# Genomics of Secondarily Temperate Adaptation in the Only Non-Antarctic Icefish

Angel G. Rivera-Colón [1], Niraj Rayamajhi [1], Bushra Fazal Minhas [2], Giovanni Madrigal [1], Kevin T. Bilyk [3], Veronica Yoon,[1] Mathias Hüne [4], Susan Gregory [5,6], C.-H. Christina Cheng [1], and Julian M. Catchen [1,*]

[1]Department of Evolution, Ecology, and Behavior, University of Illinois at Urbana-Champaign, Urbana, IL
[2]Informatics Program, University of Illinois at Urbana-Champaign, Urbana, IL
[3]Department of Biology, Montclair State University, Montclair, NJ
[4]Centro de Investigación para la Conservación de los Ecosistemas Australes, Punta Arenas, Chile
[5]British Antarctic Survey, Cambridge, United Kingdom
[6]Government of South Georgia and the South Sandwich Islands, Stanley, Falklands

***Corresponding author:** E-mail: jcatchen@illinois.edu.

**Associate editor:** Joanna Kelley

## Abstract

**White-blooded Antarctic icefishes, a family within the adaptive radiation of Antarctic notothenioid fishes, are an example of extreme biological specialization to both the chronic cold of the Southern Ocean and life without hemoglobin. As a result, icefishes display derived physiology that limits them to the cold and highly oxygenated Antarctic waters. Against these constraints, remarkably one species, the pike icefish *Champsocephalus esox*, successfully colonized temperate South American waters. To study the genetic mechanisms underlying secondarily temperate adaptation in icefishes, we generated chromosome-level genome assemblies of both *C. esox* and its Antarctic sister species, *Champsocephalus gunnari*. The *C. esox* genome is similar in structure and organization to that of its Antarctic congener; however, we observe evidence of chromosomal rearrangements coinciding with regions of elevated genetic divergence in pike icefish populations. We also find several key biological pathways under selection, including genes related to mitochondria and vision, highlighting candidates behind temperate adaptation in *C. esox*. Substantial antifreeze glycoprotein (AFGP) pseudogenization has occurred in the pike icefish, likely due to relaxed selection following ancestral escape from Antarctica. The canonical *AFGP* locus organization is conserved in *C. esox* and *C. gunnari*, but both show a translocation of two *AFGP* copies to a separate locus, previously unobserved in cryonotothenioids. Altogether, the study of this secondarily temperate species provides an insight into the mechanisms underlying adaptation to ecologically disparate environments in this otherwise highly specialized group.**

***Key words:*** **icefish, genome assembly, RADseq, temperate adaptation, chromosomal inversions, population genomics.**

## Introduction

Antarctic notothenioid fish, or cryonotothenioids, have experienced extreme biological specialization, evolving and diversifying in the chronically cold environment of the Southern Ocean. Following Antarctica's thermal and geographical isolation over the last 35 My (Eastman 2005), cryonotothenioids evolved molecular and physiological adaptations to the extreme cold, most notably antifreeze glycoproteins (AFGPs; Chen et al. 1997; DeVries and Cheng 2005) that allowed them to thrive in an otherwise inhospitable environment. As other Antarctic fish taxa became extinct, this group displayed rapid rates of diversification and underwent an adaptive radiation (Matschiner et al. 2011; Near et al. 2012; Colombo et al. 2015) dominating modern fish fauna in the Southern Ocean and playing a key role in the ecosystem's food web (La Mesa et al. 2004). However, cold specialization has its tradeoffs. Cryonotothenioids are stenothermal, with physiological performance limited to low and narrow temperature windows (Beers and Sidell 2011), and are unable to mount the classic heat shock response (Hofmann et al. 2000; Bilyk and Cheng 2014; Bilyk et al. 2018).

Among Antarctic notothenioids, further specialization is observed in the family of the white-blooded icefishes (Channichthyidae). Characterized by a complete loss of both hemoglobin and red blood cells (Sidell and O'Brien 2006), icefish have evolved various mechanisms to compensate for this lack of active oxygen transport including larger hearts and thus greater cardiac stroke volume (Sidell and O'Brien 2006), changes in the morphology and density of mitochondria (Johnston et al. 1998; O'Brien and Mueller 2010), a novel molecular method to increase $CO_2$ excretion (Harter et al. 2018), and increased cellular lipids (Palmerini et al. 2009). Even after the







Article

evolution of these phenotypes, icefish still exhibit limited cardiac performance, with <10% of the oxygen carrying capacity of red-blooded notothenioids (Holeton 1970) and far larger cardiac energy expenditures (Tota et al. 1991; Tota and Gattuso 1996). Given these biological limitations, the loss of hemoglobin in icefish was often considered to be maladaptive (Sidell and O'Brien 2006), although recent work has proposed alternative hypotheses related to adaptive changes in the presence of sparse iron and alternative functional roles for partial hemoglobin fragments (Corliss et al. 2019). Regardless of the evolutionary mechanism behind the loss, the derived physiology stemming from the lack of hemoglobin is predicted to limit icefishes' adaptive potential to migrate to warmer waters or to respond to rapidly changing ocean temperatures.

Despite these constraints, a single icefish lineage succeeded in establishing itself north of the Antarctic Polar Front (Stankovic et al. 2002; Kock 2005). The pike icefish, *Champsocephalus esox*, is distributed along the Patagonia coast and Strait of Magellan in South America, and around the Falkland Islands (fig. 1), nonfreezing and seasonally variable environments that are significantly warmer, and much lower in oxygen concentration than the Southern Ocean. *Champsocephalus esox* shares the defining traits of the Antarctic icefish clade, including the lack of hemoglobin (Grove et al. 2004; Kock 2005), loss of myoglobin (like its Antarctic congener, *Champsocephalus gunnari*; Grove et al. 2004), and with AFGP coding sequences apparently still present in its DNA (Miya et al. 2016). This prompts the question: how does adaptation and survival in a temperate environment occur from a highly derived and cold-specialized Antarctic ancestral state? The pike icefish presents an ideal model to address this evolutionary question. Its recent divergence (~1.6 My; Stankovic et al. 2002; Near et al. 2012; Dornburg et al. 2017) from its Antarctic sister species, the mackerel icefish *C. gunnari*, enables a direct comparison of closely related Antarctic and temperate genomes. *Champsocephalus gunnari* serves as a proxy of the Antarctic ancestral state, against which we can identify genetic changes in *C. esox* related to the adaptation to a new temperate environment.

*Champsocephalus esox* remains a poorly researched species. The few published studies provide some examples of phenotypic differences in *C. esox* compared with other icefish, including different reproductive biology (Calvo et al. 1999), increased mitochondrial densities (Johnston et al. 1998), and mitochondrial cristae morphology that is more similar to that of red-blooded notothenioid species (Johnston et al. 1998). The underlying genome and architecture associated with temperate adaptation in this species is completely unknown. Given that cryonotothenioids share a temperate common ancestor (Matschiner et al. 2011; Near et al. 2012), and other polar-to-temperate transitions have also occurred in several red-blooded cryonotothenioid lineages (Stankovic et al. 2002; Dornburg et al. 2017), understanding adaptive processes active in *C. esox* will uncover whether a single or a diversity of adaptive processes enabled the recolonization of temperate environments in this group.

To explore the genomic signatures of secondarily temperate adaptation in the pike icefish, we sequenced the genome of both *C. esox* as well as its Antarctic sister species *C. gunnari* using PacBio continuous long-read (CLR) sequencing, and generated high-quality, chromosome-level genome assemblies for whole-genome comparative analyses. In addition to *C. gunnari*, we included comparisons with six available notothenioid genomes—two icefish, three red-blooded species across two different families, and one temperate outgroup to the Antarctic clade—to detect *C. esox*-specific changes in genome architecture and traits putatively under selection. In addition to the genome assemblies, we generated measures of population-level genetic diversity and divergence for *C. esox* and *C. gunnari*, which we used to uncover patterns of divergence and selection between Antarctic and temperate icefish populations. We found that the *C. esox* genome is largely conserved in structure and organization to that of other characterized icefish genomes; however, we also found evidence of structural variation specific to the *C. esox* lineage. These structural variants often co-occur in genomic regions with elevated genetic divergence and directional selection, highlighting the complex genetic architecture underlying secondarily temperate adaptation. We observe several genes under selection related to the mitochondria and visual perception, suggesting possible mechanisms for adapting to temperate environments subsequent to the specialization to Antarctic cold and oxygen-rich conditions, and unique polar light–dark regime.

## Results

### Genome Assembly and Annotation

In order to characterize the genetics underlying temperate adaptation in an icefish, we first sequenced the genomes of both the secondarily temperate icefish, *C. esox*, and its Antarctic sister species, *C. gunnari*, using PacBio CLR and scaffolded using Hi-C chromosome conformation capture libraries, generating chromosome-level reference genome assemblies for each species. For *C. esox*, a single male individual from Puerto Natales, Chile (fig. 1) was used for genome sequencing. The resulting genome assembly was 987 Mb in length (table 1), comparable with the assembly size of around 1 Gb described in other icefish genome assemblies (Bargelloni et al. 2019; Kim et al. 2019; Bista et al. 2022; Lu et al. 2022). The scaffold and contig N50 were 43.6 and 2.6 Mb, respectively, and 97.72% of the total bases were assembled into 24 putative chromosomes. The final assembly was 96.3% gene complete, according to analysis by *BUSCO* v5.1.3, using the Actinopterygii reference gene set. In addition, we sequenced RNA extracted and pooled from several different tissues, which we used as transcript-level evidence for annotating the genome using the *BRAKER* software (Brůna et al. 2021; Gabriel et al. 2021). A total of 28,257 protein-coding genes were annotated in this *C. esox* assembly.

The *C. gunnari* reference assembly, obtained from a single male individual collected from the West Antarctic Peninsula (fig. 1) was 994 Mb in length, with a scaffold and contig N50 of 44.1 and 3.2 Mb, respectively (table 1).

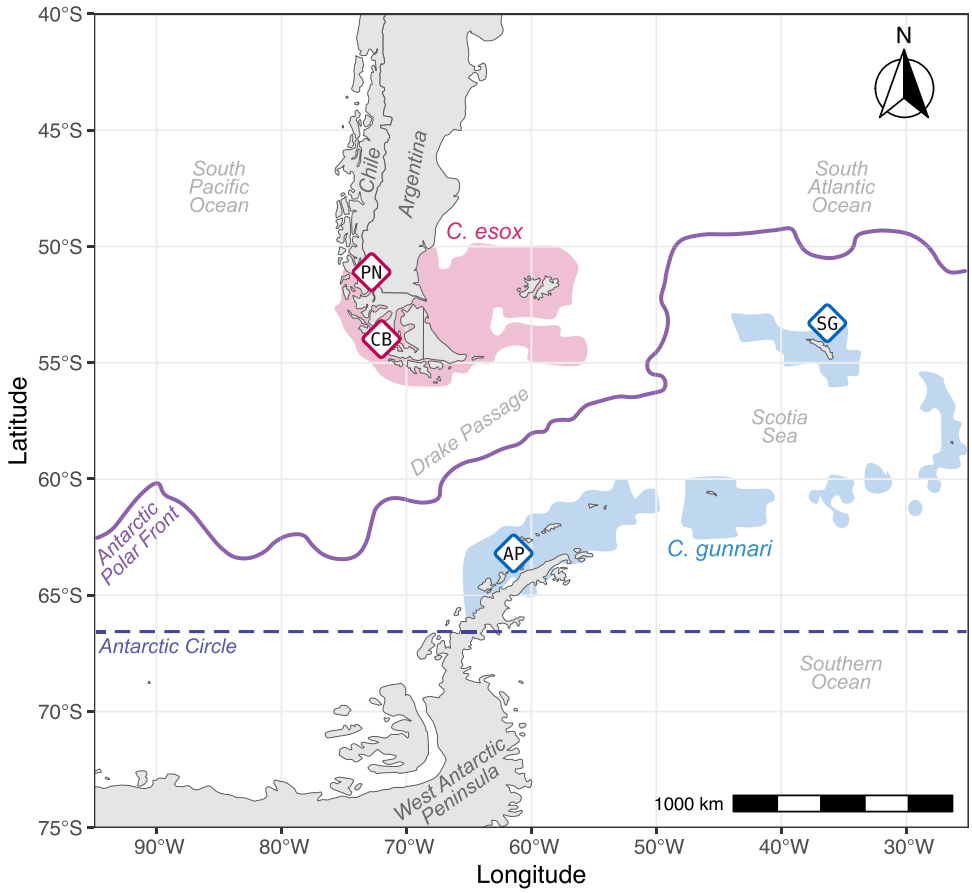


**Fig. 1.** Species ranges and sampling of temperate and Antarctic icefish species. Map showing the approximate species range of the temperate pike icefish *Champsocephalus esox* and the Antarctic mackerel icefish *Champsocephalus gunnari* (AquaMaps 2019). The approximate boundary of the Antarctic Polar Front is shown by the solid line (Orsi et al. 1995). *Champsocephalus esox* individuals were sampled from Puerto Natales, Chile (PN, $n = 16$) and Canal Bárbara, Chile (CB, $n = 12$), whereas the *C. gunnari* individuals were collected from the Gerlache Strait in the West Antarctic Peninsula (AP, $n = 20$) and South Georgia island (SG, $n = 48$). Genomes for *C. esox* and *C. gunnari* were assembled from individuals collected from the PN and AP populations, respectively.

A total of 98.4% of bases were assembled into 24 chromosome-level scaffolds, consistent with a haploid chromosome number of 24 by cytogenic characterization of the *C. gunnari* karyotype (Ozouf-Costaz et al. 1996). According to *BUSCO* v5.1.3, this assembly was 96.3% gene complete. A total of 27,918 protein-coding genes were annotated based on transcript-level evidence obtained from RNAseq data. For both the *C. esox* and *C. gunnari* assemblies, a complementary assessment of gene completeness at the protein and genome levels is available on supplementary table S1, Supplementary Material online.

## Conserved Synteny Between Temperate and Antarctic Icefish Genomes

We performed a conserved synteny analysis using *Synolog* (Catchen et al. 2009; Small et al. 2016) to identify patterns of genomic organization between the genomes of *C. esox*, *C. gunnari*, and that of other available, high-quality notothenioid assemblies (Kim et al. 2019; Jo et al. 2021; Bista et al. 2022; supplementary table S2, Supplementary Material online). In addition to size and chromosome number, the *C. esox* and *C. gunnari* genomes display conservation in large-scale genome organization, showing one-to-one correspondence between 24 orthologous chromosomes (fig. 2A). Similar patterns can be observed between *C. esox* and other chromosome-scale icefish genomes (supplementary figs. S1 and S2, Supplementary Material online), demonstrating that the conservation in genome structure in *C. esox* is not specific to the *Champsocephalus* lineage and might be conserved across icefishes. However, several examples of intrachromosomal rearrangements were observed between the two *Champsocephalus* genomes (fig. 2B and C), with 12 out of the 24 chromosomes displaying genomic rearrangements of considerable length (>1 Mb). Since we had applied a two-assembly approach for each genome, these chromosomal rearrangements were validated by looking at the patterns of within-species synteny (see Materials and Methods). By comparing conserved synteny against the chromosome-scale genome assembly of an outgroup, the non-Antarctic notothenioid *Eleginops maclovinus*—the closest sister to the Antarctic clade (Cheng et al., *in preparation*), we observed examples of rearrangements specific to the *C. esox* lineage, originating in or during its split from the Antarctic *C. gunnari* around 1.6 My (Stankovic et al. 2002; Near et al. 2012; Dornburg et al. 2017). One remarkable example is chromosome 21 in *C. esox* (fig. 2C), which displays a double chromosomal inversion when compared with both the icefish sister species *C. gunnari* and the non-Antarctic *E. maclovinus*.

## Repeat Distributions Along the Genome

Repeat elements have been associated with patterns of genome evolution in cryonotothenioids, including an

**Table 1.** Genome Assembly and Annotation Statistics.

| | *C. esox* | | *C. gunnari* | |
|---|---|---|---|---|
| ***Genome assembly*** | | | | |
| **Scale** | **Contig** | **Scaffold** | **Contig** | **Scaffold** |
| **Total size (bp)** | **986,435,181** | **987,108,782** | **993,506,282** | **994,201,109** |
| **Number of fragments** | **3,499** | **2,067** | **2,992** | **1,536** |
| **Largest fragment (bp)** | **9,841,188** | **56,719,351** | **15,571,738** | **55,323,500** |
| **N50 (bp)** | **2,611,138** | **43,613,322** | **3,182,242** | **44,089,792** |
| **L50** | **107** | **11** | **84** | **11** |
| **Number of chromosome-scale Scaffolds** | **–** | **24** | **–** | **24** |
| **Total bases in chromosomes (bp)** | **–** | **964,606,985** | **–** | **978,689,797** |
| **Percent of Assembly in Chromosomes** | **–** | **97.72%** | **–** | **98.44%** |
| ***Gene completeness (BUSCO v5.1.3)*** | | | | |
| **Complete** | **3,506 (96.3%)** | | **3,506 (96.3%)** | |
| **Complete and single copy** | **3,463 (95.1%)** | | **3,469 (95.3%)** | |
| **Complete and duplicated** | **43 (1.2%)** | | **37 (1.0%)** | |
| **Fragmented** | **14 (0.4%)** | | **19 (0.5%)** | |
| **Missing** | **120 (3.3%)** | | **115 (3.2%)** | |
| **Total** | **3,640** | | **3,640** | |
| **Reference gene set** | `actinopterygii_odb10` | | `actinopterygii_odb10` | |
| ***Genome annotation*** | | | | |
| **Number of protein-coding genes** | **28,257** | | **27,918** | |
| **Number of exons** | **246,908** | | **249,138** | |

expansion in genome size and chromosomal diversification (Detrich et al. 2010; Auvinet et al. 2018; Chen et al. 2019). Therefore, we performed a *de novo* annotation of the repeat elements present in our assemblies to identify changes in repeat content putatively linked to temperate adaptation. Both the *C. esox* and *C. gunnari* genomes appear to contain nearly identical repeat distributions, composed of 59.38% and 59.45% repeat content, respectively (fig. 3*A*, supplementary table S3, Supplementary Material online). Highly similar values were also observed for the major interspersed repeat element types, that is DNA, SINE, LINE, and LTR (fig. 3*A*, supplementary table S3, Supplementary Material online). These repeat distributions match those observed in both other icefish genomes (Kim et al. 2019; Bista et al. 2022), and red-blooded Antarctic notothenioids (Jo et al. 2021; Bista et al. 2022; fig. 3*A*), suggesting that repeats in *C. esox* are largely not related to the secondarily temperate adaptation process and might instead be present across species of the Antarctic clade.

Both *C. esox* and *C. gunnari* display repeats that are more concentrated along the termini of chromosomes than in their centers (fig. 3*B* and *C*). One clear exception is the high repeat content in the center of chromosome 21 in *C. esox* (fig. 3*B*), where a species-specific rearrangement is found (fig. 2*C*). The double inversion pattern observed would cause the two repeat-rich terminal ends in the *C. gunnari* chromosome to be located in the center of the *C. esox* chromosome.

## Genome-wide Diversity and Divergence

To measure patterns of genetic diversity and divergence between *C. esox* and *C. gunnari* populations we genotyped individuals using restriction site-associated DNA sequencing (RADseq). For each species, we collected individuals across two sites (fig. 1; hereafter referred to as populations); 28 individuals for *C. esox* (16 from Puerto Natales and 12 from Canal Bárbara), and 68 for *C. gunnari* (48 from South Georgia Island and 20 from the West Antarctic Peninsula). Sequenced reads from these individuals were aligned to our new *C. esox* assembly and genotyped using *Stacks* v2 (Rochette et al. 2019). After filtering missing data, we retained ~56K RAD loci (average locus length of 639 bp) along the genome, comprised of 288K variant sites. The two *C. gunnari* populations, from South Georgia Island and the West Antarctic Peninsula, were more divergent from one another than the two populations of *C. esox* from Puerto Natales and Canal Bárbara (supplementary table S4, Supplementary Material online). To limit the effects of this within-species population structure, we removed the West Antarctic Peninsula *C. gunnari* samples from later steps of the analysis, assaying *C. esox* samples in a single, species-level population (see Materials and Methods). At this species-level comparison, we observe higher genetic diversity in the *C. gunnari* populations (variant-site $\pi$ of 0.12) than in the *C. esox* populations ($\pi$ of 0.088). We also find high genetic divergence between *C. esox* and *C. gunnari* populations (mean $F_{ST} =$ 0.40, mean $D_{XY} = 0.0042$). Given the high proportion of elevated $F_{ST}$ values observed between the populations of both species (supplementary fig. S3, Supplementary Material online), likely due to differentially fixed sites resulting from a recent history of isolation, we decided to use the absolute divergence as a comparative metric for subsequent analysis. Regions of high genetic divergence were spread across the genome, with 23 out of 24 chromosomes containing at least one genomic window with outlier $D_{XY}$ values.

Patterns of elevated genetic divergence between populations can be the product of both selective and neutral

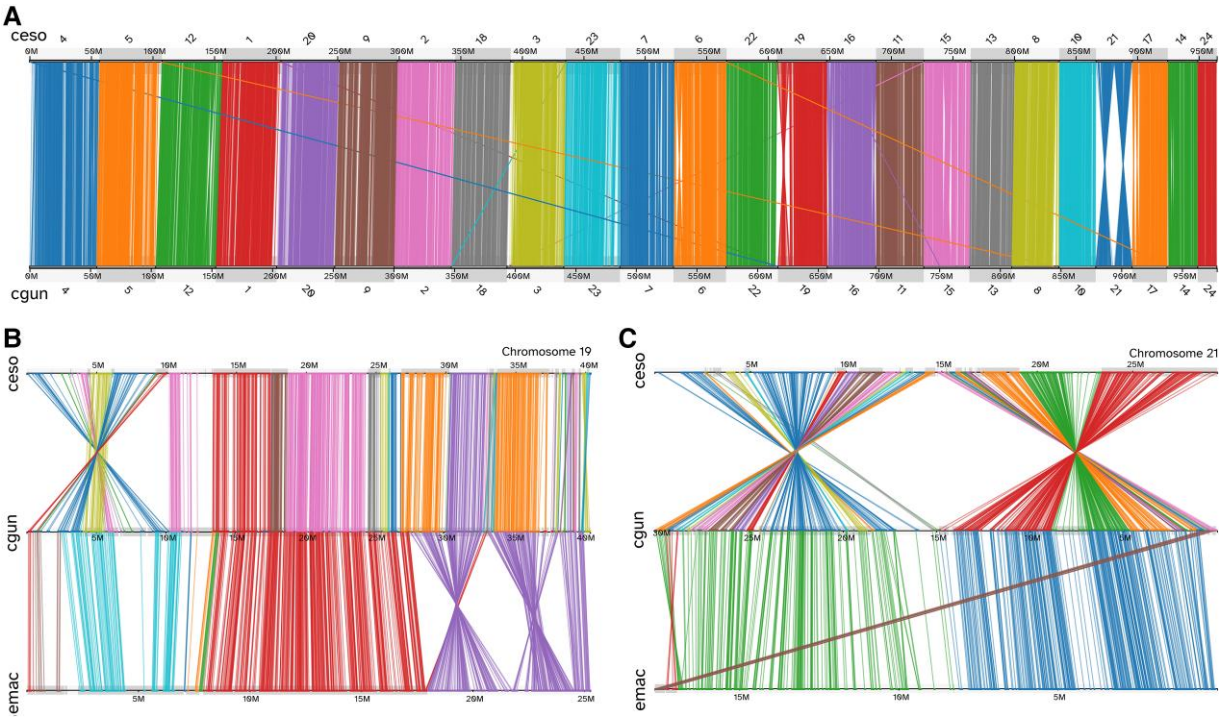


FIG. 2. Patterns of conserved synteny between the genomes of temperate and Antarctic icefish species. (*A*) Genome-wide synteny plot shows one-to-one correspondence between the 24 chromosomes of *Champsocephalus esox* (top) and *Champsocephalus gunnari* (bottom). Each line represents an orthologous gene between the two genomes, color-coded according to the chromosome of origin. (*B*) Conserved synteny between chromosome 19 in *C. esox* (top), *C. gunnari* (middle), and *Eleginops maclovinus* (bottom). Lines are color-coded according to the genomic contig of origin in either *C. esox* or *E. maclovinus*. A *C. esox*-specific chromosomal inversion is seen in the chromosome, located between 1 and 10 Mb of the *C. esox* chromosome. There are two additional inversions between the icefish and *E. maclovinus* lineages. (*C*) Conserved synteny between chromosome 21 in *C. esox* (top), *C. gunnari* (middle), and *E. maclovinus* (bottom). This chromosome displays two large chromosomal inversion specific to the *C. esox* lineage.

processes. To further test the patterns of selection behind regions of genetic divergence, we calculated extended haplotype homozygosity (*EHH*; Sabeti et al. 2002) separately in each species. The *EHH* statistic looks for signatures of positive selection in a population by identifying changes in homozygosity and linkage disequilibrium across the genome. We then compared the patterns of *EHH* between *C. esox* and *C. gunnari* populations using the cross-population *EHH* statistic (*XP-EHH*; Sabeti et al. 2007), which allowed us to identify changes in *EHH*—and thus differences in positive selection—between temperate and Antarctic icefish populations. Patterns of population-level directional selection appeared to be more common in *C. esox* than in *C. gunnari*, spanning all 24 chromosomes. Out of the 976 sites displaying outlier *XP-EHH* values, 939 (96.2%) showed directional positive selection in temperate populations. There was also strong co-localization between patterns of divergence and selection, as windows of divergence ($D_{XY}$) and *EHH* outliers overlapped in 20 chromosomes. Furthermore, there was a correlation with the presence of genomic rearrangements and these selection/divergence outliers (fig. 4). Although not all selection/divergence outlier windows were co-localized with structural differences—and not all structural differences display divergence and/or selection outliers—there were examples of these outlier windows limited to the boundaries of a genomic inversion, for example chromosome 12 (fig. 4*A*), or outlier windows located within the span of a larger chromosomal rearrangement, for example chromosome 21 (fig. 4*B*).

## Candidate Genes Under Selection

We used a two-pronged approach to identify candidate genes under selection in *C. esox*. First, we applied an analysis of molecular evolution using the branch-site model implemented in the software *PAML* (Yang 2007), to identify significant differences in the rate of nonsynonymous to synonymous substitutions ($\omega$) in the protein-coding genes of *C. esox* relative to orthologs of five other cryonotothenioids. With this approach, we identified a set of genes displaying elevated $\omega$ values, a measure of positive selection in the focal lineage (supplementary fig. S4, Supplementary Material online). By comparing the annotations of six cryonotothenioid taxa, including four icefish and two red-blooded Antarctic species, we identified 14,093 orthogroups composed of one single-copy ortholog per-species. Out of these 14K orthogroups, 258 genes displayed significantly elevated $\omega$ values (hereafter referred to as d*N*/d*S* candidates, supplementary table S5, Supplementary Material online).

We also identified protein-coding genes that are located within regions displaying population-level signatures of

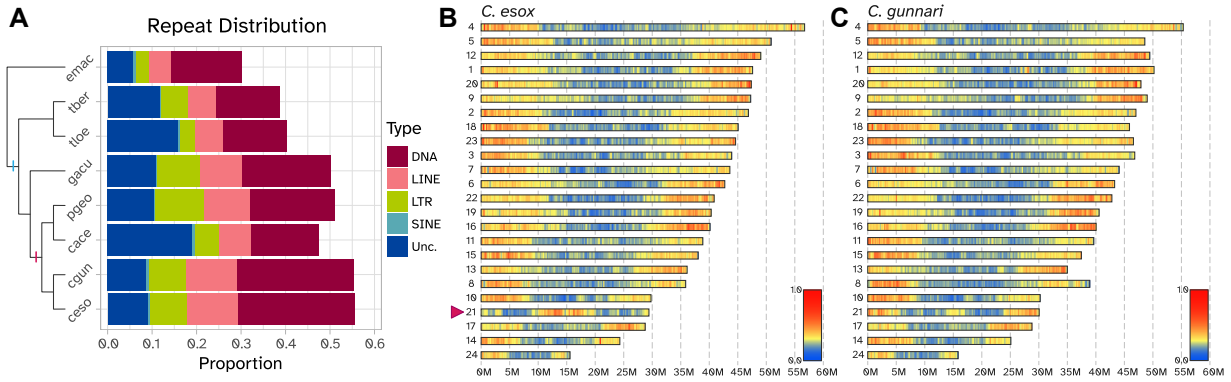


FIG. 3. Repeat distribution is conserved along icefish genomes. (*A*) Interspersed repeat distributions for *Champsocephalus esox* (ceso), *Champsocephalus gunnari* (cgun), *Chaenocephalus aceratus* (cace), *Pseudochaenichthys georgianus* (pgeo), *Gymnodraco acuticeps* (gacu), *Trematomus loennbergii* (tloe), *Trematomus bernacchii* (tber), and *Eleginops maclovinus* (emac). The cladogram on the left shows the phylogenetic relationship among the species. The top and bottom vertical hatches on the cladogram represent the locations of Crynotothenia and Channichthyidae, respectively. The colored bars represent proportions across five different repeat classes (Unc. = Unclassified). The repeat distributions for *C. gunnari* and *C. esox* were generated in this study. We used the published repeat distributions for the remaining species: *E. maclovinus* (Cheng et al., *in prep*), *T. loennbergii* (Jo et al. 2021), *C. aceratus* (Kim et al. 2019), *P. georgianus*, *G. acuticeps*, and *T. bernacchii* (Bista et al. 2022). (*B*) Relative distribution of interspersed repeats along the *C. esox* genome. Chromosomes are ordered according to their size, largest to smallest. The color gradient represents the relative repeat content, plotted for each 250 kb window along the genome. Repeats appear to have a higher distribution along the terminal ends of chromosomes, except for chromosome 21 (highlighted by the triangle) which shows evidence of genomic rearrangements in the lineage. (*C*) Relative distribution of interspersed repeats along the *C. gunnari* genome. Chromosomes are ordered according to the length of the orthologous chromosome in *C. esox*. The color gradient represents the relative repeat plotted for each 250 kb window. As in *C. esox*, higher repeat distribution is observed in the terminal ends of chromosomes.

selection, as measured by *XP-EHH* and genetic divergence. This approach found 1,490 protein-coding genes, hereafter referred to as "divergence/*XP-EHH* candidates" (supplementary table S6, Supplementary Material online). Although not all these genes have substitutions in their coding sequences, we corroborated the presence of nearby SNPs with outlier *EHH* values, which may reflect cis-regulatory selection (supplementary fig. S5, Supplementary Material online).

The two analyses yielded 1,744 genes under selection representing candidates for secondarily temperate adaptation in *C. esox*. Assessing the functional annotation of these candidates, obtained during genome annotation by comparing the gene sequences to the *InterProScan* database (Quevillon et al. 2005), we observed genes associated with mitochondrial morphology and maintenance, with the mitochondrial electron transport chain, with light sensing and vision, and with the heat shock response (table 2). A complete list of all identified candidate genes and their predicted function based on their zebrafish orthologs is presented in supplementary table S7, Supplementary Material online.

## Expanded and Contracted Gene Families

In addition to specific genes exhibiting signatures of positive or directional selection in *C. esox*, we identified gene families showing elevated rates of evolution based on the gain and loss of genes within the family. Across seven notothenioids species—six cryonotothenioids including *C. esox* and a temperate outgroup (see Materials and Methods), the *OrthoFinder* software (Emms and Kelly 2015, 2019) identified 26,838 phylogenetic hierarchical orthogroups or gene families. *CAFE* v5 (Mendes et al. 2021) retained 21,019 families that were present at the root of the phylogeny of the sampled species and could be used to infer gene family expansion and contraction events. Out of these 21K gene families, 6,011 exhibit significant events of expansions and/or contractions across one or several points of the species phylogeny (supplementary fig. S6, Supplementary Material online). In the temperate pike icefish genome in particular, we observe 495 gene family expansion and 1,080 contraction events (1,575 total). Among the significant expanded/contracted gene families in *C. esox* (supplementary table S8, Supplementary Material online), we observe the contraction of *zona pellucida*, odorant receptor, and PR/SET domain-containing genes, as well as the expansion of several myosin H families, histone methyltransferases, and BED zinc finger protein genes.

## Evolution of Antifreeze Glycoproteins

The evolution of AFGPs has been regarded as the key evolutionary innovation that allowed notothenioids to survive in the inhospitable Antarctic environment. The cryonotothenioid *AFGP* genotype evolved once (Chen et al. 1997), and its genomic location and gene neighborhood appear to be conserved in a "canonical" *AFGP/TLP* locus across characterized species (Nicodemus-Johnson et al. 2011; Kim et al. 2019; Bista et al. 2022). Given the secondary escape of the *C. esox* lineage into a temperate environment, studying its *AFGP* genotype contributes to understanding gene evolution following the loss of a selective constraint. Even after the temperate transition, *C. esox* displays conservation in the organization of the canonical *AFGP/TLP* locus flanked by *hsl* and *tomm40*, on

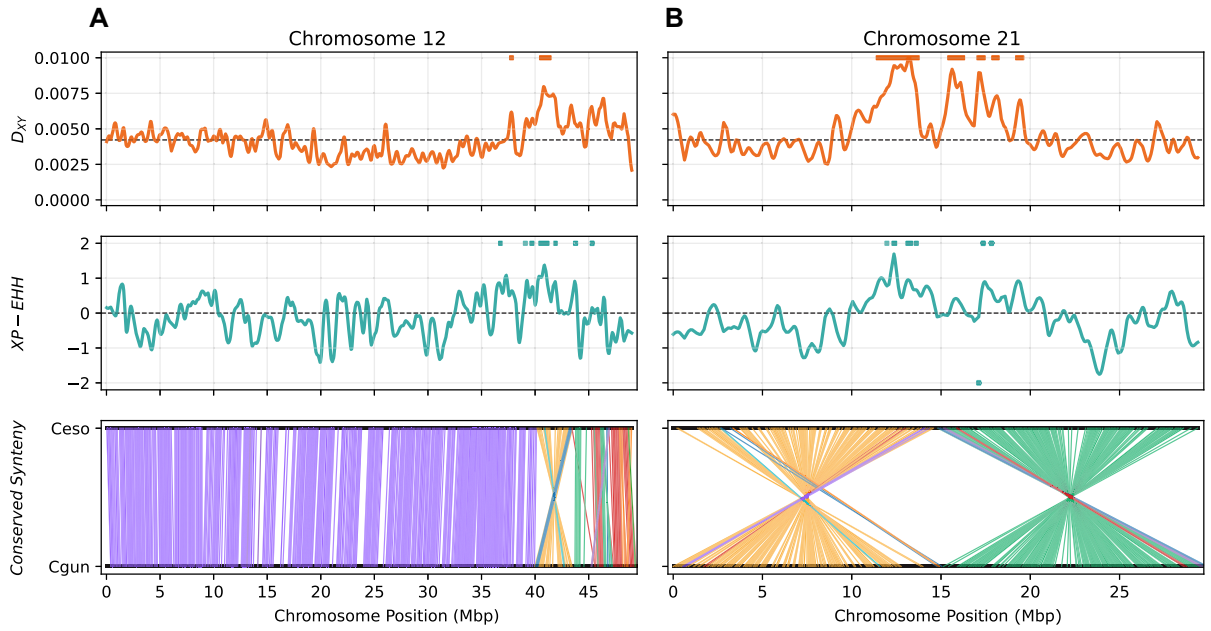


**Fig. 4.** Patterns of genetic divergence and positive selection in populations of temperate and Antarctic icefish species. (*A*) Population-level comparisons for *Champsocephalus esox* chromosome 12. Top panel shows the distribution of absolute genetic divergence ($D_{XY}$) between *C. esox* and *Champsocephalus gunnari* populations. Solid line shows the kernel-smoothed $D_{XY}$ values. Top horizontal bars indicate genomic windows displaying outlier genomic divergence ($P < 5 \times 10^{-4}$). Middle panel shows the distribution of *XP-EHH* values between *C. esox* and *C. gunnari* populations. Positive values indicate positive selection in *C. esox*, negative values selection in *C. gunnari*. Solid line shows the kernel-smoothed *XP-EHH*. Horizontal bars indicate variant sites displaying outlier *XP-EHH* values (FDR < 5%). Bottom panel shows patterns of conserved synteny between the *C. esox* and *C. gunnari* orthologous chromosome. Each line represents an orthologous gene, color coded according to its conserved synteny cluster. In chromosome 12, we can observe a region (at 40–45 Mb) displaying outlier divergence and *XP-EHH* values coinciding with a chromosomal inversion. (*B*) Population-level comparisons for *C. esox* chromosome 21. Top, middle, and bottom panels show $D_{XY}$, *XP-EHH*, and conserved synteny, as described in (*A*). This chromosome displays two large chromosomal inversions (one from 0 to15 Mb, second one from 15 to 30 Mb), each co-localized with regions with outlier $D_{XY}$ and *XP-EHH* values.

its 5′ and 3′ ends, respectively, as in other cryonotothenioids (fig. 5). The locus contains five tandem duplicates of *tryp1*, one copy each of *tryp3* and *TLP*, and six *AFGP* copies. Five of the *AFGP* copies appear nonfunctional (denoted with Ψ, fig. 5). Two of these are partial copies composed of only exon 2, containing the coding sequence of the repetitive AFGP tripeptides $(\text{Thr-Ala/Pro-Ala})_n$. The missing exon 1 encodes the signal peptide, thus these partial copies lack an appropriate translation start and the signal sequence for extracellular export of the protein. The other three *AFGP* copies exhibit inactivating frameshift mutations and/or premature stop codons. This high number of pseudogenized *AFGP* copies (five of six) likely reflects a history of relaxed selective pressure following secondary colonization of temperate habitats in this lineage. The canonical *AFGP*/*TLP* locus in the Antarctic sister species *C. gunnari* exhibits the same conserved organization, with variation in the number of copies and direction of several genes. There are eight tandemly duplicated *tryp1* copies, typical of other notothenioids including the basal *E. maclovinus*, versus five in *C. esox*. Three functional *AFGP* copies are observed in the canonical *AFGP*/*TLP* locus of *C. gunnari* along with one additional, nonfunctional copy.

Of the six notothenioids we examined, excluding the non-Antarctic outgroup *E. maclovinus* (which lacks *AFGPs*), *C. gunnari* and *C. esox* had the fewest *AFGP* copies. However, in the genomes of both species we found two complete and (at a sequence-level) putatively functional *AFGP* copies about 8 Mb away from the canonical locus on the same chromosome. In both *C. esox* and *C. gunnari*, these additional copies are flanked by *zgc:193726* (193726 zebrafish transcript homolog) and *bbs9* (Bardet–Biedl syndrome 9), on the 5′ and 3′ ends, respectively. These two additional *AFGP* copies have not been identified in any notothenioid species studied to date and, in combination with their conserved synteny across both genomes, might reflect a translocation forming a noncanonical *AFGP* locus unique to the genus *Champsocephalus*.

## Discussion

In this study, we describe two high-quality, chromosome-scale genome assemblies for a pair of sister icefish species, the temperate *C. esox* and the Antarctic *C. gunnari*. This species pair represents the only example of major ecological divergence in the Channichthyidae—two closely related species occurring in disparate environments across the Antarctic Polar Front. Their close evolutionary relationship presents a unique opportunity for studying the genetic basis of secondary adaptation to warm environments, which would otherwise be lethal to the cold-specialized cryonotothenioids. The hemoglobin-less phenotype precluding

**Table 2:.** Examples of Biological Categories Displaying top Candidates Under Selection.

| Category | Gene name | Gene description | Type[a] |
|---|---|---|---|
| Heat Shock Response | *dnajb6b* | DnaJ heat shock protein family (Hsp40) member B6b | w |
| | *dnajc5b* | DnaJ (Hsp40) homolog, subfamily C, member 5 beta | w |
| | *dnajc5ga* | DnaJ (Hsp40) homolog, subfamily C, member 5 gamma a | w |
| | *hikeshi* | heat shock protein nuclear import factor hikeshi | w |
| | *hspbap1* | hspb associated protein 1 | w |
| | *serpinh1b* | serpin peptidase inhibitor, clade H (heat shock protein 47), member 1b | d |
| | *zgc:122979* | zgc:122979 (DnaJb5 ortholog) | w |
| Mitochondria | *cox20* | cytochrome c oxidase assembly factor COX20 | w |
| | *cox5b* | cytochrome c oxidase subunit 5b | w |
| | *cox7a1* | cytochrome c oxidase subunit 7A1 | w |
| | *cox7a2b* | cytochrome c oxidase subunit 7A2b | w |
| | *gdap1* | ganglioside induced differentiation associated protein 1 | w |
| | *mfn1a* | Mitofusin 1a | w |
| | *rpl37* | ribosomal protein L37 | w |
| | *sdhaf3* | succinate dehydrogenase complex assembly factor 3 | w |
| | *trak1a* | trafficking protein, kinesin binding 1a | d |
| | *uqcc1* | ubiquinol-cytochrome c reductase complex assembly factor 1 | w |
| Vision | *cryba4* | crystallin, beta A4 | w |
| | *crybb1* | crystallin, beta B1 | w |
| | *gnat2* | guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 2 | d |
| | *ift172* | intraflagellar transport 172 | w |
| | *ift80* | intraflagellar transport 80 homolog | d |
| | *impg1a* | interphotoreceptor matrix proteoglycan 1a | w |
| | *irbpl* | interphotoreceptor retinoid-binding protein like | w |
| | *rbp3* | retinol binding protein 3 | w |
| | *rdh10a* | retinol dehydrogenase 10a | w |

[a]Candidate type: w = Divergence/*XP-EHH* windows, *d* = dN/dS.

active transport of oxygen in icefishes makes them particularly vulnerable to warm temperatures, in turn making *C. esox* an even more remarkable example of adaptation and diversification from a cold-specialized and stenothermal ancestor.

## Conservation in Genome Structure Despite Ecological Disparity

We observed a high degree of conservation in genome structure and organization in the temperate *C. esox* and Antarctic *C. gunnari* genomes, despite drastic ecological disparity and specialization. Both genomes show conservation in architecture and chromosome number, including a strong one-to-one correspondence across orthologous chromosomes. These results are in agreement with the $2n = 48$ karyotype previously characterized in *C. gunnari* (Ozouf-Costaz et al. 1996) and the patterns of conservation in chromosome number and organization described in other cryonotothenioid genome assemblies (Kim et al. 2019; Bista et al. 2020; Bista et al. 2022). Although changes in chromosome number have been observed in cryonotothenioids (Ozouf-Costaz et al. 1991; Morescalchi et al. 1992; Ghigliotti et al. 2015)—some of them quite extreme, as seen in the genus *Notothenia* (Amores et al. 2017)—the ancestral teleost karyotype number ($N = 24$) is largely conserved. However, other changes to genome organization, like changes in karyotype structure, have been extensively observed (Mazzei et al. 2006; Ghigliotti et al. 2007; Ghigliotti et al. 2015), giving precedence to the patterns of intrachromosomal changes observed between *C. esox* and *C. gunnari*.

The distribution of transposable repeat elements is also conserved between temperate and Antarctic icefish genomes. This distribution has long been thought to be important to genome evolution and lineage-specific diversity in vertebrates (Warren et al. 2015). Genomic repeats comprise a large portion of cryonotothenioid genomes (Auvinet et al. 2018; Kim et al. 2019; Bista et al. 2022), with a lineage-specific expansion in transposable elements likely linked to environmental stressors in the extreme Antarctic environment. Note that, although the percentage of the genome composed of transposable elements is higher for the two *Champsocephalus* species when compared with other published notothenioid assemblies (fig. 3), this is likely a technical artifact explained by methodological differences. For example, when we recomputed the repeat annotation for *Chaenocephalus aceratus* (Kim et al. 2019) using the analyses described in this study (see Materials and Methods), the repeat distribution increased to that observed in the two *Champsocephalus* assemblies (supplementary fig. S7, Supplementary Material online). This technical issue aside, given the known link between environmental stress and the expansion of transposable elements (Oliver and Greene 2009; Lanciano and Mirouze 2018), it is striking that no substantive change in the transposable element distribution is observed in *C. esox*. The transition to a temperate

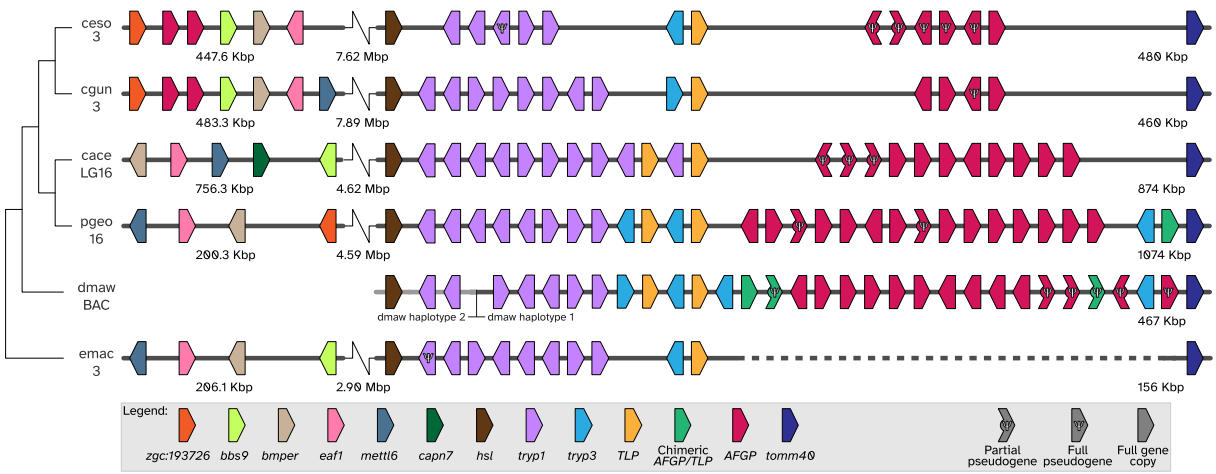


**Fig. 5.** Evolution of the antifreeze glycoprotein locus. Conserved synteny in the *AFGP/TLP* locus for six notothenioids (ceso, *Champsocephalus esox*; cgun, *Champsocephalus gunnari*; cace, *Champsocephalus aceratus*; pgeo, *Pseudochaenichthys georgianus*; dmaw, *Dissostichus mawsoni*; emac, *Eleginops maclovinus*). Cladogram on the left shows the phylogenetic relationship across the six species. The ID for the chromosome containing the *AFGP* is displayed under the species ID. Each horizontal line represents the span of the chromosome, arranged from 5′ to 3′ orientation, with the diagonal line representing a break in the represented sequence. For *D. mawsoni*, the sequence represents the bacterial artificial chromosome (BAC) sequence assembled in (Nicodemus-Johnson et al. 2011), with the majority of the sequence (from the third *tryp1* copy to *tomm40*) corresponding to haplotype 1. The horizontal light gray line on the 5′ end of this sequence, from *hsl* to the second *tryp1* copy, corresponds to the 5′ end boundary of the *D. mawsoni AFGP/TLP* haplotype 2. Each colored arrow represents a different gene, color-coded according to their corresponding gene family, and arranged according to their relative position in the annotated sequence. Distances between genes are not drawn to scale. Pentagonal arrows represent full gene copies, whereas truncated arrows represent partial copies. The Greek letter Ψ denotes pseudogenes. Both *C. esox* and *C. gunnari* have relative conservation in the organization of the canonical *AFGP/TLP* locus (flanked by the genes *hsl* and *tomm40*) when compared with other Antarctic notothenioid fishes, including both white-blooded (*C. aceratus*, *P. georgianus*) and red-blooded (*D. mawsoni*) species, but display a reduction in *AFGP* gene copies in the canonical locus. *Champsocephalus esox* additionally shows an increase in pseudogenized *AFGP* copies, with 5 out of the 6 genes appearing nonfunctional. Both *Champsocephalus* species display a small non-canonical locus with two complete *AFGP* copies in the same chromosome but about 8 Mb away from the canonical locus. Gene synteny in this noncanonical locus is highly conserved between *C. gunnari* and *C. esox*, with both *AFGP* copies flanked by the genes *zgc:193726* and *bbs9*. Although the majority of flanking genes (e.g., *bbs9*, *bmper*, *eaf1*) are observed across all six species, changes in gene order and orientation are observed between both *Champsocephalus* species and other notothenioids, suggesting ancestral rearrangements in this section of the chromosome.

environment following extreme cold specialization, both in terms of novel environmental stressors and possible demographic changes, could have otherwise been predicted as a prime driver for transposable element expansion and diversification. However, the relatively recent divergence time between *C. esox* and the Antarctic icefish lineage (~1.6 My; Stankovic et al. 2002; Near et al. 2012; Dornburg et al. 2017) may not be sufficient for extensive diversification, resulting in a genome that still retains many of its ancestral Antarctic properties and organization. This highlights the need for the study of other secondarily temperate notothenioid genomes, particularly for species exhibiting more ancient polar-to-temperate transitions, to determine if the patterns described above are specific to *C. esox* or shared across independent polar-to-temperate transition events.

## The Genomic Architecture of Secondarily Temperate Adaptation

Colonization of temperate Patagonian waters by the ancestral *C. esox* lineage likely occurred over the last 2 My (Stankovic et al. 2002), matching the Quaternary colonization of temperate environments observed in other red-blooded notothenioid species (Hüne et al. 2015; Ceballos et al. 2019). Given this relatively recent window of time, it is remarkable to find a complex architecture underlying polar-to-temperate transitions in this system, as it would require the organization and interaction of many individual loci for adaptation to take place. We identified divergence outlier loci present in most *C. esox* chromosomes (supplementary fig. S3, Supplementary Material online), indicating high sequence differentiation between temperate and Antarctic *Champsocephalus* populations. This fact alone is unsurprising since these two species are separated by the Antarctic Polar Front, which serves as a strong barrier for gene flow between population residing in the waters of Patagonia and the Southern Ocean (Poulin et al. 2014; Moore et al. 2018). However, neutral processes cannot fully account for the divergence as many of these regions also house candidate genes under selection, signaling that changes in selective pressures during the polar-to-temperate transition acted across the whole genome and, in part, mediated the observed patterns of divergence.

Genomic rearrangements are likely playing an important role in shaping the genomic architecture underlying this adaptive process. We identified several examples of

genomic rearrangements between the *C. esox* and *C. gunnari* genomes co-localized with the candidate regions described above (fig. 4). Structural variations, particularly chromosomal inversions, are mediators of genomic processes underlying adaptation, preventing recombination between inverted chromosomes, and maintaining divergent adaptive haplotypes (reviewed in Kirkpatrick and Barton [2006] and Wellenreuther et al. [2019]). Chromosomal inversions have played a role in several adaptive processes in wild populations, including thermal adaptation in *Drosophila* (Giribets et al. 2019), local climatic adaptation in seaweed flies (Mérot et al. 2021), the maintenance of social morphs in *Formica* ants (Brelsford et al. 2020), differentiation between migratory and nonmigratory Atlantic cod populations (Matschiner et al. 2022), and divergence between coat color and tail length in deer mice ecotypes (Hager et al. 2022). Observing these candidates of selection within inversions implies that these structural rearrangements are acting (or acted in the past) as islands of divergence, allowing for the maintenance and segregation between cold- and temperate-adapted haplotypes in the emerging *C. esox* populations after their isolation from cold-specialized populations. This would suggest that the structural variants, and the regions contained within that gave rise to temperate-adapted haplotypes, may have existed in the Antarctic icefish population prior to speciation. These two factors of the genomic architecture—structural variants co-located with signatures of selection alongside high baseline differentiation in the absence of gene flow—illustrates a contrast between early and late evolutionary forces. Instead of the classic pattern of divergence with gene flow, that is a handful of regions in the genome displaying elevated patterns of divergence surrounded by a low divergence background (Cruickshank and Hahn 2014), based on our data, we hypothesize that early adaptive divergence was facilitated by structural variants, after which neutral variation independently accumulated between Antarctic and temperate populations following isolation due to geographic and environmental barriers. Following this early isolation and the processes of temperate adaptation, we may be presently observing the remnants of these ancient islands of divergence on this now differentially adapted and fully speciated icefish populations.

The key role of inversions as islands of divergence is primarily observed in other systems as polymorphic variants in a single population or multiple populations in sympatry or with some degree of gene flow. Although our reconstructions of single genomes cannot show if these inversions were (or still are) polymorphic in either *C. gunnari* or *C. esox*, the overlap in the genome between structural variation and candidates of selection suggests a possible link between the two. Whether they served as mechanisms of isolation in the emerging populations or facilitated the maintenance of novel adaptive haplotypes following speciation, these chromosomal inversions are a key component of the genomic architecture of secondarily temperate adaptation and speciation in *C. esox*. In order to fully understand their role, however, future studies should focus on determining both the polymorphic status of these rearrangements, as well as determining their age in relation to the time of speciation between *C. esox* and *C. gunnari*.

## Physiological and Molecular Changes in a Polar-to-temperate Transition

The process of adaptation to stable temperature environments has been experimentally shown to limit physiological plasticity in response to future changes to those temperatures (Morgan et al. 2022). Thus, the ability of a cold-specialized icefish species to adapt to temperate environments is a remarkable example of adaptation following specialization. Given the candidate genes, we identified in *C. esox*, we can observe how secondarily temperate adaptation has impacted an already unique icefish physiology.

## Antifreeze Glycoproteins

Previous work detected the presence of AFGP coding sequences in *C. esox* (Miya et al. 2016), but it was unknown if they were functional. Our genome assemblies demonstrate that, in the absence of chronic cold, the *AFGP* copies in *C. esox* have followed a different evolutionary trajectory from those in the Antarctic notothenioids. A canonical *AFGP*/*TLP* locus remains in *C. esox* (fig. 5), which is not unexpected given the recent age of divergence from its Antarctic ancestor. Yet, although similar in copy number, the majority of *C. esox AFGP* coding sequences exhibit evidence of pseudogenization, showing a combination of exon loss, premature stop codons, or frameshift mutations that would prevent the production of AFGP polyproteins. Although pseudogenization of *AFGP* copies has occurred in both red- and white-blooded Antarctic species (fig. 5), the particularly high frequency in *C. esox* likely resulted from the absence of freezing-related selective pressures in their temperate environment. Interestingly, the sequence of three *AFGP* copies appears to remain complete and intact (one in the canonical locus, and two in the novel, translocated locus). However, our preliminary results indicate absence of functional AFGPs in *C. esox* (unpublished), and the mechanisms of AFGP trait loss in this system remain under investigation.

In notothenioids, the AFGP gene is known to have evolved once from a trypsinogen-like protease (*TLP*) precursor (Chen et al. 1997), and the extant gene family has a conserved gene neighborhood across different cryonotothenioid families (Nicodemus-Johnson et al. 2011; Kim et al. 2019; Bista et al. 2022). Variation in copy number appears to evolve via tandem duplications within the *AFGP*/*TLP* locus. Strikingly, our results show that both *C. esox* and *C. gunnari* have two additional *AFGP* copies translocated outside the canonical site (fig. 5), a feature not observed in all characterized notothenioids to date. This noncanonical *AFGP* locus displays conserved synteny between the two species (and the locus placement is more broadly supported by conserved synteny of the surrounding gene

neighborhood of more basal taxa), suggesting that the translocation predates speciation and originated in the Antarctic ancestor of both *C. esox* and *C. gunnari*. Interestingly, we find evidence of additional, small-scale rearrangements in the chromosome containing the canonical and novel *AFGP* loci, both when comparing *C. gunnari* to *C. esox* and to other species (fig. 5, supplementary fig. S8, Supplementary Material online). Although none of these changes disrupt the *AFGP*/*TLP* locus, their presence suggests chromosomal volatility. Previous studies have described the presence of polymorphic *AFGP* loci in cryonotothenioids, with different haplotypes displaying copy number variation in *AFGP* and surrounding genes (Nicodemus-Johnson et al. 2011). Therefore, future studies on this noncanonical *AFGP* locus should explore the presence of polymorphic *AFGP* haplotypes in *Champsocephalus*, particularly with respect to these additional copies. For example, determining if the noncanonical locus is polymorphic itself or if it is linked to a specific *AFGP*/*TLP* haplotype.

## Mitochondria

White-blooded icefishes display divergent mitochondrial form and function when compared with other red-blooded notothenioid species (O'Brien and Mueller 2010), arising from evolutionary changes of the organelle to compensate at least for the loss of hemoglobin (Sidell and O'Brien 2006; Bargelloni et al. 2019). Previous studies described *C. esox* as possessing higher mitochondrial density than other Antarctic icefish (Johnston et al. 1998) and having cristae (i.e., inner membrane) morphology more similar to that of red-blooded notothenioids (Johnston et al. 1998; O'Brien and Mueller 2010). In addition, recent analysis of the *C. esox* mitochondrial genome identified unique patterns of gene duplication and the likely presence of heteroplasmy in the species (Minhas et al. 2022). Correlating with these previously described changes, we found several candidate genes under selection that are related to mitochondrial morphology, potentially mediating the unique mitochondria observed in this species. The gene *mfn1a* encodes for a GTPase essential for mitochondrial fusion (Chen et al. 2003), whereas *gdap1* is known to affect several mitochondrial phenotypes including changes to fragmentation and elongation (Niemann et al. 2005). The interaction between the gene products of *mfn1a* and *gdap1* are known to maintain mitochondrial morphology and stability across the cellular mitochondrial network (Niemann et al. 2005). In addition, the gene *trak1a*, for which we found nonsynonymous substitutions under positive selection in *C. esox* (supplementary table S5, Supplementary Material online), plays a role in mitochondrial trafficking (Koutsopoulos et al. 2010). The trak1 protein has been found to colocalize with mitofusin proteins, like those encoded by *mfn1a*, putatively facilitating mitochondrial fusion (Lee et al. 2018).

Aside from mitochondrial morphology, we identified other candidate genes under selection in *C. esox* related to cellular respiration and the organization of the electron transport chain. More specifically, we identified several genes that form or are associated with cytochrome *c* oxidase (COX), that is complex IV of the electron transport chain. Two genes that appeared within our divergence/*XP-EHH* windows (supplementary table S6, Supplementary Material online) are *cox7a1* and *cox7a2b*, two paralogs of COX subunit 7. Each paralog produces a different, tissue-specific protein isoform (reviewed in Sinkler et al. 2017), which, in mammals, appears to fine tune complex IV activity via the formation of different COX dimer complexes (Zong et al. 2018). Another gene within this set of candidates is *cox5b*, which encodes for a hypoxia-inducible isoform of COX subunit 5 (Hodge et al. 1989; Kwast et al. 1998). Although not a component of complex IV itself, *cox20* is another candidate of selection involved in the assembly and maturation of the COX supermolecule (Bourens et al. 2014). In addition to complex IV, we also identified other related genes as putative candidates of selection. For example, the ribosomal protein *rpl37* (located in one of the large inversions on chromosome 21) plays a role in the assembly of complexes I and III of the electron transport chain (reviewed in Soto et al. 2012). The gene *sdhaf3* is important for the maturation of complex II (Na et al. 2014), whereas *uqcc1* acts on the assembly of complex III (Brockerhoff et al. 2003). Although none of these candidate genes contain nonsynonymous substitutions that could be altering their canonical function, mutations over nearby noncoding elements could alter their regulation and expression patterns. The supramolecular organization of the oxidative phosphorylation molecular complexes is dynamic. Different molecular assemblages of the different electron transport chain complexes can be found across different tissue types, oxygen levels, and resource conditions (reviewed in Enríquez 2016). Therefore, temperate adaptation in *C. esox* may have acted upon the regulation of components of the electron transport chain, taking advantage of the variability in the configuration of these complexes to maintain or maximize metabolic output in its new warmer, less oxygenated environment.

## Circadian Rhythms and Vision

One prior study reported the loss of genes belonging to the canonical circadian regulation in cryonotothenioids (particularly, in the blackfin icefish *C. aceratus*), likely as a result of the unique polar light–dark regimes experienced by Antarctic organisms (Kim et al. 2019). We find that many of the same canonical circadian regulation genes are also apparently lost in *C. esox* (supplementary fig. S9, Supplementary Material online), including the absence of several paralogs of the period (e.g., *per1a*, *per2a*, *per3*) and cryptochrome circadian regulators (e.g., *cry1b*, *cry2*, *cry3b*, *cry4*). Many of these putative loses are in fact shared between *C. esox*, its Antarctic sister species *C. gunnari*, and other cryonotothenioids (supplementary fig. S9 and table S9, Supplementary Material online), implying that they are shared across the lineage and not a product of the

temperate adaptation process. The molecular and physiological changes resulting from these gene losses in icefish are presently unknown, as well as their interaction with other abiotic factors (e.g., light, temperature) that might be interacting with biological rhythms in these organisms. Particularly, given the complex patterns of loss across circadian regulator paralogs across teleosts (Toloza-Villalobos et al. 2015), these loses alone are likely not sufficient to indicate a disruption in the canonical circadian regulation in these organisms. We observe a single circadian-related candidate gene under selection in *C. esox*, a serotonin receptor (*htr7a*). Although this gene is not part of the canonical circadian regulation pathways, the neurotransmitter serotonin does exhibit circadian rhythmicity (Morin 1999), and might thus indicate some yet uncharacterized changes to circadian-associated pathways in *C. esox*.

Changes in light patterns can impact the biological clocks of organisms (Beale et al. 2013). *Champsocephalus esox* originated from a cold-specialized ancestor that might have been subjected to polar regimes of extended months of light and dark periods in the Antarctic. It is important to note that *C. gunnari* has an extensive range and can be found on both the Atlantic and Indian sectors of the Southern Ocean (Kock and Everson 2003). Within the Atlantic sector, *C. gunnari* can be found along a 10°–15° longitudinal range, from the West Antarctic Peninsula in the south, to the Scotia Sea in the north (fig. 1). South Georgia island, located at the northernmost point of this distribution, is at a similar latitude to that of Patagonia, and thus would have similar periods of daylight (Smith and Walton 1975). This creates environmental differences among these different *C. gunnari* populations, which likely include differences in daily light levels. In turn, these different *C. gunnari* populations might experience different selective pressures related to photoperiod and circadian rhythms. However, although the patterns of population structure and isolation among these *C. gunnari* populations has been studied (Kock and Everson 2003; Damerau et al. 2014; Young et al. 2015), their evolutionary relationship to extant *C. esox* populations remain unknown.

Although the light environment *C. esox* experiences may or may not resemble that of other closely related Antarctic icefishes, we do find evidence of selection acting on genes related to light sensing and vision. Among our candidates in divergence/*XP-EHH* windows (supplementary table S6, Supplementary Material online) we found two crystallin genes, *cryba4* and *crybb1*, both encoding for β-crystallin proteins. Crystallins are integral structural components of the vertebrate eye lenses (reviewed in Wistow et al. 2005; Wistow 2012), whose development is mediated by the complex regulation of crystallin genes across cell types and developmental stages (Farnsworth et al. 2021). Compared with other teleosts, Antarctic notothenioids possess cold-stable crystallin proteins that remain transparent at subzero temperatures (Kiss et al. 2004). It is unknown how the lens of temperate notothenioids compares with the Antarctic species. However, changes in the regulation of *cryba4* and *crybb1* in *C. esox* might allow for the development of functional lenses at higher temperatures, counteracting the effects of the cold-specialized phenotypes observed in other notothenioids.

In addition to structural components of the lens, we observed several candidates under selection related to visual processes. One example among the d*N*/d*S* candidates (supplementary table S5, Supplementary Material online) is *gnat2*, which encodes the alpha subunit (GNAT2) of a cone-specific transducin (Morris and Fong 1993), a heterotrimeric G protein coupled to the photoreceptor in the visual phototransduction cascade (reviewed in Arshavsky et al. 2002). Transducin is activated by a light-activated opsin, whereby its GNAT2 subunit binds a GTP, dissociates from the heterotrimer and activates the next component—a phosphodiesterase. The codon in *gnat2* under selection in *C. esox* is a Thr, which substitutes a Gly that is conserved across most of the background notothenioid group (supplementary fig. S10, Supplementary Material online). This substitution suggests a novel protein phenotype present in the secondarily temperate *C. esox*; however, the functional outcome of this substitution in the signaling activity of *C. esox* GNAT2 awaits further exploration. Two genes among our divergence/*XP-EHH* window candidates are *irbpl* and *rbp3*, both interphotoreceptor binding protein homologs that are important for the transport of visual pigments in the light cycle (Jin et al. 2009; Parker et al. 2011). Another candidate, *impg1a*, encodes a proteoglycan protein part of the interphotoreceptor matrix, an extracellular structure that plays a role in the maintenance of photoreceptor cells and cellular adhesion to the retinal pigmented epithelium (Ishikawa et al. 2015). In *Astyanax* cavefish, *impg1a* is among the differentially expressed genes associated with loss of vision (Stahl and Gross 2017), although the functional role of *impg1a* on this phenotype is unknown. Additionally, we observed two intraflagellar transport genes, *ift172* and *ift80*, as candidates of selection. Intraflagellar transport proteins are essential for flagellar assembly, which then has an ubiquitous role in the transport of a variety of biomolecules (reviewed in Rosenbaum and Witman 2002). Both *ift172* and *ift80* have been specifically described as important for the transport of opsin molecules in zebrafish (Sukumaran and Perkins 2009; Hudak et al. 2010). Lastly, another gene located in the chromosome 21 inversion and among our list of divergence/*XP-EHH* window candidates is *rdh10a*. In mice, the enzyme encoded by this gene plays a role in the regeneration of 11-*cis*-retinal, an important reaction in the recycling of retinoids during the visual cycle (Sahu et al. 2015).

Our candidate genes do not intrinsically indicate differences in the circadian rhythm and biological periodicity between *C. esox* and other cryonotothenioid species. However, we find evidence for selection in the maintenance, transport, and activation of visual pigments and tissue. Although the genetic differences described above could simply be related to sensing in its light environment, detection of light alone can be important for the

entrainment of the internal biological clock of the organism to its environment (Hunter-Ensor et al. 1996; Hurd and Cahill 2002). The presence and detection of light plays a key role upstream of the core circadian rhythm molecular feedback loop in zebrafish (Vatine et al. 2011). Therefore, changes in light detection, as a result of selection in these vision-related genes, could lead to them playing a larger role in the entrainment of biological rhythms in the temperate pike icefish.

### Heat Shock Response and Molecular Chaperones

Associated with cold specialization in cryonotothenioids is their inability to mount the classic, inducible heat shock response (HSR; Hofmann et al. 2000; Bilyk and Cheng 2014; Bilyk et al. 2018). Instead, increased constitutive expression of *hsp6a*, the major ancestrally inducible member of the HSP70 family, is observed across both red- and white-blooded Antarctic species, suggesting an increase in native expression to mitigate cold denaturation of proteins (Bilyk et al. 2021). In a studied icefish (*Chionodraco rastrospinosus*), exposure to thermal stress triggers an inflammatory response instead of the inducible HSR (Bilyk et al. 2018). *Champsocephalus esox* lives in an environment that can reach temperatures of ~14 °C in the austral summer, so biological pathways related to stress-induced chaperones and HSR represent prime candidates for selection. Among our divergence/*XP-EHH* window candidates (supplementary table S6, Supplementary Material online) we observed a suite of cytosolic molecular chaperones, among them several DnaJ/HSP40 paralogs, including *dnajc5b*, *dnaj5ga*, *dnajb6b*, and *zgc:122979* (a *dnajb5* homolog). DnaJ/HSP40 are a family of molecular co-chaperones that bind to and stabilize HSP70s and other chaperones (reviewed in Qiu et al. 2006). Furthermore, we also identified *hspbap1*, which encodes a HSP27-associated protein. HSP27s are a suite of small HSP molecular chaperones that have been associated with responses to oxidative stress (reviewed in Arrigo 2001). Another HSR-associated gene found in divergence/*XP-EHH* windows is *hikeshi*, a HSP70 nuclear import carrier important for HSR gene transcription and the reversal of stress-induced phenotypes (Kose et al. 2012). Lastly, we found evidence of nonsynonymous substitutions under selection in *serpinh1b*, which produces an inducible procollagen-specific HSP47. Unlike other molecular chaperones, HSP47 recognizes the folded state of its target, the collagen triple helix (Widmer et al. 2012). Expression of *serpinh1b* has been associated with thermal stress response in the rainbow trout (Wang et al. 2016; Pandey et al. 2021). Although most of these genes do not show coding changes under selection, mutations over nearby noncoding sites might be driving the evolution of HSR regulation. The state of inducible HSR in *C. esox* is presently unknown; however, evidence of selection on these genes suggest possible changes in HSR pathways and other chaperones following colonization of temperate environments and are good candidates for future work regarding the functional characterization of secondarily temperate adaptation.

## Conclusion

Here, we describe new genomic resources for *C. esox*, the only non-Antarctic icefish, and its Antarctic congener, *C. gunnari*. We characterize the primary changes to the genomic architecture of this temperate icefish, including several major structural variants, and show where key signals of evolutionary forces, including measures of divergence, linkage, and genic selection co-localize with these physical, chromosomal changes. In addition, we identified genes displaying evidence of positive selection in the species, providing examples of major biological and cellular pathways likely underlying the physiological changes behind secondary adaptation to temperate environments. These lines of evidence paint a picture of how *C. esox* has evolved since its Antarctic departure, allowing for a polar-to-temperate transition in this group and illuminating an important evolutionary question of how adaptation can occur after specialization. Future studies should explore the genotypes and candidate genes identified here in more detail, measuring tissue-specific gene expression of these genes, characterizing the effect of any nonsynonymous mutations altering protein function, and complimenting these results with the physiological studies of live individuals. Nonetheless, the work presented here sets the stage for future projects to further research the mechanisms behind secondarily temperate adaptation in the cryonotothenioid system.

## Materials and Methods

### Specimens and Sampling

A total of 28 *C. esox* and 68 *C. gunnari* individuals were collected and used in this study (fig. 1). Twelve of the 28 *C. esox* were captured with a gill net from Canal Bárbara (fig. 1—CB), a channel connecting the Strait of Magellan to the Pacific Ocean, in April 2016. The other 16 were obtained by rod and reel or gill net in the Patagonia waters near Puerto Natales, Chile (fig. 1—PN) in 2008, 2009, 2016, and from December 2017 to early January 2018. *Champsocephalus gunnari* was collected from two Antarctic regions. Forty-eight of the 68 were collected from waters around South Georgia Island and Shag Rock Shelf (fig. 1—SG) by trawls from onboard the British Antarctic Survey chartered Falkland Islands F/V *Sil* (ZDLR1) during the South Georgia groundfish survey from January to February 2017. The other 20 individuals were collected from several coastal sites along the West Antarctic Peninsula (fig. 1—AP) using otter trawl from onboard the US R/V Laurence M. Gould during austral winters (July—August) of 2008 and 2014. Detailed metadata for the specimens collected and used for the various sequencing effort in this study are given in supplementary table S10, Supplementary Material online.

For genome sequencing, white muscles from a single male *C. gunnari* (collected in 2014 from the Gerlache Strait in West Antarctic Peninsula) were flash frozen with liquid nitrogen and stored at −80 °C until used for high molecular weight (HMW) DNA extraction. For *C. esox*, fresh liver from a single male (collected in January 2018 from Puerto Natales) was perfused to obtain isolated hepatocytes. Perfusion was carried out on ice with collagenase (Sigma) at 0.5 mg/mL in heparinized, $Ca^{+2}$-free notothenioid physiological buffer (260 mM NaCl, 5 mM KCl, 2.5 mM $MgCl_2$, 2.5 mM $NaHCO_3$, 2 mM $NaH_2PO_4$, 250 units heparin/mL, pH 8.4), following a published protocol (O'Grady et al. 1982). Cell density was assessed by counting with a hemacytometer, and appropriate volumes of the hepatocyte suspension were embedded in 1% agarose blocks using disposable plug molds (BioRad). Hepatocytes were thoroughly lysed in-agarose using 1% lithium dodecyl sulfate in notothenioid ringer (100 mM sodium phosphate buffer, pH 8.0, adjusted to 420 mOsm with NaCl). The blocks were then fully equilibrated with a preservation buffer (100 mM EDTA, 10 mM Tris, 0.2% *N*-laurylsarcosyl, pH 9.0) that permitted return transport at ambient temperature, and were subsequently kept at 4 °C until use. For RAD and RNA sequencing, tissue samples for both species were dissected, preserved in ice-cold or −20 °C 90% ethanol, and stored at −20 °C until use.

Specimen collection and sampling were conducted with permits granted by the Chilean Fisheries Service under Technical Memorandum P.INV N° 244-2016, the Government of South Georgia and the South Sandwich Islands Regulated Activity Permit 2016-026, and the University of Illinois at Urbana-Champaign IACUC approved protocols 07053 and 17148, for the respective field collection and sampling effort.

## HMW DNA and Long-read Genome Sequencing

HMW DNA of the single male *C. gunnari* was extracted from frozen white muscle of the single male using the Nanobind Tissue Big DNA kit (Circulomics) following manufacturer protocol. For *C. esox*, agarose blocks of embedded hepatocytes from the single male were treated with proteinase K (final concentration 800 µg/µl, 52 °C, 2 h). The treated blocks were then melted (65 °C, 20 min), and digested with 2 units of β-agarase (NEB) per block (1 h, 42 °C). The liberated HMW DNA was gently extracted once with phenol-chloroform (1:1) and precipitated with isopropanol. The recovered DNA was further purified using the Nanobind Tissue Big DNA Kit following manufacturer protocol starting at the DNA binding step. Quality check of HMW DNA included fluorometric (Qubit v.3; Invitrogen) quantification of concentration, spectrophotometric (Nanodrop One; Fisher Scientific) determination of purity, and HMW assessment using pulsed field electrophoresis (CHEF Mapper XA; BioRad), and Fragment Analyzer (Advanced Analytical).

Pacific Biosciences CLR library preparation and sequencing were carried out at the University of Oregon Genomics and Cell Characterization Core Facility. The HMW DNA was lightly sheared with Megaruptor (Diagenode) at 60 kb target length for library construction using PacBio SMRTbell Express Template Prep kit 2.0. The resulting library was selected for inserts approximately >30 kb with the BluePippin (Sage Science) and sequenced on two SMRT cells 8M for each species on Sequel II for 30 h of data capture. A total of 12.6 million and 10.7 million CLR reads were obtained for *C. esox* and *C. gunnari*, respectively. Statistics of sequence reads are given in supplementary table S11, Supplementary Material online.

## Hi-C Library and Sequencing

For scaffolding the genome assembly, Hi-C chromosome conformation capture libraries were constructed for the same individual of *C. esox* and *C. gunnari* using ethanol-preserved spleen and liver, respectively. Library constructions were performed by Phase Genomics, Inc. using its proprietary Proximo Hi-C kit that utilizes *Dpn*II as the restriction enzyme. The libraries were sequenced by Phase Genomics on an Illumina NovaSeq6000, which yielded 257.8 million and 161.2 million 2 × 150 bp reads for *C. esox* and *C. gunnari*, respectively.

## Transcriptome Library and Sequencing

RNAseq libraries for *C. esox* and *C. gunnari* were constructed and sequenced to provide transcript-level evidence for annotating protein-coding genes in the genome. RNA from a broad range of tissues (14 for *C. esox*, 19 for *C. gunnari*) derived from multiple individuals were isolated using Ultraspec II (Biotecx) or Trizol (Invitrogen) reagent, DNaseI treated, purified, and checked for RNA quality on Agilent Bioanalyzer. RNA from all tissues achieved a high RNA Quality Number, as shown in supplementary table S12, Supplementary Material online. The individual, tissue-specific RNA samples from each species were then pooled at 1 or 0.5 µg to form a final pool. The library was then constructed from 2 µg of the final pool using the TruSeq-stranded mRNAseq Sample Prep kit (Illumina). The two libraries were multiplexed and sequenced for 2 × 250 bp reads on one Illumina HiSeq2500 lane. Raw reads for *C. esox* were processed for quality using *Trimmomatic* v0.32 (Bolger et al. 2014), retaining 98.7 million read-pairs. For *C. gunnari*, raw reads were independently processed using *AfterQC* v0.9.6 (Chen et al. 2017), leading to 99.9 million read-pairs retained.

## RADseq Library and Sequencing

To genotype *C. esox* and *C. gunnari* individuals, a restriction site-associated DNA (RADseq) library was constructed. Genomic DNA from the 28 *C. esox* and 68 *C. gunnari* individuals were isolated from ethanol-preserved muscle or other tissues using standard tissue lysis and phenol-chloroform extraction protocol. DNA concentrations were measured using a Qubit fluorometer (Invitrogen). A single enzyme digest RADseq library was constructed following published protocols (Baird et al. 2008; Etter et al.

2011) with slight modifications. Briefly, the DNA of the 96 individuals was diluted to equimolar concentration (33.3 g/µl), and 1 µg (30 µl) of each sample was transferred to and digested in a 96-well plate with the restriction enzyme *SbfI*-HF (NEB). Custom P1 adapters (Hohenlohe et al. 2012), each containing a 7-bp unique barcode identifier, were then ligated to the digested DNA of each individual. The 96 barcoded samples were pooled at equimolar concentration for a total of 2 µg per pool to generate four replicate pools. Each pool was sheared using a Covaris M220 focused ultrasonicator, after which the fragmented DNA was size selected within a 300–600 bp range using AMPure XP beads (Beckman Coulter). The size-selected pools were then end-repaired and A-tailed, followed by ligation of P2 adapters, forming the library. To minimize PCR duplicates, replicate libraries were pooled to maximize the amount of template DNA amplified, pooling ~150 ng of template DNA per replicate. A 100 µl PCR reaction was used to amplify the samples for 12 PCR cycles. The final amplified library was cleaned with AMPure XP beads and quantified using Qubit fluorometry. The library was sequenced on an Illumina NovaSeq6000 SP 2 × 150 bp lane at the University of Illinois Roy J. Carver Biotechnology Center, producing 821 million raw reads.

## Genome Assembly

PacBio CLR raw reads were subsampled at different depths of coverage and size distributions (i.e., removing reads shorter or longer than a specified range of length). This subsampling of reads served as an assembly optimization process (Rayamajhi et al. 2022). See supplements for more details on the subsampling optimization. For *C. esox*, the highest quality assembly was derived from a subsample of reads 10–50 kb in length that provided 70× coverage, based on an initial genome size estimation of 1.1 Gb, yielding 3.8 million reads with a 20.4 kb mean length and 23 kb read N50. For *C. gunnari*, the most optimal subsample contained reads over 15 kb and 80× coverage, yielding a 29.5 kb mean length and a 32 kb N50. Contig-level assemblies were made using two different assemblers, *Flye* (Kolmogorov et al. 2019) and *wtdbg2* (Ruan and Li 2020). Each assembly was assessed for contiguity using *QUAST* v4.4 (Gurevich et al. 2013) and gene completeness using *BUSCO* v3.0.1 (Simão et al. 2015), using the `actinopterygii_odb9` lineage data set. For both species, *Flye* v2.5 generated superior contig-level assemblies based on quality and completeness scores (supplementary methods) and was selected as the primary assembly for downstream analyses. The *wtdbg2* v2.5 assembly is referred to as the secondary assembly and was used for comparison purposes (supplementary table S13, Supplementary Material online).

The Hi-C data were then integrated into both the primary and secondary assemblies to generate chromosome-level super-scaffolds using *Juicer* v1.6.2 (Durand et al. 2016) to identify the Hi-C junctions, and then with *Juicer*'s `3d-dna` program to complete the integration of scaffolds. In addition, conserved synteny analysis (described below) between primary and secondary assemblies was used to perform manual curation of the scaffolding process. For example, we employed manual insertions, inversions, or translocations of whole contigs when discrepancies were found between the primary and secondary assemblies, and evidence, such as contig or scaffold boundaries, that supported one assembly to be correct. We used a custom Python program to propagate these changes through the constituent assembly files, such as the structure (AGP), annotation (GFF), and sequence files (FASTA). Following scaffolding and manual curation, a reassessment of the contiguity of the final assembly was done using *QUAST* v4.4 and an assessment of gene completeness was performed with both *BUSCO* v3.0.1 and *BUSCO* v5.1.3 (Manni et al. 2021), using the `actinopterygii_odb9` and `actinopterygii_odb10` gene data sets, respectively.

## Genome Annotation

To annotate repeat elements in the genome, a *de novo* repeat library was built for each reference assembly separately using the *RepeatModeler* v2.0.2a pipeline (Flynn et al. 2020). First, a database was generated for each genome with `BuildDatabase`, using the NCBI database as input (`-engine ncbi`). The final library was then created using `RepeatModeler`, enabling the optional discovery of LTR elements using LTRharvest (`-LTRStruct`) (Ellinghaus et al. 2008). The *de novo*, species-specific repeat libraries were then combined with the teleost-specific repeat libraries available in *Repbase* release 27.02 (Bao et al. 2015). This combined repeat library was then used as the input of *RepeatMasker* v4.1.2-p1 (Smit et al. 2013), which was used for the final repeat annotation and masking of the genome assemblies.

For the annotation of protein-coding genes, we first indexed the masked reference genome for alignment using *STAR* v2.7.1.a (Dobin et al. 2013) `--runMode genomeGenerate`, after which we aligned the processed RNAseq reads to the chromosome-level genome assembly using `--runMode alignReads`. Final annotation of protein-coding genes was done using *TSEBRA* (Gabriel et al. 2021), which selects curated transcripts generated by the *BRAKER* annotation pipeline (Hoff et al. 2016; Brůna et al. 2021). For integration with *TSEBRA*, we ran *BRAKER* v2.1.6 in both protein and transcript modes. For protein mode (`--prot_seq`), we used the zebrafish protein sequences from *OrthoDB* v10.1 (Kriventseva et al. 2019), whereas the transcript mode (`--bam`) used the species-specific RNAseq alignments. The output of both *BRAKER* runs was then processed using *TSEBRA* v1.0.1 to retain the annotation of protein-coding genes supported at both the protein and transcript level. For both species, we verified the gene completeness of the annotation using *BUSCO* v5.1.3, providing the amino acid sequences of the annotated protein-coding genes as input, running in protein mode (`--mod prot`)

and comparing against the `actinopterygii_odb10` reference set. Finally, the functional annotation of the curated protein-coding genes was done using the *InterProScan* database (Quevillon et al. 2005).

For both *C. esox* and *C. gunnari*, *BRAKER* failed to annotate the AFGP genes. Thus, genomic coordinates and sequences for all *AFGP* copies were determined by manual annotation. We used `blastn` from *BLAST* + v2.4.0 (Camacho et al. 2009) to locate the coding sequences of both *AFGP* and other genes in the canonical *AFGP/TLP* locus (Nicodemus-Johnson et al. 2011), including trypsinogen 1 (*tryp1*), trypsinogen 3 (*tryp3*), trypsinogen-like protease (*TLP*), chimeric *AFGP/TLP* genes, and the two flanking genes, *hsl* (hormone sensitive lipase, *lipeb* homolog) and *tomm40* (translocase of outer mitochondrial membrane 40). For *AFGP* and *AFGP/TLP* sequences, we used the *Dissostichus mawsoni*-specific orthologs (Nicodemus-Johnson et al. 2011) as queries in the search. Coordinates of the `blastn` alignments were used to guide manual inspection and mapping of complete gene structures, including intron–exon boundaries, and the precise start and end of the highly repetitive tripeptide $(\text{Thr-Ala/Pro-Ala})_n$ *AFGP* coding sequences (Chen et al. 1997). The *AFGP* copies were then translated and manually inspected to evaluate the integrity of the gene, or whether mutations had occurred resulting in pseudogenization. We used comparative synteny (see Materials and Methods below) to assess conservation in the position of flanking genes surrounding the *AFGP/TLP* locus. We compared the annotation between both *C. esox* and *C. gunnari*, as well as *C. aceratus* (Kim et al. 2019) and *Pseudochaenichthys georgianus* (Bista et al. 2022), to confirm their location in orthologous chromosomes. In addition, we discovered a noncanonical *AFGP* locus during manual annotation (see Results), which we validated by applying a similar conserved synteny assessment of the surrounding genomic neighborhood between *C. esox* and *C. gunnari*.

Given both the repetitive nature and tandem duplication of *AFGP* copies in notothenioid genomes (Nicodemus-Johnson et al. 2011; Kim et al. 2019; Bista et al. 2022), which could lead to their misassembly, we further validated the regional assembly of the *AFGP/TLP* locus by tiling the raw CLR reads in this portion of the genome. Raw reads were aligned using *minimap2* v2.24-r1122 (Li 2018). In addition, we identified *AFGP*-specific k-mers in both the reference sequences and raw reads (supplementary methods). We visualized both the read alignments and presence of *AFGP* k-mer "clumps" (series of sequential k-mers) using a custom Python program, confirming that: (1) the read alignments spanned the totality of the corresponding *AFGP* locus, (2) no major disruptions in read alignments were present, and (3) that the location of *AFGP* clumps in the raw reads matched the location of *AFGP* sequences manually annotated in the reference sequences (supplementary figs. S11–S14, Supplementary Material online). Further validation of the noncanonical *AFGP* copies was done by inspecting for significant secondary/supplementary alignments between both *AFGP* loci, confirming that raw reads are uniquely mapping to their corresponding *AFGP* locus of origin.

## Conserved Synteny Analysis

An analysis of conserved synteny was performed at different stages of the assembly using the *Synolog* software (Catchen et al. 2009; Small et al. 2016). Annotated coding sequences from *C. esox* and *C. gunnari* primary and secondary assemblies, and other species of interest, were first reciprocally matched using the `blastp` algorithm from *BLAST* + v2.4.0. The *BLAST* results and genome annotation coordinates were fed to *Synolog*, which (1) establishes reciprocal best *BLAST* hits to identify orthologous genes, (2) uses the genome coordinates of each best hit to define clusters of conserved synteny between the genomes, (3) refines ortholog assignments using the defined clusters, and (4) defines orthology between chromosomes/scaffolds based on the conserved synteny patterns. In addition to the *C. esox* and *C. gunnari* genomes described here, for comparative purposes, we included other high-quality notothenioid assemblies including *C. aceratus* (Kim et al. 2019), *P. georgianus* (Bista et al. 2022), *Gymnodraco acuticeps* (Bista et al. 2022), and *Trematomus bernacchii* (Bista et al. 2022; supplementary table S2, Supplementary Material online).

This conserved synteny analysis was initially used to manually curate the assemblies, by identifying discrepancies between primary and secondary assemblies, and/or to identify structural variants located within scaffold boundaries indicating a misassembly. Once the curated chromosome-level sequences were generated, conserved synteny was determined between the genomes of *C. esox* and *C. gunnari* against the genome of *E. maclovinus*, the closest sister species to the Antarctic clade, in order to identify and name orthologous chromosomes. The *E. maclovinus* assembly (Cheng et al., in preparation) was first compared against the *Xiphophorus maculatus* reference assembly, which contains the ancestral teleost karyotype number of 24 chromosomes (Amores et al. 2014). Chromosomes were named accordingly to these patterns of conserved synteny.

## Molecular Evolution on Protein-coding Genes

To identify patterns of molecular evolution and positive selection on protein-coding genes, we first identified single-copy orthologous genes (orthogroups) with *Synolog*, as described above. A total of 14,093 orthogroups were identified across the six notothenioid species (i.e., *C. esox*, *C. gunnari*, *C. aceratus*, *P. georgianus*, *G. acuticeps*, and *T. bernacchii*). To guide alignments, we used the published notothenioid phylogenetic tree (Near et al. 2018), using the *ape* v5.5 R package (Paradis and Schliep 2019) to sample the tree to include only the taxa of interest (supplementary fig. S4, Supplementary Material online). The coding sequences for all genes in each orthogroup were aligned using *PRANK* v.170427 (Löytynoja and Goldman 2005), using the codon alignment option (`-codon`), a guide phylogenetic tree (`-t`), and forcing

insertions to be skipped (`-F`). Alignments were then filtered using *Gblocks* v 0.91b (Castresana 2000) under a codon model (`-t = c`). We used *GWideCodeML* v 1.1 (Macías et al. 2020), a wrapper for *PAML* v4 (Yang 2007), designed for large-scale genomic data sets, to identify genes under positive selection in *C. esox* based on the ratio ($\omega$) of nonsynonymous (d*N*) to synonymous substitutions (d*S*). *GWideCodeML* was run under the branch-site model (`-model BS`), supplying a species tree (`-tree`), defining *C. esox* as the foreground species (using the `-branch` file), and generating a d*N*/d*S* summary for all tested orthogroups (`-dnds`). We used a custom Python script to further parse the *GWideCodeML* outputs and obtain additional information needed to filter the results (e.g., site class proportions, model likelihoods, and per-site posterior probabilities). We then selected orthologs showing $\omega$ values >1, with alignment lengths greater than 150 amino acid sites, and showing significant ($P < 0.05$) differences in the likelihood ratio test between alternative and null models. For these selected genes, we also looked for individual sites showing posterior probabilities greater than 95% for *2a* or *2b* site classes. In addition, we filtered comparisons for which both the *2a* and *2b* site class proportions were zero, as these can result in unreliable $\omega$ estimates (Yang and dos Reis 2011). The gene ontology information for all the retained genes under selection was then categorized using the *PANTHER* database (Mi et al. 2021).

## Expansion and Contraction of Gene Families

We used the *CAFE* v5 (Mendes et al. 2021) software to identify evolutionary patterns of gene family expansion and contraction in *C. esox* relative to the other studied cryonotothenioid species. To do this, we first selected the annotated protein sequences for six cryonotothenioids (*C. esox*, *C. gunnari*, *C. aceratus*, *P. georgianus*, *G. acuticeps*, and *T. bernacchii*) and one temperate outgroup (*E. maclovinus*), which were then clustered into orthogroups using the *OrthoFinder* v2.5.4 software (Emms and Kelly 2015; Emms and Kelly 2019). When running the *OrthoFinder* pipeline, we first inferred trees based on multiple sequence alignments (`-M msa`). Alignments were made using *MAFFT* v7.310 (Katoh and Standley 2013), and the final gene trees inferred using *FastTree* v2.1.11 (Price et al. 2010). Once the orthogroup inference was complete, we used a custom Python script to calculate a per-species tally of the number of genes in each *OrthoFinder* phylogenetic hierarchical orthogroup. These tallied orthogroups, alongside the pruned phylogenetic tree from (Near et al. 2018), were then used as input for *CAFE* v5.0.0. As described in the *CAFE* documentation, we ran the software in three separate iterations in order to (1) estimate the global error model attributed to issues in genome assembly and annotation (`--error_model`) (Han et al. 2013); (2) calculate the mean rate of gene gain and loss ($\lambda$) accounting for the estimated error (`-e <error_file>`); (3) to perform an estimation of clade-wise gene family expansion and contraction using this fixed, error-adjusted $\lambda$ value (`--fixed_lambda`). Orthogroups with significant expansion and/or contraction events were determined according to the *P*-value reported by *CAFE* ($P < 0.05$).

## RADseq Data Processing

The 821.4 million paired-end reads from the sequenced library were processed with *Stacks* (Rochette et al. 2019). Raw reads were analyzed with `process_radtags` v2.54 to demultiplex samples, removing those with low quality (`--quality`), containing uncalled bases (`--clean`), lacking the *SbfI* cutsite (`--renz-1 sbfI`), containing adapters (`--adapter_1`, `--adapter_2`), and rescuing barcodes and cutsites (`--rescue`). After filtering, 737,838,374 reads (89.8%) were retained, with a mean 7.7 million reads retained per sample. The processed reads were then aligned to the *C. esox* reference assembly generated in this study with *BWA* mem v0.7.17-r1188 (Li 2013), using *SAMtools* v1.12 (Li et al. 2009) to process and sort alignments.

Following alignment, samples were genotyped using `gstacks` v2.59, removing PCR duplicate reads (`--rm-pcr-duplicates`). After removing 46.5% PCR duplicates, the library contained a nonredundant average coverage of 23.8×. A total of 9 samples were removed due to low coverage (<5×). The resulting 87 samples (24 for *C. esox* and 63 for *C. gunnari*) were processed using the populations module, defining the *C. esox* and *C. gunnari* samples as two distinct populations. To filter for missing data, we retained only loci present in the two populations (`-p 2`), in 80% of samples per population (`-r 80`), and variant sites with a minimum allele count of 3 (`--min-mac 3`). The final *Stacks* catalog kept 56,275 loci and 288,631 variant sites.

## Genetic Divergence and Signatures of Selection

To determine patterns of genetic divergence across the genome between *C. esox* and *C. gunnari*, the *Stacks* catalog described above was reanalyzed using `populations` v2.60. To account for the effects of genetic divergence between the two *C. gunnari* populations (see Results), the West Antarctic Peninsula individuals were removed from the analysis. The final comparisons for divergence and selection were made between 69 individuals, 24 *C. esox* (combined across Canal Bárbara and Puerto Natales) and 45 *C. gunnari* individuals from South Georgia Island. In addition to filtering for missing data, we calculated *F*-statistics between the populations (`--fstats`) and averaged the resulting statistics along the chromosomes using a kernel-smoothing algorithm (`--smooth`). The `populations` program was run twice using the parameters described above. First, using the `--bootstrap-archive` flag to generate an archive of randomly selected loci. Once the archive was generated, we applied bootstrap resampling 10,000 times over all kernel-smoothed statistics (`--bootstrap`, `--bootstrap-reps 10000`). Genomic windows displaying outlier values of genetic divergence between *C.*

*esox* and *C. gunnari* populations were identified based on the *P*-values from bootstrap replicates ($P < 5 \times 10^{-4}$).

In addition to genetic divergence, this RADseq data set was used to calculate *EHH* (Sabeti et al. 2002) to identify population-level patterns of positive selection. The *Stacks* catalog was again filtered using `populations`, this time increasing the stringency of filtering parameters to minimize the imputation of missing genotypes during phasing (`-r 90`, `-p 2`, `--min-mac 3`), and exporting the output as a sorted VCF (`--ordered-export`, `--vcf`). This stringent filtering still retained 53,464 loci and 239,008 variant sites. Before phasing, we split the *Stacks* VCF output into two species-specific VCFs using *VCFtools* v0.1.15 (Danecek et al. 2011) (`--keep`) in order to phase sites for each species separately. Each species-specific VCF was phased using *beagle* v5.3 (Browning and Browning 2007; Browning et al. 2018). We provided an initial effective population size of 5,000 (`ne = 5000`), based on the $N_e$ estimation for the related blackfin icefish (Kim et al. 2019). Population size and error parameters were estimated (`em = true`) by running for 12 iterations (`niterations = 12`). The phased VCF files were then indexed and merged using *BCFtools* v1.12 (Danecek et al. 2021). *EHH* calculations were performed using the *rehh* v3.2.1 *R* package (Gautier et al. 2017). First, the phased VCF was loaded using the `data2haplohh` () function, creating a subset of the data for each species, and filtering sites with a minimum allele frequency <5%. The integrated, site-specific *EHH* (*iES*; Tang et al. 2007) was calculated in each species using the `scan_hh` () function. To account for gaps of ungenotyped sequences in the genome in the RADseq data set, gaps were scaled (`scalegap`) to 20.3 kb (median distance between RAD loci in the data set), with a max gap (`maxgap`) distance of 104.5 kb (95$^{th}$ percentile of the interlocus distance) and keeping the per-site *EHH* integration in the boundaries of gaps (`discard_integration_at_border = FALSE`). Species-specific *iES* values were compared using the cross-population *EHH* statistic (*XP-EHH*; Sabeti et al. 2007) using the `ies2xpehh` () function, obtaining both *XP-EHH* and a significance value (−log10(*P*-value)) at each site. To account the effects of multiple testing, a Benjamini–Hochberg false discovery rate (FDR) correction was applied to the *XP-EHH* significance *P*-value using the `p.adjust` () function in *R* v4.0.1 (R Core Team 2020). Sites having corrected *P*-values lower than 0.05 (i.e., FDR < 5%) were classified as *XP-EHH* outliers.

We additionally extracted information on the identity and functional annotation of all protein-coding genes annotated within these divergence/*XP-EHH* outlier windows (450 kb from the variant site). To further validate the identified candidate genes, we identified the location and ID of the nearest outlier *XP-EHH* site and calculated site-specific *EHH* in order to observe the span of *EHH* decay of that SNP in relation to the location of the gene. Gene ontology information for these genes was then categorized using the *PANTHER* database (Mi et al. 2021) and overlaps were identified between the candidates of selection identified via molecular evolution analysis.

## Acknowledgments

This work was supported by the National Science Foundation (NSF DGE 10-69157 IGERT to A.G.R.C., NSF OPP Grant 1645087 to J.M.C. and C.H.C.C., and NSF ANT Grant 11-42158 to C.H.C.C.). The authors thank Kira M. Long, Alida de Flamingh, and Carlos Ortiz-Alvarado for comments and discussions on the manuscript, Francesco Zapelloni and Jacob Anderson for assisting with the DNA extractions used for RAD sequencing, Shehani Gunawardena and Eliana Eng for their assistance in the bioinformatic analyses, Ernesto Davis Seguic for assistance on location in Chile, and Arthur DeVries for assistance with field collection and sampling.

## Data Availability

The the raw PacBio CLR and Hi-C Illumina reads for the *C. esox* and *C. gunnari* genome assemblies are available on NCBI under BioProject PRJNA857989. Per-sample, raw RADseq reads are available on NCBI BioProject PRJNA857737. RNAseq reads for *C. gunnari* and *C. esox* are available on NCBI BioProject PRJNA859929. Copies of the bioinformatic scripts used for analysis, including custom software, are available on BitBucket (`bitbucket.org/angelgr2/icefish_genome_paper/`). An additional copy of the finalized assemblies and associated annotations for *C. esox*, *C. gunnari*, and the temperate outgroup *E. maclovinus*, are hosted on Dryad (`doi.org/10.5061/dryad.zgmsbccfd`).

## Authors' Contributions

A.G.R.C. designed the experiments, collected specimens, constructed RADseq libraries, implemented bioinformatic data analyses, and wrote the manuscript. N.R., B.F.M., and G.M. implemented bioinformatic data analyses. K.T.B. and V.Y. led RNAseq transcript sequencing. M.H. and S.G. collected and contributed specimens. C.H.C.C. designed the experiments, collected specimens, led genome sequencing, analyzed the data, and edited the manuscript. J.M.C. designed the experiments, implemented the bioinformatic data analyses, and edited the manuscript. All authors read and approved the manuscript.

## Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

## References

Amores A, Catchen J, Nanda I, Warren W, Walter R, Schartl M, Postlethwait JH. 2014. A RAD-tag genetic map for the platyfish (*Xiphophorus maculatus*) reveals mechanisms of karyotype evolution among teleost fish. *Genetics* **197**:625–641.

Amores A, Wilson CA, Allard CAH, Detrich HW, Postlethwait JH. 2017. Cold fusion: massive karyotype evolution in the Antarctic bullhead notothen *Notothenia coriiceps*. *G3 (Bethesda)*. **7**:2195–2207.

AquaMaps. 2019. Computer generated distribution maps for *Champsocephalus esox* (Pike icefish), with modelled year 2050 native range map based on IPCC RCP8.5 emissions scenario. *AquaMaps* [Internet]. Available from: https://www.aquamaps.org

Arrigo A-P. 2001. Hsp27: novel regulator of intracellular redox state. *IUBMB Life Int Union Biochem Mol Biol Life*. **52**:303–307.

Arshavsky VY, Lamb TD, Pugh EN. 2002. G proteins and phototransduction. *Annu Rev Physiol*. **64**:153–187.

Auvinet J, Graça P, Belkadi L, Petit L, Bonnivard E, Dettaï A, Detrich WH, Ozouf-Costaz C, Higuet D. 2018. Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*. *BMC Genomics*. **19**:339.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. **3**:e3376.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. **6**:11.

Bargelloni L, Babbucci M, Ferraresso S, Papetti C, Vitulo N, Carraro R, Pauletto M, Santovito G, Lucassen M, Mark FC, *et al.* 2019. Draft genome assembly and transcriptome data of the icefish *Chionodraco myersi* reveal the key role of mitochondria for a life without hemoglobin at subzero temperatures. *Commun Biol*. **2**:443.

Beale A, Guibal C, Tamai TK, Klotz L, Cowen S, Peyric E, Reynoso VH, Yamamoto Y, Whitmore D. 2013. Circadian rhythms in Mexican blind cavefish astyanax mexicanus in the lab and in the field. *Nat Commun*. **4**:2769.

Beers JM, Sidell BD. 2011. Thermal tolerance of Antarctic notothenioid fishes correlates with level of circulating hemoglobin. *Physiol Biochem Zool*. **84**:353–362.

Bilyk KT, Cheng C-HC. 2014. RNA-seq analyses of cellular responses to elevated body temperature in the high Antarctic cryopelagic nototheniid fish *Pagothenia borchgrevinki*. *Mar Genomics*. **18**:163–171.

Bilyk KT, Vargas-Chacoff L, Cheng C-HC. 2018. Evolution in chronic cold: varied loss of cellular response to heat in Antarctic notothenioid fish. *BMC Evol Biol*. **18**:143.

Bilyk KT, Zhuang X, Vargas-Chacoff L, Cheng C-HC. 2021. Evolution of chaperome gene expression and regulatory elements in the Antarctic notothenioid fishes. *Heredity (Edinb)*. **126**:424–441.

Bista I, McCarthy SA, Wood J, Ning Z, Detrich HW III, Desvignes T, Postlethwait J, Chow W, Howe K, Torrance J, *et al.* 2020. The genome sequence of the channel bull blenny, *Cottoperca gobio* (Günther, 1861). *Wellcome Open Res*. **5**:148.

Bista I, Wood JMD, Desvignes T, McCarthy SA, Matschiner M, Ning Z, Tracey A, Torrance J, Sims Y, Chow W, *et al.* 2022. Genomics of cold adaptations in the Antarctic notothenioid fish radiation. Available from: http://biorxiv.org/content/early/2022/06/09/2022.06.08.494096.abstract

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120.

Bourens M, Boulet A, Leary SC, Barrientos A. 2014. Human COX20 cooperates with SCO1 and SCO2 to mature COX2 and promote the assembly of cytochrome c oxidase. *Hum Mol Genet*. **23**:2901–2913.

Brelsford A, Purcell J, Avril A, Tran Van P, Zhang J, Brütsch T, Sundström L, Helanterä H, Chapuisat M. 2020. An ancient and eroded social supergene is widespread across formica ants. *Curr Biol*. **30**:304–311.e4.

Brockerhoff SE, Rieke F, Matthews HR, Taylor MR, Kennedy B, Ankoudinova I, Niemi GA, Tucker CL, Xiao M, Cilluffo MC, *et al.* 2003. Light stimulates a transducin-independent increase of cytoplasmic ca $^{2+}$ and suppression of current in cones from the zebrafish mutant *nof*. *J Neurosci*. **23**:470–480.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. **81**:1084–1097.

Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. **103**:338–348.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma*. **3**:lqaa108.

Calvo J, Morriconi E, Rae GA. 1999. Reproductive biology of the icefish *Champsocephalus esox* (Günther, 1861) (Channichthyidae). *Antarct Sci*. **11**:140–149.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. **10**:421.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. **17**:540–552.

Catchen JM, Conery JS, Postlethwait JH. 2009. Automated identification of conserved synteny after whole-genome duplication. *Genome Res*. **19**:1497–1505.

Ceballos SG, Roesti M, Matschiner M, Fernández DA, Damerau M, Hanel R, Salzburger W. 2019. Phylogenomics of an extra-Antarctic notothenioid radiation reveals a previously unrecognized lineage and diffuse species boundaries. *BMC Evol. Biol*. **19**:13.

Chen H, Detmer SA, Ewald AJ, Griffin EE, Fraser SE, Chan DC. 2003. Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J Cell Biol*. **160**:189–200.

Chen L, DeVries AL, Cheng C-HC. 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 S, Huang T, Zhou Y, Han Y, Xu M, Gu J. 2017. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*. **18**:80.

Chen L, Lu Y, Li W, Ren Y, Yu M, Jiang S, Fu Y, Wang J, Peng S, Bilyk KT, *et al.* 2019. The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. *GigaScience* [Internet] 8. Available from: https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz016/5304890

Colombo M, Damerau M, Hanel R, Salzburger W, Matschiner M. 2015. Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. *J Evol Biol*. **28**:376–394.

Corliss BA, Delalio LJ, Stevenson Keller TC, Keller AS, Keller DA, Corliss BH, Beers JM, Peirce SM, Isakson BE. 2019. Vascular expression of hemoglobin alpha in Antarctic icefish supports iron limitation as novel evolutionary driver. *Front Physiol*. **10**:1389.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol*. **23**:3133–3157.

Damerau M, Matschiner M, Salzburger W, Hanel R. 2014. Population divergences despite long pelagic larval stages: lessons from crocodile icefishes (Channichthyidae). *Mol Ecol*. **23**:284–299.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008.

Detrich HW, Stuart A, Schoenborn M, Parker SK, Methé BA, Amemiya CT. 2010. Genome enablement of the notothenioidei: genome size estimates from 11 species and BAC libraries from 2 representative taxa. *J Exp Zoolog B Mol Dev Evol*. **314B**:369–381.

DeVries AL, Cheng C-HC. 2005. Antifreeze Proteins and Organismal Freezing Avoidance in Polar Fishes. In: Fish Physiology. Vol. 22. Elsevier. p. 155–201. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1546509804220040

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21.

Dornburg A, Federman S, Lamb AD, Jones CD, Near TJ. 2017. Cradles and museums of Antarctic teleost biodiversity. *Nat. Ecol Evol.* **1**: 1379–1384.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. **3**:95–98.

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

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*. **9**:18.

Emms DM, Kelly S. 2015. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. **16**:157.

Emms DM, Kelly S. 2019. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. **20**:238.

Enríquez JA. 2016. Supramolecular organization of respiratory complexes. *Annu Rev Physiol.* **78**:533–561.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2011. SNP Discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, Rockman MV, editors. *Molecular methods for evolutionary genetics.* Totowa (NJ): Humana Press. p. 157–178.

Farnsworth DR, Posner M, Miller AC. 2021. Single cell transcriptomics of the developing zebrafish lens and identification of putative controllers of lens development. *Exp Eye Res*. **206**:108535.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. **117**: 9451–9457.

Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. **22**:566.

Gautier M, Klassmann A, Vitalis R. 2017. Rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol Ecol Resour.* **17**:78–90.

Ghigliotti L, Cheng CC-H, Ozouf-Costaz C, Vacchi M, Pisano E. 2015. Cytogenetic diversity of notothenioid fish from the Ross sea: historical overview and updates. *Hydrobiologia*. **761**:373–396.

Ghigliotti L, Mazzei F, Ozouf-Costaz C, Bonillo C, Williams R, Cheng C-HC, Pisano E. 2007. The two giant sister species of the Southern Ocean, dissostichus eleginoides and dissostichus mawsoni, differ in karyotype and chromosomal pattern of ribosomal RNA genes. *Polar Biol*. **30**:625–634.

Giribets MP, Guerreiro MPG, Santos M, Ayala FJ, Tarrío R, Rodríguez-Trelles F. 2019. Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup. *Mol Ecol.* **28**: 1316–1332.

Grove TJ, Hendrickson JW, Sidell BD. 2004. Two species of Antarctic icefishes (genus *Champsocephalus*) share a common genetic lesion leading to the loss of myoglobin expression. *Polar Biol*. **27**: 579–585.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, Neugeboren B, Turner KM, Jensen JD, Hoekstra HE. 2022. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**: 399–405.

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* **30**:1987–1997.

Harter TS, Sackville M, Wilson JM, Metzger DCH, Egginton S, Esbaugh AJ, Farrell AP, Brauner CJ. 2018. A solution to nature's haemoglobin knockout: a plasma-accessible carbonic anhydrase catalyses CO2 excretion in Antarctic icefish gills. *J Exp Biol.* **221**: jeb190918.

Hodge MR, Kim G, Singh K, Cumsky MG. 1989. Inverse regulation of the yeast COX5 genes by oxygen and heme. *Mol Cell Biol.* **9**: 1958–1964.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS: table 1. *Bioinformatics* **32**:767–769.

Hofmann GE, Buckley BA, Airaksinen S, Keen JE, Somero GN. 2000. Heat-shock protein expression is absent in the Antarctic fish *Trematomus bernacchii* (family Nototheniidae). *J Exp Biol.* **203**: 2331–2339.

Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc B Biol Sci.* **367**:395–408.

Holeton GF. 1970. Oxygen uptake and circulation by a hemoglobinless Antarctic fish (*Chaenocephalus aceratus* Lonnberg) compared with three red-blooded Antartic fish. *Comp Biochem Physiol.* **34**:457–471.

Hudak LM, Lunt S, Chang C-H, Winkler E, Flammer H, Lindsey M, Perkins BD. 2010. The intraflagellar transport protein Ift80 is essential for photoreceptor survival in a zebrafish model of Jeune asphyxiating thoracic dystrophy. *Invest Ophthalmol Vis Sci.* **51**: 3792–3799.

Hüne M, González-Wevar C, Poulin E, Mansilla A, Fernández DA, Barrera-Oro E. 2015. Low level of genetic divergence between Harpagifer fish species (Perciformes: Notothenioidei) suggests a quaternary colonization of Patagonia from the Antarctic Peninsula. *Polar Biol*. **38**:607–617.

Hunter-Ensor M, Ousley A, Sehgal A. 1996. Regulation of the *Drosophila* protein timeless suggests a mechanism for resetting the circadian clock by light. *Cell* **84**:677–685.

Hurd MW, Cahill GM. 2002. Entraining signals initiate behavioral circadian rhythmicity in larval zebrafish. *J Biol Rhythms.* **17**:307–314.

Ishikawa M, Sawada Y, Yoshitomi T. 2015. Structure and function of the interphotoreceptor matrix surrounding retinal photoreceptor cells. *Exp Eye Res*. **133**:3–18.

Jin M, Li S, Nusinowitz S, Lloyd M, Hu J, Radu RA, Bok D, Travis GH. 2009. The role of interphotoreceptor retinoid-binding protein on the translocation of visual retinoids and function of cone photoreceptors. *J Neurosci.* **29**:1486–1495.

Jo E, Lee SJ, Kim J-H, Parker SJ, Choi E, Kim J, Han S-R, Oh T-J, Park H. 2021. Chromosomal-level assembly of Antarctic scaly rockcod, *Trematomus loennbergii* genome using long-read sequencing and chromosome conformation capture (Hi-C) technologies. *Diversity (Basel).* **13**:668.

Johnston IA, Calvo J, Guderley H, Fernandez D, Palmer L. 1998. Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. *J Exp Biol.* **201**:1–12.

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

Kim B-M, Amores A, Kang S, Ahn D-H, Kim J-H, Kim I-C, Lee JH, Lee SG, Lee H, Lee J, *et al.* 2019. Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nat Ecol Evol.* **3**: 469–478.

Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434.

Kiss AJ, Mirarefi AY, Ramakrishnan S, Zukoski CF, DeVries AL, Cheng C-HC. 2004. Cold-stable eye lens crystallins of the Antarctic notothenid toothfish *Dissostichus mawsoni* Norman. *J Exp Biol.* **207**:4633–4649.

Kock K-H. 2005. Antarctic icefishes (Channichthyidae): a unique family of fishes. A review, part I. *Polar Biol*. **28**:862–895.

Kock K-H, Everson I. 2003. Shedding new light on the life cycle of mackerel icefish in the Southern Ocean: life history of mackerel icefish. *J. Fish Biol*. **63**:1–21.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* **37**:540–546.

Kose S, Furuta M, Imamoto N. 2012. Hikeshi, a nuclear import carrier for Hsp70s, protects cells from heat shock-induced nuclear damage. *Cell* **149**:578–589.

Koutsopoulos OS, Laine D, Osellame L, Chudakov DM, Parton RG, Frazier AE, Ryan MT. 2010. Human Miltons associate with mitochondria and induce microtubule-dependent remodeling of mitochondrial networks. *Biochim Biophys Acta BBA—Mol Cell Res*. **1803**:564–574.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. **47**:D807–D811.

Kwast KE, Burke PV, Poyton RO. 1998. Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. *J Exp Biol.* **201**:1177–1195.

La Mesa M, Eastman JT, Vacchi M. 2004. The role of notothenioid fish in the food web of the Ross Sea shelf waters: a review. *Polar Biol*. **27**:321–338.

Lanciano S, Mirouze M. 2018. Transposable elements: all mobile, all different, some stress responsive, some adaptive? *Curr Opin Genet Dev*. **49**:106–114.

Lee CA, Chin L-S, Li L. 2018. Hypertonia-linked protein Trak1 functions with mitofusins to promote mitochondrial tethering and fusion. *Protein Cell*. **9**:693–716.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: https://arxiv.org/abs/1303.3997

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. **102**:10557–10562.

Lu Y, Li W, Li Y, Zhai W, Zhou X, Wu Z, Jiang S, Liu T, Wang H, Hu R, *et al*. 2022. Population genomics of an icefish reveals mechanisms of glacier-driven adaptive radiation in Antarctic notothenioids. *BMC Biol*. **20**:231.

Macías LG, Barrio E, Toft C. 2020. GWideCodeML: a python package for testing evolutionary hypotheses at the genome-wide level. *G3* **10**:4369–4372.

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.

Matschiner M, Barth JMI, Tørresen OK, Star B, Baalsrud HT, Brieuc MSO, Pampoulie C, Bradbury I, Jakobsen KS, Jentoft S. 2022. Supergene origin and maintenance in Atlantic cod. *Nat Ecol Evol*. **6**:469–481.

Matschiner M, Hanel R, Salzburger W. 2011. On the origin and trigger of the notothenioid adaptive radiation. *PLoS One*. **6**:e18911.

Mazzei F, Ghigliotti L, Lecointre G, Ozouf-Costaz C, Coutanceau J-P, Detrich W, Pisano E. 2006. Karyotypes of basal lineages in notothenioid fishes: the genus bovichtus. *Polar Biol*. **29**:1071.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2021. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**:5516–5518.

Mérot C, Berdan EL, Cayuela H, Djambazian H, Ferchaud A-L, Laporte M, Normandeau E, Ragoussis J, Wellenreuther M, Bernatchez L. 2021. Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol Biol Evol.* **38**:3953–3971.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER Version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*. **49**:D394–D403.

Minhas BF, Beck EA, Cheng C-HC, Catchen JM. 2022. Novel mitochondrial genome rearrangements including duplications and extensive heteroplasmy in Antarctic Notothenioid Fishes. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.09.19.508608

Miya T, Gon O, Mwale M, Cheng C-HC. 2016. Multiple independent reduction or loss of antifreeze trait in low Antarctic and sub-Antarctic notothenioid fishes. *Antarct Sci.* **28**:17–28.

Moore JM, Carvajal JI, Rouse GW, Wilson NG. 2018. The Antarctic circumpolar current isolates and connects: structured circumpolarity in the sea star *Glabraster antarctica*. *Ecol Evol.* **8**:10621–10633.

Morescalchi A, Hureau JC, Olmo E, Ozouf-Costaz C, Pisano E, Stanyon R. 1992. A multiple sex-chromosome system in Antarctic ice-fishes. *Polar Biol*. **11**:655–661.

Morgan R, Andreassen AH, Åsheim ER, Finnøen MH, Dresler G, Brembu T, Loh A, Miest JJ, Jutfelt F. 2022. Reduced physiological plasticity in a fish adapted to stable temperatures. *Proc Natl Acad Sci U S A*. **119**:e2201919119.

Morin LP. 1999. Serotonin and the regulation of mammalian circadian rhythmicity. *Ann Med.* **31**:12–33.

Morris TA, Fong S-L. 1993. Characterization of the gene encoding human cone transducin α-subunit (GNAT2). *Genomics* **17**:442–448.

Na U, Yu W, Cox J, Bricker DK, Brockmann K, Rutter J, Thummel CS, Winge DR. 2014. The LYR factors SDHAF1 and SDHAF3 mediate maturation of the iron-sulfur subunit of succinate dehydrogenase. *Cell Metab*. **20**:253–266.

Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernández DA, Jones CD. 2012. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc Natl Acad Sci U S A*. **109**:3434–3439.

Near TJ, MacGuigan DJ, Parker E, Struthers CD, Jones CD, Dornburg A. 2018. Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. *Mol Phylogenet Evol.* **129**:268–279.

Nicodemus-Johnson J, Silic S, Ghigliotti L, Pisano E, Cheng C-HC. 2011. Assembly of the antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish *Dissostichus mawsoni* (Norman). *Genomics* **98**:194–201.

Niemann A, Ruegg M, La Padula V, Schenone A, Suter U. 2005. Ganglioside-induced differentiation associated protein 1 is a regulator of the mitochondrial network. *J Cell Biol*. **170**:1067–1078.

O'Brien KM, Mueller IA. 2010. The unique mitochondrial form and function of Antarctic channichthyid icefishes. *Integr Comp Biol.* **50**:993–1008.

O'Grady SM, Clarke A, DeVries AL. 1982. Characterization of glycoprotein antifreeze biosynthesis in isolated hepatocytes from *Pagothenia borchgrevinki*. *J Exp Zool.* **220**:179–189.

Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* **31**:703–714.

Orsi AH, Whitworth T, Nowlin WD. 1995. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Res Part Oceanogr Res Pap*. **42**:641–673.

Ozouf-Costaz C, Hureau JC, Beaunier M. 1991. Chromosome studies on fish of the suborder Notothenioidei collected in the Weddell Sea during EPOS 3 cruise. *Cybium Paris*. **15**:271–289.

Ozouf-Costaz C, Pisano E, Bonillo C, Williams R. 1996. Ribosomal RNA location in the Antarctic fish *Champsocephalus gunnari* (Notothenioidei, Channichthyidae) using banding and fluorescencein situ hybridization. *Chromosome Res*. **4**:557–561.

Palmerini CA, Mazzoni M, Giovinazzo G, Arienti G. 2009. Blood lipids in Antarctic and in temperate-water fish Species. *J Membr Biol.* **230**:125–131.

Pandey A, Rajesh M, Baral P, Sarma D, Tripathi PH, Akhtar MS, Ciji A, Dubey MK, Pande V, Sharma P, *et al.* 2021. Concurrent changes in thermal tolerance thresholds and cellular heat stress response

reveals novel molecular signatures and markers of high temperature acclimation in rainbow trout. *J Therm Biol.* **102**:103124.
Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.
Parker R, Wang J-S, Kefalov VJ, Crouch RK. 2011. Interphotoreceptor retinoid-binding protein as the physiologically relevant carrier of 11-cis-retinol in the cone visual cycle. *J Neurosci.* **31**:4714–4719.
Poulin E, González-Wevar C, Díaz A, Gérard K, Hüne M. 2014. Divergence between Antarctic and South American marine invertebrates: what molecular biology tells us about Scotia Arc geodynamics and the intensification of the Antarctic circumpolar current. *Glob Planet Change.* **123**:392–399.
Price MN, Dehal PS, Arkin AP. 2010. Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. **5**:e9490.
Qiu X-B, Shao Y-M, Miao S, Wang L. 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol Life Sci*. **63**:2560–2570.
Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. Interproscan: protein domains identifier. *Nucleic Acids Res*. **33**:W116–W120.
Rayamajhi N, Cheng C-HC, Catchen JM. 2022. Evaluating illumina-, nanopore-, and PacBio-based genome assembly strategies with the bald notothen, *Trematomus borchgrevinki*. *G3 (Bethesda)*. **12**:jkac192.
R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Available from: https://www.R-project.org/
Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol.* **28**:4737–4754.
Rosenbaum JL, Witman GB. 2002. Intraflagellar transport. *Nat Rev Mol Cell Biol.* **3**:813–825.
Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods.* **17**:155–158.
Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, *et al.* 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**:832–837.
Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, *et al.* 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.
Sahu B, Sun W, Perusek L, Parmar V, Le Y-Z, Griswold MD, Palczewski K, Maeda A. 2015. Conditional ablation of retinol dehydrogenase 10 in the retinal pigmented epithelium causes delayed dark adaption in mice. *J Biol Chem.* **290**:27239–27247.
Sidell BD, O'Brien KM. 2006. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. *J Exp Biol.* **209**:1791–1802.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.
Sinkler CA, Kalpage H, Shay J, Lee I, Malek MH, Grossman LI, Hüttemann M. 2017. Tissue- and condition-specific isoforms of mammalian cytochrome *c* oxidase subunits: from function to human disease. *Oxid Med Cell Longev*. **2017**:1–19.
Small CM, Bassham S, Catchen J, Amores A, Fuiten AM, Brown RS, Jones AG, Cresko WA. 2016. The genome of the Gulf pipefish enables understanding of evolutionary innovations. *Genome Biol*. **17**:258.
Smit AF, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org
Smith RIL, Walton DWH. 1975. South Georgia, subantarctic. In: Rosswall T, Heal OW, editors. *Structure and function of tundra ecosystems. Ecological bulletins*. Lund, Sweden. Oikos Editorial Office. p. 399–423. Available from: http://www.jstor.org/stable/45331787
Soto IC, Fontanesi F, Liu J, Barrientos A. 2012. Biogenesis and assembly of eukaryotic cytochrome c oxidase catalytic core. *Biochim Biophys Acta BBA—Bioenerg*. **1817**:883–897.
Stahl BA, Gross JB. 2017. A comparative transcriptomic analysis of development in two *Astyanax* cavefish populations. *J Exp Zoolog B Mol Dev Evol*. **328**:515–532.
Stankovic A, Spalik K, Kamler E, Borsuk P, Weglenski P. 2002. Recent origin of sub-antarctic notothenioids. *Polar Biol*. **25**:203–205.
Sukumaran S, Perkins BD. 2009. Early defects in photoreceptor outer segment morphogenesis in zebrafish ift57, ift88 and ift172 intraflagellar transport mutants. *Vision Res*. **49**:479–489.
Tang K, Thornton KR, Stoneking M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol*. **5**:e171.
Toloza-Villalobos J, Arroyo JI, Opazo JC. 2015. The circadian clock of teleost fish: a comparative analysis reveals distinct fates for duplicated genes. *J Mol Evol.* **80**:57–64.
Tota B, Acierno R, Agnisola C. 1991. Mechanical performance of the isolated and perfused heart of the haemoglobinless Antarctic icefish *Chionodraco hamatus* (lönnberg): effects of loading conditions and temperature. *Philos Trans R Soc Lond B Biol Sci*. **332**: 191–198.
Tota B, Gattuso A. 1996. Heart ventricle pumps in teleosts and elasmobranchs: a morphodynamic approach. *J Exp Zool.* **275**:162–171.
Vatine G, Vallone D, Gothilf Y, Foulkes NS. 2011. It's time to swim! Zebrafish and the circadian clock. *FEBS Lett.* **585**:1485–1494.
Wang Y, Liu Z, Li Z, Shi H, Kang Y, Wang J, Huang J, Jiang L. 2016. Effects of heat stress on respiratory burst, oxidative damage and SERPINH1 (HSP47) mRNA expression in rainbow trout Oncorhynchus mykiss. *Fish Physiol Biochem*. **42**:701–710.
Warren IA, Naville M, Chalopin D, Levin P, Berger CS, Galiana D, Volff J-N. 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Res*. **23**:505–531.
Wellenreuther M, Mérot C, Berdan E, Bernatchez L. 2019. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol.* **28**:1203–1209.
Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Drögemüller C, Leeb T, Baumann U. 2012. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc Natl Acad Sci U S A*. **109**:13243–13247.
Wistow G. 2012. The human crystallin gene families. *Hum Genomics*. **6**:26.
Wistow G, Wyatt K, David L, Gao C, Bateman O, Bernstein S, Tomarev S, Segovia L, Slingsby C, Vihtelic T. 2005. γN-crystallin and the evolution of the βγ-crystallin superfamily in vertebrates: γN-crystallin. *FEBS J*. **272**:2276–2291.
Yang Z. 2007. PAML 4: phylogenetic analysis by Maximum likelihood. *Mol Biol Evol.* **24**:1586–1591.
Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol.* **28**:1217–1228.
Young EF, Belchier M, Hauser L, Horsburgh GJ, Meredith MP, Murphy EJ, Pascoal S, Rock J, Tysklind N, Carvalho GR. 2015. Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. *Evol Appl.* **8**:486–509.
Zong S, Wu M, Gu J, Liu T, Guo R, Yang M. 2018. Structure of the intact 14-subunit human cytochrome c oxidase. *Cell Res*. **28**: 1026–1034.