

OPEN ACCESS

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

Evolution of Antifreeze Protein Genes in the Diatom Genus *Fragilariopsis*: Evidence for Horizontal Gene Transfer, Gene Duplication and Episodic Diversifying Selection

Ulf Sorhannus

Department of Biology and Health Services, Edinboro University of Pennsylvania, Edinboro, Pennsylvania 16444, USA.
Corresponding author email: usorhannus@edinboro.edu

Abstract: Hypotheses about horizontal transfer of antifreeze protein genes to ice-living diatoms were addressed using two different statistical methods available in the program Prunier. The role of diversifying selection in driving the differentiation of a set of antifreeze protein genes in the diatom genus *Fragilariopsis* was also investigated. Four horizontal gene transfer events were identified. Two of these took place between two major eukaryote lineages, that is from the diatom *Chaetoceros neogracile* to the copepod *Stephos longipes* and from a basidiomycete clade to a monophyletic group, consisting of the diatom species *Fragilariopsis curta* and *Fragilariopsis cylindrus*. The remaining two events included transfers from an ascomycete lineage to the proteobacterium *Stigmatella aurantiaca* and from the proteobacterium *Polaribacter irgensii* to a group composed of 4 proteobacterium species. After the *Fragilariopsis* lineage acquired the antifreeze protein gene from the basidiomycetes, it duplicated and went through episodic evolution, characterized by strong positive selection acting on short segments of the branches in the tree. This selection pattern suggests that the paralogs differentiated functionally over relatively short time periods. Taken together, the results obtained here indicate that the group of antifreeze protein genes considered here have a complex evolutionary history.

Keywords: diatom, antifreeze protein, horizontal gene transfer, episodic evolution, positive selection, sea ice

Evolutionary Bioinformatics 2011:7 279–289

doi: [10.4137/EBO.S8321](https://doi.org/10.4137/EBO.S8321)

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

Genes coding for antifreeze proteins (AFPs) have been discovered in various taxa that occur in cold temperatures; these include unicellular eukaryotes, plants, bacteria, fungi, fish, crustaceans and insects.^{1–9} Two major groups of AFPs are recognized. One type prevents cell/body fluids from freezing while the other kind consists of proteins that make it possible for organisms to survive cell/body fluid freezing.^{1,5} It is thought that both AFPs and protein ice nucleators play an important role in both freeze avoidance and tolerance.^{1,5} The former prevents freezing while the latter limits supercooling and induce freezing.¹ In some prokaryotes and eukaryotes that survive cell/body fluid freezing, the AFPs act as cryoprotectants. Even though this process is not fully understood, it may take place through ice recrystallization inhibition and possibly also by cell membrane stabilization.^{5,14} Those organisms that survive through freeze avoidance have AFPs that lower the freezing temperature of body fluids noncolligatively without affecting the melting temperature.^{1,5}

Genes coding for proteins with antifreeze properties have probably evolved independently in a number of lineages¹⁰ while some organisms have most likely acquired the genes through horizontal gene transfer (HGT).^{4,11,12} The evolution of prokaryote genomes is thought to have been profoundly influenced by lateral gene transfer.¹³ The effect of this process on eukaryote genome evolution appear to be much more prevalent than previously thought. This is evidenced by an increasing number of documented HGTs in eukaryotes, particularly in unicellular eukaryotes.^{13,14} For instance, the nuclear genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* appear to contain a large number of prokaryote genes.^{15–17} This has probably also contributed to the great success of the diatoms in different ecosystems.¹⁶

All the currently known AFPs in ice-living diatoms appear to act as cryoprotectants.^{4,11} AFPs, which show similar amino acid sequences to those found in diatoms, have also been described for different prokaryote, fungal and crustacean lineages.^{4,11} This observation suggests that diatoms may have acquired the AFP genes from distantly related taxa through HGTs.^{4,11} However, another possibility, which cannot

completely be ruled out, is convergent evolution of AFP genes in diatoms and certain prokaryote, fungal and crustacean lineages.^{4,11}

In this study, hypotheses about horizontal transfer of AFP genes to sea-ice living diatoms from distantly related taxa were tested using the program Prunier—an algorithm developed by¹⁸ for inferring HGT events based on statistical criteria. The role of diversifying selection in driving the differentiation of a set of duplicated AFP genes in the diatom genus *Fragilariopsis* was also investigated.¹⁹ The results indicated that at least four statistically significant HGT events, involving AFP genes and genes coding for antifreeze-like proteins (AFLPs) took place. Two of these events included diatoms. It was also shown that episodic diversifying selection has most likely been instrumental in driving the differentiation of the paralogs in the two *Fragilariopsis* species, suggesting that the duplicated AFP genes have changed functionally over relatively short time periods.

Material and Methods

Sequences

The choice of taxa in the current study was largely influenced by the paper of,¹¹ that is one of the major goals was to test hypotheses of HGT using the taxa shown in Figure 2 in¹¹ Moreover, amino acid sequences similar to the various AFPs of the two *Fragilariopsis* species were used. The degree of similarity was determined through multiple NCBI blasts. Another aspect considered when choosing taxa was to include representatives of different higher level taxa (especially eukaryotes), that is as long as they were reasonably similar to the different AFP genes of the two *Fragilariopsis* species. The SSU rRNA data and the amino acid sequences (+CDS) of the AFPs considered here were downloaded from the GenBank. The accession numbers of all the downloaded SSU rRNA sequences and proteins are shown in Table 1. The amino acid sequences of the following taxa have been demonstrated to have antifreeze activities: *Stephos longipes*, *Flammulina populicola*, *Fragilariopsis curta*, *Fragilariopsis cylindrus*, *Typhula ishikariensis*, *Lentinula edodes*, *Navicula glaciei*, *Colwellia sp.*,¹¹ and references therein and *Chaetoceros neogracile*.²⁰ The remaining amino

**Table 1.** Taxa and GenBank accession numbers of the antifreeze/antifreeze-like proteins (AFP/AFLP) and the small-subunit ribosomal RNA (SSU rRNA) sequences used in the study.

Higher taxon	Species	AFP/AFLP #	SSU rRNA #
Euryarchaeota	<i>Candidatus Methanoregula boonei</i>	YP_001404652 YP_001403476	EU887826
Proteobacteria	<i>Shewanella denitrificans</i>	YP_562920 YP_562921	AY771743
	<i>Shewanella frigidamarina</i>	YP_749708	Y13699
	<i>Colwellia</i> sp.	ABH08428	HM771242
	<i>Psychromonas ingrahamii</i>	YP_944155	U73721
	<i>Rhodiferax ferrireducens</i>	YP_523138	AF435948
	<i>Stigmatella aurantiaca</i>	ZP_01462925	GU207882
Bacterioidetes	<i>Cytophaga hutchinsonii</i>	YP_676864	AB517710
	<i>Polaribacter irgensii</i>	ZP_01118128	EU000226
	<i>Marivirga tractuosa</i>	YP_004052221	CP002349
Bacillariophyta	<i>Fragilariopsis curta</i>	GQ265833 GQ265834 GQ265835 GQ265836 GQ265838 GQ265837 GQ265839 GQ265840	
Bacillariophyta	<i>Fragilariopsis curta</i>	GQ265842 GQ265843 GQ265841	
	<i>Fragilariopsis cylindrus</i>	GQ232744 GQ232745 GQ232746 GQ232747 GQ232748 GQ232749 GQ232750 GU001148 GU001149 GU001150 GU001151 GU001152 DR026070 EL737258	EF140624
	<i>Navicula glaciei</i>	AAZ76252 AAZ76253	EF106789
	<i>Chaetoceros neogracile</i>	EL622418	EU090012
Copepods	<i>Stephos longipes</i>	ACL00837 ACL00838	
Basidiomycota	<i>Lentinula edodes</i>	ACL27145	FJ379280
	<i>Flammulina populicola</i>	ACL27144	NG_013173
	<i>Leucosporidium</i> sp.	ACU30806	GQ336996
	<i>Typhula ishikariensis</i>	BAD02891	AF026630
	<i>Typhula ishikariensis</i>	BAD02897	
Ascomycota	<i>Talaromyces stipitatus</i>	EED17205	AY526487
	<i>Phaeosphaeria nodorum</i>	XP_001806212	EU189213

acid sequences are considered to be AFLPs because it is not known if they possess antifreeze properties despite showing a high degree of similarity to those that exhibit such properties.

Data analyses

The software package DAMBE (Version 5.0.5),²¹ was implemented to manage the data and to match the codons against the aligned amino acid sequences. The alignments of the amino acid and SSU rRNA sequences, which are available upon request, were generated using the MAFFT (Version 6²²) Web Server available at <http://mafft.cbrc.jp/alignment/server/>. The alignment mode G-INS-i²³ with an offset value of 0.1 and the MAFFT homolog function “turned on” (all the other parameter settings were default values), was used for the proteins. When aligning the SSU rRNA sequences option Q-INS-i²⁴ was implemented. Only the highly conserved regions in the SSU rRNA alignment were used in the phylogenetic analysis.

A species phylogeny was created based on the NCBI taxonomic classification (http://itol.embl.de/other_trees.shtml). Since the relationships among some of the eubacteria lineages (Fig. 1) were

unresolved, a SSU rRNA phylogeny was inferred using BayesPhylogenies (available from <http://www.evolution.rdg.ac.uk/BayesPhy.html>). This program implemented a joint model that accommodated shifts in site-specific substitution rates over time (ie, heterotachy) and among-site rate heterogeneity (ie, “pattern heterogeneity”)^{25,26}; these phenomena are expected to occur in SSU rRNA-based phylogenetic analyses of prokaryote/eukaryote lineages. If heterotachy and “pattern heterogeneity” are not accounted for in phylogeny reconstruction, the resulting relationships are likely to be distorted.²⁶ A reversible-jump Markov chain Monte Carlo (rjMCMC) algorithm was used to determine how many distinct among-site rate-variation patterns, and branch length parameters (with a maximum of two parameters for each branch) were required to optimally describe the empirical data matrix. This approach is appealing because it requires far fewer parameters than conventional mixture models to describe heterotachy and “pattern heterogeneity”.^{25,26} A General Time Reversible (GTR) model of nucleotide substitution with discretized gamma-distributed rate variability (with 4 rate categories; Γ_4) was employed. Three independent MCMC (Markov chain

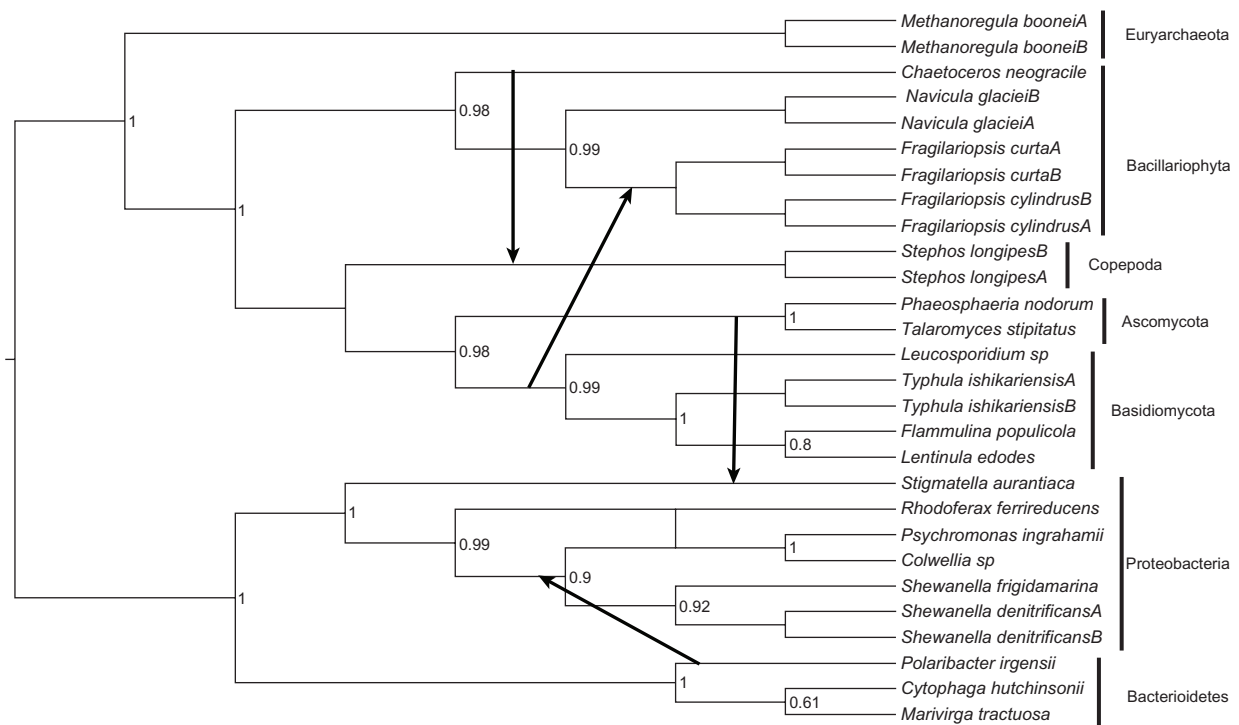


Figure 1. A rooted species phylogeny with the 4 horizontal gene transfer (HGT) events shown (see text for details).

Notes: Arrows indicate the direction of gene transfer. The HGTs were inferred based on minimum LR-ELW edge support values of 95% for recognizing topological conflict between the gene (Fig. 2) and species trees. The numbers next to each node show the posterior clade probabilities obtained in the Bayesian phylogenetic analysis of SSU rRNA sequences (note that all the taxa were not included in the analysis).



Monte Carlo) analyses (each with 3 chains running for 2×10^7 generations, sampling every 10^3 generations) were carried out to estimate the posterior distribution of phylogenetic trees, and post-burnin samples (with burnin set to 10%) from all analyses were combined. Convergence of the MCMC runs was determined by visually examining the cumulative posterior and between-run variation in split frequencies²⁷ using the on-line tool AWTY (“Are We There Yet”).²⁸ The software package FigTree (available from <http://tree.bio.ed.ac.uk/software/figtree/>) was used to generate the trees.

The program Prunier (version 2.0)¹⁸ which works in conjunction with Treefinder²⁹ was used to detect potential HGTs between the species included in this investigation. The “slow” Prunier method uses Kishino-Hasegawa,³⁰ Shimodaira-Hasegawa³¹ expected likelihood weights³² and the approximately unbiased³³ tests to infer whether topological differences between the gene and species trees are statistically significant. The “fast” Prunier method uses LR-ELW (Expected-Likelihood Weights applied to Local Rearrangements) edge support values³² to identify such discrepancies. HGT events are inferred when statistically significant topological conflicts between the species and gene trees were identified.¹⁸ Both the “fast” and “slow” methods, which in simulations have been shown to perform equally well¹⁸ were used in the current study. The aligned amino acid sequences and the species phylogeny were provided as input for the Prunier runs. In each Prunier run, the gene trees were inferred by the Treefinder program.²⁹ For the “slow” method the default settings were used. The following settings were implemented for the “fast” method: `boot.thresh.conflict = 95` (ie, support value threshold for topological conflict); `fwd.depth = 1` (ie, maximal depth at which Prunier looks forward to find a significant HGT when the current HGT is not significant). The “`boot.thresh.conflict`” number is the minimum LR-ELW edge support value for a given node in the gene tree used for recognizing topological conflict between the gene and species trees. The “`fwd.depth`” (for further details see <http://pbil.univ-lyon1.fr/software/prunier/>) is the maximal depth at which the “fast” method Prunier looks forward to find a significant HGT when the current event is not significant. A depth value of 1 implies that if no significantly supported conflict can be removed with one HGT event

the algorithm looks one step “forward” to see if the next HGT in the list will remove a significant conflict. If Prunier finds a better solution with a depth value of 2, it provides this solution in the output.

A newly developed unrestricted random effects branch-site model¹⁹ was implemented for the purpose of detecting episodic diversifying selection on codons in the paralogs of the two *Fragilariopsis* species. This method, which is available in a free public web implementation,³⁴ has been shown to have low error rates and to be powerful in identifying episodic adaptive evolution restricted to a few amino acid sites in a gene tree.¹⁹ Without prior assumptions of which lineages have gone through episodic adaptive evolution, this approach can, through the use of likelihood ratio tests, identify all the branches in a tree which have a certain proportion of sites evolving with dN/dS values significantly higher than 1.¹⁹ The gene tree required for this branch-site analysis was inferred through a partitioned maximum likelihood analysis in Treefinder.²⁹ The following optimal models, elucidated through a testing procedure in Treefinder, were used for the 1st, 2nd and 3rd codon positions, respectively: J1[optimum,Optimum]:G[Optimum]:5; GTR[optimum,Optimum]:G[Optimum]:5; TVM[optimum,Optimum]: G[Optimum]:5.

Results

Species and gene trees

The initial species phylogeny was generated based on the NCBI taxonomic classification information. This tree showed unresolved relationships for *Candidatus Methanoregula boonei*, *Colwellia* sp., *Psychromonas ingrahamii*, *Stigmatella aurantiaca* and *Rhodoferrax ferrireducens*. Since the methods in Prunier require a fully resolved species tree, a Bayesian phylogenetic analysis of a sample of SSU rRNA sequences (Table 1) was conducted. In the absence of other data, SSU rRNA sequences can serve as a first approximation to a species phylogeny because these genes seem to undergo lateral transfers at relatively low frequencies. The resulting consensus tree resolved the branching orders of *Candidatus Methanoregula boonei*, *Colwellia* sp., *Psychromonas ingrahamii*, *Stigmatella aurantiaca* and *Rhodoferrax ferrireducens* (Fig. 1) and did not show any conflicts with the NCBI taxonomic classification information. The resolved species



tree, with posterior clade probabilities for the taxa included in the Bayesian phylogenetic analysis are shown in Figure 1. The only poorly supported relationship is the one between *Cytophaga hutchinsonii* and *Marivirga tractuosa*. The species phylogeny was rooted on the branch between Eubacteria and Archaea/Eukaryota (Fig. 1) and the phylogenetic relationships are as expected, that is the Bacillariophyta (ie, the diatoms) form a well supported clade among the eukaryotes and so do the Ascomycota and Basidiomycota clades (Fig. 1). Among Eubacteria two major groups were inferred, namely Bacteroidetes and Proteobacteria (Fig. 1).

Figure 2 shows the optimal unrooted maximum likelihood tree derived from the aligned amino acid sequences of the AFPs/AFLPs. A number of major conflicts are apparent between the species and gene trees (compare Fig. 1 with Fig. 2). The prokaryotes *Rhodospirillum rubrum*, *Methanoregula boonei*, *Stigmatella aurantica* are found within a eukaryote group that consists of *Navicula glaciei*, *Chaetoceros neogracile*, *Stephos longipes*, *Phaeosphaeria nodorum* and *Talaromyces stipitatus* (Fig. 2). Furthermore, within this eukaryote clade, the proteobacterium *Stigmatella aurantica* form a sister relationship with the ascomycete *Talaromyces stipitatus* and the copepod *Stephos longipes* share a common node with the diatom *Chaetoceros neogracile*. Rather than being grouped together with the other eukaryotes, the *Fragilariopsis* and basidiomycete lineages are found among a number of proteobacteria and bacteroidetes species (Fig. 2).

Horizontal gene transfer

The outcome of the Prunier analyses, based on the “slow” and the “fast” methods, were the same, namely the same HGT events were inferred (see Fig. 1). From here on the results of the “fast” method is described and discussed. The arrows in Figure 1 show the direction of HGTs. For these events to be inferred the topological conflicts between the gene and species phylogenies had to have a minimum LR-ELW edge support values of 95% in the former tree (Fig. 2). Thus, two transfers occurred among eukaryotes, that is from *Chaetoceros neogracile* to *Stephos longipes* and from the ancestor of the basidiomycete group, composed of *Flammulina populicola*, *Lentinula edodes*, *Typhula ishikariensis*, *Leucosporidium sp.*,

to the ancestral lineage of the *Fragilariopsis* clade. In the other two transmissions, the genes moved from the ascomycete lineage (ie, *Phaeosphaeria nodorum* and *Talaromyces stipitatus*) to a proteobacterium (ie, *Stigmatella aurantica*) and from *Polaribacter irgensii* to the ancestral lineage of a group, consisting of *Shewanella denitrificans*, *Shewanella frigidamarina*, *Colwellia sp.*, *Psychromonas ingrahamii*. The result of the “fast” HGT analysis was not affected by the position of the root, that is whether it was assumed to be positioned between the Eubacteria and the Archaea/eukaryote clade or between the Eubacteria/Archaea clade and the eukaryotes. Lowering the minimum support value for recognizing topological conflict between the gene and species trees to 90% (ie, the default setting in Prunier for the “fast” method) did not affect the results.

Episodic diversifying selection analysis

Episodic evolution took place in 32% (15/47) of the branches in the *Fragilariopsis* gene tree (Fig. 3 and Table 2). The strength of diversifying selection (ω^+) and the length of the branch segments (q^+) affected by this selection regime showed substantial variation across the topology (Fig. 3 and Table 2). The former parameter (ω^+) was in most cases much larger than 8.0 whereas the latter parameter (q^+) ranged from 0.004 to 0.16. Thus, in the majority of the cases, “strong” positive selection acted on very “short” segments of the branches (Fig. 3 and Table 2).

Discussion

Horizontal gene transfer

Conflicts between species and gene trees can be due to lateral gene transfers between distantly related species.¹⁸ The program Prunier (see Material and Methods) has been demonstrated to perform well when it comes to identifying incongruences between species and gene trees caused by horizontal gene transfers.¹⁸ This method allowed me to further investigate hypotheses about horizontal transfers of AFP and AFLP genes between distantly related taxa.

The Prunier analyses of the amino acid alignment rejected the idea that the prokaryote genera *Shewanella* or *Cytophaga* have been instrumental in transferring genes coding for AFPs or AFLPs to the diatoms⁴—at least this appears to be the case with regard to the taxa considered here. The only

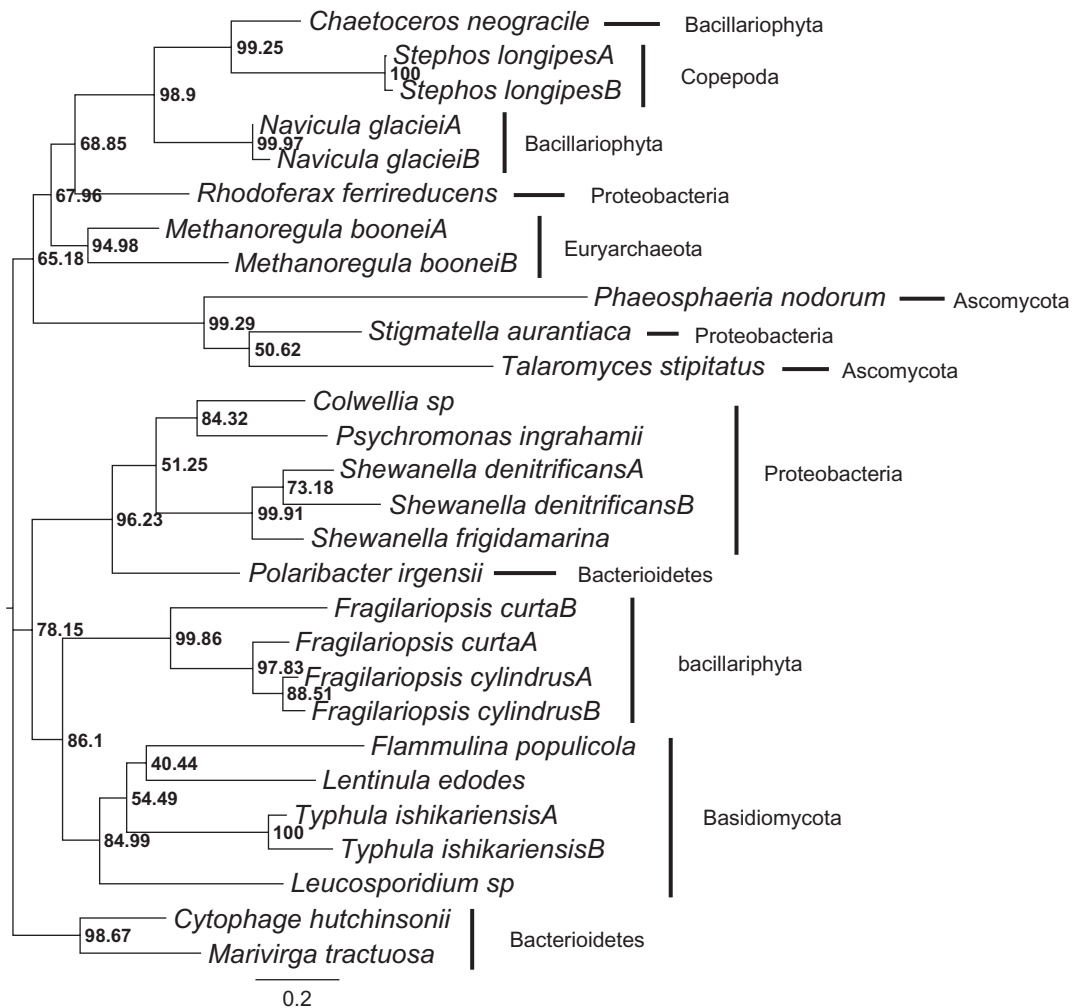


Figure 2. Unrooted maximum likelihood tree inferred by the Treefinder based on the alignment of the amino acid sequences. **Notes:** The values next to the nodes are LR-ELW edge supports. The scale bar shows the number of substitutions per nucleotide.

diatom lineages that have “picked up” the AFP or AFLP gene from another taxon are *Fragilariopsis curta* and *Fragilariopsis cylindrus* (Fig. 1). The most likely donor are the basidiomycetes (Fig. 1) which also is consistent with the hypothesis proposed by.⁴ The AFP genes of *Chaetoceros neogracile* and *Navicula glaciei* do not appear to have been acquired from other lineages. Instead the AFPs in these taxa may have evolved from ancestral genes with different functions.^{4,11} This scenario suggests that the AFPs found in diatoms could have a complex origin¹¹ which may or may not involve HGT events. To further address this issue a wider taxonomic sampling of the AFP and AFLP genes needs to be considered. The AFP gene in *Stephos longipes*, a sea-ice copepod, has been transferred to it from *Chaetoceros neogracile*,⁹ also see¹¹ (Fig. 1). This is not unexpected since *C.*

neogracile occur in the same habitat as *S. longipes* and copepods are known to “graze” on diatoms.

Few studies, that provide evidence of eukaryotic-to-prokaryotic HGTs, have been reported.¹³ However, the acquisition of an AFLP gene by the proteobacterium *Stigmatella aurantiaca* from the ascomycetes *Talaromyces stipitatus* and *Phaeosphaeria nodorum* constitutes an example of such a rare event (Fig. 1). Gene transfers from eukaryotes to prokaryotes are disfavored due to the presence of introns in the former and because of conjugation/transduction processes in the latter.¹³ Other reasons that have been mentioned are the eukaryote genes being of less adaptive significance to prokaryotes and possibly the lower frequency of interaction between representatives of the eukaryotic and prokaryotic domains relative to prokaryote-prokaryote contacts.¹³

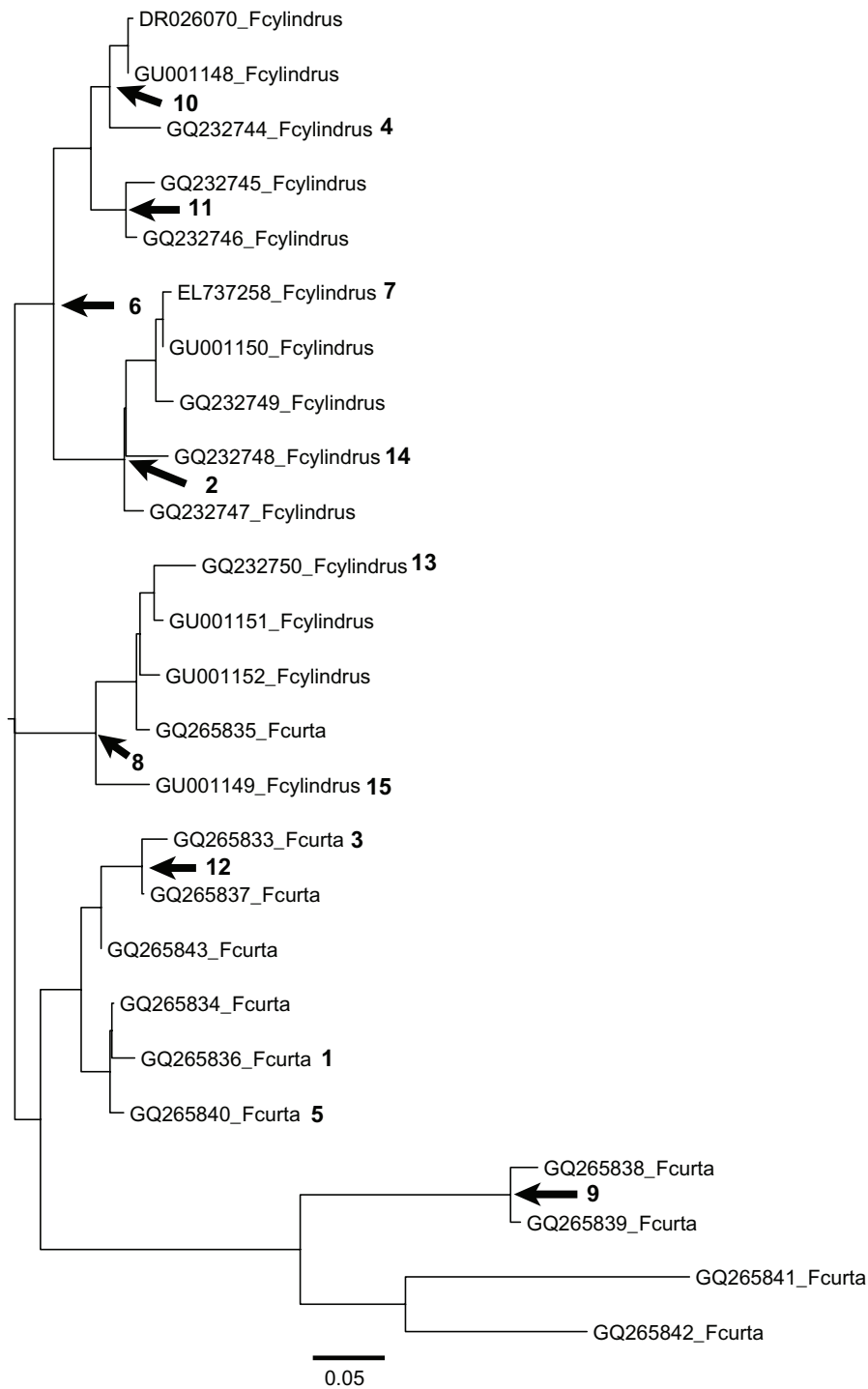


Figure 3. Relationships between the duplicated antifreeze protein genes in *Fragilariopsis curta* and *Fragilariopsis cylindrus*. **Notes:** The numbers show the branches that were found to be under episodic diversifying selection. See Table 2 for the strength of positive selection (ω^+) on each of the affected branches and for the proportion of the total branch length (q^+) influenced by this selective regime. The scale bar shows the expected number of substitutions per nucleotide.

One prokaryote-to-prokaryote transfer took place, that is between *Polaribacter irgensii* and a lineage composed of *Shewanella denitrificans*, *Shewanella frigidamarina*, *Colwellia sp.*, *Psychromonas ingrahamii* (Fig. 1). Perhaps what is the most surprising

result in this study is that not a single prokaryote-to-eukaryote transfer was detected. These are events that are thought to have been common throughout the history of life.¹³ This unexpected outcome suggests that the group of AFPs and AFLPs examined

Table 2. Branches found to be under episodic diversifying selection by the unrestricted random effects branch-site model.

^a Branch	^b Mean ω	^c ω^-	^d q^-	^e ω^N	^f q^N	^g ω^+	^h q^+	ⁱ LRT	^j p	^k Corrected p
1	1.66	0.00	0.44	0.00	0.53	1087.74	0.03	30.87	<0.0001	<0.0001
2	0.63	0.00	0.96	0.86	0.00	219.42	0.04	29.63	<0.0001	<0.0001
3	0.95	0.00	0.97	0.00	0.00	432.24	0.03	28.02	<0.0001	<0.0001
4	0.67	0.64	0.83	0.64	0.15	10000	0.02	26.82	<0.0001	<0.0001
5	0.91	0.00	0.99	0.96	0.00	1363.14	0.01	25.34	<0.0001	<0.0001
6	0.68	0.39	0.96	1.00	0.00	10000	0.03	18.35	<0.0001	0.00
7	0.53	0.32	0.98	0.34	0.02	10000	0.00	18.11	<0.0001	0.00
8	1.42	1.00	0.95	0.00	0.01	228.10	0.04	17.60	<0.0001	0.00
9	1.46	0.43	0.73	0.51	0.11	33.76	0.16	16.73	<0.0001	0.00
10	1.33	1.00	0.97	0.98	0.00	2969.33	0.03	15.30	<0.0001	0.00
11	0.47	0.20	0.94	0.19	0.03	599.70	0.03	13.49	0.00	0.00
12	0.31	0.00	0.79	1.00	0.20	213.29	0.01	13.10	0.00	0.01
13	0.19	0.08	0.98	0.08	0.01	10000	0.01	12.58	0.00	0.01
14	0.68	0.67	0.98	0.67	0.01	10000	0.01	12.48	0.00	0.01
15	0.69	0.05	0.83	0.05	0.07	8.69	0.10	9.66	0.00	0.03

Notes: ^aBranches (see Fig. 3 for branch numbers) that were inferred to be under episodic diversifying selection; ^bmean ω is the average dN/dS estimated for each branch under the free-ratio MG94 \times REV model (no site-to-site rate variation); ^c ω^- and ^d q^- values reflect the strength of negative selection (ω^-) and the proportion of the total branch length affected by negative selection (q^-); ^e ω^N and ^f q^N -values reflect (nearly) neutral evolution (ω^N) and the proportion of the total branch length affected by (nearly) neutral evolution (q^N); ^g ω^+ and ^h q^+ -values reflect the strength of positive selection (ω^+) and the proportion of the total branch length affected by positive selection (q^+); ⁱLRT is the likelihood ratio test statistic; ^jp is the uncorrected *P*-value; ^kcorrected p is the probability based on Holm's multiple testing correction.

here may have evolved independently in some prokaryotes and eukaryotes. Since the Prunier analysis did not identify the “clustering” of *Methanoregula boonei* and *Rhodiferax ferrereducens* with *N. glaciei*, *C. neogracile*, *S. longipes* in the gene tree as being the result of HGTs (Fig. 2), it is quite possible that this close association between distantly related taxa can be explained by convergent evolution. Due to the strong selection pressures for the evolution of proteins with antifreeze properties in ice-living organisms, it is conceivable that some of the cryoprotectants have originated independently in different polar taxa.

Not all the proteins included here have been shown to have antifreeze properties (see Material and Methods). In database searches a number of genes show a high degree of similarity to those with known antifreeze functions—some of those sequences were included in the current investigation (also see discussion in¹¹). Even though information about their distribution is scanty, many of the organisms that have AFLP genes are not known to occur in polar habitats nor in ice, suggesting that some of these proteins do not have a function related to cold tolerance. This is probably the case with *Talaromyces stipitatus*, *Phaeosphaeria nodorum* and *Stigmatella aurantiaca*. The former

two lineages appear to have donated a gene coding for an AFLP to the *S. aurantiaca* (Fig. 1) but none of these species are known occur in the polar regions. The antifreeze activity of the proteins, found in life forms that are known to exist in sub-zero environments, may have evolved from genes with a different function.¹¹ Bayer-Giraldi and colleagues (2010) noted that many of the AFPs show a high degree of similarity to adhesins which are surface proteins with functions that could easily be modified, through molecular evolution, to have antifreeze properties (also see discussion in¹¹). Natural selection is expected to readily favor transfers of AFP or AFLP genes to organism living in sub-zero environments if they increase the chance of becoming better adapted to these extreme conditions over time (ie, evolution of AFLP genes to AFP genes)¹³ for a general discussion on HGT and adaptations.

Episodic diversifying selection analysis

The results of the Prunier analysis suggested that the *Fragilariopsis* lineage acquired the AFP gene from the basidiomycetes after which the transferred gene(s) have gone through a number of duplication events (Fig. 3). The interesting question here is whether the duplicated genes in *Fragilariopsis* changed



functionally over time. Functional modification of a gene is expected to be associated with shifts in the selective regime acting on it.^{35,36} In fact, it has been demonstrated that positive selection can play a role in the early functional evolution of duplicate genes.³⁷ If positive selection is instrumental in driving functional changes one should observe rapid amino acid replacements in the encoded proteins, that is rates much higher than expected under neutral evolution and negative selection.^{35,36} The current analysis showed that “strong” positive selection acted on “short” segments of the branches (Fig. 3 and Table 2). This scenario provides support for the *Fragilariopsis* paralogs having differentiated functionally over short time periods. Duplication of AFP genes in association with functional differentiation is thought to be an adaptive strategy for coping with cold stress^{10,37,38} because it can expand the range of functions in AFPs.¹⁰

Acknowledgments

I thank James Raymond for introducing me to questions associated with horizontal transfer of antifreeze protein genes in diatoms and Vincent Daubin for advice on using Prunier. Three anonymous reviewers provided comments that improved the content of the paper.

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contributorship, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

- Duman JG. Antifreeze and ice nucleator proteins in terrestrial arthropods. *Ann Rev Physiol.* 2001;63:327–57.
- Hoshino T, Kiriaki M, Ohgiya S, et al. Antifreeze proteins from snow mold fungi. *Can J Bot.* 2003;81:1175–81.
- Griffith M, Yaish MWF. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Sci.* 2004;9:399–405.
- Janech MG, Krell A, Mock T, Kang JS, Raymond JA. Ice-binding proteins from sea ice diatoms (Bacillariophyceae). *J Phycol.* 2006;42:410–6.
- Liu Y, Li Z, Lin Q, et al. Structure and evolutionary origin of Ca²⁺-dependent herring type II antifreeze protein. *PLoS ONE.* 2007;2(6):e548.
- Raymond JA, Fritsen CH, Shen K. An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol Ecol.* 2007;61:214–21.
- Raymond JA, Janech MG, Fritsen CH. Novel ice-binding proteins from a psychrophilic antarctic alga (Chlamydomonadaceae, Chlorophyceae). *J Phycol.* 2009;45:130–6.
- Raymond JA, Janech MG. Ice-binding proteins from enoki and shiitake mushrooms. *Cryobiology.* 2009;58:151–6.
- Kiko R. Acquisition of freeze protection in a sea-ice crustacean through horizontal gene transfer? *Polar Biol.* 2010;33:543–56.
- Kelley JL, Aagaard JE, MacCoss MJ, Swanson WJ. Functional diversification and evolution of antifreeze proteins in the antarctic fish. *Lycodichthys dearborni.* *J Mol Evol.* 2010;71:111–8.
- Bayer-Giraldi M, Uhlig C, John U, Mock T, Valentin K. Antifreeze proteins in polar sea ice diatoms: diversity and gene expression in the genus. *Fragilariopsis.* *Environ Microbiol.* 2010;12:1041–52.
- 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.
- Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 2008;9:605–18.
- Armbrust EV. The life of diatoms in the world's oceans. *Nature.* 2009;459:185–92.
- Armbrust EV, Berges JA, Bowler C, et al. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science.* 2004;306:79–86.
- Bowler C, Allen AE, Badger JH, et al. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature.* 2008;456:239–44.
- Bowler C, Vardi A, Allen EA. Oceanographic and biogeochemical insights from diatom genomics. *Annu Rev Mar Sci.* 2010;2:355–65.
- Abby SS, Tannier E, Gouy M, Daubin V. Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. *BMC Bioinformatics.* 2010;11:324.
- Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delpont W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 2011;28:3033–43.
- Gwak IG, Jung WS, Kim HJ, Kang SH, Jin E. Antifreeze protein in Antarctic marine diatom, *Chaetoceros neogracile.* *Mar Biotechnol.* 2010;12:630–9.
- Xia X. Data analysis in molecular biology and evolution. Boston Kluwer Academic Publishers; 2001.
- Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res.* 2002;30:3059–66.
- 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.
- Katoh K, Toh H. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics.* 2008;9:212.
- Pagel M, Meade A. A phylogenetic mixture model for detecting pattern heterogeneity in gene sequence or character-state data. *Syst Biol.* 2004;53:571–81.
- Pagel M, Meade A. Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Phil Trans R Soc B.* 2008;363:3955–64.
- Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (Are We There Yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics.* 2008;24:581–3.
- Wilgenbusch JC, Warren DL, Swofford DL. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty>. 2004.
- Jobb G, von Haeseler A, Strimmer K. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology.* 2004;4:18.



30. 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.
31. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 1999;16:1114–6.
32. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc Roy Soc B.* 2002;269:137–42.
33. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 2002;51:492–508.
34. Delpont W, Poon AF, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics.* 2010;26:2455–7.
35. Arnau V, Gallach M, Lucas JI, Marín I. UVPAR: fast detection of functional shifts in duplicate genes. *BMC Bioinformatics.* 2006;7:174.
36. Beisswanger S, Stephan W. Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. *Proc Natl Acad Sci U S A.* 2008;105:5447–52.
37. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol.* 2003;18:292–8.
38. Carginale V, Trinchella F, Capasso C, Scudiero R, Parisi E. Gene amplification and cold adaptation of pepsin in Antarctic fish. A possible strategy for food digestion at low temperature. *Gene.* 2004;336:195–205.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>