029841	16,524	25.6	18.8	27.3	28.4	52.9	−0.033	−0.204
	Stichaeidae	*Chirolophis japonicus*	NA	LC081980	16,522	25.6	18.2	28.5	27.7	54.1	−0.054	−0.206

(*continued on next page*)

Table 1 (*continued*)

Infraorder	Family	Species	Depth zone	Accession number	Size (bp)	Whole genome composition						
						A%	G%	T%	C%	A + T%	AT skew	GC skew
		Leptoclinus maculatus	2–607	NC_027588	16,521	27.1	17.5	27.7	27.8	54.7	−0.011	−0.227
		Xiphister atropurpureus	44–538	NC_034669	16,518	25.4	18.7	27.6	28.3	53	−0.041	−0.206
	Zoarcidae	*Lycodes tanakae*	10–1100	NC_034649	16,594	25.6	18.7	25.2	30.6	50.8	0.007	−0.242
		Lycodes ygreknotatus	50–386	NC_034751	16,486	26.3	18.0	25.3	30.4	51.6	0.019	−0.258

† genomes inferred and assembled as described in Supplementary Information.

[a] Originally published as *Ulcina olriki* (Lütken, 1877), misspelling of the species name i.e., should be *U. olrikii* not *U. olriki.*

[b] Presence of an 830 bp repeat region (*CR - tRNA^{Pro}*) record has not yet been subject to final NCBI review (submitted 20-MAR-2009), but identical to FJ848374.

[c] Reported on GenBank as L. *toyamensis* but the SOURCE field states "mitochondrion *I. toyamensis* (*Petroschmidtia toyamensisi* [unaccepted])" while the ORGANISM field states "*I. toyamensis*".

[d] Based on our sequence and phylogenetic analysis we found that the specimen of *Liparis tanakae* sequenced by Wang et al. [108] is actually *L. ochotensis* (Fig. 3; **Supporting Information S2**).

[e] Complete genome except for partial control region (D-loop). For further information on *P. swirei* (Liparidae_sp_YYS-2017 [108] or MT2016 No. 1 [93]) see the Supplementary Information S3. Labelled as *Liparidae_sp_MT2016_YYS2017_KY242356* and *Pseudoliparis_swirei _MT2016_YYS2017_KY242356* in Fig. 3 and Fig. 4, respectively.

[f] *Pseudoliparis amblystomopsis* and *P. swirei* were found to be synonomous according to NCBI data related to Wang et al. [108].

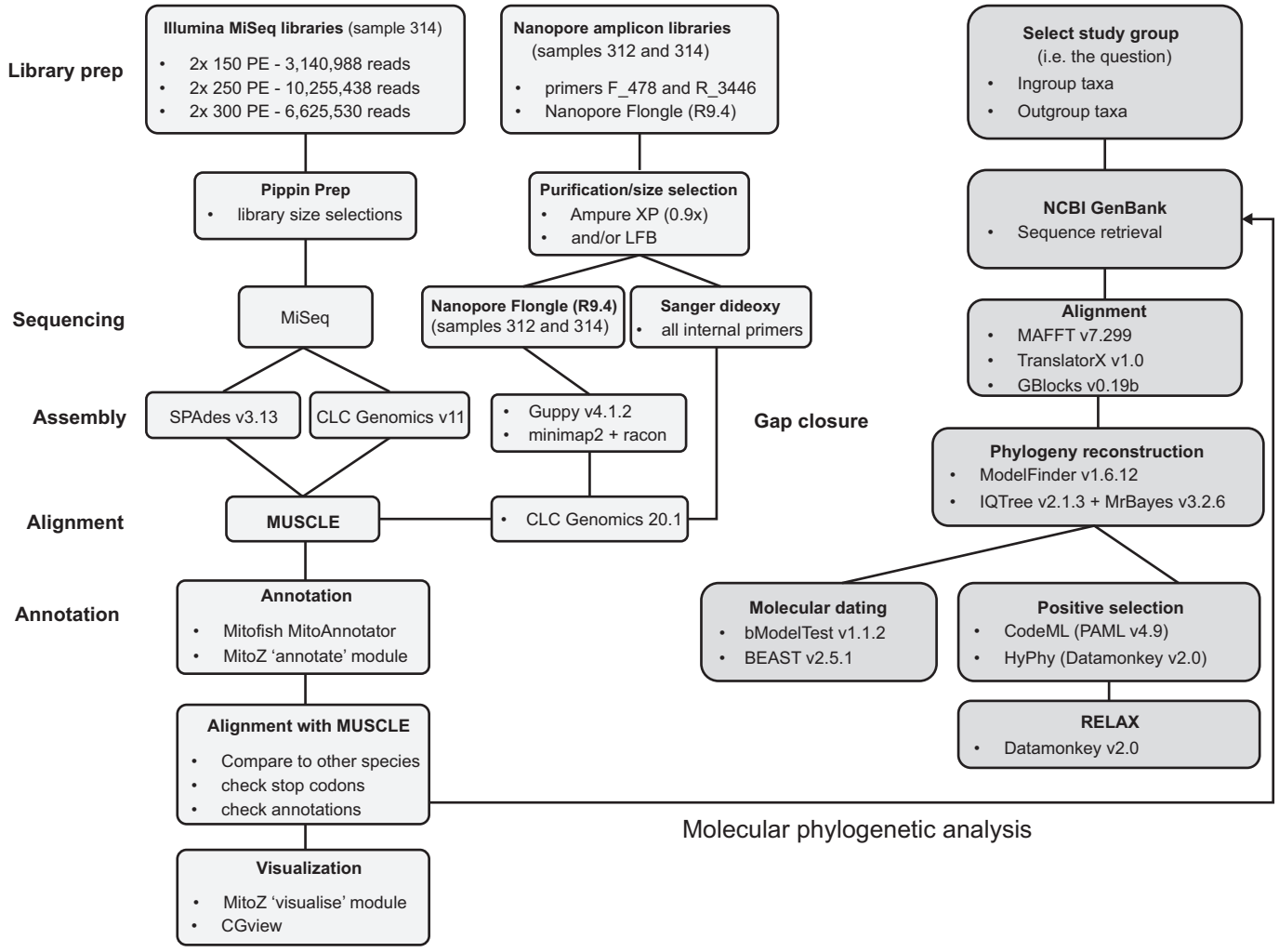

Fig. 1. Schematic summary and representation of *Cyclopterus lumpus* mitogenome assembly and phylogenetic analysis.

fish may result from an ancestral inversion event (**Table S3**).

In this study, the *CR* of the related snailfish mitogenomes was found to be especially long (23.6% to 23.8% of the mitogenome length) and difficult to assemble, and having a very high AT-rich composition for both species (see **Supplementary Information**). In contrast, the *C. lumpus CR* is of a moderate size, comparable to other species being a total of 6.6.% in length (**Fig. S3**). Within the *CR* of *C. lumpus* and all study taxa we identified the TASs (TACAT) with varying number per domain, consisting of three CSBs within the conserved central domain (CSB-D, CSB-E and CSB-F), three CSBs within the conserved sequences in blocks domain (CSB-I, CSB-II, and CSB-III), a T-homopolymer spanning 10–12 nucleotides located in the upper region of CSB-I, and AT-rich

Table 2
The organisation and characteristics of the complete mitochondrial genome of *Cyclopterus lumpus*. IGN values represent intergenic nucleotides and overlapping nucleotides (−). H-strand – heavy strand; L-strand – low strand.

Gene	Strand	Position		Size		Codon		Anticodon	IGN
		Start	End	Nucleotide (bp)	Amino acid	Start	Stop		
tRNA^{Phe} (F)	H	1	68	68				GAA	0
12S rRNA	H	69	1012	944					0
tRNA^{Val} (V)	H	1013	1084	72				TAC	0
16S rRNA	H	1085	2775	1691					0
tRNA^{Leu2} (L2)	H	2776	2849	74				TAG	0
pseudo-tRNA^{Ser}(S)	H	3268	3332	65				TCT	418
ND1	H	3333	4307	975	324	ATG	TAG		0
tRNA^{Ile} (I)	H	4312	4381	70				GAT	4
tRNA^{Gln} (Q)	L	4381	4451	71				TTG	−1
tRNA^{Met} (M)	H	4451	4519	69				CAT	−1
ND2	H	4520	5565	1046	348	ATG	TA		0
tRNA^{Trp} (W)	H	5566	5636	71				TCA	0
tRNA^{Ala} (A)	L	5638	5706	69				TGC	1
tRNA^{Asn} (N)	L	5708	5780	73				GTT	1
O_L	L	5782	5820	39					1
tRNA^{Cys} (C)	L	5818	5883	66				GCA	−3
tRNA^{Tyr} (Y)	L	5884	5952	69				GTA	0
COI	H	5954	7504	1551	516	GTG	TAA		1
tRNA^{Ser2} (S2)	L	7505	7575	71				GCT	0
tRNA^{Asp} (D)	H	7578	7650	73				GTC	2
COII	H	7656	8346	691	230	ATG	T		5
tRNA^{Lys} (K)	H	8347	8420	74				TTT	0
ATPase 8	H	8422	8589	168	55	ATG	TAA		1
ATPase 6	H	8580	9262	683	227	ATG	TA		−10
COIII	H	9263	10,047	785	261	ATG	TA		0
tRNA^{Gly} (G)	H	10,048	10,120	73				TCC	0
ND3	H	10,121	10,469	349	116	ATG	T		0
tRNA^{Arg} (R)	H	10,470	10,538	69				TCG	0
ND4L	H	10,539	10,835	297	98	ATG	TAA		0
ND4	H	10,829	12,209	1381	460	ATG	T		−7
tRNA^{His} (H)	H	12,210	12,278	69				GTG	0
tRNA^{Ser1} (S1)	H	12,279	12,346	68				GCT	0
tRNA^{Leu1} (L1)	H	12,351	12,423	73				TAG	4
ND5	H	12,424	14,262	1839	612	ATG	TAA		0
ND6	L	14,259	14,780	522	173	ATG	TAG		−4
tRNA^{Glu} (E)	L	14,781	14,848	68				TTC	0
Cytb	H	14,854	15,994	1141	380	ATG	T		5
tRNA^{Thr} (T)	H	15,995	16,066	72				TGT	0
tRNA^{Pro} (P)	L	16,066	16,135	70				TGG	−1
Putative *CR*	H	16,136	17,266	1131					0

variable number tandem repeats (Fig. 2B). The nucleotide composition of the mitogenome of *C. lumpus* is 26.5% A, 17.6% G, 26.8% T, and 29% C bases. The genome has an overall A + T content of 53.3%, a negative AT-skew (−0.006) and a negative GC-skew (−0.245), which appears to be a common feature in Cottioidei (Table 1). With the exception of *ATP8*, all PCGs have a negative AT-skew, indicating that thymines occur more frequently than adenines, similar to other members of Cottioidei. Likewise, all the PCGs had a negative GC-skew, suggestive of C biased nucleotide composition. Similar to other fish, *ATP8* has the highest AT-skew while *ND6* had the highest GC-skew.

2.2. Protein-coding genes

Among the 13 protein-coding genes, we found two cases of overlapping reading frames on the same strand: *ATP6* and *ATP8* share 10 nucleotides, and *ND4* and *ND4L* share seven nucleotides. Moreover, *ND5* and *ND6* share four nucleotides on the opposite strand. All PCGs share the start codon ATG, except for *COI*, which has GTG as the start codon. The stop codon TAA is present in *COI*, *ATP8*, *ND4L* and *ND5*; TAG, is present in *ND6*; and truncated stop codon "TA" is found in *ND2*, *ATP6* and *COIII*, and "T" is found in *COII*, *ND3*, *ND4*, and *Cytb* (Table 2).

We found a significant bias towards A/T in the codon usage of the mitochondrial genome of *C. lumpus*, concordant with the nucleotide composition estimates. Among PCGs, leucine (17.63%) and cysteine (0.71%) are the most and the least frequently used amino acids, respectively. Leucine (Leu1) had the highest relative synonymous codon usage (RSCU = 4.73) followed by serine (Ser2, RSCU = 4.3). The Leu2 had the lowest RSCU of 1.27.

2.3. Ribosomal RNA and transfer RNA genes

The *12S rRNA* and *16S rRNA* genes in the mitogenome of *C. lumpus* are positioned between *tRNA^{Phe}* and *tRNA^{Val}*, and between *tRNA^{Val}* and *tRNA^{lue2}*, respectively. The length of *12S rRNA* and *16S rRNA* is 944 bp and 1691 bp, respectively. Both rRNAs are encoded by the H-strand, with A + T contents of 51.69% and 53.64%, respectively. Out of 22 tRNAs, 14 genes are encoded by the H-strand, and the remaining 8 tRNAs are encoded by the L-strand (Table 1). The tRNAs varied in size from 66 bp (*tRNA^{Cys}*) to 74 bp (*tRNA^{Lue2}* and *tRNA^{Lys}*). Each of these tRNAs could be folded into a secondary cloverleaf structure as predicted by ARWEN and tRNAscan-SE, except for *tRNA^{Ser1}*, which lacked a dihydrouridine (DHU) arm. Anticodon sequences of these tRNA genes were also identical with ones previously reported in other members of Cottioidei (Table 1). Finally, a *pseudo-tRNA^{Ser}*-like structure was predicted immediately prior to the *ND1* PCG, at positions 3268–3332 bases, on the heavy stand, by Infernal v. 1.1.1 as implemented in the MitoZ annotation and visualization modules. tRNAscan-SE, however, did not support such a model, but a partial cloverleaf and series of secondary RNA structures were reported using RNAfold v. 2.0 (**Fig. S2**).

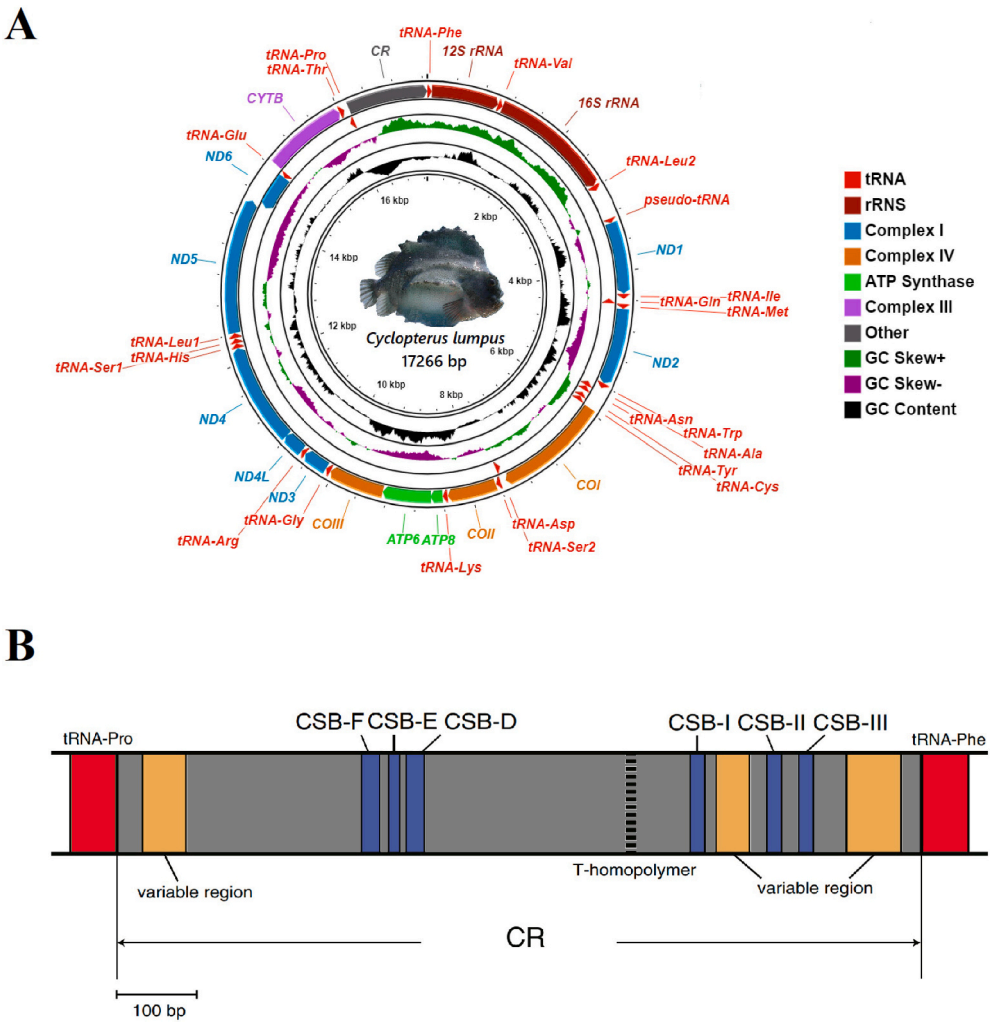

Fig. 2. (A) Circular map of the complete mitochondrial genome depicting gene order and protein complexes. Arrows indicate the direction of gene transcription. The GC content and GC-skew are plotted as the deviation from the average GC content and GC-skew of the entire sequence, respectively. (B) Schematic diagram of the control region (D-loop) model of fish mitogenome modified from Satoh et al. [90]. Locations of conserved sequence block domains and variable regions are mapped. The location of the T-homopolymer region is represented by broken lines.

2.4. Phylogenetic reconstruction and divergence times

The final concatenated alignment from 13 PCGs and 2 rRNA genes comprised 13,970 sites of which 8021 were variable and 6757 were parsimony-informative. Phylogenetic relationships among five infraorders and 14 families of Cottoidei were similar in model-based phylogenetic reconstructions based on Maximum likelihood and Bayesian inference (Fig. 3). These phylogenetic reconstructions placed *C. lumpus* as sister to *A. ventricosus* in the family Cyclopteridae, and this family as sister to the Liparidae (Fig. 3). Our results provided good support for the monophyly of each infraorder and family, except for the family Cottidae and Stichaeidae. The results placed the cottid *Hemilepidotus gilberti* (Jordan and Starks 1904) closest to *Aspidophoroides olrikii* (Lütken 1877), which is a member of Agonidae, with a posterior probability of 99% (Fig. 3). *Leptoclinus maculatus* (Fries 1838), a member of Stichaeidae, has a closer relationship to with members of Pholidae and Zoarcidae. The best supported phylogenetic relationship of infraorders and families found in this study is represented in Fig. 3.

Based on the fossil record and a maximum crown age of Gasterosteidae, the minimum time since divergence of members of the family was constrained to 27.15 MYA, corresponding to the Chattian stage of the Oligocene epoch. A different tree topology from BI/ML was obtained from molecular dating analyses in relation to the placement of Gasterosteales and Anoplopomatales (Fig. 4). Our molecular dating results suggested that Gasterosteales diverged from the other cottoid infraorders around 44.31 MYA (95% HDP: 25.37–69.64 MYA), while the Zoarcales and Anoplopomatales diverged around 34.1 MYA (18.89–53.99 MYA) and 31.23 MYA (17.17–49.38 MYA), respectively. The divergence time between Cottales and Hexagrammales is estimated at around 39.48 MYA (14.30–39.26 MYA). Our results also reveal that Cyclopteridae and Liparidae diverged 15.86 MYA in the middle Miocene epoch (8.89–24.94 MYA).

2.5. Positive selection and strength of natural selection

The ω (dN/dS) ratio calculated under the basic model (i.e., one-ratio model (M0)) was 0.04026 for the 13 concatenated PCGs. Then, in the comparison of the 'one-ratio' model (M0) and the 'free-ratio' model (M1), the LRT indicated that the free-ratio model fit the data significantly better than did the M0 model ($p < 0.001$, Table 3). Further, the "two-ratios" (BM2) model was found fitting the data significantly better than the M0 model ($p < 0.001$, Table 3), when the deep-sea marine lineages of Cottales was set as a foreground branch (Fig. 3, branch R), indicating that ω of the deep-sea marine lineages ($\omega_1 = 0.18785$) was significantly higher than that of other lineages ($\omega_0 = 0.03994$). Moreover, the "three-ratios" (BM3) model was found fitting the data significantly better than the BM2 model ($p < 0.001$, Table 3), when the freshwater and deep-sea marine lineages of Cottales was set as foreground branches, indicating that ω of the deep-sea marine ($\omega_1 = 0.18785$) and freshwater ($\omega_2 = 0.01274$) lineages were both significantly higher and lower, respectively, from the background ratio ($\omega_0 = 0.04017$) (Fig. 3, branch R and T). The branch-site model for detecting

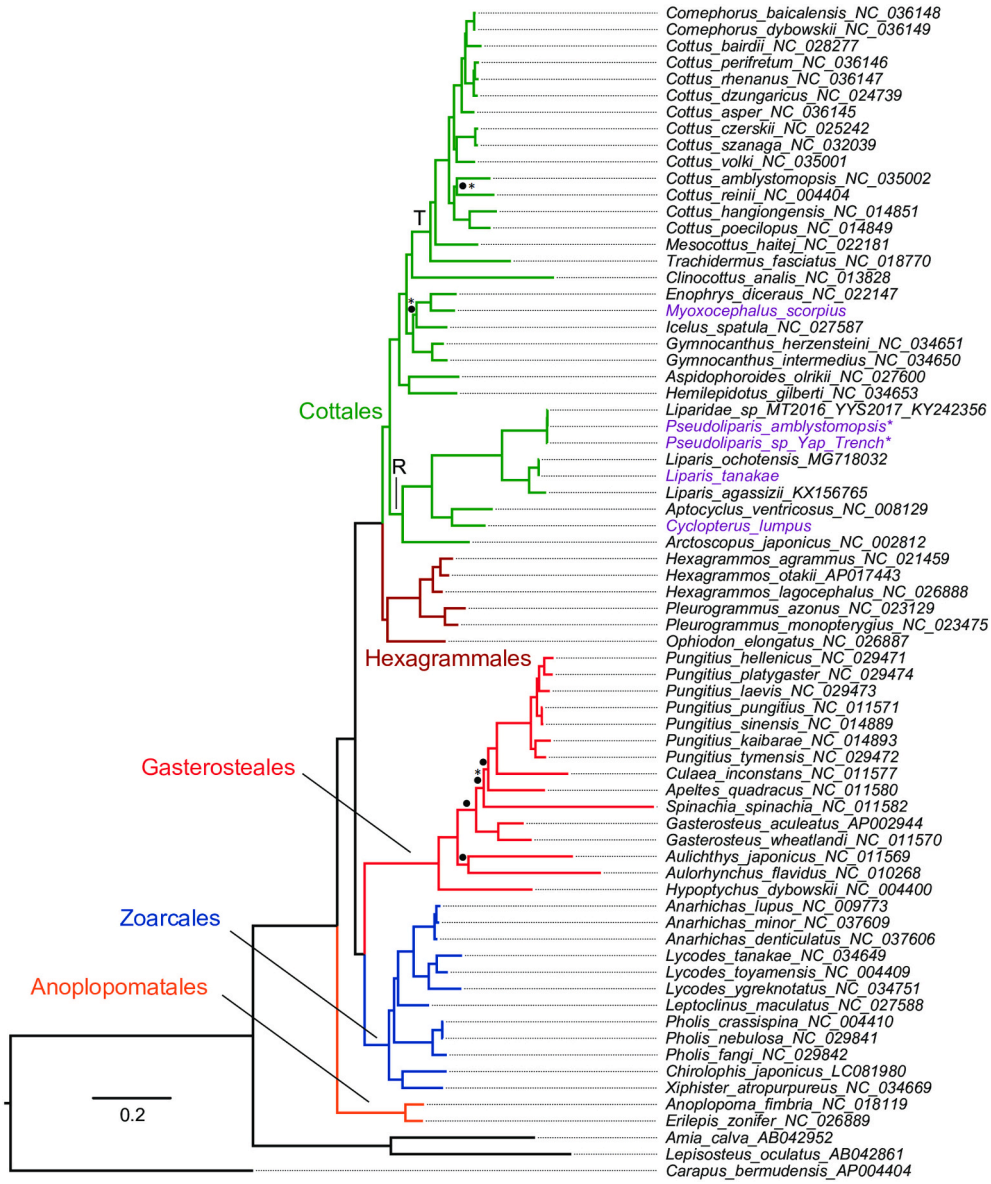

Fig. 3. Maximum likelihood phylogeny for the reclassified Cottioidei based on the 13 mitochondrial protein-coding genes and two rRNA genes. Posterior probabilities and ultrafast bootstrapping SH-aLRT branch supports with less than 95% are indicated at nodes with an asterisk and a black dot, respectively. T and R represent branches and lineages used in selection analyses. Bayesian inference yielded the same topology. Taxa in purple text are from the present study. **Pseudoliparis amblystomopsis* and *P. swirei* were found to be synonymous according to NCBI data related to Wang et al. [108]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the sites affected by positive selection found 36 sites independently subjected to positive selection in the deep-sea marine lineages of Cottales, although 1 site had a high BEB value of >95% (2162 D 0.982*), located within the *ND4L* gene (Table 3). We found no clear linear relationship between ω and depth inhabited by study taxa ($r^2 = 0.019$, $p = 0.22$).

Within the concatenated PCG dataset, FUBAR found evidence of episodic negative/purifying selection at 3676 sites and episodic positive/diversifying selection at three sites (Table 3). The latter were located within the *ND5* (site 3172) and *ND6* (sites 3244 and 3246) genes. MEME found evidence of episodic positive selection at 19 sites located within the genes *ND1* (site 166), *ND2* (sites 609 and 612), *COI* (sites 816 and 1077), *COII* (sites 1158, 1405, and 1407), *ATP8* (site 1464), *COIII* (sites 1747, 1750, and 1804), *ND4* (sites 2256 and 2597), *ND5* (sites 3047 and 3246), *ND6* (sites 3284, and 3364), and *CYTB* (site 3609) (Table 3).

The RELAX analysis found significant intensified selection ($k = 1.61$, $p < 0.01$) when the recently radiated freshwater lineage of Cottales was taken as a test branch (Fig. 3, branch T) and the deep-sea marine lineage as the reference branch (Fig. 3, branch R) (Table 3).

3. Discussion

Here, we report the complete mitochondrial genome assemblies for *C. lumpus* and several other Liparidae, including the extraordinary hadal snailfish *Pseudoliparis swirei*. Whilst the control region was difficult to assemble in the latter species due to the presence of a vast array of repeats and complex structure (see **Figs. S6 and S7**), another GC-rich structure located in the intergenic spacer between *tRNA-*^{Leu2} and *ND1* was problematic for both Illumina and Sanger-based sequencing approaches in *C. lumpus* (**Fig. S1**). The formation of secondary conformation structures, including hairpins, tandem repeat arrays, and GC and AT-rich tracts, likely prevent effective library generation or population of short-read sequences on Illumina flow cells. Even at very high coverage, paired-end reads ranging up to 830 bp in insert length could not provide sufficient scaffolding. Complete sequencing of the non-coding control region was also unsuccessful in previous attempts focusing on *A. ventricosus* and the hadal snailfish *Pseudoliparis* sp. 1 MT-2016 (KY242356) due to the length and complexity of that region ([93]; **Fig. S6**). Although the size and expansion of the control region in *C. lumpus* (1131 bp) is not as extreme as in the hadal and abyssal snailfishes (e.g., 4899 bp for *P. swirei*, 4844 bp for *P. yapensis*; [68]), a

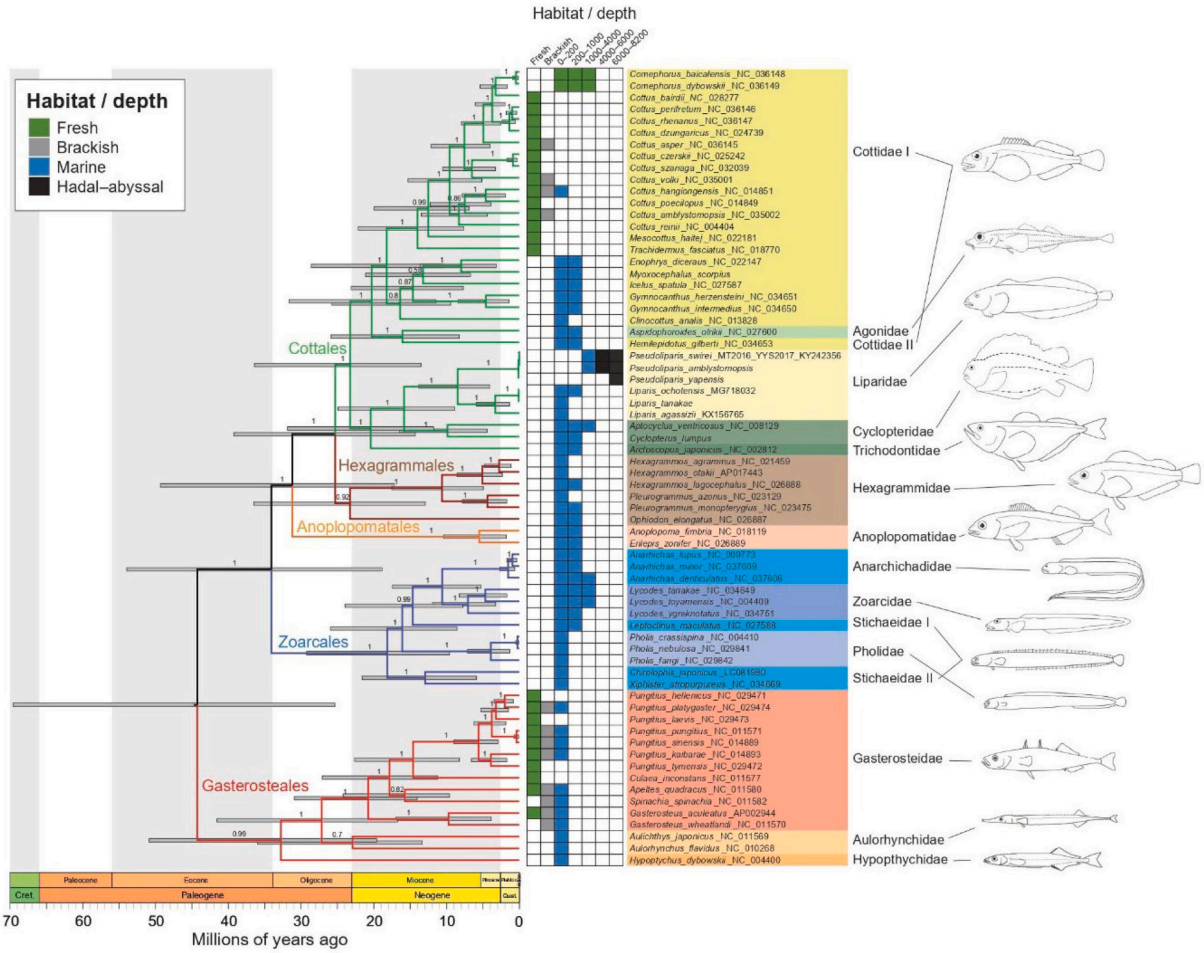

Fig. 4. Maximum clade credibility (MCC) chronogram from molecular dating analysis of the systematically revised Cottioidei. Values above the branches indicate the posterior probabilities, horizontal blue bars show the 95% highest posterior density (HPD) intervals around mean node ages. Habitat and depth are represented by the centre panel. This is generalized for mature fish but *Pseudoliparis swirei* larval stages may populate depths from 1000 to the hadal zone [33]. Mature body plan and form for selected families is shown on the right-hand side. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

correct length estimation is warranted for other cyclopterid species, since this attribute may not be restricted to deep-living fish [93]. Together, these results also highlight the requirement for long-read sequencing technologies like PacBio and Oxford Nanopore sequencing, which operate unimpeded on native or amplified DNAs.

The large intergenic spacer between $tRNA^{Leu2}$ and *ND1* gene has not been reported in fish species (sensu [90]). Our investigations uncovered that this intergenic spacer DNA varies in size and was present in several closely related species. A homopolymer G-tract of variable length was also present in *Liparis* species, but was not found within *Pseudoliparis* (**Table S3**). The function of the intergenic spacer DNA between the $tRNA^{Thr}$ and $tRNA^{Pro}$ genes of cod fishes was proposed to be related to heavy-strand transcript termination [9,45]. The intergenic spacer could possibly also have additional functional roles. Whilst this remains to be investigated in detail, we show that bases spanning positions 3268–3332 may encode a transcribed RNA structure reminiscent of a *pseudo-*$tRNA^{Ser}$ molecule (**Fig. S3**). The information provides novel genomic information, which, when combined with the available microsatellite database [62] can help delineate the phylogeographic structure of *C. lumpus*. In fact, all available population genetic studies for *C. lumpus* are based solely on genotypic data from nuclear microsatellite loci, potentially overlooking phylogeographic intraspecific variation [31,46,77,114]. Intra-specific comparative analysis of the *C. lumpus* mitogenomes revealed that informative genetic variation is present and might have potential application for provenance sourcing and the determination of lumpfish broodstocks used in aquaculture (**Table S2**).

Our molecular phylogenetic reconstruction of 68 mitogenomes of fish within the revised suborder Cottoidei (sensu [14]) provides a fresh insight and deeper clarity on patterns of radiation and adaptive evolution within this group of fish. This is true especially for the deep-sea families, such as Ophidiidae and Liparidae, which make the greatest contribution to the fish fauna at depths greater than 6000 m [82]. The results of the phylogenetic relationships of infraorders and their families in the Cottoidei appear most congruent with the hypothesis of Betancur-R et al. [14]. Additionally, our study recovered the family Stichaeidae and Cottidae as polyphyletic, consistent with previous conclusions [14,52,53,94,95,102,103]. Our results firmly support a sister-group relationship between Cyclopteridae and Liparidae, and show that these families diverged between the middle and late Miocene (15.86 MYA, 95% HPD: 8.89–24.94 MYA).

The confirmed placement of these taxa is significant and is consistent with the presence of major morphological synapomorphies such as the specialized pelvic sucking disc, which is present in all but a few species from both families [54,75,76,107]. The sucking disc is however lost in deep hadal and abyssal liparid fish and it is also absent from the basal liparid genus *Nectoliparis* [54]. Similarly, *Aptocyclus japonicus* (Trichodontae), which was recovered basally to all other liparid and cyclopterid fish, does not possess this adaptation. The sister relationship of

Table 3
Selection pressure on mitochondrial genes in the Cottoidei lineages.

CODEML

Model	lnL	Parameter estimates	Model compared	2ΔlnL	LRT *p*-value	Positively selected sites
(a) Branch-specific models						
M0 (one-ratio)	−232,000.987826	$\omega = 0.04026$				
Two-ratio (BM2)	−231,984.035755	$\omega_0 = 0.03994$ $\omega_1 = 0.18785$	BM2 vs. M0	33.904142	0.000000*	
Three-ratio (BM3)	−231,972.688698	$\omega_0 = 0.04017$ $\omega_1 = 0.18785$ $\omega_2 = 0.01274$	BM3 vs. BM2	22.694114	0.000002*	
Free-ratio (M1)	−231,152.336638		M1 vs. M0	1697.302376	0.000000*	
(b) Branch-site models						
Model A_{null}	−230,250.049960	Proportion: $P_0 = 0.88084$ $P_1 = 0.03929$ $P_{2a} = 0.07645$ $P_{2b} = 0.00341$ Background ω: $\omega_0 = 0.03088$ $\omega_1 = 1$ $\omega_{2a} = 0.03088$ $\omega_{2b} = 1$ Foreground ω: $\omega_0 = 0.03088$ $\omega_1 = 1$ $\omega_{2a} = 0.03088$ $\omega_{2b} = 1$				
Model A	−230,248.026308	Proportion: $P_0 = 0.93959$ $P_1 = 0.04190$ $P_{2a} = 0.01771$ $P_{2b} = 0.00079$ Background ω: $\omega_0 = 0.03088$ $\omega_1 = 1$ $\omega_{2a} = 0.03088$ $\omega_{2b} = 1$ Foreground ω: $\omega_0 = 0.03088$ $\omega_1 = 1$ $\omega_{2a} = 6.35337$ $\omega_{2b} = 6.35337$	Model A vs. Model A_{null}	4.047304	0.044242	2162 D 0.982*

FUBAR (Positively selected sites)

Site	α	β	β-α	*PP*[α > β]	PP[α < β]	BF[α < β]
3172	0.374	0.927	0.554	0.039	0.937*	479.207
3244	0.402	0.949	0.547	0.047	0.925*	399.626
3246	0.17	0.954	0.783	0.009	0.986*	2202.937

MEME (Positively selected sites)
166, 609, 612, 816, 1077, 1158, 1405, 1407, 1464, 1747, 1750, 1804, 2256, 2597, 3047, 3246, 3284, 3364, 3609

RELAX

Model	lnL	Branch set	Parameter (*k*)	ω_1	ω_2	ω_3	2ΔlnL	LRT *p*-value
RELAX alternative	−227,614.0	Reference (deep-sea)	1.61	0.01 (89.67%)	0.55 (10.33%)	68.66 (0.00%)	49.05	0.000000
		Test (freshwater)		0.00 (89.67%)	0.38 (10.33%)	892.52 (0.00%)		
RELAX null	−227,638.6	Reference (deep-sea)	1.00	0.00 (89.48%)	0.44 (10.52%)	12.62 (0.00%)		
		Test (freshwater)		0.00 (89.48%)	0.44 (10.52%)	12.62 (0.00%)		

* $p < 0.01$ or posterior probability (*PP*) > 90%, ** *PP* > 95%.

Trichodontidae to Cyclopteridae + Liparidae is also consistent with previous conclusions [14,94,95]. Moreover, the specialized pelvic sucking disc is consistently absent in the Agonidae group, which consists of *H. gilberti* and *Aspidophroroides olrikii*. While this confirms the earlier placement of *Hemilepidotus* within the family of Agonidae [94,95], the evolution of the sucking disc likely arose in the Liparidae and Cyclopteridae, and was subsequently lost in specific lineages of the Liparidae.

Our analysis of selection pressures operating on the mitogenome of broader Cottoidei revealed selection constraints. We found that purifying (negative) selection was the predominant force in the evolution of mitogenomes of Cottoidei fish. Nonetheless, there was evidence for weak and/or episodic positive selection occurring in this background of purifying selection. The 'three-ratio' model showed that the deep-sea marine lineage had a significantly higher ω (d_N/d_S) ratio than did the freshwater lineage of Cottales. As well as that, the RELAX analysis revealed significant evidence of relaxed selective pressures along the lineage of deep-sea fish, whereas selection was intensified in the freshwater lineage. Our findings were concordant with Shen et al. [93] in that the deep-sea lineage had accumulated more nonsynonymous mutations, and our study further revealed that this was due to relaxed selective constraints rather than to positive selection along this lineage. However, even with our denser dataset, the positive linear relationship between the magnitude of natural selection and increasing depth alluded to Shen et al. [93] could not be confirmed for cottoid fish. In fact, Shen et al. [92] showed that even though Perciformes had at least three independent events of deep-sea adaptation, there were no deep-sea lineages with significant signals of positive selection, corroborating our findings. Marine and freshwater environments have marked differences in salinity, the frequency of anoxic events, dissolved oxygen levels, pH, and amounts of hydrogen sulphide. Under those extreme environmental conditions, survival requires a modified and adapted energy metabolism, or assistance provided by mutualistic organisms that enable colonization of this niche.

Having full and completely annotated mitochondrial DNA genomes for both commercially important species such as *Cyclopterus* and rarely observed marine ecosystem species like the hadal snailfish *Pseudoliparis* enables effective management strategies, and provides the capacity to monitor ecological ecosystems non-invasively through the use of environmental DNA (eDNA) techniques. Difficulties in the accurate assembly of these fish mitogenomes highlight the utility of long-read sequencing technologies like Oxford Nanopore and PacBio sequencing. Whilst a handful of mitogenomes have been assembled in this study, the reference genomes combined with eDNA analysis are likely to open new avenues to study population structure, selection pressure and reproduction, or provide capabilities for species detection. Having the mitochondrial genomes described here will help with the significant difficulties posed by assessing the presence and identity of organisms in deep-sea ecosystems.

4. Materials and methods

4.1. Fish sampling and genomic DNA extraction

We obtained fin clips from individuals of *C. lumpus* caught and sacrificed during commercial fishing activity in Austevoll (N 60.02 E 05.16), southwestern Norway. As this was a commercial fishing operation, no ethics committee approval was necessary. We stored the samples at 4°C in absolute ethanol until total genomic DNA was extracted using the DNeasy Tissue Kit following the manufacturer's instructions (Qiagen; Cat. No. 69504). The Qubit Broad Range dsDNA assay (Thermofisher Scientific; Cat. No. Q32853) was used to quantify the extracted DNA.

4.2. Illumina and nanopore library preparation

Three short-read paired-end DNA libraries of differing insert sizes were prepared (using sample No. 314), each with 100 ng of DNA as starting material using the MuSeek library preparation kit (Cat. No. K1361; Thermo Fisher Scientific). Purification was performed at intermediate steps with 1.1× volume of Agencourt AMPure XP beads (Cat. No. A63881; Beckman Coulter). Amplification and barcoding of the libraries was completed with separate MuSeek barcodes (Cat. No. K1551), using either the Platinum PCR Supermix High Fidelity polymerase (Cat. No. 12532-016; Thermo Fisher Scientific; in 130 μl volumes; 1 library), or with Phusion HotStart II polymerase (part of Cat. No. K1361; in 50 μl volumes; 2 libraries) for 8 to 12 cycles. Reactions were purified with AMPure XP, and the three libraries were size-selected using a pre-cast 1.5% agarose gel cassette containing ethidium bromide (Sage Science; Cat. No. CSD1510) on the Pippin Prep system. The short and long insert libraries were quantified and quality checked on a BioAnalyzer 2100 High Sensitivity chip (Agilent; Cat. No. 5067-4626). These had a total combined insert size range from 455 to 830 bp. The individual libraries were then diluted to a concentration of 4 nM before being populated on a MiSeq flow cell as per the manufacturer's instructions for paired-end sequencing using the MiSeq Reagent Kit v3 for 600 cycles (Cat. No. MS-102-3003), or the MiSeq Reagent Kit v2 for 300 (MS-102-2002) or 500 cycles (Cat. No. MS-102-2003). Sequencing data were demultiplexed with the MiSeq instrument.

A gap remained in the intergenic space between $tRNA^{Leu2}$ and *ND1* after MiSeq sequencing (**Fig. S1**). This required a different strategy for gap closing (Fig. 1), so a Nanopore amplicon sequencing library was prepared using the primers Cl_mt_F_478 and Cl_mt_R_3754 using DNA from two different samples (**Table S1**). The 3.4 kb amplicon was sequenced on an Oxford Nanopore Flongle flow cell (R9.4.1, FLO-FLG001) using the Ligation Sequencing Kit (SQK-LSK109) protocol to attach sequencing adapters. The final step before flow cell loading included washing the DNA library with LFB buffer as described in the SQK-LSK109 protocol. Long reads were signal processed with MinKNOW v4.1.2 and Guppy v4.2.2 (https://github.com/nanoporetech/). The purified PCR product was also Sanger sequenced (Eurofins) as described in the Supplementary Information.

4.3. Mitogenome assemblies and annotation

A total of 20.2 million reads generated by the MiSeq sequencer were used in the assembly of the *C. lumpus* mitogenome. Two main methods of independent assembly from quality-filtered and trimmed reads were performed with CLC Genomics Workbench v. 12.x (www.qiagenbioinformatics.com) and the software package SPAdes v3.12 and v3.13 [7,11]. For CLC Genomics Workbench assemblies, the trim quality limit was set to 0.01, and the number of ambiguous bases set to 1, prior to de novo assembly with scaffolding and read mapping (options were changed for the length fraction of 0.9, and similarity fraction of 0.92). The de novo assembly was then queried by BLAST. The complete data set was also independently de novo assembled with the SPAdes assembler (v13.1; [7]), with the following parameters; *spades.py –k* 21,33,55,77,99,127 *–careful –large –dataset* SPAdes_lumpfish_input.yaml *–o* SPAdes_lumpfish_output *-t* 32 *–m* 120. The complete output in GFA format was then viewed in the software package BANDAGE v0.8.1 [115], and mitogenomic sequences identified by an integrated BLAST analysis within BANDAGE.

BLAST identification of the *C. lumpus* mitogenome used the *M. scorpius* and *Anoplopoma fimbria* mitogenomes as query sequences (Table 1). The *M. scorpius* mitogenomic sequence had been identified after a separate assembly of data from the NCBI Sequence Read Archive (SRA; ERR1473910; [63]) using a similar strategy to *C. lumpus*. While a mitogenomic sequence was found in the NCBI assembly (accession number GCA_900312955.1; Dryad), the independent assembly with SPAdes and CLC Genomics provided a complete and corrected sequence (Table 1). We additionally compared the mitogenomes of *Cyclopterus* and *Aptocyclus* to published genomic data from other Liparidae species, together with their respective supporting information from the NCBI SRA database. Significant misassemblies were found for the mitogenomes of *Pseudoliparis swirei* [108], and *C. lumpus* (please see Supplementary Information for methods, data, and accession numbers). In these cases, the published genomes used FALCON [22] and MaSuRCA [120] as primary assemblers (respectively), and the mitochondrial genomes had been incorrectly merged into the published autosomal scaffolds (see **Supplementary Information S1**). Furthermore, despite the recent genome assembly of L. *tanakae* from 851 Gb of clean sequence [108], the mitochondrial genome was found across more than 20 scaffolds. Therefore, we inferred the full mitogenome of L. *tanakae* by assembly with the CLC Genomics Workbench from raw data (PRJNA472846; see **Supplementary Information S2**).

Mitogenome sequences were assessed for completeness and low-coverage areas before and after circularization. Due to the difficulty of assembling the *CR* and the intergenic spacer between $tRNA^{Leu2}$ and *ND1*, different assemblies and read-mapping methods were assessed, and the resulting assemblies compared to alignments in MUSCLE v. 3.8.31 [29], and/or analysed by dot-plot analysis in CLC Genomics Workbench v. 12.x.

Prediction and annotation of the protein-coding genes (PCGs), ribosomal RNA (rRNA) genes, transfer RNA (tRNA) genes, and non-coding regions of the mtDNA sequences, was performed by two different methods due to differences in prediction (Fig. 1). Firstly, we used the automated online MitoFish MitoAnnotator v. 3.4.1 tool (http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html, [42]). Secondly, we ran the annotation and visualization modules of MitoZ separately (v2.3; [65]). Errors and incomplete stop codons were identified in both MitoFish MitoAnnotator pipeline and by the MitoZ annotate module. These were reconciled by examining the liparid and cyclopterid nucleotide alignments (in MUSCLE), with their respective amino acid codon usage, and if necessary, by then re-examining the original read map alignments in CLC Genomics Workbench v. 21.0 to detect errors in assembly or non-canonical alternative protein coding terminations. Circular maps of complete mitochondrial genomes were generated using the beta version of the CGView server (http://cgview.ca; [34]).

To infer the secondary (cloverleaf) structure of putative tRNAs, we used ARWEN v. 1.2.3 [59] and the tRNAscan-SE web server v. 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/, [61]), following the generalized vertebrate mitochondrial tRNA settings.

We calculated the A + T content of sequences and relative synonymous codon usage (RSCU) using MEGA v. 7.0 [56]. To measure the base composition skewness of nucleotide sequences, we used the formulae of Perna and Kocher [80]: AT-skew = $[A\text{-}T]/[A + T]$ and GC-skew = $[G\text{-}C]/[G + C]$. We manually counted the overlapping regions and intergenic spacers between PCGs, rRNAs, tRNAs, and non-coding regions. We deposited the complete annotated mitochondrial genomic DNA sequences of *C. lumpus* and the reassembled mitogenomes of additional Liparidae species into the NCBI GenBank database under the accession numbers listed in Table 1.

4.4. Sequence alignment

We obtained 65 mitochondrial genomes of fish within the suborder Cottoidei sensu Betancur-R et al. [14] from GenBank (Table 1). We included only one representative per valid species (longest mitogenome sequence) when more than one known sequence was available on GenBank. The ingroup taxa, including the one sequenced in this study, represent five infraorders and 14 families (Table 1). After a series of preliminary analyses, we selected three outgroup taxa for phylogenetic analyses: the pearlfish *Carapus bermudensis* (Jones 1874; order Ophidiiformes), the bowfin *Amia calva* (Linnaeus 1766; order Amiiformes) and the spotted gar *Lepisosteus oculatus* (Winchell 1864; order Lepisosteiformes).

Alignments of protein-coding genes were guided by amino acid translations and codon position as implemented in the online version of TRANSLATORX v. 1.0 (http://translatorx.co.uk/; [1]). Briefly, nucleotide sequences were first translated into amino acid sequences using the vertebrate mitochondrial genetic code, and these were then aligned using the MAFFT L-INS-i algorithm (accurate for alignment of $\leq$200 sequences; [48,50]) implemented in TRANSLATORX. The amino acid alignments were then reverted into the original nucleotide sequences, and the remaining ambiguously aligned sites were removed using GBLOCKS v. 0.19b [20] also implemented in TRANSLATORX with default settings. We checked for the correct translation of nucleotide sequences to amino acids in GENEIOUS PRIME v. 2019.0.4 [51]. We aligned rRNA genes using the online version of MAFFT v. 7.299 (https://mafft.cbrc.jp/alignment/server/; [48,50]), with the Q-INS-i algorithm for ribosomal fragments [49], and removed ambiguously aligned sites using the online version of GBLOCKS with default settings. Before phylogenetic analysis, we concatenated the aligned individual PCGs and rRNA datasets using GENEIOUS PRIME. The dataset was initially gene-partitioned into 41 partitions. We designated two partitions for the rRNA genes (*12S* and *16S*, treated each as a single partition) and 39 partitions covering the three codon positions in each of the 13 protein coding genes. We used ModelFinder v. 1.6.12 ([47] to select the best-fitting partitioning scheme and models of evolution using the corrected Akaike Information Criterion (AICc) and the edge-linked partition model [21] as implemented in IQ-TREE v. 2.1.3 [67]. We opted for the new model selection procedure (*−m MF + MERGE*), which additionally implements the FreeRate heterogeneity model inferring the site rates directly from the data instead of being drawn from a gamma distribution [96]. The top 30% partition schemes were checked using the relaxed clustering algorithm (*−rcluster* 30), as described in Lanfear et al. [58]. The best-fit partitioning scheme contained 11 partitions (Table 4).

4.5. Phylogeny reconstruction and molecular dating

We reconstructed phylogenies based on the Maximum-likelihood (ML) criterion in IQ-TREE and Bayesian inference (BI) in MRBAYES v. 3.2.6 [38,86]. In both approaches, we used the substitution models indicated by ModelFinder (Table 4). For ML analysis, we used the Nearest Neighbor Interchange (NNI) approach to search for tree topology and for computing branch supports with 1000 replicates of the Shimodaira-Hasegawa approximate likelihood-ratio test (*SH-aLRT*; [5]) and 1000 bootstrapped replicates of the ultrafast bootstrapping (*UFBoot2*) approach [37]. For BI, we ran a pair of independent searches for 10 million generations, with trees saved every 1000 generations and the first 2500 sampled trees of each search discarded as burn-in. Finally, a consensus tree showing all compatible groupings was constructed. We performed the BI analysis on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway portal v. 3.3 (www.phylo.org) at the San Diego Supercomputer Center [66].

We estimated divergence times using BEAST v. 2.6.6 [17], implementing a Birth–Death tree prior [84] and the uncorrelated lognormal relaxed molecular clock model [27]. We conducted molecular dating analysis based on the PCGs-rRNAs partitioned dataset using the package bModelTest v. 1.1.2 (part of the BEAST package; [16]) to calculate the best substitution model for each partition, excluding the outgroup taxa. bModelTest estimates the phylogeny and the substitution models jointly using a reversible jump Monte Carlo Markov Chain (rjMCMC) algorithm, which allows the chain to analyze substitution models with different numbers of parameters [16]. For computational efficiency, we opted to limit our rjMCMC search to models containing different transition-transversion rates, which includes the Jukes-Cantor model where all rates are equal, the General Time Reversible (GTR) model where all rates are different, and all models where the rate of transitions is different from the rate of transversions—a total of 31 models (see [16] and details in their Additional file 1: Appendix). Moreover, bModelTest estimates the posterior probability for gamma-distributed rate heterogeneity among sites (+Γ), a proportion of invariable sites (+I) and unequal base frequencies. bModelTest requires a priori selection of the partitioning scheme, therefore, we used the 11 partitions identified by MODELFINDER as a prior partition scheme for the analysis. Because the fossil record of Cottoidei is sparse, we employed a reliable fossil from the family Gasterosteidae. The oldest known Gasterosteidae fossil belongs to the Stickleback *Gasterosteus aculeatus* (Linnaeus 1758) species complex from the North East Pacific Ocean in the Serravallian (13.0–13.3 MYA), the middle age of the Miocene epoch [12]. Although the fossil record of Gasterosteidae is restricted to the Neogene epoch, the maximum age for the crown node of Gasterosteidae is suggested to be 41.30 MYA [78,93]. We applied a fossil calibration prior to the well-supported crown node of Gasterosteidae by setting a log-normally distributed age prior (M = 2.395, S = 0.72 and *offset* = 13) specifying a distribution centered at about 27.15 million years with a standard deviation of about 0.5 million years. After a preliminary series of test runs, we conducted the analysis with default parameters, while the MCMC chain was run for 40 million generations with sampling every 4000 generations, a 10% burn-in. To ensure that runs had converged, we performed the same run three times. Log and tree files were combined using LOGCOMBINER v. 2.6.6 (part of the BEAST package), and TRACER v. 1.7.1 [83] was used to confirm that

Table 4
Dataset partition schemes.

Partition	Best-fit substitution model	
	ModelFinder	bModelTest
ND1_pos1_ND3_pos1_ND4L_pos1_CYTB_pos1	GTR + F + I + G4	TVM + I + G4
ND1_pos2_ND2_pos2_ATP8_pos2_ND3_pos2_ND4L_pos2_ND4_pos2_ND5_pos2_ND6_pos2	GTR + F + R8	GTR + I + G4
ND1_pos3_ATP6_pos3_ND5_pos3_CYTB_pos3	TN + F + R8	TIM + I + G4
ND2_pos1_ATP6_pos1_ND4_pos1_ND5_pos1	GTR + F + I + G4	TIM + I + G4
ND2_pos3_ND4_pos3	TIM3 + F + I + G4	TN93 + I + G4
COI_pos1_COII_pos3_COIII_pos3_ND3_pos3	TN + F + R7	TIM + I + G4
COI_pos2_COII_pos1_COIII_pos1	GTR + F + R3	TVM + I + G4
COI_pos3_COII_pos2_ATP6_pos2_COIII_pos2_CYTB_pos2	GTR + F + R3	TIM + I + G4
ATP8_pos1_ATP8_pos3_ND4L_pos3_ND6_pos1	SYM + I + G4	TIM + I + G4
ND6_pos3	HKY + F + G4	TN93 + I + G4
12SrRNA_16SrRNA	GTR + F + R10	TN93 + I + G4

chains had reached stationarity and that estimated sample sizes (ESS) of sampled parameters were > 200. TREEANNOTATOR v. 2.6.6 (part of the BEAST package) was then used to summarize the post burn-in set of trees and to generate a maximum clade credibility chronogram showing mean divergence time estimates with 95% highest posterior density (HPD) intervals. Finally, we visualized the results in FIGTREE v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

4.6. Selection analyses

We test the alternate hypotheses of neutral evolution versus adaptive evolution by inferring the d_N/d_S ratio (ω), which compares the number of nonsynonymous substitutions per nonsynonymous site (d_N) to the number of synonymous substitutions per synonymous site (d_S) across the 13 PCGs of the mitogenome [72,116]. We searched for selection operating on the 13 mitochondrial PCGs by fitting codon models within a phylogenetic framework using the concatenated sequences of coding regions of all PCGs and individual genes while excluding stop codons. To assess selection on a codon-by-codon basis, we employed three approaches implemented in CODEML (PAML v. 4.9; [117]), and HYPHY as implemented in the webserver DATAMONKEY v. 2.0 (https://www.datamonkey.org/; [81,111]).

First, we applied the ML approach in CODEML to conduct branch and branch-site models [119]. To understand if specific lineages are more likely to be characterised by a ω value differing from the rest of the phylogeny, the branch-model M0 ('one-ratio' model, all branches have the same ω), was compared with two other branch models, BM2 and BM3. In the latter models, BM2 is a 'two-ratio' model, where a single branch of interest has a different ω from the background ratio of the phylogeny, and BM3 is a 'three-ratio' model, where two branches of interest have a different ω from the background ratio of the phylogeny. Here, we compared the BM2 and BM3 models to M0 to investigate if the ω of freshwater and deep-sea marine lineages of Cottales were significantly greater than that of other species. Finally, in order to compare our results against previous findings by Shen et al. [93], we also fitted the branch model M1, a 'free-ratio' model, which assumes that branches have independent ω. Then we constructed a linear model using M1 ω estimates as response variable and the maximum recorded depth for a taxa as the predictor variable. To identify whether specific lineages evolve at different rates and potentially could be under unique signatures of selection, branch-site models, implemented in CODEML (Model A and Model A_{null}) were used, which allow ω to vary among lineages [119]. When the LRT was significant ($0.01 < p < 0.05$), we used the Bayes Empirical Bayes (BEB; [118]) methods to identify amino acid residues that were likely to evolve under positive selection based on a posterior probability threshold of 0.95.

Second, we used the Fast-Unconstrained Bayesian Approximation (FUBAR) implemented in HYPHY, for estimating the number of nonsynonymous and synonymous substitutions at each codon given a phylogeny [70]. Moreover, FUBAR computes the posterior probability (*PP*) of every codon belonging to a set of classes of ω (including $\omega = 1$, $\omega < 1$ or $\omega > 1$) [70]. We implemented FUBAR such that the codons with a $PP > 0.9$ and $\omega < 1$ or $\omega > 1$ are inferred to be evolving putatively under pervasive negative or positive selection, respectively, across all branches of the predefined phylogeny.

Last, we executed the Mixed Effects Model of Evolution (MEME; [70,71]) also implemented in HYPHY, for estimating the probability for a codon to have undergone episodes of positive evolution, allowing the ω ratio distribution to vary across codons and branches in the phylogeny. We applied the MEME method such that codons evolving under episodic selection in some lineages are inferred to be significant under $p < 0.05$.

4.7. RELAX analysis

We tested for differences in selective pressures between freshwater and marine lineages using the RELAX method [113] as implemented in the webserver DATAMONKEY. RELAX is a hypothesis-testing framework that determines whether the strength of natural selection has been relaxed or intensified along a specified set of test branches [113]. Thus, it allows for identifying trends and/or shifts in the stringency of natural selection on a given gene between lineages and should not be used for explicitly testing for positive selection [113]. This test performs a likelihood ratio test (LRT) by comparing the null model in which the relation parameter (k; selection intensity) is constrained to 1 for all branches, to the alternative model in which k is a free, branch-specific, parameter, and one subset of branches is designated as the reference ($k = 1$). RELAX models each branch-site combination as a draw from a distribution of ω with three values ($\omega^- < 1$, $\omega^N = 1$ or $\omega^+ > 1$). For a specific set of branches (test branches), these values are modified by the k parameter via $\omega_T = \omega^k$. Hence, estimates of $k > 1$ imply intensification of selection, while $k < 1$ implies a relaxation of selection, relative to the reference subset of branches. In our analysis, branches of the freshwater lineage of Cottales were assigned as test branches, and branches of the marine lineages of Cottales, inhabiting variable bathymetric ranges, were assigned as reference branches.

Authorship contribution statement

SM, AVS, SBH, ODBJ and AKDI conceived and designed the study. AVS performed the library preparation, sequencing and assembly of the mitochondrial genomes. SM and AVS performed the bioinformatics analysis for the annotation of the mitochondrial genomes, wrote, and edited the first draft of the manuscript. SM conducted the molecular phylogenetic analysis with inputs from TN. SBH, ODBJ and AKDI provided the funding for the project. All other co-authors reviewed and polished the paper.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgement

The financial assistance Lerøy Seafood Group, Research Council of Norway (RFFNord project 282460) and the Icelandic Research Council (Rannís, 186971-0611) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the funding bodies. We also thank the National e-Infrastructure for Research Data (UNINETT Sigma2) for providing high-end computing resources under project number NN9614K.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110297.

References

[1] F. Abascal, R. Zardoya, M.J. Telford, TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations, Nucleic Acids Res. 38 (suppl_2) (2010) W7–W13, https://doi.org/10.1093/nar/gkq291.
[2] T. Abe, H. Munehara, Adaptation and evolution of reproductive mode in copulating cottoid species, in: Reproductive Biology and Phylogeny of Fishes (Agnathans and Bony Fishes), 2009, pp. 221–246, https://doi.org/10.1201/b10257-7.
[3] E.A. Alacs, A. Georges, N.N. FitzSimmons, J. Robertson, DNA detective: a review of molecular approaches to wildlife forensics, Forensic Sci. Med. Pathol. 6 (3) (2010) 180–194, https://doi.org/10.1007/s12024-009-9131-7.
[4] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I. C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and organisation of the human mitochondrial genome, Nature 290 (5806) (1981) 457–465, https://doi.org/10.1038/290457a0.

[5] M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative, Syst. Biol. 55 (4) (2006) 539–552, https://doi.org/10.1080/10635150600755453.

[6] W.J. Ansorge, Next-generation DNA sequencing techniques, New Biotechnol. 25 (4) (2009) 195–203, https://doi.org/10.1016/j.nbt.2008.12.009.

[7] D. Antipov, A. Korobeynikov, J.S. McLean, P.A. Pevzner, hybridSPAdes: an algorithm for hybrid assembly of short and long reads, Bioinformatics (Oxford, England) 32 (7) (2016) 1009–1015, https://doi.org/10.1093/bioinformatics/btv688.

[8] J.C. Avise, Phylogeography: retrospect and prospect, J. Biogeogr. 36 (1) (2009) 3–15, https://doi.org/10.1111/j.1365-2699.2008.02032.x.

[9] I. Bakke, G.F. Shields, S. Johansen, Sequence characterization of a unique intergenic spacer in Gadiformes mitochondrial DNA, Mar. Biotechnol. (New York, N.Y.) 1 (5) (1999) 411–415, https://doi.org/10.1007/PL00011797.

[10] J.W.O. Ballard, M.C. Whitlock, The incomplete natural history of mitochondria, Mol. Ecol. 13 (4) (2004) 729–744, https://doi.org/10.1046/j.1365-294X.2003.02063.x.

[11] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (5) (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.

[12] M.A. Bell, J.D. Stewart, P.J. Park, The world's oldest fossil threespine stickleback fish, Copeia 2009 (2) (2009) 256–265, https://doi.org/10.1643/CG-08-059.

[13] R. Betancur-R, R.E. Broughton, E.O. Wiley, K. Carpenter, J.A. López, C. Li, N. I. Holcroft, D. Arcila, M. Sanciangco, J.C. Cureton Ii, F. Zhang, The tree of life and a new classification of bony fishes, PLoS Curr. Tree Life 5 (2013), https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

[14] R. Betancur-R, E.O. Wiley, G. Arratia, A. Acero, N. Bailly, M. Miya, G. Lecointre, G. Orti, Phylogenetic classification of bony fishes, BMC Evol. Biol. 17 (1) (2017) 162, https://doi.org/10.1186/s12862-017-0958-3.

[15] J.L. Boore, Animal mitochondrial genomes, Nucleic Acids Res. 27 (8) (1999) 1767–1780, https://doi.org/10.1093/nar/27.8.1767.

[16] R.R. Bouckaert, A.J. Drummond, bModelTest: Bayesian phylogenetic site model averaging and model comparison, BMC Evol. Biol. 17 (1) (2017) 42, https://doi.org/10.1186/s12862-017-0890-6.

[17] R. Bouckaert, T.G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, A.J. Drummond, BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis, PLoS Comput. Biol. 15 (4) (2019), https://doi.org/10.1371/journal.pcbi.1006650 e1006650.

[18] G.G. Brown, G. Gadaleta, G. Pepe, C. Saccone, A.E. Sbis, Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA, J. Mol. Biol. 192 (3) (1986) 503–511, https://doi.org/10.1016/0022-2836(86)90272-X.

[19] M.S. Busby, D.M. Blood, A.J. Fleischer, D.G. Nichol, Egg deposition and development of eggs and larvae of Bigmouth Sculpin (*Hemitripterus bolini*), Northwest. Nat. (Olympia, Wash.) 93 (1) (2012) 1–16, https://doi.org/10.1898/11-13.1.

[20] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (4) (2000) 540–552, https://doi.org/10.1093/oxfordjournals.molbev.a026334.

[21] O. Chernomor, A. von Haeseler, B.Q. Minh, Terrace aware data structure for phylogenomic inference from supermatrices, Syst. Biol. 65 (6) (2016) 997–1008, https://doi.org/10.1093/sysbio/syw037.

[22] C.S. Chin, P. Peluso, F.J. Sedlazeck, M. Nattestad, G.T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G.R. Cramer, M. Delledonne, C. Luo, J.R. Ecker, D. Cantu, D.R. Rank, M.C. Schatz, Phased diploid genome assembly with single-molecule real-time sequencing, Nat. Methods 13 (12) (2016) 1050–1054, https://doi.org/10.1038/nmeth.4035.

[23] S. Consuegra, E. John, E. Verspoor, C.G. de Leaniz, Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species, Genet. Sel. Evol. 47 (1) (2015) 58, https://doi.org/10.1186/s12711-015-0138-0.

[24] A.C. Dalziel, C.D. Moyes, E. Fredriksson, S.C. Lougheed, Molecular evolution of cytochrome c oxidase in high-performance fish (Teleostei: Scombroidei), J. Mol. Evol. 62 (3) (2006) 319–331, https://doi.org/10.1007/s00239-005-0110-7.

[25] K. Deiner, H.M. Bik, E. Mächler, M. Seymour, A. Lacoursière-Roussel, F. Altermatt, S. Creer, I. Bista, D.M. Lodge, N. Vere, M.E. Pfrender, L. Bernatchez, Environmental DNA metabarcoding: transforming how we survey animal and plant communities, Mol. Ecol. 26 (21) (2017) 5872–5895, https://doi.org/10.1111/mec.14350.

[26] K. Deiner, M.A. Renshaw, Y. Li, B.P. Olds, D.M. Lodge, M.E. Pfrender, Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA, Methods Ecol. Evol. 8 (12) (2017) 1888–1898, https://doi.org/10.1111/2041-210X.12836.

[27] A.J. Drummond, S.Y.W. Ho, M.J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence, PLoS Biol. 4 (5) (2006), https://doi.org/10.1371/journal.pbio.0040088 e88.

[29] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (5) (2004) 1792–1797, https://doi.org/10.1093/nar/gkh340.

[30] N. Galtier, B. Nabholz, S. Glémin, G.D.D. Hurst, Mitochondrial DNA as a marker of molecular diversity: a reappraisal, Mol. Ecol. 18 (22) (2009) 4541–4550, https://doi.org/10.1111/j.1365-294X.2009.04380.x.

[31] E. Garcia-Mayoral, M. Olsen, R. Hedeholm, S. Post, E.E. Nielsen, D. Bekkevold, Genetic structure of West Greenland populations of lumpfish Cyclopterus lumpus, J. Fish Biol. 89 (6) (2016) 2625–2642, https://doi.org/10.1111/jfb.13167.

[32] M.R. Garvin, J.P. Bielawski, A.J. Gharrett, Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon, PLoS One 6 (9) (2011), https://doi.org/10.1371/journal.pone.0024127 e24127.

[33] M.E. Gerringer, T.D. Linley, A.J. Jamieson, E. Goetze, J.C. Drazen, *Pseudoliparis swirei sp. nov.*: a newly-discovered hadal snailfish (Scorpaeniformes: Liparidae) from the Mariana Trench, Zootaxa 4358 (1) (2017) 161–177, https://doi.org/10.11646/zootaxa.4358.1.7.

[34] J.R. Grant, P. Stothard, The CGView server: a comparative genomics tool for circular genomes, Nucleic Acids Res. 36 (suppl_2) (2008) W181–W184, https://doi.org/10.1093/nar/gkn179.

[35] M. Hajibabaei, G.A. Singer, P.D. Hebert, D.A. Hickey, DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends Genet. 23 (4) (2007) 167–172, https://doi.org/10.1016/j.tig.2007.02.001.

[36] T.M. Healy, R.S. Burton, Strong selective effects of mitochondrial DNA on the nuclear genome, Proc. Natl. Acad. Sci. U. S. A. 117 (12) (2020) 6616–6621, https://doi.org/10.1073/pnas.1910141117.

[37] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. (2017), https://doi.org/10.1093/molbev/msx281.

[38] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics 17 (8) (2001) 754–755, https://doi.org/10.1093/bioinformatics/17.8.754.

[39] A.K.D. Imsland, A. Hanssen, A.V. Nytrø, P. Reynolds, T.M. Jonassen, T. A. Hangstad, T.A. Elvegård, T.C. Urskog, B. Mikalsen, It works! Lumpfish can significantly lower sea lice infestation in large-scale salmon farming, Biol. Open 7 (9) (2018), https://doi.org/10.1242/bio.036301 bio036301.

[40] A.K.D. Imsland, P. Reynolds, T.A. Hangstad, L. Kapari, S.N. Maduna, S.B. Hagen, Ó.D.B. Jónsdóttir, F. Spetland, K.S. Lindberg, Quantification of grazing efficacy, growth and health score of different lumpfish (*Cyclopterus lumpus* L.) families: possible size and gender effects, Aquaculture 530 (2021) 735925, https://doi.org/10.1016/j.aquaculture.2020.735925.

[41] A.K. Imsland, P. Reynolds, G. Eliassen, T.A. Hangstad, A. Foss, E. Vikingstad, T. A. Elvegård, The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.), Aquaculture 424 (2014) 18–23, https://doi.org/10.1016/j.aquaculture.2013.12.033.

[42] W. Iwasaki, T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, M. Nishida, MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline, Mol. Biol. Evol. 30 (11) (2013) 2531–2540, https://doi.org/10.1093/molbev/mst141.

[43] M.W. Jacobsen, R.R. Da Fonseca, L. Bernatchez, M.M. Hansen, Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (*Coregonus* ssp.), Mol. Phylogenet. Evol. 95 (2016) 161–170, https://doi.org/10.1016/j.ympev.2015.11.008.

[44] E. Jemt, Ö. Persson, Y. Shi, M. Mehmedovic, J.P. Uhler, M. Dávila López, C. Freyer, C.M. Gustafsson, T. Samuelsson, M. Falkenberg, Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element, Nucleic Acids Res. 43 (19) (2015) 9262–9275, https://doi.org/10.1093/nar/gkv804.

[45] S. Johansen, T. Johansen, Sequence analysis of 12 structural genes and a novel non-coding region from mitochondrial DNA of Atlantic cod, Gadus morhua, Biochim. Biophys. Acta (BBA) Gene Struct. Expr. 1218 (2) (1994) 213–217, https://doi.org/10.1016/0167-4781(94)90015-9.

[46] Ó.D.B. Jónsdóttir, J. Schregel, S.B. Hagen, C. Tobiassen, S.G. Aarnes, A. K. Imsland, Population genetic structure of lumpfish along the Norwegian coast: aquaculture implications, Aquac. Int. 26 (1) (2018) 49–60, https://doi.org/10.1007/s10499-017-0194-2.

[47] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (6) (2017) 587–589, https://doi.org/10.1038/nmeth.4285.

[48] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[49] K. Katoh, H. Toh, Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework, BMC Bioinforma. 9 (1) (2008) 1–13, https://doi.org/10.1186/1471-2105-9-212.

[50] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Brief. Bioinform. 20 (4) (2019) 1160–1166, https://doi.org/10.1093/bib/bbx108.

[51] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organisation and analysis of sequence data, Bioinformatics (Oxford, England) 28 (12) (2012) 1647–1649, https://doi.org/10.1093/bioinformatics/bts199.

[52] A.P. Kinziger, R.M. Wood, D.A. Neely, Molecular systematics of the genus Cottus (Scorpaeniformes: Cottidae), Copeia 2005 (2) (2005) 303–311, https://doi.org/10.1643/CI-03-290R1.

[53] M.L. Knope, Phylogenetics of the marine sculpins (Teleostei: Cottidae) of the North American Pacific coast, Mol. Phylogenet. Evol. 66 (1) (2013) 341–349, https://doi.org/10.1016/j.ympev.2012.10.008.

[54] S.W. Knudsen, P.R. Møller, P. Gravlund, Phylogeny of the snailfishes (Teleostei: Liparidae) based on molecular and morphological data, Mol. Phylogenet. Evol. 44 (2) (2007) 649–666, https://doi.org/10.1016/j.ympev.2007.04.005.

[55] J. Korlach, G. Gedman, S.B. Kingan, C.S. Chin, J.T. Howard, J.N. Audet, L. Cantin, E.D. Jarvis, De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads, GigaScience 6 (10) (2017), https://doi.org/10.1093/gigascience/gix085 gix085.

[56] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874, https://doi.org/10.1093/molbev/msw054.

[57] Z. Lajbner, R. Pnini, M.F. Camus, J. Miller, D.K. Dowling, Experimental evidence that thermal selection shapes mitochondrial genome evolution, Sci. Rep. 8 (1) (2018) 1–12, https://doi.org/10.1038/s41598-018-27805-3.

[58] R. Lanfear, B. Calcott, D. Kainer, C. Mayer, A. Stamatakis, Selecting optimal partitioning schemes for phylogenomic datasets, BMC Evol. Biol. 14 (1) (2014) 82, https://doi.org/10.1186/1471-2148-14-82.

[59] D. Laslett, B. Canbäck, ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences, Bioinformatics 24 (2) (2008) 172–175, https://doi.org/10.1093/bioinformatics/btm573.

[60] T.D. Linley, A.L. Stewart, P.J. McMillan, M.R. Clark, M.E. Gerringer, J.C. Drazen, T. Fujii, A.J. Jamieson, Bait attending fishes of the abyssal zone and hadal boundary: community structure, functional groups and species distribution in the Kermadec, New Hebrides and Mariana trenches, Deep-Sea Res. I Oceanogr. Res. Pap. 121 (2017) 38–53, https://doi.org/10.1016/j.dsr.2016.12.009.

[61] T.M. Lowe, P.P. Chan, tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res. 44 (W1) (2016) W54–W57, https://doi.org/10.1093/nar/gkw413.

[62] S.N. Maduna, A. Vivian-Smith, Ó.D.B. Jónsdóttir, A.K. Imsland, C.F. Klütsch, T. Nyman, H.G. Eiken, S.B. Hagen, Genome-and transcriptome-derived microsatellite loci in lumpfish Cyclopterus lumpus: molecular tools for aquaculture, conservation and fisheries management, Sci. Rep. 10 (1) (2020) 1–11, https://doi.org/10.1038/s41598-019-57071-w.

[63] M. Malmstrøm, M. Matschiner, O.K. Tørresen, B. Star, L.G. Snipen, T.F. Hansen, H.T. Baalsrud, A.J. Nederbragt, R. Hanel, W. Salzburger, N.C. Stenseth, K. S. Jakobsen, S. Jentoft, Evolution of the immune system influences speciation rates in teleost fishes, Nat. Genet. 48 (10) (2016) 1204–1210, https://doi.org/10.1038/ng.3645.

[64] A. Margaryan, C.L. Noer, S.R. Richter, M.E. Restrup, J.L. Bülow-Hansen, F. Leerhøi, E.M.R. Langkjær, S. Gopalakrishnan, C. Carøe, M.T.P. Gilbert, K. Bohmann, Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark, Environ. DNA 3 (2) (2021) 472–480, https://doi.org/10.1002/edn3.138.

[65] G. Meng, Y. Li, C. Yang, S. Liu, MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization, Nucleic Acids Res. 47 (11) (2019) e63, https://doi.org/10.1093/nar/gkz173.

[66] M.A. Miller, W. Pfeiffer, T. Schwartz, The CIPRES science gateway: A community resource for phylogenetic analyses, in: Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery, 2011, July, pp. 1–8, https://doi.org/10.1145/2016741.2016785.

[67] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol. 37 (5) (2020) 1530–1534, https://doi.org/10.1093/molbev/msaa015.

[68] M. Miya, H. Takeshima, H. Endo, N.B. Ishiguro, J.G. Inoue, T. Mukai, T.P. Satoh, M. Yamaguchi, A. Kawaguchi, K. Mabuchi, S.M. Shirai, M. Nishida, Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences, Mol. Phylogenet. Evol. 26 (1) (2003) 121–138, https://doi.org/10.1016/S1055-7903(02)00332-9.

[69] M. Muñoz, Reproduction in Scorpaeniformes, Univ. Calif. Press, 2010.

[70] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S.L. Kosakovsky Pond, K. Scheffler, FUBAR: a fast, unconstrained bayesian approximation for inferring selection, Mol. Biol. Evol. 30 (5) (2013) 1196–1205, https://doi.org/10.1093/molbev/mst030.

[71] B. Murrell, J.O. Wertheim, S. Moola, T. Weighill, K. Scheffler, S.L.K. Pond, Detecting individual sites subject to episodic diversifying selection, PLoS Genet. 8 (7) (2012), https://doi.org/10.1371/journal.pgen.1002764 e1002764.

[72] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (5) (1986) 418–426.

[73] J.S. Nelson, T.C. Grande, M.V. Wilson, Fishes of the World, John Wiley & Sons, 2016, https://doi.org/10.1002/9781119174844.

[74] K. Oguri, T. Noguchi, Deepest Fish Ever Recorded, Documented at Depths of 8,178 m in Mariana Trench, 2017. JAMSTEC Press Release.

[75] K. Oku, H. Imamura, M. Yabe, Phylogenetic relationships and a new classification of the family Cyclopteridae (Perciformes: Cottoidei), Zootaxa 4221 (1) (2017) zootaxa-4221.

[76] J.W. Orr, I. Spies, D.E. Stevenson, G.C. Longo, Y. Kai, S. Ghods, M. Hollowed, Molecular phylogenetics of snailfishes (Cottoidei: Liparidae) based on MtDNA and RADseq genomic analyses, with comments on selected morphological characters, Zootaxa 4642 (1) (2019) 1–79, https://doi.org/10.11646/zootaxa.4642.1.1.

[77] C. Pampoulie, S. Skirnisdottir, G. Olafsdottir, S.J. Helyar, V. Thorsteinsson, S.Þ. Jónsson, A. Fréchet, C.M.F. Durif, S. Sherman, M. Lampart-Kałużniacka, R. Hedeholm, H. Ólafsson, A.K. Daníelsdóttir, J.M. Kasper, Genetic structure of the lumpfish *Cyclopterus lumpus* across the North Atlantic, ICES J. Mar. Sci. 71 (9) (2014) 2390–2397, https://doi.org/10.1093/icesjms/fsu071.

[78] C. Patterson, Osteichthyes: Teleostei, in: M.J. Benton (Ed.), The Fossil Record 2, Chapman & Hall, 1993, pp. 621–656.

[79] S.L. Pereira, Mitochondrial genome organisation and vertebrate phylogenetics, Genet. Mol. Biol. 23 (4) (2000) 745–752, https://doi.org/10.1590/S1415-47572000000400008.

[80] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J. Mol. Evol. 41 (3) (1995) 353–358, https://doi.org/10.1007/BF01215182.

[81] S.L.K. Pond, S.D. Frost, Datamonkey: rapid detection of selective pressure on individual sites of codon alignments, Bioinformatics 21 (10) (2005) 2531–2533, https://doi.org/10.1093/bioinformatics/bti320.

[82] I.G. Priede, R. Froese, Colonization of the deep sea by fishes, J. Fish Biol. 83 (6) (2013) 1528–1550, https://doi.org/10.1111/jfb.12265.

[83] A. Rambaut, A.J. Drummond, D. Xie, G. Baele, M.A. Suchard, Posterior summarisation in Bayesian phylogenetics using Tracer 1.7, Syst. Biol. 67 (2018), https://doi.org/10.1093/sysbio/syy032 syy032. Advance online publication.

[84] B. Rannala, Z. Yang, Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference, J. Mol. Evol. 43 (3) (1996) 304–311, https://doi.org/10.1007/BF02338839.

[85] H.C. Rees, B.C. Maddison, D.J. Middleditch, J.R. Patmore, K.C. Gough, The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology, J. Appl. Ecol. 51 (5) (2014) 1450–1459, https://doi.org/10.1111/1365-2664.12306.

[86] F. Ronquist, Bayesian inference of character evolution, Trends Ecol. Evol. 19 (9) (2004) 475–481, https://doi.org/10.1016/j.tree.2004.07.002.

[87] C. Saccone, M. Attimonelli, E. Sbisa, Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA, J. Mol. Evol. 26 (3) (1987) 205–211, https://doi.org/10.1007/BF02099853.

[88] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (12) (1977) 5463–5467, https://doi.org/10.1073/pnas.74.12.5463.

[89] Y. Sato, M. Miya, T. Fukunaga, T. Sado, W. Iwasaki, MitoFish and MiFish pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding, Mol. Biol. Evol. 35 (6) (2018) 1553–1555, https://doi.org/10.1093/molbev/msy074.

[90] T.P. Satoh, M. Miya, K. Mabuchi, M. Nishida, Structure and variation of the mitochondrial genome of fishes, BMC Genomics 17 (1) (2016) 1–20, https://doi.org/10.1186/s12864-016-3054-y.

[91] J.C. Schroeter, A.P. Maloy, C.B. Rees, M.L. Bartron, Fish mitochondrial genome sequencing: expanding genetic resources to support species detection and biodiversity monitoring using environmental DNA, Conserv. Genet. Resour. (2019) 1–14.

[92] X. Shen, Z. Pu, X. Chen, R.W. Murphy, Y. Shen, Convergent evolution of mitochondrial genes in deep-sea fishes, Front. Genet. 10 (2019) 10, https://doi.org/10.3389/fgene.2019.00925.

[93] Y. Shen, W. Dai, Z. Gao, G. Yan, X. Gan, S. He, Molecular phylogeny and divergence time estimates using the mitochondrial genome for the hadal snailfish from the Mariana trench, Sci. Bull. 61 (2017) 1106–1108, https://doi.org/10.1016/j.scib.2017.07.010.

[94] W.L. Smith, M.S. Busby, Phylogeny and taxonomy of sculpins, sandfishes, and snailfishes (Perciformes: Cottoidei) with comments on the phylogenetic significance of their early-life-history specializations, Mol. Phylogenet. Evol. 79 (2014) 332–352, https://doi.org/10.1016/j.ympev.2014.06.028.

[95] W.L. Smith, W.C. Wheeler, Polyphyly of the mail-cheeked fishes (Teleostei: Scorpaeniformes): evidence from mitochondrial and nuclear sequence data, Mol. Phylogenet. Evol. 32 (2) (2004) 627–646, https://doi.org/10.1016/j.ympev.2004.02.006.

[96] J. Soubrier, M. Steel, M.S. Lee, C. Der Sarkissian, S. Guindon, S.Y. Ho, A. Cooper, The influence of rate heterogeneity among sites on the time dependence of molecular rates, Mol. Biol. Evol. 29 (11) (2012) 3345–3358, https://doi.org/10.1093/molbev/mss140.

[98] Y.B. Sun, Y.Y. Shen, D.M. Irwin, Y.P. Zhang, Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution, Mol. Biol. Evol. 28 (1) (2011) 39–44, https://doi.org/10.1093/molbev/msq256.

[99] P. Sunnucks, H.E. Morales, A.M. Lamb, A. Pavlova, C. Greening, Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation, Front. Genet. 8 (2017) 25, https://doi.org/10.3389/fgene.2017.00025.

[100] M.K. Tilak, F. Justy, M. Debiais-Thibaud, F. Botero-Castro, F. Delsuc, E. J. Douzery, A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species, Conserv. Genet. Resour. 7 (1) (2015) 37–40, https://doi.org/10.1007/s12686-014-0338-x.

[101] M. Tobler, J.L. Kelley, M. Plath, R. Riesch, Extreme environments and the origins of biodiversity: adaptation and speciation in sulphide spring fishes, Mol. Ecol. 27 (4) (2018) 843–859, https://doi.org/10.1111/mec.14497.

[102] S.V. Turanov, Y.P. Kartavtsev, V.V. Zemnukhov, Molecular phylogenetic study of several eelpout fishes (Perciformes, Zoarcoidei) from Far Eastern seas on the basis of the nucleotide sequences of the mitochondrial cytochrome oxidase 1 gene (Co-1), Russ. J. Genet. 48 (2) (2012) 208–223, https://doi.org/10.1134/S1022795412020159.

[103] S.V. Turanov, Y.P. Kartavtsev, Y.H. Lee, D. Jeong, Molecular phylogenetic reconstruction and taxonomic investigation of eelpouts (Cottoidei: Zoarcales) based on Co-1 and Cyt-b mitochondrial genes, Mitochondrial DNA A DNA Mapp. Sequencing Anal. 28 (4) (2017) 547–557, https://doi.org/10.3109/24701394.2016.1155117.

[104] R. Van der Laan, Freshwater fish list, Almere, the Netherlands 20 (2019) 2019.
[105] R. Van Der Laan, W.N. Eschmeyer, R. Fricke, Family-group names of recent fishes, Zootaxa 3882 (1) (2014) 230, https://doi.org/10.11646/zootaxa.3882.1.1.
[106] R. Van Der Laan, W.N. Eschmeyer, R. Fricke, Family-group names of recent fishes, Zootaxa 3882 (1) (2018) 1–230, https://doi.org/10.11646/zootaxa.3882.1.1.
[107] O.S. Voskoboinikova, O.Y. Kudryavtseva, A.M. Orlov, S.Y. Orlova, M.V. Nazarkin, N.V. Chernova, O.A. Maznikova, Relationships and evolution of Lumpsuckers of the family Cyclopteridae (Cottoidei), J. Ichthyol. 60 (2020) 154–181, https://doi.org/10.1134/S0032945220020204.
[108] K. Wang, Y. Shen, Y. Yang, X. Gan, G. Liu, K. Hu, Y. Li, Z. Gao, L. Zhu, G. Yan, L. He, X. Shan, L. Yang, S. Lu, H. Zeng, X. Pan, C. Liu, Y. Yuan, C. Feng, S. He, Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation, Nat. Ecol. Evol. 3 (5) (2019) 823–833, https://doi.org/10.1038/s41559-019-0864-8.
[110] R.D. Ward, T.S. Zemlak, B.H. Innes, P.R. Last, P.D. Hebert, DNA barcoding Australia's fish species, Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1462) (2005) 1847–1857, https://doi.org/10.1098/rstb.2005.1716.
[111] S. Weaver, S.D. Shank, S.J. Spielman, M. Li, S.V. Muse, S.L. Kosakovsky Pond, Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes, Mol. Biol. Evol. 35 (3) (2018) 773–777, https://doi.org/10.1093/molbev/msx335.
[112] K. Weitemier, B.E. Penaluna, L.L. Hauck, L.J. Longway, T. Garcia, R. Cronn, Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding, Mol. Ecol. (2021), https://doi.org/10.1111/mec.15811 mec.15811.
[113] J.O. Wertheim, B. Murrell, M.D. Smith, S.L. Kosakovsky Pond, K. Scheffler, RELAX: detecting relaxed selection in a phylogenetic framework, Mol. Biol. Evol. 32 (3) (2015) 820–832, https://doi.org/10.1093/molbev/msu400.
[114] B.A. Whittaker, S. Consuegra, C.G. de Leaniz, Genetic and phenotypic differentiation of lumpfish (*Cyclopterus lumpus*) across the North Atlantic: implications for conservation and aquaculture, PeerJ 6 (2018), https://doi.org/10.7717/peerj.5974 e5974.
[115] R.R. Wick, M.B. Schultz, J. Zobel, K.E. Holt, Bandage: interactive visualization of de novo genome assemblies, Bioinformatics (Oxford, England) 31 (20) (2015) 3350–3352, https://doi.org/10.1093/bioinformatics/btv383.
[116] Z. Yang, Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution, Mol. Biol. Evol. 15 (5) (1998) 568–573, https://doi.org/10.1093/oxfordjournals.molbev.a025957.
[117] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (8) (2007) 1586–1591, https://doi.org/10.1093/molbev/msm088.
[118] Z. Yang, W.S. Wong, R. Nielsen, Bayes empirical Bayes inference of amino acid sites under positive selection, Mol. Biol. Evol. 22 (2005) 1107–1118, https://doi.org/10.1093/molbev/msi097.
[119] J. Zhang, R. Nielsen, Z. Yang, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level, Mol. Biol. Evol. 22 (2005) 2472–2479, https://doi.org/10.1093/molbev/msi237.
[120] A.V. Zimin, D. Puiu, M.C. Luo, T. Zhu, S. Koren, G. Marçais, J.A. Yorke, J. Dvořák, S.L. Salzberg, Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm, Genome Res. 27 (5) (2017) 787–792, https://doi.org/10.1101/gr.213405.116.