

ORIGINAL RESEARCH

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Evolution of Type II Antifreeze Protein Genes in Teleost Fish: A Complex Scenario Involving Lateral Gene Transfers and Episodic Directional Selection

Ulf Sorhannus

Department of Biology and Health Services, Edinboro University of Pennsylvania, Edinboro, Pennsylvania, USA.
Corresponding author email: usorhannus@edinboro.edu

Abstract: I examined hypotheses about lateral transfer of type II antifreeze protein (AFP) genes among “distantly” related teleost fish. The effects of episodic directional selection on amino acid evolution were also investigated. The strict consensus results showed that the type II AFP and type II antifreeze-like protein genes were transferred from *Osmerus mordax* to *Clupea harengus*, from the ancestral lineage of the *Brachyopsis rostratus*—*Hemitripterus americanus* clade to the ancestor of the *Hypomesus nipponensis*—*Osmerus mordax* group and from the ancestral lineage of *Brachyopsis rostratus*—*Hemitripterus americanus*—*Siniperca chuatsi*—*Perca flavescens* to *Perca flavescens*. At the present time, the available evidence is more consistent with the LGT hypothesis than with other alternative explanations. The overall results indicate that evolutionary history of the type II AFP gene is complex, and that episodic directional selection was instrumental in the evolution of this freeze-preventing protein from a C-type lectin precursor.

Keywords: teleostei, antifreeze protein, lateral gene transfer, episodic directional selection

Evolutionary Bioinformatics 2012:8 535–544

doi: [10.4137/EBO.S9976](https://doi.org/10.4137/EBO.S9976)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.

Introduction

Many prokaryotes and eukaryotes that are found in cold environments survive freezing because their cells produce antifreeze proteins (AFPs).^{1–9} These molecules function by either preventing endogenous fluids from freezing, or by enabling survival despite the freezing of these fluids. Based on structural differences, five different types of AFPs have been described for teleost fish.^{5,10,11} Type II AFP,^{5,11} shows a high degree of structural similarities among “distantly” related taxa, such as between *Clupea harengus* (herring) and the *Osmerus mordax* (rainbow smelt)—*Hypomesus nipponensis* (Japanese smelt) clade, *Hemitripterus americanus* (sea raven)—*Brachyopsis rostratus* (longsnout poacher) lineage, *Siniperca chuatsi*. *C. harengus* is a member of Ostarioclupeomorpha (order: Clupeiformes) whereas *O. mordax* (order: Osmeriformes), *H. nipponensis* (order: Osmeriformes), *H. americanus* (order: Scorpaeniformes), *B. rostratus* (order: Scorpaeniformes) and *Siniperca chuatsi* (order: Perciformes, see below) are classified in the Euteleosteomorpha (also known as Euteleostei) (see Fig. 1).¹² According to Wiley and Johnson,¹² the infraclass Teleostei is divided into 4 major monophyletic taxa or cohorts, including Elopomorpha, Osteoglossomorpha, Ostarioclupeomorpha and Euteleosteomorpha (=Euteleostei).

Herring, rainbow and Japanese smelts produce type II AFPs that require calcium ions for their antifreeze activity (ie, Ca^{2+} -dependent) whereas type II AFPs produced by the sea raven and longsnout poacher do not (ie, Ca^{2+} -independent).^{5,11} Due to this,

it has been speculated that the calcium-dependent and independent type II AFPs have different mechanisms for preventing ice formation.⁵ It is also noteworthy that this group of freeze-preventing proteins, which are thought to have evolved from the sugar-binding domain of C-type (Ca^{2+} -dependent) lectins, have 10 cysteine residues, giving rise to 5 disulfide bridges in identical positions.^{5,11,13,14} Five disulfide bridges are unique to the type II AFP, whereas C-type lectins typically form 2 to 3 disulfide bridges (see Fig. 1 by Graham and colleagues¹¹).

Evolutionary analyses have been conducted to further understand the remarkable structural similarity of the type II AFP genes in “distantly” related teleost fish species. One investigation, that was based on several independent pieces of evidence, proposed lateral gene transfer (LGT) as a possible explanation.¹¹ However, the common origin and independent evolution (ie, convergent evolution) hypotheses can potentially also account for the taxonomic distribution/conservation of type II AFP genes. The common origin hypothesis holds that the type II AFP gene evolved once in the most recent common ancestor of all the type II AFP producing species, after which it remained highly static and was lost independently at least 3 different times.⁵ The convergent evolution hypothesis explains the taxonomic distribution of the similar type II AFP genes as being the result of at least three independent evolutionary events. In the current study, the notion of LGT was further tested using methods implemented in the program Prunier,¹⁵ and further discussed in light of the convergent evolution

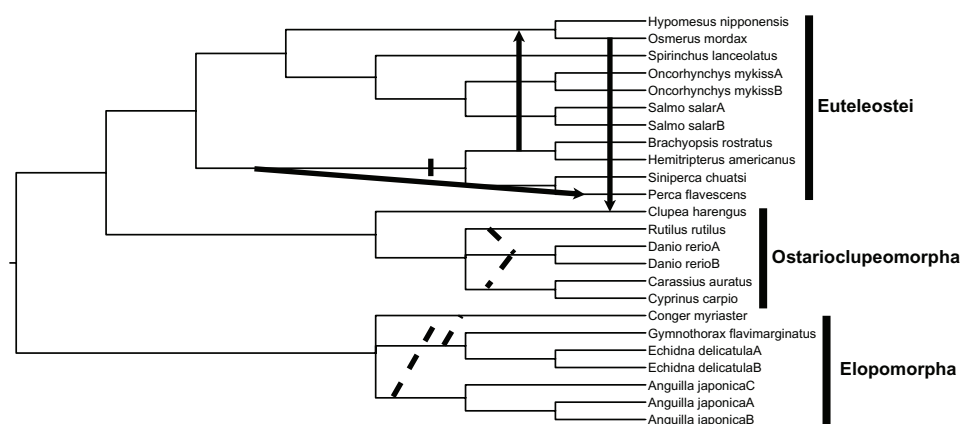


Figure 1. A rooted species phylogeny showing 3 lateral gene transfers (LGTs).

Notes: Arrows indicate the direction of the gene transfers. The dotted lines highlight the uncertain phylogenetic positions of *D. rerio* and *C. myriaster* (see text for additional details). The cross hatch specifies the internal branch on which the cysteine in amino acid site 110 (see table 2) most likely evolved; that is, when assuming that the LGT events happened as shown here (see text for details).

and common ancestor hypotheses. In addition, effects of episodic directional selection on codon evolution, especially those sites coding for the 10 cysteines involved in disulfide bridge formation, were assessed using a mixed effects model of evolution (MEME)¹⁶ available on the datamonkey website.¹⁷ The results of the various analyses indicated that at least three lateral transfer events involving genes coding for the type II AFP/antifreeze-like protein (AFLP) have occurred, and that episodic directional selection has affected the evolution of the two cysteine-containing sites that are unique to the type II AFP.

Material and Methods

Sequences

The nucleotide sequences of all the taxa shown in Figure 2 in the paper by Liu and colleagues⁵ were downloaded from the GenBank, except for *Pimephales promelas*, for which no codon data was available (Table 1). To increase taxon sampling, nucleotide sequences of *Siniperca chuatsi* and *Perca flavescens*, which were determined through multiple NCBI blasts

to be similar to the type II AFPs of *C. harengus*, *O. mordax*, *H. nipponensis*, *H. americanus* and *B. rostratus*, were also included in the current study (Table 1). All these taxa are characterized by having 10 cysteine sites, except *P. flavescens*, which has only 9 cysteine sites, suggesting that it is uncertain whether this fish species produces a type II AFP. Due to this circumstance, this particular protein will here be referred to as a type II AFLP.

Data analyses

The software package DAMBE (Version 5.2.57)¹⁸ was implemented to manage the data and to match the codons against the aligned amino acid sequences. The alignments of the amino acid (Table 2), which are available upon request, were obtained using the MAFFT (Version 6)¹⁹ web server (available at <http://mafft.cbrc.jp/alignment/server/>). The alignment mode G-INS-i,²⁰ with an offset value of 0.2 and the MAFFT homolog function “turned on” (all the other parameter settings were default values), was used for the amino acids.

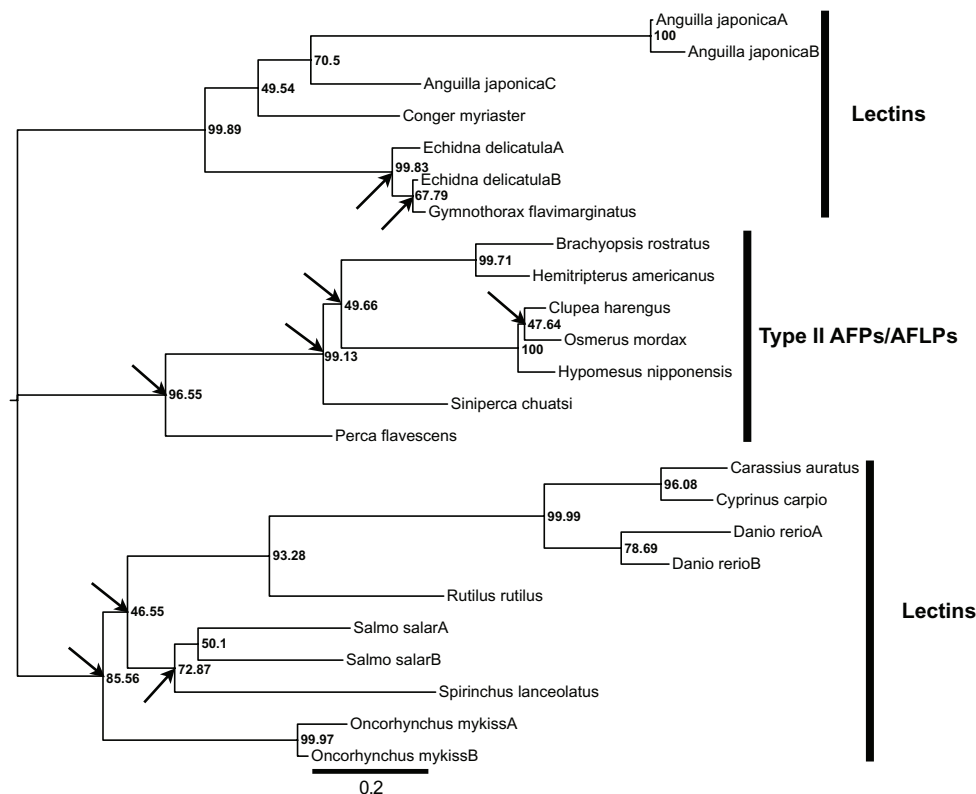


Figure 2. An unrooted maximum likelihood tree of the type II AFP/AFLPs and lectins.

Notes: The incongruent nodes (ie, bipartitions) between the gene and species trees are indicated by arrows. The values next to the nodes are LR-ELW edge supports. The scale bar shows the number of substitutions per nucleotide.

Table 1. Taxa and GenBank accession numbers/source references of the type II AFPs/AFLPs and lectins used in the study.

Species	GenBank accession number/source references
<i>Clupea harengus</i>	L14722
<i>Hypomesus nipponensis</i> ³⁵	
<i>Osmerus mordax</i>	M96154
<i>Brachyopsis rostratus</i>	AB283044
<i>Hemimipterus americanus</i>	J05100
<i>Siniperca chuatsi</i>	EU719616
<i>Perca flavescens</i>	FJ826540
<i>Oncorhynchus mykiss</i> A	AF363271
<i>Oncorhynchus mykiss</i> B	AM041212
<i>Salmo salar</i> A	AY191312
<i>Salmo salar</i> B	CB504037
<i>Spirinchus lanceolatus</i>	AB114832
<i>Rutilus rutilus</i>	EG543056
<i>Cyprinus carpio</i>	AB034807
<i>Carassius auratus</i>	AY157616
<i>Danio rerio</i> A	NM_001245031
<i>Danio rerio</i> B	XM_003199234
<i>Anguilla japonica</i> A	AB060539
<i>Anguilla japonica</i> B	AB060538
<i>Anguilla japonica</i> C	AB050703
<i>Echidna delicatula</i> A	AB183016
<i>Echidna delicatula</i> B	AB099491
<i>Gymnothorax flavimarginatus</i>	AB099490
<i>Conger myriaster</i>	AB239175

A species phylogeny was created, using the ITOL web server, based on the NCBI taxonomic classification (available at http://itol.embl.de/other_trees.shtml) and the information reported by Liu and colleagues.⁵ In the resulting tree, the phylogenetic relationships of *Rutilus rutilus*, *Danio rerio* and *Conger myriaster* (outgroup lineage) were unresolved (Fig. 1). Since a resolved species topology is required for the LGT analyses, additional systematic information of the aforementioned species was used in an effort to infer their most likely phylogenetic positions in the teleost tree. Comprehensive phylogenetic studies of Cypriniformes, based on the mitochondrial genome data or/and multiple nuclear genes, indicated that *D. rerio* is more closely related to a large cyprinid clade, made up of taxa such as Leuciscinae than to the monophyletic group consisting of representatives of Cyprininae.^{21,22} However, the opposite relationship is also possible, that is that *D. rerio* is more closely related to the monophyletic group consisting of representatives of

Cyprininae than to the large cyprinid clade, made up of taxa such as Leuciscinae.²¹ *R. rutilus* is considered to belong to the latter subfamily, whereas *Cyprinus carpio* and *Carassius auratus* are part of the former subfamily.^{21,23} No trees reported by Mayden and colleagues²¹ or Saitoh and colleagues²² suggested that *R. rutilus* (ie, the Leuciscinae) shares a more recent common ancestor with the clade composed of *Cyprinus carpio* and *Carassius auratus* than it does with the *D. rerio* lineage. It was also deduced from a study, using mitochondrial gene order information, that the phylogenetic position of *C. myriaster* is unresolved with regard to Anguillidae (ie, *Anguilla japonica*) and to the Muraenidae (*Echidna delicatula*, *Gymnothorax flavimarginatus*) clades.²⁴ Consequently, based on the currently available phylogenetic studies of *Rutilus rutilus*, *Danio rerio* and *Conger myriaster*, it is not possible to come up with a single resolved species tree, but rather four different resolved species trees (Fig. 1). All 4 phylogenies were the used in the LGT analyses. The software package FigTree (available at <http://tree.bio.ed.ac.uk/software/figtree/>) was used to generate the trees shown in Figures 1 and 2.

Prunier (version 2.0),¹⁵ a program that works in conjunction with Treefinder,²⁵ was used to detect potential LGT events. Both the “slow” and the “fast” methods were implemented in this investigation. The “slow” method uses Kishino-Hasegawa²⁶ Shimodaira-Hasegawa²⁷ expected likelihood weights²⁸ and the approximately unbiased²⁹ tests to identify statistically significant topological differences between a gene and a species tree. LGT events are inferred in the part of the tree where statistically significant topological conflicts between the species and gene trees exist (see details by Abby and colleagues¹⁵). The gene tree was inferred, in both the “fast” and “slow” Prunier analyses, based on the nucleotide alignment and the model GTR + G8 + I. The “fast” method works by finding a maximum statistical agreement forest (MSAF) between the gene and species trees (see details by Abby and colleagues¹⁵). MSAF is defined as the minimum number of branches that are required to be eliminated in order to obtain statistical agreement between the two topologies.¹⁵ LGT events were inferred in situations in which branches with statistically significant support (ie, support of 95% or higher) must be cut in order to achieve MSAF.

Table 2. Amino acid alignment of type II AFPs/AFLPs and lectins.

Clupea_harengus
 Osmerus_mordax
 Hypomesus_hipponensis
 Brachyopsis_rostratus
 Hemitripterus_americanus
 Siniperca_chuatsi
 Perca_flavescens
 Anguilla_japonicaA
 Anguilla_japonicaB
 Anguilla_japonicaC
 Carassus_auratus
 Conger_myriaster
 Cyprinus_carpio
 Danio_rerioA
 Danio_rerioB
 Echidna_delicatulaA
 Echidna_delicatulaB
 Gymnothorax_flavimarginatus
 Oncorhynchus_mykissA
 Oncorhynchus_mykissB
 Rutilus_rutilus
 Salmo_salarA
 Salmo_salarB
 Spirinchus_lanceolatus

Clupea_harengus
 Osmerus_mordax
 Hypomesus_hipponensis
 Brachyopsis_rostratus
 Hemitripterus_americanus
 Siniperca_chuatsi
 Perca_flavescens
 Anguilla_japonicaA
 Anguilla_japonicaB
 Anguilla_japonicaC
 Carassus_auratus
 Conger_myriaster
 Cyprinus_carpio
 Danio_rerioA
 Danio_rerioB
 Echidna_delicatulaA
 Echidna_delicatulaB
 Gymnothorax_flavimarginatus
 Oncorhynchus_mykissA
 Oncorhynchus_mykissB
 Rutilus_rutilus
 Salmo_salarA
 Salmo_salarB
 Spirinchus_lanceolatus

Note: The numbers specify the amino acid sites (see text for additional information) and (*) shows sites with no variability across taxa.

The “fast” Prunier method calculates Expected-Likelihood Weights applied to Local Rearrangements (LR-ELW) edge support values²⁸ (see Fig. 2) for the gene tree. The aligned codon sequences and the four different species phylogenies (see previous paragraph) were provided as input for all Prunier analyses. For the “slow” method the default parameter values were used. In the “fast” analyses, the following parameters were assigned values that were different from the default settings: `boot.thresh.conflict` = 95 (ie, cut-off support value for topological conflict); `fwd.depth` = 1 (ie, maximal depth at which Prunier looks forward to find a significant LGT event when the current LGT is not significant).

A MEME¹⁶ was implemented for the purpose of detecting possible episodic directional selection on codon sites that has occurred in a small number of branches in the type II AFP gene tree. MEME models variable dN/dS across lineages at a given codon site, such that a certain fraction of branches are allowed

to evolve neutrally or under negative selection, whereas the remaining proportion is permitted to evolve under episodic directional selection. To test for evidence of this type of selection, a likelihood ratio test is performed between the aforementioned model and the nested null model that forces parameter values for episodic directional selection to vary between 0 and 1. Prior to running the MEME analysis the codon alignments of type II AFP were screened for recombination events using a program called GARD³⁰ with the following settings: Site-to-site variation = general discrete and rate classes = 4. Recombination has the potential to confound inferences of codon selection and, consequently, needs to be accounted for before running the analyses. No significant break points in the current alignment were detected by the GARD algorithm. The model selection tool, available in Datamonkey,¹⁷ chose the following optimal model for the MEME analysis: 001102 with AIC of 15118.9.

Results

Species and gene trees

The rooted species phylogeny, generated based on the NCBI taxonomic classification and published systematic studies of teleost fish, is shown in Figure 1. As noted in the material and methods section, the phylogenetic positions of *R. rutilus*, *D. rerio* and *C. myriaster* (outgroup lineage) are, at the present time, uncertain. The dotted lines in Figure 1 show the most probable relationships of the aforementioned species. Due to these uncertainties, 4 possible resolved species phylogenies were generated for the Prunier analysis.

Figure 2 shows the unrooted gene tree that was derived from the aligned nucleotide data in the 4 Prunier runs. A number of major conflicts, indicated by the arrows in Figure 2, are apparent between the gene and species trees. In the former, all of the type II AFP and AFLP producing taxa form a monophyletic group, which is not the case in the latter, in which *C. harengus* constitutes a deep branch within the Ostarioclupeomorpha clade and the *H. nipponensis*—*O. mordax* cluster share a most recent common ancestor with the *S. salar*—*O. mykiss*—*S. lanceolatus* lineage. Moreover, *P. flavescens* and *S. chuatsi* show a sister relationship in the species tree but not in the gene tree, in which they form a paraphyletic group. The outgroup, or the Elopomorpha clade, is similar in both topologies, except that the phylogenetic position of *C. myriaster* is unresolved in the species tree. The *S. salar*—*O. mykiss*—*S. lanceolatus* lineage share a common ancestor with Ostarioclupeomorpha in the gene tree but not in the species phylogeny. The branching order of the taxa within Ostarioclupeomorpha are resolved in the gene tree but not in the species tree.

Lateral gene transfer

The results of the Prunier runs differed somewhat depending on the species phylogeny and method used. The solid arrows, which indicate the direction of LGTs (Fig. 1), constitute the strict consensus results obtained from all 8 analyses; that is, when implementing 4 different species trees for each method (ie, either the “fast” or “slow” method). These events included the transfer of type II AFP gene from *O. mordax* to *C. harengus*, from the ancestral lineage of the *B. rostratus*—*H. americanus* clade to the ancestor of

the *H. nipponensis*—*O. mordax* group and from the ancestral lineage of *B. rostratus*—*H. americanus*—*S. chuatsi*—*P. flavescens* to *P. flavescens* (Fig. 1). Because of the low support value (ie, 47.6%) for the node uniting *O. mordax* and *C. harengus* (Fig. 2), the relationship among *O. mordax*, *C. harengus*, and *H. nipponensis* can be interpreted as being unresolved. Due to this circumstance, one can consider either *O. mordax* or *H. nipponensis* as possible donors (the specific donor being unresolved) of the type II AFP gene to *C. harengus*. Four out of the 8 Prunier runs suggest that *R. rutilus* has received the C-type lectin gene from the ancestral lineage of the Cypriniformes clade within Ostarioclupeomorpha. This scenario was obtained from both the “fast” and “slow” analyses; that is, when the input species tree in which *R. rutilus* is the sister lineage of *D. rerio* (Fig. 1) was used. There are two versions of this phylogeny because *C. myriaster* most likely constitutes a sister lineage of either the Anguillidae or the Muraenidae (see Fig. 1).²⁴ When one of these two possible species phylogenies was used in conjunction with the “slow” method, the *E. delicatula*—*G. flavimarginatus* clade received the C-lectin gene from the *A. japonica*—*C. myriaster* cluster.

Episodic directional selection

The MEME¹⁶ analysis showed evidence for 16 codon/amino acid sites in the alignment having been influenced by episodic directional selection (significance level = 0.05) (see Tables 2 and 3). Two of them are cysteine-containing sites (ie, sites 71 and 110; see Table 2) that are involved in forming the unique 5th disulfide bridge of the type II AFP.^{5,11}

Discussion

Lateral gene transfer and the evolution of the type II AFP

The evolutionary history of the type II AFP gene appears to be rather complex, involving at least 3 events of LGT.¹¹ When it comes to improving our understanding of the evolution of this freeze-preventing protein, a key question involves the point in its phylogeny at which the 5th disulfide bridge evolved. The number of disulfide bridges in the C-type lectins and/or type II AFLPs vary quite a bit but are always less than 5.¹¹ Thus, The 5th disulfide bond, in addition to the other 4 disulfide bridges, makes the

Table 3. Codon (amino acid) sites found to be under episodic diversifying selection.

Codon	α	β^-	$\text{Pr}[\beta = \beta^-]$	β^+	$\text{Pr}[\beta = \beta^+]$	<i>P</i> -value	<i>q</i> -value
3	0.71267	0	0.87743	33.051	0.12257	0.0357189	0.446487
26	0.594053	0.219727	0.83171	8.61427	0.16829	0.0427582	0.493364
48	0.592732	0.0233223	0.499489	7.0793	0.500511	0.0111048	0.23796
58	0.626032	0.153589	0.886189	208.891	0.113811	0.0471447	0.441982
64	0.578003	0	0.743857	3.43805	0.256143	0.0465647	0.465647
71	0.134445	0	0.846411	32.9029	0.153589	9.0959e-05	0.0136439
76	0.706383	0.043358	0.572077	7.49808	0.427923	0.0137215	0.257279
80	1.14485	0	0.823905	8.58932	0.176095	0.044903	0.481103
84	5e-09	5e-09	0.694404	8.03456	0.305596	0.000501674	0.0376255
95	5e-09	2. 5e-17	0.589632	3.53825	0.410368	0.0140873	0.234789
99	0.608495	0	0.854247	10.7894	0.145753	0.00485977	0.242988
107	0.320033	0.320033	0.938072	40.1285	0.061928	0.0289885	0.434828
110	0.330462	0.282797	0.966787	190.886	0.033213	0.0102668	0.25667
116	5e-09	2. 5e-17	0.725722	7.7696	0.274278	0.00678347	0.25438
126	0.934956	0.871756	0.913601	222.694	0.086399	0.0303443	0.413787
130	0.637836	0.00500758	0.600822	6.96584	0.399178	0.0100967	0.3029

Notes: Codon = codon (or amino acid) number in the alignment; *alpha* = the maximum likelihood estimate (MLE) of the synonymous substitution rate *alpha*; β^- = the maximum likelihood estimate (MLE) of the non-synonymous rate for the branch class with β smaller or equal to *alpha*; $\text{Pr}[\beta = \beta^-]$ = the MLE of the proportion of sites evolving at β^- ; β^+ = the MLE of the unconstrained β non-synonymous rate; $\text{Pr}[\beta = \beta^+]$ = the MLE of the proportion of sites evolving at β^+ ; *P*-value = the probability value for the likelihood ratio test statistic for $\beta^+ = \alpha$ (null) versus β^+ *unrestricted* (alternative); *q*-value = false discovery rate under the strict neutral null.

type II AFPs unique relative to the closely related C-type lectins and/or type II AFLPs.^{5,11} The evolution of the 5th disulfide bridge requires the appearance of a cysteine in amino acid position 110 (Table 2 and Fig. 1), provided that it already existed in site 71. In fact, *C. harengus*, *O. mordax*, *H. nipponensis*, *H. americanus*, *B. rostratus* and *S. chuatsi* have this amino acid present in both of these sites, whereas *P. flavescentis* does not (ie, only in site 71 but not in 110) (Table 2), suggesting that *P. flavescentis* does not produce a protein with the 5th disulfide bridge. When optimizing the interspecific amino acid variation in site 110 (Table 2) on the species tree in Figure 1, it becomes evident—under the assumption of 3 LGT events—that the cysteine evolved in this position once in the ancestral lineage of the *B. rostratus*—*H. americanus*—*S. chuatsi*—*P. flavescentis* clade. Consequently, the occurrence of the 10 cysteines and the 5th disulfide bridge in the type II AFPs of *C. harengus*, *O. mordax* and *H. nipponensis*, and the absence of it in *P. flavescentis*, can be explained by 3 separate LGT events. However, it should be noted that this scenario assumes that *P. flavescentis* received a type II AFLP or C-type lectin gene, through LGT, from the ancestral lineage of the *B. rostratus*—*H. americanus*—*S. chuatsi*—*P. flavescentis* clade before the type II AFP gene evolved (Fig. 1). This observation further

suggests that the evolution of the 9th and 10th cysteine were temporally displaced, such that the former (ie, site 71) appeared prior to latter (ie, in site 110).

Based on the currently available data, it is not possible to determine—under the assumption of 3 LGTs—whether the calcium-dependent type II AFP gene evolved prior to the calcium-independent one or vice versa. Thus, two equally most-parsimonious scenarios, requiring a minimum of two changes when optimized on the species tree (Fig. 1), are possible. In the first situation the calcium-dependent type II AFP gene evolved first and was subsequently transferred from the ancestral lineage of the *B. rostratus*—*H. americanus*—*S. chuatsi*—*P. flavescentis* cluster to that of the *H. nipponensis*—*O. mordax* group (Fig. 1). This scenario assumes that calcium-independence evolved after the aforementioned transfer event (Fig. 1). In the second situation the calcium-independent type II AFP gene evolved first and was subsequently transferred from the ancestor of the *B. rostratus*—*H. americanus*—*S. chuatsi*—*P. flavescentis* clade to that of the *H. nipponensis*—*O. mordax* group. This interpretation assumes that the calcium-dependent type II AFP gene evolved in the ancestral lineage of *H. nipponensis*—*O. mordax*; that is, after it had been received from presumed donor.



In addition to the LGT scenario, two other hypotheses can be invoked to explain the distribution/conservation of type II AFP genes in the taxa included in the current study. Those are the common origin and the independent evolution (ie, convergent evolution) hypotheses. The common origin hypothesis, which is advocated by Liu and colleagues,⁵ holds that the type II AFP gene evolved once in the most recent common ancestor of all the type II AFP producing species after which it was lost independently at least 3 different times.⁵ To explain the remarkable taxonomic conservation/distribution and the evolution of the calcium-independent type II AFP gene, the common origin hypothesis assumes two independent losses of calcium-binding sites⁵ and that the gene has remained highly static since it evolved about 280 million years ago (ie, the estimated divergence time between Ostarioclupeomorpha and Euteleostei³¹). This is especially true with regard to the *C. harengus* and *O. mordax* divergence. Given the large amount of time that has passed since its origin in the common ancestor (approximately 280 million years ago), the introns of the type II AFP genes in *C. harengus* and *O. mordax*/*H. nipponensis* are expected to have differentiated to a degree similar to that seen in, for instance, the intron of the conserved spliceosomal protein gene (Prp8p). Graham and colleagues¹¹ have shown that the former sequences are much more conserved than the latter. This observation, in addition to the assumption that the type II AFP gene has remained largely unchanged for over 200 million years, makes the common origin hypothesis a less likely explanation at this time for the distribution/conservation of the type II AFP genes among the taxa included here.

The convergent evolution hypothesis explains the taxonomic distribution/conservation of the type II AFP genes analyzed here as being the result of a number of independent origination events. Independent evolution of a gene in “distantly” related lineages can sometimes be difficult to discern from LGT because both processes can give rise to sequences with similar nucleotide compositions. However, transferred nucleotide sequences (ie, as opposed to convergent evolution) are more likely to show distinctive evolutionary patterns relative to the other genes (regions) in the genome of the putative recipient species, while at the same time exhibiting substitution

patterns more similar to those in the assumed donor genome. In fact, Graham and colleagues¹¹ investigated molecular evolutionary patterns, such as intron substitution rates and codon usage bias in exons, in the *C. harengus*—*O. mordax*—*H. nipponensis*—*H. americanus* complex and found an unexpectedly high degree of sequence conservation among introns in the type II AFP genes in the above taxa, but also a lack of intron sequence conservation in the otherwise well-conserved spliceosomal protein gene (Prp8p). With regard to codon usage and GC₃ content in the amino acid coding regions of the type II AFP gene, Graham and colleagues¹¹ found patterns largely consistent with random use, while for the Prp8p genes there was evidence for nonrandom use or selection for increased GC content in the third codon positions. Moreover, the convergent evolution hypothesis, as opposed to the LGT scenario, assumes that some of the disulfide bridges in the type II AFP genes arose independently in the same location at least three different times.¹¹ Taken together, it can be concluded that at the present time, the available evidence is more consistent with the LGT hypothesis than with the common origin and the independent evolution (ie, convergent evolution) hypotheses (also see the paper by Graham and colleagues¹¹ for additional evidence and discussion).

Episodic directional selection

The MEME analysis showed that the cysteine containing sites 71 and 110 (Table 2), which are responsible for the formation of the 5th disulfide bridge, have been affected by episodic directional selection (Table 3). This result suggests that this mechanism was instrumental in driving the evolution of the type II AFP gene from a C-type lectin precursor. Such an evolutionary event is highly likely to have been adaptive in nature since directional positive selection is sometimes thought to be associated with adaptive (functional) changes in protein-coding genes;^{32,33} that is, by favoring fitness-enhancing mutations. An earlier study suggested that episodic positive selection also was important in driving the differentiation of duplicated antifreeze protein genes in two *Fragilariopsis* species after the ancestral lineage acquired the sequence from the basidiomycetes.³³ It is also noteworthy that none of the ice-binding sites (here Thr104 and Thr106)

and Ca²⁺-coordinating residues (here Asp102 and Glu109) identified in the *C. harengus* type II AFP by Graham and colleagues⁵ showed any evidence of having been influenced by episodic directional selection (Table 2).

Acknowledgements

I thank Laurie Graham, Peter L. Davies and the authors (ie, Yang Liu, Zhengjun Li, Qingsong Lin, Jan Kosinski, J. Seetharaman, Janusz M. Bujnicki, J. Sivaraman, Choy-Leong Hew) of the paper entitled “Structure and Evolutionary Origin of Ca²⁺-Dependent Herring Type II Antifreeze Protein” for sending me the data used in their studies of the type II antifreeze protein. The author is grateful for the constructive comments provided by two anonymous reviewers.

Funding Sources

Edinboro University of Pennsylvania provided financial support for the research project.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Author Contributions

Conceived and designed the experiments: US. Analyzed the data: US. Wrote the first draft of the manuscript: US. Contributed to the writing of the manuscript: US. Agree with manuscript results and conclusions: US. Jointly developed the structure and arguments for the paper: US. Made critical revisions and approved final version: US.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any

disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Duman JG. Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annu Rev Physiol.* 2001;63:327–57.
2. Hoshino T, Kiriaki M, Ohgiya S, et al. Antifreeze proteins from snow mold fungi. *Can J Bot.* 2003;81:1175–81.
3. Griffith M, Yaish MW. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Sci.* 2004;9:399–405.
4. Janech MG, Krell A, Mock T, Kang JS, Raymond JA. Ice-binding proteins from sea ice diatoms (Bacillariophyceae). *J Phycol.* 2006;42:410–6.
5. Liu Y, Li Z, Lin Q, et al. Structure and evolutionary origin of Ca(2+)-dependent herring type II antifreeze protein. *PLoS ONE.* 2007;2(6):e548.
6. Raymond JA, Fritsen C, Shen K. An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol Ecol.* 2007;61:214–21.
7. Raymond JA, Janech MG, Fritsen CH. Novel ice-binding proteins from a psychrophilic antarctic alga (Chlamydomonadaceae, Chlorophyceae). *J Phycol.* 2009;45:130–6.
8. Raymond JA, Janech MG. Ice-binding proteins from enoki and shiitake mushrooms. *Cryobiology.* 2009;58:151–6.
9. Kiko R. Acquisition of freeze protection in a sea-ice crustacean through horizontal gene transfer? *Polar Biol.* 2009;33:543–56.
10. Fletcher GL, Hew CL, Davies PL. Antifreeze proteins of teleost fishes. *Annu Rev Physiol.* 2001;63:359–90.
11. Graham LA, Loughheed SC, Ewart KV, Davies PL. Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PLoS ONE.* 2008;3(7):e2616.
12. Wiley EO, Johnson GD. A Teleost Classification Based On Monophyletic Groups. In: Nelson JS, Schultze HP, Wilson MVH, editors. *Origin and Phylogenetic Relationships of Teleosts*. Munich: Verlag Dr. Friedrich Pfeil.
13. Drickamer K. C-type lectin-like domains. *Curr Opin in Struct Biol.* 1999;9:585–90.
14. Ewart KV, Rubinsky B, Fletcher GL. Structural and functional similarity between fish antifreeze proteins and calcium-dependent lectins. *Biochem Biophys Res Commun.* 1992;185:335–40.
15. Abby SS, Tannier E, Gouy M, Daubin V. Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. *BMC Bioinformatics.* 2010;11:324.
16. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8(7):E1002764.
17. Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics.* 2010;26:2455–7.
18. Xia X. *Data Analysis in Molecular Biology and Evolution*. Boston: Kluwer Academic Publishers; 2001.
19. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
20. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33:511–8.
21. Mayden RL, Chen WJ, Bart HL, et al. Reconstructing the phylogenetic relationships of the Earth's most diverse clade of freshwater fishes – order Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the mitochondrial genome. *Mol Phylogenet Evol.* 2009;51:500–14.
22. Saitoh K, Sado T, Mayden RL, et al. Mitogenomic Evolution and Interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The First Evidence Toward Resolution of Higher-Level Relationships of the World's Largest Freshwater Fish Clade Based on 59 Whole Mitogenome Sequences. *J Mol Evol.* 2006;63:826–41.
23. He S, Mayden RL, Wang X, et al. Molecular phylogenetics of the family Cyprinidae (Actinopterygii: Cypriniformes) as evidenced by sequence variation in the first intron of S7 ribosomal protein-coding gene: further evidence from a nuclear gene of the systematic chaos in the family. *Mol Phylogenet Evol.* 2008;46:818–29.



24. Inoue JG, Miya M, Tsukamoto K, Nishida M. Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: Anguilliformes): novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for anguilliform families. *J Mol Evol*. 2001;52:311–20.
25. Jobb G, von Haeseler A, Strimmer K. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol*. 2004;4:18.
26. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol*. 1989;29:170–9.
27. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol*. 1999;16:1114–6.
28. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc Biol Sci*. 2002;269:137–42.
29. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. 2002;51:492–508.
30. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 2006;23:1891–901.
31. Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evol Biol*. 2008;8:215.
32. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nat Rev Genet*. 2002;3:137–44.
33. Beisswanger S, Stephan W. Evidence that strong positive selection drives neofunctionalization in the tenderly duplicated polyhomeotic genes in *Drosophila*. *Proc Natl Acad Sci U S A*. 2008;105:5447–52.
34. Sorhannus U. Evolution of antifreeze protein genes in the diatom genus *fragilariopsis*: evidence for horizontal gene transfer, gene duplication and episodic diversifying selection. *Evol Bioinform Online*. 2011;7:279–89.
35. Yamashita Y, Miura R, Takemoto Y, Tsuda S, Kawahara H, Obata H. Type II antifreeze protein from a mid-latitude freshwater fish, Japanese Smelt (*Hypomesus nipponensis*). *Biosci Biotechnol Biochem*. 2003;67:461–6.