



OPEN

Origin of an antifreeze protein gene in response to Cenozoic climate change

Laurie A. Graham, Sherry Y. Gauthier & Peter L. Davies

Antifreeze proteins (AFPs) inhibit ice growth within fish and protect them from freezing in icy seawater. Alanine-rich, alpha-helical AFPs (type I) have independently (convergently) evolved in four branches of fishes, one of which is a subsection of the righteye flounders. The origin of this gene family has been elucidated by sequencing two loci from a starry flounder, *Platichthys stellatus*, collected off Vancouver Island, British Columbia. The first locus had two alleles that demonstrated the plasticity of the AFP gene family, one encoding 33 AFPs and the other allele only four. In the closely related Pacific halibut, this locus encodes multiple *Gig2* (antiviral) proteins, but in the starry flounder, the *Gig2* genes were found at a second locus due to a lineage-specific duplication event. An ancestral *Gig2* gave rise to a 3-kDa “skin” AFP isoform, encoding three Ala-rich 11-a.a. repeats, that is expressed in skin and other peripheral tissues. Subsequent gene duplications, followed by internal duplications of the 11 a.a. repeat and the gain of a signal sequence, gave rise to circulating AFP isoforms. One of these, the “hyperactive” 32-kDa Maxi likely underwent a contraction to a shorter 3.3-kDa “liver” isoform. Present day starry flounders found in Pacific Rim coastal waters from California to Alaska show a positive correlation between latitude and AFP gene dosage, with the shorter allele being more prevalent at lower latitudes. This study conclusively demonstrates that the flounder AFP arose from the *Gig2* gene, so it is evolutionarily unrelated to the three other classes of type I AFPs from non-flounders. Additionally, this gene arose and underwent amplification coincident with the onset of ocean cooling during the Cenozoic ice ages.

Ocean waters freeze near $-2\text{ }^{\circ}\text{C}$, but fish blood and lymph is less salty and freezes at around $-0.8\text{ }^{\circ}\text{C}$ ¹. Any contact with ice in seawater increases the freezing risk, so some fishes produce antifreeze proteins (AFPs) or antifreeze glycoproteins (AFGPs)^{2–4}. These AF(G)Ps bind to the surface of ice crystals, preventing growth and decreasing the non-equilibrium freezing point to below $-2\text{ }^{\circ}\text{C}$ ^{5,6}. As a result, any internal ice crystals that arise due to contact through the skin, gills or alimentary canal remain small in a quasi-stable supercooled state⁷, thereby allowing the fish to live in an icy ecosystem.

Four different types of fish AF(G)Ps, type I, II and III as well as AFGP, occur in species within the clade Percomorpha (Fig. 1). Both type I and type III AFPs are restricted to this clade. Type I AFPs are found within four groups within three separate orders (Perciformes^{8,9}, Labriformes¹⁰ and Pleuronectiformes^{11–14}), interspersed with groups producing the three other AFP types. This patchy taxonomic distribution was attributed to convergent evolution of these Ala-rich alpha-helical peptides, but their origins were not known^{10,15}. Type II AFPs arose from a lectin-like precursor¹⁶, but the presence of this globular, non-repetitive protein in three distantly related fish groups that diverged over 200 million years ago (Ma) (Fig. 1), came about via horizontal gene transfer (HGT)¹⁷ rather than convergence. Type III appears to have arisen only once, within infraorder Zoarcales, from a domain of sialic-acid synthase^{18–20}. Finally, the AFGPs, composed primarily of simple Ala-Ala-Thr repeats where the Thr is glycosylated, arose convergently in northern cods (not shown) and Antarctic notothenioid fishes (such as the Antarctic toothfish) from non-coding DNA and a *trypsinogen* gene, respectively^{21–23}.

The appearance of these different AF(G)Ps within various groups of fish is correlated with past climate history (Fig. 1). After the warming period culminated by the Paleocene–Eocene thermal maximum at 55 Ma (red on color bar), when the oceans were perpetually ice-free^{27,29,30}, fish would have had no need of AF(G)Ps for many Ma, and if they were present in prior epochs, they were likely lost. Southern glaciation commenced ~ 34 Ma (blue on color bar), but continental-scale northern glaciation lagged by ~ 30 Ma, beginning at ~ 3 Ma²⁷. Nevertheless, there is evidence for sea ice and localized ephemeral northern glaciation far earlier, roughly coincident

Department of Biomedical and Molecular Sciences, Queen’s University, Kingston, ON K7L 3N6, Canada. ✉email: daviesp@queensu.ca

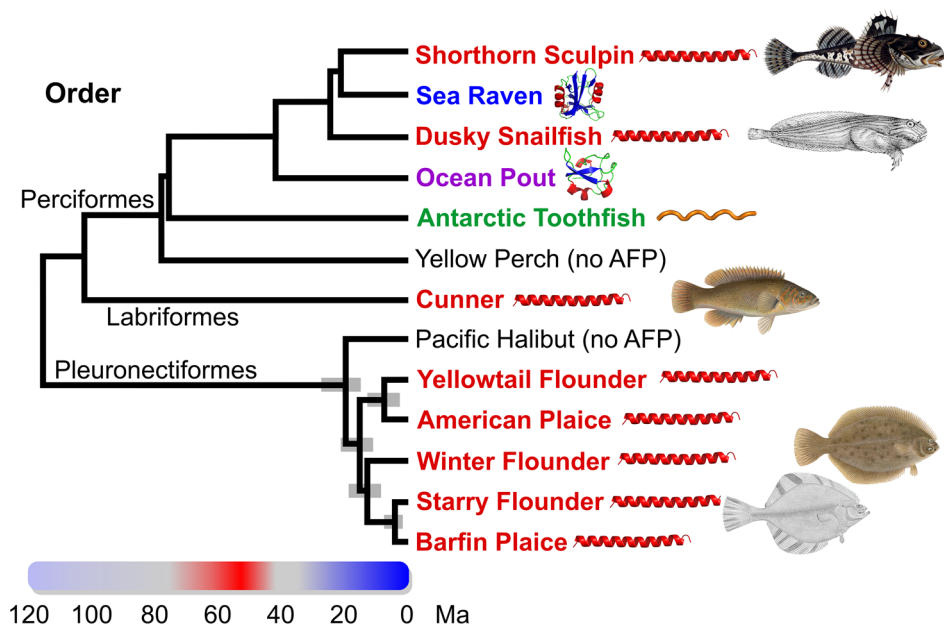


Figure 1. Phylogenetic relationships amongst type I AFP-producing fishes and several other species within the clade, Percomorpha, that produce different AFPs^{24,25}. The common name of species that produce AFPs are coloured red (type I), blue (type II), purple (type II) or green (antifreeze glycoprotein). The 95% highest posterior credibility intervals within the Pleuronectiformes are indicated with grey bars²⁴. Pacific halibut and yellow perch (black) do not produce AFPs²⁶. The coloured bar spanning 120 Ma indicates relative ocean temperatures with red corresponding to ice-free oceans and blue corresponding to glacial periods²⁷. Schematics of the AFP types were generated in PyMOL²⁸ and fish images/drawings for shorthorn sculpin, dusky snailfish, cunner, winter flounder and starry flounder are from Wikimedia Commons (see Supplementary Material and Methods). Binomial names for the species are as follows; *Myoxocephalus scorpius* (shorthorn sculpin), *Hemirhamphus intermedius* (sea raven), *Liparis gibbus* (dusky snailfish), *Zoarces americanus* (ocean pout), *Dissostichus mawsoni* (Antarctic toothfish), *Perca flavescens* (yellow perch), *Tautoglabrus adspersus* (cunner), *Hippoglossus stenolepis* (Pacific halibut), *Limanda ferruginea* (yellowtail flounder), *Hippoglossoides platessoides* (American plaice), *Pseudopleuronectes americanus* (winter flounder), *Platichthys stellatus* (starry flounder), *Pleuronectes pinnifasciatus* (barfin plaice). Some species were not analyzed in the studies cited above, so the position of the following species with the same genus (dusky snailfish, Antarctic toothfish, Pacific halibut, yellowtail flounder, American plaice, barfin plaice) or family (sea raven) was used as a proxy. Other fish, including Atlantic herring and rainbow smelt that are outside Percomorpha, also produce AFPs.

with southern glaciation^{27,31}. The patchy distribution of AF(G)P types in groups that diverged prior to 20 Ma is consistent with the hypothesis that these proteins arose anew, allowing these species to inhabit a new icy niche as cooling intensified. It is only within recently diverged groups, such as the type I AFP-producing Pleuronectiformes, that AFPs are homologous due to descent from a common ancestor.

Type I AFPs have been best characterized in the winter flounder. There are three isoform classes, all of which are Ala-rich, with Thr appearing at 11 a.a. intervals¹⁵. The abundant small serum isoform HPLC6, produced primarily by the liver (hereafter called a liver isoform), is processed by removal of the secretory signal peptide and pro-region, plus removal of the C-terminal Gly during amidation³². The mature 37-a.a. peptide is 62% Ala by content and forms a single α helix with three 11 a.a. repeats, delineated by four evenly-spaced Thr residues that lie along one side of the peptide³³. Subsequently, a second class was isolated from skin, hereafter called skin isoforms, although they are expressed in a variety of tissues. These 37–40 a.a. isoforms lack a signal peptide, and their only modification is acetylation of the N-terminal Met residue³⁴. The third isoform is the much larger hyperactive AFP, hereafter called Maxi, whose only modification is removal of the signal peptide^{35,36}. This 195 a.a. α -helical peptide folds in half to form an antiparallel homodimer via clathrate water interactions³⁷.

The presence of type I AFPs in four groups within Percomorpha (Fig. 1) could potentially be explained by the presence of the gene in the common ancestor of these groups, followed by its loss in most branches and subsequent gain of different AFPs in a subset of branches. The 76% sequence identity between a winter flounder skin isoform and a longhorn sculpin isoform would seem to support this hypothesis¹⁵. However, other type I AFPs are far less like those from flounders, including the 113-residue dusky snailfish AFP which lacks the 11-a.a. Thr periodicity⁸. Additionally, the stark differences in the Ala codon usage in the AFP genes of three of the four groups and the complete lack of similarity of their non-coding sequences led to the hypothesis that they arose via convergent evolution^{10,15}. Convergence of the AFGPs in northern and southern fishes has been clearly demonstrated following the determination of their progenitors as mentioned above^{21–23}, but until now such analysis was lacking for any of the type I AFPs.

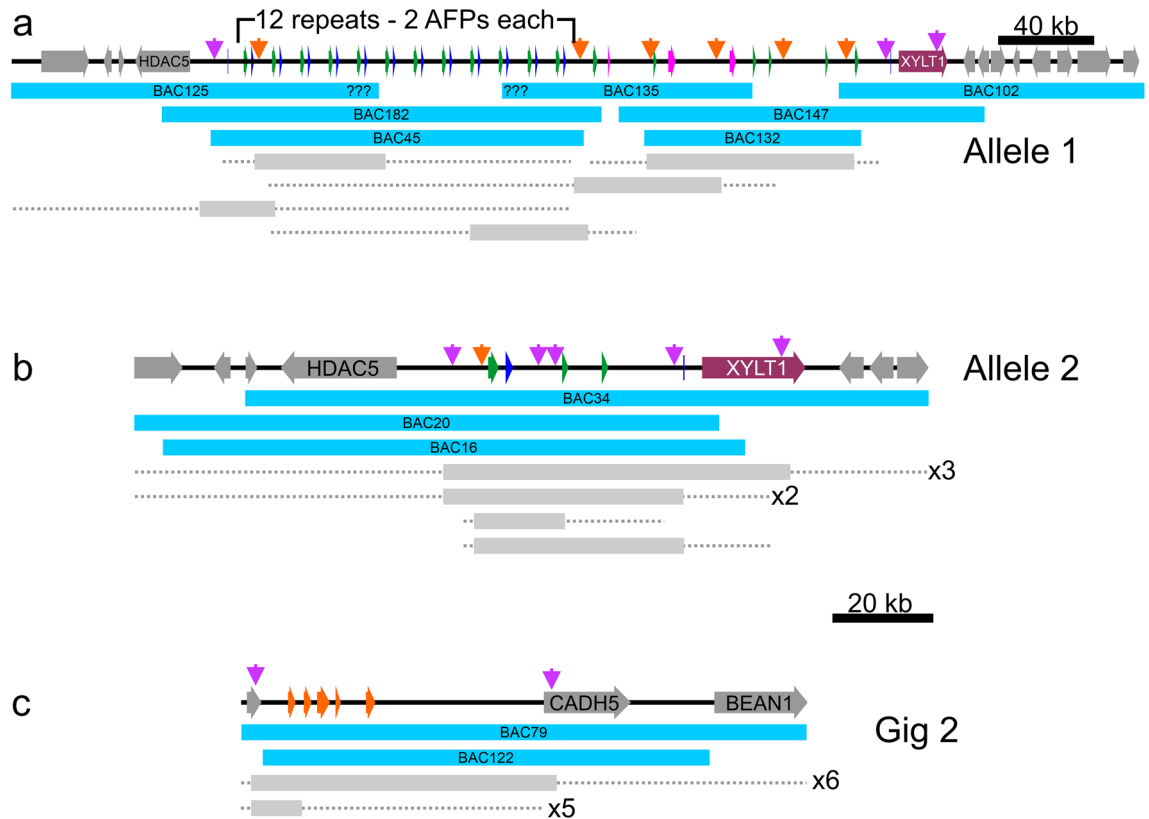


Figure 2. Schematic diagram of BAC clones which overlap (a) AFP allele 1, (b) AFP allele 2 and (c) the Gig2 locus. The 33 AFP genes in allele 1 and the four in allele 2 are indicated in blue (liver isoforms), green (skin isoforms) and pink (intermediate length “Midi” isoform and long “Maxi” isoforms). The deduced number of tandem repeats is indicated for allele 1. The sequenced BAC clones are indicated with cyan bars. The span of other BAC clones (grey bars with dashed lines indicating uncertainty) were determined by PCR using location-specific primers (purple arrows) and primers that were location and allele specific (orange arrows) (Supplementary Table 1). All clones were PCR positive using primers specific to the 3’ UTR found in both the Gig2 and AFP genes.

The starry flounder, *Platichthys stellatus*, is a flatfish that inhabits shallow waters of the Northern Pacific Ocean from South Korea, up through the Bering Sea and down to California, as well as portions of the Arctic Ocean^{38,39}. It is known to produce type I AFPs, but their sequences were previously unknown^{40,41}. Loci containing AFP-like sequences were cloned from BAC libraries and both AFPs and the progenitor gene, *Gig2* (grass carp reovirus-induced gene 2)⁴², were identified. Similarity between the loci is restricted to non-coding regions and *Gig2* has a different function, related to viral resistance⁴³. This demonstrates that the AFPs of Pleuronectiformes arose recently and independently of the type I AFPs of other fishes. The two alleles at the AFP locus are very different, containing 4 and 33 AFPs with Southern blotting demonstrating that gene copy number increases with latitude.

Results

Part 1: flounder loci. *Starry flounder AFP genes reside at a single locus.* Two BAC libraries made from a single starry flounder caught off Vancouver Island, British Columbia were screened using a probe to the well-conserved 3’ UTRs found in flounder AFPs. The tiling paths of 35 positive BACs were determined by PCR screening with a variety of primers (Fig. 2, Supplementary Table 1) and corresponded to two loci. The first locus was represented by 22 clones corresponding to two remarkably divergent alleles from a single multigene AFP locus (Fig. 2a,b). The two banks of AFPs are allelic as they share the same four flanking genes on each side, including those coding for collagen type 1, $\alpha 1$ (COL1A1) and histone deacetylase 5 (HDAC5) on the upstream side and xylosyltransferase 1 (XYLT1) downstream. The remaining 13 clones contained five closely spaced *Gig2* genes (Fig. 2c) with partial sequence similarity to AFPs. Based on the starry flounder genome size obtained from the Animal Genome Size Database (6.5×10^8 bp) (<http://www.genomesize.com/index.php>) this is consistent with a single gene locus. The greater number of clones for the AFP locus is consistent with the AFPs spanning a much larger DNA length (31 or 240 kb) than the *Gig2* locus (17 kb) (Fig. 2).

The two AFP alleles contain a vastly different number of AFP genes. The number of genes within both copies of the locus from this single fish differ greatly as one allele contains 33 AFPs, whereas the smaller contains only four (Fig. 3a). The difference between the two alleles is not a cloning artefact for two reasons. First, multiple BAC inserts were sequenced for each allele (Fig. 2a,b), and they were exact matches where they overlapped. Second,

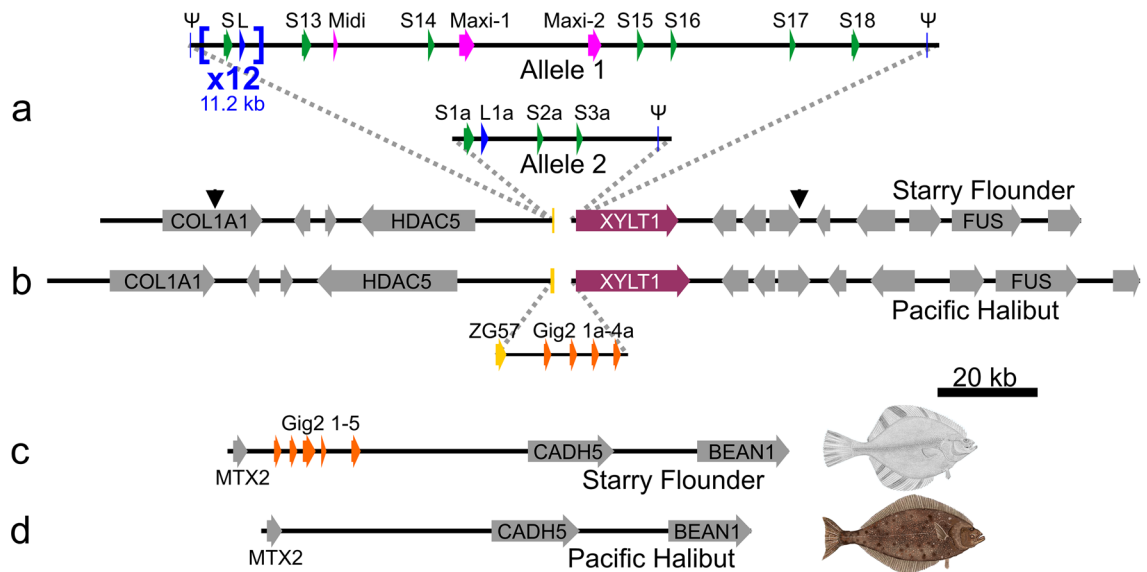


Figure 3. Low-resolution schematic of the *AFP* and *Gig2* loci of starry flounder and Pacific halibut. A solid arrow spans each gene, across all exons and introns, from the start to stop codon, except for the *AFP* and *Gig2* genes where non-coding exons were included in the span. Syntenic genes that are not germane to the evolution of the *AFP* are in grey with the acronyms of the shorter genes omitted. All schematics are at the same scale. (a) The *AFP* locus from the single fish used to generate the BAC library is shown with the *AFP*-containing segment that differs from Pacific halibut and between the two alleles shown as a pop out. The *AFP* genes are colored as in Fig. 2 and are numbered sequentially by type. The *ZG57* gene that was partially deleted at this location is in dark yellow and the *XYLT1* gene is in maroon. The first 24 *AFP* genes (12 liver and 12 skin) occur in pairs within twelve nearly identical tandem repeats that are each 11.2 kb in length (shown compressed to one repeat $\times 12$). These are flanked by two short segments (Ψ) that are highly similar to portions of the *AFP* genes. The second locus contains four *AFP*s denoted with the suffix “a” and one pseudogene. The black arrows show the boundaries of the locus 2 assembly. (b) The segment of Pacific halibut DNA on chromosome 16 that corresponds to the *AFP* locus. The pop out shows the region that differs with respect to the starry flounder locus with the four *Gig2* genes shown in orange. (c) The *Gig2* locus of starry flounder with five *Gig2* genes and (d) the corresponding region from chromosome 12 of Pacific halibut. GenBank accession numbers for these sequences, top to bottom, are OK041463, OK041464, NC_048942 (845,791 bp to 1,041,091 bp), OK041465, NC_048938 (22,286,642 bp to 22,384,527 bp).

the flanking regions of the two alleles are not identical, with around 3% divergence in DNA sequence, primarily within low-complexity regions. However, the protein sequences of the two genes immediately flanking the *AFP*s, *HDAC5* and *XYLT1* (Fig. 3a), are 100% identical.

The structure of the larger allele (allele 1) is complex. Its 33 *AFP*s are flanked on both sides with partial gene sequences (pseudogenes) whereas the single pseudogene in allele 2 is downstream of the four *AFP*s (Fig. 3a). The downstream pseudogenes retain some of the coding sequence (Fig. 4a). Allele 1 contains twelve (Supplementary Fig. 1) nearly-identical 11.2 kb tandem repeats, each encoding both a skin and a liver *AFP* isoform, *L1-L12* and *S1-S12* (Fig. 3a, see “Nomenclature” in “Materials and Methods” for further details about gene/protein names). These are followed by nine additional *AFP*s; six skin isoforms (*S13-S18*), one longer liver isoform (*Midi*) and two long isoforms (*Maxi-1*, *Maxi-2*). Allele 2 lacks *Maxi* sequences and contains a single pair of genes encoding a skin and liver isoform (*S1a*, *L1a*), with high similarity to the pairs within the tandem repeats of allele 1 (Fig. 4b,c). This region of allele 2 is 94% identical, over 11.9 kb, to the repeat region of allele 1, and the two skin isoforms that follow, *S2a* and *S3a*, closely resemble *S15* and *S16*, respectively (Supplementary Fig. 2). Allele 2 could have arisen from allele 1 via two large deletions, the first removing 11 of 12 repeats through to *Maxi-2*, and the second removing *S17* through *S18*. Alignments between these two alleles can share up to 98% identity over several kb, but all of these contain a few base insertions or deletions in addition to mismatches (not shown). A comparison of the four coding sequences in allele 2 to their closest matches in allele 1 show an average identity of 98.4%.

***AFP* gene structure.** All the *AFP* genes, with the exceptions of the pseudogenes that flank the locus, possess two exons (Fig. 5, partial data shown), the first of which is non-coding in the case of the *skin* isoforms, but which encodes most of the signal peptide in all other isoforms. The basis for identifying the flanking sequences as pseudogenes are as follows. The 5' pseudogene of allele 1 lacks a coding sequence but is identical over 80 bp to the 3'-end of the 3' UTR of the *liver*, *Maxi* and some *skin* genes. The 3' pseudogenes of both alleles contain partial coding sequences (16 a.a. or 33 a.a.) that are shorter than the shortest *skin* isoform (37 a.a.), and the Thr are not spaced at 11 a.a. intervals (Fig. 4a). Additionally, they lack the first exon due to the insertion of an ~ 2 kb LINE1 transposon (not shown), which would likely interfere with expression.

a

```

Ψ2-Allele 1      MDAATIAAAAA-----TATAT 16
Ψ-Allele 2      MDAATIAAAAAATAAKAAEDAASATAAATAT 33
*****
:****
    
```

b

```

S1,2,1a      MDAPARAAAAATAAAKA--AAEAT-----KAAAAKAAATKAAR----- 37
S13          MDAPAKAAAAATAAAKA--AAEAT-----KAAAAKAAATKAAR----- 37
S12          MDAPARAAAAATAAAKA--AAEAT-----AAAAKAAATKAAR----- 37
S17          MDAPARAAAAATAAAKA--AAEAT-----AAAAKAAADTKAAAAAAL-- 43
WF-S1        MDAPAKAAAAATAAAKA--AAEAT-----AAAAKAAATKAGAAAR----- 39
S16          MDAPAKAAAAATAAAKA--AAEAT-----AAAAKAAATKAGR----- 37
S3a,WF-S2   MDAPAKAAAAATAAAKA--AAEAT-----AAAAKAAATKAAAR----- 39
S14          MDAPAAAAATAAAKAKAAAEAT-----AAAKAAAAKAAAAALLGSWAS 47
S2a          MDAPAAAAATAKAAKA--AAEAT-----AAAAKAAATKAAAR----- 39
S15          MDAPAAAAATAAAKA--AAEAT-----AAAAKAAATKAAAR----- 39
S18          MDAPAAAAATAAAKA--AAEATAAAAAATAEAAKAAATKAAAAAAR-- 54
*****
*****
    
```

c

```

L1,2,11,12   malslftvgqliflftirineaNPDPAAKAAAVADPAAAAVAPAADAFSAADTASDAAAAAA
L1a          malslftvgqliflftirineaNPDPAAKAAAVADPAAAAVAPAADAFSAADTASDAAAAAA
WF-L1        malslftvgqliflftiriteaSPDPAKAA---PAAAAA-PAA---AAPDTASDAAAAAAL
WF-L2        malslftvgqliflftiriteaNPDPAAKAAV---PAA---AAPDTASDAAAAAA
Midi         malslftvgqliflftiriteaNPDPAAKAAAA---DAFSAADTASDAAAAAA
*****
*****

L1,2,11,12   T-----AAAAKAAEK TARDAAAAAATA---RG 91
L1a          T-----AAAAKAAEK NARDAAAAAATA---RG 91
WF-L1        T-----AANAKAAAEI LTAANAAAAATA---RG 82
WF-L2        TAATAAAAAAT-----AATAAKAAL LTAANAAAAATAAAAR 91
Midi         TAATAAAAAATATAAKAAATAATAKAAATAATAAAMA TIEAAKAAAAATAAAAR 116
*
*
*
*
*
*
*
*
*
*
    
```

d

```

Maxi-2       malslftvgqiflflstiriteaNIDPAKAAAAAAAAAATAADAAAAATIAASAAA 60
WF-Maxi      malslftvgqiflflwtisiteaNIDPAAKAAAAAAASKAAVTAADAAAAATIAASAA 60
Maxi-1       malslftvgqiflflwtisitea-IDPAAQAAAAAAAAAVTAADAAAAAANAANA 59
WF-5a        malslftvgqiflflwtisitea-IDPAAKAAAAAAATAVVTAAAAAAAAAAIAATAAA 59
*: *****:**** * * * * * :*****: * . * * * * * * * * * * * *

Maxi-2       VAAATAADAAAAS IATINAAAAATATAAAAAIAAKETATAAASAAAAASAAAAAQATI 120
WF-Maxi      VAAATAADDAAS IATINAA SAAKSI AAAAA MAAKDTA AASAAAAVA SAAKLETTI 120
Maxi-1       VAAATAADVATASIATIKANAAAAATAAAAAIAAEAAATAAATAAAAAATAAQAATI 119
WF-5a        VAGATAADAAAASIASINANTAAAAATAAAAAKAAEAAATAAAAAATAAATAQAATI 119
**.* ** * :***: * * : * * * : * * * * * : * * * * * : * * * * * : *

Maxi-2       NDKAAAAAATTATTAATAAATAATTAATAAATAATIE NAAATKAAAAAVAVAAAVTAAT 180
WF-Maxi      NVKAAVAAATTANTAAAAATAATTAATAAATAKATIDNAAAKAAAVATAVSDAAATAAT 180
Maxi-1       FDKAAAAASTATTAATAAATAATTAATAAATAETIDKAAAAAATAAATAVATAAATAAT 179
WF-5a        KDKAAAAASTATNAAAAATAATTAATAAATAVAKTIDKAAAAAVAVAAATAVAAAAATAAT 179
* * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Maxi-2       AAATAAATLEAAAAKAAAVAVSAAAAAAAAAIAAAAA-- 217
WF-Maxi      AAAVAAATLEAAAAKAAATAVSAAAAAAAAAIAFAAAP- 218
Maxi-1       AAATAAATLGAAAAKAAATAVAAAAAATAAAAAAAP 218
WF-5a        AAATAAATLGAAATVKAATAVNAAAAAAAAAATAAAAAAP 218
* * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *
    
```

Figure 4. Alignments of the AFP sequences of starry flounder along with selected sequences from winter flounder. Variable a.a. are highlighted yellow (variation 1) or grey (variation 2) where they occur. (a) The translations of the remnant coding sequences of the two presumptive pseudogenes at the 3' end of the AFP alleles. (b) Skin isoforms, including two from winter flounder (WFs1, WF2, GenBank accessions M63479.1 and M63478.1 respectively). (c) Liver isoforms, including two from winter flounder (WFL1, WFL2, GenBank accessions M62416 and DQ062445 respectively) with the secretory signal peptide in lowercase font and the pro-peptide region in italics. Arrows indicate cleavage sites. Underlining indicates the residue interrupted by the intron (phase 2). (d) Hyperactive isoforms, including two from winter flounder (WF-Maxi, WF-5a GenBank accessions EU188795.1 and AH002489.2 respectively) with the signal peptide and intron location indicated as above.

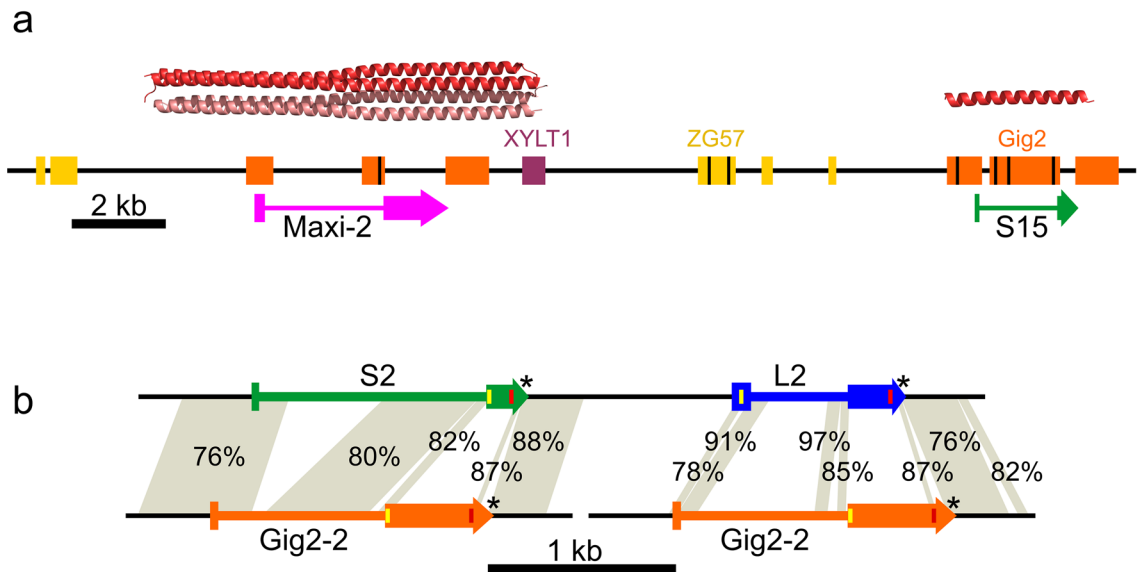


Figure 5. Higher-resolution view of selected *AFP* genes with similarities to the *non-AFP* progenitor genes indicated. **(a)** A 24 kb segment of Allele 1 containing the *Maxi-2* and *S-15* genes, coloured as in Fig. 2, with exons indicated by thicker bars. Blocks show regions of similarity to conspecific *XYLT1* (maroon) and *Gig2* (orange) as well as *ZG57* from Pacific halibut (dark yellow). Black lines within blocks indicate the location of deletions within the *AFP* genes relative to the *non-AFP* genes. Identities range from 70 to 96%. The structure of the winter flounder *Maxi* dimer (PBD 4KE2), which is the same length as *Maxi-2*, and a simplified AlphaFold 2.0⁴⁴ model of *S-15*, are shown above their genes at the same relative scale. **(b)** Detailed comparison of repeat 2 (bases 87,151–91,650), containing the skin (*S2*) and liver (*L2*) genes, with the *Gig2-2* gene (bases 12,051–14,425). Yellow and red lines within exons represent the start and stop codons respectively and the introns are indicated with a thinner bar. Asterisks denote conserved AATAAA polyadenylation motifs. The shading connects regions of similarity between the two loci with percent identities indicated.

There are twelve 11.2 kb AFP-containing repeats in allele 1. The 11.2-kb repeats at the 5' end of allele 1 were almost identical. By selecting and anchoring the longest reads to polymorphisms in the outer repeats, as described in supplementary materials and methods, the first 2.4 repeats and the last 1.5 repeats were unambiguously assembled. The interior repeats appeared virtually identical, so they were counted using a different method (Supplementary Fig. 1). A subset of raw sequence reads, from two clones that overlapped the entire region (BAC45 and BAC182, Fig. 2) were analyzed. The number of reads corresponding to either the BAC vector or the repeat was compared. The larger BAC45 dataset indicated that there were likely 12 repeats (11.9 ± 0.6), overlapping the estimate of 11 repeats (11.2 ± 0.9) from the smaller BAC182 dataset. The lack of divergence of the internal repeats suggests that they may be undergoing rounds of expansion and contraction through unequal crossing over.

The near identity of the twelve tandem 11.2 kb repeats is mirrored in the protein sequences of the repeats that were assembled. The four liver AFPs (L1, L2, L11, L12) are identical and the last of the three skin isoforms (S12) differs at just one a.a. residue from S1 and S2 (Fig. 4b,c).

The AFPs fall into three main groups. The shortest encoded isoforms are the skin isoforms that lack both a signal peptide and propeptide (Fig. 4b). Most are 37–39 a.a. long with an acidic residue (Asp) at position 2 and a C-terminal basic residue (Arg) to interact with the helix dipole, as well as three Thr residues at 11 a.a. intervals. The exceptions have a C-terminal extension lacking Arg (S17, S14), a two-residue internal insertion (S14) and both a C-terminal extension and an additional 11 a.a. repeat (S18, 54 a.a.). One winter flounder skin isoform is identical to S3a and a second differs at a single residue⁴⁵.

The second group are secreted isoforms that have both a signal peptide and a propeptide that are cleaved from the mature AFP (Fig. 4c). The starry flounder liver isoforms in the 11.2 kb repeats are 38 residues long after processing, similar in length to the skin isoforms. The liver isoform of the second allele (L1a) has a single Asn mutation at one of the periodic Thr residues. These isoforms have several substitutions relative to their winter flounder counterparts^{46,47} and a longer propeptide region. The sequence designated *Midi* is like the liver isoforms with a signal sequence and propeptide region that are thought to undergo the same N-terminal processing. However, instead of three 11-a.a. repeats, this isoform has six and the mature protein is intermediate in length (76 a.a.) between the shortest (37 a.a.) and longest (195 a.a.) isoforms (Fig. 4).

The third group are the hyperactive *Maxi* isoforms (Fig. 4d), found only in allele 1, where they are adjacent to one another. These isoforms have a signal peptide, but they lack the propeptide domain found in the other liver isoforms. These 194–195 a.a. proteins are over five times longer than most of the skin and liver isoforms and align well with the two known hyperactive isoforms from winter flounder (Fig. 4d)^{35,45}. The identity between the

Species name ^a		Taxonomic level shared with starry flounder	Type I AFPs?	Gig2 genes	Genes flanking Gig2	
Common	Binomial				5'	3'
Starry Flounder	<i>Platichthys stellatus</i>		Yes	4 + 1Ψ	MTX2	CADH5
Pacific Halibut	<i>Hippoglossus stenolepis</i>	Family	No	2 + 2Ψ	HDAC5 ZG57	XYLT1
Greenland Halibut	<i>Reinhardtius hippoglossoides</i>	Family	No	3 + 1Ψ	HDAC5 ZG57	XYLT1
Spotted Halibut	<i>Verasper variegatus</i>	Family	No	2	HDAC5 ZG57	XYLT1
Olive Flounder	<i>Paralichthys olivaceus</i>	Suborder	No	1 ^b	HDAC5 ZG57	XYLT1
Turbot	<i>Scophthalmus maximus</i>	Suborder	No	5	RAB40C ZG57	UNK93
Barramundi	<i>Lates calcarifer</i>	Series	No	3	RAB40C ZG57	XYLT1
Yellow Perch	<i>Perca flavescens</i>	Subdivision	No	1 + 1Ψ	RAB40C ZG57	XYLT1

Table 1. Characteristics of the Gig2 loci of selected fishes. ^aSpecies shown in order of relatedness to starry flounder based on phylogeny from The Fish Tree of Life²⁵. ^bNumber uncertain as there is a long gap in the assembly near the Gig2 gene.

two starry flounder sequences, Maxi-1 and Maxi-2, is 82%. When compared to the winter flounder sequences, Maxi-1 is more like 5a (82%) than WF-Maxi (79%), whereas the opposite is true for Maxi-2 (79% to 5a vs. 84% to WF-Maxi). Maximum-likelihood phylogenetic analysis (Supplementary Fig. 3) groups Maxi-1 with WF-5a and Maxi-2 with WF-Maxi, indicating that these two isoforms may have arisen prior to the separation of the winter flounder and starry flounder lineages, over 13 MA ago (Fig. 1). This is also consistent with the divergence (18%), between Maxi-1 and Maxi-2.

The second cloned locus contains five copies of Gig2. The two BACs that were sequenced (Fig. 2c) from the Gig2 locus (Fig. 3c) were identical, suggesting they originated from the same allele. The Gig2 genes lie between the *metaxin-2* (MTX2) and *cadherin-5* (CADH5) genes, so they reside at a different locus than the AFP genes. This locus was isolated because the Gig2 genes share up to 92% identity to a 252 bp segment of the 3' UTR AFP probe used to screen the library.

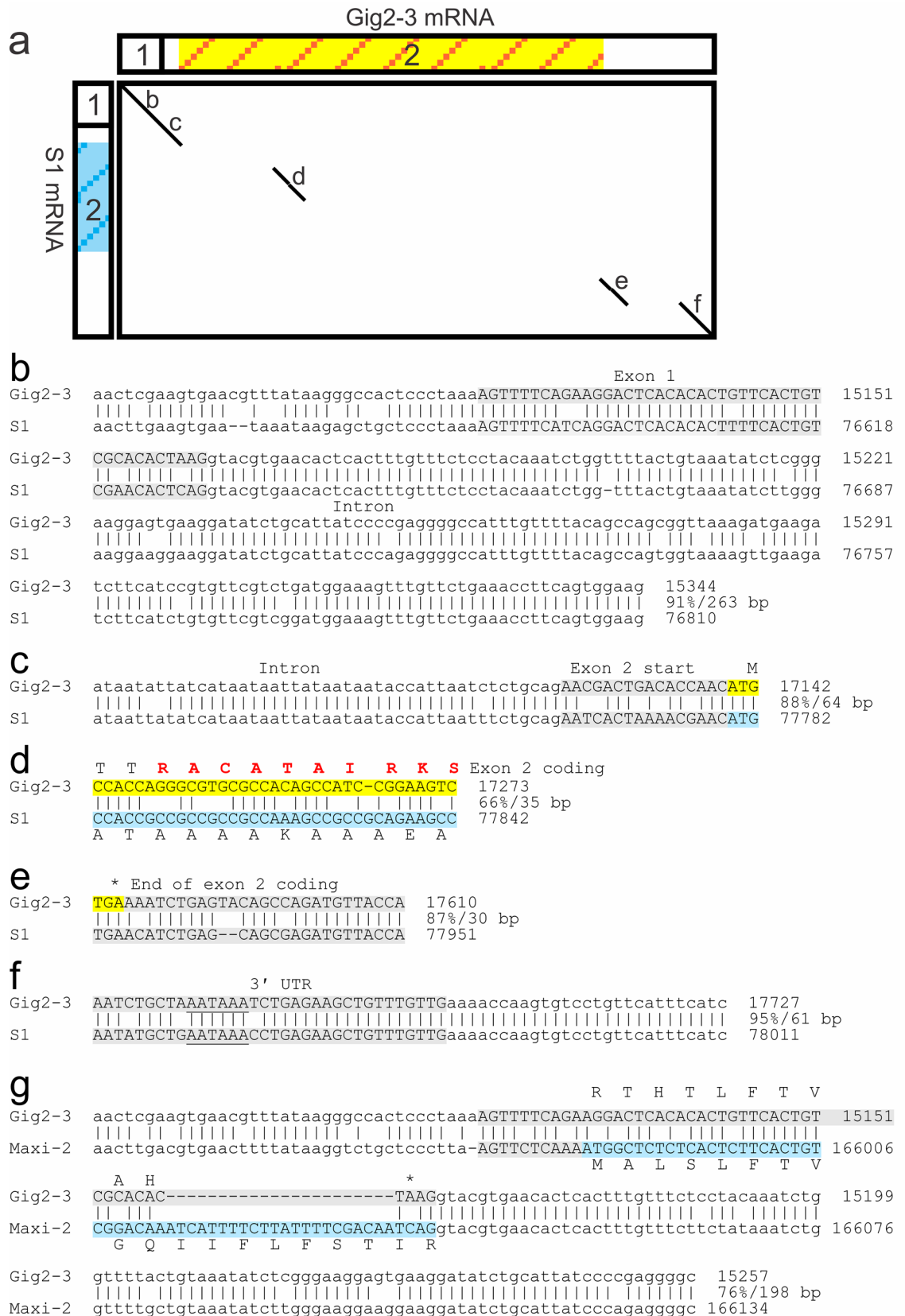
The five Gig2 genes in this locus were identified and annotated by comparison with well-characterized Gig2 genes from other fishes⁴². Gig2 has been shown to protect fish kidney cells in culture from viral infection⁴³. One of the isoforms (Gig2-4) is 40 residues shorter than the others and may be a pseudogene. The four isoforms that are 147 a.a. long were aligned (Supplementary Fig. 4) and they share 73–86% sequence identity. Notably, the sequence of these proteins does not resemble that of the AFPs as they contain little Ala. SMART analysis (<http://smart.embl-heidelberg.de/>) suggests that residues 20–115 of Gig2-3 are similar to the poly(ADP-ribose) polymerase catalytic domain (expect value of 1.6×10^{-6}).

Part 2: similar loci in other fishes. *A syntenic Pacific halibut locus lacks AFPs but contains Gig2 and ZG57 genes.* A high-quality genome sequence is available for the Pacific halibut (GenBank Assembly GCA_013339905.1)⁴⁸, a species in the same family (Pleuronectidae) as starry flounder. These species shared a common ancestor around 20 MA ago (Fig. 1). The region of its genome corresponding to where the AFP locus is in the starry flounder shares the same flanking genes on either side, including *COL1A1*, *HDAC5*, *XYLT1* and *FUS*, but it completely lacks AFP genes (Fig. 3b). Instead, it contains four Gig2 genes. These were annotated in the GenBank deposition (XM_035180664.1) as one combined Gig2 gene with adjustments for frameshifts. Conspecific transcriptomic sequences in the Sequence Reads Archive database at NCBI⁴⁹ were inconsistent with this combined gene model, so they were reannotated to show four copies of Gig2, each with a small non-coding exon followed by a coding exon as in the starry flounder Gig2 genes. The first two genes encode proteins that are highly similar (71–80% identity) to the starry flounder Gig2 proteins (Supplementary Fig. 4). The next two contain frameshifts that disrupt the reading frames, so like Gig2-4 in starry flounder, these may be pseudogenes.

There was one gene found downstream of *HDAC5* in Pacific halibut, just upstream of the Gig2 genes, that was not found in starry flounder (Fig. 3b). This gene is well conserved, contains two exons, and encodes gastrula zinc finger protein XICGF57.1 (ZG57), a 56.3-kDa protein that shares no similarity with AFPs.

The Pacific halibut locus that is syntenic to the Gig2 locus in starry flounder lacks Gig2 genes. The region of the genome in Pacific halibut that corresponds to the Gig2 locus of starry flounder was also characterized (Fig. 3d). Although the flanking genes, *MTX2*, *CADH5* and *BEAN1*, were well conserved, there is a complete absence of Gig2-like sequences at this location.

The microsynteny of Gig2 genes varies among fishes but is unique in starry flounder. The Gig2 loci of species closely related to starry flounder, with genome assemblies sufficiently long to span Gig2 and neighbouring genes, were characterized (Table 1). Species within the same family (Pleuronectidae) as the starry flounder and Pacific



◀ **Figure 6.** Alignments between *Gig2-3* and AFPs. (a) Dot plot comparison of the mRNA sequences of *Gig2-3* to *S1* generated using YASS⁵¹. The two exons are indicated by rectangles and the coding sequence of *Gig2* by the yellow/orange striped background and that of the AFP with a blue striped background. (b–f) Exon-spanning alignments of the gene sequences of *Gig2-3* and *S1*, corresponding to the segments identified in (a). Exons are in uppercase font, highlighted grey if non-coding or as in (a) if coding. Percent identities and alignment length are at the end of each aligned segment. Genic matches not overlapping exons are not shown. Residues modelled as helical within *Gig2-3* (Fig. 7) are shown in (d) in red, the stop codon for *S1* is 31 bp upstream (not shown) of the *Gig2-3* stop codon in (e), and the polyadenylation signal is underlined in (f). (g) Match between *Gig2-3* and *Maxi-2* spanning exon 1 only. The signal peptide sequence is shown along with a translation of the corresponding region of the non-coding *Gig2* exon. The base numbers shown correspond to GenBank Accessions OK041465 (*Gig* locus) and OK041463 (*AFP* locus 1).

halibut share microsynteny with the halibut, with *HDAC5* and *ZG57* upstream and *XYLT1* downstream of the *Gig2* locus (Table 1 and Fig. 3b). More variability is found in selected species outside the Pleuronectidae, with *RAB40C* in place of *HDAC5* in several species and *UNK93* in place of *XYLT1* in one (Table 1). However, none of these *Gig2* loci are flanked by either *MTX2* or *CADH5*, as in starry flounder (Fig. 3c). These observations support the hypothesis expanded on below, that the *AFP* arose from the original *Gig2*, following the latter's gene duplication and relocation in an ancestor of the starry flounder.

Starry flounder AFPs are homologous to AFPs from other Pleuronectiformes. The homology of the winter flounder and starry flounder AFPs is apparent from the similarity of their non-coding sequences. A 2.9 kb portion of a 7.8 kb tandemly-repeated gene from winter flounder encodes a liver isoform⁵⁰. Most (88%) of this sequence, which is primarily non-coding, has over 84% identity to the starry flounder 11.2 kb repeat (Supplementary Fig. 5). It was not determined if this winter flounder repeat DNA also contained a *skin* isoform.

Additional winter flounder genomic sequences, initially identified as pseudogenes⁴⁵, are also highly similar to starry flounder sequences. Two *skin* genes [GenBank accessions M63478.1 (1.4 kb) and M63479.1 (1.2 kb)], are most like *S14*, with 90% and 85% identity respectively. Additionally, the *WF-5a* gene (GenBank accession AH002489.2) is over 80% identical to both *Maxi-1* and *Maxi-2* over most of its length.

The non-coding sequence of the mRNA encoding an AFP (GenBank accession X06356.1) from the more distantly-related yellowtail flounder (Fig. 1)¹², is also highly similar to that of the starry flounder *liver* isoform within the repeats. The 5' UTR (30 bp) is 93% identical and the 5' UTR is (96 bp) is 96% identical to the *liver* isoforms in the 11.2 kb repeat. Similar comparisons to the non-coding regions of the *type I AFPs* of other orders (Fig. 1) failed to identify any similarity, as was found when comparisons were done using winter flounder sequences¹⁵.

Part 3: the origin of the flounder AFP genes. *Remnants of three genes indicate that the AFP genes arose at their current location.* The region containing the starry flounder AFPs was compared to the flanking sequences and to the Pacific halibut *ZG57* locus (Fig. 3). A portion of the *ZG57* gene containing the first exon and part of the intron is found just upstream of the first AFP pseudogene in allele 1 (Fig. 3a, yellow bar). This segment encodes 22 a.a. that closely resemble the N-terminal sequence of the halibut protein, but several frameshifts thereafter disrupt the reading frame, and the second exon is absent, so this gene is no longer functional (not shown). Sequences similar to various regions of *ZG57* are found scattered throughout the *AFP* region and some of these are indicated in dark yellow in Fig. 5a. Similarly, segments corresponding to the 5' region of the downstream *XYLT1* gene are also found scattered about, and while only one small segment is found in the region shown in Fig. 5a in maroon, three segments totaling 2.2 kb are found within the 11.2 kb repeats (not shown). Some *AFP*s, such as *Maxi-2* (Fig. 5a), are flanked by both *ZG57* and *XYLT1* segments. *ZG57* segments are always upstream and *XYLT1* segments are always downstream of *AFP*s. This suggests that a single *AFP* gene arose between *ZG57* and *XYLT1* and that when the *AFP* locus expanded, portions of these flanking genes were duplicated along with the *AFP*.

Gig2 was likely the AFP progenitor. A comparison of the *Gig2* and *AFP* loci of starry flounder indicated that there were many stretches of similar sequence, some of which are shown in Fig. 5a. As these matches cover a significant portion of the *AFP* gene, except for the coding sequence, this suggests that the *AFP* gene arose from the *Gig2* gene. Furthermore, the greater number of matches to *S15* than to *Maxi-2* suggests that the *skin* gene likely arose first and that subsequent alterations, in which regions similar to *Gig2* were lost, gave rise to the *Maxi* genes.

A more detailed comparison is shown between the skin and liver *AFP*s within the 11.2 kb repeat and the *Gig2-2* locus (Fig. 5b). Here again, the skin *AFP* is more like *Gig2* with regions of similarity beginning before and extending across the non-coding exon 1, continuing throughout much of the intron and into exon 2, up to and including the start codon. The coding sequences of *S2* and *Gig2-2* share no significant similarity, but similarity begins again downstream of the coding sequence. The matches between *Gig2* and the liver *AFP* are more limited, including in the presumptive promoter/enhancer region upstream of the gene, and resemble those between *Gig2* and *Maxi-2*.

A dot plot comparison of the predicted mRNA sequences of *S1* and a second *Gig2* gene, *Gig2-3* showed four segments with similarity (Fig. 6a). Sequence alignments between the genes in these vicinities are shown in Fig. 6b–f. The similarity between the non-coding first exon of both genes is evident with a match of 39 out of 44 bp, with the similarity extending further, both 5' of the gene and downstream into the intron (Fig. 6b). The match at the start of exon 2 also extends into the intron, but the sequences diverge downstream of the start codon (Fig. 6c). There is but one short segment showing 66% identity within the coding region (Fig. 6a,d).

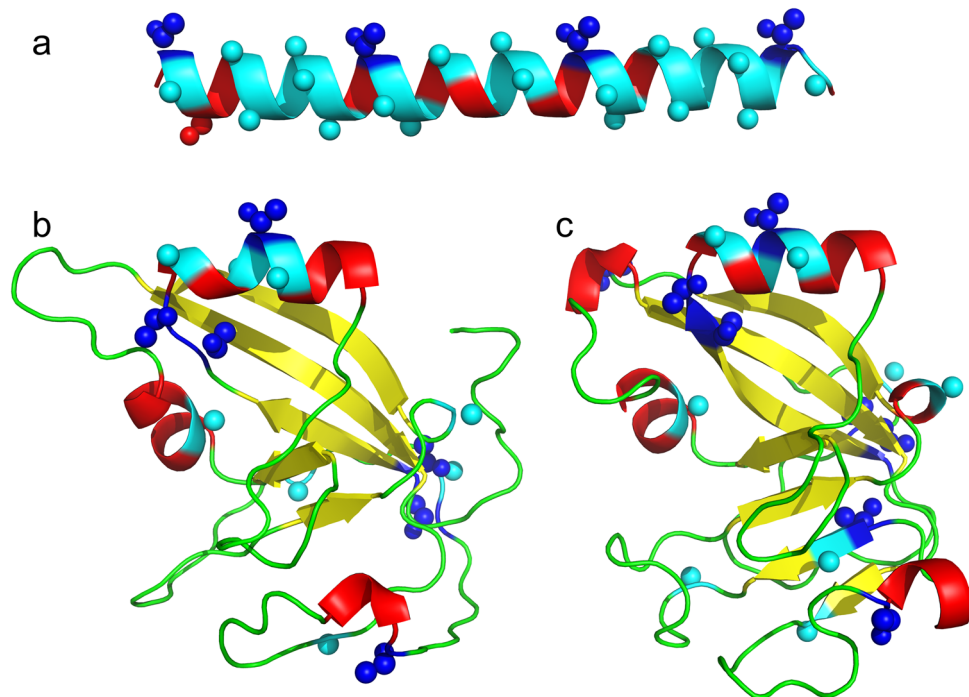


Figure 7. Homology model of Gig2 compared to type I AFP from winter flounder. (a) Type I AFP (PDB:1WEA). (b) Gig2-3, modelled using Phyre2⁵², was aligned with 100% confidence over 89% of its length to the template PDB:3C4H. (c) Gig2-3 modelled without a template using simplified AlphaFold 2.0⁴⁴. The first eight residues (5%) were removed as they were modelled with low confidence. The images were generated using PyMOL²⁸ and are shown in cartoon mode with small spheres representing side chains for Ala residues (cyan) and Thr residues (blue). The other residues are coloured by secondary structure with α -helices in red, β -strands in yellow and coils in green.

The last two matches are downstream of the coding sequence, the first of which starts right at the stop codon of *Gig2-3* and 31 bp downstream of the stop codon of *S1* (Fig. 6e). The second extends into the 3' region and overlaps a presumptive poly-adenylation signal (Fig. 6f). As mentioned previously, exon 1 of both *Gig2* and skin *AFPs* is non-coding, but for the liver and Maxi *AFPs*, it encodes a signal peptide. Despite this, an alignment of the *Maxi-2* and *Gig2-3* regions spanning this exon shows that a limited number of mutations, such as AGG to ATG to introduce a start codon, along with a small insertion of 23 bases, were sufficient to convert the exon to a signal-peptide encoding sequence (Fig. 6g). This indicates that the signal peptide arose in situ, from the non-coding exon of *Gig2*.

Possible origins of the AFP coding sequence. Flounder AFP is Ala rich and these straight α helices provide a flat surface that interacts with ice^{33,37}. In contrast, *Gig2* has a lower-than-average Ala content (~5%), with only one 5 a.a. segment, ACATA, found in two isoforms (Supplementary Fig. 4) that resembles the Ala-rich AFP sequence. This sequence is encoded by the region of similarity detected by dot matrix analysis (Fig. 6a,d). If this region gave rise to a type I AFP, it would be expected to reside within a surface-exposed α helix. Fortunately, the structure of a homolog, poly(ADP-ribose) polymerase catalytic domain, is known and the Phyre2⁵² homology model of *Gig2* (Fig. 7) shows that this ACATA segment is likely surface exposed and is located on the longest helical segment predicted for this globular protein. The AlphaFold2⁴⁴ de novo model is very similar and predicts the same surface exposed helix. Deletion of most of the coding sequence, followed by amplification of this short segment, could have given rise to a primordial AFP. Alternatively, a GC-rich sequence encoding numerous Ala residues, such as (GCC)_n, could have replaced the *Gig2* coding sequence.

Deduced steps in the generation of the flounder AFPs. The comparisons between the various loci of the starry flounder and Pacific halibut, as well as the location of the *Gig2* loci in other closely-related fish, make it clear that the ancestor of the flounder had *Gig2* genes lying between the *ZG57* and *XYLT1* genes (Figs. 3 and 5, Table 1). Within the flounder lineage, a gene duplication event led to additional copies of the *Gig2* gene at the second locus, between *MTX2* and *CADH5* (Fig. 3c). The original *Gig2* genes were then redundant, and one underwent changes that generated a skin AFP. This could have come about if the short Ala-containing segment within the α -helix region expanded (Fig. 6d) or if a segment of repetitive, GC-rich DNA replaced the coding sequence. The gene was then duplicated an unknown number of times, at this location, as shown by the many segment within the AFP locus that are similar to the *ZG57* and *XYLT1* genes (Fig. 5a). Eventually, the non-coding exon 1 of one duplicate evolved into encode a signal peptide (Fig. 6g). Further gene duplications and/or gene losses (as can be

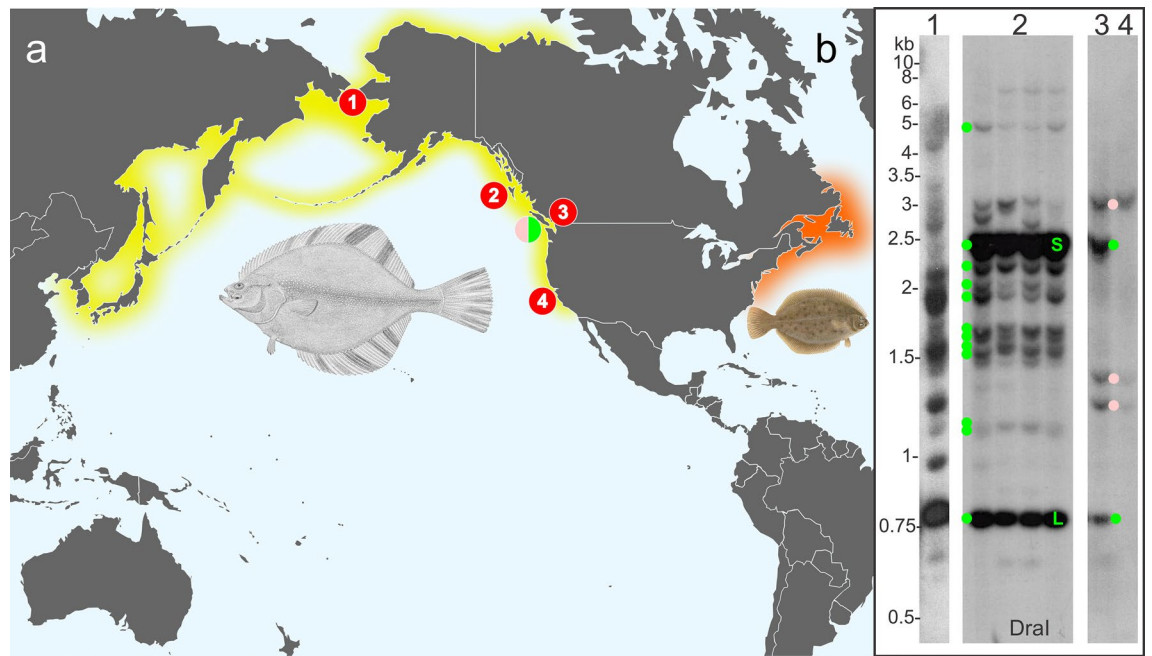


Figure 8. Southern blots of genomic DNA from starry flounder collected from various regions throughout its range. **(a)** The native ranges of starry flounder and winter flounder are indicated with yellow and orange shading, respectively. The locations where starry flounder were harvested for Southern blotting are indicated with the red numbered circles while the location of the fish harvested off Vancouver Island used to generate the BAC library is indicated with the split pink/green circle. **(b)** Southern blots of DNA digested with *DraI* for individual fish from the Bering Strait, Alaska (1) English Bay, British Columbia (3), Monterey Bay, California (4) and from four fish from Haida Gwaii (2). The blots were probed with a sequence from allele 1 (bases 77,759 to 77,972) that overlaps the second exon of the skin *AFP* within the 11.2 kb repeat. The expected locations of fragments from allele one are indicated by green dots with S (skin) and L (liver) for the genes within the repeats. Pink dots correspond to fragment sizes expected to arise from allele 2. The flounder images and map were obtained from Wikimedia Commons (see Supplementary Materials and Methods).

postulated from Supplementary Figs. 2 and 6), as well as expansions and contractions of the repetitive coding sequences, gave rise to the extant complex alleles due to the selective pressure (or lack thereof) of living around sea ice.

Allele 2 is more prevalent in starry flounders from warmer waters. The fish that was used to construct the library, and which had the two differing *AFP* alleles, was caught in southerly Canadian waters of the North Pacific, off the western side of Vancouver Island (pink/green circle, all locations are shown in Fig. 8a). In contrast, a genomic Southern blot of four fish collected from the Haida Gwaii, approximately 300 km further north (location 1), showed that the larger *AFP* allele 1 was prevalent at this location (Fig. 8b-2). Two intense bands, corresponding to the *skin* and *liver* genes within the 11.2 kb repeat, confirm the repetitive nature of this repeat. Bands corresponding to the predicted sizes of all the other genes from allele 1 were also observed, further confirming the accuracy of our assembly. A more detailed analysis of the correspondence between these bands and the two *AFP* alleles is shown in Supplementary Figure 7. There is some evidence of limited polymorphism as a few unexplained bands were present in one or two of the fish, but all these fish appear to be homozygous for alleles very similar to allele 1, as bands corresponding to the unique and well-separated fragment sizes expected for *S2a*, *S3a* and *S4a* were not observed.

In contrast to the large *AFP* copy number of the more northerly starry flounder, a fish caught in Monterey Bay, California (location 4), only has bands consistent with allele 2 (Fig. 8b-4). Although at a similar latitude as the sequenced flounder from the west coast of Vancouver Island, the fish caught in the warmer slightly brackish waters of English Bay, off Vancouver (location 3), had bands consistent with allele 2, along with some moderately intense bands consistent with the *skin* and *liver* genes within the 11.2 kb repeats (Fig. 8b-3). We speculate that it contains an allele similar to allele 2 that still has a small number of 11.2 kb repeats remaining. A fish from Alaska (location 1), approximately 1500 km further north from Haida Gwaii, had many intense bands with sizes that were not consistent with either allele (Fig. 8b-1). Together, these results suggest that gene copy number is correlated with risk of ice exposure and that numerous alleles with differing numbers of *AFP* genes can be found within this species.

Discussion

Taxonomically restricted genes (TRGs) confer phenotypic novelty on their hosts and the selective pressures of new environments often provide the driving force for their development^{53,54}. For example, water striders have colonized the water surface due in part to TRGs that generate a “fan” on the middle leg that provides propulsion across the surface⁵⁵. Similarly, the climate cooling that intensified during the latter half of the Cenozoic Era generated an icy sea environment that had been absent for at least tens of Ma^{27,31}, and which would have excluded fish from shallow water niches where ice is found until the *AFP* genes arose in certain species, including the recent ancestors of the starry flounder. These and other TRGs arise in a variety of ways⁵³, including via duplication and divergence of existing genes, as for example with *AFGP*, type II and type III *AFP*^{22,18,16}, or de novo from non-coding DNA (*AFGP*^{21,23}). It can be difficult to determine the mechanism, as selection for a new function can lead to rapid divergence, erasing the similarity to the progenitor sequence⁵⁶. This erasure likely occurred with the coding sequence of the flounder *AFP* gene as it bears little similarity to the *Gig2* progenitor. Fortunately, the *AFP* arose recently, so extensive similarity between the flanking regions of the two genes was retained (Figs. 5 and 6). Additionally, the lineage-specific duplication of the *Gig2* genes at a second locus, as well as sequential duplications of segments of the flanking genes at the original locus (Figs. 3 and 5), shows that the *AFP* gene arose, in situ, at the original *Gig2* locus via gene duplication and divergence.

It is now clear that the *AFPs* of Pleuronectiformes, such as starry flounder, are not homologous to the type I *AFPs* found in the other three lineages (snailfish, cunner and sculpin) within Perciformes and Labriformes, as these other *AFPs* lack similarity to *Gig2*. It was proposed that the snailfish *AFP* could have arisen from a frameshifting of the Gly-rich region of either keratin or chorion cDNAs that were inadvertently cloned along with the *AFP* genes⁵⁷. However, the similarity did not extend into non-coding segments. As all these genes arose within the last ~20 Ma, they would be expected, like the flounder's, to retain some evidence of their origins in their non-coding regions, since diversifying selection would be lower here. Currently, the origin of the three other type I *AFPs* remains unknown.

The convergence of the *AFPs* from four lineages to Ala-rich helices, sometimes with Thr residues at 11 a.a. residues^{9,10,15,34}, suggests that this motif is well-suited to interacting with ice. Similar convergence, albeit with a different structural framework, was seen with arthropod *AFPs* that adopt a β -helical conformation. A beetle (yellow mealworm) and a fly (midge) produce tight, disulfide-stabilized solenoids, with an ice-binding surface composed of a double row of Thr residues or a single row of Tyr residues, respectively^{58,59}. The looser solenoid of the moth (spruce budworm) is more triangular and lacks bisecting disulfide bonds, but like the beetle *AFP*, its ice-binding surface consists of a double row of Thr residues⁶⁰. This suggests that there are nascent structures with propensities to evolve into *AFPs*, but that different types are more likely to arise in marine versus terrestrial environments because of the vastly different requirements for freezing point depression.

When a novel gene arises from a pre-existing one, non-coding sequences are thought to be almost as important as coding sequences⁶¹. It is likely that the promoter and enhancer sequences controlling expression of the *Gig2* gene were co-opted, for two reasons. First, the *skin* genes and *Gig2* share high identity upstream of the first exon. Second, the expression patterns of *Gig2* in zebrafish⁴² and the winter flounder skin *AFPs*³⁴ are similar as they are expressed in a variety of tissues. The tissue- and season-specific enhancement of the liver *AFPs*⁶² may have arisen later, given that its gene lacks similarity to the upstream regions of the *Gig2* gene. However, all the genes retain the two exons and the polyadenylation signal.

The rapid divergence of the starry flounder *AFP* coding sequence from the *Gig2* progenitor is reminiscent of that observed for the *AFGP* that was derived from the *trypsinogen* gene²². For the *AFP*, a 35 bp segment, corresponding to 10 a.a. in a helical region of the protein, was likely retained and amplified (Figs. 6 and 7). For *AFGP*, the amplified segment was only 9 bp long and it overlapped the acceptor splice junction at the start of exon 2. Both gene types retained the first exon, which is non-coding in skin *AFPs* and *Gig2*, but which encodes a signal peptide in both *AFGP* and *trypsinogen*. However, the first exon of the flounder *liver*, *Midi* and *Maxi* genes does encode a signal peptide and similarity with the *Gig2* non-coding exon shows that it arose, in situ. This is reminiscent of the origin of the signal peptide of type III *AFP*¹⁸, where an additional 54 bp in exon 1 gained coding potential, generating a signal peptide. One explanation for rapid divergence of specific portions of DNA sequence, such as the signal peptides mentioned above, is positive Darwinian selection, where the rate of non-synonymous (missense) to synonymous (silent) mutations at certain positions is higher than expected under either a neutral or negative model of selection⁶³. Such selection has also been observed in numerous surface-exposed residues of the globular type III *AFP* sequences from fish and the solenoid *AFP* from beetles⁶⁴. Given that there are far fewer structural constraints on isolated α -helical peptides than on the two aforementioned *AFPs*, any mutations that increased helical content or the ability to bind to ice could be subject to strong positive selection in fishes exposed to ice in a cooling ocean. The result would be higher divergence of the coding sequences relative to non-coding sequences, as seen between the *AFP* and *Gig2* sequences of the starry flounder.

The number of *AFP* genes was higher in starry flounders from the northern waters of Alaska and British Columbia than in flounders from more southerly waters (Fig. 8). Variation in gene copy number was also observed in winter flounder from different regions along the Atlantic coast, with animals from warmer waters having fewer genes⁶⁵. The same pattern has been observed for ocean pout, which can have up to ~150 genes that produce type III *AFP*⁶⁶. As many of the *AFP* genes are arranged in tandem arrays, they are likely prone to rapid expansion and contraction via unequal crossing over⁶⁷, providing variation that would be subject to environmental selection.

Gene duplication also provides additional copies that can undergo neofunctionalization⁶⁷, which is how the three main classes of type I *AFPs* found in flounders (*Maxi*, *liver* and *skin*) arose. The properties of these isoforms differ dramatically as *Maxi* is far more active than either the *skin* or *liver* isoforms³⁶, and expression of the *liver* isoform is extremely high in this tissue⁶⁸. Unequal crossing over likely led to the loss of the *Maxi* genes and the

majority of the *skin* and *liver* genes in the shorter starry flounder AFP allele. A similar process may have occurred in the American plaice. Despite being closely related to the yellowtail flounder that possesses both liver and Maxi isoforms^{12,14,24} (Fig. 1), American plaice serum only contains Maxi-like AFPs¹⁴. This suggests that the common ancestor of both of these fish had the liver isoform and that the plaice locus may have undergone contraction, losing the small liver-specific AFP genes. Similar processes, working on a smaller scale, may also be responsible for the generation of isoform variation. For example, liver-like isoforms with extra copies of the 11-a.a. repeat are found in both starry flounder (Midi with three extra repeats) and yellowtail (one extra repeat¹²). This plasticity may also explain why the banding pattern from the Alaskan starry flounder observed by Southern blotting is so different from that of fish from Haida Gwaii (Fig. 8), despite both having large numbers of AFP genes.

In summary, the origin of the flounder AFP from the gene encoding the globular, antiviral Gig2 protein, via gene duplication and divergence, has been determined. Detailed comparisons between the two loci elucidate the steps involved in the evolution of the AFP. Although the flounder AFP is superficially similar to the type I AFPs of other groups, all of which are extended alanine-rich alpha-helical proteins of varying length, it clearly arose by convergent evolution. The two extended loci that were characterized from starry flounder encode either the AFP genes or five of the *Gig2* progenitor genes. The two AFP alleles sequenced contain either four or 33 AFP genes, indicating that gene copy number can vary dramatically. These genes encode skin, liver and Maxi AFPs, with the number of AFP genes being higher in fish that inhabit colder waters.

Materials and methods

BAC library construction, screening and sequencing. A BAC (bacterial artificial chromosome) library was constructed by Amplicon Express (Pullman, Washington, USA) from genomic DNA from an individual starry flounder captured off the west coast of British Columbia. Fish tissues were harvested from euthanized fish in accordance with the Canadian Council on Animal Care Guidelines and Policies with approval from the Animal Care and Use Committee at Queen's University. A total of 12 clones that hybridized to the 3' untranslated region (UTR) of an AFP transcript were sequenced at the Génome Québec Innovation Centre (Montreal, Quebec, Canada) using the PacBio RS II single molecule real-time (SMRT[®]) sequencing technology (Pacific Biosciences, Menlo Park, California, USA).

DNA assembly, gene annotation and Southern blotting. The initial assembly was done by the Génome Québec Innovation using the Celera assembler⁶⁹. The overlapping regions of different clones were identical except at longer homopolymer or dinucleotide repeat regions. A region containing near-identical 11.2 kb repeats was assembled and evaluated separately, yielding 3.9 assembled repeats out of 12 total, as described in Supplementary Materials and Methods. Genes were annotated using homologs from other fish.

DNA from starry flounders collected at various locations from California to Alaska was Southern blotted and the blots were evaluated using various ³²P-labelled various probes to AFP genes. A more detailed description of all procedures can be found in Supplementary Materials and Methods.

Nomenclature. Genes are differentiated from proteins using italics. For simplicity, AFPs from starry flounder are named by class with “liver” for small circulating isoforms, “skin” for small isoforms first isolated from skin, “Midi” for an isoform of intermediate size and Maxi for the large circulating isoforms. Numbering is used for classes with multiple isoforms, such as S1 and L1 for the first skin and liver gene at allele 1 respectively. Isoforms from allele 2 are differentiated by letter a (S1a, L1a for example) whereas those from winter flounder are preceded by WF.

Data availability

The starry flounder sequences generated during the current study and the Pacific halibut sequences they were compared to are available from GenBank under accession numbers OK041463, OK041464 and OK041465, NC_048942 (845791 bp to 1041091 bp) and NC_048938 (22286642 bp to 22384527 bp). The structure of type I AFP was obtained from the Protein Data Bank, accession 1WFA.

Received: 8 February 2022; Accepted: 28 April 2022

Published online: 20 May 2022

References

- DeVries, A. L. Glycoproteins as biological antifreeze agents in antarctic fishes. *Science* **172**, 1152–1155. <https://doi.org/10.1126/science.172.3988.1152> (1971).
- Davies, P. L. & Graham, L. A. Protein evolution revisited. *Syst. Biol. Reprod. Med.* **64**, 403–416. <https://doi.org/10.1080/19396368.2018.1511764> (2018).
- Bar Dolev, M., Braslavsky, I. & Davies, P. L. Ice-binding proteins and their function. *Annu. Rev. Biochem.* **85**, 515–542. <https://doi.org/10.1146/annurev-biochem-060815-014546> (2016).
- Kim, H. J. *et al.* Marine antifreeze proteins: Structure, function, and application to cryopreservation as a potential cryoprotectant. *Mar. Drugs* **15**, 27. <https://doi.org/10.3390/md15020027> (2017).
- Raymond, J. A. & DeVries, A. L. Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2589–2593. <https://doi.org/10.1073/pnas.74.6.2589> (1977).
- Pertaya, N. *et al.* Fluorescence microscopy evidence for quasi-permanent attachment of antifreeze proteins to ice surfaces. *Biophys. J.* **92**, 3663–3673. <https://doi.org/10.1529/biophysj.106.096297> (2007).
- Praebel, K., Hunt, B., Hunt, L. H. & DeVries, A. L. The presence and quantification of splenic ice in the McMurdo Sound notothenioid fish, *Pagothenia borchgrevinki* (Boulenger, 1902). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **154**, 564–569. <https://doi.org/10.1016/j.cbpa.2009.09.005> (2009).

8. Evans, R. P. & Fletcher, G. L. Type I antifreeze proteins expressed in snailfish skin are identical to their plasma counterparts. *FEBS J.* **272**, 5327–5336. <https://doi.org/10.1111/j.1742-4658.2005.04929.x> (2005).
9. Low, W. K. *et al.* Isolation and characterization of skin-type, type I antifreeze polypeptides from the longhorn sculpin, *Myoxocephalus octodecemspinosus*. *J. Biol. Chem.* **276**, 11582–11589. <https://doi.org/10.1074/jbc.M009293200> (2001).
10. Hobbs, R. S., Shears, M. A., Graham, L. A., Davies, P. L. & Fletcher, G. L. Isolation and characterization of type I antifreeze proteins from cunner, *Tautoglabrus adspersus*, order Perciformes. *FEBS J.* **278**, 3699–3710. <https://doi.org/10.1111/j.1742-4658.2011.08288.x> (2011).
11. Mahatabuddin, S. *et al.* Concentration-dependent oligomerization of an alpha-helical antifreeze polypeptide makes it hyperactive. *Sci. Rep.* **7**, 42501. <https://doi.org/10.1038/srep42501> (2017).
12. Scott, G. K., Davies, P. L., Shears, M. A. & Fletcher, G. L. Structural variations in the alanine-rich antifreeze proteins of the pleuronectinae. *Eur. J. Biochem.* **168**, 629–633. <https://doi.org/10.1111/j.1432-1033.1987.tb13462.x> (1987).
13. Scott, G. K., Davies, P. L., Kao, M. H. & Fletcher, G. L. Differential amplification of antifreeze protein genes in the pleuronectinae. *J. Mol. Evol.* **27**, 29–35. <https://doi.org/10.1007/BF02099727> (1988).
14. Gauthier, S. Y., Marshall, C. B., Fletcher, G. L. & Davies, P. L. Hyperactive antifreeze protein in flounder species. The sole freeze protectant in American plaice. *FEBS J.* **272**, 4439–4449. <https://doi.org/10.1111/j.1742-4658.2005.04859.x> (2005).
15. Graham, L. A., Hobbs, R. S., Fletcher, G. L. & Davies, P. L. Helical antifreeze proteins have independently evolved in fishes on four occasions. *PLoS ONE* **8**, e81285. <https://doi.org/10.1371/journal.pone.0081285> (2013).
16. Ewart, K. V., Rubinsky, B. & Fletcher, G. L. Structural and functional similarity between fish antifreeze proteins and calcium-dependent lectins. *Biochem. Biophys. Res. Commun.* **185**, 335–340. [https://doi.org/10.1016/s0006-291x\(05\)90005-3](https://doi.org/10.1016/s0006-291x(05)90005-3) (1992).
17. Graham, L. A. & Davies, P. L. Horizontal gene transfer in vertebrates: A fishy tale. *Trends Genet.* **37**, 501–503. <https://doi.org/10.1016/j.tig.2021.02.006> (2021).
18. Deng, C., Cheng, C. H., Ye, H., He, X. & Chen, L. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21593–21598. <https://doi.org/10.1073/pnas.1007883107> (2010).
19. Baardsnes, J. & Davies, P. L. Sialic acid synthase: The origin of fish type III antifreeze protein? *Trends Biochem. Sci.* **26**, 468–469. [https://doi.org/10.1016/S0968-0004\(01\)01879-5](https://doi.org/10.1016/S0968-0004(01)01879-5) (2001).
20. Hobbs, R. S., Hall, J. R., Graham, L. A., Davies, P. L. & Fletcher, G. L. Antifreeze protein dispersion in eelpouts and related fishes reveals migration and climate alteration within the last 20 Ma. *PLoS ONE* **15**, e0243273. <https://doi.org/10.1371/journal.pone.0243273> (2020).
21. Baalsrud, H. T. *et al.* De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. *Mol. Biol. Evol.* **35**, 593–606. <https://doi.org/10.1093/molbev/msx311> (2018).
22. Chen, L., DeVries, A. L. & Cheng, C. H. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3811–3816. <https://doi.org/10.1073/pnas.94.8.3811> (1997).
23. Zhuang, X., Yang, C., Murphy, K. R. & Cheng, C. C. Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4400–4405. <https://doi.org/10.1073/pnas.1817138116> (2019).
24. Ribeiro, E., Davis, A. M., Rivero-Vega, R. A., Orti, G. & Betancur, R. R. Post-cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. *Proc. Biol. Sci.* **285**, 20182010. <https://doi.org/10.1098/rspb.2018.2010> (2018).
25. Betancur, R. R. *et al.* Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17**, 162. <https://doi.org/10.1186/s12862-017-0958-3> (2017).
26. King, M. J., Kao, M. H., Brown, A. & Fletcher, G. L. Lethal freezing temperatures of fish: Limitations to seapen culture in Atlantic Canada. *Bull. Aquacult. Assoc. Can.* 47–49 (1989).
27. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283. <https://doi.org/10.1038/nature06588> (2008).
28. Schrödinger, L. & DeLano, W. <http://www.pymol.org/pymol> (2020).
29. Pross, J. *et al.* Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature* **488**, 73–77. <https://doi.org/10.1038/nature11300> (2012).
30. Sluijs, A. *et al.* Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* **441**, 610–613. <https://doi.org/10.1038/nature04668> (2006).
31. Tripati, A. & Darby, D. Evidence for ephemeral middle Eocene to early Oligocene Greenland glacial ice and pan-Arctic sea ice. *Nat. Commun.* **9**, 1038. <https://doi.org/10.1038/s41467-018-03180-5> (2018).
32. Hew, C. L. *et al.* Biosynthesis of antifreeze polypeptides in the winter flounder. Characterization and seasonal occurrence of precursor polypeptides. *Eur. J. Biochem.* **160**, 267–272. <https://doi.org/10.1111/j.1432-1033.1986.tb09966.x> (1986).
33. Sigheri, F. & Yang, D. S. Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature* **375**, 427–431. <https://doi.org/10.1038/375427a0> (1995).
34. Gong, Z., Ewart, K. V., Hu, Z., Fletcher, G. L. & Hew, C. L. Skin antifreeze protein genes of the winter flounder, *Pleuronectes americanus*, encode distinct and active polypeptides without the secretory signal and prosequences. *J. Biol. Chem.* **271**, 4106–4112. <https://doi.org/10.1074/jbc.271.8.4106> (1996).
35. Graham, L. A., Marshall, C. B., Lin, F. H., Campbell, R. L. & Davies, P. L. Hyperactive antifreeze protein from fish contains multiple ice-binding sites. *Biochemistry* **47**, 2051–2063. <https://doi.org/10.1021/bi7020316> (2008).
36. Marshall, C. B., Fletcher, G. L. & Davies, P. L. Hyperactive antifreeze protein in a fish. *Nature* **429**, 153. <https://doi.org/10.1038/429153a> (2004).
37. Sun, T., Lin, F. H., Campbell, R. L., Allingham, J. S. & Davies, P. L. An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science* **343**, 795–798. <https://doi.org/10.1126/science.1247407> (2014).
38. Froese, R. P., D. *FishBase*. www.fishbase.org (2021).
39. Allen, M. J., Smith, G. B. & United States. National marine fisheries service. In *Atlas and Zoogeography of Common Fishes in the Bering Sea and Northeastern Pacific*. Vol. 66 151 (U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, 1988).
40. Nabeta, K. K. *The Type I Antifreeze Protein Gene Family in Pleuronectidae*, Queen's University Graduate Thesis, (2009).
41. Hincha, D. K., DeVries, A. L. & Schmitt, J. M. Cryotoxicity of antifreeze proteins and glycoproteins to spinach thylakoid membranes—comparison with cryotoxic sugar acids. *Biochim. Biophys. Acta* **1146**, 258–264. [https://doi.org/10.1016/0005-2736\(93\)90364-6](https://doi.org/10.1016/0005-2736(93)90364-6) (1993).
42. Zhang, Y. B. *et al.* Identification of a novel Gig2 gene family specific to non-amniote vertebrates. *PLoS ONE* **8**, e60588. <https://doi.org/10.1371/journal.pone.0060588> (2013).
43. Sun, C. *et al.* Gig1 and Gig2 homologs (CiGig1 and CiGig2) from grass carp (*Ctenopharyngodon idella*) display good antiviral activities in an IFN-independent pathway. *Dev. Comp. Immunol.* **41**, 477–483. <https://doi.org/10.1016/j.dci.2013.07.007> (2013).
44. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2> (2021).
45. Davies, P. L. & Gauthier, S. Y. Antifreeze protein pseudogenes. *Gene* **112**, 171–178. [https://doi.org/10.1016/0378-1119\(92\)90373-w](https://doi.org/10.1016/0378-1119(92)90373-w) (1992).
46. Davies, P. L. Conservation of antifreeze protein-encoding genes in tandem repeats. *Gene* **112**, 163–170. [https://doi.org/10.1016/0378-1119\(92\)90372-v](https://doi.org/10.1016/0378-1119(92)90372-v) (1992).

47. Cheng, C. C., Cziko, P. A. & Evans, C. W. Nonhepatic origin of notothenioid antifreeze reveals pancreatic synthesis as common mechanism in polar fish freezing avoidance. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10491–10496. <https://doi.org/10.1073/pnas.0603796103> (2006).
48. Planas, J. V., Jasonowicz, A., Simeon, A., Zahm, M., Klopp, C., Guiguen, Y. *First Complete Chromosome Level Assembly of the Pacific Halibut (*Hippoglossus stenolepis*) Genome* (International Pacific Halibut Commission).
49. Sayers, E. W. *et al.* Database resources of the National Center for biotechnology information. *Nucleic Acids Res.* **48**, D9–D16. <https://doi.org/10.1093/nar/gkz899> (2020).
50. Scott, G. K., Hew, C. L. & Davies, P. L. Antifreeze protein genes are tandemly linked and clustered in the genome of the winter flounder. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2613–2617. <https://doi.org/10.1073/pnas.82.9.2613> (1985).
51. Noe, L. & Kucherov, G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* **33**, W540–543. <https://doi.org/10.1093/nar/gki478> (2005).
52. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. <https://doi.org/10.1038/nprot.2015.053> (2015).
53. Rodelsperger, C., Prabh, N. & Sommer, R. J. New gene origin and deep taxon phylogenomics: Opportunities and challenges. *Trends Genet.* **35**, 914–922. <https://doi.org/10.1016/j.tig.2019.08.007> (2019).
54. Johnson, B. R. Taxonomically restricted genes are fundamental to biology and evolution. *Front. Genet.* **9**, 407. <https://doi.org/10.3389/fgene.2018.00407> (2018).
55. Santos, M. E., Le Bouquin, A., Crumiere, A. J. J. & Khila, A. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science* **358**, 386–390. <https://doi.org/10.1126/science.aan2748> (2017).
56. Schlotterer, C. Genes from scratch—the evolutionary fate of de novo genes. *Trends Genet.* **31**, 215–219. <https://doi.org/10.1016/j.tig.2015.02.007> (2015).
57. Evans, R. P. & Fletcher, G. L. Type I antifreeze proteins: Possible origins from chorion and keratin genes in Atlantic snailfish. *J. Mol. Evol.* **61**, 417–424. <https://doi.org/10.1007/s00239-004-0067-y> (2005).
58. Basu, K., Wasserman, S. S., Jeronimo, P. S., Graham, L. A. & Davies, P. L. Intermediate activity of midge antifreeze protein is due to a tyrosine-rich ice-binding site and atypical ice plane affinity. *FEBS J.* **283**, 1504–1515. <https://doi.org/10.1111/febs.13687> (2016).
59. Liou, Y. C., Tocilj, A., Davies, P. L. & Jia, Z. Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein. *Nature* **406**, 322–324. <https://doi.org/10.1038/35018604> (2000).
60. Tyshenko, M. G., Doucet, D., Davies, P. L. & Walker, V. K. The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nat. Biotechnol.* **15**, 887–890. <https://doi.org/10.1038/nbt0997-887> (1997).
61. Klasberg, S., Bitard-Feidel, T. & Mallet, L. Computational identification of novel genes: Current and future perspectives. *Bioinform. Biol. Insights* **10**, 121–131. <https://doi.org/10.4137/BBI.S39950> (2016).
62. Gong, Z., King, M. J., Fletcher, G. L. & Hew, C. L. The antifreeze protein genes of the winter flounder, *Pleuronectes americanus*, are differentially regulated in liver and non-liver tissues. *Biochem. Biophys. Res. Commun.* **206**, 387–392. <https://doi.org/10.1006/bbrc.1995.1053> (1995).
63. Hurst, L. D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486. [https://doi.org/10.1016/s0168-9525\(02\)02722-1](https://doi.org/10.1016/s0168-9525(02)02722-1) (2002).
64. Swanson, W. J. & Aquadro, C. F. Positive Darwinian selection promotes heterogeneity among members of the antifreeze protein multigene family. *J. Mol. Evol.* **54**, 403–410. <https://doi.org/10.1007/s00239-001-0030-0> (2002).
65. Hayes, P. H., Davies, P. L. & Fletcher, G. L. Population differences in antifreeze protein gene copy number and arrangement in winter flounder. *Genome* **34**, 174–177 (1991).
66. Hew, C. L. *et al.* Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *J. Biol. Chem.* **263**, 12049–12055 (1988).
67. Eirin-Lopez, J. M., Rebordinos, L., Rooney, A. P. & Rozas, J. The birth-and-death evolution of multigene families revisited. *Genome Dyn.* **7**, 170–196. <https://doi.org/10.1159/000337119> (2012).
68. Pickett, M. H., Hew, C. L. & Davies, P. L. Seasonal variation in the level of antifreeze protein mRNA from the winter flounder. *Biochim. Biophys. Acta* **739**, 97–104. [https://doi.org/10.1016/0167-4781\(83\)90049-0](https://doi.org/10.1016/0167-4781(83)90049-0) (1983).
69. Myers, E. W. *et al.* A whole-genome assembly of Drosophila. *Science* **287**, 2196–2204. <https://doi.org/10.1126/science.287.5461.2196> (2000).

Acknowledgements

We thank Eric Clelland, Dave Riddell, and other staff at the Bamfield Marine Sciences Centre, Bamfield, BC for collecting and shipping starry flounder blood. Amplicon Express (Pullman, WA) made two BAC libraries and corresponding nylon filters. The McGill University and Génome Québec Innovation Centre provided high-quality PacBio sequencing and assembly services without which this project would not have been possible. We are grateful to Nick Ostan for mapping the BACs and to Gary K. Scott and Kyra K. Nabeta for providing genomic DNA and Virginia K. Walker for comments on the manuscript. This work was supported by Canadian Institutes of Health Research Foundation award (FRN 148422) to P.L.D., who holds the Canada Research Chair in Protein Engineering.

Author contributions

L.A.G., S.Y.G. and P.L.D. designed research; L.A.G. and S.Y.G. performed research; L.A.G. analyzed data; and L.A.G. and P.L.D. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12446-4>.

Correspondence and requests for materials should be addressed to P.L.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022