

1 **Supplementary Material:**

2

3 **Supplementary Methods:**

4 Sampling

5 Sea ice samples from the Southern Ocean were collected on cruise ANTXXIII-7 of RV
6 Polarstern in the Weddell Sea in 2006 (sampling sites ANT-B1 and ANT-B2) and on the SIPEX
7 cruise of RV Aurora Australis in the Dumont d'Urville Sea in 2007 (sampling sites ANT-A1 and
8 ANT-A2) (Supplementary Table 1). The Arctic sample was retrieved from Kongsfjord,
9 Svalbard, in spring 2009 (sampling site ARC) (Supplementary Table 1).

10 Ice samples were retrieved by collection of biomass rich ice pieces which were freshly
11 broken by passage of the ship (ice fishing, for station ANT-B1) or by drilling (Kovacs drill 9 cm
12 diameter, all other samples). The biomass rich sections of the ice cores (lower 1 - 10 cm)
13 were taken, cut into slices, crushed and washed with cold sterile brine or sea water.

14 Organisms were collected on polycarbonate filters with pore size 1.2 μm for Kongsfjord
15 (ARC) and Weddell Sea (ANT-B1, ANT-B2) samples and 0.2 μm for Dumont d'Urville (ANT-A1,
16 ANT-A2) at 4 °C under vacuum not exceeding -200 mbar. Samples were prefiltered through a
17 50 μm mesh for ANT-A1 and ANT-A2 and a 200 μm mesh for ARC samples, to remove larger
18 organisms. ANT-B1 and ANT-B2 were not prefiltered, but filters were checked visually for
19 larger organisms and those were eventually removed. Filters were either treated with
20 RNeasy[®] (Applied Biosystems) (0.5 to 1.5 mL) and stored at -80 °C (ANT-A1, ANT-A2, ARC)
21 or frozen dry and stored in liquid nitrogen (ANT-B1, ANT-B2, 454-ANT-B) until RNA
22 extraction.

23

24 RNA preparation

25 For samples stored in *RNAlater*[®], *RNAlater*[®] solution was removed by repeated
26 centrifugation for 10 min at 10000 g to 16000 g at 4 °C. RNA was isolated using TRIreagent[®]
27 (Sigma) according to the manufacturer's recommendations except the following
28 modifications. TRIreagent[®] was heated to 60 °C before addition to the cell pellet, and glass
29 beads (diameter 212 – 600 µm) were used to facilitate homogenization of the cells.
30 Isopropanol precipitation was carried out at -20 °C and for samples stored in *RNAlater*[®] 200
31 µL RNase-free water was added to improve mixing of the isopropanol and the aqueous
32 phase. DNA was digested using the RNase-Free DNase Set (Qiagen) and RNA purified either
33 by using the RNeasy kit (Qiagen) or by ammonium acetate precipitation.

34

35 Construction of cDNA libraries

36 Construction of cDNA libraries for all stations was accomplished by vertis Biotechnologie AG
37 (Munich, Germany). First strand synthesis from total RNA was conducted using an oligo(dT)-
38 linker primer and cDNA was further amplified with high fidelity polymerase. The plasmid
39 vector pBS II sk+ was used for ligation of cDNA. For cloning, the ligations were
40 electroporated into T1 Phage resistant TransforMax[™] EC100[™]-T1R (Epicentre) electro-
41 competent cells. Sanger sequencing of the complete cDNA libraries was performed by the
42 Max-Planck genome centre Berlin/Cologne (Germany) using the M13 forward primer (TGT
43 AAA ACG ACG GCC AGT).

44 Sanger reads are available under GenBank accession numbers JZ733060 to JZ761128.

45

46

47 454 raw data preparation

48 Samples for the 454 metatranscriptome of station ANT-B were treated as described in
49 Toseland *et al.* (2013) and sequencing performed with Roche 454 GS-FLX and GS-Titanium
50 techniques. Raw reads were assembled with Newbler (Roche, version 2.6) with default
51 parameters for transcriptomic sequences. Only isotigs longer than 250bp were considered
52 for analysis.

53 454 sequence data will soon be available at the NCBI Sequence Read Archive accession
54 number SRR1752079.

55

56 Quantitative analysis with BLAST:

57 All datasets (contigs and singletons of the Sanger EST libraries and isotigs larger 250bp of the
58 454 metatranscriptome) were filtered for the presence of potential IBP sequences and
59 sequences of reference genes responsible for core cellular functions (actin, fucoxanthin-
60 chlorophyll-binding proteins (fcps), protochlorophyllide reductase (por), oxygen-evolving
61 enhancer protein 1 of photosystem II (psbO) and 40S ribosomal protein S4 (RS4)) using local
62 TBLASTN (BLAST 2.2.25) (Altschul *et al.*, 1997). Datasets were used as databases whereas
63 sequences of the target genes listed in Supplementary Table 2 were used as queries, with a
64 cutoff E-value of 0.1 in the blast run.

65 The resulting sequences for IBPs and reference genes were applied in an online BLASTX
66 2.2.29+ analysis against the refseq and swissprot databases, respectively, excluding models
67 and uncultured samples. Queries that produced hits of the expected gene (e.g. IBP, RS4)
68 with E-values $\leq 10^{-2}$ were kept for further analysis. For quantification the number of reads
69 building one contig/isotig was taken into account (i.e. a contig with five reads was counted

70 as five and not one). Quantification of reads per 100,000 was based on unassembled reads
71 for the EST databases and the number of assembled reads for the 454 metatranscriptome.

72

73 Phylogenetic analysis with pplacer:

74 Phylogeny of IBP transcripts was analyzed using the phylogenetic placement program
75 pplacer v1.1alpha10 (Matsen *et al.*, 2010). For the backbone tree the Pfam-alignment of
76 domain DUF3494 (Pfam A) was downloaded from the Pfam database
77 (<http://pfam.xfam.org/>). Six sequences with unknown taxonomy (from a mine drainage
78 sample) were removed, resulting in an alignment of 175 domains with a length of 476bp. A
79 maximum likelihood tree was calculated with PhyML 20120412 (Guindon and Gascuel, 2003)
80 using default parameters (LG model for amino acid substitutions) and 1000 bootstraps. All
81 translated IBP sequences retrieved from the metatranscriptomic libraries were aligned to a
82 profile HMM calculated with HMMER 2.4 (Durbin *et al.*, 1998) and placed into the reference
83 tree with pplacer 1.1alpha10 (Matsen *et al.*, 2010). Graphical output was generated using
84 guppy and the trees were displayed and modified in Archaeopteryx (Han and Zmasek, 2009).
85 Phylogenetic assignments were chosen according to the best placement (pplacer output
86 parameter ML likelihood weight ratio: MLratio). Additionally, the posterior probability value
87 of each placement was recorded (Supplementary Figure S1).

88 The placement was conducted with the singletons and contigs (EST libraries) as well as the
89 isotigs (454 dataset) that were identified as IBPs in the BLAST analysis.

90

91

92 Principal coordinate analysis of phylogenetic diversity of IBP transcripts

93 Principal coordinate analysis (PCO) was calculated based on the phyloassigner placements
94 using Kantorovich-Rubinstein-distances (Figure 2c) as implemented by Evans and Matsen
95 (2012). PCO was performed with the R-package ade4 (Dray and Dufour, 2007). A fit of
96 station data for salinity, temperature, ice thickness, daylight-hours as well as maximum and
97 minimum filter size was performed for samples ANT-A1, ANT-A2, ANT-B1 (without
98 temperature and salinity), ANT-B2 and ARC.

99

100 De novo analysis of phylogenetic diversity of the environmental IBP transcripts

101 We calculated a Profile-Alignment with HMMER 2.4 (Durbin *et al.*, 1998) from the DUF3494
102 PfamA alignment including the environmental IBP transcripts larger 150 amino acids. The
103 resulting alignment was used to calculate a maximum likelihood tree with PhyML 20120412
104 (Guindon and Gascuel, 2003) using default parameters (LG model for amino acid
105 substitutions) and 100 bootstraps. Environmental sequences smaller than 150 amino acids
106 were placed onto this backbone tree using pplacer as described in material and methods.

107

108 **Supplementary Table 1:** List of samples used in this study including sampling date, position and physical properties of sea ice, as well as the size
 109 fractions collected by filtration. Data of stations ANT-A1 and ANT-A2 and J were adapted from Meiners *et al.* (2011) and for station ANT-B1 and
 110 ANT-B2 from Haas *et al.* (2009).
 111

Station	Date	Latitude	Longitude	Ice thickness (m)	Salinity (range or mean)	Temp (°C)	Ice type at lower section	Age of ice	Size fraction (µm)	Identifier in Suppl. Fig. 1 and 2
Antarctic										
Weddell Sea	060923	60°07.150 S	47°54.550 W	1.46 (± 0.05)	n.d.	n.d.	n.d.	1 st year ice	>1,2	awig5
ANT-B2	061008	65°06.117 S	57°23.551 W	1.51 (± 0.57)	3.71	-1.9 ²	columnar	1 st year ice	>1.2	awis5
454-ANT-B	pool of parallel filters to samples ANT-B1 and ANT-B2									
Dumont d'Urville Sea	070911	64°13.773 S	127°57.132 E	0.59 (0.52–0.70)	5.0 – 11.4	-5.7 (-9.7 to -2.3)	granular - columnar - granular	1 st year pack ice	0.2-50	awiA4
ANT-A2	071003	65°01.496 S	117°42.015 E	1.08 (1.05-1.09)	2.1 – 8.1	-4.5 (-6.8 to -2.0)	columnar	1 st year pack ice	0.2-50	awiJ4
Arctic										
Kongsfjord ARC	090504	78°57.550N	12°20.023 E	0.50	5.4 – 9.9	-2.01 (-2.1 to -1.6)	n. d.	1 st year ice (spring melt)	1.2-200	awikF1

112 ¹ due to the sampling method (ice fishing) no information on physical ice properties is available for station ANT-B1, ice thickness and age were
 113 assumed to be similar to another station from 060923 and values adapted
 114 ² temperature at the ice water interface, where the sample was taken

115 **Supplementary Table 2:**

116 Table of the genes that were used for the blast search and full name for genes.
 117 Abbreviations for the genes, the full names, lengths of the sequences in amino acids, NCBI
 118 accession numbers as well as the originating organisms are given. For IBPs only part of the
 119 DUF3494 domain was used.

	Length [aa]	NCBI Acc. No.	organism
type 1 IBPs	140-155	ABH08428	<i>Colwellia sp.</i> SLW05
type 1 ice-binding		ACL00837	<i>Stephos longipes</i>
proteins (DUF3494		ACL00838	<i>Stephos longipes</i>
IBPs)		ACL27143	<i>Flammulina populicola</i>
		ACL27145	<i>Lentinula edodes</i> (shiitake mushroom)
		ACU09498	<i>Chaetoceros neogracile</i>
		YP_003095014	<i>Flavobacteriaceae bacterium</i> 3519-10
		ACU30806	<i>Leucosporidium sp.</i> AY30
		ACX36851	<i>Fragilariopsis cylindrus</i>
		ACX36853	<i>Fragilariopsis cylindrus</i>
		AEY75833	<i>Nitzschia stellata</i>
		AEY75834	<i>Amphora sp.</i> CCMP2378
		AEY75837	<i>Attheya sp.</i> CCMP212
		AEY75838	<i>Phaeocystis antarctica</i>
		BAD02891	<i>Typhula ishikariensis</i>
		AFK64811	<i>Pyramimonas gelidicola</i>
		AGC91914	<i>Chlamydomonas raudensis</i>
		AAZ76251	<i>Navicula glaciei</i>
		YP_943880	<i>Psychromonas ingrahamii</i> 37
type 2 IBPs	353-359	ABY64758	<i>Chlamydomonas sp.</i> CCMP681
<i>Chlamydomonas sp.</i>		ABY64759	<i>Chlamydomonas sp.</i> CCMP681
CCMP 681 ice binding		ABY64761	<i>Chlamydomonas sp.</i> CCMP681
proteins		ABY64760	<i>Chlamydomonas sp.</i> CCMP681
por	420-440	XP_002294544	<i>Thalassiosira pseudonana</i> CCMP1335
protochlorophyllide		XP_002179689	<i>Phaeodactylum tricornutum</i> CCAP 1055/1
reductase		XP_005853690	<i>Nannochloropsis gaditana</i> CCMP526
		XP_005784495	<i>Emiliana huxleyi</i> CCMP1516
psbO	305-314	XP_002180309	<i>Phaeodactylum tricornutum</i> CCAP 1055/1
oxygen-evolving		XP_002291225	<i>Thalassiosira pseudonana</i> CCMP1335
enhancer protein 1		XP_005855446	<i>Nannochloropsis gaditana</i> CCMP526
of		XP_005761454	<i>Emiliana huxleyi</i> CCMP1516
photosystem II			

Supplementary Table 2 continuing:

	Length [aa]	NCBI Acc.	organism
RS4 40S ribosomal protein S4	260	XP_002288080	<i>Thalassiosira pseudonana</i> CCMP1335
		XP_002177120	<i>Phaeodactylum tricornutum</i> CCAP 1055/1
		XP_001691218	<i>Chlamydomonas reinhardtii</i>
fcps fucoxanthin chl a/c light-harvesting protein	205-400	XP_002292153	<i>Thalassiosira pseudonana</i> CCMP1335
		XP_002292353	<i>Thalassiosira pseudonana</i> CCMP1335
		XP_002289005	<i>Thalassiosira pseudonana</i> CCMP1335
		XP_002288517	<i>Thalassiosira pseudonana</i> CCMP1335
		XP_005767752	<i>Emiliana huxleyi</i> CCMP1516
		XP_005786132	<i>Emiliana huxleyi</i> CCMP1516
		XP_005778279	<i>Emiliana huxleyi</i> CCMP1516
		XP_005778485	<i>Emiliana huxleyi</i> CCMP1516
		XP_001698519	<i>Chlamydomonas reinhardtii</i>
		XP_001693987	<i>Chlamydomonas reinhardtii</i>
		XP_001700243	<i>Chlamydomonas reinhardtii</i>
		XP_001694115	<i>Chlamydomonas reinhardtii</i>
		XP_001701405	<i>Chlamydomonas reinhardtii</i>
		XP_001698542	<i>Chlamydomonas reinhardtii</i>
		XP_001701405	<i>Chlamydomonas reinhardtii</i>
		XP_001698542	<i>Chlamydomonas reinhardtii</i>
		XP_001694115	<i>Chlamydomonas reinhardtii</i>
		XP_001695467	<i>Chlamydomonas reinhardtii</i>
		XP_001695344	<i>Chlamydomonas reinhardtii</i>
		XP_001695353	<i>Chlamydomonas reinhardtii</i>
		XP_001703699	<i>Chlamydomonas reinhardtii</i>
		XP_001697526	<i>Chlamydomonas reinhardtii</i>
		XP_001695466	<i>Chlamydomonas reinhardtii</i>
		XP_001691959	<i>Chlamydomonas reinhardtii</i>
		XP_001696202	<i>Chlamydomonas reinhardtii</i>
		XP_001692548	<i>Chlamydomonas reinhardtii</i>
		XP_001699932	<i>Chlamydomonas reinhardtii</i>
XP_002958754	<i>Volvox carteri f. nagariensis</i>		
NP_173034	<i>Arabidopsis thaliana</i> (thale cress)		
actin	280-380	XP_002294917	<i>Thalassiosira pseudonana</i> CCMP1335
		AFO84294	<i>Ditylum brightwellii</i>
		ABC54738	<i>Skeletonema costatum</i>
		XP_002183424	<i>Phaeodactylum tricornutum</i> CCAP 1055/1
		ABQ45363	<i>Nitzschia closterium f. minutissima</i>
		AAO92429	<i>Phytophthora brassicae</i>

Supplementary Table 2 continuing:

	Length [aa]	NCBI Acc.	organism
INP	1200-2145	P16239	<i>Pantoea agglomerans</i>
Ice nucleation proteins		P06620	<i>Pseudomonas syringae</i> pv. <i>syringae</i>
		AAK70465	<i>Pantoea ananatis</i>
		P20469	<i>Pantoea ananatis</i>
		BAK13807	<i>Pantoea ananatis</i> AJ13355
		ACB59244	<i>Pseudomonas borealis</i>
		CBI99147	<i>Pseudomonas syringae</i>
		YP_001033817	<i>Rhodobacter sphaeroides</i> 2.4.1
		YP_003576802	<i>Rhodobacter capsulatus</i> SB 1003
		YP_419798	<i>Magnetospirillum magneticum</i> AMB-1

122

123

124 **Supplementary Table 3:**

125 Read and contig/isotig length (in basepairs) of Sanger and 454 data

<i>Sanger sequenced datasets</i>			
	number of reads	average read length	average contig length
ARC	5228	654	772
ANT-A1	5002	665	829
ANT-A2	6254	699	813
ANT-B1	5812	584	715
ANT-B2	5773	580	767
<i>454 sequenced dataset</i>			
	number of aligned reads	average read length	average isotig length
454-ANT-B¹ total	206983	179	391
454-ANT-B >250bp			495

126 ¹454-ANT-B is a sequence pool obtained from parallel filters of samples ANT-B1 and ANT-B2
 127 (Toseland *et al.*, 2013)

128

129 **Supplementary Newbler assembly statistics of sample 454-ANT-B:**

130 Input

131	Number of reads	391614	
132	Number of bases	65831723	
133	Number of reads trimmed	290013	74.1%
134	Number of bases trimmed	51875003	78.8%

135 Consensus results

136	Number of reads assembled	157097	54.2%
137	Number partial	49886	17.2%
138	Number singleton	31652	10.9%
139	Number repeat	3468	1.2%
140	Number outlier	1063	0.4%
141	Number too short	46847	16.2%

142 Isogroup Metrics

143	Number of isogroups	1725	
144	Average contig count	1.0	
145	Largest contig count	1	
146	Number with one contig	1725	
147	Average isotig count	1.0	
148	Largest isotig count	1	
149	Number with one isotig	1725	

150 Isotig Metrics

151	Number of Isotigs	1725	
152	Average contig count	1.0	
153	Largest contig count	1	
154	Number with one contig	1725	
155	Number of bases	675898	
156	Average isotig size	391	
157	N50 isotig size	451	
158	Largest isotig	2329	

159 Large Contig Metrics

160	Number of contigs	347	
161	Number of bases	267138	
162	Average contig size	769	
163	N50 contig size	772	
164	Largest contig size	2329	
165	Q40 plus bases	253142	94.76%

166 All Contig Metrics

167	Number of contigs	1725	
168	Number of bases	675898	
169	Average contig size	392	

170

171 **Caption to Supplementary Figure S1:**

172 Placements of all environmental IBPs into the backbone tree are shown on the red branches.
173 Sequence names carry a sample identifier (awiKF1: ARC, awiA4: ANT-A1, awiJ4: ANT-A2,
174 awig5: ANT-B1, awis5: ANT-B2, 454(g3/s3): 454-ANT-B). Numbers in brackets show the
175 maximum likelihood weight ratio (like weight ratio) and the posterior probability (PP) of a
176 placement on an edge from the pplacer analysis. Sequences marked in yellow are placed
177 with high robustness (PP >75%). Sequences marked in blue have a low PP support and
178 alternative placements of the respective read were found in a different group of the
179 backbone tree. Five reads are alternating between the “Microalgae and crustacean” clade
180 and a directly neighbouring *Phaeocystis antarctica* sequence. One read from the “Diatom”
181 clade has the alternative placement in a group of closely related fungal sequences. The latter
182 could be explained by horizontal gene transfer of IBP sequences from a basidiomycete to
183 *Fragilariopsis sp.* (Sorhannus, 2011).

184

185 **Caption to Supplementary Figure S2:**

186 *De novo* analysis of the phylogenetic diversity of the environmental IBP transcripts. The
187 backbone tree (PhyML 20120412, LG model for amino acid substitutions, 100 bootstraps)
188 was constructed from the DUF3494 PfamA alignment and the environmental IBP transcripts
189 longer than 150 amino acids (HMMER2.4 Profile-Alignment). Transcripts shorter than 150
190 amino acids were placed with pplacer and are shown on red branches. Environmental
191 sequences are shown in blue and carry a sample identifier in the name (awiKF1: ARC, awiA4:
192 ANT-A1, awiJ4: ANT-A2, awig5: ANT-B1, awis5: ANT-B2, 454(g3/s3): 454-ANT-B). A previously
193 unknown diversity of the environmental IBPs emerges in the “Microalgae and copepod”
194 clade which is shadowed in grey.

195 **Supplementary References:**

196 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997). Gapped
197 BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids*
198 *Res* 25:3389–3402. doi: 10.1093/nar/25.17.3389.

199 Durbin R, Eddy S, Krogh A, Mitchison G. (1998). *Biological sequence analysis: probabilistic*
200 *models of proteins and nucleic acids*. Cambridge University Press.

201 Dray S, Dufour A (2007). The ade4 package: implementing the duality diagram for ecologists.
202 *Journal of Statistical Software* 22:1–20.

203 Evans SN, Matsen FA (2012). The phylogenetic Kantorovich-Rubinstein metric for
204 environmental sequence samples. *Journal of the Royal Statistical Society: Series B (Statistical*
205 *Methodology)* 74:569–592. doi: 10.1111/j.1467-9868.2011.01018.x.

206 Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large
207 phylogenies by maximum likelihood. *Systematic Biology* 52:696–704. doi:
208 10.1080/10635150390235520.

209 Haas C, Friedrich A, Li NM Zijun, Pfaffling A, Toyota T. (2009). Regional variability of sea ice
210 properties and thickness in the northwestern Weddell Sea obtained by in-situ and satellite
211 measurements. In: Lemke P. (ed.) *The expedition of the research vessel “Polarstern” to the*
212 *Antarctic in 2006 (ANT-XXIII/7)*, vol. 586 of *Berichte zur Polar- und Meeresforschung*. Alfred-
213 *Wegener-Institut für Polar- und Meeresforschung*, Bremerhaven, p. 36–74.

214 Han M, Zmasek C. (2009). phyloXML: XML for evolutionary biology and comparative
215 genomics. *BMC Bioinformatics* 10:356. doi: 10.1186/1471-2105-10-356.

216 Matsen F, Kodner R, Armbrust EV. (2010). pplacer: linear time maximum-likelihood and
217 bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC
218 Bioinformatics 11:538. doi: 10.1186/1471-2105-11-538.

219 Meiners KM, Norman L, Granskog MA, Krell A, Heil P, Thomas DN. (2011). Physico-
220 ecobiogeochemistry of East Antarctic pack ice during the winter-spring transition. Deep Sea
221 Research Part II: Topical Studies in Oceanography 58:1172 – 1181. doi: DOI:
222 10.1016/j.dsr2.2010.10.033. Antarctic Sea Ice Research during the International Polar Year
223 2007-2009.

224 Sorhannus U. (2011). Evolution of antifreeze protein genes in the diatom genus
225 *Fragilariopsis*: Evidence for horizontal gene transfer, gene duplication and episodic
226 diversifying selection. Evolutionary Bioinformatics 7:279–289. doi: 10.4137/EBO.S8321.

227 Toseland A, Daines SJ, Clark JR, Kirkham A, Strauss J, Uhlig C, et al. (2013). The impact of
228 temperature on marine phytoplankton resource allocation and metabolism. Nature Clim
229 Change 3:979–984. doi: 10.1038/nclimate1989.



