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*], Julian M. Catchen[1†]

[1]Department of Evolution, Ecology, and Behavior, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[2]Informatics Program, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[3]Department of Biology, Montclair State University, Montclair, New Jersey, USA
[4]Centro de Investigación para la Conservación de los Ecosistemas Australes, Punta Arenas, Chile
[5]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK
[6]Government of South Georgia and the South Sandwich Islands, Stanley FIQQ 1ZZ, Falklands
Co-corresponding authors (*c-cheng@illinois.edu, †jcatchen@illinois.edu)

Abstract

White-blooded Antarctic icefishes 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 of *C. esox*, 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. Our study presents the first whole genome characterization of a secondarily temperate notothenioid to date, providing a key resource and platform for understanding the adaptive potential of the highly vulnerable icefishes, and of cryonotothenioids in general, in face of a warming Antarctic environment.

Introduction

Antarctic notothenioid fish, or cryonotothenioids, have experienced extreme biological specialization, evolving and diversifying in the chronically cold environment of the Southern Ocean (SO). Following Antarctica's thermal and geographical isolation over the last 35 million years (Mya) (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 (Colombo *et al.* 2015, but see Matschiner *et al.* 2011; Near *et al.* 2012), dominating modern fish fauna in the SO and playing a key role in the 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 (Hb) 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 evolution of these phenotypes, icefish still exhibit limited cardiac performance, with less than 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 Hb 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 Hb fragments (Corliss *et al.* 2019). Regardless of the evolutionary mechanism behind the loss, the derived physiology stemming from the lack of Hb 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 SO (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 Falklands –non-freezing and seasonally variable environments that are significantly warmer, and much lower in oxygen concentration than the SO. *C. esox* shares the defining traits of the Antarctic icefish clade, including the lack of Hb (Grove *et al.* 2004; Kock 2005), loss of myoglobin (Mb, 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 adaption 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 Mya, 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. *C. 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.

C. esox is a highly elusive species to capture, and therefore poorly researched. The few published studies provide some examples of phenotypic differences in *C. esox* compared to other icefish, including different reproductive biology (Calvo *et al.* 1999), increased mitochondrial densities (Johnston *et al.* 1998), and mitochondrial cristate 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), understanding adaptive processes active in *C. esox*, and in the red-blooded lineages when those genomes are available, will uncover whether a single or a diversity of adaptive processes enabled the re-colonization of temperate environments in this group.

To explore the genomic signatures of secondarily temperate adaptation in the pike icefish, we sequenced *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. To date, the *C. esox* assembly is the only reference genome available for a secondarily temperate notothenioid, enabling us to study the evolution of adaptation following extreme cold specialization. In addition to *C. gunnari,* we included comparisons to 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 specialization to Antarctic cold and oxygen-rich conditions, and unique polar light-dark regime.

Methods

Specimens and sampling

A total of 28 *C. esox* and 68 *C. gunnari* individuals were sampled. *C. esox* was collected from two locations in Southern Chile. Twelve of the 28 were captured with a gill net from Canal Bárbara (a channel connecting the Strait of Magellan to the Pacific) in April 2016. The other 16 were obtained by rod and reel or gill net in the Patagonia waters near Puerto Natales in 2008, 2009, 2016, and from December 2017 to early January 2018. *C. gunnari* was collected from two Antarctic regions. Forty-eight of the 68 were collected from waters around South Georgia and Shag Rock Shelf by trawls from onboard the British Antarctic Survey chartered Falkland Islands F/V *Sil* (ZDLR1) during the South Georgia groundfish survey from January – February 2017. The other 20 individuals were collected from several coastal sites along the West Antarctic Peninsula using otter trawl from onboard the US R/V Laurence M. Gould during austral winters (July – August) of 2008 and 2014. The metadata for the specimens collected and used are given in Table S1.

For genome sequencing, white muscles from a single male *C. gunnari* (collected in 2014) 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) 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^{-1} , 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 4ºC 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 & 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 *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 hr.). The treated blocks were then melted (65°C, 20 min.), and digested with 2 units of β-agarase (NEB) per block (1 hr., 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 continuous long read (PacBio CLR) library preparation and sequencing were carried out at the University of Oregon Genomics & Cell Characterization Core Facility. The HMW DNA was lightly sheared with Megaruptor (Diagenode) at 60 kilobasepairs (Kbp) target length for library construction using PacBio SMRTbell Express Template Prep Kit 2.0. The resulting library was selected for inserts approximately greater than 30 Kbp with the BluePippin (Sage Science) and sequenced on two SMRT cells 8M for each species on Sequel II for 30 hours 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 Table S2.

Hi-C library and sequencing

For scaffolding the genome assembly, Hi-C 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 *DpnII* as the restriction enzyme. The libraries were sequenced by Phase Genomics on an Illumina NovaSeq6000, which yielded 257.8 million and 161.2 million paired-end 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 (RQN), as shown in Table S3. The individual, tissue-specific RNA samples from each species were then pooled at 1µg 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 2x250 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 reads-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 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 1µg (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 2x150 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 Kbp in length that provided 70x coverage, based on an initial genome size estimation of 1.1 gigabasepairs (Gbp), yielding 3.8 million reads with a 20.4 Kbp mean length and 23 Kbp read N50. For *C. gunnari*, the most optimal subsample contained reads over 15 Kbp and 80x coverage, yielding a 29.5 Kbp mean length and a 32 Kbp 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). For both species, *FLYE* v2.5 generated the optimal contig-level assemblies based on quality and completeness scores (Supplements – 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 (Table S4).

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 fixes 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 re-assessment of the contiguity of the final assembly was done using *QUAST* v4.4. An assessment of gene-completeness was performed with both *BUSCO* v3.0.1 and *BUSCO* v5.1.3 (Manni *et al.* 2021).

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* (Ellinghaus *et al.* 2008) (`-LTRStruct`). 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), while 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. 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 flanking genes, including trypsinogen 1 (*tryp1*), trypsinogen 3 (*tryp3*), trypsinogen-like protease (*TLP*), and chimeric *AFGP/TLP* genes, using the *Dissostichus mawsoni* orthologs (Nicodemus-Johnson *et al.* 2011) as queries. Two more distant neighbors, *hsl* (hormone sensitive lipase, *lipeb* homolog) and *tomm40* (translocase of outer mitochondrial membrane 40) bracket the two ends of this region in characterized cryonotothenioids, hereafter referred to as the canonical *AFGP/TLP* locus. 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 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 *Chaenocephalus aceratus* (Kim *et al.* 2019) and *Pseudochaenichthys georgianus* (Bista *et al.* 2022), to confirm their location in orthologous chromosomes. In addition, we discovered a non-canonical *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 (Supplements – 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 (Fig. S1-4). Further validation of the non-canonical *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).

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 *Eleginops 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 prep*) was first compared against the *Xiphophorous 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 (Fig. S5). 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 datasets, to identify genes under positive selection in *C. esox* based on the ratio (ω) of non-synonymous (*dN*) to synonymous substitutions (*dS*). *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 *dN*/*dS* 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 ω values greater than one, with alignment lengths greater than 150 amino acid sites, and showing significant (p-value < 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 ω 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).

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 non-redundant average coverage of 23.8x. A total of 9 samples were removed due to low coverage (<5x). 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. 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 ($p < 5\times10^{-4}$).

In addition to genetic divergence, this RADseq dataset was used to calculate extended haplotype homozygosity (*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 less than 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 dataset, gaps were scaled (`scalegap`) to 20.3 Kbp (median distance between RAD loci in the dataset), with a max gap (`maxgap`) distance of 104.5 Kbp (95th percentile of the inter-locus 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 Kbp 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.

Data Accessibility

The raw PacBio CLR and Hi-C Illumina reads used for the for *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. Additional bioinformatic scripts used for analysis are available on BitBucket (https://bitbucket.org/angelgr2/icefish_genome_paper/).

Genome assemblies and associated annotations are hosted on Dryad (https://doi.org/10.5061/dryad.zgmsbccfd).

Results

Genome assembly and annotation

We generated chromosome-level reference genome assemblies for the secondarily temperate icefish, *C. esox*, and its Antarctic sister species, *C. gunnari.* For *C. esox*, the genome assembly was 987 megabasepairs (Mbp) in length (Table 1), comparable with the assembly size of around 1 Gbp described in other icefish genome assemblies (Kim *et al.* 2019; Bargelloni *et al.* 2019; Bista *et al.* 2022). The scaffold and contig N50 were 43.6 Mbp and 2.6 Mbp, 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. An assessment of gene completeness using *BUSCO* v3.0.1 can be seen in (Table S5). A total of 28,257 protein coding genes were annotated. The *C. gunnari* reference assembly was 994 Mbp in length, with a scaffold and contig N50 of 44.1 Mbp and 3.2 Mbp, respectively. A total of 98.4% 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*, this assembly was 96.3% gene-complete, and a total of 27,918 protein-coding genes were annotated.

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
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 & Single Copy	3,463 (95.1%)		3,469 (95.3%)	
Complete & 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	
Genome Annotation				
Number of Protein Coding Genes	28,257		27,918	
Number of Exons	246,908		249,138	

Conserved synteny between temperate and Antarctic icefish genomes

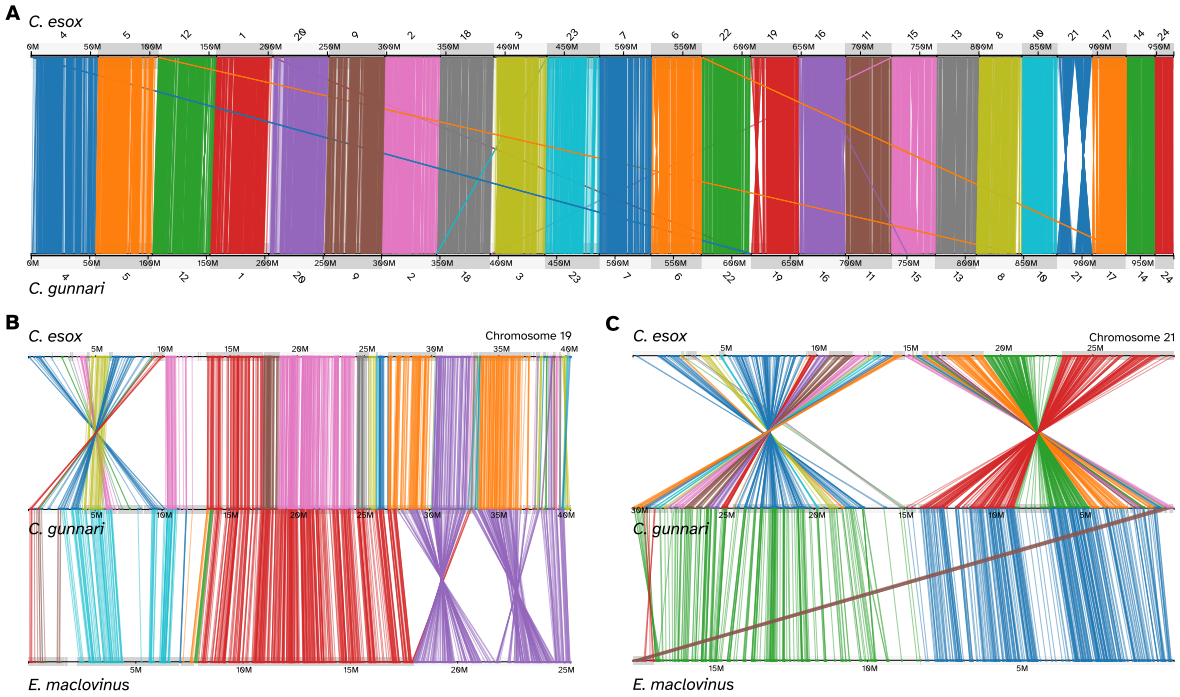

Figure 1. Patterns of conserved synteny between the genomes of temperate and Antarctic icefish species. A) Genome-wide synteny plot shows one-to-one correspondence between the 24 chromosomes of *C. esox* (top) and *C. gunnari* (bottom). Each line represents an orthologous gene between the two genomes, color-coded according to the chromosome of origin. B) Conserved synteny between chromosome 19 in *C. esox* (top), *C. gunnari* (middle), and *E. maclovinus* (bottom). Lines are color-coded according to the genomic scaffold of origin in either *C. esox* or *E. maclovinus*. A *C. esox*-specific chromosomal inversion is seen in the chromosome, located between 1-10 Mbp 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.

In addition to size and chromosome number, the *C. esox* and *C. gunnari* genomes display conservation in large-scale genome organization and synteny, showing one-to-one correspondence between 24 orthologous chromosomes (Fig. 1A). Similar patterns can be observed between *C. esox* and other chromosome-scale icefish genomes (Fig. S6-7), demonstrating that the conservation in genome structure in *C. esox* is not specific to the *Champsocephalus* lineage and might be conserved across icefish. However, several examples of intra-chromosomal rearrangements were observed between the two *Champsocephalus* genomes (Fig. 1B-C), with 12 out of the 24 chromosomes displaying genomic rearrangements of considerable length (>1 Mbp). By comparing conserved synteny against the chromosome scale genome assembly of an outgroup, the non-Antarctic notothenioid *E. maclovinus* – the closest sister to the Antarctic clade (Cheng *et al.*, *in prep*), 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 Mya (Stankovic *et al.* 2002; Near *et al.* 2012; Dornburg *et al.* 2017). One remarkable example is chromosome 21 in *C. esox* (Fig. 1C), which displays a double chromosomal inversion when compared to 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 expansion in genome size and chromosomal diversification (Detrich *et al.* 2010; Auvinet *et al.* 2018; Chen *et al.* 2019). 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. 2A, Table S6). Highly similar values were also observed for the major interspersed repeat element types, *i.e.*, DNA, SINE, LINE, and LTR (Fig. 2A, Table S6). 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. 2A), 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.

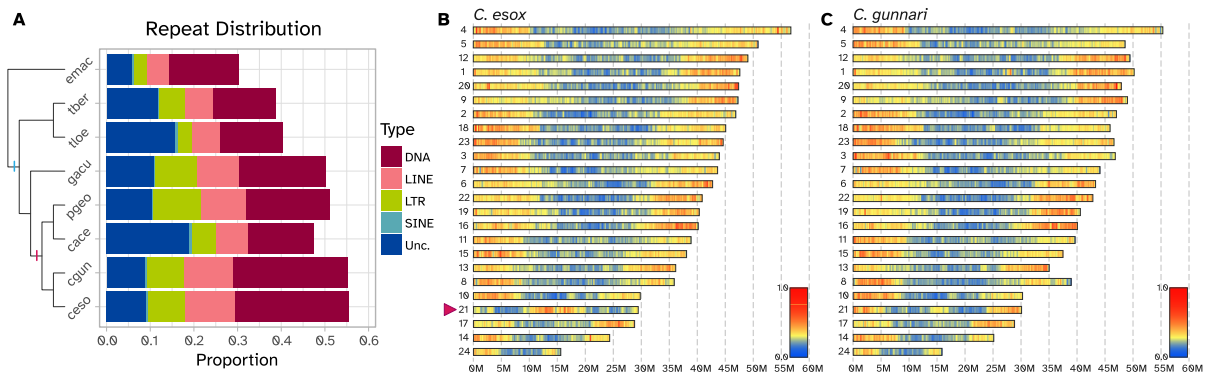

Figure 2. Repeat distribution is conserved along icefish genomes. A) Interspersed repeat distributions for *C. esox* (ceso), *C. gunnari* (cgun), *C. aceratus* (cace), *P. georgianus* (pgeo), *G. acuticeps* (gacu), *T. loennbergii* (tloe), *T. bernacchii* (tber), and *E. maclovinus* (emac). The cladogram on the left shows the phylogenetic relationship among the species. Blue hatch in the cladogram represents the location of Cryonotothenia, while the red hatch shows the location of Channichthyidae. The colors in the 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 Kbp window along the genome. Repeats appear to have a higher distribution along the terminal ends of chromosomes, except for chromosome 21 (red arrow) 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 Kbp window. As in *C. esox*, higher repeat distribution is observed in the terminal ends of chromosomes.

Both *C. esox* and *C. gunnari* display repeats that are more concentrated along the termini of chromosomes than in their centers (Fig. 2B-C). One clear exception is the high repeat content in the center of chromosome 21 in *C. esox* (Fig. 2B), where a species-specific rearrangement is found (Fig. 1C). 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 multiple individuals of each species using RADseq. With this technique we can efficiently and cost-effectively obtain genome-wide genetic diversity by sequencing sites of the genome adjacent to restriction enzyme cutsites, sampling each major linkage block present across the population. Using our RADseq dataset, we were able to analyze ~56K RAD loci (average locus length of 639 bp) along the *C. esox* genome, comprised of 288K variant sites. At the species level, we observe higher genetic diversity in the *C. gunnari* populations (variant-site π of 0.12) than in the *C. esox* populations (π of 0.088). Geographically, the two *C. gunnari* populations, one from South Georgia and the other from the West Antarctic Peninsula, are also more divergent from one another than the two populations of *C. esox* from Puerto Natales and Canal Bárbara (Table S7). At a species level, 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, 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 processes. Tests for changes in homozygosity in combination with linkage disequilibrium (in the form of *EHH*) can be performed to distinguish between neutral and adaptive changes in genetic diversity. When applying this method across our icefish populations (*XP-EHH*), we observed similar patterns of positive selection spanning all 24 chromosomes in *C. esox*. Patterns of population-level directional selection appeared to be more common in *C. esox* than in *C. gunnari.* 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. 3). While 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, *e.g.*, chromosome 12 (Fig. 3A), or outlier windows located within the span of a larger chromosomal rearrangement, *e.g.*, chromosome 21 (Fig. 3B).

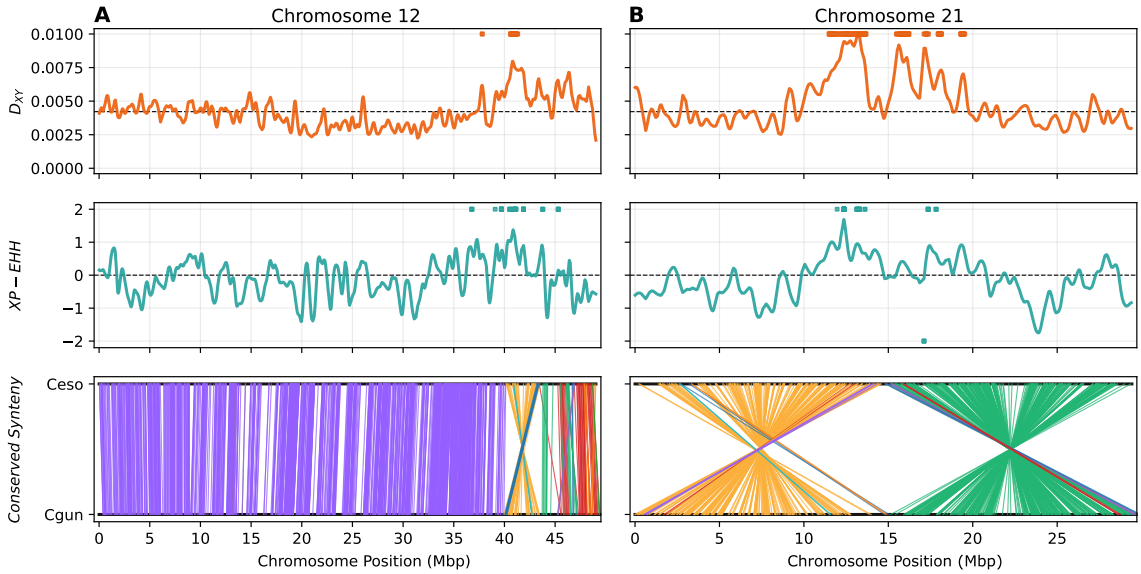

Figure 3: Patterns of genetic divergence and positive selection in populations of temperate and Antarctic icefish species. A) Population-level comparisons for *C. esox* chromosome 12. Top panel shows the distribution of absolute genetic divergence (D_{XY}) between *C. esox* and *C. gunnari* populations. Orange 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*. Cyan 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 Mbp) 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 panel A. This chromosome displays two large chromosomal inversions (one from 0 to15 Mbp, second one from 15 to 30 Mbp), each co-localized with regions with outlier D_{XY} and *XP-EHH* values.

Candidate genes under selection

We used a two-pronged approach to identify candidate genes under selection in *C. esox*. First, using an analysis of molecular evolution, we identified a set of genes displaying elevated signatures of positive selection. By comparing the annotations of six notothenioid 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 ω values (hereafter referred to as *dN/dS* candidates, Table S8). In addition to the analysis of molecular evolution, we also identified genome-wide signatures of selection at a population level using an *XP-EHH* analysis. We then identified 1,490 protein-coding genes annotated within these outlier divergence/*XP-EHH* windows, hereafter referred to as 'divergence/*XP-EHH* candidates' (Table S9). While not all these genes might have substitutions in their coding sequences, we corroborated the presence of nearby SNPs with outlier *EHH* values, which may reflect selection over nearby regulatory sites (Fig. S8).

Among the identified candidates, we observed genes associated with mitochondrial morphology and maintenance, genes associated 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 Table S10.

Table 2: Examples of biological categories displaying top candidates under selection

Category	Gene Name	Gene Description	Type*
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

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

Evolution of antifreeze glycoproteins

The evolution of AFGPs has been regarded as the key evolutionary innovation that allowed notothenioids to survive in the otherwise inhospitable Antarctic environment. 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 its 5'- and 3'-end respectively, as in other cryonotothenioids (Fig. 4). The locus contains 5 tandem duplicates of *tryp1*, one copy each of *tryp3* and *TLP*, and six *AFGP* copies. Five of the *AFGP* copies appear non-functional (denoted with *Ψ*, Fig.4). Two of these are partial copies composed of only exon 2,

containing the coding sequence of the repetitive AFGP tripeptides (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 (5 of 6) 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, non-functional copy.

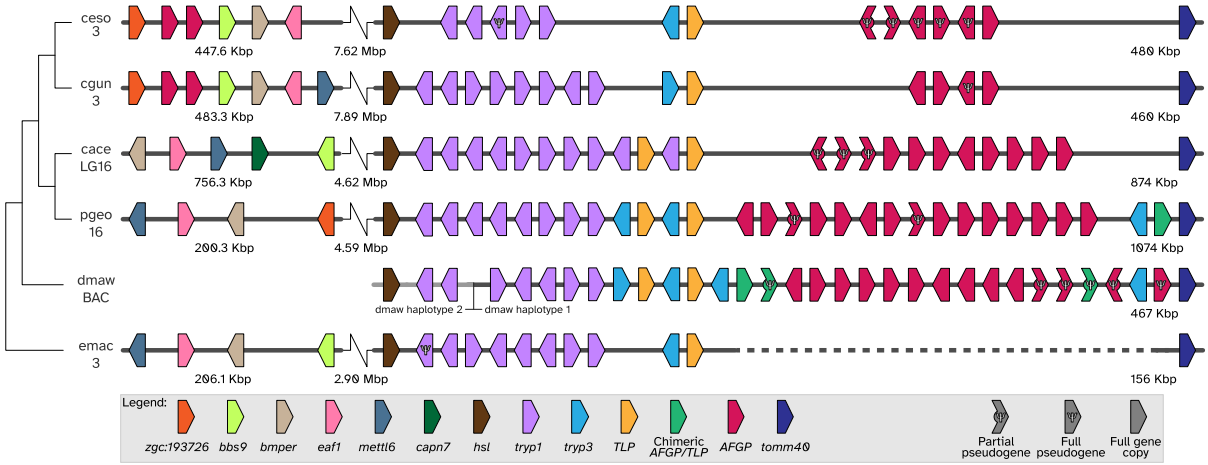

Figure 4: Evolution of the antifreeze glycoprotein locus. Conserved synteny in the *AFGP/TLP* locus for six notothenioids (ceso, *C. esox*; cgun, *C. gunnari*; cace, *C. aceratus*; pgeo, *P. georgianus*; dmaw, *D. mawsoni*; emac, *E. maclovinus*). Cladogram on the left shows the phylogenetic relationship across the six species. The ID for the chromosome containing the *AFGP* is displayed under the species ID. Each horizontal grey 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 grey 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, while 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 to 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. *C. 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 Mbp away from the canonical locus. Gene synteny in this non-canonical locus is highly conserved between *C. gunnari* and *C. esox,* with both *AFGP* copies flanked by the genes *zgc:193726* and *bbs9*. While 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.

Of the six notothenioids we examined, excluding the non-Antarctic outgroup *E. maclovinus,* which does not have *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 Mbp 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 non-canonical *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*. To date, this is the first reference genome available for a secondarily temperate notothenioid fish. This species pair represents the only example of major ecological divergence in the Channichthyidae – two closely related species occurring in disparate thermal 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 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 for other notothenioid species in which, aside from extreme examples in the genus *Notothenia* (Amores *et al.* 2017), the ancestral teleost karyotype number (N=24) is largely conserved (Kim *et al.* 2019; Bista *et al.* 2020, 2022; Jae Lee *et al.* 2021; Jo *et al.* 2021).

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 Antarctic notothenioid 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, while the percentage of the genome composed of transposable elements is higher for the two *Champsocephalus* species when compared to other published notothenioid assemblies (Fig. 2), this is likely a technical artifact explained by methodological differences. For example, when we recomputed the repeat annotation for *C. aceratus* (Kim *et al.* 2019) using the analyses described in this study (see Methods), the repeat distribution increased to that observed in the two *Champsocephalus* assemblies (Fig. S9). 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 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 Mya, 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

Patterns of genetic divergence and population-level positive selection between temperate and Antarctic icefish populations reveal a complex genomic architecture underlying secondarily temperate adaptation in *C. esox*. We identified candidate loci – outliers in both divergence and selection – present in most *C. esox* chromosomes, signaling that changes in selective pressures during the polar-to-temperate transition acted across the whole genome. Therefore, secondarily temperate adaptation from an Antarctic icefish ancestor might not be mediated by a single (or a few) large genetic “switches” and is instead the outcome of complex selection over many genomic regions. Colonization of temperate Patagonian waters by the ancestral *C. esox* lineage likely occurred over the last 2 Mya (Stankovic *et al.* 2002), matching the Quaternary colonization of temperate waters 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.

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. 3). Structural variation, 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; Wellenreuther *et al.* 2019). Chromosomal inversions have played a role in several adaptive processes in wild populations, including thermal adaptation in *Drosophila* (Puig 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), and differentiation between migratory and non-migratory Atlantic cod populations (Matschiner *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 and speciation, allowing for the maintenance and segregation between cold- and temperate-adapted haplotypes in the emerging *C. esox* populations. This would suggest the structural variants, and associated

temperate-adapted haplotypes, existed in the Antarctic icefish population prior to speciation. However, while the divergence between these two species is relatively recent (~1.6 Mya), the currently observed patterns of high and pervasive genetic divergence throughout the genome indicate that there has been little-to-no gene flow between Antarctic and secondarily temperate populations in recent time. These two factors of the genomic architecture – structural variants co-located with signatures of selection alongside high baseline differentiation between the two species – illustrates a contrast between early and late evolutionary forces. Instead of the classic pattern of divergence with gene flow, *i.e.*, a handful of regions in the genome displaying elevated patterns of divergence surrounded by a low divergence background (Cruickshank and Hahn 2014), our data paints a picture where early adaptive divergence was facilitated by structural variants, after which neutral variation independently accumulated between the temperate and Antarctic populations in the absence of gene flow. Following this process of temperate adaptation and isolation, 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 populations with some degree of gene flow. While 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 adaption following specialization. Studying candidates of selection identified across both the *C. esox* reference genome and at the population level, we can observe how secondarily temperate adaptation has impacted an already unique icefish physiology.

Antifreeze glycoproteins

Previous work had detected the presence of AFGP coding sequences in *C. esox* (Miya *et al.* 2016), but it is unknown if they are 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. 4), which is not unexpected given the recent age of divergence from its Antarctic ancestor. However, while 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. While pseudogenization of *AFGP* copies has occurred in both red- and white-blooded Antarctic species (Fig. 4), the particularly high abundance in *C. esox* likely resulted from the absence of freezing-related selective pressures in their temperate environment. Interestingly, three apparently *AFGP* copies (one in the canonical locus, and two in the translocated locus discussed below) without inactivating mutations are still present. Further work is required to determine if they are transcriptionally active or would produce functional proteins.

In notothenioids, the AFGP gene is known to have evolved once from a trypsinogen-like protease precursor (Chen *et al.* 1997) and 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. 4), a feature not observed in all characterized notothenioids to date. This non-canonical *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 it occurred in the ancestral lineage. Interestingly, we find evidence of additional, small-scale rearrangements in the chromosome containing the canonical and novel *AFGP* loci (chromosome 3), both when comparing *C. gunnari* to *C. esox* and to other species (Fig. 4, Fig. S10). While 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 non-canonical *AFGP* locus should explore the presence of polymorphic *AFGP* haplotypes in *Champsocephalus*, particularly with respect to these additional copies. For example, determining if the non-canonical 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 to other red-blooded notothenioid species (O'Brien and Mueller 2010), arising from evolutionary changes of the organelle to compensate for the loss of Hb (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.*, *in prep*). 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), while *gdap1* is known to affect several mitochondrial phenotypes including changes to fragmentation and elongation (Niemann *et al.* 2005). Expression of both gene products have been observed to interact, maintaining mitochondrial morphology and stability across the cellular mitochondrial network (Niemann *et al.* 2005). In addition, the gene *trak1a*, for which we found non-synonymous substitutions under positive selection in *C. esox* (Table S8), 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), *i.e.*, complex IV of the electron transport chain. Two genes that appeared within our divergence/*XP-EHH* windows (Table S9) 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). While 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), while *uqcc1* acts on the assembly of complex III (Brockerhoff *et al.* 2003). While none of these candidate genes contain non-synonymous substitutions that could be altering their canonical function, mutations over nearby non-coding 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 this complexes to maintain or maximize metabolic output in its new warmer, less oxygenated environment.

Vision and Circadian Rhythms

In addition to their different climates, Antarctica and Patagonia possess drastically different light regimes. We found several candidate genes for secondarily temperate adaptation in *C. esox* related to the eye and to vision, likely suggesting selective changes in response to different light environments and photoperiods when compared to Antarctic icefish populations. Among our candidates in divergence/*XP-EHH* windows (Table S9) 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 to 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 to 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 *dN*/*dS* candidates (Table S8) 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 cone 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 the background cryonotothenioid group (Fig. S11). Interestingly, the homologous position in zebrafish *gnat2* is Ser, which possesses similar physico-chemical properties, which may suggest a reversal to temperate phenotype at this position in *C. esox*. The functional outcome of this substitution in the signaling activity of *C. esox* GNAT2 in temperate environments 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). In combination, these genes suggest selection in the maintenance of visual pigments and visual tissue in *C. esox* likely as a response to increased light availability in its new temperate environment.

One prior study reported the loss of genes belonging to the canonical circadian regulation in cryonotothenioids, likely as a result of the unique polar light-dark regimes experienced by Antarctic organisms (Kim *et al.* 2019). While the functional changes resulting from these gene losses in icefish are presently unknown, changes in light patterns can impact the biological clocks of organisms (Beale *et al.* 2013). *C. esox* originated from an Antarctic ancestor subjected to polar regimes of extended months of light and dark periods – and potentially disrupted circadian regulation – but presently experiences more typical daily light-dark cycles at temperate latitudes. Except for a receptor (*htr7a*) for serotonin – a neurotransmitter that exhibits circadian rhythmicity (Morin 1999) – none of our candidate genes in *C. esox* are explicitly related to circadian activity. While the genetic differences described above could simply be related to sensing in this new 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). Changes in light exposure and detection, as a result of selection in these vision-related genes, could lead to it playing a larger role in the entrainment of biological rhythms in the temperate pike icefish.

Heat shock response and molecular chaperones

As a consequence of cold specialization, Antarctic notothenioids are unable to mount the canonical heat shock response (HSR) (Hofmann *et al.* 2000; Bilyk and Cheng 2014; Bilyk *et al.* 2018). Constitutive expression of *hsp6a,* an 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 icefishes in particular, exposure to thermal stress triggers an inflammatory response instead of the canonical HSR (Bilyk *et al.* 2018). *C. 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 (Table S9) 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 non-synonymous 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). While most of these genes do not show coding changes under selection, mutations over nearby non-coding sites might be driving the evolution of HSR regulation. The state of HSR in *C. esox* is presently unknown; however, evidence of selection on these genes suggests 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 further characterize the candidates of selection identified here by taking a functional approach, measuring tissue-specific gene expression of these genes, as well as characterizing the effect of any nonsynonymous mutations altering protein function, in order to fully describe the mechanisms behind secondarily temperate adaption in an icefish.

Acknowledgements

The authors would like to 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 extraction 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. Work was supported by the NSF DGE 10-69157 IGERT Grant to AGR-C, by NSF OPP Grant 1645087 to JC and C-HCC, and NSF ANT Grant 11-42158 to C-HCC.

Author Contributions

AGR-C designed the experiments, collected specimens, constructed RADseq libraries, implemented bioinformatic data analyses, and wrote the manuscript. NR, BFM, and GM implemented bioinformatic data analyses. KTB and VY led RNAseq transcript sequencing. MH and SG collected and contributed specimens. C-HCC designed the experiments, collected specimens, led genome sequencing, analyzed the data, and edited the manuscript. JMC designed the experiments, implemented the bioinformatic data analyses, and edited the manuscript. All authors read and approved the manuscript.

References

Amores A., J. Catchen, I. Nanda, W. Warren, R. Walter, *et al.*, 2014 A RAD-Tag Genetic Map for the Platyfish (*Xiphophorus maculatus*) Reveals Mechanisms of Karyotype Evolution Among Teleost Fish. Genetics 197: 625–641. https://doi.org/10.1534/genetics.114.164293

Amores A., C. A. Wilson, C. A. H. Allard, H. W. Detrich, and J. H. Postlethwait, 2017 Cold Fusion: Massive Karyotype Evolution in the Antarctic Bullhead Notothen *Notothenia coriiceps*. G3 GenesGenomesGenetics 7: 2195–2207. https://doi.org/10.1534/g3.117.040063

Arrigo A.-P., 2001 Hsp27: Novel Regulator of Intracellular Redox State. IUBMB Life Int. Union Biochem. Mol. Biol. Life 52: 303–307. https://doi.org/10.1080/152165401317291156

Arshavsky V. Y., T. D. Lamb, and E. N. Pugh, 2002 G Proteins and Phototransduction. Annu. Rev. Physiol. 64: 153–187. https://doi.org/10.1146/annurev.physiol.64.082701.102229

Auvinet J., P. Graça, L. Belkadi, L. Petit, E. Bonnivard, *et al.*, 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. https://doi.org/10.1186/s12864-018-4714-x

Baird N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, *et al.*, 2008 Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers, (J. C. Fay, Ed.). PLoS ONE 3: e3376. https://doi.org/10.1371/journal.pone.0003376

Bao W., K. K. Kojima, and O. Kohany, 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6: 11. https://doi.org/10.1186/s13100-015-0041-9

Bargelloni L., M. Babbucci, S. Ferraresso, C. Papetti, N. Vitulo, *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. https://doi.org/10.1038/s42003-019-0685-y

Beale A., C. Guibal, T. K. Tamai, L. Klotz, S. Cowen, *et al.*, 2013 Circadian rhythms in Mexican blind cavefish Astyanax mexicanus in the lab and in the field. Nat. Commun. 4: 2769. https://doi.org/10.1038/ncomms3769

Beers J. M., and B. D. Sidell, 2011 Thermal Tolerance of Antarctic Notothenioid Fishes Correlates with Level of Circulating Hemoglobin. Physiol. Biochem. Zool. 84: 353–362. https://doi.org/10.1086/660191

Bilyk K. T., and C.-H. C. Cheng, 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. https://doi.org/10.1016/j.margen.2014.06.006

Bilyk K. T., L. Vargas-Chacoff, and C.-H. C. Cheng, 2018 Evolution in chronic cold: varied loss of cellular response to heat in Antarctic notothenioid fish. BMC Evol. Biol. 18: 143. https://doi.org/10.1186/s12862-018-1254-6

Bilyk K. T., X. Zhuang, L. Vargas-Chacoff, and C.-H. C. Cheng, 2021 Evolution of chaperome gene expression and regulatory elements in the antarctic notothenioid fishes. Heredity 126: 424–441. https://doi.org/10.1038/s41437-020-00382-w

Bista I., S. A. McCarthy, J. Wood, Z. Ning, H. W. Detrich III, *et al.*, 2020 The genome sequence of the channel bull blenny, Cottoperca gobio (Günther, 1861). Wellcome Open Res. 5: 148. https://doi.org/10.12688/wellcomeopenres.16012.1

Bista I., J. M. D. Wood, T. Desvignes, S. A. McCarthy, M. Matschiner, *et al.*, 2022 Genomics of cold adaptations in the Antarctic notothenioid fish radiation. bioRxiv 2022.06.08.494096. https://doi.org/10.1101/2022.06.08.494096

Bolger A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bourens M., A. Boulet, S. C. Leary, and A. Barrientos, 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. https://doi.org/10.1093/hmg/ddu003

Brelsford A., J. Purcell, A. Avril, P. Tran Van, J. Zhang, *et al.*, 2020 An Ancient and Eroded Social Supergene Is Widespread across Formica Ants. Curr. Biol. 30: 304-311.e4. https://doi.org/10.1016/j.cub.2019.11.032

Brockerhoff S. E., F. Rieke, H. R. Matthews, M. R. Taylor, B. Kennedy, *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. https://doi.org/10.1523/JNEUROSCI.23-02-00470.2003

Browning S. R., and B. L. Browning, 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. https://doi.org/10.1086/521987

Browning B. L., Y. Zhou, and S. R. Browning, 2018 A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. 103: 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Brůna T., K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky, 2021 BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinforma. 3: lqaa108. https://doi.org/10.1093/nargab/lqaa108

Calvo J., E. Morriconi, and G. A. Rae, 1999 Reproductive biology of the icefish *Champsocephalus esox* (Günther, 1861) (Channichthyidae). Antarct. Sci. 11: 140–149. https://doi.org/10.1017/S0954102099000206

Camacho C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421. https://doi.org/10.1186/1471-2105-10-421

Castresana J., 2000 Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17: 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Catchen J. M., J. S. Conery, and J. H. Postlethwait, 2009 Automated identification of conserved synteny after whole-genome duplication. Genome Res. 19: 1497–1505. https://doi.org/10.1101/gr.090480.108

Ceballos S. G., M. Roesti, M. Matschiner, D. A. Fernández, M. Damerau, *et al.*, 2019 Phylogenomics of an extra-Antarctic notothenioid radiation reveals a previously unrecognized lineage and diffuse species boundaries. BMC Evol. Biol. 19: 13. https://doi.org/10.1186/s12862-019-1345-z

Chen L., A. L. DeVries, and C.-H. C. Cheng, 1997 Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. 94: 3811–3816. https://doi.org/10.1073/pnas.94.8.3811

Chen H., S. A. Detmer, A. J. Ewald, E. E. Griffin, S. E. Fraser, *et al.*, 2003 Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. J. Cell Biol. 160: 189–200. https://doi.org/10.1083/jcb.200211046

Chen S., T. Huang, Y. Zhou, Y. Han, M. Xu, *et al.*, 2017 AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics 18: 80. https://doi.org/10.1186/s12859-017-1469-3

Chen L., Y. Lu, W. Li, Y. Ren, M. Yu, *et al.*, 2019 The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. GigaScience 8. https://doi.org/10.1093/gigascience/giz016

Colombo M., M. Damerau, R. Hanel, W. Salzburger, and M. Matschiner, 2015 Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. J. Evol. Biol. 28: 376–394. https://doi.org/10.1111/jeb.12570

Corliss B. A., L. J. Delalio, T. C. Stevenson Keller, A. S. Keller, D. A. Keller, *et al.*, 2019 Vascular Expression of Hemoglobin Alpha in Antarctic Icefish Supports Iron Limitation as Novel Evolutionary Driver. Front. Physiol. 10: 1389. https://doi.org/10.3389/fphys.2019.01389

Cruickshank T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23: 3133–3157. https://doi.org/10.1111/mec.12796

Danecek P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, *et al.*, 2021 Twelve years of SAMtools and BCFtools. GigaScience 10: giab008. https://doi.org/10.1093/gigascience/giab008

Detrich H. W., A. Stuart, M. Schoenborn, S. K. Parker, B. A. Methé, *et al.*, 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. https://doi.org/10.1002/jez.b.21341

DeVries A. L., and C. -H. C. Cheng, 2005 Antifreeze Proteins and Organismal Freezing Avoidance in Polar Fishes, pp. 155–201 in *Fish Physiology*, Elsevier.

Dobin A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dornburg A., S. Federman, A. D. Lamb, C. D. Jones, and T. J. Near, 2017 Cradles and museums of Antarctic teleost biodiversity. Nat. Ecol. Evol. 1: 1379–1384. https://doi.org/10.1038/s41559-017-0239-y

Durand N. C., M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, *et al.*, 2016 Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3: 95–98. https://doi.org/10.1016/j.cels.2016.07.002

Eastman J. T., 2005 The nature of the diversity of Antarctic fishes. Polar Biol. 28: 93–107. https://doi.org/10.1007/s00300-004-0667-4

Ellinghaus D., S. Kurtz, and U. Willhoeft, 2008 LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9: 18. https://doi.org/10.1186/1471-2105-9-18

Enríquez J. A., 2016 Supramolecular Organization of Respiratory Complexes. Annu. Rev. Physiol. 78: 533–561. https://doi.org/10.1146/annurev-physiol-021115-105031

Etter P. D., S. Bassham, P. A. Hohenlohe, E. A. Johnson, and W. A. Cresko, 2011 SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing, pp. 157–178 in *Molecular Methods for Evolutionary Genetics*, edited by Orgogozo V., Rockman M. V. Humana Press, Totowa, NJ.

Farnsworth D. R., M. Posner, and A. C. Miller, 2021 Single cell transcriptomics of the developing zebrafish lens and identification of putative controllers of lens development. Exp. Eye Res. 206: 108535. https://doi.org/10.1016/j.exer.2021.108535

Flynn J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, *et al.*, 2020 RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gabriel L., K. J. Hoff, T. Brůna, M. Borodovsky, and M. Stanke, 2021 TSEBRA: transcript selector for BRAKER. BMC Bioinformatics 22: 566. https://doi.org/10.1186/s12859-021-04482-0

Gautier M., A. Klassmann, and R. Vitalis, 2017 REHH 2.0: a reimplementation of the R package REHH to detect positive selection from haplotype structure. Mol. Ecol. Resour. 17: 78–90. https://doi.org/10.1111/1755-0998.12634

Grove TheresaJ., JamieW. Hendrickson, and BruceD. Sidell, 2004 Two species of antarctic icefishes (genus Champsocephalus) share a common genetic lesion leading to the loss of myoglobin expression. Polar Biol. 27. https://doi.org/10.1007/s00300-004-0634-0

Gurevich A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013 QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Harter T. S., M. Sackville, J. M. Wilson, D. C. H. Metzger, S. Egginton, *et al.*, 2018 A solution to nature's haemoglobin knockout: a plasma-accessible carbonic anhydrase catalyses CO2 excretion in Antarctic icefish gills. J. Exp. Biol. jeb.190918. https://doi.org/10.1242/jeb.190918

Hodge M. R., G. Kim, K. Singh, and M. G. Cumsky, 1989 Inverse regulation of the yeast COX5 genes by oxygen and heme. Mol. Cell. Biol. 9: 1958–1964. https://doi.org/10.1128/mcb.9.5.1958-1964.1989

Hoff K. J., S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke, 2016 BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1. Bioinformatics 32: 767–769. https://doi.org/10.1093/bioinformatics/btv661

Hofmann G. E., B. A. Buckley, S. Airaksinen, J. E. Keen, and G. N. Somero, 2000 Heat-shock protein expression is absent in the antarctic fish Trematomus bernacchii (family Nototheniidae). J. Exp. Biol. 203: 2331–2339. https://doi.org/10.1242/jeb.203.15.2331

Hohenlohe P. A., S. Bassham, M. Currey, and W. A. Cresko, 2012 Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. Philos. Trans. R. Soc. B Biol. Sci. 367: 395–408. https://doi.org/10.1098/rstb.2011.0245

Holeton G. F., 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. https://doi.org/10.1016/0010-406X(70)90185-4

Hudak L. M., S. Lunt, C.-H. Chang, E. Winkler, H. Flammer, *et al.*, 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. https://doi.org/10.1167/iovs.09-4312

Hüne M., C. González-Wevar, E. Poulin, A. Mansilla, D. A. Fernández, *et al.*, 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. https://doi.org/10.1007/s00300-014-1623-6

Hunter-Ensor M., A. Ousley, and A. Sehgal, 1996 Regulation of the Drosophila Protein Timeless Suggests a Mechanism for Resetting the Circadian Clock by Light. Cell 84: 677–685. https://doi.org/10.1016/S0092-8674(00)81046-6

Hurd M. W., and G. M. Cahill, 2002 Entraining Signals Initiate Behavioral Circadian Rhythmicity in Larval Zebrafish. J. Biol. Rhythms 17: 307–314. https://doi.org/10.1177/074873002129002618

Ishikawa M., Y. Sawada, and T. Yoshitomi, 2015 Structure and function of the interphotoreceptor matrix surrounding retinal photoreceptor cells. Exp. Eye Res. 133: 3–18. https://doi.org/10.1016/j.exer.2015.02.017

Jae Lee S., J.-H. Kim, E. Jo, E. Choi, J. Kim, *et al.*, 2021 Chromosomal assembly of the Antarctic toothfish (*Dissostichus mawsoni*) genome using third-generation DNA sequencing and Hi-C technology. Zool. Res. 42: 124–129. https://doi.org/10.24272/j.issn.2095-8137.2020.264

Jin M., S. Li, S. Nusinowitz, M. Lloyd, J. Hu, *et al.*, 2009 The Role of Interphotoreceptor Retinoid-Binding Protein on the Translocation of Visual Retinoids and Function of Cone Photoreceptors. J. Neurosci. 29: 1486–1495. https://doi.org/10.1523/JNEUROSCI.3882-08.2009

Jo E., S. J. Lee, J.-H. Kim, S. J. Parker, E. Choi, *et al.*, 2021 Chromosomal-Level Assembly of Antarctic Scaly Rockcod, Trematomus loennbergii Genome Using Long-Read Sequencing and Chromosome Conformation Capture (Hi-C) Technologies. Diversity 13: 668. https://doi.org/10.3390/d13120668

Johnston I. A., J. Calvo, H. Guderley, D. Fernandez, and L. Palmer, 1998 Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. J. Exp. Biol. 201: 1–12. https://doi.org/10.1242/jeb.201.1.1

Kim B.-M., A. Amores, S. Kang, D.-H. Ahn, J.-H. Kim, *et al.*, 2019 Antarctic blackfin icefish genome reveals adaptations to extreme environments. Nat. Ecol. Evol. 3: 469–478. https://doi.org/10.1038/s41559-019-0812-7

Kirkpatrick M., and N. Barton, 2006 Chromosome Inversions, Local Adaptation and Speciation. Genetics 173: 419–434. https://doi.org/10.1534/genetics.105.047985

Kiss A. J., A. Y. Mirarefi, S. Ramakrishnan, C. F. Zukoski, A. L. DeVries, *et al.*, 2004 Cold-stable eye lens crystallins of the Antarctic nototheniid toothfish *Dissostichus mawsoni* Norman. J. Exp. Biol. 207: 4633–4649. https://doi.org/10.1242/jeb.01312

Kock K.-H., 2005 Antarctic icefishes (Channichthyidae): a unique family of fishes. A review, Part I. Polar Biol. 28: 862–895. https://doi.org/10.1007/s00300-005-0019-z

Kolmogorov M., J. Yuan, Y. Lin, and P. A. Pevzner, 2019 Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37: 540–546. https://doi.org/10.1038/s41587-019-0072-8

Kose S., M. Furuta, and N. Imamoto, 2012 Hikeshi, a Nuclear Import Carrier for Hsp70s, Protects Cells from Heat Shock-Induced Nuclear Damage. Cell 149: 578–589. https://doi.org/10.1016/j.cell.2012.02.058

Koutsopoulos O. S., D. Laine, L. Osellame, D. M. Chudakov, R. G. Parton, *et al.*, 2010 Human Miltons associate with mitochondria and induce microtubule-dependent remodeling of mitochondrial networks. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1803: 564–574. https://doi.org/10.1016/j.bbamcr.2010.03.006

Kriventseva E. V., D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, *et al.*, 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. https://doi.org/10.1093/nar/gky1053

Kwast K. E., P. V. Burke, and R. O. Poyton, 1998 Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. J. Exp. Biol. 201: 1177–1195. https://doi.org/10.1242/jeb.201.8.1177

La Mesa M., J. T. Eastman, and M. Vacchi, 2004 The role of notothenioid fish in the food web of the Ross Sea shelf waters: a review. Polar Biol. 27: 321–338. https://doi.org/10.1007/s00300-004-0599-z

Lanciano S., and M. Mirouze, 2018 Transposable elements: all mobile, all different, some stress responsive, some adaptive? Curr. Opin. Genet. Dev. 49: 106–114. https://doi.org/10.1016/j.gde.2018.04.002

Lee C. A., L.-S. Chin, and L. Li, 2018 Hypertonia-linked protein Trak1 functions with mitofusins to promote mitochondrial tethering and fusion. Protein Cell 9: 693–716. https://doi.org/10.1007/s13238-017-0469-4

Li H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/ARXIV.1303.3997

Li H., 2018 Minimap2: pairwise alignment for nucleotide sequences, (I. Birol, Ed.). Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Löytynoja A., and N. Goldman, 2005 An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. 102: 10557–10562. https://doi.org/10.1073/pnas.0409137102

Macías L. G., E. Barrio, and C. Toft, 2020 GWideCodeML: A Python Package for Testing Evolutionary Hypotheses at the Genome-Wide Level. G3 GenesGenomesGenetics 10: 4369–4372. https://doi.org/10.1534/g3.120.401874

Manni M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, 2021 BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes, (J. Kelley, Ed.). Mol. Biol. Evol. 38: 4647–4654. https://doi.org/10.1093/molbev/msab199

Matschiner M., R. Hanel, and W. Salzburger, 2011 On the Origin and Trigger of the Notothenioid Adaptive Radiation, (H. Escriva, Ed.). PLoS ONE 6: e18911. https://doi.org/10.1371/journal.pone.0018911

Matschiner M., J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, *et al.*, 2022 Supergene origin and maintenance in Atlantic cod. Nat. Ecol. Evol. 6: 469–481. https://doi.org/10.1038/s41559-022-01661-x

Mérot C., E. L. Berdan, H. Cayuela, H. Djambazian, A.-L. Ferchaud, *et al.*, 2021 Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly, (D. Charlesworth, Ed.). Mol. Biol. Evol. 38: 3953–3971. https://doi.org/10.1093/molbev/msab143

Mi H., D. Ebert, A. Muruganujan, C. Mills, L.-P. Albou, *et al.*, 2021 PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 49: D394–D403. https://doi.org/10.1093/nar/gkaa1106

Miya T., O. Gon, M. Mwale, and C.-H. C. Cheng, 2016 Multiple independent reduction or loss of antifreeze trait in low Antarctic and sub-Antarctic notothenioid fishes. Antarct. Sci. 28: 17–28. https://doi.org/10.1017/S0954102015000413

Morgan R., A. H. Andreassen, E. R. Åsheim, M. H. Finnøen, G. Dresler, *et al.*, 2022 Reduced physiological plasticity in a fish adapted to stable temperatures. Proc. Natl. Acad. Sci. 119: e2201919119. https://doi.org/10.1073/pnas.2201919119

Morin L. P., 1999 Serotonin and the regulation of mammalian circadian rhythmicity. Ann. Med. 31: 12–33. https://doi.org/10.3109/07853899909019259

Morris T. A., and S.-L. Fong, 1993 Characterization of the Gene Encoding Human Cone Transducin α-Subunit (GNAT2). Genomics 17: 442–448. https://doi.org/10.1006/geno.1993.1345

Na U., W. Yu, J. Cox, D. K. Bricker, K. Brockmann, *et al.*, 2014 The LYR Factors SDHAF1 and SDHAF3 Mediate Maturation of the Iron-Sulfur Subunit of Succinate Dehydrogenase. Cell Metab. 20: 253–266. https://doi.org/10.1016/j.cmet.2014.05.014

Near T. J., A. Dornburg, K. L. Kuhn, J. T. Eastman, J. N. Pennington, *et al.*, 2012 Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Natl. Acad. Sci. 109: 3434–3439. https://doi.org/10.1073/pnas.1115169109

Near T. J., D. J. MacGuigan, E. Parker, C. D. Struthers, C. D. Jones, *et al.*, 2018 Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. Mol. Phylogenet. Evol. 129: 268–279. https://doi.org/10.1016/j.ympev.2018.09.001

Nicodemus-Johnson J., S. Silic, L. Ghigliotti, E. Pisano, and C.-H. C. Cheng, 2011 Assembly of the antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish Dissostichus mawsoni (Norman). Genomics 98: 194–201. https://doi.org/10.1016/j.ygeno.2011.06.002

Niemann A., M. Ruegg, V. La Padula, A. Schenone, and U. Suter, 2005 Ganglioside-induced differentiation associated protein 1 is a regulator of the mitochondrial network. J. Cell Biol. 170: 1067–1078. https://doi.org/10.1083/jcb.200507087

O'Brien K. M., and I. A. Mueller, 2010 The Unique Mitochondrial Form and Function of Antarctic Channichthyid Icefishes. Integr. Comp. Biol. 50: 993–1008. https://doi.org/10.1093/icb/icq038

O'Grady S. M., A. Clarke, and A. L. DeVries, 1982 Characterization of glycoprotein antifreeze biosynthesis in isolated hepatocytes fromPagothenia borchgrevinki. J. Exp. Zool. 220: 179–189. https://doi.org/10.1002/jez.1402200207

Oliver K. R., and W. K. Greene, 2009 Transposable elements: powerful facilitators of evolution. BioEssays 31: 703–714. https://doi.org/10.1002/bies.200800219

Ozouf-Costaz C., E. Pisano, C. Bonillo, and R. Williams, 1996 Ribosomal RNA location in the Antarctic fishChampsocephalus gunnari (Notothenioidei, Channichthyidae) using banding and fluorescencein situ hybridization. Chromosome Res. 4: 557–561. https://doi.org/10.1007/BF02261718

Pandey A., M. Rajesh, P. Baral, D. Sarma, P. H. Tripathi, *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. https://doi.org/10.1016/j.jtherbio.2021.103124

Paradis E., and K. Schliep, 2019 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R, (R. Schwartz, Ed.). Bioinformatics 35: 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parker R., J.-S. Wang, V. J. Kefalov, and R. K. Crouch, 2011 Interphotoreceptor Retinoid-Binding Protein as the Physiologically Relevant Carrier of 11-cis-Retinol in the Cone Visual Cycle. J. Neurosci. 31: 4714–4719. https://doi.org/10.1523/JNEUROSCI.3722-10.2011

Puig Giribets M., M. P. García Guerreiro, M. Santos, F. J. Ayala, R. Tarrío, *et al.*, 2019 Chromosomal inversions promote genomic islands of concerted evolution of *Hsp70* genes in the *Drosophila subobscura* species subgroup. Mol. Ecol. 28: 1316–1332. https://doi.org/10.1111/mec.14511

Qiu X.-B., Y.-M. Shao, S. Miao, and L. Wang, 2006 The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell. Mol. Life Sci. 63: 2560–2570. https://doi.org/10.1007/s00018-006-6192-6

Quevillon E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, *et al.*, 2005 InterProScan: protein domains identifier. Nucleic Acids Res. 33: W116–W120. https://doi.org/10.1093/nar/gki442

R Core Team, 2020 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rayamajhi N., C.-H. C. Cheng, and J. M. Catchen, 2022 Evaluating Illumina-, Nanopore-, and PacBio-based genome assembly strategies with the bald notothen, *Trematomus borchgrevinki*. G3 GenesGenomesGenetics jkac192. https://doi.org/10.1093/g3journal/jkac192

Rochette N. C., A. G. Rivera-Colón, and J. M. Catchen, 2019 Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28: 4737–4754. https://doi.org/10.1111/mec.15253

Rosenbaum J. L., and G. B. Witman, 2002 Intraflagellar transport. Nat. Rev. Mol. Cell Biol. 3: 813–825. https://doi.org/10.1038/nrm952

Ruan J., and H. Li, 2020 Fast and accurate long-read assembly with wtdbg2. Nat. Methods 17: 155–158. https://doi.org/10.1038/s41592-019-0669-3

Sabeti P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837. https://doi.org/10.1038/nature01140

Sabeti P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913–918. https://doi.org/10.1038/nature06250

Sahu B., W. Sun, L. Perusek, V. Parmar, Y.-Z. Le, *et al.*, 2015 Conditional Ablation of Retinol Dehydrogenase 10 in the Retinal Pigmented Epithelium Causes Delayed Dark Adaption in Mice. J. Biol. Chem. 290: 27239–27247. https://doi.org/10.1074/jbc.M115.682096

Sidell B. D., and K. M. O'Brien, 2006 When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. J. Exp. Biol. 209: 1791–1802. https://doi.org/10.1242/jeb.02091

Simão F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sinkler C. A., H. Kalpage, J. Shay, I. Lee, M. H. Malek, *et al.*, 2017 Tissue- and Condition-Specific Isoforms of Mammalian Cytochrome *c* Oxidase Subunits: From Function to Human Disease. Oxid. Med. Cell. Longev. 2017: 1–19. https://doi.org/10.1155/2017/1534056

Small C. M., S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, *et al.*, 2016 The genome of the Gulf pipefish enables understanding of evolutionary innovations. Genome Biol. 17: 258. https://doi.org/10.1186/s13059-016-1126-6

Smit A. F., R. Hubley, and P. Green, 2013 RepeatMasker Open-4.0

Soto I. C., F. Fontanesi, J. Liu, and A. Barrientos, 2012 Biogenesis and assembly of eukaryotic cytochrome c oxidase catalytic core. Biochim. Biophys. Acta BBA - Bioenerg. 1817: 883–897. https://doi.org/10.1016/j.bbabio.2011.09.005

Stahl B. A., and J. B. Gross, 2017 A Comparative Transcriptomic Analysis of Development in Two *Astyanax* Cavefish Populations. J. Exp. Zoolog. B Mol. Dev. Evol. 328: 515–532. https://doi.org/10.1002/jez.b.22749

Stankovic A., K. Spalik, E. Kamler, P. Borsuk, and P. Weglenski, 2002 Recent origin of sub-Antarctic notothenioids. Polar Biol. 25: 203–205. https://doi.org/10.1007/s00300-001-0327-x

Sukumaran S., and B. D. Perkins, 2009 Early defects in photoreceptor outer segment morphogenesis in zebrafish ift57, ift88 and ift172 Intraflagellar Transport mutants. Vision Res. 49: 479–489. https://doi.org/10.1016/j.visres.2008.12.009

Tang K., K. R. Thornton, and M. Stoneking, 2007 A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome, (K. H. Wolfe, Ed.). PLoS Biol. 5: e171. https://doi.org/10.1371/journal.pbio.0050171

Tota B., R. Acierno, and C. Agnisola, 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. https://doi.org/10.1098/rstb.1991.0049

Tota B., and A. Gattuso, 1996 Heart ventricle pumps in teleosts and elasmobranchs: A morphodynamic approach. J. Exp. Zool. 275: 162–171. https://doi.org/10.1002/(SICI)1097-010X(19960601/15)275:2/3<162::AID-JEZ8>3.0.CO;2-B

Vatine G., D. Vallone, Y. Gothilf, and N. S. Foulkes, 2011 It's time to swim! Zebrafish and the circadian clock. FEBS Lett. 585: 1485–1494. https://doi.org/10.1016/j.febslet.2011.04.007

Wang Y., Z. Liu, Z. Li, H. Shi, Y. Kang, *et al.*, 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. https://doi.org/10.1007/s10695-015-0170-6

Warren I. A., M. Naville, D. Chalopin, P. Levin, C. S. Berger, *et al.*, 2015 Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome Res. 23: 505–531. https://doi.org/10.1007/s10577-015-9493-5

Wellenreuther M., C. Mérot, E. Berdan, and L. Bernatchez, 2019 Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Mol. Ecol. 28: 1203–1209. https://doi.org/10.1111/mec.15066

Widmer C., J. M. Gebauer, E. Brunstein, S. Rosenbaum, F. Zaucke, *et al.*, 2012 Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. Proc. Natl. Acad. Sci. 109: 13243–13247. https://doi.org/10.1073/pnas.1208072109

Wistow G., K. Wyatt, L. David, C. Gao, O. Bateman, *et al.*, 2005 γN-crystallin and the evolution of the βγ-crystallin superfamily in vertebrates: γN-crystallin. FEBS J. 272: 2276–2291. https://doi.org/10.1111/j.1742-4658.2005.04655.x

Wistow G., 2012 The human crystallin gene families. Hum. Genomics 6: 26. https://doi.org/10.1186/1479-7364-6-26

Yang Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24: 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang Z., and M. dos Reis, 2011 Statistical Properties of the Branch-Site Test of Positive Selection. Mol. Biol. Evol. 28: 1217–1228. https://doi.org/10.1093/molbev/msq303

Zong S., M. Wu, J. Gu, T. Liu, R. Guo, *et al.*, 2018 Structure of the intact 14-subunit human cytochrome c oxidase. Cell Res. 28: 1026–1034. https://doi.org/10.1038/s41422-018-0071-1