Nucleotide sequence alignment of the C-terminal domain coding exon (exon6) of Ld sialic acid synthase (SAS)-A and LdSAS-B and exon2 of three typical antifreeze protein (AFPIII) genes, LdAfPIII, LdAfPIII-1, and LdAfPIII-2, from the Lygodichthys dearborni AFPIII locus. Among AFPIII genes, LdAfPIII shares the greatest sequence similarity with the LdSAS genes, including having six additional nucleotides (shown in red) not present in the other two AFPIII genes, suggesting that LdAfPIII is the earliest member of the AFPIII gene family. Asterisks indicate identical nucleotides between the six sequences, and the asterisks in red indicate identical nucleotides shared between LdAfPIII and the LdSAS genes. The stop codon (TGA for SAS and TAG for AFPIII) is underlined.

**Fig. S1.** Nucleotide sequence alignment of the homologous regions between LdSAS-A, LdSAS-B, and LdAfPIII, the putative primitive AFPIII gene members from L. dearborni. (A) Nucleotide sequence alignment and deduced amino acid sequence of LdSAS-A and LdSAS-B. The proximal 5’ flanking sequences (~600 nt) of the two genes share no sequence similarity and thus, are not homologous. The locations of the six exons are labeled, and exon sequences and translated amino acids are bold. Amino acids and codons of LdSAS-B that differ from LdSAS-A are shown in blue. (B) Nucleotide sequence alignment of the homologous regions between LdSAS-B and LdAfPIII. The proximal 5’ flanking regions of the two genes are homologous. The color highlighted LdSAS-B sequence regions (in the ancestral LdSAS-B) gave rise to the AFPIII gene. The signal peptide of AFPIII (exon1) was derived from partial 5’ flanking sequences and the first six codons of exon1 of LdSAS-B (yellow), AFPIII intron 1 from LdSAS-B intron 5 (turquoise), and the mature AFPIII (exon2) from LdSAS-B C-terminal domain (exon6; yellow). (C) Nucleotide sequence alignment of the homologous regions between LdSAS-A and LdAfPIII. LdSAS-A and LdAfPIII share lower nucleotide identities (66% id) than LdSAS-B and LdAfPIII (70% identity) (Fig. 2). The proximal 5’ flanking regions of the two genes are not homologous, and without this homologous region, LdSAS-A could not have given rise to LdAfPIII. In all sequences, the stop codon is in red, nucleotides not shown are indicated by number of nucleotides, asterisks indicate identical nucleotides, and dashes indicate gaps. Sequences have been deposited in GenBank under accession numbers GQ368894 (LdAfPIII) and GQ368892 (LdSAS-A and LdSAS-B).

**Fig. S2.** Sequence alignments of LdSAS-A, LdSAS-B, and LdAfPIII, the putative primitive AFPIII gene members from L. dearborni. (A) Nucleotide sequence alignment and deduced amino acid sequence of LdSAS-A and LdSAS-B. The proximal 5’ flanking sequences (~600 nt) of the two genes share no sequence similarity and thus, are not homologous. The locations of the six exons are labeled, and exon sequences and translated amino acids are bold. Amino acids and codons of LdSAS-B that differ from LdSAS-A are shown in blue. (B) Nucleotide sequence alignment of the homologous regions between LdSAS-B and LdAfPIII. The proximal 5’ flanking regions of the two genes are homologous. The color highlighted LdSAS-B sequence regions (in the ancestral LdSAS-B) gave rise to the AFPIII gene. The signal peptide of AFPIII (exon1) was derived from partial 5’ flanking sequences and the first six codons of exon1 of LdSAS-B (yellow), AFPIII intron 1 from LdSAS-B intron 5 (turquoise), and the mature AFPIII (exon2) from LdSAS-B C-terminal domain (exon6; yellow). (C) Nucleotide sequence alignment of the homologous regions between LdSAS-A and LdAfPIII. LdSAS-A and LdAfPIII share lower nucleotide identities (66% id) than LdSAS-B and LdAfPIII (70% identity) (Fig. 2). The proximal 5’ flanking regions of the two genes are not homologous, and without this homologous region, LdSAS-A could not have given rise to LdAfPIII. In all sequences, the stop codon is in red, nucleotides not shown are indicated by number of nucleotides, asterisks indicate identical nucleotides, and dashes indicate gaps. Sequences have been deposited in GenBank under accession numbers GQ368894 (LdAfPIII) and GQ368892 (LdSAS-A and LdSAS-B).

**Fig. S3.** The synthetic E6 domain of LdSAS-A as well as the full-length LdSAS-A showed similar ice-binding activities as those of LdSAS-B. Single ice-crystal growth morphology in (A) water and solutions of (B) LdSAS-A E6 (12 mg/mL), (C) LdSAS-B E6 (12 mg/mL), and (D) AFPIII (0.12 mg/mL) at their respective equilibrium freezing points. Single ice-crystal growth morphology in (E) water and solutions of (F) LdSAS-A E6 (12 mg/mL) and (G) LdSAS-B (12 mg/mL) at temperatures below the freezing points. (H) In contrast, in native AFPIII (0.12 mg/mL), the seed ice crystal did not grow until the temperature was lowered to −0.104 °C. (I) Initial seed ice crystal at the equilibrium freezing point (fp) of the recombinantly expressed full-length LdSAS-A solution. Ice growth is inhibited during 20 min of holding the ice crystal at the equilibrium fp (I) or small cooling (−0.004 °C to 0.074 °C) below the equilibrium fp (K). When the nonequilibrium fp is reached, (L) ice expands rapidly as a hexagonal disk. The detectable ice-inhibition activity of LdSAS-A is slightly greater than that of LdSAS-B.

**Fig. S4.** The expression and sialic acid synthetic activity of LdSAS-A, LdSAS-B, and LdSAS-B mutant (LdSAS-Bm4). (A) SDS/PAGE of the expressed enzymes after purification and dialysis. (B) Sialic acid synthetic activity of LdSAS-A, LdSAS-B, and LdSAS-Bm4. The LdSAS-Bm4 is a mutant form of the LdSAS-B gene, of which four amino acids (V299, G304, V306, and T334) in the C-terminal domain were substituted with their homologs (Q9, N14, A16, and Q44) in the AFPIII ice-binding surface, respectively. Together with the T305, T308 preexisted in the LdSAS-B, the mutant mimic of the evolutionary conversion of the SAS-B C-terminal to AFPIII in the active site. One unit (U) of sialic acid synthetic activity was defined as the amount of the enzyme that synthesizes 1 µmol sialic acid per 1 min. Sialic acid synthetic activity was assayd using a classical method (1). Briefly, 160 µL substrate solution (containing required amounts of ManNAc-6-P, 8.3 mM phosphoenolpyruvate (PEP), and 12.5 mM MgCl₂ in 50 mM bicine, pH 7.5) were added to 90 µL enzyme solution (1 mg/mL). The reactions were carried out at 37 °C for 1 h and stopped by boiling for 3 min. After centrifugation for 3 min at 13,000 × g, the samples were analyzed using the thiobarbituric acid (TBA) assay (2). The enzyme activity was calculated using a molar absorption coefficient for sialic acid of 57,000 M⁻¹·cm⁻¹ at 549 nm. Sialic acid synthetic activity was not detected for the mutant LdSAS-Bm4 protein, even when enzyme concentration was increased to 3 mg/mL under the same reaction conditions.

The phylogenetic trees used in the nonsynonymous substitution per nonsynonymous site (Ka)/synonymous substitution per synonymous site (Ks) ratio tests. (A) The phylogenetic relationship of SAS genes from two AFPIII-bearing Antarctic eelpouts, L. dearborni and Pachycara brachycephalum, and other five teleost species inferred by neighbor-joining, maximum likelihood, and Bayesian algorithms. The nonsynonymous substitution per nonsynonymous site (dN)/
synonymous substitution per synonymous site (dS) (ω) value was calculated using the branch model of codeml in the Phylogenetic Analysis by Maximum Likelihood (PAML) program for each branch, and the values of the SAS-B branch of L. dearborni and P. brachycephalum and the two SAS genes of Gasterosteus aculeatus were shown. The ω values of all other branches, like those of the G. aculeatus branch, are smaller than 1 and therefore, are not shown. The eelpout SAS-B branch was then assigned as the foreground branch in the branch-site model and tested for selection against other branches of the tree as the background (Table 1 has detailed results). (B) The phylogenetic tree used to test positive selection in the AFPIII branch. The tree was constructed using the three homologous regions between the AFPIII and SAS genes, including the signal peptide, the intron, and the C-terminal/AFP regions. Sequences from each species were aligned with Clustalw, and evolutionary relationship was inferred by the same three algorithms as in A. A ω value greatly larger than 1 was detected in the branch leading to the AFPIII genes, whereas all other branches showed ω values less than 1. The AFPIII branch was then tested for selection using the branch-site model against the background of other branches (results shown in Table 1). Ld, L. dearborni; Pb, P. brachycephalum; Ga, G. aculeatus; Dr, Danio rerio; Ss, Salmo salar; Tr, Tetraodon nigroviridis; Ts, Takifugu rubripes; Ma, Macrozoarces americanus. The teleost SAS sequences were retrieved from GenBank or Ensembl databases. The GenBank accession numbers are LdSAS-A and LdSAS-B, GQ368899; DrSAS, BC009953; GaSAS-1, BT028736; GaSAS-2, groupIX, 133638390:13682490; ssSAS, BT045230; TsSAS, AJ705104; LdAFPIII, U20439; MaAFPIII, J03924; AiAFPIII, M22125; and LdyAFPIII, GQ368894. The Ensembl Gene IDs are TnSAS, ENSTNIT00000021471; and OsSAS, ENSORLG000000020815.

Amino acid alignment of the SAS C-terminal (exon6) domain and mature AFPIII (exon2). The position numbers of amino acids for AFPIII (first line) and SAS (second and third lines) are based on the length of the mature proteins. Underlined residues are detected to experience positive Darwinian selection with posterior probability greater than 95% (Table 1). Bold residues constitute a putative flat ice-binding surface (1). Residues in red are identical between AFPIII and LdSAS-A but different from LdSAS-B. Asterisks indicate identical residues in AFPIII and LdSAS-A and -B.

The monomer and homodimer structures of sialic acid synthase reported by Gunawan et al. (1). (A) View of the SAS monomer showing the triose-phosphate isomerase (TIM) barrel (N-terminal) and the antifreeze-like (C-terminal) domain. (B) Arrangement of the SAS holoenzyme with two monomers (one in yellow and the other in purple) in a domain-swapped manner. Each subunit has the surface area of about 12,900 Å², with the interface area of about 3,880 Å². The majority of the residues within the active site come from loop regions and S2 and S4 of the C-terminal end of the TIM barrel, with additional contacts made by the helix and coiled region of the linker domain and an extended 11-residue loop of the antifreeze-like domain (residues 285–314) (1). The structure was viewed with the PyMOL program (2) based on deposit (Protein Data Bank (PDB) ID = 1XUZ).
Fig. S8. Expression analysis showing that the precursor sequence for the AFPIII signal peptide found in the SAS-B gene possesses the same potential as the native AFPIII signal peptide in secreting AFPIII. (A) The expression construct containing \textit{L. dearborni} AFPIII second exon without any signal peptide sequence but with an initiation codon [non-signal peptide (SP) construct]. (B) The expression construct containing \textit{L. dearborni} AFPIII cDNA including the native signal peptide and the AFPIII second exon (pre-AFPIII construct). (C) The expression construct containing the AFPIII signal peptide precursor from the corresponding \textit{LdSAS-B} gene and the AFPIII second exon (SP precursor construct). Each construct was linked to a flag tag at the C terminal. (D) Western blot analysis to detect the flag tag of the expressed protein exported to culture media from transfected HEK293T cells. Lane 1: HEK293T transfected with the empty pCS2-flag\textsuperscript{4} vector. Lane 2: HEK293T transfected with the non-SP construct (AFPIII second exon without any leading sequence). Lane 3: HEK293T transfected with the pre-AFPIII construct (the native AFPIII signal peptide and the AFPIII second exon). Lane 4: HEK293T transfected with the SP precursor construct (the SP precursor from \textit{LdSAS-B} gene and AFPIII second exon). Confluent HEK293T cells cultured on six-well plates were transiently transfected with one of the constructs and incubated at 37 °C for 3 d. Culture media (about 1 mL) from each well were collected and centrifuged at 106 × g to remove any suspended cells; 12 μL spun culture medium from each transfection experiment were electrophoresed on a 15% SDS-PAGE gel and transblotted to a PVDF transfer membrane. Rabbit antiflag antibody (Beyotime) and HRP-labeled goat anti-rabbit IgG secondary antibody were used to detect the presence or absence of secreted AFPIII (D Top). As controls, the same amount of culture media (12 μL) and the total proteins extracted from the transfected cells were electrophoresized and transblotted under the same conditions as above and probed with rabbit anti-human β-actin antibody (Beyotime) and HRP-labeled goat anti-rabbit IgG secondary antibody. It shows that the cytoplasmic β-actin was not detectable in any culture media and excludes the presence of intracellular protein in culture media because of cell lysis (D Middle). In contrast, similar amounts of cytoplasmic β-actin were detectable in the intracellular components of the four transfected cells, indicating that the antibodies used are valid (D Bottom). All of the photos were produced by exposing the blots to X-ray films (Kodak) for the same amount of time (1 min).
**Fig. S1**

**LdSAS-A**

CTGGGTAAGTCCTTGCTGCACCGGTCAAGATCCCCAAAGGCACCGTCCTGACTCAGGACATGTTGACGGTGAAGGTGGCCGAGCCGATGGGCGTCGCGGCCGAGGACATCTTCCAAATG

**LdSAS-B**

CTGGTTAAAATCTGTGGTGGCCAAGGTCAAGATCCCCAAAGGAACCGTCCTGACTCAGGACATGTTGGCGGTGAAGGCGGCCGAGCCGATGGGTATCGCGGCCGAGGACTTGTGCAAAATG

**LdAFPIII**

---GCGTCCGTGGTGGCCAACCAGCTGATCCCCATAAATACTGCCCTGACTCTGATAATGATGAAGGCGGAGGTGGTCACCCCAATGGGCATCCCCGCCGAGGACATTCCCCGAATA

**LdAFPIII-1**

---GCGTCCGTGGTGGCCAACCAGCTGATCCCCATAAATACTGCCCTGACTCTGATAATGATGAAGGCGGAGGTGGTCACCCCAATGGGCATCCCCGCCGAGGACATTCCCCGAATA

**LdAFPIII-2**

---GCGTCCGTGGTGGCCAACCAGCTGATCCCCATAAATACTGCCCTGACTCTGATAATGATGAAGGCGGAGGTGGTCACCCCAATGGGCATCCCCGCCGAGGACATTCCCCGAATA

* ******* ******* ******* *   ** * ******* ******* *   *** **  ** * * * ** *   ** *****  ** * ******* *   *  ***

**LdSAS-A**

GTGGGAAAGACCGTGACGAAGGACGTGGAGGAGGACGGCAGCCTCTTGCCAGAGGTGGTGGACGGGTACTGCAAGAAGAGGAAGTGCTGA

**LdSAS-B**

GTGGGAAAGACCGTGACGGAGGACGTGGAGGAGGACGACAGCGTCATGCCAGAGATGGTGAAAGGGTACTGCAAGAACAAGATGCTGA

**LdAFPIII**

ATCGGAATGCAAGTGAACAGGGCAGTGCCGTTGGGCACAACCCTCATGCCAGACATGGTGAAAACGTACCGCGTGGCGAAGTAGTTCTGA

**LdAFPIII-1**

ATCGGAATGCAAGTGAACAGGGCAGTGCCGTTGGGCACAACCCTCATGCCAGACATGGTGAAAACGTACCGCGTGGCGAAGTAGTTCTGA

**LdAFPIII-2**

ATCGGAATGCAAGTGAACAGGGCAGTGCCGTTGGGCACAACCCTCATGCCAGACATGGTGAAAACGTACCGCGTGGCGAAGTAGTTCTGA

* ******* ******* ******* *   ** * ******* ******* *   *** **  ** * * * ** *   ** *****  ** * ******* *   *  ***
LdSAS-A and LdSAS-B were analyzed for sialic acid synthetic activity. The putative dimer was detected in LdSAS-Bm4, while putative dimers were also present in LdSAS-A and LdSAS-B. The monomer was consistently seen in all three samples.

<table>
<thead>
<tr>
<th></th>
<th>LdSAS-A</th>
<th>LdSAS-B</th>
<th>LdSAS-Bm4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sialic acid synthetic activity</td>
<td>1.4±0.2mU/mg</td>
<td>2.7±0.2mU/mg</td>
<td>NO activity detected</td>
</tr>
</tbody>
</table>
AFPIII   E2  ASVVANQLIPIINTAILLIMMKAEVVTPMGIAPEDIPRIIGMQVNRAVPLGTTLMPDMVKNYEK-
LdSAS-B  E6  KSVVAKVKIPGTKTVLTQDMLAVKAEPMMGIAAEDELCMVGTDMVDEEDDSSVMPEMVGYCKNNKC-
LdSAS-A  E6  KSLVATVKIPGTKTVLTQDMLTVKVAPMGVAEDIFQMVGTDTKVKEEDGSLPEVDGYCKKRKC-

Fig. S6
Fig. S7

A

TIM barrel domain

Antifreeze-like domain

B

Active site
Fig. S8

A

pCS2-flag4 vector

BamHI

non-SP

flag tag

(XYKDDDDKTS)\(^4\)

pCS2-flag4 vector

B

pCS2-flag4 vector

BamHI

SP

flag tag

(XYKDDDDKTS)\(^4\)

pCS2-flag4 vector

C

pCS2-flag4 vector

BamHI

SP precursor

flag tag

(XYKDDDDKTS)\(^4\)

pCS2-flag4 vector

D

medium

medium

transfected
293T

1 2 3 4

AFPIII with Flag-tag

β-Actin

β-Actin

medium

β-Actin