Temperature and Pressure Shaped the Evolution of Antifreeze Proteins in Polar and Deep Sea Zoarcoid Fishes

Authors: Samuel N. Bogan[1], Nathan Surendran[1], Scott Hotaling[2], Thomas Desvignes[3], Iliana Bista[4,5,6], Luana S.F. Lins[7], Mari H. Eilertsen[8], Nathalie R. Le François[9], Tait Algayer[10], Scott L. Hamilton[11], Paul B. Frandsen[12], Federico G. Hoffmann[13,14], and Joanna L. Kelley[1]

Affiliations:

[1] Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA USA

[2] Department of Watershed Sciences, Utah State University, Logan, UT, USA

[3] Institute of Neuroscience, University of Oregon, Eugene, OR, USA

[4] LOEWE Centre for Translational Biodiversity Genomics, Frankfurt, 60325, Germany

[5] Senckenberg Research Institute and Natural History Museum, Frankfurt, 60325, Germany

[6] Wellcome Sanger Institute, Tree of Life, Wellcome Genome Campus, Hinxton CB10 1SA, United Kingdom

[7] CSIRO, Health and Biosecurity, Canberra, Australia

[8] Department of Biological Sciences, Centre for Deep Sea Research, University of Bergen, Norway

[9] Laboratoire de Physiologie et Aquaculture de la Conservation (LPAC), Division des collections vivantes, de la conservation et de la recherche, Biodôme de Montréal/ Espace pour la vie, 4777, av. Pierre-De Coubertin, Montréal, Québec H1V 1B3

[10] School of Biological Sciences, Washington State University, Pullman, WA, USA

[11] Moss Landing Marine Laboratories, Moss Landing, CA, USA

[12] Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, USA

[13] Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University, Mississippi State, MS, USA

[14] Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS, USA

Keywords: Long-read sequencing; antifreeze protein; adaptation; genomics; temperature; gene copy number

Running head: Long reads reveal antifreeze protein evolution

Abstract

Antifreeze proteins (AFPs) have enabled teleost fishes to repeatedly colonize polar seas. Four AFP types have convergently evolved in several fish lineages. AFPs inhibit ice crystal growth and lower cellular freezing point. In lineages with AFPs, species inhabiting colder environments may possess more AFP copies. Elucidating how differences in AFP copy number evolve is challenging due to the genes' tandem array structure and consequently poor resolution of these repetitive regions. Here we explore the evolution of type III AFPs (AFP III) in the globally distributed suborder Zoarcoidei, leveraging six new long-read genome assemblies. Zoarcoidei has fewer genomic resources relative to other polar fish clades while it is one of the few groups of fishes adapted to both the Arctic and Southern Oceans. Combining these new assemblies with additional long-read genomes available for Zoarcoidei, we conducted a comprehensive phylogenetic test of AFP III evolution and modeled the effects of thermal habitat and depth on AFP III gene family evolution. We confirm a single origin of AFP III via neofunctionalization of the enzyme sialic acid synthase B and show that AFP gains and losses were not only correlated with temperature, but also with depth, which reduces freezing point via pressure. Associations between the environment and AFP III copy number were driven by duplications of paralogs that were translocated out of the ancestral locus at which Zoarcoidei AFP arose. Our results reveal novel environmental effects on AFP evolution and demonstrate the value of high-quality genomic resources for studying how structural genomic variation shapes convergent adaptation.

1. Introduction

The evolution of gene family copy number variation can contribute to environmental adaptation (Schmidt et al. 2010) and be stimulated by environmental change (Hull et al. 2017). Studying the causes and consequences of copy number's evolution is difficult when duplications produce large tandem repeats or affect synteny, but the advent of long-read sequencing has helped resolve repetitive variants clouding the origins of copy number evolution (Marx 2023). The evolution of antifreeze proteins (AFPs) and antifreeze glycoproteins (AFGPs) are remarkable examples of convergent environmental adaptation in ectotherms living at low temperatures (A. L. DeVries 1971; Baskaran et al. 2021). Increased AFP 'dosage' driven by gene duplication can facilitate adaptation to chronic cold (Desjardins et al. 2012). In theory, temperate species derived from cold-adapted ancestors should undergo relaxed selection on AFP copy number and subsequent contractions (Hayes, Davies, and Fletcher 1991; Hew et al. 1988). Studying contractions and expansions of AFPs across thermally-distributed clades is critical for understanding thermal adaptation by ectotherms but has been hampered by poor resolution of their copy number due to low quality reference genome assemblies.

Antifreeze proteins were first discovered in Antarctic notothenioid fish in 1971, which possess antifreeze glycoproteins (AFGPs) and lower both the freezing and boiling point of cellular fluid (A. L. DeVries 1971; Raymond and DeVries 1977; A. L. DeVries and Wohlschlag 1969). Since this discovery, AFPs that induce thermal hysteresis, an increase in degrees of difference between a fluid's freezing and boiling points (Barrett 2001), have been discovered in microbes (Gilbert, Davies, and Laybourn-Parry 2005), plants (Gupta and Deswal 2014), invertebrates (Duman 2015), and vertebrates (Arthur L. DeVries and Cheng 2005). AFPs generally function by binding to the leading edges of newly-crystalized nuclei of ice, attenuating their expansion and

88 reducing cellular freezing point (Jia et al. 1996). In fishes, AFPs have independently evolved at
89 least three times (C. Cheng and Devries 2002). These convergently evolved AFPs make up four
90 known categories termed AFGPs, and AFP types I - III. The evolution of AFPs by ectotherms is a
91 classical example of environmental adaptation, but we still know little about the environmental
92 processes that shaped their evolution and the diversity and variation they exhibit across species
93 and their genomes.

94 It has been repeatedly suggested that, among lineages possessing AFPs, species or
95 populations inhabiting colder environments should exhibit more AFP copies (Hew et al. 1988;
96 Hayes, Davies, and Fletcher 1991; Kelley et al. 2010). Comparative studies of AFP copy number
97 in population- and species-pairs of fishes from nominally warmer versus colder habitats have
98 confirmed this expectation (Hayes, Davies, and Fletcher 1991; Desjardins et al. 2012). Studying
99 AFGPs in Arctic cod and Antarctic notothenioids, Baalsrud et al. (2018) observed losses of AFGPs
100 in secondarily sub-polar species of both groups. Long-read sequencing of Notothenioids has
101 revealed pseudogenization and contraction of AFGP copies in individual species inhabiting milder
102 latitudes suggestive of relaxed selection on antifreeze function (Bista et al. 2023; Rivera-Colón et
103 al. 2023). However, an analysis of AFP copy number across continuous thermal clines has not
104 been reported. Whether and how AFP copy number and gene dosage scales with freezing risk
105 remains unknown. One environmental variable that has been proposed to affect freezing risk and
106 AFP evolution is pressure. Because increased pressure in seawater affects the stability of ice
107 crystals and decreases phase change temperatures, lowering freezing point (Fujino, Lewis, and
108 Perkin 1974), it has been proposed that polar fishes inhabiting different depths and exhibiting
109 divergence in AFP copy number have experienced differential positive selection for AFP copy
110 number (Desjardins et al. 2012).

111 AFPs of bony fishes have evolved from neofunctionalization, and AFGPs by both
112 neofunctionalization and *de novo* gene birth. These events were driven in part by repetitive
113 genomic regions that provided substrate for *de novo* gene birth of an AFGP in cods (Baalsrud et
114 al. 2018; Zhuang et al. 2019) and translocations of ancestral genes that underwent
115 neofunctionalization and expanded into tandem arrays of AFPs (Deng et al. 2010; Hotaling et al.
116 2023). In addition to influencing AFP origins, structural variation such as translocations,
117 inversions, and repetitive element content have the potential to influence expansions and
118 contractions of AFPs across environmental gradients (Lauer and Gresham 2019). For example,
119 cytogenetically-balanced and unbalanced translocations can both facilitate variation in gene copy
120 number (Harewood et al. 2010; S. K. Watson et al. 2007). The accurate detection of translocations
121 and measurement of gene copy number in repetitive regions require long-read sequencing
122 approaches, which enable identification of structural variants such as translocations and better
123 resolve complex genomic regions relative to short read sequencing (Marx 2023). Long-read
124 sequencing has shown success in assembling repetitive AFGP arrays in notothenioid fishes (Bista
125 et al. 2023).

126 The bony fish suborder Zoarcoidei is an ideal system in which to study the joint effects of
127 environmental and structural genomic variation on AFP evolution. Zoarcoidei likely originated in
128 north temperate seas, followed by stepwise invasions of (i) the Arctic Ocean, (ii) south temperate
129 seas and (iii) the Southern Ocean (Hotaling et al. 2021) where minimum temperatures can fall
130 below -2.0 °C (Auger et al. 2021; Fetterer 2017). Sanger sequencing of BAC clones in the zoarcoid
131 *Lycodichthys dearborni* suggested that type III AFP arose within Zoarcoidei via
132 neofunctionalization of the enzyme sialic acid synthase B (*sasb*). *sasb* likely duplicated before
133 undergoing deletions of all exons except 1 and 6, which are homologous to AFP III exons 1 and 2

134 (Deng et al. 2010). Zoarcoidei repeatedly invaded and adapted to the poles (Hotaling et al. 2021)
135 while Zoarcoidei AFP III likely evolved once (Deng et al. 2010; Hobbs et al. 2020). AFP III copies
136 have likely expanded and contracted multiple times in Zoarcoidei (Hotaling et al. 2023), but the
137 extent of gene family evolution of AFP III in zoarcoids, of AFPs across fishes, and how genomic
138 and environmental variation interactively affected AFP evolution remain unclear. Zoarcoidei
139 diversified across depth faster than any group of fishes (Friedman and Muñoz 2023), potentiating
140 a complex landscape of selection for AFP function and dosage across the suborder's distributions
141 through latitude and depth. Previously identified tandem arrays of zoarcoid AFP III exhibited
142 conserved synteny with several flanking genes (Hotaling et al. 2023), enabling identification of
143 translocations out of this region.

144 Here we report high-quality, long-read genome assemblies from six species of Zoarcoidei
145 from diverse thermal habitats and depths (Fig. 1; Table 1). We integrated annotations of AFP III
146 genes with oceanographic measures of thermal minima, pressure, and depth within species'
147 habitats to test whether temperature and pressure are associated with AFP III copy number
148 evolution. Using analyses of synteny, we traced AFP III copies to ancestral and translocated arrays
149 and determined whether translocation drove AFP expansions under high freezing risk. These new
150 reference genome assemblies provide a valuable resource for studying how abiotic environments
151 and structural variation jointly influence convergent adaptation.

152

153 **2. Results and Discussion**

154 *2.1. Genome assembly* - Reference genomes were assembled with Pacific Biosciences
155 (PacBio) long-read sequence data for six zoarcoid fishes spanning a range of temperatures and
156 depths: the ronquil *Bathymaster signatus*, the wolffishes *Anarhichas lupus* and *Anarhichas minor*,

157 and the eelpouts *Lycenchelys platyrhina*, *Lycodes diapterus*, and *Ophthalmolycus amberensis*.
158 These species inhabit polar seas with latitudes as northerly as 73.0 °N and southerly as -65.8 °S.
159 Assembled genomes included temperate species with absolute latitudes as low as 37.8 ° (Fig. 1).
160 Species' maximum depths spanned 24 - 2,229 meters. These mark the first long-read genome
161 assemblies for the families Bathymasteridae (ronquils) and Anarhichadidae (wolffishes), and
162 increase the number of long-read genome assemblies among Zoarcidae (eelpouts) by 2.5x.

163 Genome sizes, sequencing coverage, N50 values, BUSCO complete percentages, and
164 percent repetitive elements are reported in Table 1 for the six new assemblies, five existing
165 zoarcoid long-read assemblies (Bista, Collins, and Wellcome Sanger Institute Tree of Life
166 Management, Samples and Laboratory team 2024), and one scorpaeniform outgroup. Genome
167 sizes of all Zoarcoidei examined equaled a mean of 647.80 ± 74.3 Mb. Genome assemblies
168 reported here averaged 11.7x greater coverage compared to publicly available, long-read zoarcoid
169 assemblies. Contig N50 values were an average of 6.5 Mb longer than available assemblies (Table
170 1). As expected, repetitive element content linearly and positively correlated with genome size
171 across all species (Kidwell 2002) as shown in Fig. S1 (r^2 = 0.93).

172 Zoarcoidei families are among the few fish groups inhabiting both the Arctic and Southern
173 Oceans (Møller, Nielsen, and Anderson 2005), positioning this suborder to be a valuable system
174 in which to study convergent evolution and adaptation to extreme cold. However, a paucity of
175 genomic resources for zoarcoid fishes stands as an obstacle to their value in research. The
176 improved reference genome quality of these new assemblies among other members of Zoarcoidei
177 and the extension of thermal habitats and depths represented among them significantly strengthen
178 Zoarcoidei as a tractable, valuable system in which to study convergent adaptation.
179

Species	Genome size (Mb)	Sequencing coverage	Contig N50 (Mb)	Scaffold N50 (Mb)	% Repetitive elements	BUSCO % complete (*n* orthologs)
Anarhichas lupus	627.7	80x	23.8	*na*	27.0%	98.5% (3587)
Anarhichas minor	643.9	74x	13.2	*na*	28.0%	98.5% (3587)
Bathymaster signatus	714.3	42x	3.1	*na*	32.2%	98.3% (3587)
Lycodes diapterus	711.9	43x	6.0	*na*	34.1%	98.5% (3585)
Lycenchelys platyrhina	718.5	34x	2.8	*na*	31.0%	98.3% (3579)
Ophthalmolycus amberensis	712.6	95x	5.8	*na*	30.6%	98.5% (3586)
Cebidichthys violaceus	575.7	31x	1.0	16.4	25.4%	93.3% (3396)
*Gasterosteus aculeatus**	471.9	75x	0.5	20.5	18.6%	97.0% (3534)
Leptoclinus maculatus	699.9	65x	2.4	23.5	32.8%	98.0% (3569)
Lycodes pacificus	646.4	68x	4.7	27.7	29.0%	98.1% (3573)
Melanostigma gelatinosum	662.0	39x	1.9	28.0	29.3%	98.1% (3569)
Pholis gunnellus	588.7	45x	3.5	25.4	23.3%	98.6% (3591)

Table 1 | *Genome assembly metrics.* Assemblies reported by the authors are shaded in gray. Publicly available zoarcoid genomes leveraged in this study are unhighlighted. A non-zoarcoid outgroup is designated by an asterisk. BUSCO analysis was performed on universal single-copy orthologs of Actinopterygii.

191 *2.2. Variation in type III antifreeze protein copy number* - The number of AFP III copies
192 in the 11 zoarcoids studied ranged from 0 to 42 putatively functional copies. Between 3 – 16
193 pseudogenized AFP III copies were detected across all zoarcoids (Fig. 1A). The copy number of
194 putatively-functional and pseudogenized AFP genes correlated non-linearly such that pseudogenes
195 increased in species with more AFP III toward a plateau (Fig. S2; 2nd order polynomial $r^2 = 0.31$).
196 This association was explained by phylogeny and became insignificant after controlling for species
197 relatedness (95% posterior interval = -1.54 – 0.67). The presence of pseudogenized AFP III copies
198 in all zoarcoids indicates that the gene's origin predates the most-recent common ancestor of
199 zoarcoids included in this study (Fig. 1A).
200 We evaluated whether AFP III may have arisen in a common ancestor predating Zoarcoidei
201 by screening for the gene in long-read assemblies of 12 species of order Scorpaeniformes
202 representing 7 families and found no evidence of complete or partial AFP III paralogs (Table S1).
203 Past phylogenetic analyses of zoarcoid AFP III also detected a partial, pseudogenized AFP III
204 sequence in Bathymasteridae, the earliest-branching family of Zoarcoidei (Hobbs et al. 2020).
205 However, Hobbs et al. (2020) and others have not ruled out an emergence of AFP III prior to the
206 branching of Zoarcoidei. Our results therefore strengthen evidence that Zoarcoidei AFP III likely
207 arose in an early branch of Bathymasteridae as opposed to a common ancestor within
208 Scorpaeniformes predating Zoarcoidei. Long-read sequencing of Zoarcoidei offers the potential to
209 trace the origin of AFP III within Zoarcoidei by mapping structural variants contributing to the
210 gene's evolution. These events include duplication-translocation of *sasb* and deletions of *sasb*
211 exons 2 – 5.
212
213

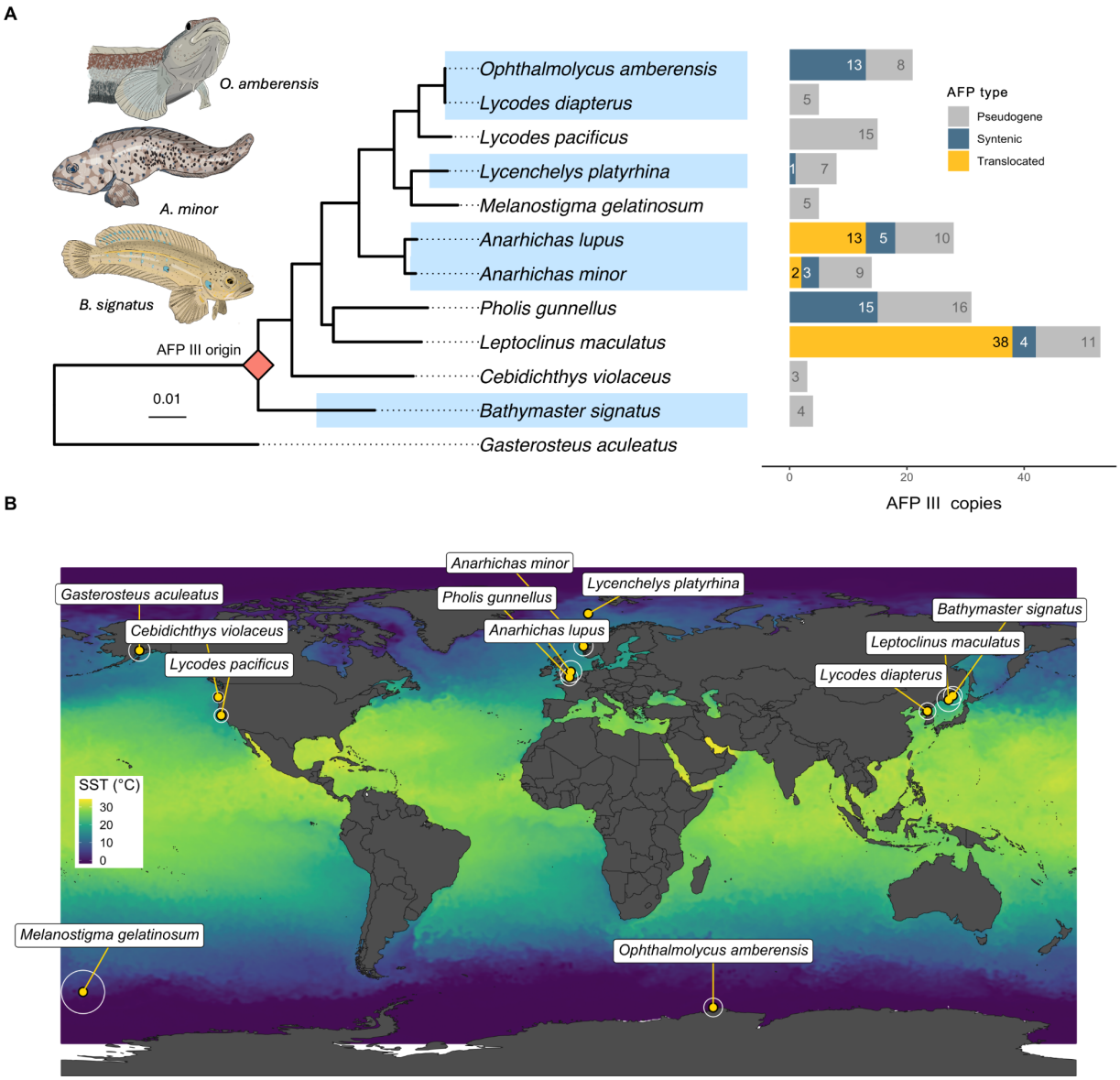


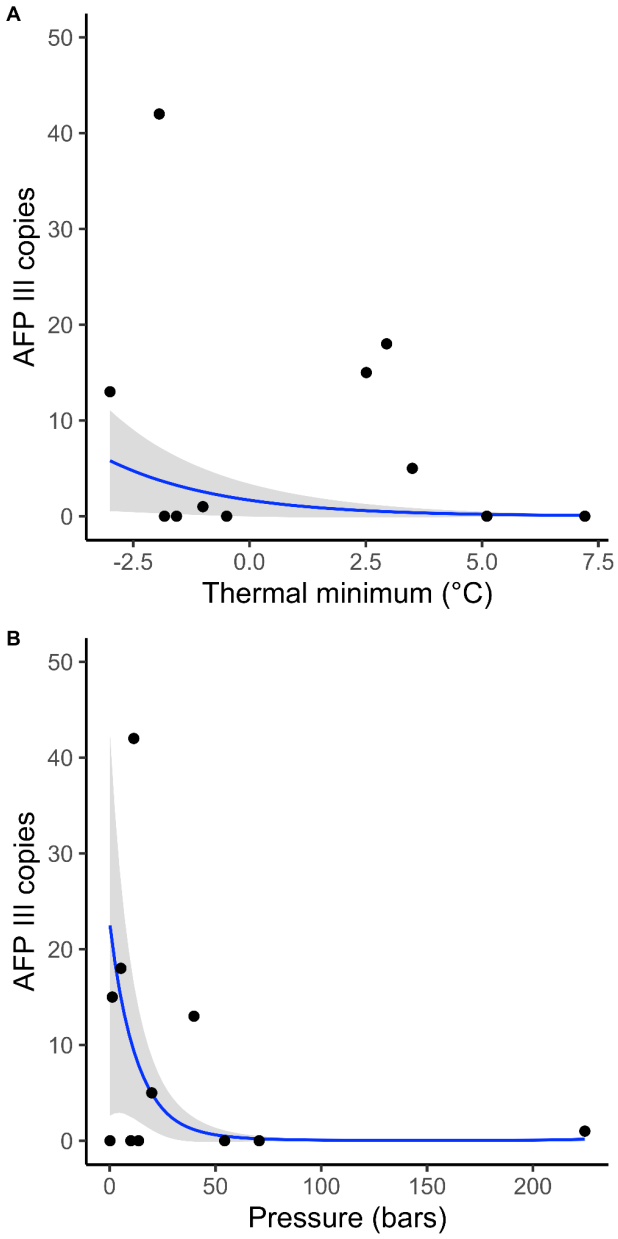
Figure 1 | *Species' phylogeny, AFP copy numbers, and distributions.* (A) A species tree of Zoarcoidei included in this study including outgroup *Gasterosteus aculeatus* is plotted adjacent to AFP copy number variation colored by pseudogenization or location of homologs in the ancestral syntenic region or translocation(s). The red diamond signifies the origin of type III AFP. Species highlighted in blue denote genome assemblies reported by the authors. The scale bar denotes substitutions per amino acid. *Ophthalmolycus amberensis* and *Lycodes diapterus* branch tip lengths equal 0.000069 and 0.000074 substitutions per amino acid, respectively. (B) Species' distributions are plotted as the point of maximum likelihood of occurrence from AquaMaps standardized distributions. The size of white circles around occurrences are scaled according to species abundance. Mean sea surface temperature (SST, °C) recorded in August, 2020, is visualized according to color. Illustrations by SNB.

227 *2.3. Environmental effects on type III antifreeze protein copy number* - Habitat temperature
228 and pressure jointly influence freezing risk (Fujino, Lewis, and Perkin 1974) and were additively
229 correlated with AFP copy number. A phylogenetically corrected Bayesian model revealed that
230 AFP copy number significantly increased with decreases in (i) thermal minima of species' habitats
231 and (ii) mean pressure of species' depth ranges (Fig. 2). Thermal minima equaled the lowest
232 recorded temperature at a species' coordinate of maximum likelihood of occurrence within their
233 depth range. Pressure was derived from habitat depth and latitude. Thermal minima of species
234 habitats spanned -3.00 – 7.21 °C (Fig. S2). AFP copy number significantly increased in species
235 occupying lower thermal minima (Fig. 2A). This effect was significant after controlling for
236 phylogeny (scaled effect = -1.55 ± 0.97).

237 AFP III gene copy number declined in species inhabiting greater pressure (Fig. 2B).
238 Pressure had a stronger effect on AFP copy number than temperature and was non-linear,
239 occupying a negative exponential curve. Pressure affected AFP copy number after controlling for
240 habitat depth as well as phylogeny (scaled linear effect = -4.72 ± 2.58). As expected, habitat
241 temperatures correlated with depth and pressure (Fig. S4 – S5). To account for this autocorrelation,
242 predicted effects of thermal minima and pressure on AFP copy number were modeled in a
243 statistically independent manner.

244 The finding that AFP III copy number is reduced in deep sea species experiencing high
245 pressure is pertinent to the biogeography of Zoarcoidei. Zoarcoidei includes the eelpout family
246 Zoarcidae, which has exhibited faster rates of evolution across depth than any group of fishes
247 (Friedman and Muñoz 2023). Associations between depth and AFP copy number were previously
248 proposed by Desjardins et al., who observed that the wolffishes *A. lupus* and *A. minor* exhibit
249 higher AFP copy number in the shallow water species *A. lupus* rather than the species occupying

250 a colder thermal environment (Desjardins et al. 2012). Our results mark the first quantitative
251 evidence that selection on antifreeze proteins across thermal environments was shaped by depth
252 and pressure.



253

254 **Figure 2 |** *Associations between temperature, pressure, and type III antifreeze protein copy number.*
255 Relationships between AFP III copy number and (A) thermal minima of species habitats and (B) mean
256 pressure of species' habitats are plotted. Fitted lines represent fit from a phylogenetically controlled model
257 of type III antifreeze protein copy number.
258

A direct explanation of pressure's association with AFP copy number is that increased energy imposed by pressure lowers the freezing point of intracellular fluids (Fujino, Lewis, and Perkin 1974). An indirect explanation also exists that is not mutually exclusive to this direct hypothesis. Deep sea fishes generally regulate the hydrostatic pressure of cells through osmoregulation such that increased pressure is countered by greater intracellular osmolality (Yancey, Blake, and Conley 2002; Gillett et al. 1997; Yancey et al. 2014). Because increased osmolality also reduces the freezing point of fluids (Bodnar 1993; Potter, Clynne, and Brown 1978), it is possible that pressure and osmolality additively affect cellular freezing point.

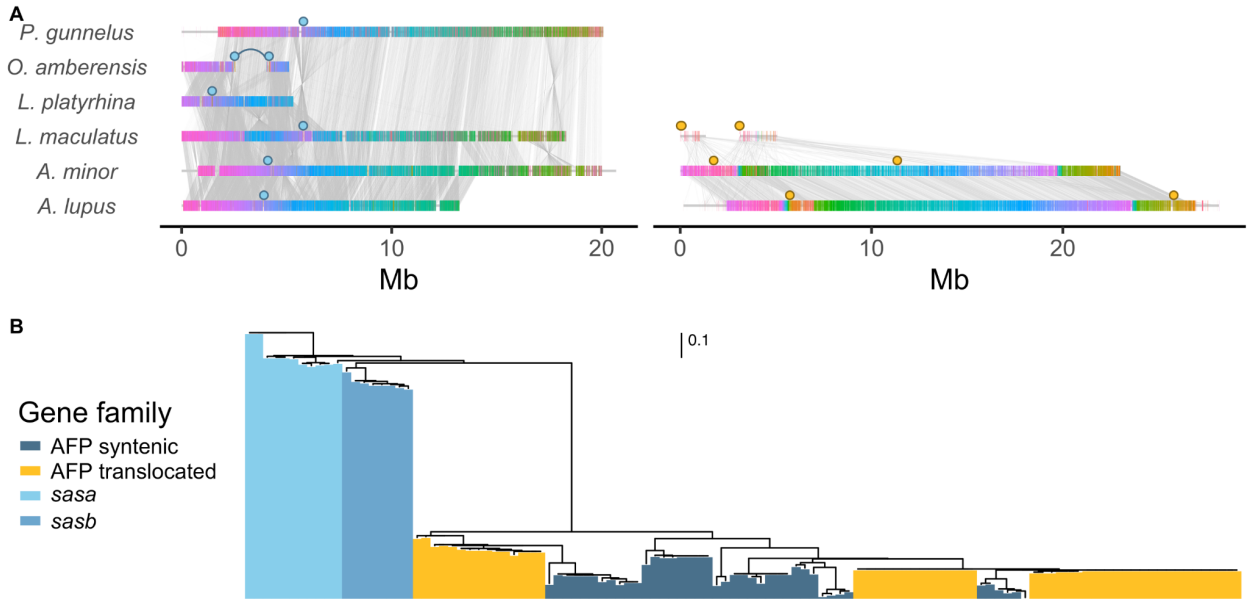


Figure 3 | *Synteny and origin of type III antifreeze proteins.* (A) Synteny of AFP III tandem arrays. Full lengths of scaffolds and contigs are shown. Colored bars represent regions conserved in at least two species. Regions sharing the same color are homologous. Gray lines connect these homologous regions. Narrow gray bars depict regions lacking homology in another species. Positions of ancestral, syntenic AFP arrays are marked by blue circles and translocated arrays by orange circles. Curved lines connect single AFP arrays split between contigs. All contigs containing syntenic AFP are homologous to *P. gunnelus* Chr 21. *L. maculatus* contigs containing translocated AFP fall within different regions of Chr 12. *A. minor* and *A. lupus* contigs with translocated AFP are homologous to Chr 2. (B) A rooted gene tree of *sasa*, *sasb* and AFP III paralogs is shown. Branches are colored by whether they denote syntenic versus translocated AFPs, *sasa*, or *sasb* genes. The scale bar represents nucleotide substitutions per base pair. The node most closely related to *sasb* branches into gene glades of translocated and syntenic AFPs, prohibiting resolution of whether ancestral Zoarcoidei AFP evolved in the syntenically conserved region. Tests of whether Zoaroidei AFP arose in the syntenic region are described under Supplemental Material.

282 *2.4. Gene family evolution and convergence of antifreeze protein copy number* - Restricted
283 maximum likelihood ancestral state reconstruction of AFP III copy number demonstrated that at
284 least 8 expansions and 12 contractions of AFP III occurred across the studied phylogeny (Fig. S6).
285 To evaluate whether AFP contractions and expansions convergently evolved during independent
286 invasions of polar or deep seas, the phylogenetically-controlled model of environmental
287 associations with AFP III copy number was expanded as a structural equation model predicting
288 both habitat thermal minima and mean pressure as a function of phylogenetic relatedness. This
289 enabled regression of AFP III copy number against phylogenetically-independent invasions of
290 colder or higher-pressure environments, which is necessary for quantifying convergence.
291 Independent invasions of lower thermal minima were significantly associated with increased AFP
292 III copy number (95% posterior interval = -1.15 – -0.03 AFP copies per °C). Despite having a
293 stronger association with pressure before testing for convergence, independent invasions of higher
294 pressure habitats were not associated with decreased AFP III copy number.

295 While the association between AFP III copy number and habitat pressure among extant
296 branches was stronger than with thermal minimum, there is little evidence that this association
297 resulted from convergent evolution between independent invasions of the deep ocean. By contrast,
298 weaker associations between temperature and AFP III are more likely to have evolved as a result
299 of convergent gains and losses of AFP III during invasions of Zoarcoidei into colder and warmer
300 environments, respectively. The pole-to-pole distribution of Zoarcoidei is marked by a likely
301 origin in the Arctic followed by invasion of the deep ocean, trans-polar migration across deep-
302 ocean regions, and subsequent reinvasion of shallow waters in the Southern Ocean (Hobbs et al.
303 2020; Hotaling et al. 2023). The ancestral state reconstruction of thermal minima shows repeated
304 invasions of sub-zero oceans across Zoarcoidei (Fig. S7). As shown in Figure S9, the

305 reconstruction of habitat pressure shows one primary lineage of deep ocean fishes belonging to the
306 eelpout family Zoarcidae. Zoarcidae is the deepest dwelling family of zoarcoid fishes. Thus, the
307 strong effect of pressure on AFP III copies across Zoarcoidei was more likely attributed to a
308 singular invasion of the deep sea by Zoarcidae rather than multiple convergent events. Decreases
309 in habitat pressure in Zoarcidae that corresponded with increased AFP III copy number represent
310 instances of parallel rather than convergent evolution.

311 Using maximum likelihood gene tree inference, we confirmed the origin of AFP III
312 evolution in Zoarcoidei. Our tree of all *sasa*, *sasb*, and AFP III homologs across study species is
313 consistent with a single origin of AFP III from *sasb (Deng et al. 2010)* as shown in Figure 3B.
314 Maximum likelihood ancestral state reconstructions were used to predict binary characters of
315 whether species possessed AFP III in the syntenic locus or in one of the six translocated loci. These
316 reconstructions demonstrated that ancestral Zoarcoidei AFP III most likely arose in the syntenic
317 region rather than one of the translocated positions (relative likelihood = 727.89 – 1440.39).

318 *2.5. Role of translocation in AFP III expansions* - The ancestral AFP III array falls within
319 a conserved, syntenic region flanked by genes including synuclein gamma (*sncgb*) and LIM-
320 binding domain protein (*ldb3b*) as previously shown by Deng et al. 2010. We identified at least
321 six independent translocations of AFP III out of the ancestral syntenic region (Fig. 3A). All
322 zoarcoids with AFP III retained this narrow scale synteny such that at least some AFP III copies
323 were flanked upstream by *sncgb* and downstream by *ldb3b*. Whole-contig alignment revealed
324 conservation of synteny across larger genomic scales. The ancestral AFP III tandem repeat region
325 flanked by *sncgb* and *ldb3b* fell in the same homologous region of chromosome 21 across
326 Zoarcoidei (Fig. 3A). In *L. maculatus*, however, the ancestral AFP III array was positioned in a ~3

327 Mb chromosomal inversion (Fig. 3A). Translocated AFPs clustered into three gene clades (Fig.
328 3B).

329 AFP III copies outside of the ancestral array mapped to two translocated regions in each of
330 three species, totalling six translocated arrays across all study species. The wolffishes *A. lupus* and
331 *A. minor* each contained two translocations on contigs homologous to *P. gunnelus* chromosome 2.
332 However, the two translocated arrays in *A. lupus* and the two in *A. minor* possessed differing
333 synteny, falling into four distinctly different regions of chromosome 2 (Fig. 3A). Thus, these arrays
334 likely represent four discrete translocation events. The two translocated arrays in *L. maculatus*
335 were distributed across two contigs mapping to different regions of *P. gunnelus* chromosome 12
336 (Fig. 3A). Translocated copies contributed significantly to AFP III expansions across Zoarcoidei,
337 making up 13 of 18 AFP copies in *A. lupus*, 2 of 5 copies in *A. minor*, and 38 of 42 copies in *L.*
338 *maculatus* (Fig. 1A). Considering the short lengths of contigs in which translocated AFP III were
339 located *L. maculatus* (Fig. 3A), there is potential for misassembly to have inflated the total number
340 of translocations, the AFP copy number within translocated arrays, and the total AFP copy number
341 in this species. Conversely, translocated arrays that border the ends of these contigs may also cause
342 underreporting of total AFP copy number at these locations (Fig. 3A).

343 A multivariate phylogenetic model predicting syntenic and translocated AFP III copy
344 number as a function of thermal minima and pressure revealed that environmental gradients in
345 AFP copy number were driven by expansions of translocated homologs. After controlling for
346 phylogeny, habitat thermal minima was associated with syntenic AFP copies at a scaled effect
347 strength of -0.62 ± 0.72 and translocated AFP copies at -1.36 ± 0.70 (Fig. 4A). Mean habitat
348 pressure yielded an effect on syntenic copies of -0.88 ± 0.54 and an effect on translocated copies
349 of -9.84 ± 3.08 (Fig. 4B).

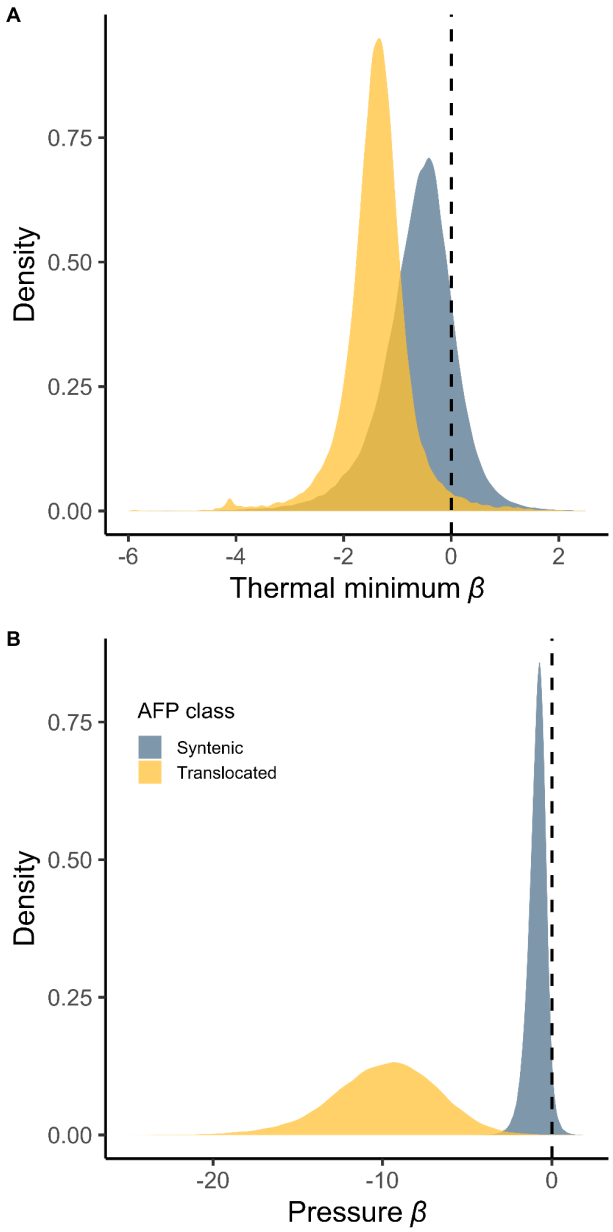


Figure 4 | *Effects of thermal minimum and pressure on copy number of syntenicically-conserved and translocated type III antifreeze protein genes*. Densities represent the posterior probabilities for fixed effects (β) of (A) thermal minimum and (B) mean pressure on antifreeze protein copy in ancestral, syntenically-conserved versus translocated sites.

Molecular repair mechanisms causing variation in copy number such as break-induced replication can directly facilitate translocation of gene copies to regions sharing microhomology with break sites (Hastings et al. 2009; Hastings, Ira, and Lupski 2009; Kramara, Osia, and Malkova

360 2018). It is possible that expansions of AFP III via translocation are simply a consequence of
361 selection for high copy number. AFP III translocations were not more likely among species with
362 a large number of AFPs, however. Rather, copy number in translocated arrays correlated with
363 habitat temperature and pressure. This is suggestive of directional selection acting more strongly
364 on AFP copy number outside of the syntenic, ancestral AFP III locus. Differences in directional
365 selection on copies may arise as a consequence of subfunctionalization stemming from
366 transcriptional changes that occur following translocation into new regulatory landscapes (Y.
367 Wang, Wang, and Paterson 2012; Harewood et al. 2010). Alternatively, copy number at the
368 ancestral AFP array may be under stronger stabilizing selection and evolutionary constraint
369 relative to translocated arrays.

370 Studies observing environmental effects on gene copy number are abundant (Kondrashov
371 2012; Lauer and Gresham 2019; Mérot et al. 2020), but the role of translocation in mediating
372 environmental gradients of copy number is less understood. In populations of salmon, a
373 chromosomal translocation shows strong genotype-environment association with temperature (K.
374 B. Watson et al. 2022) while in soybean, a reciprocal chromosomal translocation may have
375 facilitated climatic adaptation (W. Wang et al. 2021). Furthermore, translocation duplications of
376 AFGP have been observed in notothenioid icefishes (Rivera-Colón et al. 2023), indicating
377 potential for translocation to affect environmental gradients of antifreeze gene copy number in
378 other systems. Our results expand on these studies by demonstrating that not only translocation,
379 but its association with copy number, can contribute to molecular evolution across environments.

380 Population genomic studies of AFP III structural variation within Zoarcoidei species are
381 necessary for further understanding environmental gradients in AFP copy number and the potential
382 role of translocation in shaping these gradients. Intraspecific AFP copy number variation has been

observed in populations of winter flounder inhabiting different depths and extents of sea ice cover (Hayes, Davies, and Fletcher 1991). Zoarcoids included in our study inhabit ranges that, in some cases, span both non-freezing and freezing conditions. Neutral and adaptive variation shaping AFP III copy number across the Zoarcoidei suborder can be more finely resolved by resequencing of conspecifics and populations distributed across gradients of temperature, pressure, and freezing risk. Detection and modeling of adaptive and neutral variation in AFP copy number is necessary for evaluating (i) additional evidence of whether copy number-environment associations arose via drift and (ii) whether non-neutral copy number variation is greater among translocated AFP III arrays.

3. Conclusion

Structural variation observable via long-read sequencing appears to have been critical for AFP III gene family evolution during the convergent adaptation of Zoarcoidei to both poles. Our findings highlight the role of multiple environmental stressors (temperature and pressure) in shaping the genetic architecture of adaptations like antifreeze function. Furthermore, structural variants such as translocations may act as gatekeepers of adaptation by constraining or facilitating molecular evolution under the selective pressures of those stressors. Future research should address whether translocation-associated gene family expansions are driven by mechanisms associated with transposable element architecture at translocation endpoints or break-induced replication. The reported, long-read Zoarcoidei reference genomes offer further potential to study the role of structural variation and chromosomal translocation in convergent environmental adaptation, as well as the complex evolution of antifreeze proteins.

406 **4. Materials and Methods**

407 *4.1. DNA extraction, sequencing, and genome assembly* - *Lycenchelys platyrhina* was
408 collected from the Norwegian Sea, Ægir vent field, Norway (72.3286 °N, 1.5178 °E) on July 28th,
409 2021. The specimen for *Lycenchelys platyrhina* (originally described as *Lycodes platyrhinus*
410 Jensen, 1902) received a putative identification during collection which required genomic
411 identification and statistical consideration. *L. platyrhina* is a rarely observed species associated
412 with only 11 occurrences in the Global Biodiversity Information Facility. The sequenced specimen
413 was collected inside the range of *L. platyrhina* observations (67.1 - 73.8 °N, 10.3 °W - 8.2 °E). To
414 this end, mitochondrial scaffolds of the *L. platyrhina* assembly were aligned to the NCBI
415 nucleotide database using blastn (Altschul et al. 1990). No sequence of *Lycenchelys platyrhina*
416 was available in the database. The top 10 greatest percent similarity blast hits were with species of
417 the family Lycodinae to which *L. platyrhina* belongs; top 5 = *O. amberensis* (95.88%) *Lycenchelys*
418 *kolthoffi* (95.57%), *Lycenchelys raridens* (95.05%), *Lycenchelys muraena* (94.60%), and
419 *Lycenchelys sarsii* (94.60%). Skeletal muscle from a single individual was used for DNA isolation
420 using the Circulomics Nanobind, BIG DNA prep, Tissue protocol. Isolated genomic DNA was
421 quantified with a Fragment Analyzer (Agilent). Ampure-PB beads were used to purify and remove
422 short fragments from the sample prior to library prep according to manufacturer protocol. A second
423 size selection for >10 kb fragments was performed on the sample using a BluePippin system (Sage
424 Science Inc., Beverly, MA, USA) with a 10kb cutoff. A library was prepared without shearing
425 using SMRTbell® ExpressTemplate Prep Kit 3.0. Final library was size selected using BluePippin
426 with an 8 kb cutoff. The library was sequenced on one 8M SMRT cells on Sequel IIe instrument
427 using Sequel II Binding kit 2.2 and Sequencing Chemistry v2.0. Loading was performed by
428 adaptive loading, movie time: 30 hours, pre-extension 2 hours.

429 Sampling of *Anarhichas lupus* was performed in July, 2022, on a captive-reared individual
430 derived from eggs collected in Conception Bay, NewFoundland, Canada (47.697222 °N, -
431 52.921944 °N) via diving. Sampling of *Anarhichas minor* was performed in July, 2022, on a
432 captive-reared individual bred from broodstock collected in St-Lawrence Gulf, Banc Beaugé,
433 Canada (49.750000 °N, -60.166667 °W). Tissue was ground using dry ice. Cells were gently added
434 to pre-warmed 0.8% agarose and mixture was placed into the wells of a pre-chilled plug mold.
435 Following solidification, plugs were dislodged and treated with proteinase K and EDTA at 50°C.
436 HMW DNA was extracted by dissolving the plugs at 70°C to remove the agarose followed by a
437 phenol-chloroform extraction. Purification was performed on 15.0 µg of gDNA with AMPure PB
438 beads and a bead-to-template ratio of 1. The DNA libraries were prepared following the Pacific
439 Biosciences Preparing HiFi SMRTbell® Libraries using the SMRTbell Express Template Prep Kit
440 2.0 protocol. 10.0 µg of high molecular weight purified genomic DNA (final volume of 120 µl)
441 was sheared using the Diagenode Megaruptor 3 instrument (Diagenode Inc., Denville, NJ, USA)
442 and a two-cycle shear method at speed 33 followed by speed 34. The DNA Damage repair, End
443 repair and SMRT bell ligation steps were performed as described in the template preparation
444 protocol with the SMRTbell Template Prep Kit 2.0 reagents (Pacific Biosciences, Menlo
445 Park, CA, USA). The DNA libraries were size selected on a BluePippin system using a 10 kb
446 cutoff. The libraries underwent an AMPure PB bead cleanup following the SMRTlink v10
447 calculator procedure before being sequenced on a PacBio Sequel II instrument using adaptive
448 loading, Sequel II Sequencing Kit 2.0, SMRT Cell 8M and 30 hours movies with a pre-extension
449 time of 2h. Sequence data was processed using the SMRT Link Analysis
450 v. 10.1.0.119588 pipeline.

451 *Bathymaster signatus* was collected in Alaska, USA (52.1861 °N, -173.596 °W), by the
452 Alaska Trawl Survey on July 28th, 2012, via trawl net. DNA was extracted from gill filament via
453 the Circulomics Nanobind Tissue Kit. HiFi Library prep was done with SMRTbell prep kit 3.0
454 (version 102-166-600 Apr 2022). As part of the library prep, the genomic DNA was sheared to
455 25kb on a Megaruptor 2.0. The library was size selected on the Blue Pippin according to PacBio's
456 recommended settings (10kb-50kb collection). The sample was sequenced on one SMRT cell on
457 the Sequel II with v5 primer, v2.2 polymerase, 1 hr binding time, a 30 hr movie, 2 hr pre-extension,
458 and adaptive loading.

459 *Lycodes diapterus* was collected in Monterey Bay, California, USA (36.8009 °N, -
460 121.903867 °W), on October, 20th, 2021. High molecular weight DNA was extracted from
461 dissected muscle tissue using a Qiagen Genomic-tip DNA extraction kit. DNA was then sheared
462 to ~20 kbp using a Diagenode Megaruptor 3 and size selected for DNA for fragments greater than
463 10 kb using a BluePippin system. A PacBio HiFi library was prepared with the SMRTbell Express
464 Template Prep Kit 2.0. The library was then sequenced on a PacBio Sequel II instrument on an
465 8M SMRT cell in CCS mode with a 30-hour movie time. DNA extraction, library preparation, and
466 long-read sequencing of *Ophthalmolycus amberensis* are described in detail by Hotaling et al.
467 (2023).

468 Haplotype-phased genomes of all species except for *O. amberensis* were assembled from
469 long-read libraries using hifiasm v0.19.5-r592 (H. Cheng et al. 2021) set to default parameters.
470 Assemblies of primary contigs are reported here. Read processing and genome assembly of *O.*
471 *amberensis* was achieved using Canu (Koren et al. 2017) v1.8 and is described in detail in
472 Hotaling, et al. (2023). Assembly completeness was evaluated using BUSCO v5.4.7 (Simão et al.
473 2015) input with universal, single-copy Actinopterygii orthologs.

474 *4.2. AFP III and sasa/b annotation* - Homologs of AFP III were annotated in each genome
475 assembly included in the study using Exonerate v2.78.4 input with a query sequence of the
476 translated *Macrozoarces americanus* AFP III (J03924.1) set to the model parameter
477 'protein2genome' (Slater and Birney 2005). *sasa*, *sasb*, and genes flanking the 5' and 3' ends of
478 the ancestral syntenic region of Zoarcoidei AFP III (*sncgb* and *ldb3b*) were annotated using
479 Exonerate (Slater and Birney 2005) set to the same parameters input with protein sequences of *sas*
480 paralogs from *Lycodichtys dearborni* and *Gasterosteus aculeatus* protein sequences of *sncgb*
481 (XP_040031911.1) and *ldb3b* (XP_040031907.1). Exonerate BED file annotations of AFP III
482 were filtered for whether hits contained start and stop codons and complete sequences of exons 1
483 and 2. Translocated AFP III copies outside of the ancestral syntenic AFP III region were manually
484 reviewed for evidence of misassembly by aligning translocated AFP III regions to (i) known,
485 characterized AFP translocation in *A. lupus* (Desjardins et al. 2012) and (ii) additional regions of
486 translocation within single species if they possessed multiple translocation events to identify
487 artificial duplications. This ancestral region is defined as the locus of conserved synteny at which
488 AFP tandem arrays are flanked by the genes *sncgb* and *lbd3b*. Evidence of artificial AFP III
489 translocations driven by misassembly was not apparent. Pseudogenized copies of AFP III were
490 identified as regions mapping to AFP III query sequences that were complete but lacked a start
491 codon, lacked one of the two AFP III exons, exhibited truncations of the majority of an exon, or
492 conserved 5'/3' untranslated regions that lacked a proximal AFP coding sequence. Publicly-
493 available long-read genomes used in this study included *Cebidichthys violaceus*
494 (GCA_023349555.1), *Gasterosteus aculeatus* (GCA_016920845.1), *Leptoclinus maculatus*
495 (GCA_032191485.1), *Lycodes pacificus* (GCA_028022725.1), *Melanostigma gelatinosum*
496 (GCA_949748355.1), *Pholis gunnellus* (GCA_910591455.2).

497 *4.3. Environmental data* - Depth-informed temperatures and salinities were mined from the
498 Copernicus Global Ocean Reanalysis (GLORYS) database (Jean-Michel et al. 2021) within ±1°
499 latitude/longitude of species' coordinate of maximum likelihood of occurrence and the expected
500 depth range of each species spanning a 10 year period between June 1st 2011 - June 1st 2021.
501 Coordinates of maximum likelihood of occurrence were extracted from AquaMaps standardized
502 range predictions (Kaschner, Rius-Barile, and Kesner-Reyes 2013), which are probabilistic
503 predictions of abundance and occurrence derived from Global Biodiversity Information Facility
504 occurrence data ("GBIF.org," n.d.). AquaMaps ranges were constrained within coordinates of
505 known species occurrences. Species minimum and maximum habitat depths were extracted from
506 FishBase (Froese 1996). Thermal minima were calculated as the minimum temperature
507 experienced by each species within their representative latitude-longitude-depth range. Pressure
508 was calculated as a function of sea water depth and latitude using conversions defined by the
509 UNESCO Joint Panel on Oceanographic Tables and Standards (Fofonoff 1983) and expressed as
510 a mean.

511 *2.4. Species tree construction and phylogenetic modeling* - A maximum likelihood species
512 tree was constructed from MUSCLE (Edgar 2004) alignments of concatenated protein sequences
513 of universal 3,603 Actinopterygii single copy orthologs using IQ-Tree2 v1.6.12 parameterized
514 with a WAG substitution model and fixed rate heterogeneity (Minh et al. 2020; Nguyen et al.
515 2015). A Bayesian phylogenetically-corrected model was then fitted to predict AFP III copy
516 number variation as a function of species' habitat temperature and pressure. The model assumed
517 negative binomial variance for the AFP III count data and uninformative, uniform priors.
518 Phylogenetic relatedness was fitted as a random effect of the genetic distance matrix derived from
519 the species tree. The thermal minima and mean pressure of species' habitats were included as fixed

520 effects, with pressure fitted as a second order polynomial function. An iteration of this model was
521 fitted that predicted AFP III copy number as a function of both pressure and depth, revealing that
522 pressure exhibited a significant and equivalent effect after controlling for depth. Model selection
523 performed using Bayesian marginal likelihoods (Berger and Pericchi 1996) revealed that pressure
524 was a better predictor of AFP III copy number than depth.

525 A multivariate iteration of the phylogenetic model was also fitted to predict two separate
526 outcome variables for copy numbers of (i) AFP III within its ancestral, syntenic region and (ii)
527 AFP III genes translocated out of the syntenic region. This multivariate model did not permit
528 residual correlations. A third iteration of the phylogenetic model fit structural equations predicting
529 variation in AFP III copy number as a function of phylogenetically independent variation in habitat
530 thermal minimum and pressure. Phylogenetically independent variance in abiotic environments
531 were estimated by predicting environmental variables as a function of phylogeny, and AFP III
532 copy number as a function of environment and phylogeny. Thus, any effect of thermal minimum
533 or pressure on AFP III copy number is quantified after conditioning on phylogenetic variation in
534 abiotic environment.

535 Posterior sampling across all models was performed using four MCMC chains and 40,000
536 iterations including 10,000 warmups. All models were fitted using brms v2.20.4 (Bürkner 2017),
537 an R interpreter of the Stan Bayesian programming language (Carpenter et al. 2017). The
538 significance of fixed effects was assessed using a probability of direction test such that effects
539 described as 'significant' exhibited posterior probabilities with 95% confidence intervals above or
540 below 0 (Makowski et al. 2019). Because of potential misidentification of *L. platyrhina*, we
541 evaluated the effect of the identification on model predictions. The leverage (the effect of a single
542 observation on model predictions) that *L. platyrhina* imposed on parameter estimation in

543 phylogenetic models of AFP gene family evolution across environments was quantified using a
544 leave-one-out analysis. Removal of *L. platyrhina* from the univariate model imposed a 6.45%
545 reduction in the effect of thermal minima on AFP III copy number and 4.24% reduction in the
546 linear effect of pressure.

547 *4.5. Analyses of AFP III gene family evolution and synteny* - FASTA sequences of AFP III
548 exons 1 and 2 and *sasa*/*b* exons 1 and 6 were extracted from annotation bed files using bed2fasta
549 from bedtools v2.31.1 (Quinlan and Hall 2010). The two exons were concatenated per gene and
550 aligned using MAFFT v7.520 (Katoh et al. 2002). A maximum likelihood gene tree was
551 constructed from the alignment using IQ-TREE2 v1.6.12 set up fixed rate heterogeneity and
552 substitution model optimization, which selected a HKY+F+G4 model (Minh et al. 2020; Nguyen
553 et al. 2015). Restricted maximum likelihood ancestral state reconstructions of AFP III copy
554 number, habitat thermal minimum, and mean habitat pressure were performed using the 'ace'
555 function of the R package ape v5.7-1 (Paradis, Claude, and Strimmer 2004; Paradis and Schliep
556 2019) in order to model contractions and expansions of AFP III across the studied phylogeny.
557 Ancestral state reconstructions of AFP III positions at the ancestral Zoarcoidei node were
558 performed using ace(). Individual, discrete reconstructions were performed on 7 binary characters
559 associated with whether species possessed putatively functional copies of AFP III at the
560 syntenically-conserved locus or the six translocation loci. The relative likelihoods of the
561 Zoarcoidei common ancestor possessing AFP III at these regions was measured as a ratio of
562 maximum likelihoods for AFP III presence at the syntentically-conserved region divided by
563 maximum likelihood of presence at each of the six translocation loci.

564 A multiple alignment of scaffolds or contigs containing translocated AFP III copies was
565 performed using progressiveMAUVE v2.4.0 (Darling, Mau, and Perna 2010), which identified

conservation and synteny of genomic regions conserved in at least two species of Zoarcoidei that exhibited translocated AFPs. This multiple alignment and the positions of translocated AFP arrays relative to conserved, syntenic regions are plotted in Fig. 3A in order to visualize conserved synteny of the ancestral AFP III locus, as well as the locations and number of AFP III translocations.

Acknowledgements

We acknowledge funding from National Science Foundation awards OPP-1543383 and OPP-1947040 (supporting TD), and OPP-2312253 (to JLK) as well as funding from the Antarctic Science Bursary. We acknowledge computing resources from the Center for Institutional Research Computing at Washington State University, the National Science Foundation Extreme Science and Engineering Discovery Environment (XSEDE), and University of California, Santa Cruz Research Computing. We thank the veterinary clinic staff of the Biodôme de Montréal, including Karine Béland, DMV. We also thank the captain and crew of the ARSV *Laurence M. Gould*, the personnel of the U.S. Antarctic Program Support Contractors for assistance in Chile, at sea and at Palmer Station, as well as the logistics in Denver, CO, for their support related to the Antarctic fieldwork required for this study. We acknowledge funding from the Trond Mohn Science Foundation through the Centre for Deep Sea Research (grant number TMS2020TMT13) and the Norwegian Biodiversity Information Centre (the Taxonomy Initiative, project nr. 3-20-70184243) and thank the scientists and students from University of Bergen (Norway), the crew of the RV GO Sars and the pilots of the ROV ÆGIR6000 for their assistance with sampling in the Arctic. Lastly, we would like to thank the Alaskan Trawl Survey, and the Ocean Genome Legacy Center, Northeastern University, for assisting with specimen collections and archiving.

589 **Author Contributions**

590 Designed research (JLK, SNB)

591 Performed research (FH, LSFL, IB, JLK, MHE, NRLF, NS, PBF, SH, SNB, TA, TD)

592 Analyzed data (JLK, NS, SNB)

593 Wrote the paper (JLK, NS, SNB)

594 Edited the paper (FH, LSFL, IB, JLK, MHE, NRLF, NS, PBF, SH, SNB, TA, TD)

595

596 **Data Availability**

597 Genome assemblies, raw sequencing reads, and associated files are available on NCBI under the
598 following Bioprojects: PRJNA980960 (*Anarhichas lupus*), PRJNA982125 (*Anarhichas minor*),
599 PRJNA982427 (*Bathymaster signatus*), PRJNA985917 (*Lycenchelys platyrhina*), PRJNA982746
600 (*Lycodes diapterus*) and PRJNA701078 (*Omphalmolycus amberensis*). All code used for
601 bioinformatic processing and statistical analysis are available under the GitHub repository
602 https://github.com/snbogan/Long_AFP [available upon acceptance and publication] in addition to
603 intermediate data files and specimen metadata. An archived copy of this repository is available on
604 Zenodo under the DOI [to be generated upon acceptance].

605

606

607

608

609

610

611

612

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Auger, Matthis, Rosemary Morrow, Elodie Kestenare, Jean-Baptiste Sallée, and Rebecca Cowley. 2021. "Southern Ocean in-Situ Temperature Trends over 25 Years Emerge from Interannual Variability." *Nature Communications* 12 (1): 514.

Baalsrud, Helle Tessand, Ole Kristian Tørresen, Monica Hongrø Solbakken, Walter Salzburger, Reinhold Hanel, Kjetill S. Jakobsen, and Sissel Jentoft. 2018. "De Novo Gene Evolution of Antifreeze Glycoproteins in Codfishes Revealed by Whole Genome Sequence Data." *Molecular Biology and Evolution* 35 (3): 593–606.

Barrett, J. 2001. "Thermal Hysteresis Proteins." *The International Journal of Biochemistry & Cell Biology* 33 (2): 105–17.

Baskaran, Abirami, Manigundan Kaari, Gopikrishnan Venugopal, Radhakrishnan Manikkam, Jerrine Joseph, and Parli V. Bhaskar. 2021. "Anti Freeze Proteins (Afp): Properties, Sources and Applications - A Review." *International Journal of Biological Macromolecules* 189 (October):292–305.

Berger, James O., and Luis R. Pericchi. 1996. "The Intrinsic Bayes Factor for Model Selection and Prediction." *Journal of the American Statistical Association* 91 (433): 109–22.

Bista, Iliana, Martin Collins, and Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team. 2024. "The Genome Sequence of the Limp Eelpout, Melanostigma Gelatinosum Günther, 1881." *Wellcome Open Research* 9:192.

Bista, Iliana, Jonathan M. D. Wood, Thomas Desvignes, Shane A. McCarthy, Michael Matschiner, Zemin Ning, Alan Tracey, et al. 2023. "Genomics of Cold Adaptations in the Antarctic Notothenioid Fish Radiation." *Nature Communications* 14 (1): 3412.

Bodnar, R. J. 1993. "Revised Equation and Table for Determining the Freezing Point Depression of H2O-Nacl Solutions." *Geochimica et Cosmochimica Acta* 57 (3): 683–84.

Bürkner, Paul-Christian. 2017. "Brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (August):1–28.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (January). https://doi.org/10.18637/jss.v076.i01.

Cheng, C., and A. Devries. 2002. "Origins and Evolution of Fish Antifreeze Proteins." *Fish Antifreeze Proteins*, October, 83–107.

Cheng, Haoyu, Gregory T. Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. 2021. "Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm." *Nature Methods* 18 (2): 170–75.

Darling, Aaron E., Bob Mau, and Nicole T. Perna. 2010. "progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement." *PloS One* 5 (6): e11147.

Deng, Cheng, C-H Christina Cheng, Hua Ye, Ximiao He, and Liangbiao Chen. 2010. "Evolution of an Antifreeze Protein by Neofunctionalization under Escape from Adaptive Conflict." *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21593–98.

Desjardins, Marième, Laurie A. Graham, Peter L. Davies, and Garth L. Fletcher. 2012. "Antifreeze Protein Gene Amplification Facilitated Niche Exploitation and Speciation in Wolffish." *The FEBS Journal* 279 (12): 2215–30.

DeVries, A. L. 1971. “Glycoproteins as Biological Antifreeze Agents in Antarctic Fishes.” *Science* 172 (3988): 1152–55.

DeVries, A. L., and D. E. Wohlschlag. 1969. “Freezing Resistance in Some Antarctic Fishes.” *Science* 163 (3871): 1073–75.

DeVries, Arthur L., and C. -H Christina Cheng. 2005. “Antifreeze Proteins and Organismal Freezing Avoidance in Polar Fishes.” In *Fish Physiology*, 22:155–201. Academic Press.

Duman, John G. 2015. “Animal Ice-Binding (antifreeze) Proteins and Glycolipids: An Overview with Emphasis on Physiological Function.” *The Journal of Experimental Biology* 218 (Pt 12): 1846–55.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97.

Fetterer, F. ;. Knowles K. ;. Meier W. ;. Savoie M. ;. Windnagel A. 2017. “Sea Ice Index, Version 3.” NSIDC. https://doi.org/10.7265/N5K072F8.

Fofonoff, N. P. 1983. “Algorithms for the Computation of Fundamental Properties of Seawater.” https://repository.oceanbestpractices.org/handle/11329/109.

Friedman, Sarah T., and Martha M. Muñoz. 2023. “A Latitudinal Gradient of Deep-Sea Invasions for Marine Fishes.” *Nature Communications* 14 (1): 773.

Froese, R. 1996. “FishBase.” *Oceanographic Literature Review* 3 (43): 321.

Fujino, K., E. L. Lewis, and R. G. Perkin. 1974. “The Freezing Point of Seawater at Pressures up to 100 Bars.” *Journal of Geophysical Research* 79 (12): 1792–97.

“GBIF.org.” n.d. Global Biodiversity Information Facility. Accessed 2024. https://www.gbif.org.

Gilbert, Jack A., Peter L. Davies, and Johanna Laybourn-Parry. 2005. “A Hyperactive, Ca2+-Dependent Antifreeze Protein in an Antarctic Bacterium.” *FEMS Microbiology Letters* 245 (1): 67–72.

Gillett, Matthew B., Jolyn R. Suko, Freddy O. Santoso, and Paul H. Yancey. 1997. “Elevated Levels of Trimethylamine Oxide in Muscles of Deep-Sea Gadiform Teleosts: A High-Pressure Adaptation?” *The Journal of Experimental Zoology* 279 (4): 386–91.

Gupta, Ravi, and Renu Deswal. 2014. “Antifreeze Proteins Enable Plants to Survive in Freezing Conditions.” *Journal of Biosciences* 39 (5): 931–44.

Harewood, Louise, Frédéric Schütz, Shelagh Boyle, Paul Perry, Mauro Delorenzi, Wendy A. Bickmore, and Alexandre Reymond. 2010. “The Effect of Translocation-Induced Nuclear Reorganization on Gene Expression.” *Genome Research* 20 (5): 554–64.

Hastings, P. J., Grzegorz Ira, and James R. Lupski. 2009. “A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation.” *PLoS Genetics* 5 (1): e1000327.

Hastings, P. J., James R. Lupski, Susan M. Rosenberg, and Grzegorz Ira. 2009. “Mechanisms of Change in Gene Copy Number.” *Nature Reviews. Genetics* 10 (8): 551–64.

Hayes, Pliny H., Peter L. Davies, and Garth L. Fletcher. 1991. “Population Differences in Antifreeze Protein Gene Copy Number and Arrangement in Winter Flounder.” *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada* 34 (1): 174–77.

Hew, C. L., N. C. Wang, S. Joshi, G. L. Fletcher, G. K. Scott, P. H. Hayes, B. Buettner, and P. L. Davies. 1988. “Multiple Genes Provide the Basis for Antifreeze Protein Diversity and Dosage in the Ocean Pout, Macrozoarces Americanus.” *The Journal of Biological Chemistry* 263 (24): 12049–55.

Hobbs, Rod S., Jennifer R. Hall, Laurie A. Graham, Peter L. Davies, and Garth L. Fletcher.

2020. “Antifreeze Protein Dispersion in Eelpouts and Related Fishes Reveals Migration and Climate Alteration within the Last 20 Ma.” *PloS One* 15 (12): e0243273.

Hotaling, Scott, Marek L. Borowiec, Luana S. F. Lins, Thomas Desvignes, and Joanna L. Kelley. 2021. “The Biogeographic History of Eelpouts and Related Fishes: Linking Phylogeny, Environmental Change, and Patterns of Dispersal in a Globally Distributed Fish Group.” *Molecular Phylogenetics and Evolution* 162 (September):107211.

Hotaling, Scott, Thomas Desvignes, John S. Sproul, Luana S. F. Lins, and Joanna L. Kelley. 2023. “Pathways to Polar Adaptation in Fishes Revealed by Long-read Sequencing.” *Molecular Ecology* 32:1381–97.

Hull, Ryan M., Cristina Cruz, Carmen V. Jack, and Jonathan Houseley. 2017. “Environmental Change Drives Accelerated Adaptation through Stimulated Copy Number Variation.” *PLoS Biology* 15 (6): e2001333.

Jean-Michel, Lellouche, Greiner Eric, Bourdallé-Badie Romain, Garric Gilles, Melet Angélique, Drévillon Marie, Bricaud Clément, et al. 2021. “The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis.” *Frontiers of Earth Science in China* 9. https://doi.org/10.3389/feart.2021.698876.

Jia, Z., C. I. DeLuca, H. Chao, and P. L. Davies. 1996. “Structural Basis for the Binding of a Globular Antifreeze Protein to Ice.” *Nature* 384 (6606): 285–88.

Kaschner, K., J. Rius-Barile, and K. Kesner-Reyes. 2013. “About AquaMaps.” aquamaps.org. 2013. https://www.aquamaps.org/AboutAquaMaps.htm.

Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66.

Kelley, Joanna L., Jan E. Aagaard, Michael J. MacCoss, and Willie J. Swanson. 2010. “Functional Diversification and Evolution of Antifreeze Proteins in the Antarctic Fish Lycodichthys Dearborni.” *Journal of Molecular Evolution* 71 (2): 111–18.

Kidwell, Margaret G. 2002. “Transposable Elements and the Evolution of Genome Size in Eukaryotes.” *Genetica* 115 (1): 49–63.

Kondrashov, Fyodor A. 2012. “Gene Duplication as a Mechanism of Genomic Adaptation to a Changing Environment.” *Proceedings. Biological Sciences / The Royal Society* 279 (1749): 5048–57.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.

Kramara, J., B. Osia, and A. Malkova. 2018. “Break-Induced Replication: The Where, The Why, and The How.” *Trends in Genetics: TIG* 34 (7): 518–31.

Lauer, Stephanie, and David Gresham. 2019. “An Evolving View of Copy Number Variants.” *Current Genetics* 65 (6): 1287–95.

Makowski, Dominique, Mattan S. Ben-Shachar, S. H. Annabel Chen, and Daniel Lüdecke. 2019. “Indices of Effect Existence and Significance in the Bayesian Framework.” *Frontiers in Psychology* 10 (December):2767.

Marx, Vivien. 2023. “Method of the Year: Long-Read Sequencing.” *Nature Methods* 20 (1): 6–11.

Mérot, Claire, Rebekah A. Oomen, Anna Tigano, and Maren Wellenreuther. 2020. “A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation.” *Trends in Ecology & Evolution* 35 (7): 561–72.

Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37 (5): 1530–34.

Møller, Peter Rask, Jørgen G. Nielsen, and M. Eric Anderson. 2005. "Systematics of Polar Fishes." In *Fish Physiology*, 22:25–78. Academic Press.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. "APE: Analyses of Phylogenetics and Evolution in R Language." *Bioinformatics* 20 (2): 289–90.

Paradis, Emmanuel, and Klaus Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35 (3): 526–28.

Potter, R. W., M. A. Clynne, and D. L. Brown. 1978. "Freezing Point Depression of Aqueous Sodium Chloride Solutions." *Economic Geology and the Bulletin of the Society of Economic Geologists* 73 (2): 284–85.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.

Raymond, J. A., and A. L. DeVries. 1977. "Adsorption Inhibition as a Mechanism of Freezing Resistance in Polar Fishes." *Proceedings of the National Academy of Sciences of the United States of America* 74 (6): 2589–93.

Rivera-Colón, Angel G., Niraj Rayamajhi, Bushra Fazal Minhas, Giovanni Madrigal, Kevin T. Bilyk, Veronica Yoon, Mathias Hüne, Susan Gregory, C. H. Christina Cheng, and Julian M. Catchen. 2023. "Genomics of Secondarily Temperate Adaptation in the Only Non-Antarctic Icefish." *Molecular Biology and Evolution* 40 (3): msad029.

Schmidt, Joshua M., Robert T. Good, Belinda Appleton, Jayne Sherrard, Greta C. Raymant, Michael R. Bogwitz, Jon Martin, et al. 2010. "Copy Number Variation and Transposable Elements Feature in Recent, Ongoing Adaptation at the Cyp6g1 Locus." *PLoS Genetics* 6 (6): e1000998.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12.

Slater, Guy St C., and Ewan Birney. 2005. "Automated Generation of Heuristics for Biological Sequence Comparison." *BMC Bioinformatics* 6 (February):31.

Wang, Weidong, Liyang Chen, Xutong Wang, Jingbo Duan, Rachel D. Flynn, Ying Wang, Chancelor B. Clark, et al. 2021. "A Transposon-Mediated Reciprocal Translocation Promotes Environmental Adaptation but Compromises Domesticability of Wild Soybeans." *The New Phytologist* 232 (4): 1765–77.

Wang, Yupeng, Xiyin Wang, and Andrew H. Paterson. 2012. "Genome and Gene Duplications and Gene Expression Divergence: A View from Plants." *Annals of the New York Academy of Sciences* 1256 (May):1–14.

Watson, K. Beth, Sarah J. Lehnert, Paul Bentzen, Tony Kess, Antony Einfeldt, Steven Duffy, Ben Perriman, Sigbjørn Lien, Matthew Kent, and Ian R. Bradbury. 2022. "Environmentally Associated Chromosomal Structural Variation Influences Fine-Scale Population Structure of Atlantic Salmon (Salmo Salar)." *Molecular Ecology* 31 (4): 1057–75.

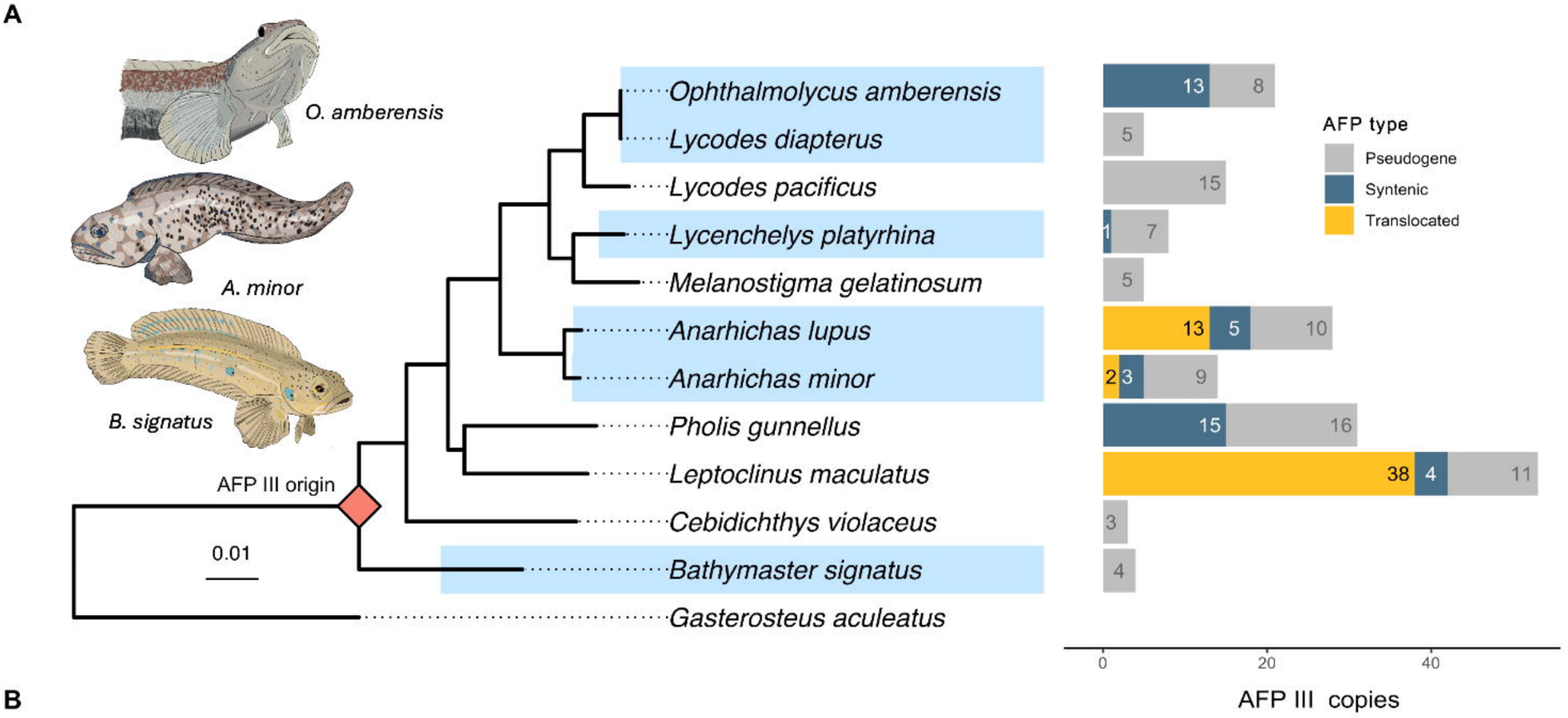
Watson, Spencer K., Ronald J. deLeeuw, Doug E. Horsman, Jeremy A. Squire, and Wan L. Lam.

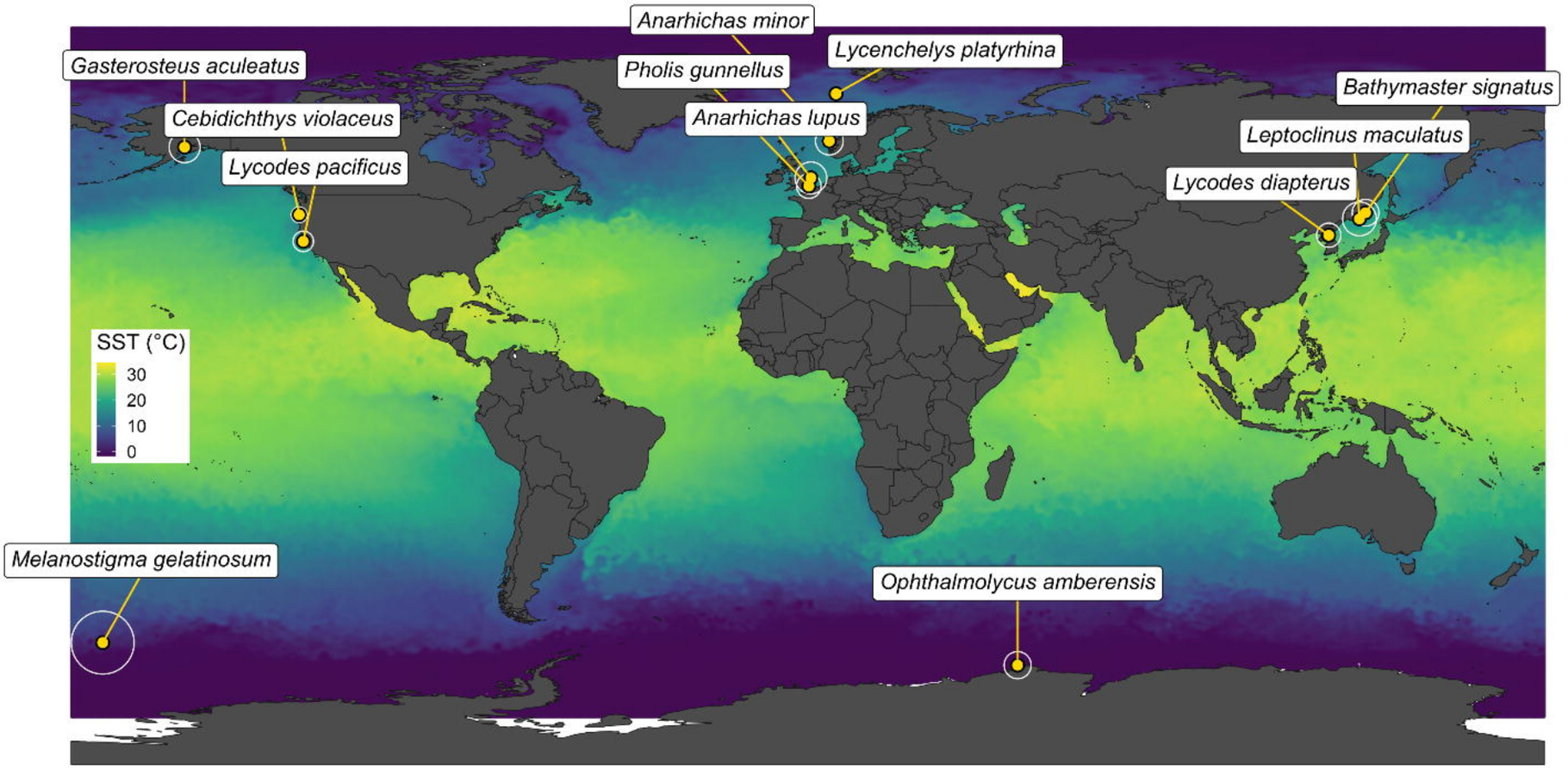
2007. “Cytogenetically Balanced Translocations Are Associated with Focal Copy Number Alterations.” *Human Genetics* 120 (6): 795–805.

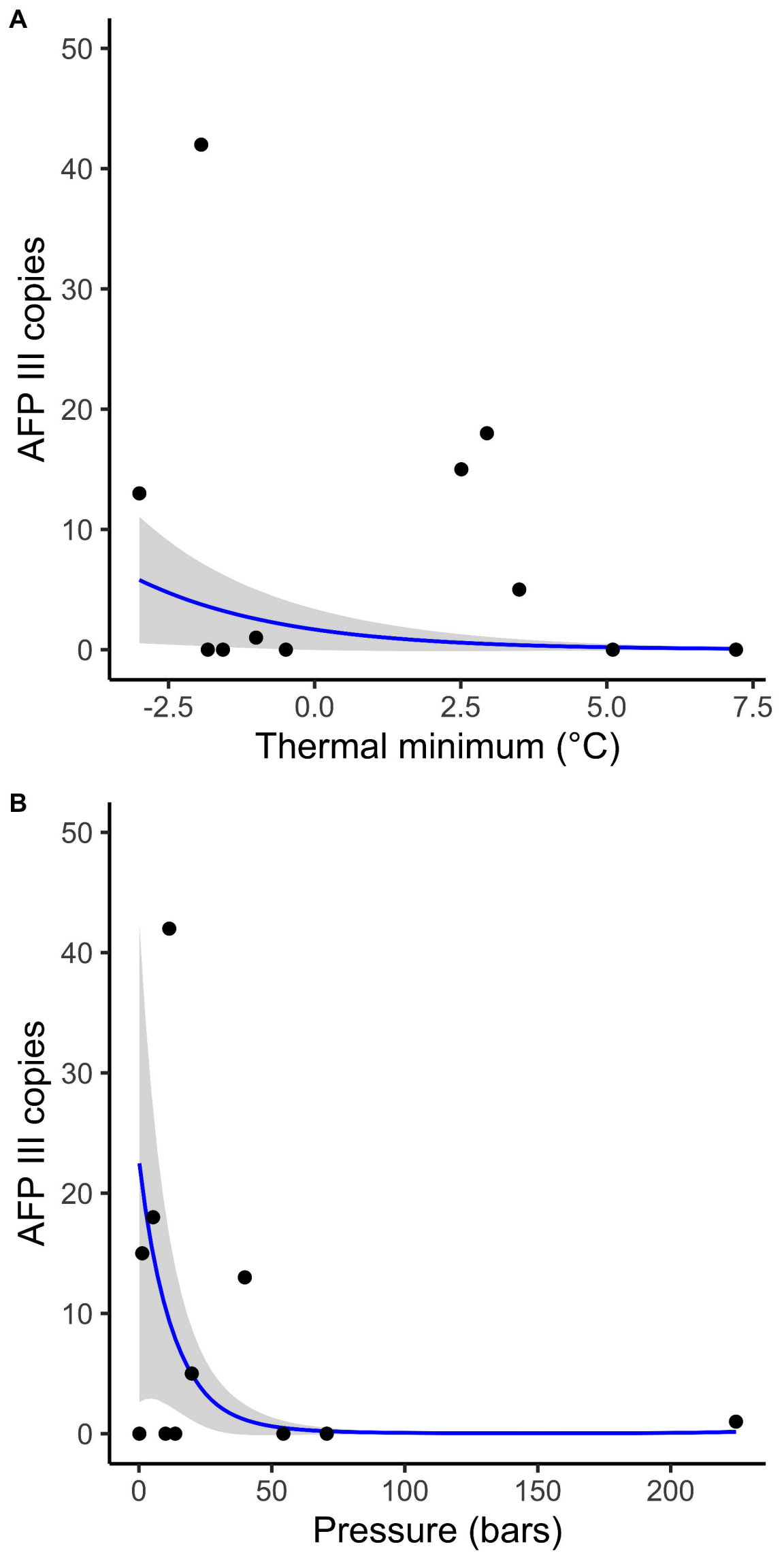
Yancey, Paul H., Wendy R. Blake, and James Conley. 2002. “Unusual Organic Osmolytes in Deep-Sea Animals: Adaptations to Hydrostatic Pressure and Other Perturbants.” *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 133 (3): 667–76.

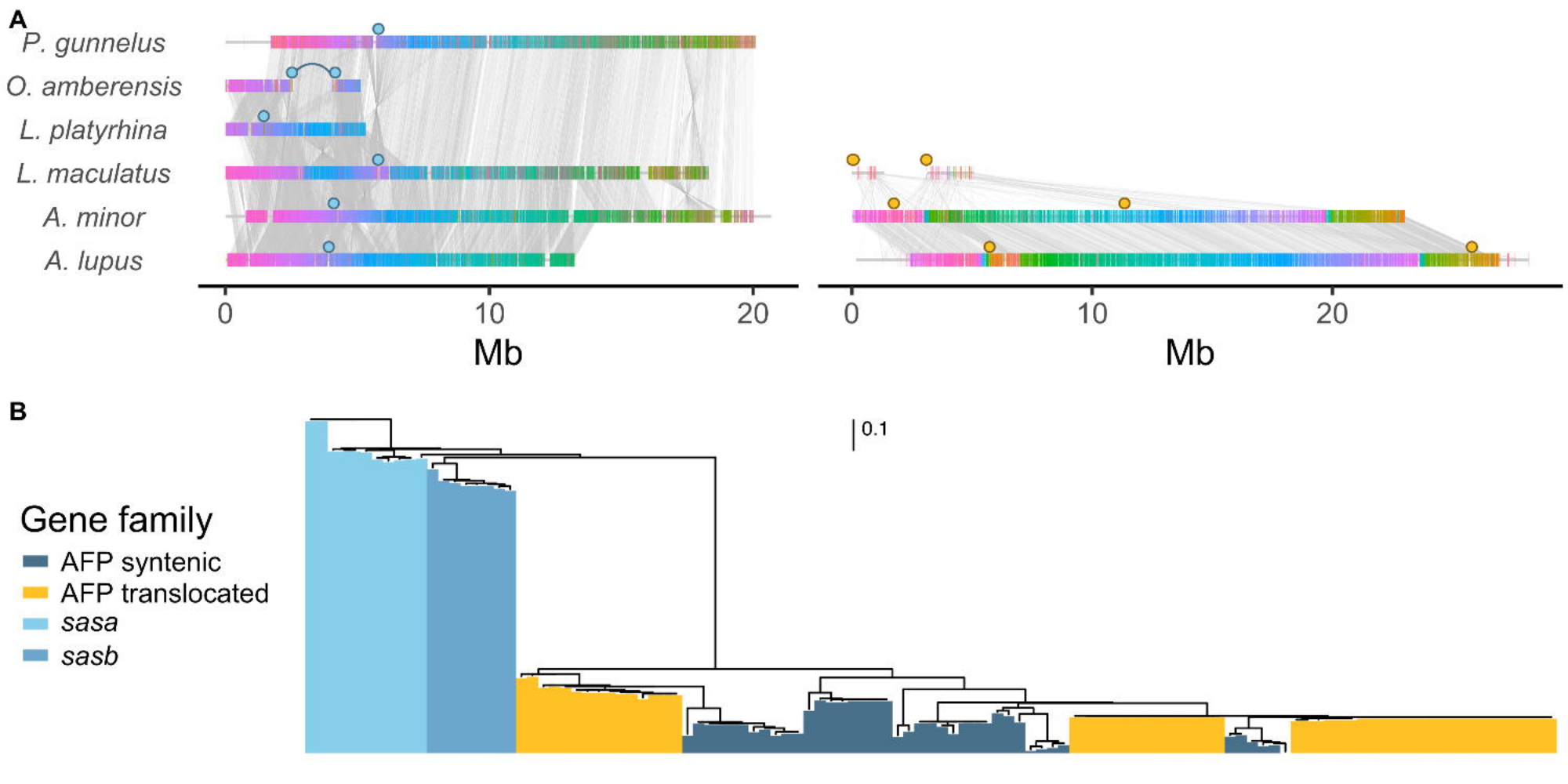
Yancey, Paul H., Mackenzie E. Gerringer, Jeffrey C. Drazen, Ashley A. Rowden, and Alan Jamieson. 2014. “Marine Fish May Be Biochemically Constrained from Inhabiting the Deepest Ocean Depths.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4461–65.

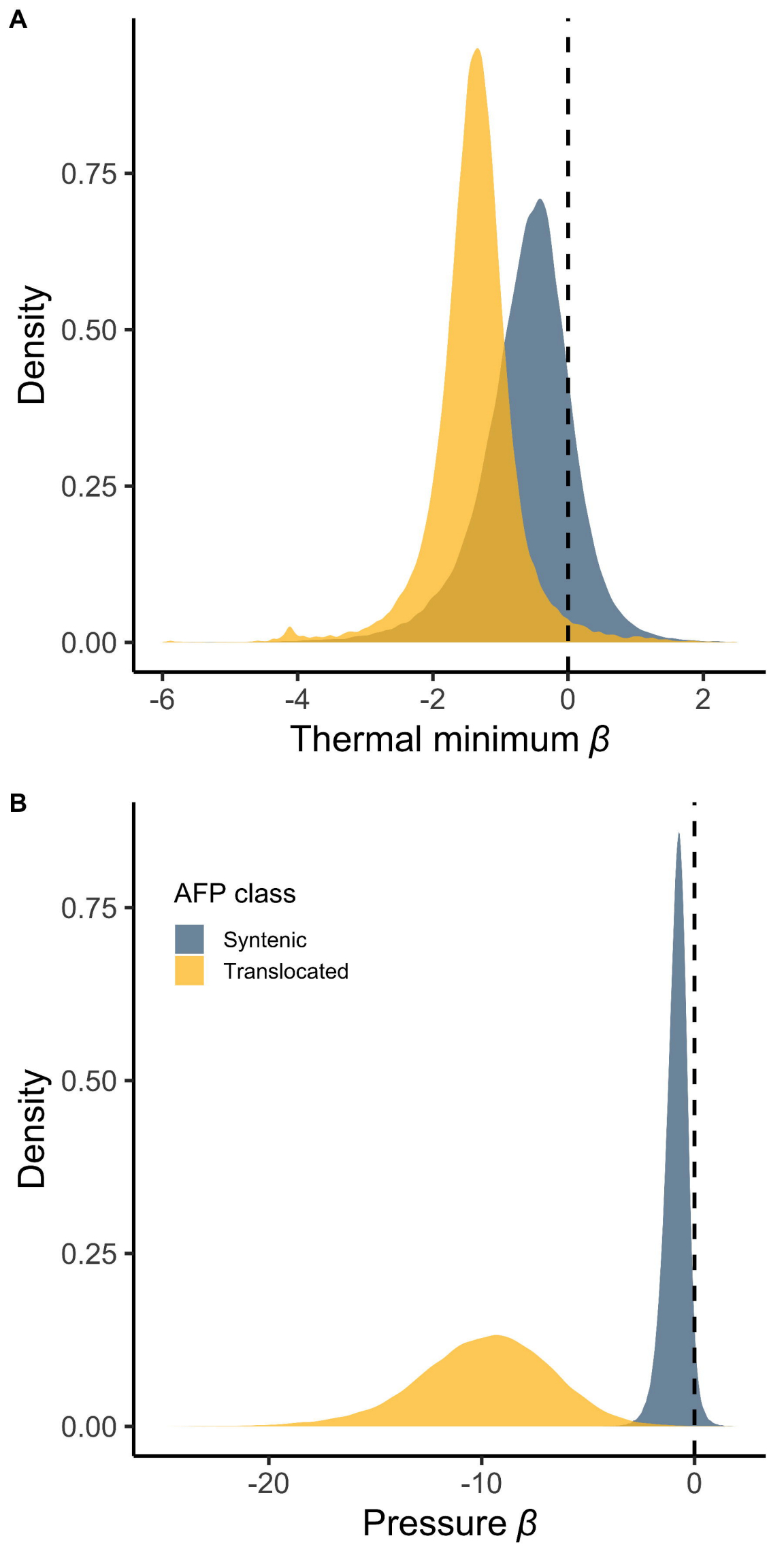
Zhuang, Xuan, Chun Yang, Katherine R. Murphy, and C-H Christina Cheng. 2019. “Molecular Mechanism and History of Non-Sense to Sense Evolution of Antifreeze Glycoprotein Gene in Northern Gadids.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (10): 4400–4405.


A
O. amberensis
A. minor
B. signatus
AFP III origin
0.01
Ophthalmolycus amberensis
Lycodes diapterus
Lycodes pacificus
Lycenchelys platyrhina
Melanostigma gelatinosum
Anarhichas lupus
Anarhichas minor
Pholis gunnellus
Leptoclinus maculatus
Cebidichthys violaceus
Bathymaster signatus
Gasterosteus aculeatus
13
8
5
15
1
7
5
13
5
10
2
3
9
15
16
38
4
11
3
4
AFP type
Pseudogene
Syntenic
Translocated
0
20
40
AFP III copies


B
Anarhichas minor
Lycenchelys platyrhina
Gasterosteus aculeatus
Pholis gunnellus
Bathymaster signatus
Cebidichthys violaceus
Anarhichas lupus
Leptoclinus maculatus
Lycodes pacificus
Lycodes diapterus
SST (°C)
30
20
10
0
Melanostigma gelatinosum
Ophthalmolycus amberensis


A
AFP III copies
50
40
30
20
10
0
-2.5
0.0
2.5
5.0
7.5
Thermal minimum (°C)
B
AFP III copies
50
40
30
20
10
0
0
50
100
150
200
Pressure (bars)


A
P. gunnelus
O. amberensis
L. platyrhina
L. maculatus
A. minor
A. lupus
0
10
20
Mb
0
10
20
Mb
B
0.1
Gene family
AFP syntenic
AFP translocated
sasa
sasb


A
Density
0.75
0.50
0.25
0.00
-6
-4
-2
0
2
Thermal minimum β
B
AFP class
Syntenic
Translocated
Density
0.75
0.50
0.25
0.00
-20
-10
0
Pressure β