

1 ***Rare but not absent: the Inverted Mitogenomes of Deep-Sea Hatchetfish***

2

3 André Gomes-dos-Santos¹, Nair Vilas-Arrondo^{2,3}, André M. Machado¹, Esther
4 Román-Marcote³, Jose Luís Del Río Iglesias³, Francisco Baldó⁴, Montse Pérez³,
5 Miguel M. Fonseca¹, L. Filipe C. Castro^{1,5*} and Elsa Froufe^{1*}

6

7 ¹CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research,
8 University of Porto, Matosinhos, Portugal

9 ² Programa de Doctorado “Ciencias marinas, Tecnología y Gestión” (Do*MAR),
10 Montse Pérez Facultad de biología, Universidad de Vigo, Vigo, Spain

11 ³ Centro Oceanográfico de Vigo Instituto Español de Oceanografía (IEO, CSIC).
12 Subida a Radio Faro, 50. 36390 Vigo (Pontevedra), Spain

13 ⁴ Centro Oceanográfico de Cádiz, Instituto Español de Oceanografía (IEO, CSIC),
14 Puerto pesquero, dársena de Levante s/n, 11006 Cádiz, Spain

15 ⁵ Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal

16 * Corresponding authors: Elsa Froufe elsafroufe@gmail.com and L. Filipe C. Castro
17 lfilipecastro@gmail.com

18

19

20

21 **Abstract**

22 Mitochondrial genomes are by definition compact and structurally stable over aeons.
23 This generalized perception results from a vertebrate-centric vision, as very few
24 types of mtDNA rearrangements have been described in vertebrates. By combining a
25 panel of sequencing approaches, including short- and long-reads, we show that
26 species from a group of illusive marine teleosts, the deep-sea hatchetfish
27 (Stomiiforms: Sternoptychidae), display a myriad of new mtDNA structural
28 arrangements. We show a never reported inversion of the coding direction of protein-
29 coding genes (PCG) coupled with a strand asymmetry nucleotide composition
30 reversal directly related to the strand location of the Control Region (which includes
31 the heavy strand replication origin). An analysis of the 4-fold redundant sites of the
32 PCGs, in thousands of vertebrate mtDNAs, revealed the rarity of this phenomenon,
33 only found in 9 fish species, five of which are deep-sea hatchetfish. Curiously, in
34 Antarctic notothenioid fishes (Trematominae), where a single PCG inversion (the only
35 other record in fish) is coupled with the inversion of the Control Region, the standard
36 asymmetry is disrupted for the remaining PCG but not yet reversed, suggesting a
37 transitory state in this species mtDNA. Together, our findings hint that a relaxation of
38 the classic vertebrate mitochondrial structural *stasis*, observed in Sternoptychidae
39 and Trematominae, promotes disruption of the natural balance of asymmetry of the
40 mtDNA. Our findings support the long-lasting hypothesis that replication is the main
41 molecular mechanism promoting the strand-specific compositional bias of this unique
42 and indispensable molecule.

43

44

45 **Keywords:** mitogenome; strand asymmetry; deep-sea fish; vertebrate; mitochondrial

46 gene order; rearrangements.

47

48 **Introduction**

49 Counterintuitively, the mitochondrial genome (mtDNA) is far more variable than
50 normally recognized, including structure, gene content, order and orientation,
51 organization and mode of expression (Shtolz and Mishmar 2023). In Metazoa,
52 several types of mtDNA are described, including linear telomeric molecules (e.g.,
53 Cnidaria), “giant” circular molecules (e.g. some Bivalvia and Placozoa) and several
54 mini-circular molecules (e.g. Insecta) (Kolesnikov and Gerasimov 2012; Shtolz and
55 Mishmar 2023). In vertebrates, the classic mtDNA is described as a compact circular
56 molecule, maternally inherited, between 16–19 kbp and two compositionally distinct
57 strands, conventionally referred to as heavy (H-strand with high G composition) and
58 light (L-strand with low G composition) (Clayton 1982; Boore 1999; Kolesnikov and
59 Gerasimov 2012). The standard vertebrate mtDNA encodes 37 genes, 13 protein-
60 coding genes (PCG); two ribosomal RNAs (rRNAs); 22 transfer RNAs (tRNAs); and
61 two differently located and strand-specific replication origins (heavy strand replication
62 origin [OH] and light strand replication origin [OL]) (Clayton 1982; Lee et al. 1995;
63 Boore 1999; Kolesnikov and Gerasimov 2012). The OH is placed within a larger non-
64 coding region, the Control Region (CR), which includes several regulators of the
65 mtDNA replication and transcription and where four evolutionary conserved
66 sequence blocks are commonly found (i.e, CSB-I, CSB-II, CSB-III, and CSB-D)
67 (Clayton 1982; Lee et al. 1995; Boore 1999; Kolesnikov and Gerasimov 2012). The
68 13 mtDNA PCG, which follows the standard architecture (i.e., number, relative
69 positioning and coding strand), play key functional roles in the oxidative
70 phosphorylation (OXPHOS) cascade (Clayton 1982; Boore 1999; Kolesnikov and
71 Gerasimov 2012).

72 Exceptions to this generalized picture, in vertebrate classes, are extremely rare, with
73 very few and small-scale deviations (Gissi et al. 2010; Zhang et al. 2020; Formenti et
74 al. 2021; Montaña-Lozano et al. 2022; Sharbrough et al. 2023; Shtolz and Mishmar
75 2023). Among the different groups, birds and reptiles show the highest distribution of
76 rearranged genes, while mammals and fish show residual examples of
77 rearrangements (Zhang et al. 2020; Montaña-Lozano et al. 2022; Shtolz and
78 Mishmar 2023). In fish, three distinct types of mtDNA gene rearrangements have
79 been described recently by Satoh et al. (2016): “shuffling”, i.e., local position change
80 maintaining coding polarity (i.e., genes coding strand); “translocation”, i.e., movement
81 to a distinct location maintaining the genes encoded strand; and the rarest event
82 “inversions”, i.e., genes (or non-coding unities) switch to their complementary strand
83 (Fonseca et al. 2008; Kong et al. 2009; Gong et al. 2013; Fonseca et al. 2014;
84 Arrondo et al. 2020; Papetti et al. 2021; Shtolz and Mishmar 2023). Inversions, when
85 acting upon the replication controlling unities (i.e., the CR or OH), have been shown
86 to promote a switch of the strand asymmetry nucleotide composition. Consequently,
87 The mtDNA replication is suggested to be involved in the differently accumulated
88 strand mutation patterns of the mtDNA (assuming an asymmetric model of mtDNA
89 replication) (Fonseca et al. 2008; Fonseca et al. 2014; Shtolz and Mishmar 2023).
90 Marine deep-sea hatchetfish from the family Sternoptychidae (Order: Stomiiformes)
91 are a group of small (less than 100 mm) and peculiar mesopelagic fishes (Nelson
92 2016). The family consists of two subfamilies, Maurolicinae and Sternoptychinae,
93 which include around 70 species distributed through 10 genera (Howell and Krueger
94 1987; Nelson 2016; Coad 2019). These species are generally found in high
95 abundance and biomass within the mesopelagic ichthyofauna and with important
96 ecological functions and a key trophic position (Gjoesaeter and Kawaguchi 1980;

97 Eduardo et al. 2020). Species from the family Sternoptychidae have been described
98 as “*some of the most bizarre stomiiforms*” and are characterized by having a
99 condensed body with a reflecting fattened silver side that allows camouflage
100 (Carnevale 2008). As in all other Stomiiformes, deep-sea hatchetfish possess
101 specialized bioluminescent organs, i.e., photophores that allow them to produce light
102 (Krönström et al. 2005; Carnevale 2008; Haddock et al. 2009).

103 Here we describe four new Sternoptychidae whole mtDNA, using Illumina PE short
104 reads and Oxford Nanopore long reads, showing that the mtDNA of Sternoptychidae
105 have a diverse and unusual gene architecture. Our findings include mtDNA gene
106 shuffling, translocation and inversions, acting on PCG, tRNA rRNA and/or CR. In
107 particular, we demonstrate that strand asymmetry nucleotide composition reversal
108 occurs when genes change their coding polarity relative to the Control Region (i.e.
109 OH, initiation of replication). Conversely, inversions of CR were shown to promote
110 the complete nucleotide strand asymmetry reversal in two deep-sea hatchetfish
111 species. Moreover, by investigating over 6000 species we determine that strand
112 asymmetry is a rare event in vertebrate mitochondrial genomes. In Antarctic
113 notothenioid fishes (Trematominae), a PCG inversion (the only case previously
114 reported in fish) coupled with the CR inversion is shown to disrupt the standard
115 asymmetry for the remaining PCG. Together, our findings provide strong evidence
116 supporting the long-lasting theory that replication is the main molecular mechanism
117 promoting the strand-specific compositional bias of the mtDNA.

118

119 **Results and Discussion**

120

121 **Deep-sea hatchetfish mtDNA show a widespread complex repetitive region**
122 **hampering full sequence circularization**

123

124 We generated four novel whole mitochondrial genomes from the marine hatchetfish,
125 two of them resulting from new nuclear genome assemblies: *Argyropelecus*
126 *aculeatus* Valenciennes, 1850 (lovely hatchetfish), *Argyropelecus hemigymnus*
127 Cocco, 1829 (half-naked hatchetfish), *Argyropelecus olfersii* (Cuvier, 1829) and
128 *Maurollicus muelleri* (Gmelin, 1789) (Mueller's pearlside). The raw sequencing reads
129 and mtDNA assemblies were deposited at NCBI, and respective SRA and assembly
130 accessions are depicted in Table 1 and linked to BioProject PRJNA977192
131 (Provisory Reviewer Link
132 <https://dataview.ncbi.nlm.nih.gov/object/PRJNA977192?reviewer=qp1erbs33ac430jh>
133 [nvu9demc1m](#)).

134 The mtDNA lengths varied from 15,230 bp in *M. muelleri* to 23,291 bp in *A.*
135 *aculeatus*, largely influenced by poorly resolved repetitive regions, which includes the
136 putative CR and long species-specific intergenic regions (figs 1-3). The only
137 previously assembled hatchet fish mtDNA with a CR annotation is *Polyipnus polli*
138 Schultz, 1961 (AP012962.1) (fig. 1). Conversely, in other publicly available mtDNAs,
139 the CR was either not annotated, i.e., in *Sternoptyx obscura* Garman, 1899
140 (OP057081.1) and in *Sternoptyx diaphana* Hermann, 1781 (MT588184.1) (fig. 2), or
141 not sequenced, i.e., in *M. muelleri* (AP012963.1) and *Argyropelecus affinis* Garman,
142 1899 (AP012964.1) (fig 1). The difficulty in obtaining this region has already been
143 observed in many other animal groups (e.g., (Ghiselli et al. 2021)). Conversely, in the
144 three novel *Argyropelecus sp.* sequenced mtDNAs, i.e., *A. aculeatus*, *A.*
145 *hemigymnus*, and *A. olfersii*, long intergenic repeats were identified (fig. 3). The

146 difficulty in resolving these repetitive regions using short reads prevented the
147 assembly of a single contig in *A. offersii* (composed of two contigs), as well as the
148 circularization of the *A. hemigygnus* assembly (fig. 3). This factor also hampered the
149 assembly of the recently published *S. diaphana* mtDNA, which required scaffolding
150 with Sanger sequencing (Arrondo et al. 2020). Nevertheless, we efficiently identified
151 the CSB-II (thus the CR) in all but one (i.e., *A. offersii*) (Supplementary File 1 and fig.
152 S1).

153 Sequencing approaches are key for resolving complex mtDNA assemblies (e.g.
154 (Calcino et al. 2020; Formenti et al. 2021)). Consequently, in *A. aculeatus* we used
155 Oxford Nanopore long-reads to determine the composition of the complete mtDNA
156 molecule. The produced reads were highly efficient in resolving the full mtDNA
157 sequence, with some reads even spanning the entire molecule, supporting the
158 inferred architecture. The final assembly was circularized, revealing two long
159 repetitive regions, one spanning ~4800 bp, which included the CSB-II (between
160 tRNA-Asp and tRNA-Pro) and the other ~1750 bp (between tRNA-Phe and a COI)
161 (figs 3 and S2). As for the other two *Argyropelecus sp.*, although the short-read
162 assemblies retrieved all the mtDNA genes, the aforementioned repetitive regions
163 were fragmented (fig. 3).

164

165 **Gene shuffling, translocations, duplications and inversions define the structure** 166 **of mtDNA structure in deep-sea hatchetfish**

167

168 We next investigated the gene content and overall structure of the mtDNA in deep-
169 sea hatchetfish. The newly sequenced genomes as well as those deposited in
170 GenBank show striking deviations from the standard vertebrate gene arrangement

171 (figs 1-3). While the mtDNA from three species, i.e., *M. muelleri*, *A. affinis* and *P.*
172 *polli*, maintain the standard vertebrate mtDNA architecture (fig. 2), most of the newly
173 sequenced mtDNA revealed considerably modified architectures (figs 2-3) (Saccone
174 et al. 2002; Satoh et al. 2016). The results show that shuffling, translocation and
175 inversion of PCGs, rRNAs and more abundantly tRNAs, are widespread in the
176 mtDNA deep-sea hatchetfish (figs 2-3). Even within the same genus, high structural
177 differences can be detected (figs 2-3). Species from the *Argyropelecus* genus show
178 four distinct mtDNA architectures: one maintains the standard vertebrate structure;
179 while the other three share a radical inversion of a large fragment composed of
180 several genes, with reciprocal additional unique features (fig. 3). This fragment
181 includes the inversion of the PCGs ND2 and ND1, the rRNA 16S and 12S and
182 several tRNAs (M, I, L1, V and F), as well as the shuffling of two tRNAs (A and C)
183 (fig. 3). The inversion of two PCGs has never been reported in fish mtDNA before. To
184 date, the only reported inversion of a PCG is of ND1 in a single clade of Antarctic
185 notothenioid fishes (Nototheniidae: Trematominae) (Papetti et al. 2021; Minhas et al.
186 2023). Interestingly, this also resulted from the inversion of a large fragment, which
187 included the rRNA genes, several tRNAs (E, I, L2, V and F) as well as the CR
188 (Papetti et al. 2021; Minhas et al. 2023).

189 Some species-specific tRNA duplications were observed (figs. 2-3). The most striking
190 of these duplications was captured by the Nanopore-based assembly in *A. aculeatus*,
191 corresponding to a tandem duplication of two tRNAs C-A, followed by one tRNA C.
192 This pattern seems to also occur in *A. offersii*, in which a shorter tandem duplication
193 of two tRNAs C-A was present, while in *A. hemigymnus* the two tRNAs are present in
194 the same location but in a single copy (fig. 3). Furthermore, *A. hemigymnus* shows a
195 unique shuffling of the ND6 and the tRNA E (fig. 2). The two *Sternoptyx sp.* revealed

196 the overall same mtDNA architecture, with *S. obscura* having duplications and
197 “*pseudogenization*” of ND3 and 16S (fig. 2). Gene duplication in mtDNA is often
198 followed by the loss of one the copies, frequently retaining fragments of the original
199 duplicated gene (Sato et al. 2016; Papetti et al. 2021).

200

201 **Phylogenetic Analysis shows poorly resolved deep-sea hatchetfish**

202 **evolutionary relationships**

203

204 We next analysed the phylogenetic relationships between the various deep-sea
205 hatchetfish species. The resulting phylogenetic inference, rooted with *Coregonus*
206 *lavaretus* (Linnaeus, 1758) (following (Ijichi et al. 2018; Arrondo et al. 2020)), is split
207 into two major groups, one composed of order Stomiiformes and the other composed
208 of families Synodontidae, Ateleopodidae, Myctophidae and Trachipteridae (fig. 4).
209 Stomiiformes' phylogenetic relationships are poorly resolved, with very low support in
210 most of the nodes. Within deep-sea hatchet fish, i.e., family Sternoptychidae, the
211 relationships among the genus are also poorly resolved and an unexpected long
212 branch is seen for the two *Sternoptyx* species (fig. 4). The low support persisted in
213 the amino acid-based phylogeny (fig. S3). The mtDNA phylogeny here presented is
214 the first to include more than one species of the deep-sea hatchet fish (see (Arrondo
215 et al. 2020)). However, the taxon representation is still incomplete, which likely
216 explains the low support observed. The relationships of the four *Argyropelecus* sp.
217 are well resolved and reflect the structural mtDNA variation here described (figs. 1-3).
218 *Argyropelecus affinis*, the only species to preserve the standard vertebrate structure
219 is the first split within the genus, followed by *A. hemigymnus*, sister to *A. offersii* + *A.*
220 *aculeatus*, the two most structurally similar (fig. 4).

221 To obtain a more taxa-representative phylogeny, all the cytochrome oxidase subunit I
222 (COI) sequences available on GenBank were used (fig. 3). The COI alignment
223 included a total of 261 sequences (total length of 652 bp) (Supplementary File 2),
224 from 26 distinct deep-sea hatchetfish species and the outgroup species, i.e.,
225 *Coregonus clupeaformis* (Mitchill, 1818). The resulting nucleotide phylogeny shows
226 once again low support among inter-genus relationships (fig. 4). Even though this
227 phylogeny includes 260 sequences from marine hatchetfish, it highlights the need to
228 increase the molecular data for the whole group.

229

230 **GC/AT skew of strand-specific 4-fold redundant sites show a strikingly**
231 **frequent strand asymmetry reversal in deep-sea hatchetfish always coupled**
232 **with CR strand relative position**

233

234 The mtDNA asymmetric strand nucleotide composition of the H and L strands is
235 generally highly demarked in vertebrates, with a strong signature in the A-T and C-G
236 composition of each stand (Saccone et al. 2002). This pronounced signature is the
237 distinguishable factor between the two strands, i.e., the H(heavy)-strand, which is
238 guanine rich, and L(light)-strand, which is guanine poor (Francino and Ochman
239 1997). Conversely, inverted genes, i.e., genes that change to the complementary
240 strand, will be exposed to the mutational bias specific of the new strand, and thus are
241 expected to change their mutational signature accordingly, as proposed by the
242 asymmetric model of mtDNA replication (Fonseca et al. 2008; Kennedy et al. 2013;
243 Fonseca et al. 2014).

244 Given the newly detected inversions of two PCGs in *A. aculeatus*, *A. hemigymnus*,
245 and *A. olfersii*, as well as the oddly long branch lengths observed in *Sternoptyx sp.*,

246 we next estimated the GC/AT skew of the PCGs at the 4-fold redundant sites (the
247 most likely to accumulate strand-specific mutations and to reflect the underlying
248 mutational processes given that selection acting on these sites is weaker) for all the
249 marine hatchetfish under study.

250 As expected, in the mtDNA that maintains the standard vertebrate architecture with
251 no detectable structural changes, i.e., *M. muelleri*, *A. affinis* and *P. polli*, shows the
252 AT/GC skew pattern of a gene encoded in the L-strand (fig. 5 and S4, i.e., AT skew <
253 0 and GC skew > 0). Interestingly, in the two *Argyropelecus sp.* with inverted PCG
254 polarity, this AT/GC skew pattern was observed in ND6, but also in ND1 and ND2,
255 the two PCGs included in the inverted fragment (fig. 5 and S4). Most strikingly, for
256 the two *Sternoptyx sp.*, the entire AT/GC skew pattern is reversed in all PCGs (fig. 5
257 and S4) but with no rearrangements or inversions detected in those genes. What
258 could be causing the AT/GC skew reversion in these genomes? To date, such
259 phenomenon, in vertebrates, has been suggested to result from CR inversion, and
260 from the replication mechanism itself with the leading strand becoming the lagging
261 strand and vice-versa during an asymmetric mode of replication (Fonseca et al.
262 2008; Fonseca et al. 2014). We tested this hypothesis by characterising the relative
263 positioning of the initiation of replication and by identifying elements that have been
264 described to be key in the replication mechanism of vertebrate mtDNA: the
265 conserved sequence blocks CSB-II and III. Typically, CSB-II consists of six or more
266 cytosines, one thymine and one adenine ending with five or more cytosines: 5'-
267 CCCCCCTACCCCC – 3'. We used this sequence as a reference, allowing for some
268 variation in the length of the poly-C flanks. CSB-III is less conserved, generally
269 identified by its positioning regarding CSB-II, and may be absent in fish CR (Sato et
270 al. 2016). The Motif Discovery analysis identified CSB-II in all the analysed

271 sequences, except in *P. polli* while CSB-III was less prevalent, as previously
272 observed by Satoh et al., (2016) (Supplementary File 1 and fig. S1). The absence of
273 both CSBs in *P. polli* needs further evaluation, as this mitogenome is a direct
274 submission to NCBI without detailed information on how it was generated.
275 Astonishingly, as previously predicted, the orientation of the CSB-II and therefore of
276 the CR relative to the PCG was the common denominator determining the AT/GC
277 skew nucleotide asymmetry pattern (Supplementary File 1 and figs 1-3 and S1). That
278 is, in the two *Sternoptyx sp.* the CSB-II is the only element changing its coding
279 polarity (inversion), having the same polarity as ND6 and contrary to all other PCGs,
280 while in the remaining species, the relative position follows the standard vertebrate
281 architecture (Supplementary File 1 and figs 1-3 and S1).
282 To date, the only reported PCG inversion in a vertebrate is in one clade of Antarctic
283 notothenioid fishes (Nototheniidae: Trematominae), consisting of the inversion of a
284 large genomic segment (~5kbp), which includes four tRNAs, two rRNAs, the PCG
285 inversion (ND1) and the CR (Kennedy et al. 2013; Papetti et al. 2021). Once again,
286 the relative position of the CR, here identified through CSB-II (Supplementary File 1
287 and fig. S1), is the determining factor shaping the nucleotide strand asymmetry, i.e.,
288 ND1 is the only PCG with the same polarity as the CR and thus it is also the only
289 PCG maintaining its AT/GC skew pattern, while the remaining genes show a
290 completely disrupted pattern. Interestingly, the fact that the latter-mentioned PCG
291 shows a disruption but not a reversal AT/GC skew pattern, suggests that, in this
292 group of organisms, the process of the mtDNA nucleotide composition strand
293 asymmetry reversal is still in a transitory state (fig. S5). A complete disruption of the
294 typical strand asymmetry signature was detected in all 15 Trematominae mtDNA

295 available, while the remaining notothenioid fishes maintained the standard pattern
296 (fig. S5).

297 To access similar putative hidden patterns in other vertebrates, we next performed a
298 vertebrate-wide assessment of the AT/GC skew at 4-fold redundant sites (Fonseca et
299 al. 2008; Fonseca et al. 2014) in a total of 6,297 mtDNA (Retrieved from RefSeq in
300 May 2022). The results demonstrate that the complete reversal of the nucleotide
301 strand asymmetry composition is observed solely in the *Sternoptyx sp.* and those
302 previously reported by Fonseca et al., (2008, 2014), totalizing seven species (figs. 6
303 and S4). Moreover, single gene asymmetry reversal is detectable for the ND1 and
304 ND2 genes of the three *Argyropelecus* species here described (figs. 6 and S4).

305 Given that there are still many un-sequenced marine hatchetfish whole mtDNA, we
306 hypothesised that calculating the AT/GC skew pattern at 4-fold redundant sites for
307 the most sequenced mitochondrial gene of the group, i.e., COI, would be an
308 informative strategy regarding the whole family. The results also show the reversed
309 pattern in all analysed *Sternoptyx* species (n=4), thus suggesting the reversal of the
310 nucleotide strand asymmetry is shared by all analysed *Sternoptyx* (figs. 6 and S4).
311 The remaining hatchetfish species maintain the standard vertebrate AT/GC skew
312 pattern, which is also in accordance with the results deduced from whole mtDNA
313 assembly and annotation.

314

315 **Genomic *Oddities*: inverted mtDNA and large nuclear genomes**

316

317 The combination of sequencing methodologies allowed the determination of *de novo*
318 four mtDNA from deep-sea hatchetfish species. Strikingly, the sequential analysis of
319 gene annotation, phylogenetics, and AT/GC skew sequence investigations support a

320 unique case of mtDNA structural plasticity (Shtolz and Mishmar 2023). We expanded
321 the analysis to include thousands of available mtDNA. Our findings show that partial
322 or full mtDNA inversions are evolutionary restricted in a minute of unrelated fish
323 clades: *Argyropelecus sp.* *Sternoptyx sp.* The immediate causes of this
324 phylogenetically restricted alteration of mtDNA structure are unknown and deserve
325 further exploration. In fact, within Metazoa, vertebrates show the strongest signs of
326 purifying selection for gene rearrangement (Shtolz and Mishmar 2023), with recent
327 studies suggesting that gene blocks inversions (only recorded in invertebrates)
328 promote sticking changes in the transcriptional pattern, thus requiring new regulatory
329 elements (Blumberg et al. 2014; Blumberg et al. 2017; Barshad et al. 2018; Shtolz
330 and Mishmar 2023). Consequently, our highly unexpected findings raise the
331 interesting possibility of an adaptive scenario of the reported mtDNA oddities.
332 Interestingly, using the high-coverage whole genome sequencing for *A. aculeatus*
333 and *S. diaphana* we were able to appraise the genome size for both species (fig. S6).
334 The results show an unexpectedly large genome size estimation, with both species
335 being approximately 2.70 Gbp long (fig. S6). Most teleost species have an average of
336 1 Gbp long genome size, with the exceptions being usually attributed to whole
337 genome duplication events (Parey et al. 2022). Together, our results suggest that like
338 the mitogenomes, the whole genome of deep-sea hatchetfish is also unusual and
339 deserves future explorations.

340

341

342 **Methods**

343

344 **Samples collection and DNA extraction**

345 In total, four specimens were captured for this study, three from the genus
346 *Argyropelecus*, i.e., *Argyropelecus aculeatus* Valenciennes, 1850, *Argyropelecus*
347 *hemigymnus* Cocco, 1829, and *Argyropelecus olfersii* (Cuvier, 1829), and one from
348 genus *Maurolicus*, i.e., *Maurolicus muelleri* (Gmelin, 1789). The specimens were
349 obtained during the scientific surveys: EU Groundfish Survey (Platuxa-2019) in North
350 Atlantic (43,1544 N- 51,4429 W, 2019); from the EU Groundfish Survey (Platuxa-
351 2019) in the Northwest Atlantic Ocean (43,3838 N- 49,0036 W, 2019); from the
352 Survey PORCUPINE20 In Porcupine Bank (51.0677 N; -14.2862 W, 2020) and from
353 EU Groundfish Survey FN3L19 in North Atlantic (-47,668 N; 47,497833 W),
354 respectively. Morphological identification was performed onboard and whole
355 specimens are stored in absolute ethanol and are stored at DNA and Tissue bank at
356 CIIMAR – Interdisciplinary Center of Marine and Environmental Research. The
357 specimen treatment has been approved by the CIIMAR ethical committee and by
358 CIIMAR Managing Animal Welfare Body (ORBEA) according to the European Union
359 Directive 2010/63/EU. Whole genomic DNA for each specimen was obtained from a
360 small portion of the muscle tissue using the Qiagen MagAttract HMW DNA extraction
361 kit, following the manufacturer's instructions. For all samples, total DNA was used for
362 Illumina paired-end (PE) library preparation and sequencing at the Macrogen, Inc.
363 (Seoul, Korea), using Illumina HiSeq X Ten platform, with 150 bp PE configuration.
364 For *A. aculeatus* and *S. diaphana* high coverage whole genome (WGS) PE
365 sequencing was performed, while the remaining samples were only sequenced at
366 low coverage. Furthermore, low-coverage Nanopore (Oxford Nanopore) genome
367 skimming was performed for *A. aculeatus*. Briefly, ~1 µg of genomic DNA was used
368 for library preparation using the LSK109 kit and after sequenced on an FLO-MIN106
369 revD SpotON R9.4 Flow Cell for 48 h.

370

371 **Whole mitogenome assemblies and annotation**

372 Raw Illumina PE reads were quality-filter and adaptors were removed using
373 Trimmomatic (version 0.38) (Bolger et al. 2014), using the parameters LEADING:5
374 TRAILING:5 SLIDINGWINDOW:5:20 MINLEN:36. Read quality was inspected before
375 and after trimming using FastQC (version 0.11.8)
376 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

377 Whole mitogenome reconstruction for each species was obtained using several
378 distinct approaches. For each species, assemblies were performed with
379 GetOrganelle v1.7.1 (Jin et al. 2020), specifying a multi k-mer approach (i.e., from
380 20-125 with a 5-mer increment). GetOrganelle is an interactive baiting pipeline that
381 filters mtDNA reads and uses the SPAdes assembler (Bankevich et al. 2012) to
382 reconstruct the mitogenome with the filtered reads. The results of the assemblies
383 were individually validated using multiple approaches. For cases where multiple
384 assemblies were generated, i.e., *A. hemigymnus*, and *A. olfersii*, the gfa files were
385 inspected using Bandage v0.8.1 (Wick et al. 2015) which revealed several
386 ambiguous disjoints. Furthermore, annotation was generated for some of the putative
387 assemblies using the module “annotate” from MitoZ v.3.4 (Meng et al. 2019), which
388 showed that the ambiguous disjoints were localized within no-coding repetitive
389 regions. Consequently, new assemblies were performed, first using larger k-mer with
390 GetOrganelle v1.7.1 and if the problem persisted, using metaSPAdes v3.12.0 (Nurk
391 et al. 2017) with the maximum K-mer size possible, i.e., 127. In the end, the selected
392 best representative assemblies for each species were as follows, for *M. muelleri* the
393 GetOrganelle assembly with multi k-mer (20-125-mer with a 5-mer increment), for *A.*
394 *olfersii* the GetOrganelle assembly single k-mer (131-mer) and for *A. hemigymnus*

395 the metaSPAdes assembly with the maximum allowed k-mer size (127-mer). Every
396 generated assembly was annotated with MitoZ (as described above). Read coverage
397 distributions were analysed by aligning PE reads to the final genome assemblies
398 using the Burrows-Wheeler Aligner (BWA) v.0.7.17-r1198 (Li 2013) and visualized in
399 Artemis v17.0.1 (Carver et al. 2012) (fig. S7).

400 The Nanopore reads of *A. aculeatus* were quality-filtered using Filtlong
401 (<https://github.com/rwick/Filtlong>). Given that long repetitive regions seem to cause
402 problems with the PE-based assemblies, the Nanopore reads were filtered by size
403 (i.e., >21,000bp) to include only reads spanning the whole mitogenome. The
404 mitogenome assembly was performed using Unicycler v.0.4.8. (Wick et al. 2017)).
405 The assembly was polished, following the author's suggestions, with the Nanopore
406 reads using medaka v1.2.2 (<https://github.com/nanoporetech/medaka>) and with
407 short-reads, first using Polypolish v0.4.3 (Wick and Holt 2022) and after using Polca
408 (Zimin and Salzberg 2020). Genome annotation was performed with MitoZ (as
409 described above). Read coverage distributions were analysed by aligning PE reads
410 (as described above).

411 Finally, since the *Sternopyx obscura* mitogenome available on NCBI (OP057081) is
412 marked as "UNVERIFIED" and therefore no annotation is provided, the mitogenome
413 was downloaded and annotated using MitoZ (as described above).

414 To identify the putative Control Region (CR) of the deep sea hatchet fish mitogenome
415 assemblies, we search for conserved motifs within non-coding regions to identify any
416 of the known Conserved Sequence Blocks (CSB) that are involved in the replication
417 initiation (Sato et al. 2016). The recently described CSBs of several fish species
418 were here used as a reference to guide the search, including one Stomiformes, i.e.
419 *Diplophos taenia* (Sato et al. 2016). We focused on the two most conserved of the

420 three CSBs, i.e., CSB-II and CSB-III. Given the lack of significant non-coding regions
421 within the mitogenomes of *M.muelleri*, *A. affinis*, and *A. olfersii* (see Results and
422 Discussion), these mitogenomes were not included in the analysis. Moreover, given
423 the highly disproportionate read coverage distribution in the mitogenome non-coding
424 regions of *A. hemigymnus* (fig. S7), this mitogenome was not considered. The
425 coverage distribution indicates regions likely represented by the collapse of a highly
426 repetitive region, thus not suited for the analysis. The non-coding regions from the
427 remaining deep-sea hatchet fish species, the Antarctic Trematominae species,
428 *Pagothenia borchgrevinkias* (with the only other record CDS mitochondrial inversion),
429 as well as all the fish CSB-II and CSB-III identified by Satoh et al. (2016) were used.
430 The sequences were uploaded to MEME v.5.5.2 (Bailey and Elkan 1994) and the
431 Motif Discovery model was used to identify conserved motifs across all sequences,
432 specifying a maximum motif width of 25.

433

434 **Phylogenetic reconstruction**

435 To produce a phylogenetic reconstruction, the whole mitogenomes of all
436 Stomiiformes available on NCBI (n=15), including four Sternoptychidae species (i.e.,
437 *S. obscura*, *M. muelleri*, *P. polli*, *A. affinis*), the mitogenomes here produced and four
438 outgroup taxa were used (Table 1). Alignments of the 13 protein-coding genes
439 (PCGs) were produced using MAFFT v7.453 (Kato and Standley 2013). Positions
440 with gaps in 50% or more of the individual alignments were removed using trimAL v.
441 1.2rev59 (Capella-Gutiérrez et al. 2009) and all alignments were concatenated using
442 FASconCAT-G (<https://github.com/PatrickKueck/FASconCAT-G>). The alignment
443 composed by the concatenated PCGs from 15 species had a total length of 11,436
444 bp. The partition scheme and the evolutionary models that best fit those schemes, as

445 well as Maximum Likelihood (ML) phylogenetic inference were produced in IQ-TREE
446 v.1.6.12 (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017).

447 For the amino acid phylogenetic reconstruction, individual alignments were translated
448 to proteins using the EMBOSS seqret V.6.6.0.0, trimmed with trimAL v. 1.2rev59 (as
449 described above) and concatenated using FASconCAT-G. The alignment composed
450 by the concatenated PCGs from 15 species had a total length of 3,801 aa. The
451 partition scheme and the evolutionary models that best fit those schemes, as well as
452 Maximum Likelihood (ML) phylogenetic inference were produced in IQ-TREE
453 v.1.6.12.

454

455 **COI Phylogeny**

456 To produce a phylogenetic reconstruction with a wider taxa representation, all COI
457 sequences available on GenBank (n=261) were downloaded (including the outgroup
458 taxa *C. clupeiformis*). Alignment was performed using MAFFT v7.453 (Kato and
459 Standley 2013) and an ML phylogenetic inference was constructed in IQ-TREE
460 v.1.6.12, also estimating the best evolutionary model for the analysis.

461

462 **Strand-specific 4-fold redundant sites GC/AT skew estimations**

463 The protein-coding genes of the complete mitogenomes of all vertebrates available in
464 GenBank ([http:// www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) in May of 2022 were retrieved. The
465 calculation of the GC skew $(G-C)/(G+C)$ and the AT skew $(A-T)/(A+T)$ at 4-fold
466 redundant sites using custom Perl scripts (Fonseca et al. 2014). The
467 redundant codons examined were alanine (GCN), proline (CCN), serine (TCN),
468 threonine (ACN), arginine (CGN), glycine (GGN), leucine (CTN), and valine (GTN).

469 Furthermore, to infer GC/AT skew inversions in other species from family
470 Sternoptychidae, for which no whole mitogenome is available, the COI sequences for
471 all Sternoptychidae species available on GenBank were downloaded and the
472 calculations applied, as described above.

473

474 **Whole Genome size estimation**

475 The high coverage WGS PE sequencing reads for *A. aculeatus* and *S. diaphana*
476 were used to estimate the overall characteristics of each species' genomes using
477 Jellyfish v.2.2.10 and GenomeScope2 (Ranallo-Benavidez et al. 2020) with a k-mers
478 length of 21.

479

480 **Data Availability Statement**

481 The raw sequencing reads and mtDNA assemblies are deposited at NCBI, and
482 respective SRA and assembly accessions are depicted in Table 1, all linked to
483 BioProject PRJNA977192.

484

485

486 **Funding**

487 AGS was funded by grant 2023_033_BI_ATLANTIDA, under the Project
488 “ATLANTIDA - Platform for the monitoring of the North Atlantic Ocean and tools for
489 the sustainable exploitation of the marine resources” (NORTE-01-0145-FEDER-
490 000040), co-financed by Portugal 2020 and the European Union through Program
491 FEDER. EF is funded by the Portuguese Foundation for Science and Technology
492 (FCT) under grant CEECINST/00027/2021. This research was developed under the

493 project ATLANTIDA - Platform for the monitoring of the North Atlantic Ocean and
494 tools for the sustainable exploitation of the marine resources” (NORTE-01-0145-
495 FEDER-000040). Additional strategic funding was provided by FCT
496 UIDB/04423/2020 and UIDP/04423/2020. EU-Spain NAFO Groundfish survey has
497 been co-funded by the European Union through the European Maritime and
498 Fisheries Fund (EMFF) within the National Program of collection, management and
499 use of data in the fisheries sector and support for scientific advice regarding the
500 Common Fisheries Policy and IEO BIOPEOPLE project. The Spanish Bottom Trawl
501 Survey on the Porcupine Bank (SP-PORC-Q3) was funded in part by the EU through
502 the European Maritime and Fisheries Fund (EMFF) within the Spanish National
503 Program of Collection, Management and Use of Data in the Fisheries Sector and
504 Support for Scientific Advice Regarding the Common Fisheries Policy.

505

506

507 **FIGURE AND TABLE LEGENDS**

508

509 FIG.1. Left: Schematic representation of circular mitochondrial molecule highlighting
510 the protein-coding genes encoded in the different strands. Right: Representation of
511 linearized mitochondrial molecule, depicting the standard vertebrate gene order
512 shared by the deep-sea hatchetfish species, *Polyipnus polli*, *Maurolicus muelleri* and
513 *Argyropelecus affinis*. Genes encoded in the L-strand are depicted in red; Genes
514 encoded in the H-strand are depicted in white. * Control Region has not been
515 sequenced in either *Maurolicus muelleri* or *Argyropelecus affinis*. The CSB-II
516 sequence is a representation of the generally expected composition of the motif.

517

518 FIG.2. Left: Schematic representation of circular mitochondrial molecule highlighting
519 the protein-coding genes encoded in the different strands. Right: Representation of
520 linearized mitochondrial molecule, depicting the standard vertebrate gene order
521 (middle) comparatively to deep-sea hatchetfish species, *Sternoptyx obscura*,
522 *Sternoptyx diaphana*. Genes encoded in the L-strand are depicted in red; Genes
523 encoded in the H-strand are depicted in white; Red shadow represents
524 macrosyntenic patterns of gene inversion. The CSB-II sequence for the typical
525 vertebrate mtDNA is a representation of the generally expected composition of the
526 motif. The CSB-II sequences for the marine hatchet fish are the ones here identified.
527

528 FIG.3. Left: Schematic representation of circular mitochondrial molecule highlighting
529 the protein-coding genes encoded in the different strands. Right: Representation of
530 linearized mitochondrial molecule, depicting the standard vertebrate gene order
531 (second from the top) comparatively to deep-sea hatchetfish species, *Argyropelecus*
532 *hemingymnus*, *Argyropelecus aculeatus* and *Argyropelecus olfersii*. Genes encoded
533 in the L-strand are depicted in red; Genes encoded in the H-strand are depicted in
534 white; Red shadow represents macrosyntenic patterns of gene inversion; Grey
535 shadow represents macrosyntenic patterns of genes in the same strand. The CSB-II
536 sequence for the typical vertebrate mtDNA is a representation of the generally
537 expected composition of the motif. The CSB-II sequences for the marine hatchet fish
538 are the ones here identified.

539

540 FIG.4. a) IQ-tree Maximum Likelihood (ML) phylogenetic inference retrieved from the
541 nucleotide alignment of the 13 mitogenomes concatenated protein-coding genes. b)
542 IQ-tree Maximum Likelihood (ML) phylogenetic inference retrieved from the

543 nucleotide alignment of all available COI sequences of deep-sea hatchetfish in NCBI.

544 Above the nodes are represented ultrafast bootstrap values above 80%.

545

546 FIG.5. Top: Strand-specific 4-fold redundant sites GC/AT skew estimations for all

547 protein-coding genes, for four model vertebrate species with standard vertebrate

548 mitogenome gene order, each species of deep-sea hatchetfish analysed and one

549 representative species from the Antarctic fish subfamily Trematominae that have an

550 inversion of the ND1 and CR (inside the box). Bottom: Strand-specific 4-fold

551 redundant sites GC/AT skew estimations for COI mitochondrial gene of all deep-sea

552 hatchetfish, available in NCBI. Red shadows highlight the inversion of the strand-

553 specific nucleotide GC/AT skew pattern. Middle: Schematic representation of circular

554 mitochondrial molecule highlighting the protein-coding genes encoded in the different

555 strands.

556

557 FIG.6. Plots of the GC and AT skews at the 4-fold redundant sites of the

558 mitochondrial protein-coding genes ND1, ND2, COX1 and ND6 from 6,297 vertebrate

559 species. Dots represent each of the 6,297 analysed vertebrate species, with red

560 (deep-sea hatchetfish) and green (other fish species) dots highlighting genes with

561 inverted AT and GC skews. The plots for the remaining protein-coding genes are

562 provided in fig.S4.

563

564 Table 1. List of whole mitogenomes used in the whole mitogenome-based phylogeny.

565 * represent mitogenomes sequenced in this study.

566

567 FIG.S1. Schematic representation of the MEME - Motif discovery results, including
568 the inferred conserved motif representing CSB-II (in red) and, in some species, CSB-
569 III (in blue). The results include several CR from fish species studied in Satoh et al
570 (2016) (using the same abbreviated nomenclature), as well as the non-coding
571 regions of the deep-sea marine hatchet fish species *Sternoptyx obscura*, *Sternoptyx*
572 *diaphana*, *Argyropelecus aculeatus*, the Stomiiformes species *Diplophos taenia*, and
573 the Antarctic Trematominae species *Pagothenia borchgrevinki*.

574

575 FIG.S2. Complete circular mitochondrial molecule of the deep-sea hatchetfish
576 species *Argyropelecus aculeatus*, produced by CGView/Proksee (<https://proksee.ca>).

577

578 FIG.S3. IQ-tree Maximum Likelihood (ML) phylogenetic inference retrieved from the
579 amino acid alignments of the 13 concatenated mitogenome protein-coding genes.

580 Above the nodes are represented ultrafast bootstrap values.

581

582 FIG.S4. Plots of the GC and AT skews at the 4-fold redundant sites of the
583 mitochondrial protein-coding genes ND3, ND4, ND4L, ND5, COX2, COX3, CYTB,
584 ATP6 ATP8 from 6,297 vertebrate species. Dots represent each of the 6,297
585 analysed vertebrate species, with red (deep-sea hatchetfish) and green (other fish
586 species) dots highlighting genes with inverted AT and GC skews. The plots for the
587 remaining protein-coding genes are provided in fig.6.

588

589 FIG.S5. Strand-specific 4-fold redundant sites GC/AT skew estimations for all
590 protein-coding genes, for all mitogenomes of notothenioid fishes available on NCBI.

591 Coloured circles represent the mitochondrial gene orders described by Papetti et al.

592 2021. Species not studied by Papetti et al. 2021 are in the black box. Mitogenomes
593 with an inversion of the ND1 gene are highlighted in a thick red box.

594

595 FIG.S6. GenomeScope2 k-mer (21) distributions displaying estimation of genome
596 size (len), homozygosity (aa), heterozygosity (ab), mean k-mer coverage for
597 heterozygous bases (kcov), read error rate (err), the average rate of read
598 duplications (dup), k-mer size used on the run (k:), and ploidy (p:) for *Argyropelecus*
599 *aculeatus* (left) and *Sternoptyx diaphana* (right).

600

601 FIG.S7. Artemis read coverage plot distribution across the assembled marine hatchet
602 fish mitogenome assemblies.

603

604 Supplementary File 1 - MEME - Motif discovery results, including the sequences of
605 the inferred conserved motif representing CSB-II and, in some species, CSB-III. The
606 results include all the CR of the fish species studied in Satoh et al (2016) (using the
607 same abbreviated nomenclature), as well as the non-coding regions deep-sea
608 marine hatchet fish species *Sternoptyx obscura*, *Sternoptyx diaphana*, *Argyropelecus*
609 *aculeatus*, the Stomiiformes species *Diplophos taenia*, and the Antarctic
610 Trematominae species *Pagothenia borchgrevinki*.

611 Supplementary File 2 - COI alignment including 261 sequences from 26 distinct
612 deep-sea marine hatchetfish species and the outgroup species, i.e., *Coregonus*
613 *clupeaformis* (Mitchill, 1818) (total length of 652 bp).

614 References

- 615 Arrondo NV, Gomes-dos-Santos A, Román Marcote E, Pérez M, Froufe E, Castro LFC. 2020. A new
616 gene order in the mitochondrial genome of the deep-sea diaphanous hatchet fish *Sternoptyx*
617 *diaphana* Hermann, 1781 (Stomiiformes: Sternoptychidae). *Mitochondrial DNA B Resour*
618 [Internet] 5:2859–2861. Available from:
619 <https://www.tandfonline.com/doi/full/10.1080/23802359.2020.1790325>
- 620 Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in
621 biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for*
622 *Molecular Biology* 2:28–36.
- 623 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S,
624 Prjibelski AD, et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to
625 Single-Cell Sequencing. *Journal of Computational Biology* [Internet] 19:455–477. Available from:
626 <http://www.liebertpub.com/doi/10.1089/cmb.2012.0021>
- 627 Barshad G, Marom S, Cohen T, Mishmar D. 2018. Mitochondrial DNA Transcription and Its
628 Regulation: An Evolutionary Perspective. *Trends in Genetics* 34:682–692.
- 629 Blumberg A, Rice EJ, Kundaje A, Danko CG, Mishmar D. 2017. Initiation of mtDNA transcription is
630 followed by pausing, and diverges across human cell types and during evolution. *Genome Res*
631 27:362–373.
- 632 Blumberg A, Sailaja BS, Kundaje A, Levin L, Dadon S, Shmorak S, Shaulian E, Meshorer E, Mishmar D.
633 2014. Transcription Factors Bind Negatively Selected Sites within Human mtDNA Genes.
634 *Genome Biol Evol* [Internet] 6:2634–2646. Available from:
635 <https://academic.oup.com/gbe/article/6/10/2634/610683>

- 636 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
637 *Bioinformatics* [Internet] 30:2114–2120. Available from:
638 <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu170>
- 639 Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res* [Internet] 27:1767–1780. Available
640 from: <https://academic.oup.com/nar/article/27/8/1767/2847916>
- 641 Calcino A, Baranyi C, 1, Wanninger A. 2020. Heteroplasmy and repeat expansion in the plant-like
642 mitochondrial genome of a bivalve mollusc. *bioRxiv* [Internet]:2020.09.23.310516. Available
643 from: <https://www.biorxiv.org/content/10.1101/2020.09.23.310516v2>
- 644 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment
645 trimming in large-scale phylogenetic analyses. *Bioinformatics* [Internet] 25:1972–1973.
646 Available from: <https://academic.oup.com/bioinformatics/article/25/15/1972/213148>
- 647 Carnevale G. 2008. Miniature deep-sea hatchetfish (Teleostei: Stomiiformes) from the Miocene of
648 Italy. *Geol Mag* [Internet] 145:73–84. Available from:
649 [https://www.cambridge.org/core/journals/geological-magazine/article/miniature-deepsea-](https://www.cambridge.org/core/journals/geological-magazine/article/miniature-deepsea-hatchetfish-teleostei-stomiiformes-from-the-miocene-of-italy/F6121BE08716E80EA7731CC94A62DAB6)
650 [hatchetfish-teleostei-stomiiformes-from-the-miocene-of-](https://www.cambridge.org/core/journals/geological-magazine/article/miniature-deepsea-hatchetfish-teleostei-stomiiformes-from-the-miocene-of-italy/F6121BE08716E80EA7731CC94A62DAB6)
651 [italy/F6121BE08716E80EA7731CC94A62DAB6](https://www.cambridge.org/core/journals/geological-magazine/article/miniature-deepsea-hatchetfish-teleostei-stomiiformes-from-the-miocene-of-italy/F6121BE08716E80EA7731CC94A62DAB6)
- 652 Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for
653 visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*
654 [Internet] 28:464–469. Available from:
655 <https://academic.oup.com/bioinformatics/article/28/4/464/213043>
- 656 Clayton DA. 1982. Replication of animal mitochondrial DNA. *Cell* 28:693–705.
- 657 Coad BW. 2019. Family Sternoptychidae – Marine Hatchetfishes, haches d’argent. *Marine Fishes of*
658 *Arctic Canada*:305–307.

- 659 Eduardo LN, Bertrand A, Mincarone MM, Santos L V., Frédou T, Assunção R V., Silva A, Ménard F,
660 Schwamborn R, Le Loc'h F, et al. 2020. Hatchetfishes (Stomiiformes: Sternoptychidae)
661 biodiversity, trophic ecology, vertical niche partitioning and functional roles in the western
662 Tropical Atlantic. *Prog Oceanogr* 187:102389.
- 663 Fonseca MM, Harris DJ, Posada D. 2014. The Inversion of the Control Region in Three Mitogenomes
664 Provides Further Evidence for an Asymmetric Model of Vertebrate mtDNA Replication. *PLoS*
665 *One* [Internet] 9:e106654. Available from:
666 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0106654>
- 667 Fonseca MM, Posada D, Harris DJ. 2008. Inverted Replication of Vertebrate Mitochondria. *Mol Biol*
668 *Evol* [Internet] 25:805–808. Available from: [https://academic.oup.com/mbe/article-](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msn050)
669 [lookup/doi/10.1093/molbev/msn050](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msn050)
- 670 Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli
671 FO, Ambrosini R, et al. 2021. Complete vertebrate mitogenomes reveal widespread repeats and
672 gene duplications. *Genome Biol* [Internet] 22:1–22. Available from:
673 <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02336-9>
- 674 Francino MP, Ochman H. 1997. Strand asymmetries in DNA evolution. *Trends in Genetics* 13:240–
675 245.
- 676 Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. 2021. Molluscan
677 mitochondrial genomes break the rules. *Philosophical Transactions of the Royal Society B:*
678 *Biological Sciences* [Internet] 376:20200159. Available from:
679 <https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0159>
- 680 Gissi C, Pesole G, Mastrototaro F, Iannelli F, Guida V, Griggio F. 2010. Hypervariability of Ascidian
681 Mitochondrial Gene Order: Exposing the Myth of Deuterostome Organelle Genome Stability.

- 682 *Mol Biol Evol* [Internet] 27:211–215. Available from:
683 <https://academic.oup.com/mbe/article/27/2/211/968133>
- 684 Gjoesaeter J, Kawaguchi K. 1980. A review of the world resources of mesopelagic fish. *FAO Fisheries*
685 *Technical Papers (FAO)*.
- 686 Gong L, Shi W, Wang ZM, Miao XG, Kong XY. 2013. Control region translocation and a tRNA gene
687 inversion in the mitogenome of *Paraplagusia japonica* (Pleuronectiformes: Cynoglossidae).
688 <http://dx.doi.org/10.3109/19401736.2013.773984> [Internet] 24:671–673. Available from:
689 <https://www.tandfonline.com/doi/abs/10.3109/19401736.2013.773984>
- 690 Haddock SHD, Moline MA, Case JF. 2009. Bioluminescence in the Sea.
691 <https://doi.org/10.1146/annurev-marine-120308-081028> [Internet] 2:443–493. Available from:
692 <https://www.annualreviews.org/doi/abs/10.1146/annurev-marine-120308-081028>
- 693 Howell WH, Krueger WH. 1987. Family Sternoptychidae, marine hatchetfishes and related species.
694 Smithsonian contributions to zoology. In: GIBBS RH, KRUEGER WH, editors. Biology of Midwater
695 Fishes of the Bermuda Ocean Acre. Vol. 452. Washington, D.C.: Smithsonian Institution Press. p.
696 32–50.
- 697 Ijichi M, Takano T, Hasegawa M, Yashiki H, Kogure K, Kojima S, Yoshizawa S. 2018. The complete
698 mitochondrial genome of the longfin dragonfish *Tactostoma macropus* (Stomiiformes:
699 Stomiidae). *Mitochondrial DNA B Resour* [Internet] 3:486–487. Available from:
700 <https://www.tandfonline.com/doi/full/10.1080/23802359.2018.1464411>
- 701 Jin JJ, Yu W Bin, Yang JB, Song Y, Depamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: A fast and
702 versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* [Internet]
703 21:241. Available from: [https://genomebiology.biomedcentral.com/articles/10.1186/s13059-](https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02154-5)
704 [020-02154-5](https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02154-5)

- 705 Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model
706 selection for accurate phylogenetic estimates. *Nat Methods* [Internet] 14:587–589. Available
707 from: <http://www.nature.com/articles/nmeth.4285>
- 708 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements
709 in performance and usability. *Mol Biol Evol* [Internet] 30:772–780. Available from:
710 <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010>
- 711 Kennedy SR, Salk JJ, Schmitt MW, Loeb LA. 2013. Ultra-Sensitive Sequencing Reveals an Age-Related
712 Increase in Somatic Mitochondrial Mutations That Are Inconsistent with Oxidative Damage.
713 *PLoS Genet* 9:e1003794.
- 714 Kolesnikov AA, Gerasimov ES. 2012. Diversity of mitochondrial genome organization. *Biochemistry*
715 (*Moscow*) [Internet] 77:1424–1435. Available from:
716 <https://link.springer.com/article/10.1134/S0006297912130020>
- 717 Kong X, Dong X, Zhang Y, Shi W, Wang Z, Yu Z. 2009. A novel rearrangement in the mitochondrial
718 genome of tongue sole, *Cynoglossus semilaevis*: Control region translocation and a tRNA gene
719 inversion. *Genome* [Internet] 52:975–984. Available from:
720 <https://cdnsiencepub.com/doi/abs/10.1139/G09-069>
- 721 Krönström J, Holmgren S, Baguet F, Salpietro L, Malfet J. 2005. Nitric oxide in control of
722 luminescence from hatchetfish (*Argyropspectus hemigymnus*) photophores. *Journal of*
723 *Experimental Biology* [Internet] 208:2951–2961. Available from:
724 [https://journals.biologists.com/jeb/article/208/15/2951/15770/Nitric-oxide-in-control-of-](https://journals.biologists.com/jeb/article/208/15/2951/15770/Nitric-oxide-in-control-of-luminescence-from)
725 luminescence-from
- 726 Lee WJ, Conroy J, Howell WH, Kocher TD. 1995. Structure and evolution of teleost mitochondrial
727 control regions. *J Mol Evol* [Internet] 41:54–66. Available from:
728 <https://link.springer.com/article/10.1007/BF00174041>

- 729 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
730 Available from: <http://arxiv.org/abs/1303.3997>
- 731 Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly,
732 annotation and visualization. *Nucleic Acids Res* [Internet] 47:e63–e63. Available from:
733 <https://doi.org/10.1093/nar/gkz173>
- 734 Minhas BF, Beck EA, Cheng C-HC, Catchen J. 2023. Novel mitochondrial genome rearrangements
735 including duplications and extensive heteroplasmy could underlie temperature adaptations in
736 Antarctic notothenioid fishes. *Sci Rep* 13:6939.
- 737 Montaña-Lozano P, Moreno-Carmona M, Ochoa-Capera M, Medina NS, Boore JL, Prada CF. 2022.
738 Comparative genomic analysis of vertebrate mitochondrial reveals a differential of
739 rearrangements rate between taxonomic class. *Scientific Reports* 2022 12:1 [Internet] 12:1–13.
740 Available from: <https://www.nature.com/articles/s41598-022-09512-2>
- 741 Nelson JS. 2016. *Fishes of the World*. 5th ed. (JohnWiley and Sons, editor.). New York, NY
- 742 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic
743 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* [Internet] 32:268–274.
744 Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300>
- 745 Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic
746 assembler. *Genome Res* [Internet] 27:824–834. Available from:
747 <https://genome.cshlp.org/content/27/5/824.full>
- 748 Papetti C, Babbucci M, Dettai A, Basso A, Lucassen M, Harms L, Bonillo C, Heindler FM, Patarnello T,
749 Negrisolo E. 2021. Not Frozen in the Ice: Large and Dynamic Rearrangements in the
750 Mitochondrial Genomes of the Antarctic Fish. *Genome Biol Evol* [Internet] 13. Available from:
751 <https://academic.oup.com/gbe/article/13/3/evab017/6133229>

- 752 Parey E, Louis A, Montfort J, Guiguen Y, Crollius HR, Berthelot C. 2022. An atlas of fish genome
753 evolution reveals delayed rediploidization following the teleost whole-genome duplication.
754 *Genome Res* [Internet] 32:1685–1697. Available from:
755 <https://genome.cshlp.org/content/32/9/1685.full>
- 756 Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-
757 free profiling of polyploid genomes. *Nat Commun* [Internet] 11:1–10. Available from:
758 <https://doi.org/10.1038/s41467-020-14998-3>
- 759 Saccone C, Gissi C, Reyes A, Larizza A, Sbisà E, Pesole G. 2002. Mitochondrial DNA in metazoa: degree
760 of freedom in a frozen event. *Gene* 286:3–12.
- 761 Satoh TP, Miya M, Mabuchi K, Nishida M. 2016. Structure and variation of the mitochondrial genome
762 of fishes. *BMC Genomics* [Internet] 17:719. Available from:
763 <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-3054-y>
- 764 Sharbrough J, Bankers L, Cook E, Fields PD, Jalinsky J, Mcelroy KE, Neiman M, Logsdonjr JM, Boore JL.
765 2023. Single-molecule Sequencing of an Animal Mitochondrial Genome Reveals Chloroplast-like
766 Architecture and Repeat-mediated Recombination. *Mol Biol Evol* [Internet] 40. Available from:
767 <https://academic.oup.com/mbe/article/40/1/msad007/6980790>
- 768 Shtolz N, Mishmar D. 2023. The metazoan landscape of mitochondrial DNA gene order and content is
769 shaped by selection and affects mitochondrial transcription. *Communications Biology* 2023 6:1
770 [Internet] 6:1–15. Available from: <https://www.nature.com/articles/s42003-023-04471-4>
- 771 Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies.
772 *PLoS Comput Biol* [Internet] 18:e1009802. Available from:
773 <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009802>

774 Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from
775 short and long sequencing reads. Phillippy AM, editor. *PLoS Comput Biol* [Internet] 13:e1005595.
776 Available from: <https://dx.plos.org/10.1371/journal.pcbi.1005595>

777 Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome
778 assemblies. *Bioinformatics* [Internet] 31:3350–3352. Available from:
779 <https://doi.org/10.1093/bioinformatics/btv383>

780 Zhang J, Kan X, Miao G, Hu S, Sun Q, Tian W. 2020. qMGR: A new approach for quantifying
781 mitochondrial genome rearrangement. *Mitochondrion* 52:20–23.

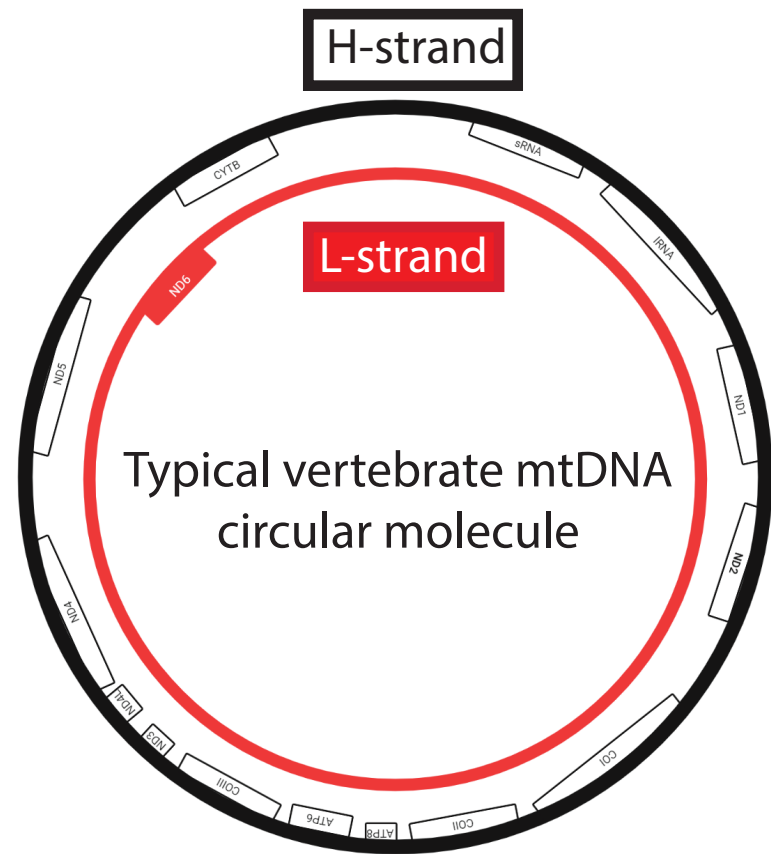
782 Zimin A V., Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections
783 in genome assemblies. *PLoS Comput Biol* [Internet] 16:e1007981. Available from:
784 <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007981>

785

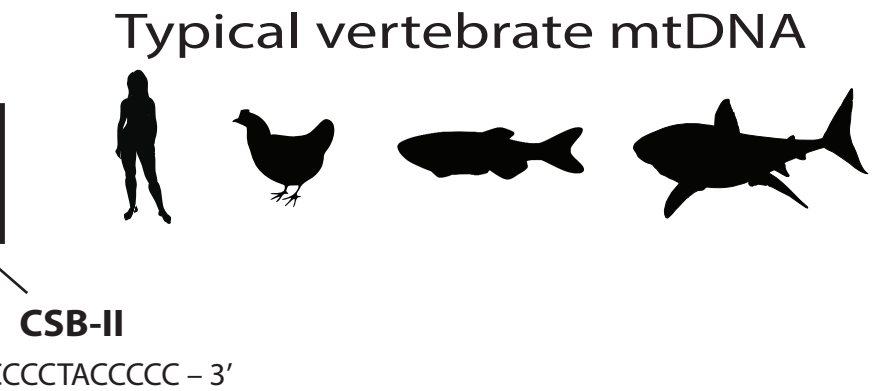
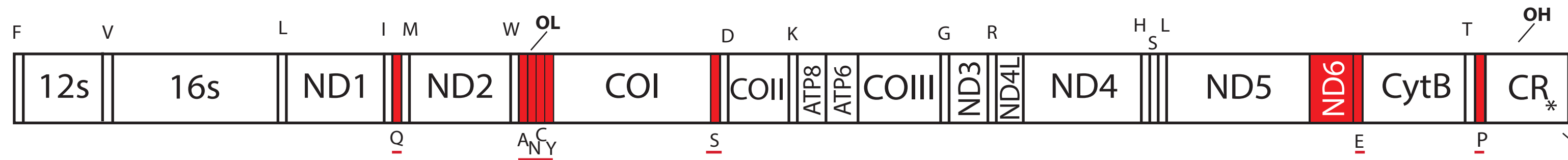
786

787

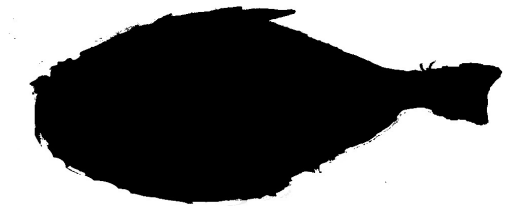
788



~16,000 kb

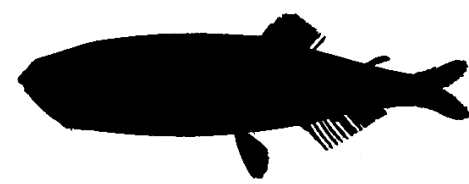


Polyipnus polli



~16,773 kb

Maurolicus muelleri

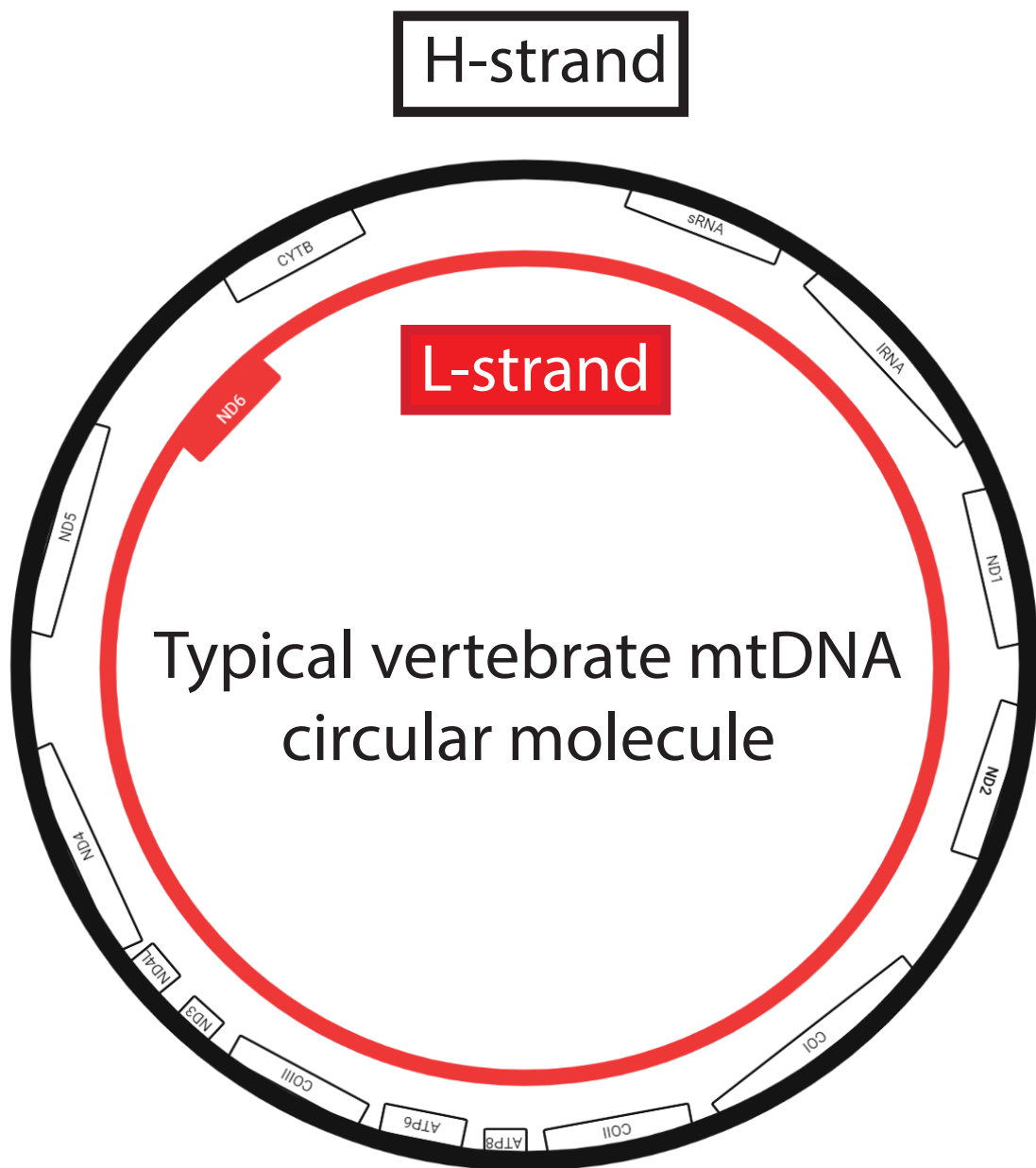


~15,230 kb

Argyroteleus affinis



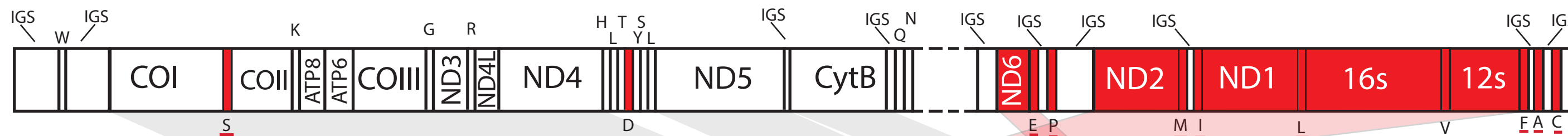
~15,489 kb



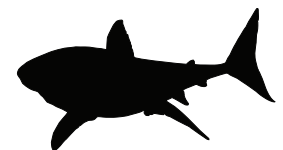
Argyropelecus hemigymnus



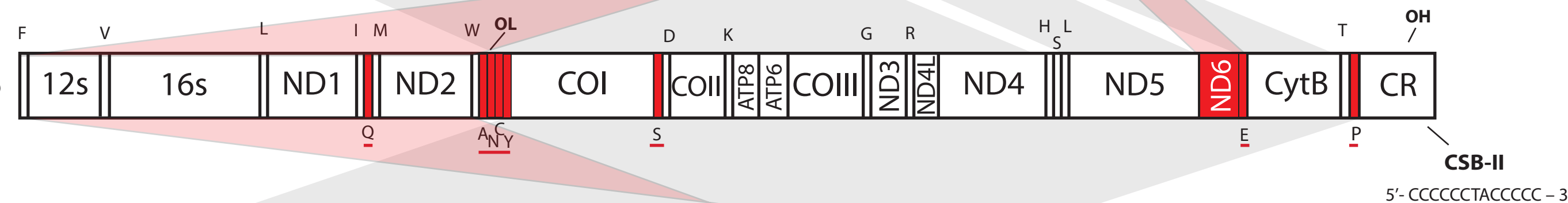
~20,747 kb



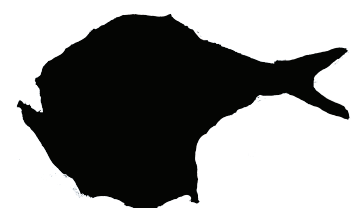
Typical vertebrate mtDNA



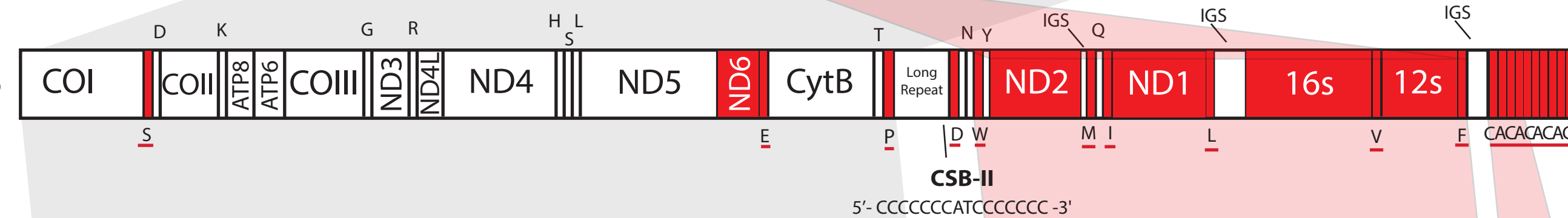
~16,000 kb



Argyropelecus aculeatus



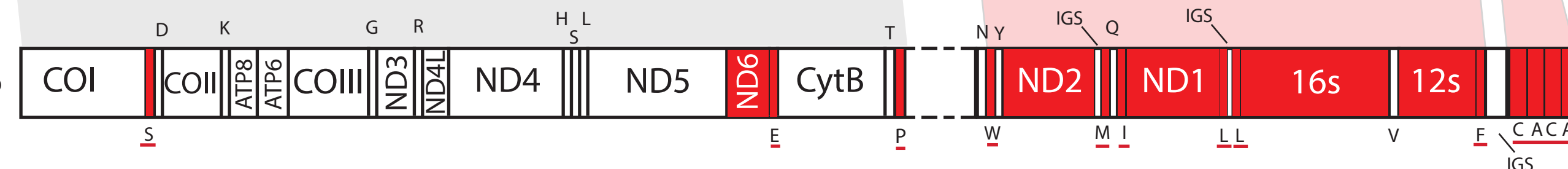
~23,291 kb

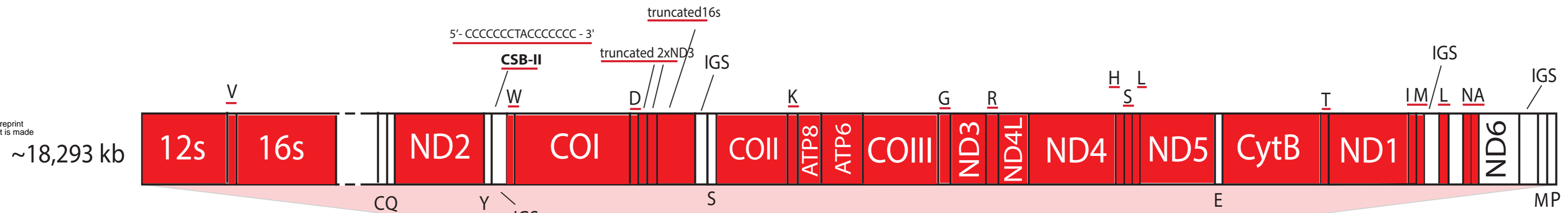
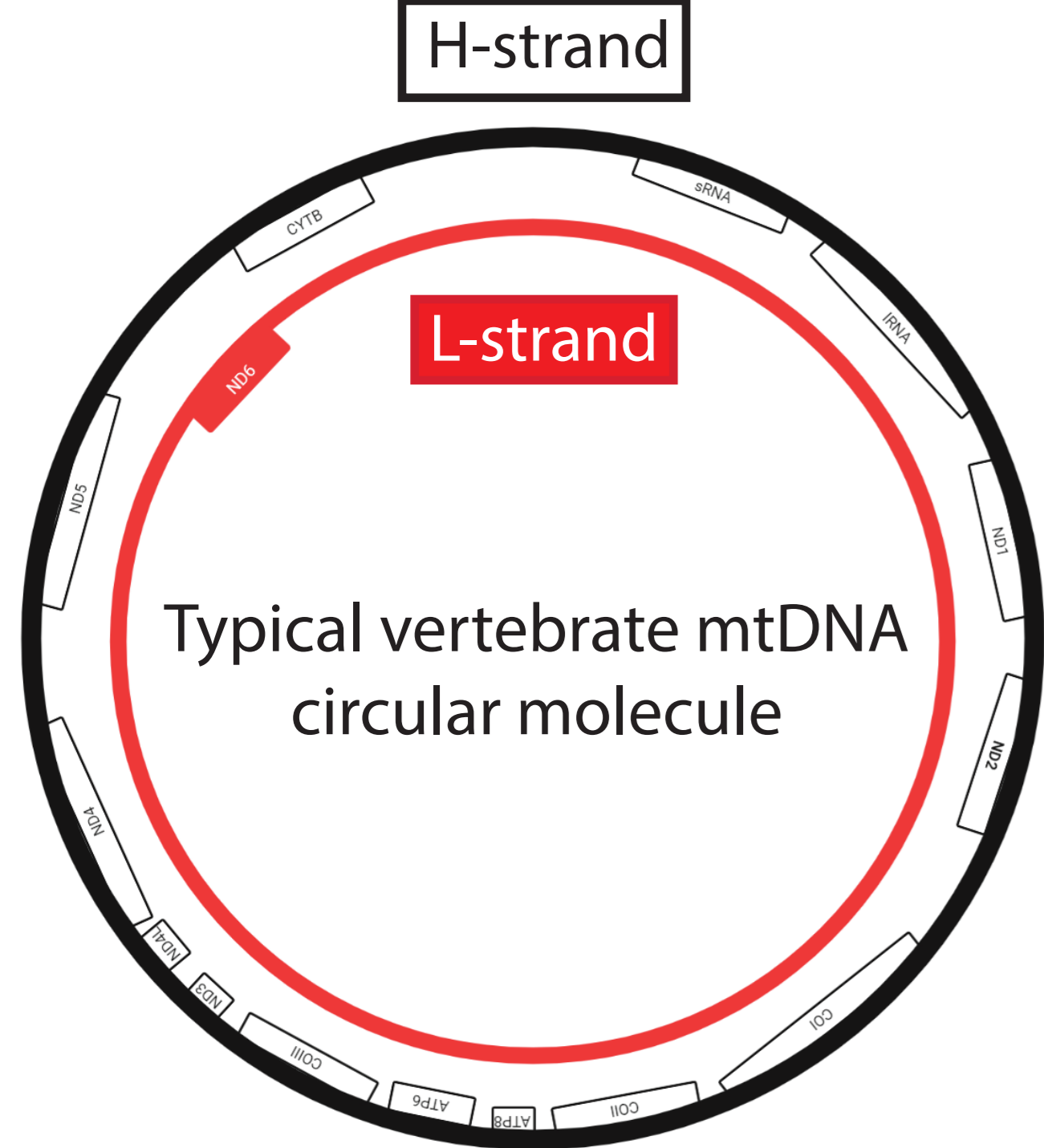


Argyropelecus olfersii

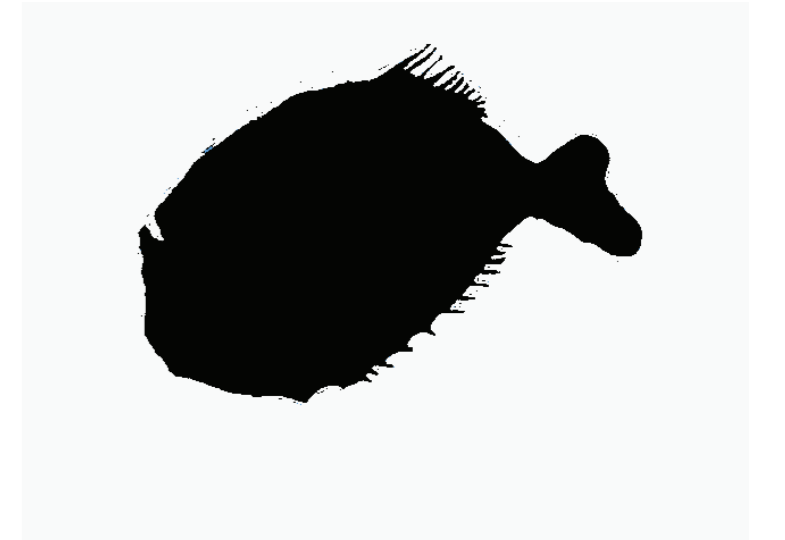


~16,414 kb

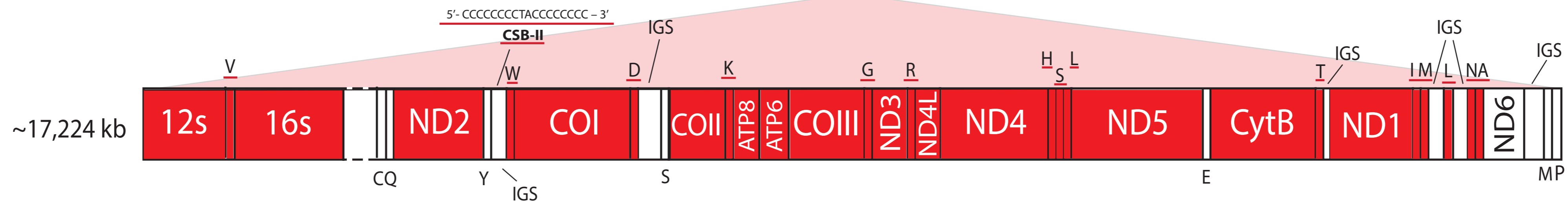
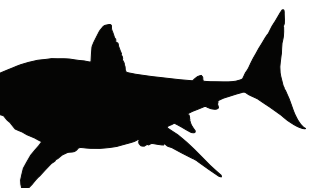
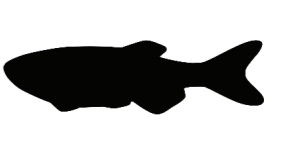
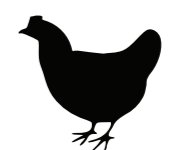
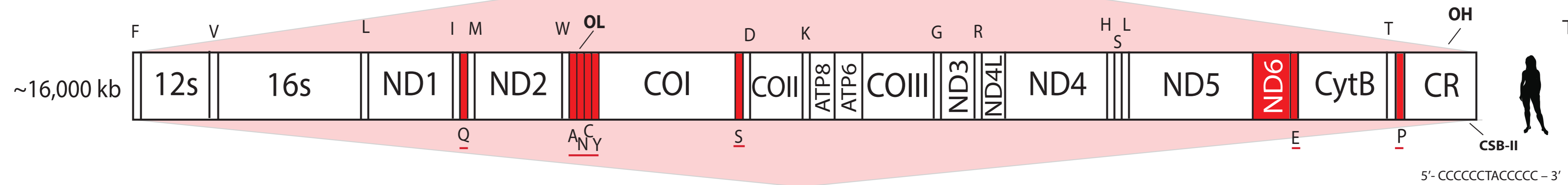




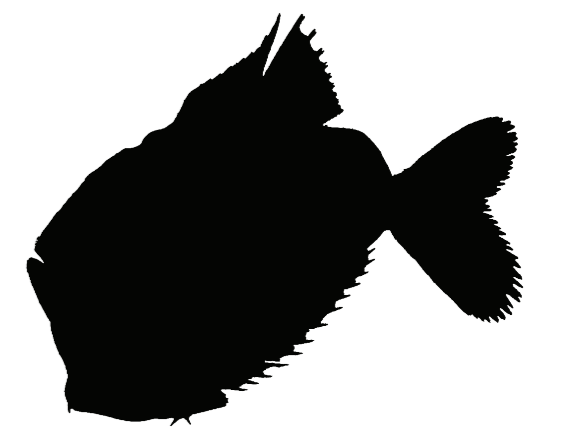
Sternoptyx obscura



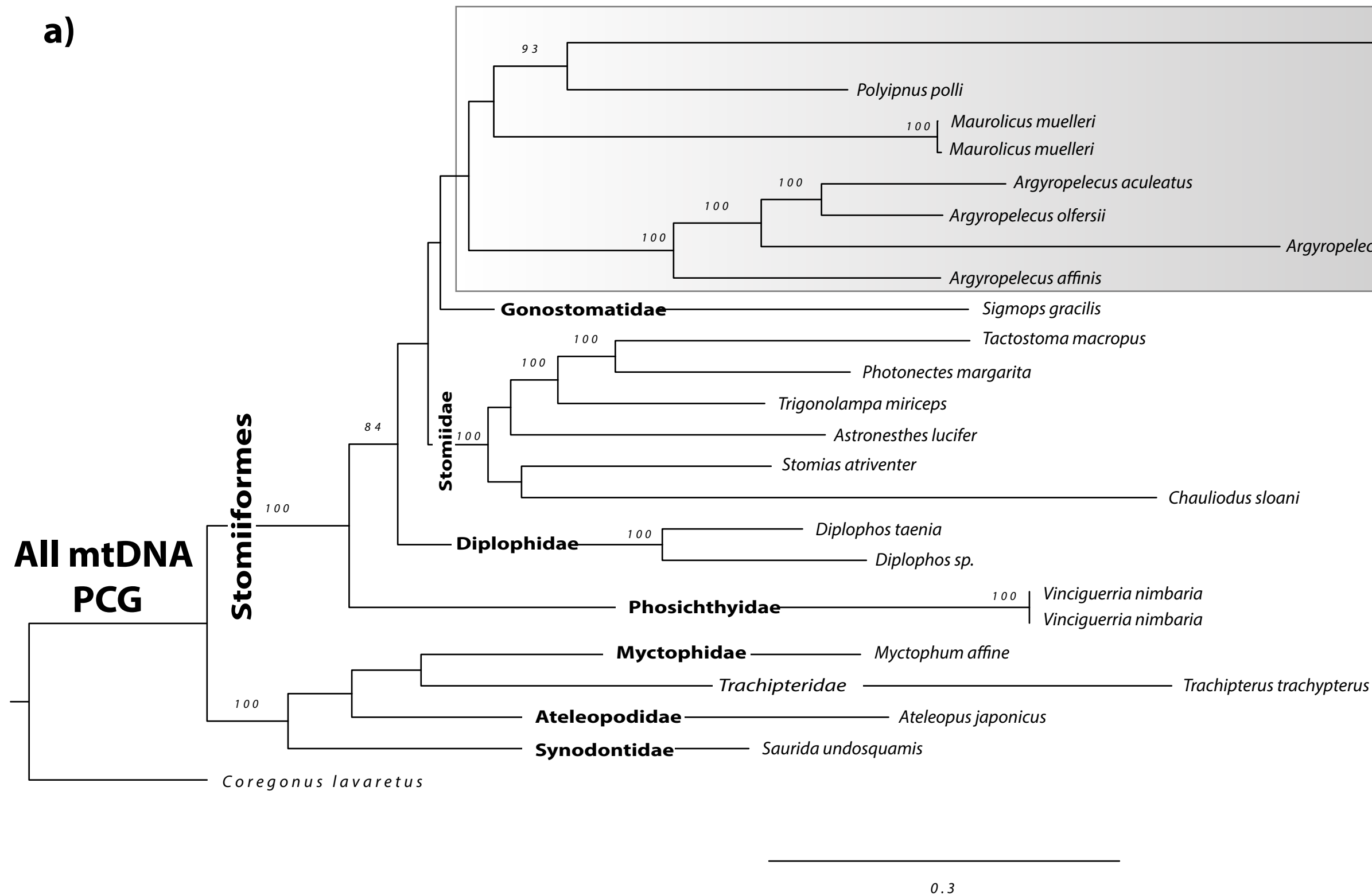
Typical vertebrate mtDNA



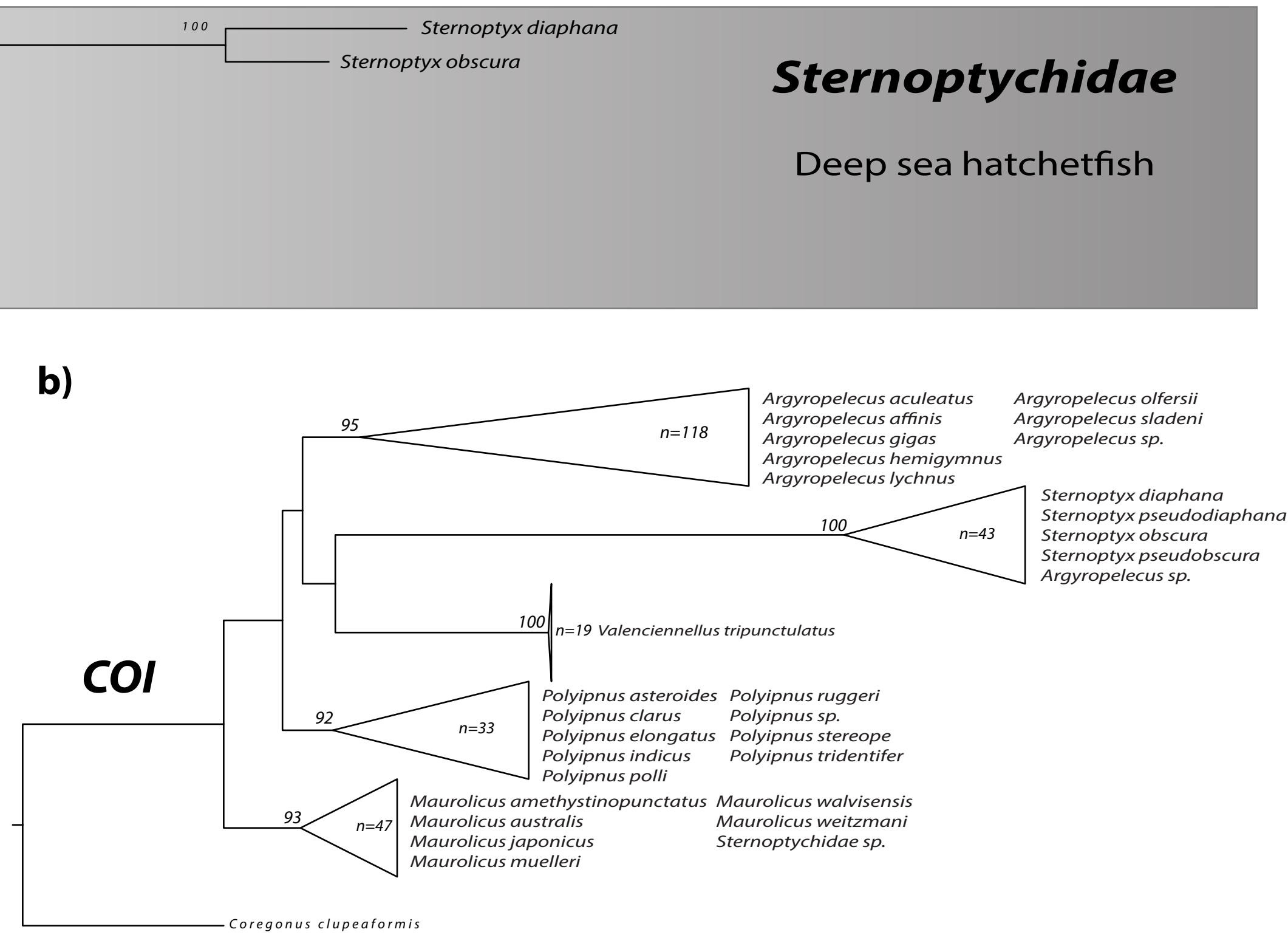
Sternoptyx diaphana



a)

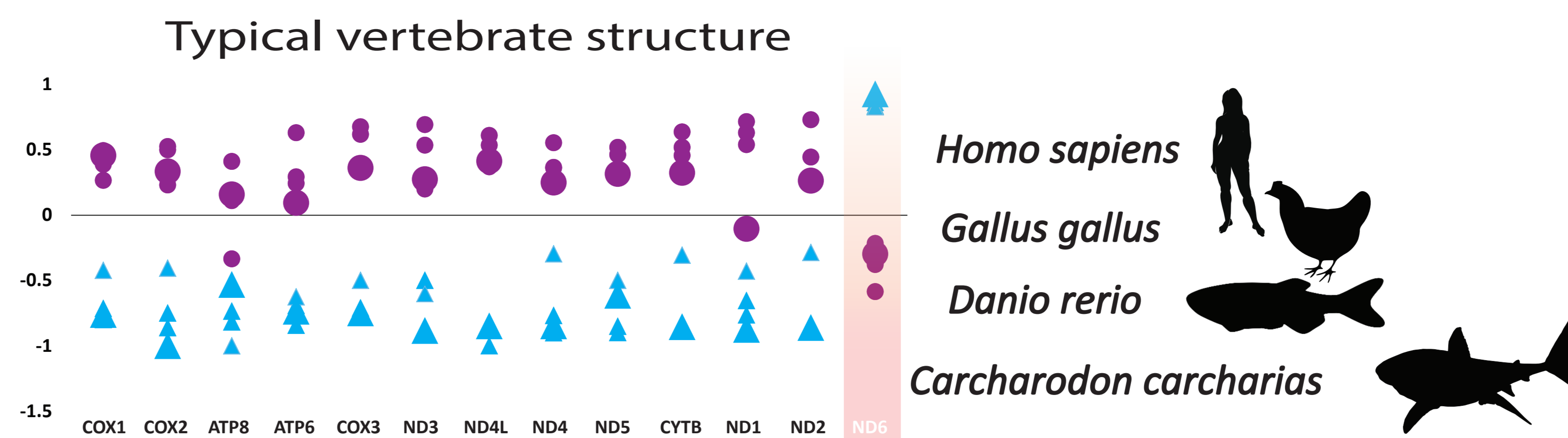
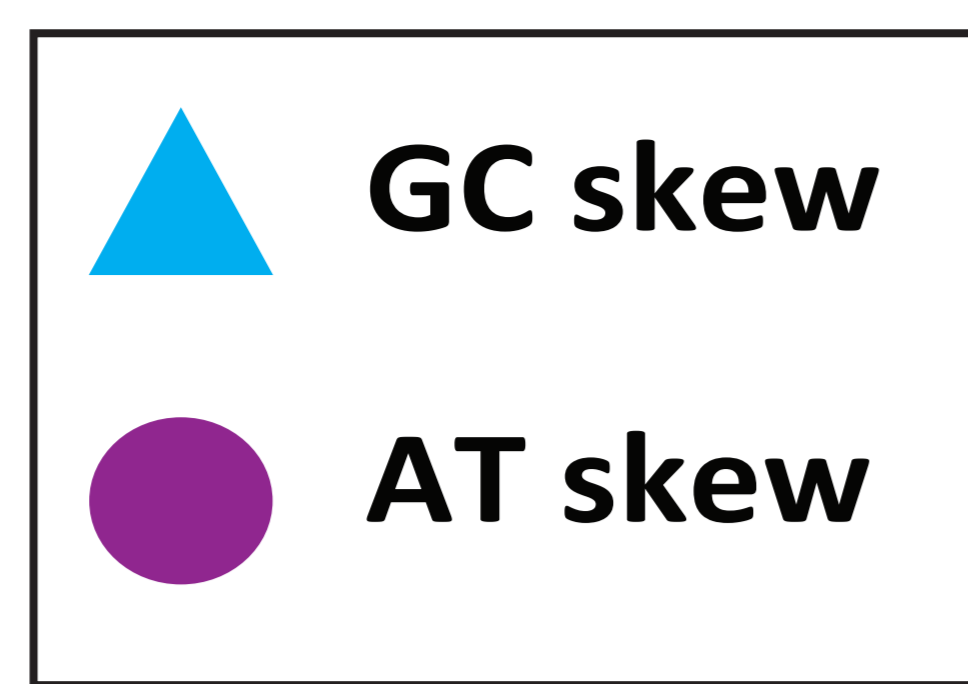


b)

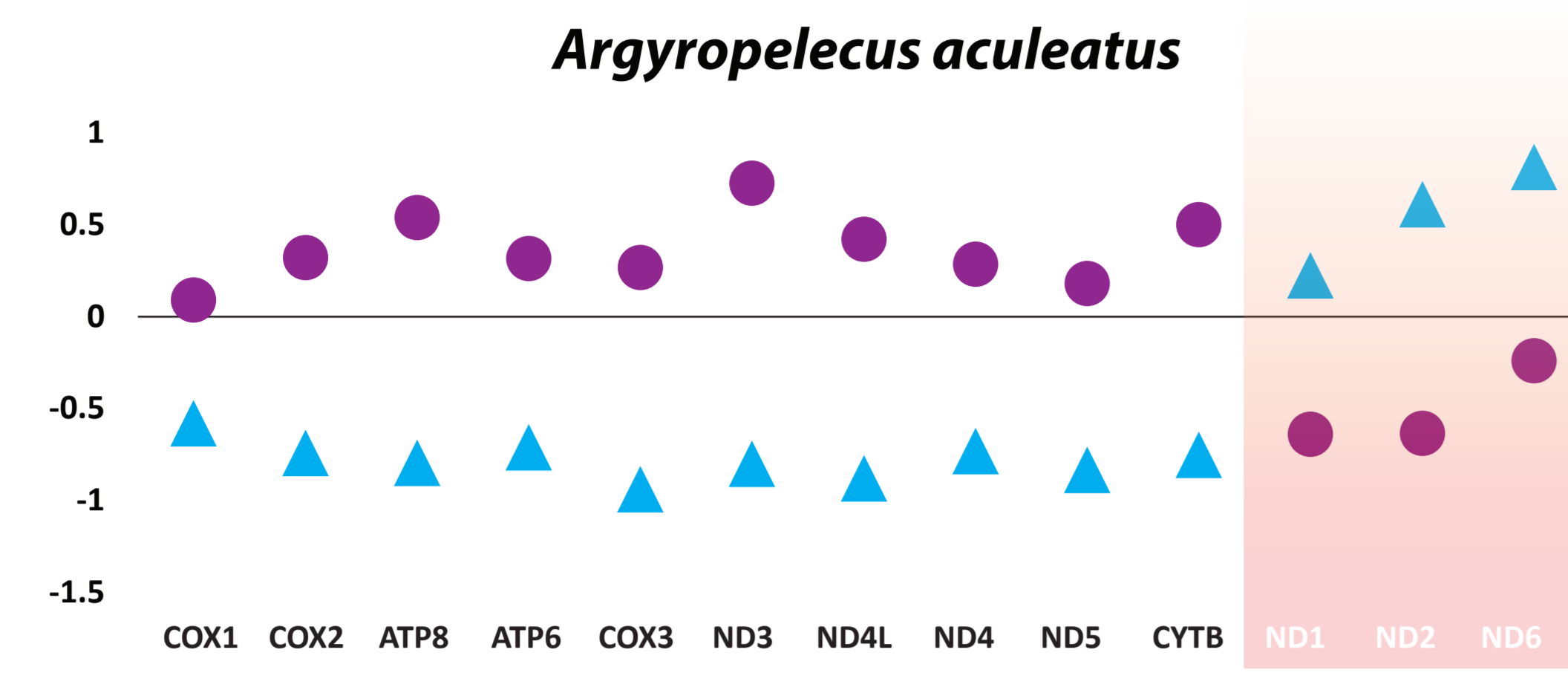
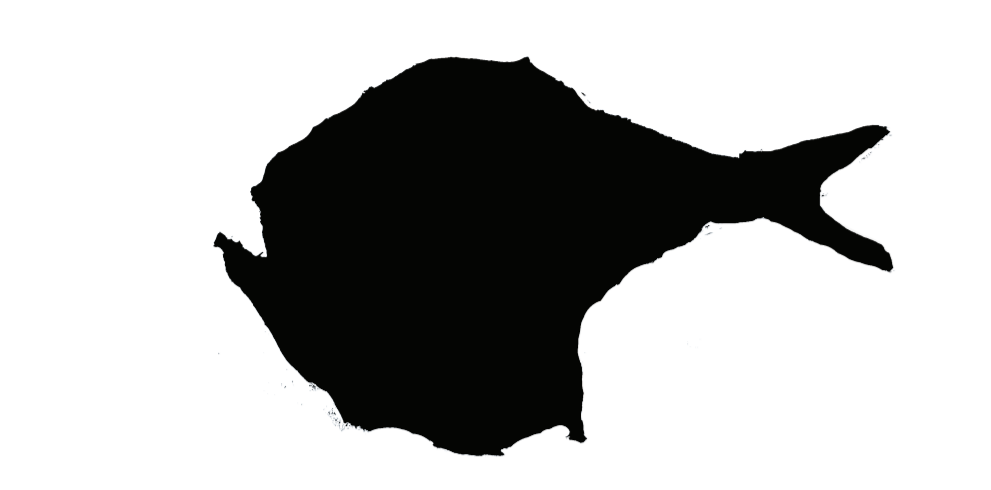


Sternoptychidae

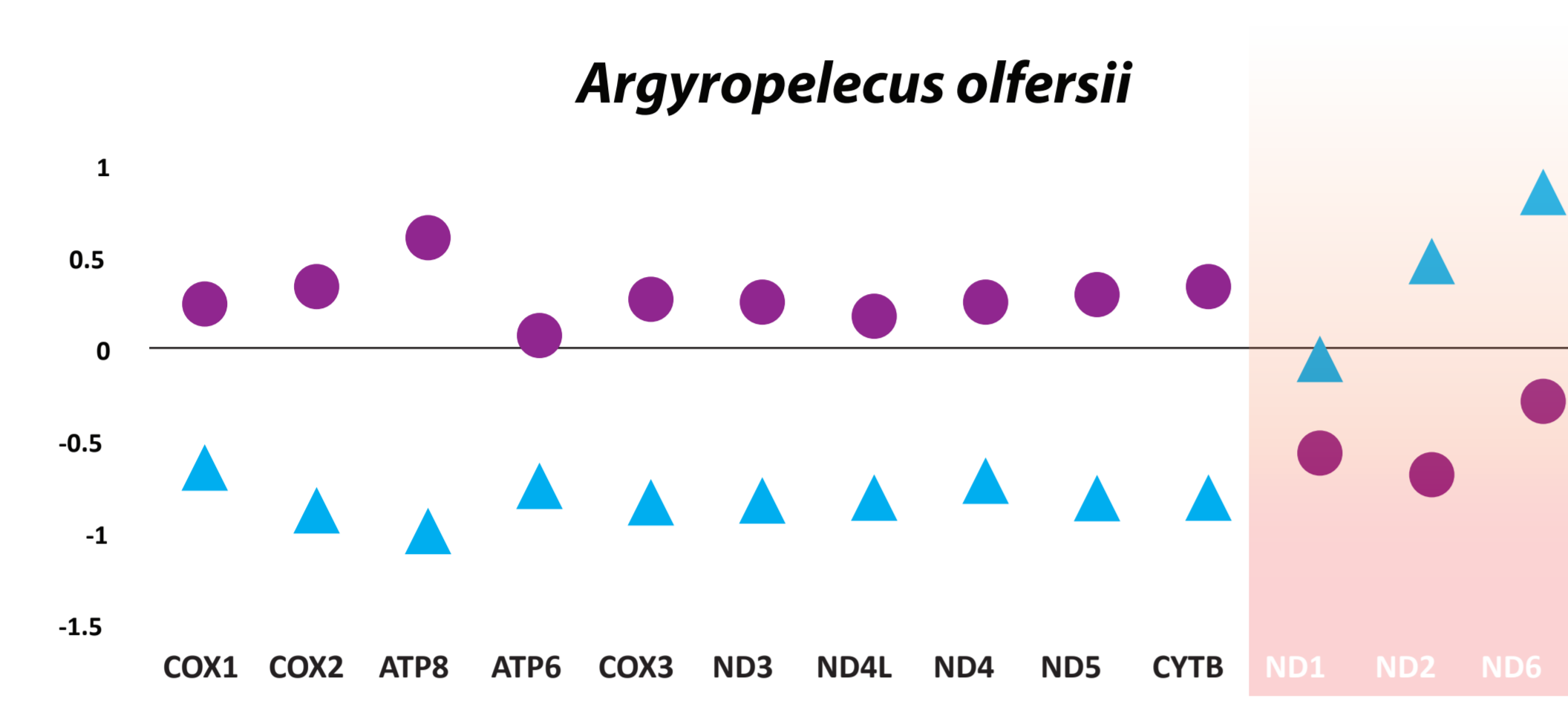
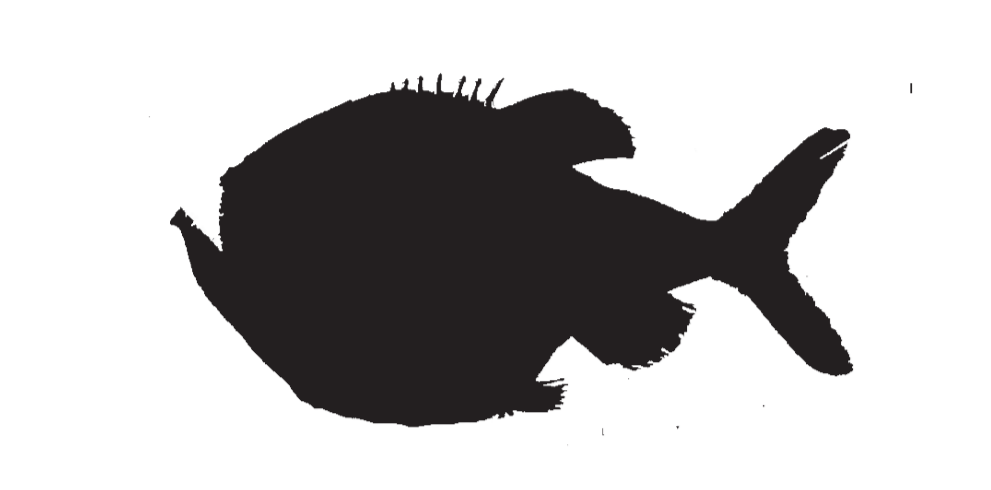
Deep sea hatchetfish



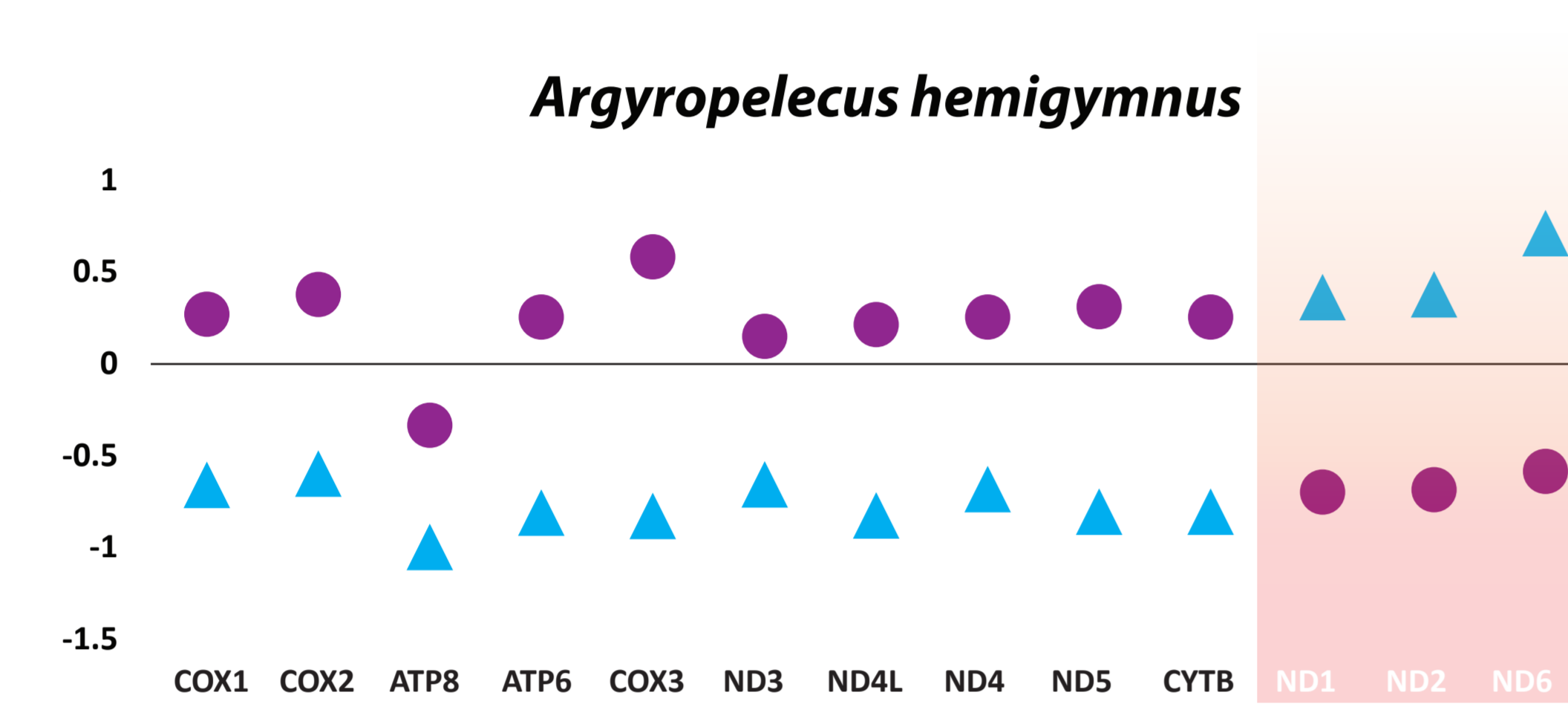
Argyrolepecus aculeatus



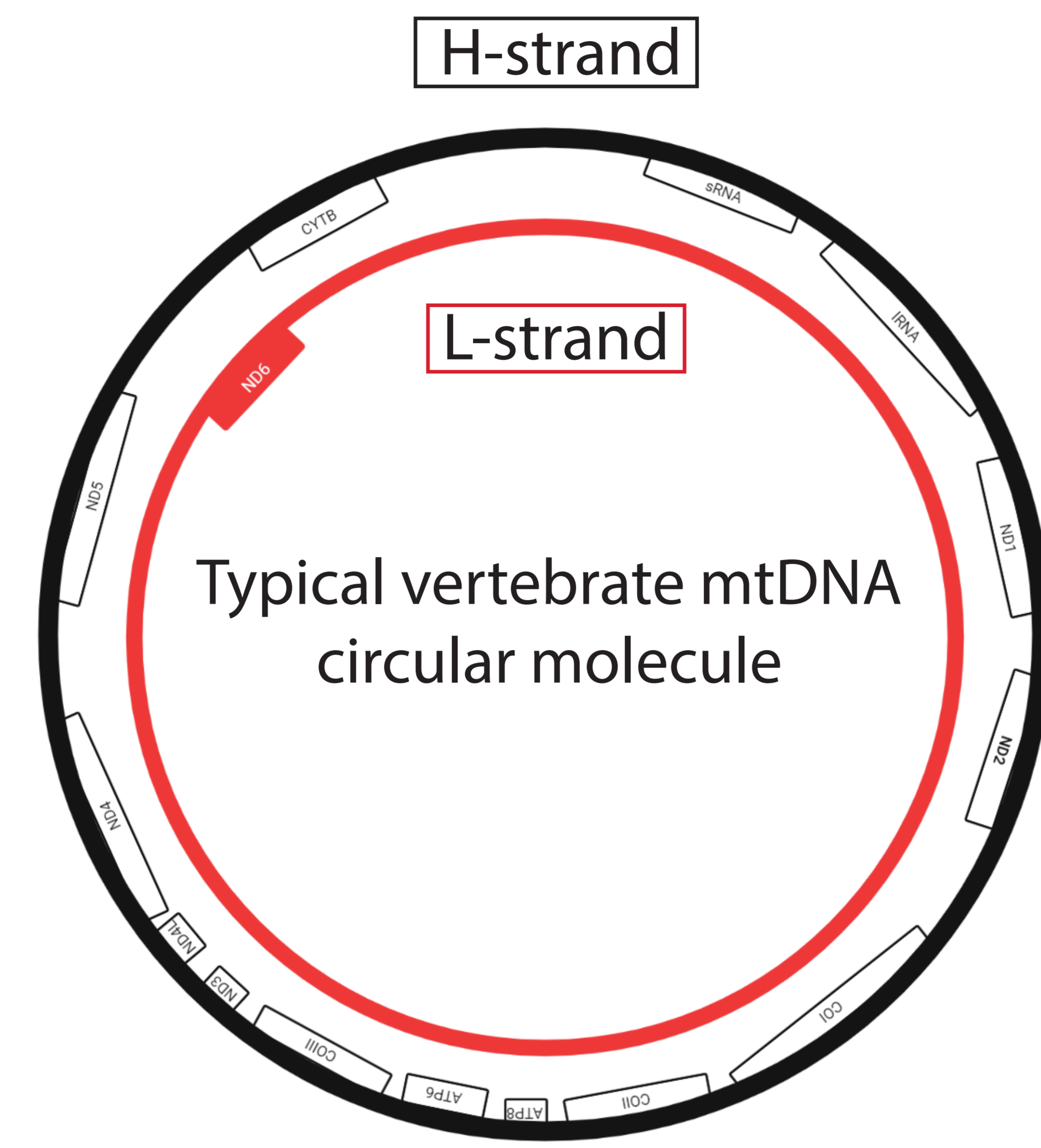
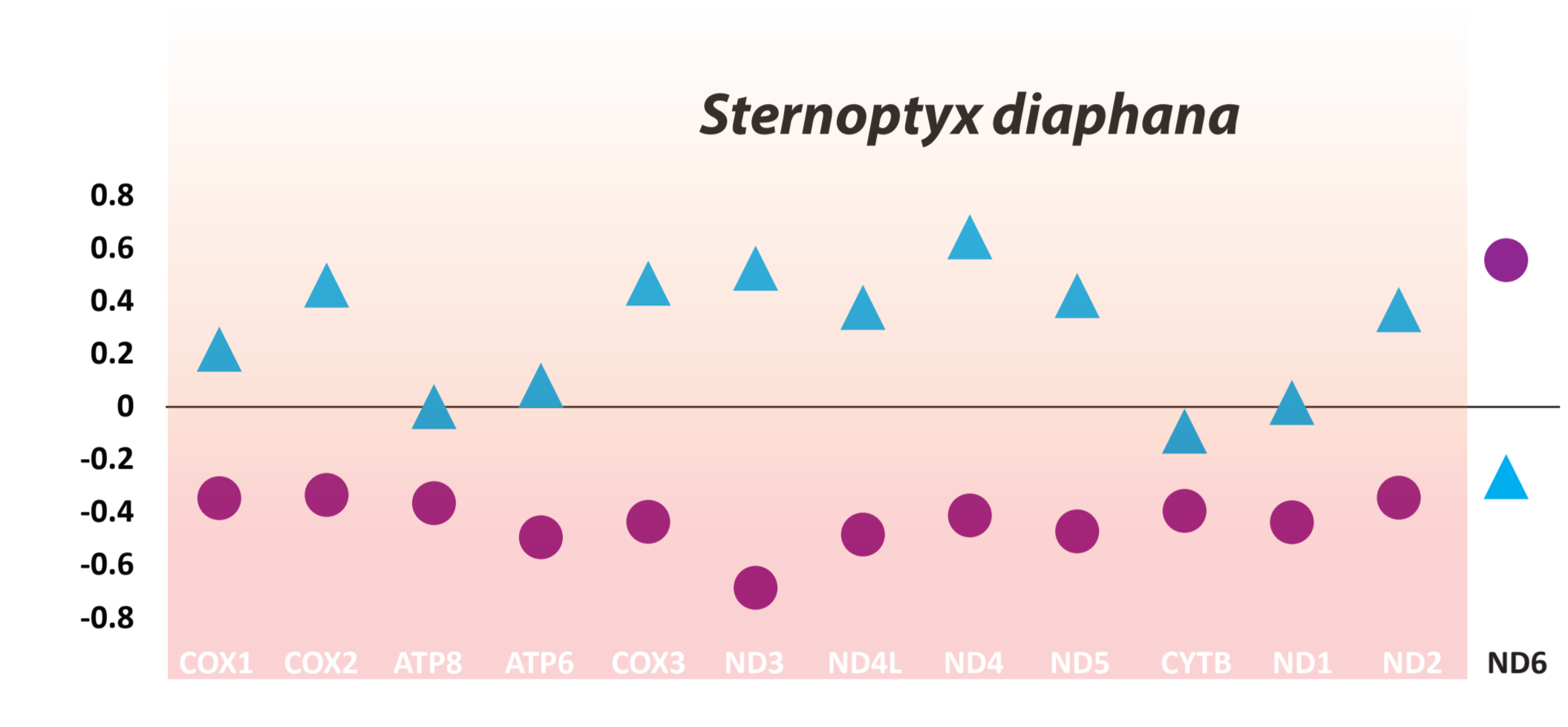
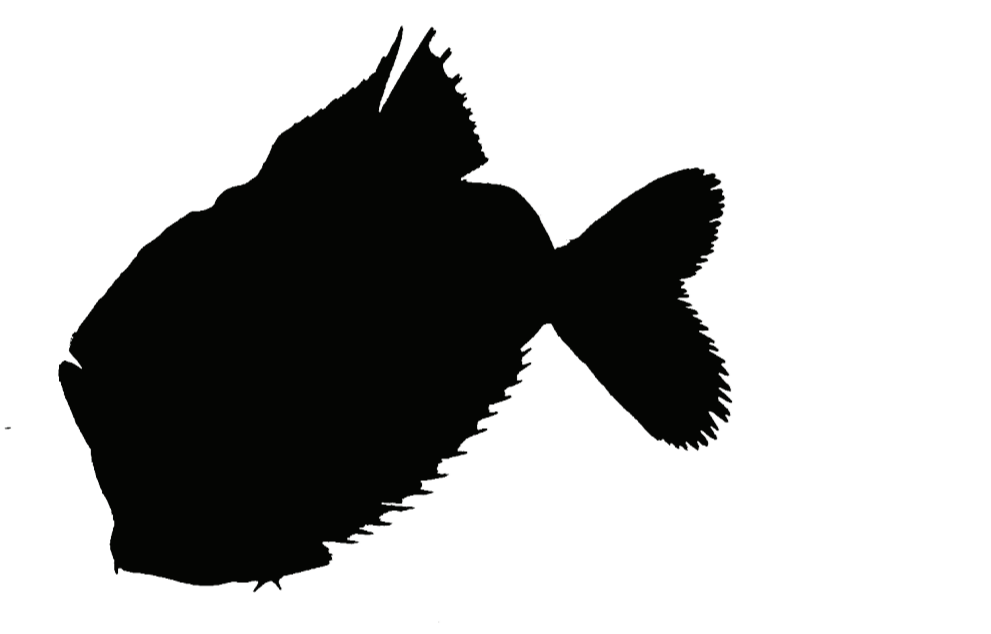
Argyrolepecus olfersii



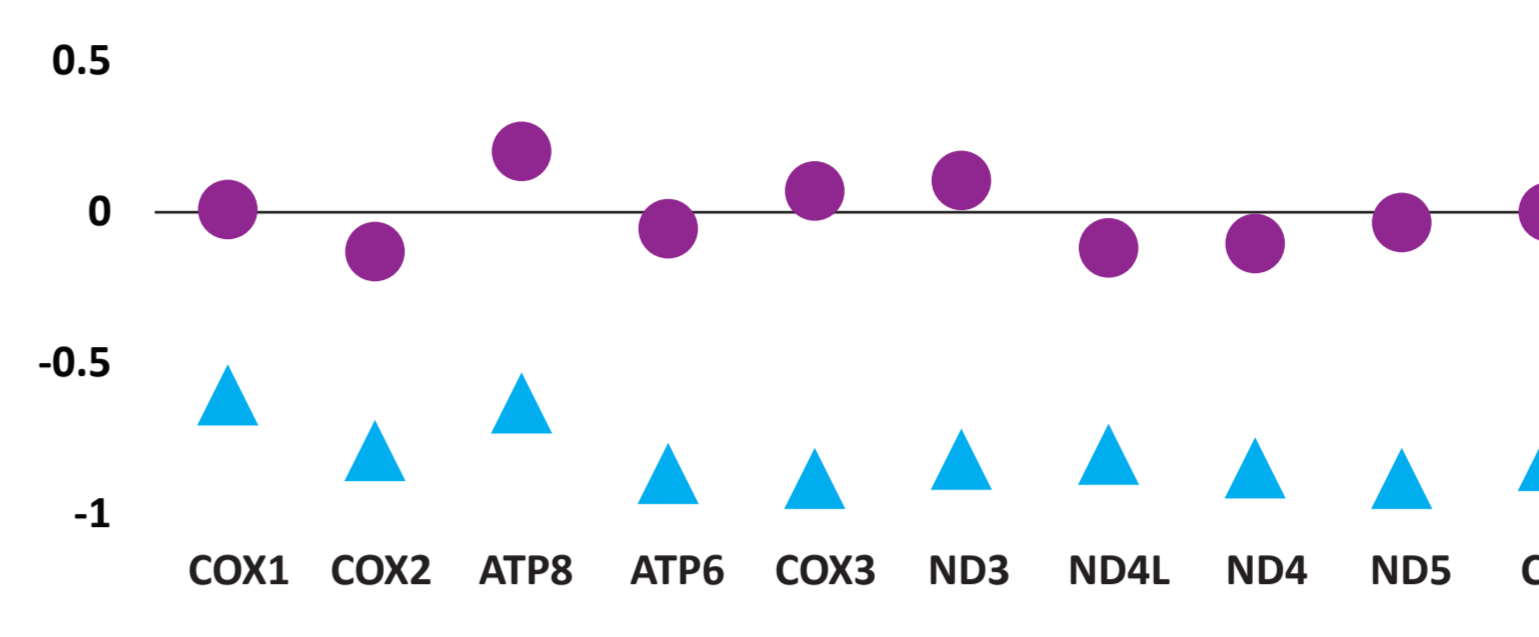
Argyrolepecus hemigymnus



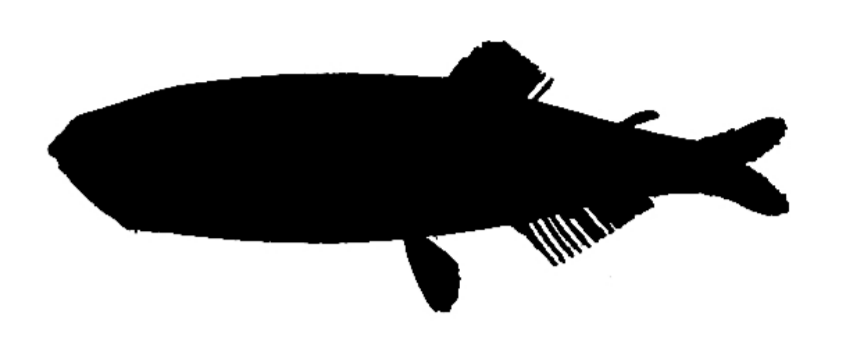
Sternoptyx diaphana (MT588184.1)



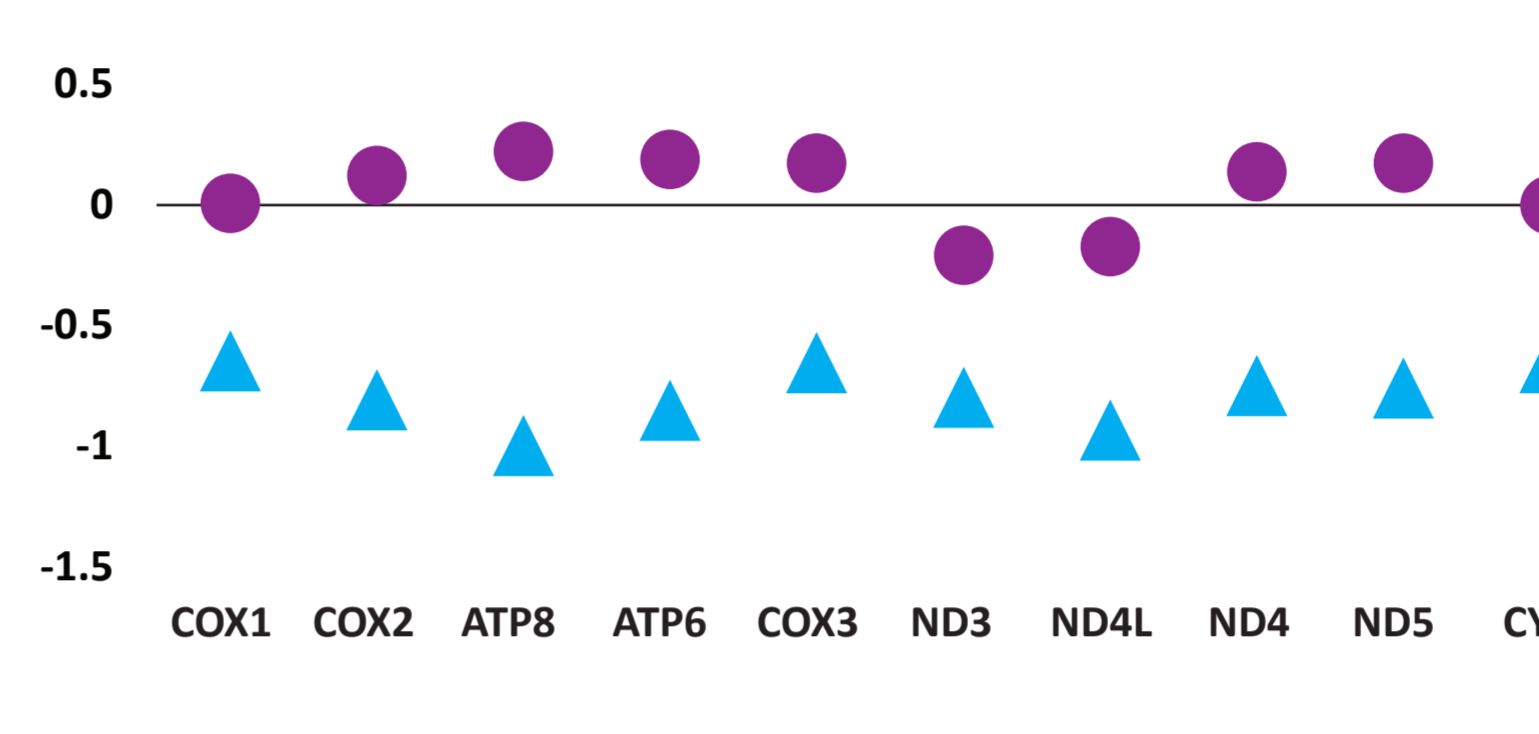
Maurolicus muelleri



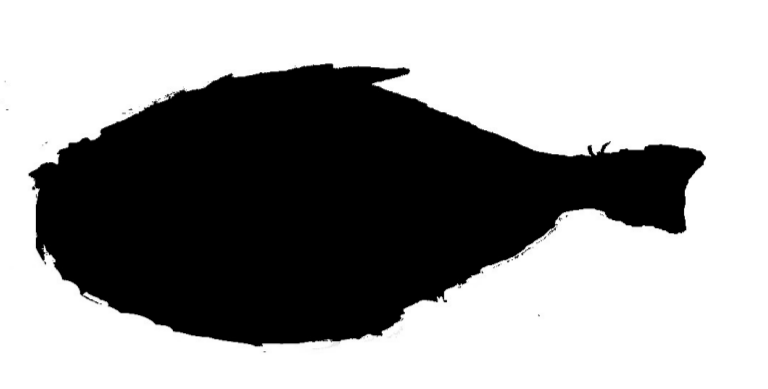
Maurolicus muelleri



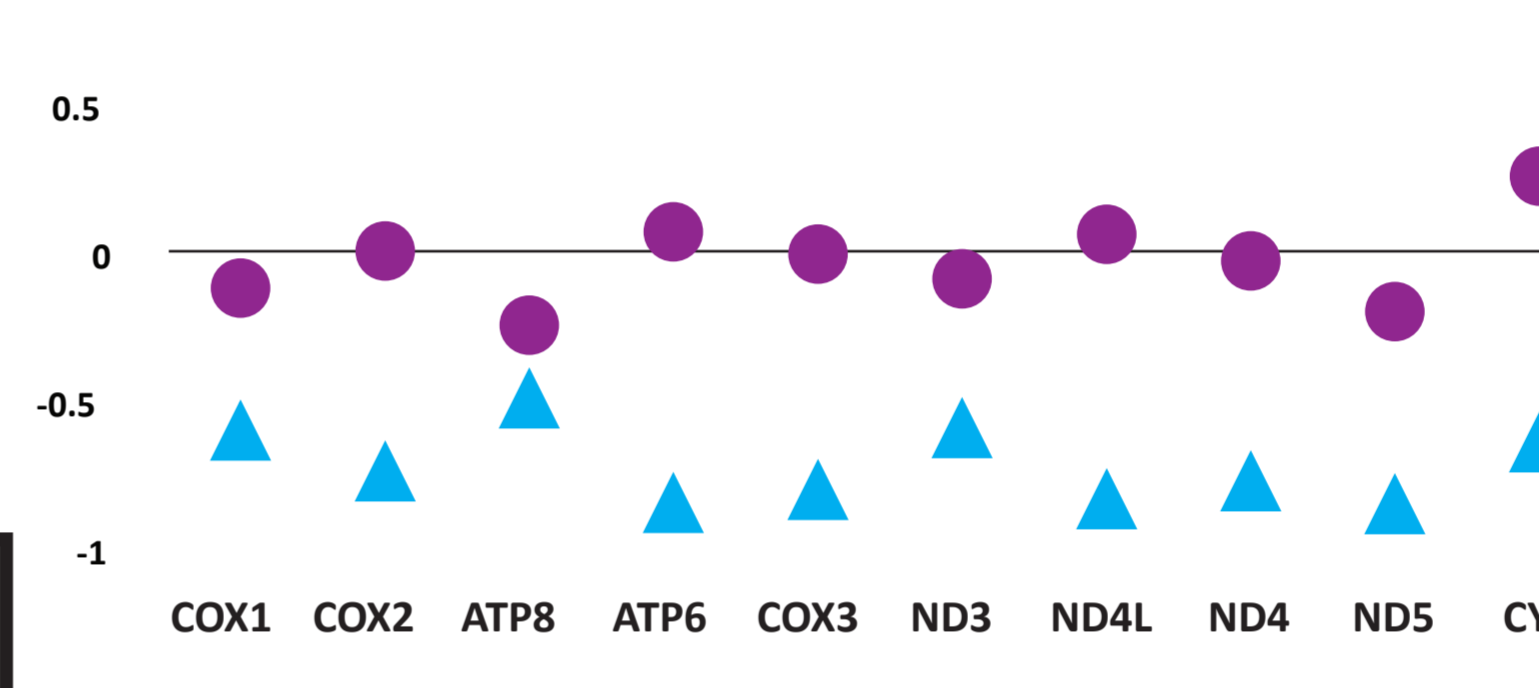
Polyipnus polli



Polyipnus polli (AP012962)



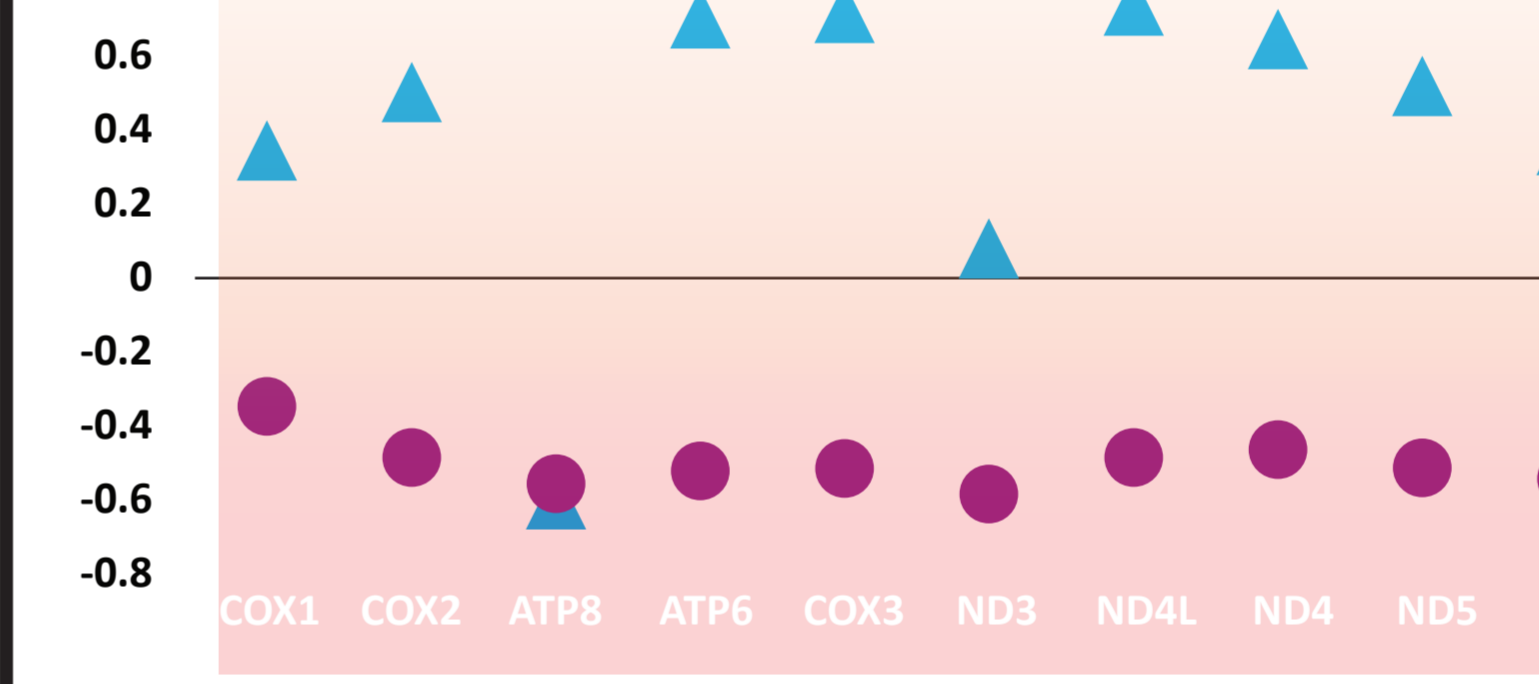
Argyrolepecus affinis



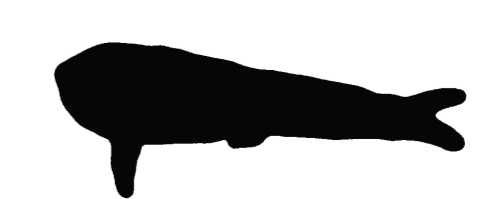
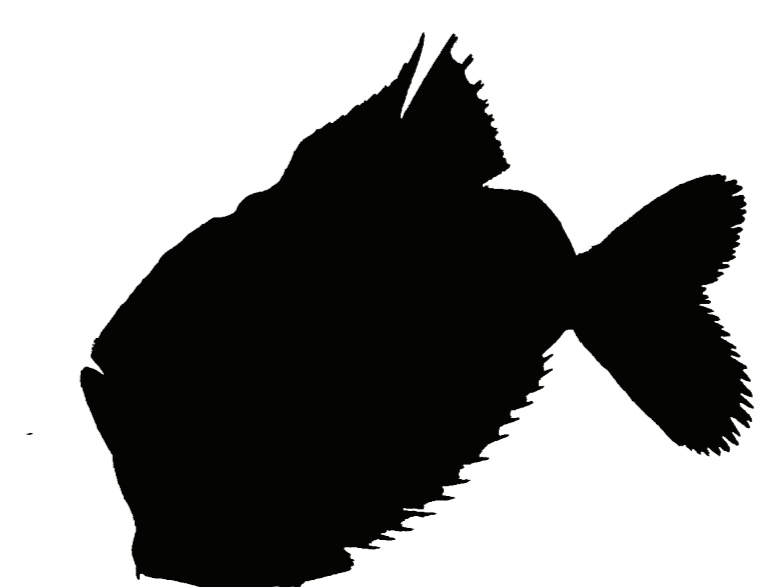
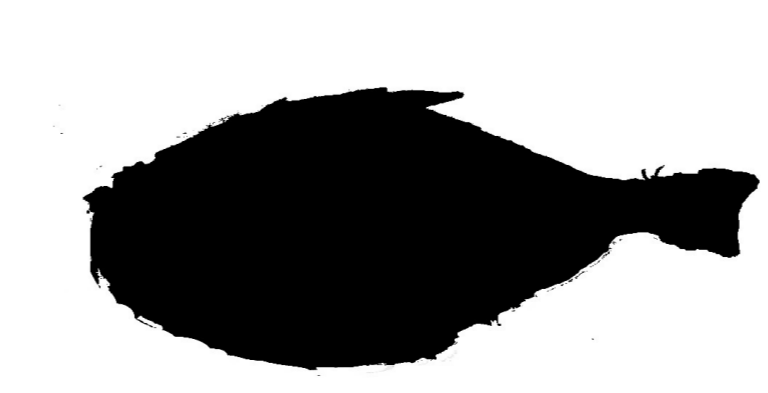
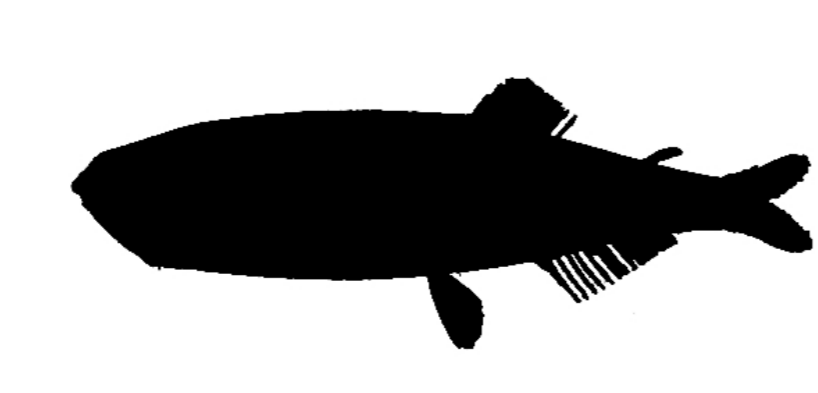
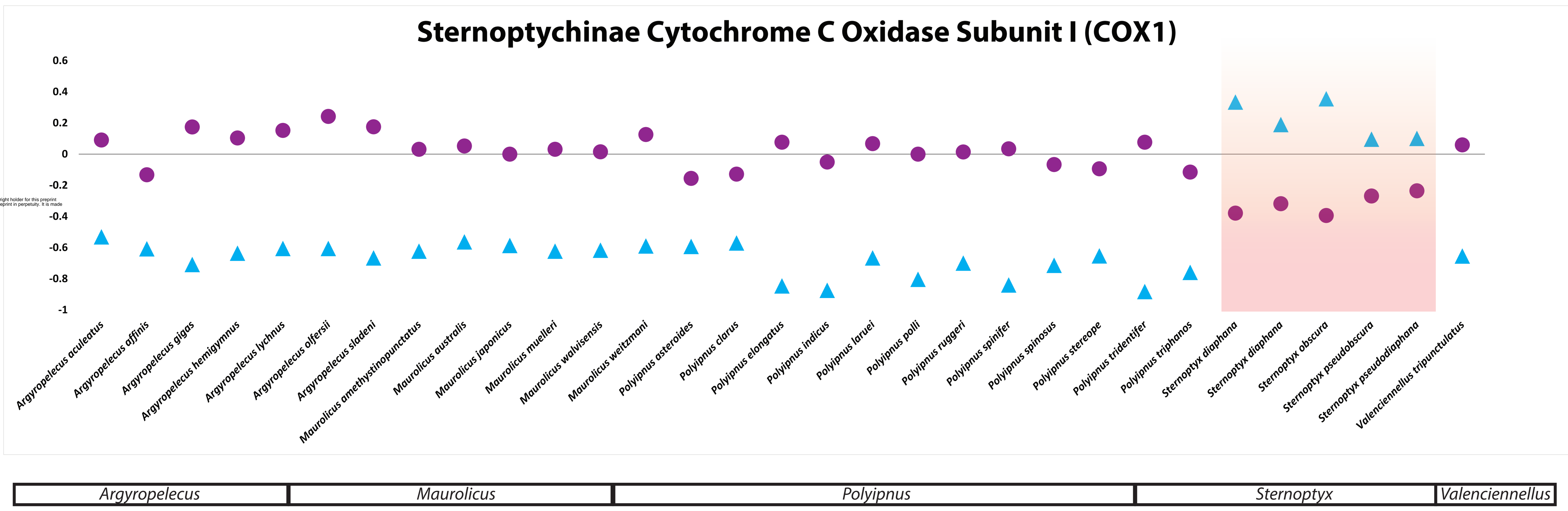
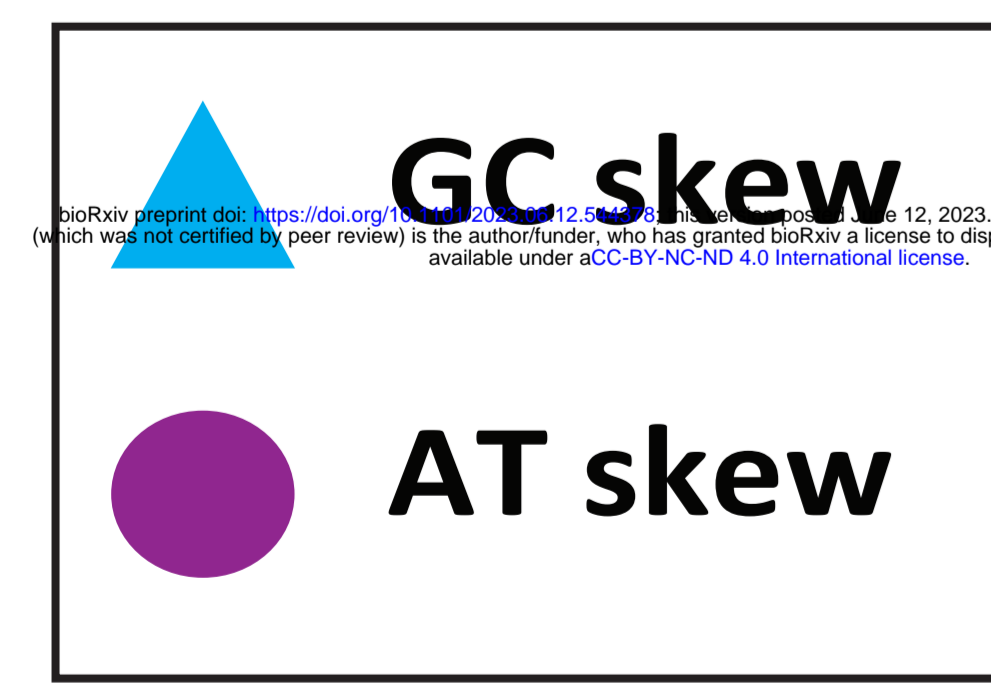
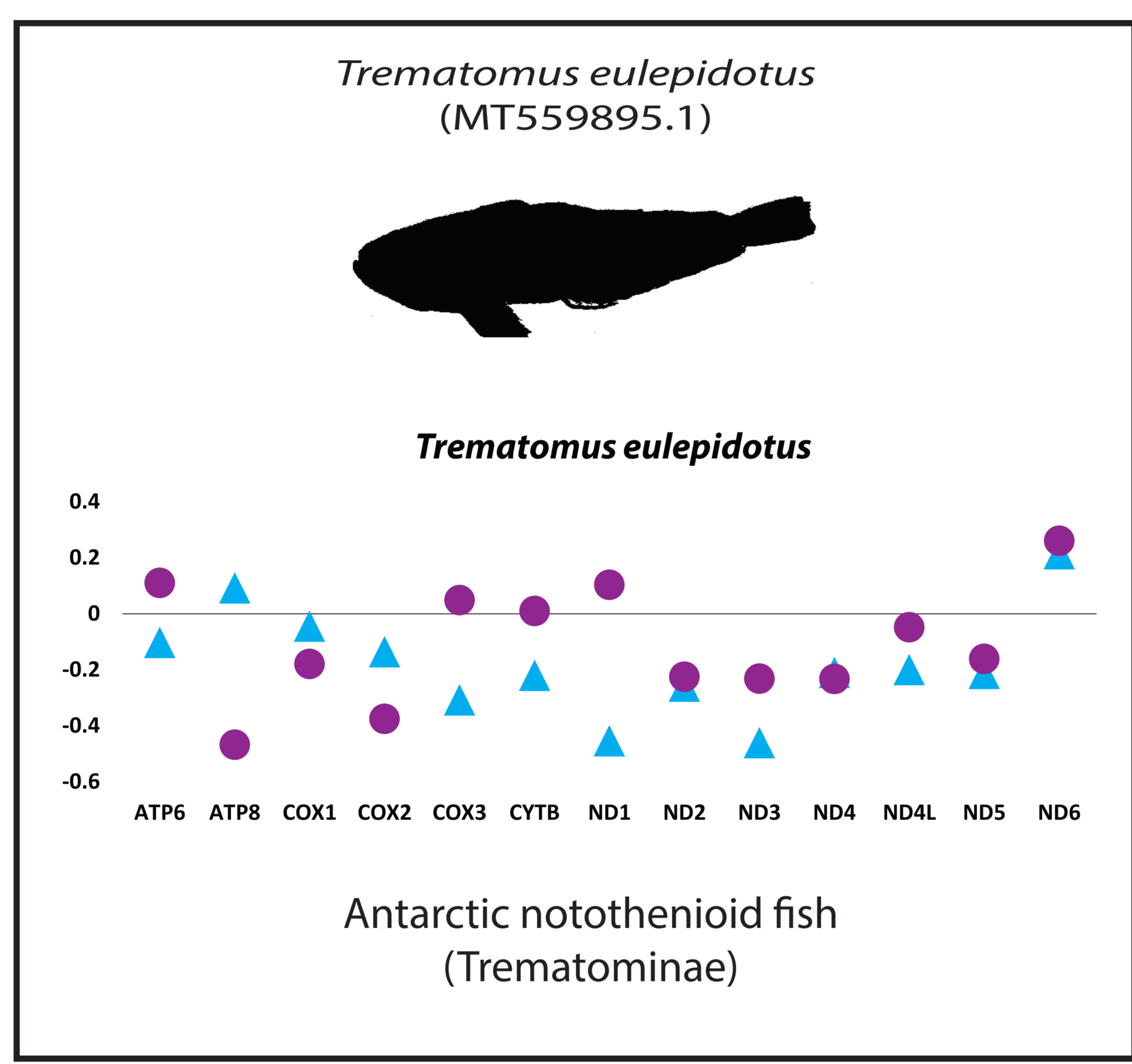
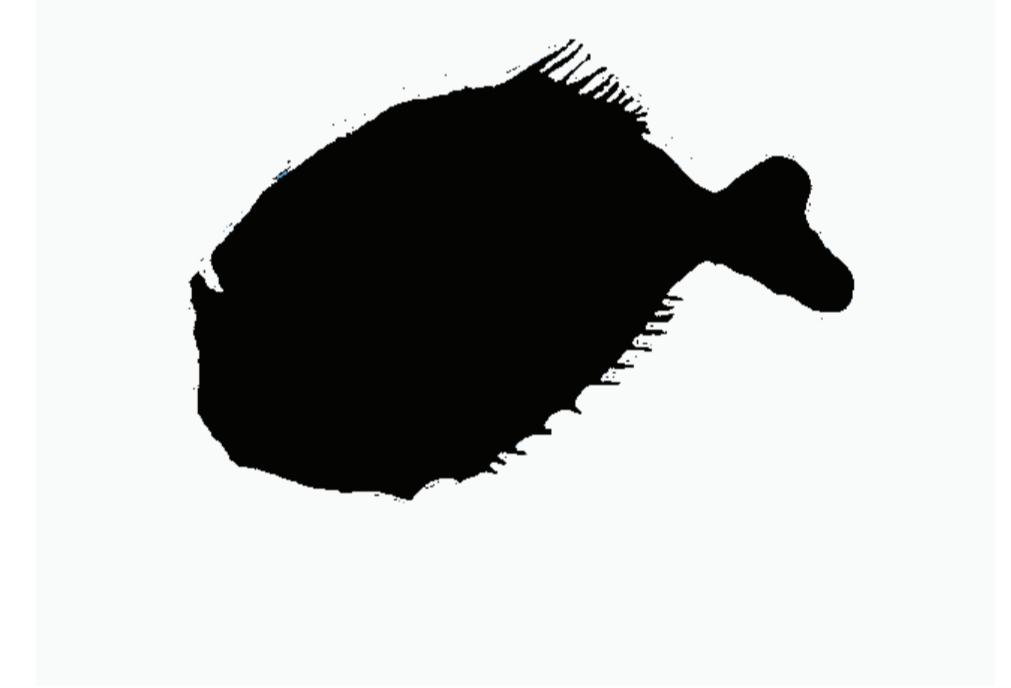
Argyrolepecus affinis (AP012964.1)

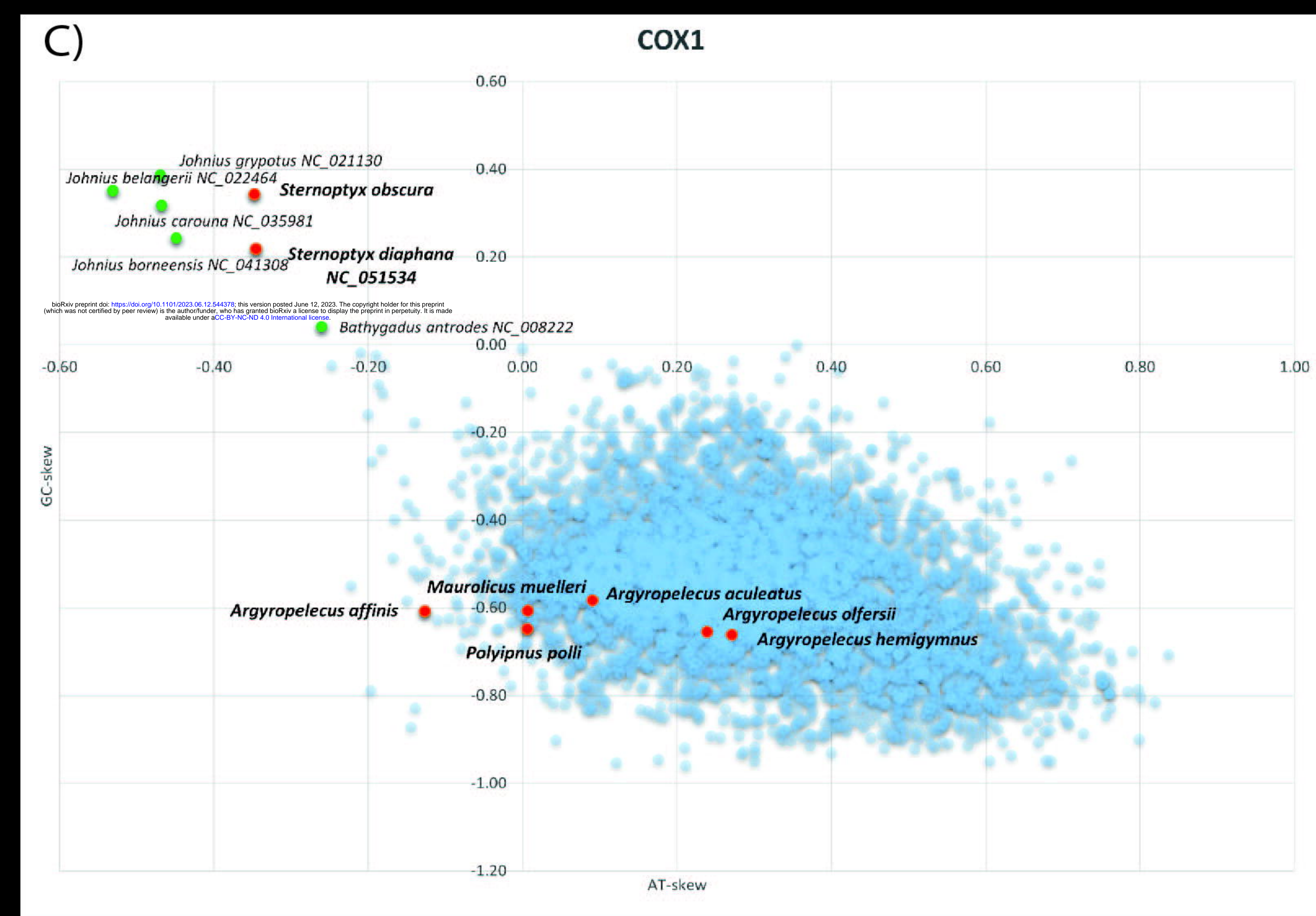
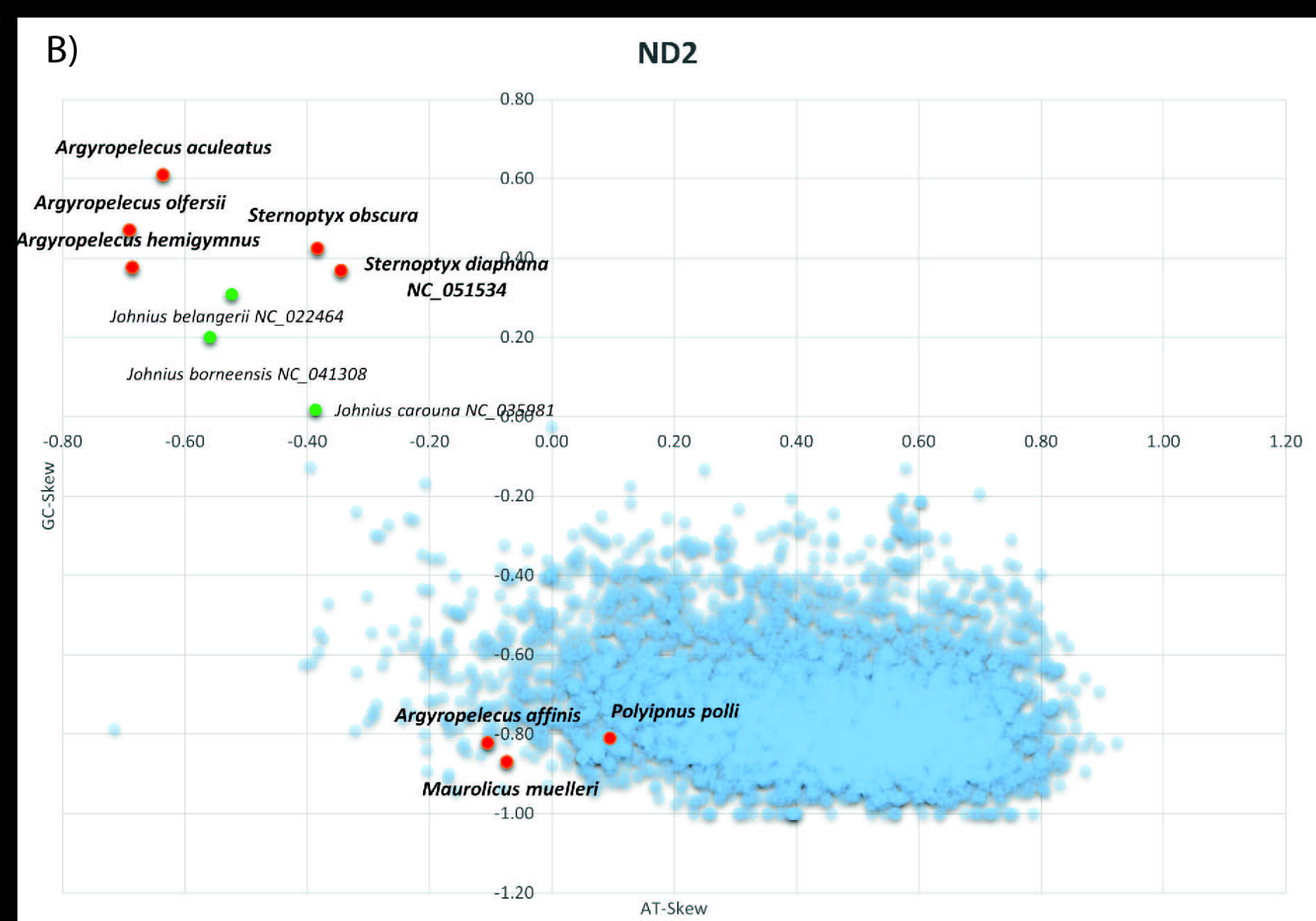


Sternoptyx obscura (OP057081.1)



Sternoptyx obscura (OP057081.1)





Sternoptychidae	<i>Argyropelecus affinis</i>	AP012964.1	-	-	15,489
Sternoptychidae	<i>Argyropelecus hemigymnus</i> *	OR062952	SAMN35453526	SRR24764802	20,747
Sternoptychidae	<i>Argyropelecus olfersii</i> *	OR062953 and OR062954	SAMN35453527	SRR24764803	16,414
Sternoptychidae	<i>Maurolicus muelleri</i>	AP012963.1	-	-	15,230
Sternoptychidae	<i>Maurolicus muelleri</i> *	OR062955	SAMN35453528	SAMN35453528	16,604
Sternoptychidae	<i>Polyipnus polli</i>	AP012962.1	-	SRR24764801	16,773
Sternoptychidae	<i>Sternoptyx diaphana</i>	MT588184.1	SAMN35453678	SRR24764797-SRR24764800	17,224
Sternoptychidae	<i>Sternoptyx obscura</i>	OP057081.1	-	-	18,293
Gonostomatidae	<i>Sigmops gracilis</i>	AB016274.1	-	-	16,436
Stomiidae	<i>Tactostoma macropus</i>	LC377784.2	-	-	17,563
Stomiidae	<i>Photonectes margarita</i>	AP018417.1	-	-	18,592
Stomiidae	<i>Trigonolampa miriceps</i>	AP012961.1	-	-	15,709
Stomiidae	<i>Astronesthes lucifer</i>	AP012959.1	-	-	15,491
Stomiidae	<i>Stomias atriventer</i>	MG321595.1	-	-	17,596
Stomiidae	<i>Chauliodus sloani</i>	AP002915.1	-	-	17,814
Diplophidae	<i>Diplophos taenia</i>	AP012960.1	-	-	16,427
Diplophidae	<i>Diplophos sp.</i>	AB034825.1	-	-	16,418
Phosichthyidae	<i>Vinciguerria nimbaria</i>	AP006769.1	-	-	16,741
Phosichthyidae	<i>Vinciguerria nimbaria</i>	AP012958.1	-	-	16,741
Myctophum affine	<i>Myctophum affine</i>	AP002922.1	-	-	16,239
Trachipteridae	<i>Trachipterus trachipterus</i>	AP002925.1	-	-	16,162
Ateleopodidae	<i>Ateleopus japonicus</i>	AP002916.1	-	-	16,650
Synodontidae	<i>Saurida undosquamis</i>	AP002920.1	-	-	15,737
Salmonidae	<i>Coregonus lavaretus</i>	AB034824.1	-	-	16,737