Transcriptomics of a Greenlandic Snailfish Reveals Exceptionally High Expression of Antifreeze Protein Transcripts

Evolutionary Bioinformatics
Volume 18: 1–9

DOI: 10.1177/11769343221118347

John A Burns[1,2], David F Gruber[1,3,4], Jean P Gaffney[3,4], John S Sparks[5] and Mercer R Brugler[1,6]

[1]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, USA. [2]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA. [3]Department of Natural Sciences, Baruch College, City University of New York, New York, NY, USA. [4]The Graduate Center, Program in Biology, City University of New York, New York, NY, USA. [5]Department of Ichthyology, Division of Vertebrate Zoology, American Museum of Natural History, New York, NY, USA. [6]Department of Natural Sciences, University of South Carolina Beaufort, Beaufort, SC, USA.

ABSTRACT: Polar fishes have evolved antifreeze proteins (AFPs) that allow them to survive in subzero temperatures. We performed deep transcriptomic sequencing on a postlarval/juvenile variegated snailfish, *Liparis gibbus* (Actinopterygii: Scorpaeniformes: Cottoidei: Liparidae), living in an iceberg habitat (−2°C) in Eastern Greenland and report detection of highly expressed transcripts that code for putative AFPs from 2 gene families, Type I and LS-12-like proteins (putative Type IV AFPs). The transcripts encoding both proteins have expression levels among the top <1% of expressed genes in the fish. The Type I AFP sequence is different from a reported Type I AFP from the same species, possibly expressed from a different genetic locus. While prior findings from related adult sculpins suggest that LS-12-like/Type IV AFPs may not have a role in antifreeze protection, our finding of very high relative gene expression of the LS-12-like gene suggests that highly active transcription of the gene is important to the fish in the iceberg habitat and raises the possibility that weak or combinatorial antifreeze activity could be beneficial. These findings highlight the physiological importance of antifreeze proteins to the survival of fishes living in polar habitats.

KEYWORDS: Antifreeze proteins, transcriptomics, Arctic, fish, high expression

RECEIVED: March 5, 2022. **ACCEPTED:** July 8, 2022.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by grants from AMNH Constantine S. Niarchos Scientific Expedition grant to JSS and DFG, NSF CAREER grant#1652731 to JPG and NSF OIA grant #1826734 to JAB.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: John A Burns, Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME 04544, USA.
Email: jburns@bigelow.org

David F Gruber, Department of Natural Sciences, Baruch College and The Graduate Center, PhD Program in Biology, City University of New York, 17 Lexington Avenue, New York, NY 10010, USA.
Email: david.gruber@baruch.cuny.edu

Introduction

Fish antifreeze proteins

The icy waters of polar oceans are an extreme environment for marine life that limits inhabitants to those with mechanisms to cope with freezing (−1.9°C) temperatures. The body fluids of fishes are hypoosmotic to seawater, and hence, have a higher freezing point than seawater.[1,2] In addition, fishes are incapable of surviving even partial freezing of their body fluids, unlike some species of reptiles and insects.[3,4] Antifreeze proteins (AFPs) were first discovered in Antarctic notothenioid fishes almost 5 decades ago[5] and many polar and subpolar fishes are known to possess ice-binding AFPs that allow them to successfully inhabit icy habitats.[6] Proteins with ice-binding properties have evolved independently in many different organisms, spanning bacteria, lichens, plants, insects, and vertebrates, to enable them to survive in sub-zero environments.[7,8] AFPs offer an advantage to animals that dwell in places below the freezing point by preventing the formation of large ice grains inside cells and body fluids, thus preserving body fluids in a liquid state. AFPs limit the growth of ice crystals to manageable sizes, rather than fully inhibiting the growth of ice crystals inside fishes. Thus, they are also referred to as ice-structuring proteins.[9] However, for the purpose of this article, we use the term AFP.

Fish AFPs, that are primarily synthesized in the liver and exported to the blood, can be seasonally regulated (eg, by water temperature or photoperiod) or maintained at high concentrations year-round.[10] Five classes of AFPs (Types I, II, III, and IV, as well as antifreeze glycoproteins (AFGPs)), that are differentiated based on their origin, amino acid composition, and the resulting 3D structure of the polypeptides, have been identified and characterized from fishes living in mid- to high-latitude seawater.[11,12] According to Graham et al,[13] Type I AFPs, that consist of single alanine-rich alpha-helices, are found in cunner, winter flounder, sculpins and snailfishes. Type II AFPs, derived from lectins, are found in smelt, herring, and sea raven. More specifically, Type II AFPs share high sequence similarity with the carbohydrate recognition domain of Ca^{2+}-dependent type (C-type) lectins.[14] However, Nishimiya et al[14] identified the AFP of the longsnout poacher, *Brachyopsis segaliensis*, as a Ca^{2+}-independent Type II AFP. Graham et al[15] hypothesized that herring acquired their Type II AFP via horizontal gene transfer of foreign extracellular DNA into herring sperm during external fertilization. The presence of a multitude of transposable elements (TEs) within the herring genome[16] allowed

Graham and Davies[17] to demonstrate that the Type II AFP was transferred from herring to smelt because the smelt AFP is bordered by 3 herring-specific TEs. Type III AFPs, derived from the C-terminal domain of the enzyme sialic acid synthase,[18] have been found in eelpouts and wolffish. LS-12 proteins, exhibiting weak antifreeze activity, were described as a new class of AFP, Type IV AFPs, and were first isolated and characterized in the longhorn sculpin, *Myoxocephalus octodecimspinosis*.[19] AFGPs have been found in Arctic codfishes (evolved from non-coding DNA)[20] and Antarctic notothenioids (derived from a trypsinogen-like protease).[21,22]

According to Deng et al,[19] Cheng,[23] and Cheng and Zhuang,[24] the evolutionary precursor of LS-12 proteins is a blood plasma apolipoprotein. The subsequent discovery of homologs of the LS-12 genes across a broad range of tropical and subtropical teleost fishes called into question whether the LS-12 protein, and therefore the new class of Type IV AFPs, are actually apolipoproteins that coincidentally have negligible antifreeze properties in polar species, or whether they have different functions altogether (eg, gastrulation; lipid metabolism).[25-28] Li et al[29] analyzed the expression of the LS-12-like gene homolog in the gibel carp (*Carassius auratus gibelio*) and found that the gene is predominantly expressed during embryogenesis; and that the protein was no longer detectable in larvae older than 13 days. Deng and Laursen[30] showed that the concentration of the LS-12 gene in the longhorn sculpin, *M. octodecimspinosis*, is so low that without concentration of the protein, antifreeze activity could not be detected. A protein homologlous to the *M. octodecimspinosis* LS-12/TypeIV AFP from the subtropical olive flounder (*Paralichthys olivaceus*),[31] which does not require AFPs for survival, has been cloned, heterologously expressed (in *E. coli*) and assayed for antifreeze activity. Based on its comparatively "very weak" thermal hysteresis activity (0.07 ± 0.01°C at 0.5 mg/mL), the *P. olivaceus* LS-12 homolog would not prevent freezing at subzero temperatures. Additionally, the thermal hysteresis activity of the LS-12 homolog in the subtropical *P. olivaceus* is comparable to that of LS-12 proteins in cold-water fishes. The authors concluded that the activity, inheritance pattern, and sequence conservation of the LS-12 gene family suggests that its antifreeze activity was not gained due to selective pressure imparted by a cold habitat, but is rather a chance activity of a protein needed for other purposes.

Fish often have more than one copy of AFP genes in their genetic repertoire. The shorthorn sculpin *Myoxocephalus scorpius* and longhorn sculpin *Myoxocephalus octodecemspinosus* have both Type I AFPs and LS-12 gene homologs. In those fishes, the Type I AFP that is present in peripheral tissues (skin) was hypothesized to provide sufficient protection from freezing,[25,32] while their LS-12-like protein was found in insufficient concentrations in fish tissues to provide freezing protection.[25] In fishes that have either Type I, II or III AFPs, the fish often carry multiple copies of genes encoding proteins from the same AFP family in their genomes.[33] The expanded gene families may account for high expression of AFPs, while LS-12 gene homologs are hypothesized to be less highly expressed because they are found in a single locus.[25] Yet, in this study we report exceptionally high expression of transcripts coding for Type I AFPs and an LS-12-like gene homolog in the same juvenile individual of the variegated snailfish, *Liparis gibbus*.

Classification and general biology of snailfishes (Liparidae)

Liparidae is a diverse family of cottoid fishes with approximately 420 species arrayed in 32 genera. Liparids are characterized by their asquamate, elongate bodies, and by a sucking disk formed by modified pelvic fins. Liparidae is bipolar in distribution, with members of the widespread genus *Liparis* (~60 spp.) restricted to the northern hemisphere. Liparids inhabit the intertidal zone to the extreme depths of the Mariana Trench, and exhibit a greater depth range than any other family of fishes.[34] They are among the most common hadal vertebrate species, and *Pseudoliparis swirei* and the undescribed "ethereal snailfish" are currently recognized as the deepest dwelling vertebrates (8076 and 8178 m, respectively).[34,35]

Liparis gibbus is a polar benthic species occurring in the Arctic, North Atlantic, and the Northeast Pacific oceans (Greenland, Canada, Russia, Svalbard, and southeastern Alaska).[36] It generally occurs between 100 and 200 m depth, but is frequently encountered at much shallower depths, has an average length of 11 cm (max: 52 cm), and typically feeds on amphipods and crabs.[37-39] *L. gibbus* is currently the only polar fish reported to have biofluorescence.[40] Prior research of *Liparis atlanticus* and *L. gibbus* AFPs discovered 1 Type I AFP in *L. atlanticus* and 2 Type I AFP proteins with confirmed antifreeze activity in blood plasma and an identical Type I AFP in the skin of *L. gibbus*. Additionally, Southern blot analysis of *L. gibbus* genomic DNA indicated that it codes for multiple copies of Type I AFP coding sequences.[41]

Results

Transcriptome statistics

In this study, we sequenced and assembled the transcriptome of a post-larval juvenile variegated snailfish, *L. gibbus* (Figure 1), collected on an iceberg in Eastern Greenland (Figure 2). The data comes from the whole animal minus the gut. After filtering based on read depth (retaining transcripts with >100 reads mapped), the transcriptome assembly consisted of 93 433 contigs with 68 031 of those with a length greater than 1000 bps. The average transcript length was 2190 bp. BUSCO completeness of the filtered assembly shows that it contained complete sequences of 84.4% of the BUSCOS from the vertebrate set, and is only missing 7.4% of them, suggesting the assembly captured the majority of expected conserved genes in vertebrates (Supplemental Table S1).

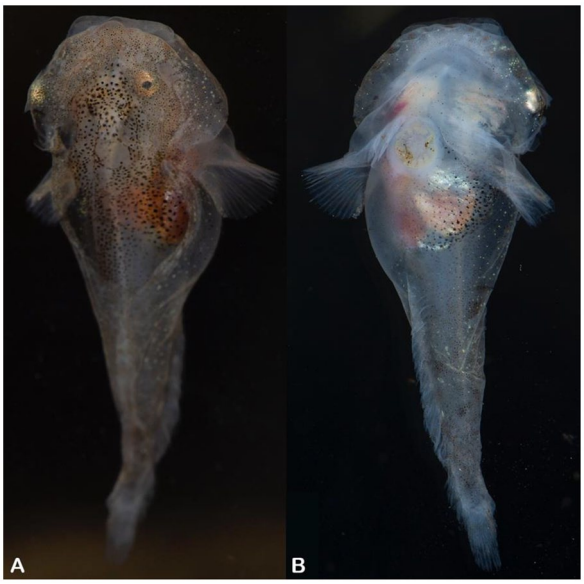

Figure 1. Post-larval/juvenile *Liparis gibbus*: (A) dorsal view and (B) ventral view. (AMNH 277096, 18 mm Standard Length (SL), measured from the tip of the snout (upper jaw) to the hypural flexure (at base of the caudal fin)).

Liparis gibbus *AFPs*

The transcriptome was screened for putative AFPs from all 5 AFP families. We detected AFPs from Type I and Type IV AFP (LS-12-like) families, but not Type II, Type III, or AFGPs. We also investigated Type I AFPs and LS-12 like proteins in available whole genome assemblies of 2 fish from the Liparidae, the hadal zone snailfish *Pseudoliparis swirei* and Tanaka's snailfish *Liparis tanakae*.

In the juvenile *L. gibbus* transcriptome assembly we identified 2 distinct full length, and 3 truncated, transcripts encoding putative Type I AFPs. The truncation in assembled contigs may be an artifact of a strong predicted secondary structure found in the 3′ end of the full-length transcripts (Supplemental Figure 1). In one of 2 available whole genome assemblies for *L. tanakae* (NCBI BioProject# PRJNA523297; https://doi.org/10.6084/m9.figshare.7819418.v1), we found 3 distinct copies of genes encoding full length Type I AFP proteins. In the other independent assembly of the *L. tanakae* genome,[34] the regions containing the Type I AFP were not found. Likewise, the *P. swirei* genome assembly[34] did not contain evidence of Type I AFP genes, however we cannot say whether that is because they are wholly absent or whether the Type I AFP genes are recalcitrant to assembly, as seems to be the case in *L. tanakae*.

Protein alignments of translations of the 5 complete or partial Type I AFPs discovered in our data with previously reported *L. gibbus* and *L. atlanticus* sequences, as well as the Type I AFP sequences discovered in the *L. tanakae* geneome, reveals that our whole transcriptome assembly uncovered the sequences of previously unreported *L. gibbus* Type I AFPs. The protein alignments show that the 2 novel full length AFPs share 66% and 62% identity, respectively, with AY455863 at the amino acid level (Figure 3). We did not detect the previously reported *L. gibbus* Type I AFP, AY455863, in our *de novo* transcriptome assembly. Additionally, the 3 Type I AFP sequences found in the *L. tanakae* genome vary from one another and are different from other reported Type I AFP sequences from fishes in the genus *Liparis* (Figure 3).

We identified 2 full length transcripts encoding LS-12-like proteins (putative Type IV AFPs) in the juvenile *L. gibbus* transcriptome assembly. Alignment of the 2 juvenile *L. gibbus* LS-12-like transcripts to the closely related *L. tanakae* genome suggests that they are splice variants. The longer contig from our juvenile *L. gibbus* assembly, contig 445 731, retains the first intron of the gene, producing an additional ATG start codon that adds a predicted secretory signal peptide on that isoform (Supplemental Figure 2). The shorter isoform, contig 445 757, has that intron spliced out and does not have any predicted signal peptides. Protein alignments showed that the LS-12-like protein from *L. gibbus* is nearly identical to the LS-12-like protein found in *L. tanakae* (Figure 3, 98.4% similarity). Additionally, unlike Type I AFPs, LS-12 proteins appear to occur at a single locus and a phylogenetic tree based on LS-12-like protein sequences suggests that the gene follows a pattern of divergence in accordance with speciation in the group (Figure 4).

We performed 3D homology modeling using Alpha-Fold[42] to infer the tertiary structure of the putative antifreeze proteins. The Type I AFP exhibits a primarily alpha-helical structure, consistent with prior examinations of Type I AFP structure[43] (Figure 5). The LS-12-like protein also folded into a primarily alpha helical structure, consistent with predicted structures of LS-12 proteins from fish in the literature.[44]

The putative full-length Type I AFP and LS-12-like transcript sequences of the snailfish *L. gibbus* were deposited under GenBank Accessions numbers MT678484 and MT678485, respectively. All raw sequence reads were deposited in NCBI's publicly available Short Read Archive under BioProject ID PRJNA624035.

Relative expression levels

We found that one of the full-length Type I AFP transcripts and the shorter contig of the LS-12-like gene transcripts were among the top <1% most highly expressed contigs in our transcriptome (Figure 6). Two of the truncated Type I AFP transcripts were also highly expressed (Supplemental Figure 3).

Discussion

The whole animal transcriptome from a juvenile variegated snailfish (*L. gibbus*) collected in Eastern Greenland revealed transcripts encoding expressed Type I AFPs and an LS-12-like protein that were among the most highly expressed transcripts

Figure 2. Study site showing the iceberg habitat in Greenland where *Liparis gibbus* was collected. The dive boat, a zodiac inflatable craft, can been seen at the bottom left, and the divers are visible near the center of the image. The image was shot using a small drone.

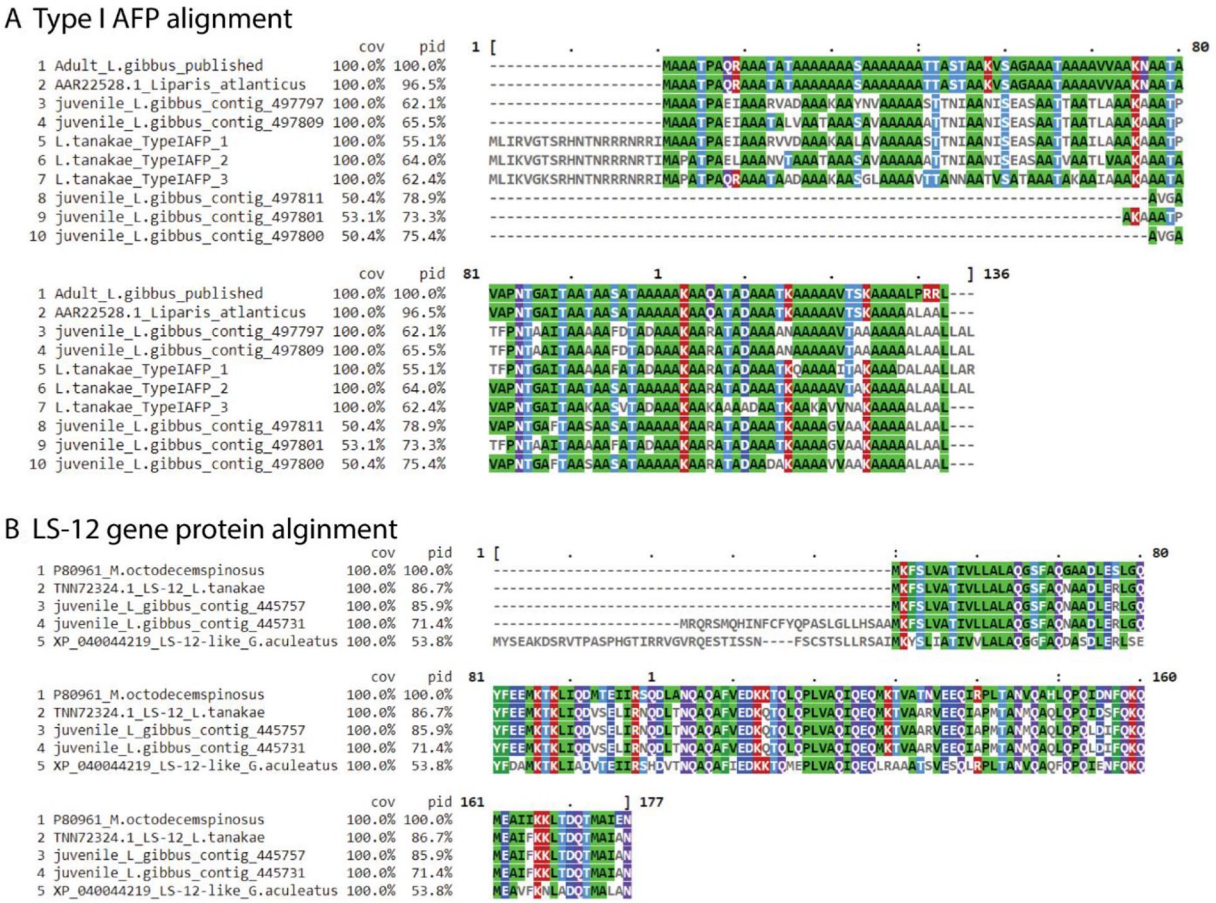

Figure 3. Protein alignments between observed juvenile *L. gibbus* Type I AFP and LS-12-like proteins and homologous proteins from the same or related species. (A) Type I AFPs. (B) Type IV AFPs. Colors correspond to amino acid identity where light and dark green indicates primarily hydrophobic residues, red indicates positively charged residues, light blue indicates polar-alcohol residues, purple indicates polar-carboxamide residues, and dark blue indicates negatively charged residues. No highlighting indicates the indicated residue is different from the reference sequence which is the top sequence in each alignment. The column "cov" indicates percent coverage of the reference sequence and "pid" indicates overall percent identity to the reference sequence.

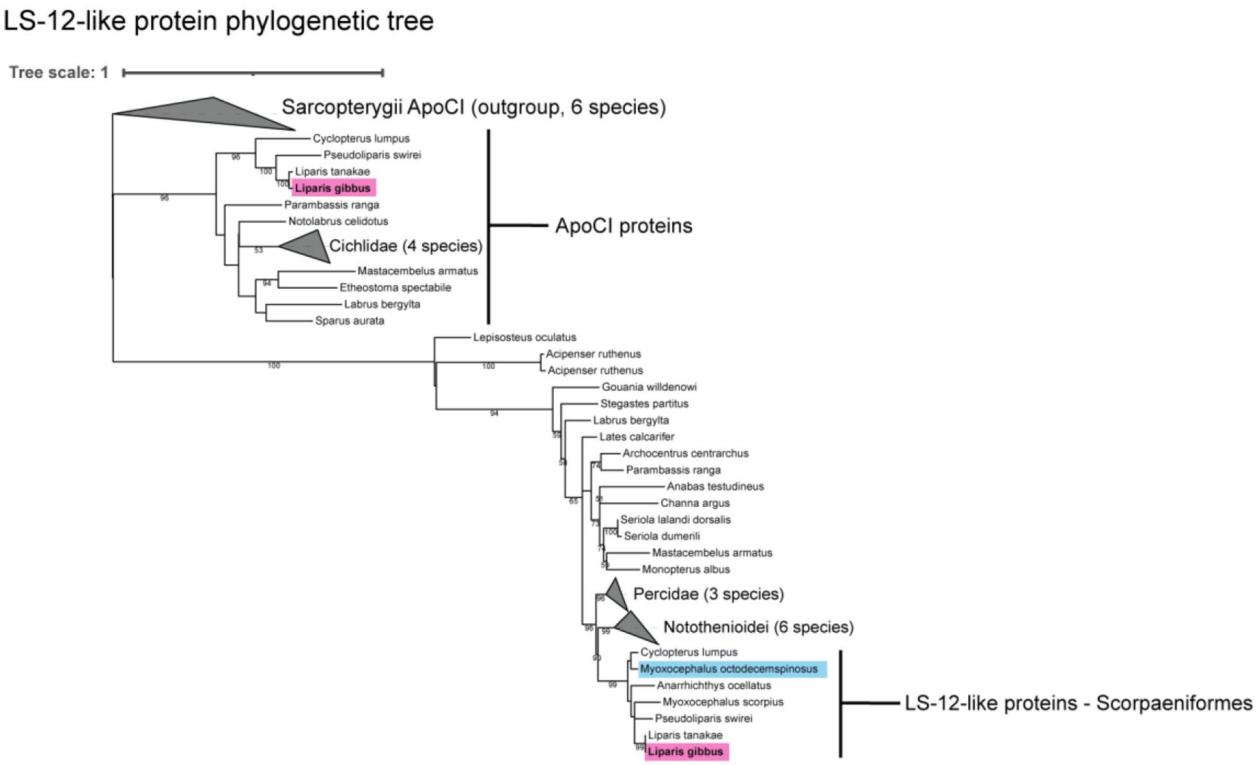

Figure 4. Phylogenetic tree depicting evolutionary relationships of the *L. gibbus* LS-12-like protein. The maximum likelihood tree was created in IQ-TREE using homologs of the longhorn sculpin, *M. octodecemspinous*, LS-12 protein (blue highlight). Support values are from 1000 ultrarapid bootstraps. *L. gibbus* sequences are highlighted in pink.

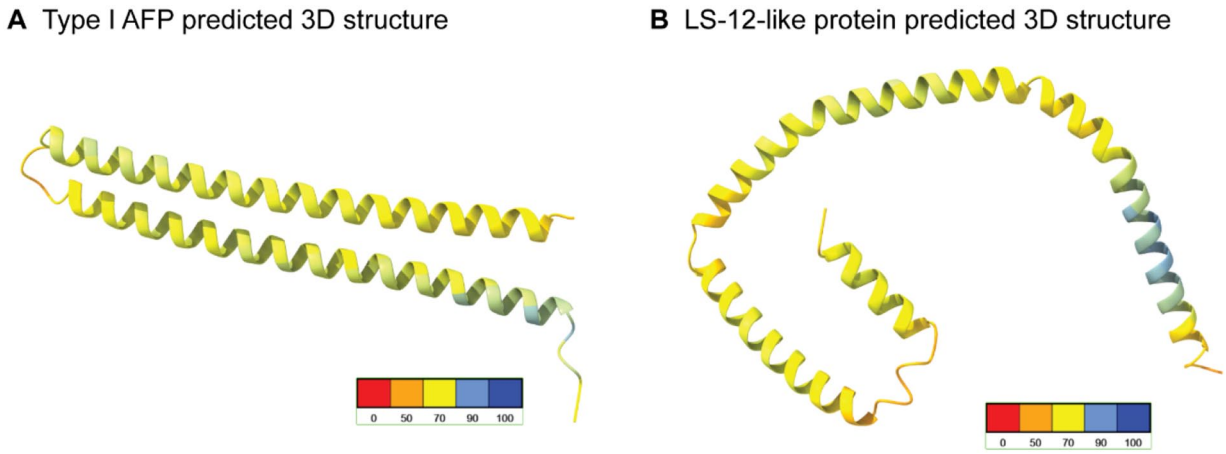

Figure 5. Structural models of the *Liparis gibbus* Type I AFP and LS-12-like protein. Models were built using Alphafold and rendered in ChimeraX. (A) Putative Type I AFP, consisting of 2 alpha helices joined by an unstructured region. (B) Putative LS-12-like protein, consisting of 3 alpha helices joined by unstructured regions. The color scale is based on the predicted Local Distance Difference Test (pLDDT) score for each residue. A yellow-blue color, as seen for the bulk of the 2 predictions, indicates a pLDDT > 70% which can be interpreted as confidence in an accurate representation of the backbone structure of the protein model. Position of individual atoms (not shown) are not predicted with confidence, and in the case of these 2 proteins likely depend on interactions with their respective substrates of ice crystals and potentially ice crystals and lipids.

in the animal (Figure 6). The Type I AFP sequences reported here are different from a previously reported Type I AFP from the same species (Figure 3), and are therefore likely expressed from a different genetic locus.

In an *L. tanakae* genome assembly (https://doi.org/10.6084/m9.figshare.7819418.v1), we found evidence of 3 genetic loci encoding full length Type I AFP genes and a single locus encoding an LS-12-like protein. The gene structure of the 2 types of genes is telling regarding their mode of inheritance and evolution. In the *L. tanakae* genome, all 3 detected Type I AFP genes lack introns and contain multiple non-synonymous differences indicative of selection on the genes. The multiple

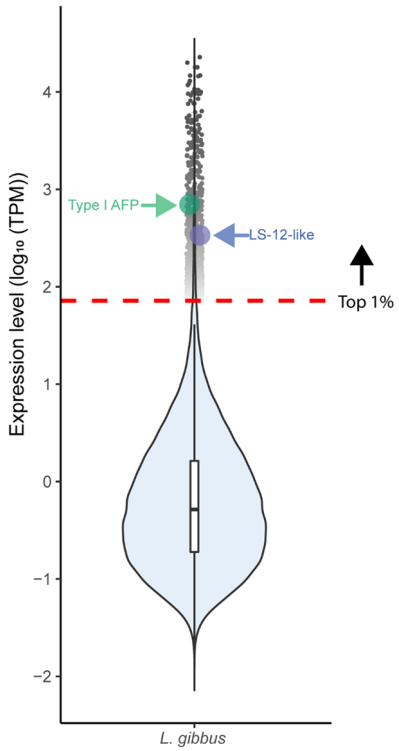

Figure 6. Expression plot indicating very high relative expression of a Type I AFP and the LS-12-like protein in the collected juvenile *L. gibbus* specimen. The Type I AFP and LS-12-like protein are among the top <1% most highly expressed genes in the *L. gibbus* transcriptome. Expression level, shown as the transcripts per million (TPM), the expected number of molecules of each transcript per million total transcripts, is plotted on a $\log_{10}$ scale. The width of the violin plot is representative of the number of genes at a given expression level.

loci of intron-less genes are indicative of recent gene duplications, possibly by retrotransposition. A BLAST search of the leading genomic sequence, 5′ to the start codon of the Type I AFP coding sequence, reveals that the leading sequence is found in multiple copies in other fish genomes (Supplemental Figure 4), but not followed by an AFP, further evidence that the Type I AFP is associated with a mobile genetic element. Active gene duplication is often correlated with gene loss of one or more copies of the gene[45] leading to multiple hypotheses regarding why we did not recover the exact previously reported *L. gibbus* Type I AFP sequence in our transcriptome. The gene could have been lost in the individual that we sequenced, or it could be repressed at this particular developmental stage, or it could be unassembled due to difficulties such as DNA repeats and RNA secondary structure.

The structure of the multiple copies of the Type I AFP gene stand in contrast to the genomic structure of the LS-12-like gene in *L. tanakae*. The LS-12-like gene is found at a single locus and contains introns. The intron/exon structure is also well conserved with related fish, like the hadal snailfish *P. swirei*, where the conserved exons are the same length with the same breakpoint in the 2 fish, but the introns, which are presumably under relaxed selection, vary in length and sequence (Figure 7). While some authors have hypothesized that the rapid expansion and multiple copies of Type I AFP genes in fish genomes explain high expression levels, it is worth noting that the LS-12 gene is present at a single, mature locus and is still among the <1% of expressed genes in the juvenile *L. gibbus* transcriptome. Transcriptional (eg, high transcription rate) and possibly post-transcriptional (eg, low degradation rate) regulation account for high expression levels in the absence of multiple gene copies.

High expression of a LS-12-like mRNA in a juvenile *L. gibbus* provides evidence that the protein product might be present at a high concentration. While a prior study concluded that LS-12-like Type IV AFPs do not exhibit strong enough antifreeze activity to be physiologically relevant, we argue that it is worth considering whether highly expressed Type I AFPs with strong antifreeze properties and LS-12-like proteins with weak antifreeze activity act synergistically to lower the freezing point of fish blood and tissues. To our knowledge a controlled experiment with a mixture of AFPs from different families (or putative families) has not been done. An alternative explanation is that LS-12 proteins do not contribute to antifreeze activity and are highly expressed for a reason other than promoting freeze tolerance. For example, in a freshwater trout that experiences seasonal variation in temperature, the proportion of unsaturated lipids carried by high-density lipoproteins (HDLs) increases with a decrease in temperature.[46] Since LS-12-like proteins are structural analogs of Apo-A1 lipoproteins[44] and Apo-A1 lipoproteins are the major component of HDLs,[47] the very high expression of the *L. gibbus* LS-12-like gene on an iceberg may be related to a temperature dependent transport of unsaturated lipids rather than, or in addition to the reported weak antifreeze activity of the protein.

Since the mid-20th century, temperatures have increased twice as fast in the Arctic as in mid-latitudes and it is predicted that if Arctic sea ice decline continues at this current rate, it will be ice-free by the next 3 decades.[48,49] Arctic seas are inhabited by only ~270 fish species,[50,51] with 15 endemic marine fish species, mostly sculpins, snailfishes and eelpouts.[52] We hypothesize that with increasingly warming oceanic temperatures in the Arctic,[52] ice-dwelling specialists such as *L. gibbus* and the few other teleosts that use substantial metabolic resources to generate AFPs and regulate lipid composition may encounter increased competition by more temperate species that were previously unable to survive in sub-zero seawater temperatures.[53]

Materials and Methods

Collection

The juvenile *Liparis gibbus* specimen (18 mm SL) was collected on August 26, 2019 in an iceberg habitat, near Qaattu (65.788, −37.879) in Eastern Greenland, at 15 m depth and was preserved in RNA Stabilization Solution (RNAlater, Qiagen) for transport back to the laboratory (Figures 1 and 2). The specimen was collected via SCUBA using a hand-net, via the

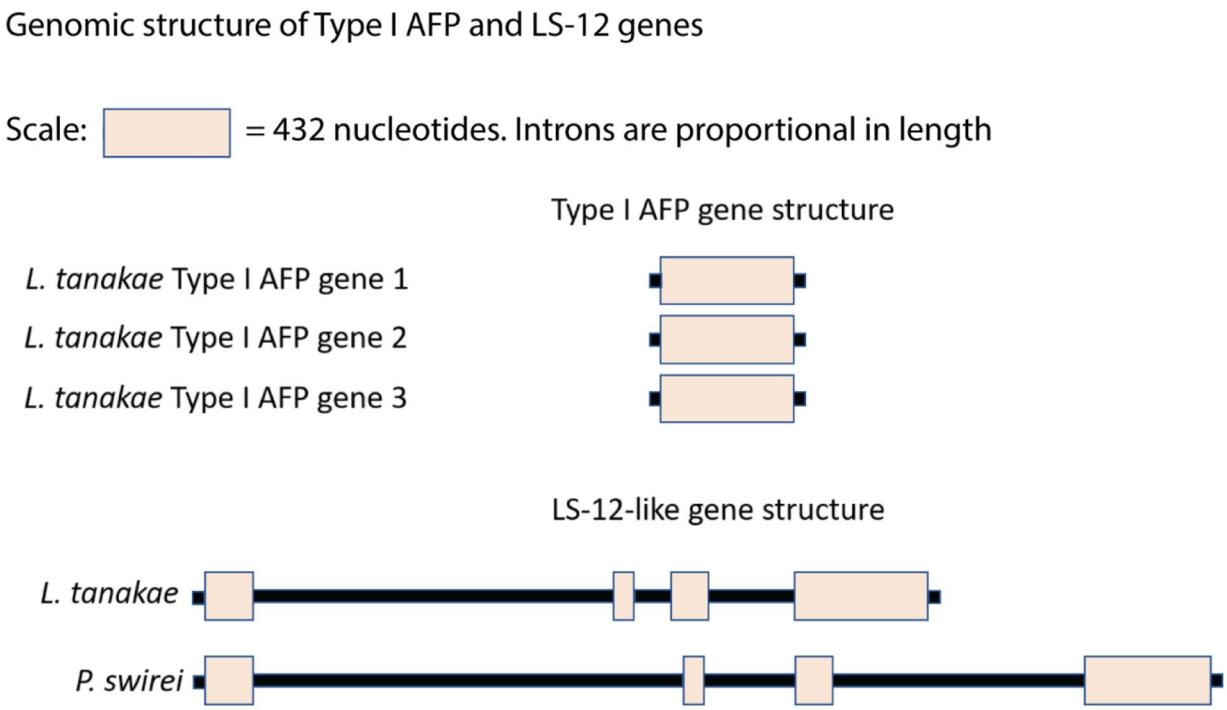

Figure 7. Genomic structure of Type I AFPs and LS-12 like genes. Three copies of a Type I AFP gene were found in 1 of 2 available *L. tanakae* whole genome assemblies. The genes were on separate contigs and all had the same intronless structure indicative of recent gene duplications, possibly by retrotransposition. In contrast, the LS-12-like gene contains introns, has a conserved structure in 2 fish genomes, and is single copy, indicative of a stable, conserved gene product.

application of quinaldine, an anesthetic used to temporarily immobilize fishes for capture, to a targeted habitat. This study was approved and carried out in strict accordance with the recommendations in the Guidelines for the Use of Fishes in Research of the American Fisheries Society and the American Museum of Natural History's Institutional Animal Care and Use Committee (IACUC).

Analysis

Given the small size of the snailfish, total RNA was extracted from the whole body (including muscle and skin) except the gut, which was dissected out to limit contamination with DNA from food. Illumina NovaSeq 2 × 150 bp sequencing and initial bioinformatic analyses were performed by Genewiz (South Plainfield, NJ). The NGS run yielded 281 412 170 reads (84 424 megabases) with a mean Q score of 35.71 (92.79% of bases ⩾*Q30*). Trimmomatic v0.36[54] was used to trim adapters and quality trim the raw data. Trinity v2.5[55] was used to assemble the transcriptome de-novo with a minimum contig length of 200 bp. SeqKit v0.11.0[56] was used to generate statistics for the transcriptome (Supplemental Table S1). Transcript quantification was accomplished by alignment and quantification of sequencing reads against the transcriptome assembly using Salmon (v1.2.0).[57] For plotting the transcript abundance distribution, contigs that recruited low numbers of reads (⩽100 reads) were removed. The threshold was determined empirically by plotting the number of contigs remaining after removing contigs below threshold. The curve flattened out at 100 reads mapped (Supplemental Figure 5). Relative abundance of transcriptome contigs was expressed using transcripts per million (TPM). The abundance distribution was visualized in R (v3.6.3)[58] using the ggplot2 package.[59]

The EMBOSS tool getorf was used to find open reading frames. The transcriptome was annotated using Diamond BLASTX. Any transcripts that were identified by Diamond BLASTX as AFPs were reciprocally blasted to nr and UniProtKB/Swiss-Prot using BLASTP. The results of the reciprocal blasts are presented. To discover additional putative AFPs, representative sequences of Type I (NCBI nucleotide: AY455862.2), Type II (UniProtKB: P84493), Type III (UniProtKB: P19614), Type IV (UniProtKB: P80961) AFPs, and AGFP (UniProtKB: P24856) were searched against the transcriptome assembly using BLASTN (nucleotide query, nucleotide database) and tBLASTN (protein query, nucleotide database) searches. The dual search strategy was needed because Type I AFPs contain repetitive amino acid motifs of 2 to 4 alanine residues in a row repeated throughout the length of the peptide sequence (65 out of 115 residues are alanine). The repetitive nature of the protein sequence leads to poor BLAST search scores. Due to the degenerate nature of the protein code, however, the nucleotide sequence that codes for the Type I AFP is not as repetitive and specific, and significant hits can be identified where present. In addition to the general BLAST search using known AFPs as a query, the transcripts in the top 1% of expression levels were searched against nr using

BLASTX (nucleotide query, protein database) to discover hits to highly expressed transcripts containing c-type lectin or apolipoprotein domains, from which AFPs have evolved in other species.

Alignments to known AFP sequences were completed using MUSCLE v3.8[60] and visualized with MView v1.63[61] accessed via EBI Web Services.[62]

Phylogenetic trees

Homologs of select *L. gibbus* transcripts and relevant antifreeze proteins were collected by BLASTP searches of potential *L. gibbus* AFPs against the NCBI database nr[63] for Type IV and Type II AFPs. The top 50 BLAST hits were recovered for phylogenetic analysis. For an outgroup, the seed proteins were searched against nr limited to Chondrichthyes sequences (cartilaginous fishes, NCBI taxon ID 7777). Canonical antifreeze protein sequences were included in the BLAST hits by searching nr using antifreeze proteins as a seed: Uniprot protein P80961 was used as a seed for LS-12-like/Type IV AFPs. For Type I AFPs, nucleotide searches using BLAST were conducted using AY455863 as a seed, and Type I AFP sequences were collected manually from the NCBI nucleotide collection by searching with the terms: "type I" antifreeze protein.

Sequences were aligned using MAFFT v7.310,[64] the alignments were trimmed with trimAl v1.2rev59,[65] ModelTest-NG v0.1.7[66] was used to find best-fit models for each alignment (Type IV: LG), and the trimmed alignments were used to build a maximum likelihood phylogenetic tree in IQ-TREE 2[67] using 1000 ultrafast bootstraps. Sequence alignments are available in respective Supplemental Files (typeIV_apoCI.trimal).

Molecular modeling

An Alphafold2 colab notebook[68] was used to generate the protein models of all putative AFPs from *Liparis gibbus*. ChimeraX[69] was used to render models.

Author Contributions

DFG and JSS performed collection of the specimen. DFG, MRB, and JSS conceived of the study. JAB, JPG, and MRB completed methods development. JAB, DFG, MRB, JPG, and JSS carried out the investigation. MRB, JAB, DFG, and JPG wrote the original draft. JAB, DFG, JPG, MRB, and JSS contributed to manuscript review and editing. JSB, JPG, DFG, and JSS contributed to data visualization.

ORCID iDs

David F Gruber https://orcid.org/0000-0001-9041-2911
Mercer R Brugler https://orcid.org/0000-0003-3676-1226

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Evans DH. Teleost fish osmoregulation: what have we learned since August Krogh, Homer Smith, and Ancel Keys. *Am J Physiol Regul Integr Comp Physiol*. 2008;295:R704-R713.
2. Ahlgren JA, Cheng CC, Schrag JD, DeVries AL. Freezing avoidance and the distribution of antifreeze glycopeptides in body fluids and tissues of Antarctic fish. *J Exp Biol*. 1988;137:549-563.
3. Storey KB, Storey JM, Brooks SP, Churchill TA, Brooks RJ. Hatchling turtles survive freezing during winter hibernation. *Proc Natl Acad Sci U S A*. 1988;85:8350-8354.
4. Storey KB, Storey JM. Freeze tolerance in animals. *Physiol Rev*. 1988;68:27-84.
5. DeVries AL. Glycoproteins as biological antifreeze agents in Antarctic fishes. *Science*. 1971;172:1152-1155.
6. Carroll SB. *Into the Jungle: Great Adventures in the Search for Evolution*. 2009.
7. Cheung RCF, Ng TB, Wong JH. Antifreeze proteins from diverse organisms and their applications: an overview. *Curr Protein Pept Sci*. 2017;18:262-283.
8. Chakraborty S, Jana B. Antifreeze proteins: an unusual tale of structural evolution, hydration and function. *Proc Indian Natl Sci Acad*. 2019;100:169-187.
9. Clarke CJ, Buckley SL, Lindner N. Ice structuring proteins - a new name for antifreeze proteins. *Cryo-Letters*. 2002;23:89-92.
10. Davies PL, Hew CL, Fletcher GL. Fish antifreeze proteins: physiology and evolutionary biology. *Can J Zool*. 1988;66:2611-2617.
11. Barrett J. Thermal hysteresis proteins. *Int J Biochem Cell Biol*. 2001;33:105-117.
12. Nath A, Chaube R, Subbiah K. An insight into the molecular basis for convergent evolution in fish antifreeze proteins. *Comput Biol Med*. 2013;43:817-821.
13. Graham LA, Hobbs RS, Fletcher GL, Davies PL. Helical antifreeze proteins have independently evolved in fishes on four occasions. *PLoS One*. 2013;8:e81285.
14. Nishimiya Y, Kondo H, Yasui M, et al. Crystallization and preliminary X-ray crystallographic analysis of Ca^{2+}-independent and Ca^{2+}-dependent species of the type II antifreeze protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2006;62:538-541.
15. Graham LA, Lougheed SC, Ewart KV, Davies PL. Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PLoS One*. 2008;3:e2616.
16. Pettersson ME, Rochus CM, Han F, et al. A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Res*. 2019;29:1919-1928.
17. Graham LA, Davies PL. Horizontal gene transfer in vertebrates: a fishy tale. *Trends Genet*. 2021;37:501-503.
18. Baardsnes J, Davies PL. Sialic acid synthase: the origin of fish type III antifreeze protein? *Trends Biochem Sci*. 2001;26:468-469.
19. Deng G, Andrews DW, Laursen RA. Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin *Myoxocephalus octodecimspinosis*. *FEBS Lett*. 1997;402:17-20.
20. Zhuang X, Yang C, Murphy KR, Cheng CC. Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *Proc Natl Acad Sci U S A*. 2019;116:4400-4405.
21. Chen L, DeVries AL, Cheng CHC. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A*. 1997;94:3811-3816.
22. Baalsrud HT, Tørresen OK, Solbakken MH, et al. De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole Genome Sequence Data. *Mol Biol Evol*. 2018;35:593-606.
23. Cheng CH. Evolution of the diverse antifreeze proteins. *Curr Opin Genet Dev*. 1998;8:715-720.
24. Cheng CHC, Zhuang X. Molecular origins and mechanisms of fish antifreeze evolution. In: Ramløv H, Friis D, eds. *Antifreeze Proteins Volume 1: Environment, Systematics and Evolution*. Springer; 2020;275-313.
25. Gauthier SY, Scotter AJ, Lin FH, Baardsnes J, Fletcher GL, Davies PL. A re-evaluation of the role of type IV antifreeze protein. *Cryobiology*. 2008;57:292-296.
26. Lee JK, Kim YJ, Park KS, et al. Molecular and comparative analyses of type IV antifreeze proteins (AFPIVs) from two Antarctic fishes, Pleuragramma antarcticum and Notothenia coriiceps. *Comp Biochem Physiol B Biochem Mol Biol*. 2011;159:197-205.
27. Xiao Q, Xia JH, Zhang XJ, et al. Type-IV antifreeze proteins are essential for epiboly and convergence in gastrulation of zebrafish embryos. *Int J Biol Sci*. 2014;10:715-732.
28. Ammar AY, El Nahas AF, Mahmoud S, Barakat ME, Hassan AM. Characterization of type IV antifreeze gene in Nile tilapia (Oreochromis niloticus) and influence of cold and hot weather on its expression and some immune-related genes. *Fish Physiol Biochem*. 2018;44:515-525.
29. Li SZ, Liu W, Li Z, et al. Molecular characterization and expression pattern of a germ cell marker gene dnd in gibel carp (Carassius gibelio). *Gene*. 2016;591:183-190.
30. Deng G, Laursen RA. Isolation and characterization of an antifreeze protein from the longhorn sculpin, Myoxocephalus octodecimspinosis. *Biochim Biophys Acta Protein Struct Mol Enzymol*. 1998;1388:305-314.

31. Lee JK, Kim HJ. Cloning, expression, and activity of type IV antifreeze protein from cultured subtropical olive flounder (Paralichthys olivaceus). *Fish Aquat Sci*. 2016;19:7.
32. Low WK, Lin Q, Stathakis C, Miao M, Fletcher GL, Hew CL. Isolation and characterization of skin-type, type I antifreeze polypeptides from the Longhorn Sculpin, Myoxocephalus octodecemspinosus. *J Biol Chem*. 2001;276:11582-11589.
33. Davies PL, Fletcher GL, Hew CL. Fish antifreeze protein genes and their use in transgenic studies. *Oxf Surv Eukaryot Genes*. 1989;6:85-109.
34. Wang K, Shen Y, Yang Y, et al. Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. *Nat Ecol Evol*. 2019;3:823-833.
35. Linley TD, Gerringer ME, Yancey PH, Drazen JC, Weinstock CL, Jamieson AJ. Fishes of the hadal zone including new species, in situ observations and depth records of Liparidae. *Deep Res Part I Oceanogr Res Pap*. 2016;114: 99-110.
36. Able KW. A revision of Arctic snailfishes of the genus Liparis (Scorpaeniformes: Cyclopteridae). *Copeia*. 1990;1990:476-492.
37. Scott MG, Scott WB. *Atlantic Fishes of Canada*. 1st ed. University of Toronto Press; 1988.
38. Eschmeyer WN, Herald ES. *A Field Guide to Pacific Coast Fishes: North America*. Houghton Mifflin Harcourt; 1999.
39. Evans RP. *Characterization of Skin and Plasma Type I Antifreeze Proteins From Atlantic (Liparis atlanticus) and Dusky (Liparis gibbus) Snailfish*. Memorial University of Newfoundland; 2003.
40. Gruber DF, Sparks JS. First Report of biofluorescence in Arctic snailfishes and rare occurrence of multiple fluorescent colors in a single species. *Am Mus Novit*. 2021;2020:1-12.
41. Evans RP, Fletcher GL. Type I antifreeze proteins expressed in snailfish skin are identical to their plasma counterparts. *FEBS J*. 2005;272:5327-5336.
42. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589.
43. Sicheri F, Yang DS. Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature*. 1995;375:427-431.
44. Mao MG, Chen Y, Liu RT, et al. Transcriptome from Pacific cod liver reveals types of apolipoproteins and expression analysis of AFP-IV, structural analogue with mammalian ApoA-I. *Comp Biochem Physiol Part D Genomics Proteomics*. 2018;28:204-212.
45. Hurles M. Gene duplication: the genomic trade in spare parts. *PLoS Biol*. 2004;2:e206.
46. Wallaert C, Babin PJ. Circannual variation in the fatty acid composition of high-density lipoprotein phospholipids during acclimatization in trout. *Biochim Biophys Acta Lipids Lipid Metab*. 1993;1210:23-26.
47. Melchior JT, Street SE, Vaisar T, et al. Apolipoprotein A-I modulates HDL particle size in the absence of apolipoprotein A-II. *J Lipid Res*. 2021;62:100099.
48. Polyak L, Alley RB, Andrews JT, et al. History of sea ice in the Arctic. *Quat Sci Rev*. 2010;29:1757-1778.
49. Yadav J, Kumar A, Mohan R. Dramatic decline of Arctic sea ice linked to global warming. *Nat Hazards*. 2020;103:2617-2621.
50. Møller PR, Nielsen JG, Knudsen SW, et al. A checklist of the fish fauna of Greenland waters. *Zootaxa*. 2010;2378:1-84.
51. Mecklenburg CW, Møller PR, Steinke D. Biodiversity of arctic marine fishes: Taxonomy and zoogeography. *Mar Biodivers*. 2011;41:109-140.
52. Christiansen JS, Reist JD. Chapter 6 Fishes. In: Meltofte H ed. *Arctic Biodiversity Assessment*. Arctic Council; 2013;193-245.
53. Hew CL, Fletcher GL, Davies PL. Transgenic salmon: tailoring the genome for food production. *J Fish Biol*. 1995;47:1-19.
54. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
55. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol Nat Biotechnol*. 2011;29:644-652.
56. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*. 2016;11:e0163962.
57. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417-419.
58. R Core Team. *A Language and Environment for Statistical Computing*. 2020; R Foundation for Statistical Computing.
59. Wickham H. *Ggplot2 Elegant Graphics for Data Analysis*. Springer; 2009.
60. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113.
61. Brown NP, Leroy C, Sander C. MView: A web-compatible database search or multiple alignment viewer. *Bioinformatics*. 1998;14:380-381.
62. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47:W636-W641.
63. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44:D733-D745.
64. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772-780.
65. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972-1973.
66. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol*. 2020;37:291-294.
67. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530-1534.
68. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19:679-682.
69. Pettersen EF, Goddard TD, Huang CC, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*. 2021;30:70-82.