Convergent Gene Duplication in Arctic and Antarctic Teleost Fishes

Authors: Nishiguchi, Tomoya, and Ishikawa, Asano

Source: Zoological Science, 42(1) : 116-126

Published By: Zoological Society of Japan

URL: https://doi.org/10.2108/zs240098

ZOOLOGICAL SCIENCE **42**: 116–126 (2025)

Convergent Gene Duplication in Arctic and Antarctic Teleost Fishes

Tomoya Nishiguchi* and Asano Ishikawa*

Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8562, Japan

Teleost fishes have independently colonized polar regions multiple times, facing many physiological and biochemical challenges due to frigid temperatures. Although increased gene copy numbers can contribute to adaptive evolution in extreme environments, it remains unclear which categories of genes exhibit increased copy numbers associated with polar colonization. Using 104 species of ray-finned fishes, we systematically identified genes with a significant correlation between copy number and polar colonization after phylogenetic correction. Several genes encoding extracellular glycoproteins, including zona pellucida (ZP) proteins, which increase their copy number in Antarctic notothenioid fishes, exhibited elevated copy numbers across multiple polar fish lineages. Additionally, some genes reported to be highly expressed under cold stress, such as cold-inducible RNA-binding protein (CIRBP), had significantly increased copy numbers in polar fishes. Further analysis will provide a fundamental basis for understanding the role of gene duplication in polar adaptations.

Key words: teleost fishes, polar fishes, antifreeze, zona pellucida, gene duplication, convergent evolution, phylogenetic comparative methods

INTRODUCTION

Convergent evolution of phenotypic traits across independent lineages in similar environments is commonly observed in nature, highlighting the role of natural selection and indicating that phenotypic evolutionary pathways may be predictable under certain ecological conditions (Schluter, 2000; Losos, 2011; Martin and Orgogozo, 2013). Extreme environments, in particular, provide valuable insights into the nature of adaptive evolution due to strong selective pressures (Hotaling et al., 2023). Polar regions, the Arctic and Antarctic, are among the most extreme environments, experiencing pronounced seasonal variation in daylight and temperature (Clarke and Harris, 2003). The Arctic is often defined by the 10°C July isotherm, whereas the Antarctic is defined by the Antarctic Polar Front (Clarke and Harris, 2003; Møller et al., 2005). In polar oceans, water temperatures drop below the freezing point of body fluids for many teleost fishes (Scholander et al., 1957; Christiansen et al., 1995; DeVries and Cheng, 2005). Frigid temperatures can disrupt various cellular processes, such as enzyme catalytic rates, protein structure, and membrane fluidity (Fields, 2001; Pörtner et al., 2007). Despite these harsh conditions, various teleost lineages have successfully colonized polar regions (Eastman, 2005; Møller et al., 2005; Mecklenburg et al., 2011). However, previous studies on molecular mechanisms have focused on a limited number of lineages or genes (DeVries and Cheng, 2005; Beers et al., 2015; Fields et al., 2015; Logan and Buckley, 2015; Daane and Detrich, 2022), leaving those underlying convergent adaptation to polar regions largely unknown.

One of the few examples of molecular mechanisms that enable polar colonization is de novo acquisition of antifreeze proteins (AFPs) (DeVries and Wohlschlag, 1969; DeVries, 1971; Duman and DeVries, 1974, 1976; Near et al., 2012; Bista et al., 2023). Polar fishes have evolved several structurally distinct types of AFPs, including AFP types I, II, and III, and antifreeze glycoproteins (AFGPs), in diverse lineages (Chen et al., 1997; Cheng, 1998; Fletcher et al., 2001; DeVries and Cheng, 2005; Deng et al., 2010; Baalsrud et al., 2018; Rives et al., 2024). These proteins lower the freezing point of blood and body fluids by adsorbing to ice surfaces and inhibiting ice growth, thereby preventing the organisms from freezing (DeVries, 1971; Raymond and DeVries, 1977; Yeh and Feeney, 1996). While de novo acquisition of AFPs is a notable example of genetic innovation in polar fishes, previous studies suggest that larvae may have additional mechanisms for freezing resistance (Cziko et al., 2006). Furthermore, because polar environments affect cellular processes beyond just freezing resistance, genes other than AFPs may also contribute to adaptation.

Gene duplication is one of the major processes that expands the genetic repertoire in teleost fishes (Brawand et al., 2014; Glasauer and Neuhauss, 2014). It can lead to phenotypic changes through various mechanisms such as increased protein dosage, alternative regulatory expression, and neofunctionalization with subsequent amino acid mutations (Magadum et al., 2013). Duplication of certain genes can enhance adaptation to different environments

* Corresponding author. E-mail: nishiguchi-tomoya605@g.ecc.u-tokyo.ac.jp (TN); ishikawa@k.u-tokyo.ac.jp (AI)

doi:10.2108/zs240098

(Kondrashov, 2012). For example, increased copy numbers of fatty acid desaturase genes in freshwater fishes can compensate for lack of docosahexaenoic acid (DHA) (Ishikawa et al., 2019). In Antarctic notothenioid fishes, genes encoding zona pellucida (ZP) proteins show increased copy numbers (Chen et al., 2008, 2019; Kim et al., 2019; Daane and Detrich, 2022; Lu et al., 2022). ZP proteins constitute a fibrillar extracellular matrix surrounding eggs (Litscher and Wassarman, 2018, 2020). ZP proteins from Antarctic notothenioids exhibit novel ice melting-promoting (IMP) activities in eggs, in addition to lowering freezing points (Cao et al., 2016). However, we do not know how prevalent convergent gene duplication is in colonization of polar regions.

In this study, we sought to comprehensively identify genes with increased copy numbers among multiple polar fish lineages. Using 104 teleost genomes, we identified 20 genes showing significant copy number increases in polar fishes after phylogenetic correction. These included genes encoding extracellular glycoproteins and genes upregulated under cold stress. We also investigated whether copy number increases occurred only in polar regions or also along latitudinal clines, and we clarified the timing of gene duplication by estimating phylogenetic trees.

MATERIALS AND METHODS

Comprehensive detection of genes with higher copy numbers in polar fishes

To investigate the association between copy numbers and colonization of polar regions for each orthologous gene, we used OrthoDB v11, a database providing orthologous relationships and functional annotations of genes from representative genomes, based on their longest isoforms (Waterhouse et al., 2013; Kuznetsov et al., 2023). Species classified in the infraclass Teleostei were selected from those available in OrthoDB via NCBI taxonomy IDs using taxize 0.9.101 (Chamberlain and Szöcs, 2013). A phylogenetic tree for ray-finned fishes was obtained using the fishtree_phylogeny function in fishtree 0.3.4 (Chang et al., 2019), selecting only those species for which a phylogenetic tree was available. Latitudinal distribution ranges for selected species were retrieved from FishBase ver. 02/2024 (Froese and Pauly, 1994), and only species with these ranges available were selected. Of the 104 species selected, 23 were classified as polar fishes, based on the checklist of polar fishes by Møller et al. (2005), who defined them as those found in Arctic or Antarctic regions (Møller et al., 2005). The remaining 81 species were classified as non-polar fishes (see Supplementary Table S1). For each of the selected polar and non-polar fishes, we calculated copy numbers of orthologous genes using the Actinopterygii level of orthology (Kuznetsov et al., 2023).

To identify orthologous genes with increased copy numbers in diverse polar fish lineages, we performed family-level downsampling. Specifically, we performed the downsampling to reduce the effect of extensively sequenced lineages (e.g., Salmonidae) when comparing copy numbers between polar and non-polar fishes. Amino acid sequences of the longest isoforms for all genes in selected species were obtained from OrthoDB. These sequences were used as inputs to calculate BUSCO (Benchmarking Universal Single-Copy Orthologs) completeness scores using BUSCO v5.6.1 with the actinopterygii_odb10 database (Manni et al., 2021). The species with the highest BUSCO completeness score in each family was selected from polar fishes when available, and from non-polar fishes otherwise (see Supplementary Table S1). Using the remaining 12 polar and 53 non-polar fishes after downsampling, we selected orthologous genes that met the following criteria: (i) the average copy number in polar fish was more than twice that in non-polar fish, and (ii) more than half of polar fish had at least two copies.

To test the hypothesis that copy numbers of the 54 orthologous genes are associated with colonization of polar regions by pre-downsampling species, we performed Bayesian inference for a generalized linear mixed model (GLMM) with a Poisson distribution using MCMCglmm 2.36 (Hadfield, 2010) with pre-downsampling of 104 species. Copy numbers of orthologous genes were used as the response variable, with habitat type (polar or non-polar) as the predictor. To account for phylogenetic relationships, a covariance structure derived from the phylogenetic tree, pruned to include only pre-downsampling species using ape 5.8 (Paradis and Schliep, 2019), was incorporated as a random effect. The MCMC process was conducted for 10,005,000 iterations with inverse Wishart priors (parameters $V = 1$ and $mu = 0.002$). The initial 5000 samples were excluded as burn-in, with subsequent samples every 1000 iterations being used to calculate p-values (pMCMC_polar). These p-values were adjusted for multiple tests using the Benjamini-Hochberg procedure, with a false discovery rate (FDR) below 0.1 considered statistically significant.

Gene ontology (GO) enrichment analysis

To assess functional information of the detected 54 orthologous genes, we performed gene ontology (GO) enrichment analysis for biological function, cellular component, and molecular function, respectively. For orthologous genes present in at least one of the 65 downsampled species, GO IDs assigned per orthologous gene (i.e., per orthologous group, not per individual gene) were retrieved from OrthoDB. To identify GO IDs overrepresented among detected orthologous genes compared to all orthologous genes in downsampled species, we performed hypergeometric tests. More specifically, we calculated a p-value for each GO ID present in detected orthologous genes using phyper function in R as follows:

$$p\ \text{value} = \text{phyper}\left(n_i, N_i, N - N_i, n, \text{lower.tail} = \text{FALSE}\right)$$

where n is the number of detected orthologous genes, n_i is the number of detected orthologous genes with GO ID i, N is the number of all orthologous genes, and N_i is the number of all orthologous genes with GO ID i. These p-values were adjusted for multiple tests using the Benjamini-Hochberg procedure. We considered GO IDs with an FDR below 0.1 and at least two detected orthologous genes as being significantly enriched. We also performed GO enrichment analysis for the identified 20 orthologous genes that showed significant associations between copy number and colonization of polar regions.

Analysis of the association between gene copy number and latitudinal clines

To examine whether copy number increased not only in polar regions, but also at higher latitudes in general, we tested the association between copy number and maximum absolute latitudes of distribution ranges for the detected 20 orthologous genes that showed significant correlations with polar colonization. We performed GLMM analysis using maximum absolute latitudes of distribution ranges as the predictor instead of habitat type (polar or non-polar) and calculated p-values (pMCMC_latitude). These p-values were adjusted for multiple tests using the Benjamini-Hochberg procedure. FDRs below 0.1 were considered statistically significant.

Phylogenetic tree estimation

To clarify the timing of gene duplication, we estimated phylogenetic trees for the identified 20 orthologous genes. Protein sequences of the longest isoform for each gene were obtained from NCBI. These sequences were aligned using MAFFT v7.525 in

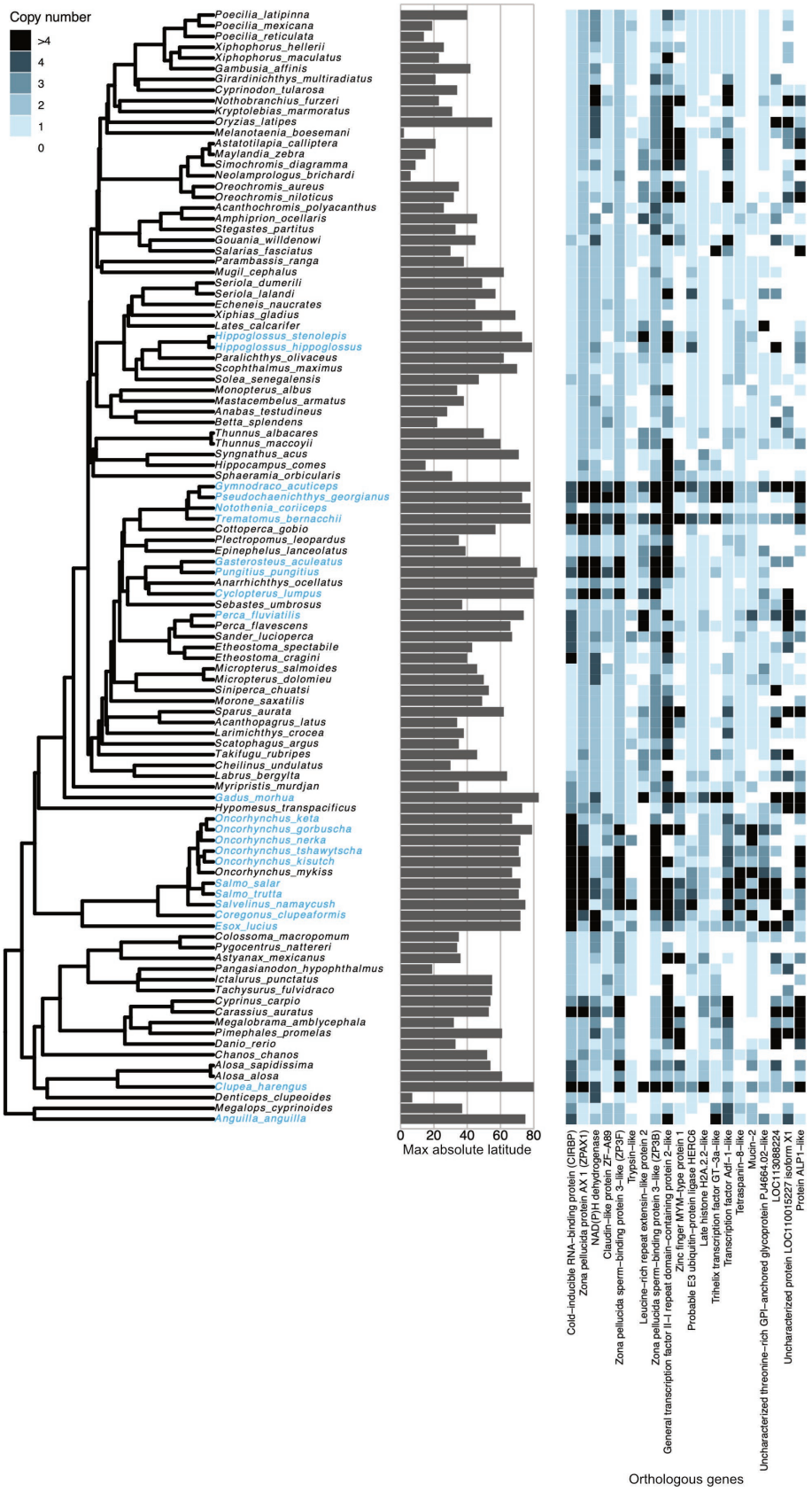

Fig. 1. Copy numbers of orthologous genes significantly increased in polar fishes. A phylogenetic tree for polar and non-polar fishes, including data on maximum latitudes of species ranges and copy numbers of the 20 orthologous genes identified. Polar fishes are labeled in light blue.

L-INS-i mode (Katoh et al., 2002; Katoh and Standley, 2013). Aligned sequences were trimmed using trimAl v1.5.rev0 with the automated1 option to remove gappy sites (Capella-Gutiérrez et al., 2009). GeneRax v2.0.4 (Morel et al., 2020) was used to infer rooted phylogenetic trees. The trimmed alignment, the pruned species tree from fishtree, and the initial tree estimated by IQ-TREE v2.3.6 with the LG+G substitution model (Minh et al., 2020) were used as inputs. Estimation was performed using the LG+G model, accounting for duplication and loss. Phylogenetic trees were visualized using ggtreeExtra v1.12.0 (Xu et al., 2021).

For the three identified ZP orthologous genes (428128at7898, 409056at7898, and 191156at7898), we included some ZP protein sequences from an Antarctic notothenioid fish, *Dissostichus mawsoni*, (DmZPAX1, DmZPC5, and DmZPC1) in the phylogenetic analysis. These ZP proteins from *D. mawsoni*, which have been experimentally verified to possess IMP activity (Cao et al., 2016), were included to clarify orthologous relationships between identified ZP genes and known antifreeze ZP proteins. Protein sequences for DmZPAX1 (AIO03056.1) and DmZPC1 (AJW66345.1) were downloaded from NCBI, while the sequence for DmZPC5 was obtained from Supplementary Figure 5 in Cao et al. (2016).

Data accessibility

Original data can be accessed from the cited sources. Code for the analysis is available at https://github.com/mkrg01/gene_dup_polar.

RESULTS

Genes that show increased copy numbers in polar fish lineages

We first screened for genes with increased copy numbers in polar fish lineages compared to non-polar fish lineages by family-level downsampling of 65 species. Our analysis detected 54 orthologous genes with increased copy numbers in multiple polar fish lineages, such as genes encoding the low choriolytic enzyme-like, pepsin A-like, E3 ubiquitin-protein ligase TRIM39-like, G2/M phase-specific E3 ubiquitin-protein ligase-like, and caspase-8-like (see Supplementary Figure S1; Supplementary Tables S2, S3). GO enrichment analysis for these 54 orthologous genes showed that genes related to proteolysis, regulation of apoptotic process, protein ubiquitination, metal ion binding, zinc ion binding, and protein dimerization activity were significantly overrepresented (Benjamini-Hochberg $FDR < 0.1$ and $N \geq 2$; see Supplementary Table S4). Of the 54 orthologous genes, 20 showed a statistically significant association between copy number and colonization of polar regions even after phylogenetic correction (*p*MCMC, $FDR < 0.1$; Fig. 1, Table 1). The gene encoding the cold-inducible RNA-binding protein (CIRBP) showed the most significant asso-

Table 1. Orthologous genes significantly increased in polar fishes.

Orthologous gene ID	Orthologous gene name	Copy number ratio (polar/non-polar, before downsampling)	Copy number ratio (polar/non-polar, after downsampling)	pMCMC_polar	adjusted_pMCMC_polar	pMCMC_latitude	adjusted_pMCMC_latitude
489137at7898	Cold-inducible RNA-binding protein (CIRBP)	3.1361	3.0749	0.0001	0.0036	0.0006	0.0108
428128at7898	Zona pellucida protein AX 1 (ZPAX1)	4.5744	5.2920	0.0002	0.0036	0.0016	0.0144
299357at7898	NAD(P)H dehydrogenase	2.1344	2.5980	0.0002	0.0036	0.0290	0.0810
465589at7898	Claudin-like protein ZF-A89	2.0389	2.1393	0.0004	0.0054	0.0020	0.0154
409056at7898	Zona pellucida sperm-binding protein 3-like (ZP3F)	2.9531	3.5630	0.0010	0.0108	0.0052	0.0312
172968at7898	Trypsin-like	2.2592	2.6768	0.0012	0.0108	0.0300	0.0810
29671at7898	Leucine-rich repeat extensin-like protein 2	2.1903	4.7321	0.0018	0.0132	0.0016	0.0144
191156at7898	Zona pellucida sperm-binding protein 3-like (ZP3B)	2.9568	3.8178	0.0020	0.0132	0.0176	0.0594
498742at7898	General transcription factor II-I repeat domain-containing protein 2-like	2.0434	2.3209	0.0022	0.0132	0.0112	0.0504
454335at7898	Zinc finger MYM-type protein 1	2.7147	4.5231	0.0034	0.0167	0.1438	0.2427
488885at7898	Probable E3 ubiquitin-protein ligase HERC6	1.8669	2.0784	0.0034	0.0167	0.0134	0.0547
428606at7898	Late histone H2A.2.2-like	1.7406	2.2083	0.0084	0.0378	0.0012	0.0144
469863at7898	Trihelix transcription factor GT-3a-like	4.6514	7.5548	0.0128	0.0532	0.2302	0.3552
485260at7898	Transcription factor Adf-1-like	2.8436	3.8850	0.0168	0.0648	0.1340	0.2334
482472at7898	Tetraspanin-8-like	2.3326	2.4537	0.0238	0.0844	0.0818	0.1841
438657at7898	Mucin-2	3.0415	2.4915	0.0250	0.0844	0.0004	0.0108
500231at7898	Uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02-like	2.8859	3.7140	0.0282	0.0896	0.0776	0.1822
307673at7898	LOC113088224	1.8818	2.0490	0.0344	0.0983	0.0142	0.0547
493746at7898	Uncharacterized protein LOC110015227 isoform X1	1.5129	2.2534	0.0350	0.0983	0.0934	0.1892
482447at7898	Protein ALP1-like	3.1345	4.6843	0.0364	0.0983	0.0082	0.0403

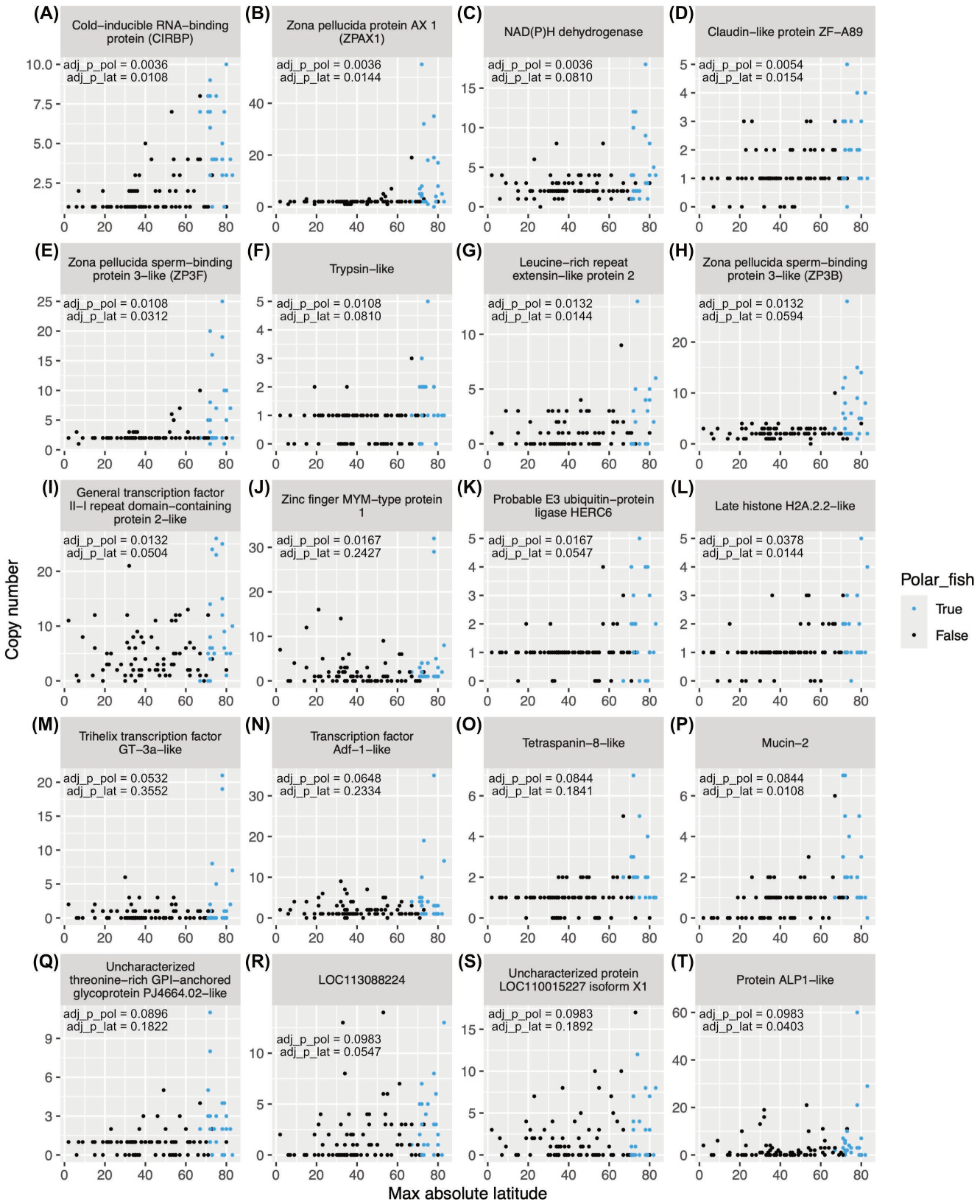

Fig. 2. Scatter plots of maximum absolute latitudes and copy numbers of the 20 orthologous genes selected. Each dot represents polar or non-polar fish before downsampling. Polar fishes are plotted in light blue. adj_p_pol, adjusted pMCMC_polar; adj_p_lat, adjusted pMCMC_latitude.

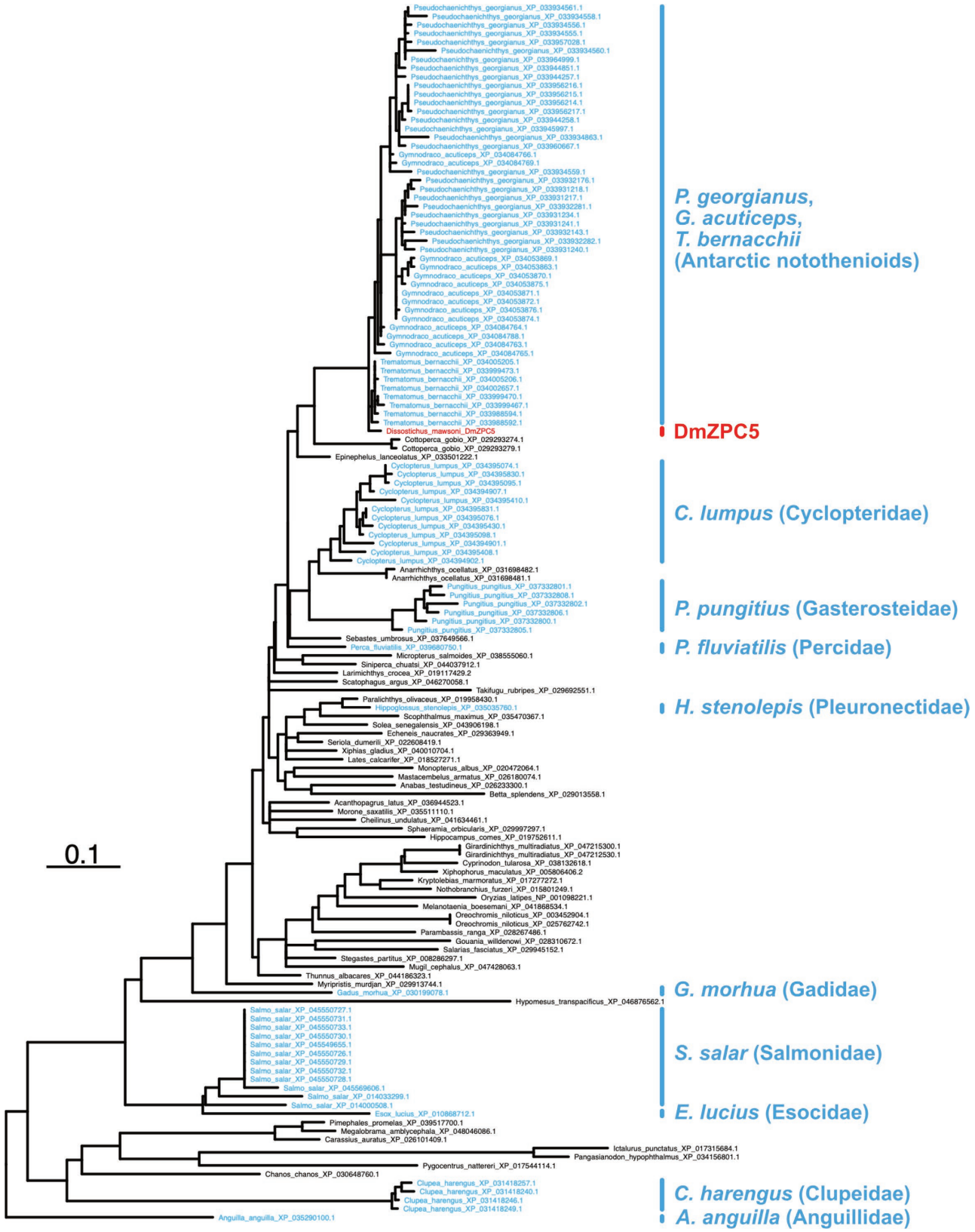

Fig. 3. An estimated phylogenetic tree of OG 191156at7898 (zona pellucida sperm-binding protein 3-like (ZP3B)) after downsampling, including DmZPC5. Accession IDs from polar fishes are labeled in light blue, and DmZPC5 in red. In addition, polar fish species with each sequence and their clade names are highlighted in light blue on the right, whereas DmZPC5 is highlighted in red for clarity. The scale of branch lengths indicates the expected number of amino acid substitutions per amino acid site. This figure shows part of the phylogenetic tree. Please see Supplementary Figure S2 for the complete tree.

ciation. Genes encoding ZP proteins, which show increased copy numbers in Antarctic notothenioid fishes (Cao et al., 2016), were among the detected genes. In addition to ZP genes, genes predicted to encode glycoproteins (leucine-rich repeat extensin-like protein 2, mucin-2, and uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02-like) and those regulating mucin release (tetraspanin-8-like) were also included. Genes related to extracellular regions and protein dimerization activity were significantly enriched in the GO enrichment analysis (Benjamini-Hochberg $FDR < 0.1$ and $N \geq 2$; see Supplementary Table S5).

Genes with and without latitudinal clines

We examined whether the identified 20 orthologous genes with increased copy number in polar regions were also associated with latitudinal clines. Of the 20 orthologous genes, 14 also showed a statistically significant association between copy number and maximum absolute latitude of distribution ranges, whereas six genes did not (pMCMC, $FDR < 0.1$; Fig. 2, Table 1). Genes with increased copy numbers in polar regions, but not significantly associated with latitudes, such as genes encoding the trihelix transcription factor GT-3a-like (Fig. 2M) and transcription factor Adf-1-like (Fig. 2N), showed exceptionally high copy numbers in some polar fishes, but not in non-polar fishes inhabiting high latitudes (Fig. 2).

Phylogenetic trees estimated for the identified 20 genes

Our phylogenetic analysis clearly showed that gene duplication occurred in multiple lineages of polar fish for the identified 20 orthologous genes (Fig. 3, and see Supplementary Figures S2–S21). Notable examples are three orthologous genes encoding ZP proteins (zona pellucida protein AX 1 (ZPAX1), zona pellucida sperm-binding protein 3-like (ZP3F), and zona pellucida sperm-binding protein 3-like (ZP3B)). In all three ZP genes, multiple gene duplications occurred not only in Antarctic notothenioids, but also in several Arctic fish lineages, such as *Cyclopterus lumpus* (Cyclopteridae), *Pungitius pungitius* (Gasterosteidae), *Salmo salar* (Salmonidae), and *Clupea harengus* (Clupeidae) (Fig. 3, and see Supplementary Figures S2–S4). Conversely, no clear evidence of duplication for the three ZP orthologous genes was detected in some polar fish lineages, such as *Perca fluviatilis* (Percidae), *Hippoglossus stenolepis* (Pleuronectidae), *Gadus morhua* (Gadidae), and *Esox lucius* (Esocidae) (Fig. 3, and see Supplementary Figures S2–S4). The three identified ZP genes were homologous to previously identified ZP proteins with antifreeze properties (ZPAX1 to DmZPAX1, ZP3F to DmZPC1, and ZP3B to DmZPC5) (Fig. 3, and see Supplementary Figures S2–S4). In addition to ZP genes, gene duplication occurred in multiple Arctic and Antarctic fish lineages in several orthologous genes, such as those encoding NAD(P)H dehydrogenase (see Supplementary Figure S6), trypsin-like (see Supplementary Figure S8), leucine-rich repeat extensin-like protein 2 (see Supplementary Figure S9), and probable E3 ubiquitin-protein ligase HERC6 (see Supplementary Figure S12).

DISCUSSION

In this study, we identified 54 genes with increased copy number in multiple polar fish lineages. GO enrichment analysis revealed that genes involved in proteolysis, ubiquitination, and apoptosis were overrepresented among these genes. At low temperatures, food consumption rates and digestive enzyme activities decrease (Ahmad et al., 2014). Therefore, the higher copy number of genes related to protein digestive enzymes, such as genes encoding pepsin A-like and trypsin-like, may be adaptive to compensate for their deficiency in cold polar oceans. The increased copy number of genes encoding the low choriolytic enzyme-like, which can digest ZP proteins in the egg extracellular matrix during hatching (Yasumasu et al., 2010; Kawaguchi et al., 2013; Sano et al., 2014), may also be adaptive to digest abundant ZP proteins in polar fish eggs. In addition to low digestive enzyme activities, cold stress is thought to denature or misfold many proteins (Place et al., 2004; Logan and Buckley, 2015). Moreover, polar fishes are thought to experience high oxidative stress, likely due to increased mitochondrial abundance and lipid polyunsaturation (Abele and Puntarulo, 2004; Guderley, 2004; Todgham et al., 2017). Thus, higher copy numbers of genes related to ubiquitination, such as genes encoding the E3 ubiquitin-protein ligase TRIM39-like, G2/M phase-specific E3 ubiquitin-protein ligase-like, and probable E3 ubiquitin-protein ligase HERC6, may help to remove proteins denatured or misfolded by cold and oxidative stress, to maintain homeostasis. Similarly, increased copy numbers of apoptosis-related genes, such as the gene encoding caspase-8-like, may facilitate removal of cells unable to maintain homeostasis under stresses in polar regions. Overall, copy number increases in genes involved in proteolysis, ubiquitination, and apoptosis may promote colonization of polar regions in teleosts.

Among the 20 genes showing a significant association between copy number and polar colonization, genes related to the extracellular region were significantly enriched. In particular, we identified three orthologous genes encoding ZP proteins that showed significant associations between copy numbers and polar colonization. ZP proteins are glycosylated proteins constituting the fibrillar extracellular matrix surrounding eggs (Litscher and Wassarman, 2018, 2020). Increases in copy numbers of ZP genes in teleost fishes are thought to enhance physical protection of eggs in harsh environments, such as alkaline environments (cyprinid fish *Leuciscus waleckii*) (Xu et al., 2017) and near-freezing conditions (the Antarctic notothenioid fish, *D. mawsoni*) (Chen et al., 2008). A previous study experimentally confirmed that some ZP proteins from Antarctic notothenioid fishes, including DmZPAX1, DmZPC1, and DmZPC5 from *D. mawsoni*, depressed melting and freezing points in eggs (Cao et al., 2016). Our phylogenetic analysis clearly showed that the identified three orthologous ZP genes were homologous to previously identified ZP proteins with antifreeze properties (ZPAX1 to DmZPAX1, ZP3F to DmZPC1, and ZP3B to DmZPC5) (Fig. 3, and see Supplementary Figures S2–S4). Gene duplications occurred repeatedly in these ZP genes not only in Antarctic notothenioids, but also in multiple Arctic fish lineages (Fig. 3, and see Supplementary Figures S2–S4). These results indicate that eggs of multiple Arctic fish lineages have independently acquired antifreeze properties through mechanisms similar to Antarctic notothenioids. In some polar fish lineages, however, no duplication was observed in these three ZP genes (Fig. 3, and see Supple-

mentary Figures S2–S4). This may be partly because some species, such as *P. fluviatilis* and *E. lucius*, spawn in warmer seasons (Casselman and Lewis, 1997; Gillet and Dubois, 2007), where freeze tolerance is likely unnecessary. Furthermore, eggs of *G. morhua* are reported to exhibit freeze tolerance (Valerio et al., 1992), suggesting additional molecular adaptations beyond the copy number increase in these three ZP genes. We also identified significant associations in genes predicted to encode glycoproteins (leucine-rich repeat extensin-like protein 2, mucin-2, and uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02-like), and a gene encoding a membrane protein regulating mucin release (tetraspanin-8-like) (Wojnacki et al., 2023). In teleosts, skin mucus helps to protect against physical and chemical attacks and alters its viscosity in response to cold stress (Guardiola et al., 2015; Sanahuja et al., 2018), suggesting that increased copy numbers of these genes may also contribute to physical and chemical adaptation to subzero environments.

We also identified significant associations between copy number and polar colonization in genes potentially related to cold stress adaptation, such as genes encoding the cold-inducible RNA-binding protein (CIRBP), NAD(P)H dehydrogenase, probable E3 ubiquitin-protein ligase HERC6, and trypsin-like. A gene encoding CIRBP increases its expression in response to cold stress (Gracey et al., 2004; Chou et al., 2008; Rebl et al., 2013; Liu et al., 2020; Ma et al., 2022) and stabilizes mRNA by binding to its 3′-untranslated regions (Zhong and Huang, 2017). A gene encoding NADH dehydrogenase that is upregulated under cold stress (Malek et al., 2004; Qian and Xue, 2016) likely compensates for decreased ATP production under reduced kinetic energy in a cold environment (O'Brien, 2011; Coppe et al., 2013). Genes involved in ubiquitination are upregulated in response to cold stress (Gracey et al., 2004; Long et al., 2013; Liu et al., 2020) and are highly expressed in Antarctic notothenioid fishes (Todgham et al., 2007; Shin et al., 2012). They probably remove proteins misfolded or damaged by cold stress besides oxidative stress, as described above (Logan and Buckley, 2015; Todgham et al., 2017). Activity of digestive enzymes, including trypsin, decreases at low temperatures (Ahmad et al., 2014). Elevated gene expression or reduced activity under cold stress suggests that increased copy numbers of these four genes in polar fishes may be adaptive to compensate for shortages of their products caused by extreme cold stress. In addition, these four orthologous genes were significantly associated with latitudinal clines and showed higher copy numbers at higher latitudes, particularly for the gene encoding CIRBP (Fig. 2). As sea surface temperatures tend to decrease at higher latitudes (Pinet, 2006), increased copy numbers of these genes in species inhabiting higher latitudes, including non-polar regions, may also indicate their importance in adaptation to relatively moderate cold stress outside polar regions.

Of the 20 genes significantly associated with polar colonization, six did not show a significant association between copy number and maximum absolute latitude. Some genes, such as those encoding trihelix transcription factor GT-3a-like and transcription factor Adf-1-like, showed a pronounced increase specifically in some polar fishes, rather than a general increase among species inhabiting high latitudes (Fig. 2). Although functions of these genes in adaptations to polar regions are unclear, the exceptional copy number increase of these transcription factors likely contributes to adaptation unique to polar ecosystems.

In conclusion, we systematically identified genes with increased copy numbers among polar fish lineages. In particular, our analysis suggests that copy number increases in genes related to extracellular function are probably involved in polar adaptation. Although further functional analysis of increased copy numbers of these genes is needed, our findings provide a basis for understanding the role of copy number increases in colonization of polar regions.

ACKNOWLEDGMENTS

We thank Steven D. Aird (https://www.sda-technical-editor.org) for editing the manuscript. This work was funded by KAKENHI 23KJ0575 from the Japan Society for the Promotion of Science. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

TN: conceptualization, data curation, formal analysis, funding acquisition, project administration, visualization, writing—original draft; AI: supervision, writing—review and editing.

SUPPLEMENTARY MATERIALS

Supplementary materials for this article are available online. (URL: https://doi.org/10.2108/zs240098)

Supplementary Table S1. Species information used in this study.

Supplementary Table S2. 54 orthologous genes identified before GLMM selection.

Supplementary Table S3. Copy numbers of the detected 54 orthologous genes for teleost species before downsampling.

Supplementary Table S4. Summary of GO enrichment analysis for the detected 54 orthologous genes.

Supplementary Table S5. Summary of GO enrichment analysis for the 20 orthologous genes that showed significant associations between copy numbers and polar colonization.

Supplementary Figure S1. Copy numbers of 54 orthologous genes identified before GLMM selection. A phylogenetic tree for polar and non-polar fishes, including copy numbers of the 54 genes.

Supplementary Figure S2. An estimated phylogenetic tree of OG 191156at7898 (zona pellucida sperm-binding protein 3-like [ZP3B]) after downsampling, including DmZPC5.

Supplementary Figure S3. An estimated phylogenetic tree of OG 428128at7898 (zona pellucida protein AX 1 [ZPAX1]) after downsampling, including DmZPAX1.

Supplementary Figure S4. An estimated phylogenetic tree of OG 409056at7898 (zona pellucida sperm-binding protein 3-like [ZP3F]) after downsampling, including DmZPC1.

Supplementary Figure S5. An estimated phylogenetic tree of OG 489137at7898 (cold-inducible RNA-binding protein [CIRBP]) after downsampling.

Supplementary Figure S6. An estimated phylogenetic tree of OG 299357at7898 (NAD[P]H dehydrogenase) after downsampling.

Supplementary Figure S7. An estimated phylogenetic tree of OG 465589at7898 (claudin-like protein ZF-A89) after downsampling.

Supplementary Figure S8. An estimated phylogenetic tree of OG 172968at7898 (trypsin-like) after downsampling.

Supplementary Figure S9. An estimated phylogenetic tree of OG 29671at7898 (leucine-rich repeat extensin-like protein 2) after downsampling.

Supplementary Figure S10. An estimated phylogenetic tree of OG 498742at7898 (general transcription factor II-I repeat domain-containing protein 2-like) after downsampling.

Supplementary Figure S11. An estimated phylogenetic tree of OG 454335at7898 (zinc finger MYM-type protein 1) after downsampling.

Supplementary Figure S12. An estimated phylogenetic tree of OG 488885at7898 (probable E3 ubiquitin-protein ligase HERC6) after downsampling.

Supplementary Figure S13. An estimated phylogenetic tree of OG 428606at7898 (late histone H2A.2.2-like) after downsampling.

Supplementary Figure S14. An estimated phylogenetic tree of OG 469863at7898 (trihelix transcription factor GT-3a-like) after downsampling.

Supplementary Figure S15. An estimated phylogenetic tree of OG 485260at7898 (transcription factor Adf-1-like) after downsampling.

Supplementary Figure S16. An estimated phylogenetic tree of OG 482472at7898 (tetraspanin-8-like) after downsampling.

Supplementary Figure S17. An estimated phylogenetic tree of OG 438657at7898 (mucin-2) after downsampling.

Supplementary Figure S18. An estimated phylogenetic tree of OG 500231at7898 (uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02-like) after downsampling.

Supplementary Figure S19. An estimated phylogenetic tree of OG 307673at7898 (LOC113088224) after downsampling.

Supplementary Figure S20. An estimated phylogenetic tree of OG 493746at7898 (uncharacterized protein LOC110015227 isoform X1) after downsampling.

Supplementary Figure S21. An estimated phylogenetic tree of OG 482447at7898 (protein ALP1-like) after downsampling.

REFERENCES

Abele D, Puntarulo S (2004) Formation of reactive species and induction of antioxidant defense systems in polar and temperate marine invertebrates and fish. Comp Biochem Physiol A 138: 405–415

Ahmad T, Singh SP, Khangembam BK, Sharma JG, Chakrabarti R (2014) Food consumption and digestive enzyme activity of *Clarias batrachus* exposed to various temperatures. Aquac Nutr 20: 265–272

Baalsrud HT, Tørresen OK, Solbakken MH, Salzburger W, Hanel R, Jakobsen KS, et al. (2018) De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. Mol Biol Evol 35: 593–606

Beers JM, Jayasundara N, Podrabsky JE, Stillman JH, Tomanek L (2015) Antarctic notothenioid fish: what are the future consequences of 'losses' and 'gains' acquired during long-term evolution at cold and stable temperatures? J Exp Biol 218: 1834–1845

Bista I, Wood JMD, Desvignes T, McCarthy SA, Matschiner M, Ning Z, et al. (2023) Genomics of cold adaptations in the Antarctic notothenioid fish radiation. Nat Commun 14: 3412

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature 513: 375–381

Cao L, Huang Q, Wu Z, Cao DD, Ma Z, Xu Q, et al. (2016) Neofunctionalization of zona pellucida proteins enhances freeze-prevention in the eggs of Antarctic notothenioids. Nat Commun 7: 12987

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973

Casselman JM, Lewis CA (1997) Habitat requirements of northern pike (*Essox lucius*). Can J Fish Aquat Sci 53: 161–174

Chamberlain SA, Szöcs E (2013) taxize: taxonomic search and retrieval in R. F1000Res 2: 191

Chang J, Rabosky DL, Smith SA, Alfaro ME (2019) An r package and online resource for macroevolutionary studies using the ray-finned fish tree of life. Methods Ecol Evol 10: 1118–1124

Chen L, DeVries AL, Cheng CH (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc Natl Acad Sci U S A 94: 3811–3816

Chen L, Lu Y, Li W, Ren Y, Yu M, Jiang S, et al. (2019) The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. Gigascience 8: giz016

Chen Z, Cheng CHC, Zhang J, Cao L, Chen L, Zhou L, et al. (2008) Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. Proc Natl Acad Sci U S A 105: 12944–12949

Cheng CH (1998) Evolution of the diverse antifreeze proteins. Curr Opin Genet Dev 8: 715–720

Chou MY, Hsiao CD, Chen SC, Chen IW, Liu ST, Hwang PP (2008) Effects of hypothermia on gene expression in zebrafish gills: upregulation in differentiation and function of ionocytes as compensatory responses. J Exp Biol 211: 3077–3084

Christiansen JS, Chernitsky AG, Karamushko OV (1995) An Arctic teleost fish with a noticeably high body fluid osmolality: a note on the navaga, *Eleginus navaga* (Pallas 1811), from the White Sea. Polar Biol 15: 303–306

Clarke A, Harris CM (2003) Polar marine ecosystems: major threats and future change. Environ Conserv 30: 1–25

Coppe A, Agostini C, Marino IAM, Zane L, Bargelloni L, Bortoluzzi S, et al. (2013) Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. Genome Biol Evol 5: 45–60

Cziko PA, Evans CW, Cheng CHC, DeVries AL (2006) Freezing resistance of antifreeze-deficient larval Antarctic fish. J Exp Biol 209: 407–420

Daane JM, Detrich HW III (2022) Adaptations and diversity of Antarctic fishes: a genomic perspective. Annu Rev Anim Biosci 10: 39–62

Deng C, Cheng CHC, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proc Natl Acad Sci U S A 107: 21593–21598

DeVries AL (1971) Glycoproteins as biological antifreeze agents in antarctic fishes. Science 172: 1152–1155

DeVries AL, Cheng CHC (2005) Antifreeze proteins and organismal freezing avoidance in polar fishes. Fish Physiol 22: 155–201

DeVries AL, Wohlschlag DE (1969) Freezing resistance in some Antarctic fishes. Science 163: 1073–1075

Duman JG, DeVries AL (1974) Freezing resistance in winter flounder *Pseudopleuronectes americanus*. Nature 247: 237–238

Duman JG, DeVries AL (1976) Isolation, characterization, and physical properties of protein antifreezes from the winter flounder, *Pseudopleuronectes americanus*. Comp Biochem Physiol B 54: 375–380

Eastman JT (2005) The nature of the diversity of Antarctic fishes. Polar Biol 28: 93–107

Fields PA (2001) Review: Protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol A 129: 417–431

Fields PA, Dong Y, Meng X, Somero GN (2015) Adaptations of protein structure and function to temperature: there is more than one way to "skin a cat." J Exp Biol 218: 1801–1811

Fletcher GL, Hew CL, Davies PL (2001) Antifreeze proteins of teleost fishes. Annu Rev Physiol 63: 359–390

Froese R, Pauly D (1994) Fishbase as a tool for comparing the life history patterns of flatfish. Neth J Sea Res 32: 235–239

Gillet C, Dubois JP (2007) Effect of water temperature and size of females on the timing of spawning of perch *Perca fluviatilis* L. in Lake Geneva from 1984 to 2003. J Fish Biol 70: 1001–1014

Glasauer SMK, Neuhauss SCF (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics 289: 1045–1060

Gracey AY, Fraser EJ, Li W, Fang Y, Taylor RR, Rogers J, et al. (2004) Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. Proc Natl Acad Sci U S A 101: 16970–16975

Guardiola FA, Cuartero M, Del Mar Collado-González M, Arizcún M, Díaz Baños FG, Meseguer J, et al. (2015) Description and comparative study of physico-chemical parameters of the teleost fish skin mucus. Biorheology 52: 247–256

Guderley H (2024) Metabolic responses to low temperature in fish muscle. Biol Rev 79: 409–427

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm r package. J Stat Softw 33: 1–22

Hotaling S, Desvignes T, Sproul JS, Lins LSF, Kelley JL (2023) Pathways to polar adaptation in fishes revealed by long-read sequencing. Mol Ecol 32: 1381–1397

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, et al. (2019) A key metabolic gene for recurrent freshwater colonization and radiation in fishes. Science 364: 886–889

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–780

Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059–3066

Kawaguchi M, Takahashi H, Takehana Y, Naruse K, Nishida M, Yasumasu S (2013) Sub-functionalization of duplicated genes in the evolution of nine-spined stickleback hatching enzyme. J Exp Zool 320: 140–150

Kim BM, Amores A, Kang S, Ahn DH, Kim JH, Kim IC, et al. (2019) Antarctic blackfin icefish genome reveals adaptations to extreme environments. Nat Ecol Evol 3: 469–478

Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc Biol Sci 279: 5048–5057

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, et al. (2023) OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res 51: D445–D451

Litscher ES, Wassarman PM (2018) The fish egg's zona pellucida. In "Current Topics in Developmental Biology" Ed by ES Litscher, PM Wassarman, Academic Press, pp 275–305

Litscher ES, Wassarman PM (2020) Zona pellucida proteins, fibrils, and matrix. Annu Rev Biochem 89: 695–715

Liu L, Zhang R, Wang X, Zhu H, Tian Z (2020) Transcriptome analysis reveals molecular mechanisms responsive to acute cold stress in the tropical stenothermal fish tiger barb (*Puntius tetrazona*). BMC Genomics 21: 737

Logan CA, Buckley BA (2015) Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. J Exp Biol 218: 1915–1924

Long Y, Song G, Yan J, He X, Li Q, Cui Z (2013) Transcriptomic characterization of cold acclimation in larval zebrafish. BMC Genomics 14: 612

Losos JB (2011) Convergence, adaptation, and constraint. Evolution 65: 1827–1840

Lu Y, Li W, Li Y, Zhai W, Zhou X, Wu Z, et al. (2022) Population genomics of an icefish reveals mechanisms of glacier-driven adaptive radiation in Antarctic notothenioids. BMC Biol 20: 231

Ma S, Luo S, Zhang K, Liu Y, Wei X, Cheng J, et al. (2022) Mitigation of low temperature stress by increased salinity is associated with multiple physiological responses in the gills of *Takifugu fasciatus*. Mar Biol 169: 141

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R (2013) Gene duplication as a major force in evolution. J Genet 92: 155–161

Malek RL, Sajadi H, Abraham J, Grundy MA, Gerhard GS (2004) The effects of temperature reduction on gene expression and oxidative stress in skeletal muscle from adult zebrafish. Comp Biochem Physiol C 138: 363–373

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38: 4647–4654

Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67: 1235–1250

Mecklenburg CW, Møller PR, Steinke D (2011) Biodiversity of arctic marine fishes: taxonomy and zoogeography. Mar Biodivers 41: 109–140

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37: 1530–1534

Møller PR, Nielsen JG, Anderson ME (2005) Systematics of polar fishes. Fish Physiol 22: 25–78

Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ (2020) GeneRax: a tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. Mol Biol Evol 37: 2763–2774

Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, et al. (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc Natl Acad Sci U S A 109: 3434–3439

O'Brien KM (2011) Mitochondrial biogenesis in cold-bodied fishes. J Exp Biol 214: 275–285

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528

Pinet PR (2006) Invitation to Oceanography. 4th ed, Jones & Bartlett Learning, Sudbury, MA

Place SP, Zippay ML, Hofmann GE (2004) Constitutive roles for inducible genes: evidence for the alteration in expression of the inducible *hsp70* gene in Antarctic notothenioid fishes. Am J Physiol Regul Integr Comp Physiol 287: R429–R436

Pörtner HO, Peck L, Somero G (2007) Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. Philos Trans R Soc Lond B 362: 2233–2258

Qian B, Xue L (2016) Liver transcriptome sequencing and de novo annotation of the large yellow croaker (*Larimichthy crocea*) under heat and cold stress. Mar Genomics 25: 95–102

Raymond JA, DeVries AL (1977) Adsorption inhibition as a mechanism of freezing resistance in polar fishes. Proc Natl Acad Sci U S A 74: 2589–2593

Rebl A, Verleih M, Köbis JM, Kühn C, Wimmers K, Köllner B, et al. (2013) Transcriptome profiling of gill tissue in regionally bred and globally farmed rainbow trout strains reveals different strategies for coping with thermal stress. Mar Biotechnol 15: 445–460

Rives N, Lamba V, Cheng CHC, Zhuang X (2024) Diverse origins of near-identical antifreeze proteins in unrelated fish lineages provide insights into evolutionary mechanisms of new gene birth and protein sequence convergence. Mol Biol Evol 41: msae182

Sanahuja I, Fernández-Alacid L, Sánchez-Nuño S, Ordóñez-Grande B, Ibarz A (2018) Chronic cold stress alters the skin mucus interactome in a temperate fish model. Front Physiol 9: 1916

Sano K, Kawaguchi M, Watanabe S, Yasumasu S (2014) Neofunctionalization of a duplicate hatching enzyme gene during the evolution of teleost fishes. BMC Evol Biol 14: 221

Schluter D (2000) The Ecology of Adaptive Radiation. Oxford University Press, Oxford

Scholander PF, van Dam L, Kanwisher JW, Hammel HT, Gordon MS (1957) Supercooling and osmoregulation in arctic fish. J Cell Comp Physiol 49: 5–24

Shin SC, Kim SJ, Lee JK, Ahn DH, Kim MG, Lee H, et al. (2012) Transcriptomics and comparative analysis of three antarctic notothenioid fishes. PLOS ONE 7: e43762

Todgham AE, Hoaglund EA, Hofmann GE (2007) Is cold the new hot? Elevated ubiquitin-conjugated protein levels in tissues of Antarctic fish as evidence for cold-denaturation of proteins in vivo. J Comp Physiol B 177: 857–866

Todgham AE, Crombie TA, Hofmann GE (2017) The effect of temperature adaptation on the ubiquitin–proteasome pathway in notothenioid fishes. J Exp Biol 220: 369–378

Valerio PF, Goddard SV, Kao MH, Fletcher GL (1992) Survival of northern Atlantic cod (*Gadus morhua*) eggs and larvae when exposed to ice and low temperature. Can J Fish Aquat Sci 49: 2588–2595

Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV (2013) OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res 41: D358–D365

Wojnacki J, Lujan AL, Brouwers N, Aranda-Vallejo C, Bigliani G, Rodriguez MP, et al. (2023) Tetraspanin-8 sequesters syntaxin-2 to control biphasic release propensity of mucin granules. Nat Commun 14: 3710

Xu J, Li JT, Jiang Y, Peng W, Yao Z, Chen B, et al. (2017) Genomic basis of adaptive evolution: the survival of Amur ide (*Leuciscus waleckii*) in an extremely alkaline environment. Mol Biol Evol 34: 145–159

Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, et al. (2021) GgtreeExtra: compact visualization of richly annotated phylogenetic data. Mol Biol Evol 38: 4039–4042

Yasumasu S, Kawaguchi M, Ouchi S, Sano K, Murata K, Sugiyama H, et al. (2010) Mechanism of egg envelope digestion by hatching enzymes, HCE and LCE in medaka, *Oryzias latipes*. J Biochem 148: 439–448

Yeh Y, Feeney RE (1996) Antifreeze proteins: structures and mechanisms of function. Chem Rev 96: 601–618

Zhong P, Huang H (2017) Recent progress in the research of cold-inducible RNA-binding protein. Future Sci OA 3: FSO246

(Received November 9, 2024 / Accepted December 1, 2024 / Published online January 22, 2025)