

## ORIGINAL ARTICLE

# Metatranscriptomes reveal functional variation in diatom communities from the Antarctic Peninsula

Gareth A Pearson<sup>1</sup>, Asuncion Lago-Leston<sup>1</sup>, Fernando Cánovas<sup>1</sup>, Cymon J Cox<sup>1</sup>, Frederic Verret<sup>1,2</sup>, Sebastian Lasternas<sup>3</sup>, Carlos M Duarte<sup>3,4</sup>, Susana Agusti<sup>3,4</sup> and Ester A Serrão<sup>1</sup>

<sup>1</sup>Centre of Marine Sciences, University of the Algarve, Faro, Portugal; <sup>2</sup>Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, Crete, Greece; <sup>3</sup>Global Change Department, IMEDEA (CSIC-UIB), Instituto Mediterráneo de Estudios Avanzados, Esporles, Mallorca, Spain and <sup>4</sup>Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

**Functional genomics of diatom-dominated communities from the Antarctic Peninsula was studied using comparative metatranscriptomics. Samples obtained from diatom-rich communities in the Bransfield Strait, the western Weddell Sea and sea ice in the Bellingshausen Sea/Wilkins Ice Shelf yielded more than 500K pyrosequencing reads that were combined to produce a global metatranscriptome assembly. Multi-gene phylogenies recovered three distinct communities, and diatom-assigned contigs further indicated little read-sharing between communities, validating an assembly-based annotation and analysis approach. Although functional analysis recovered a core of abundant shared annotations that were expressed across the three diatom communities, over 40% of annotations (but accounting for <10% of sequences) were community-specific. The two pelagic communities differed in their expression of N-metabolism and acquisition genes, which was almost absent in post-bloom conditions in the Weddell Sea community, while enrichment of transporters for ammonia and urea in Bransfield Strait diatoms suggests a physiological stance towards acquisition of reduced N-sources. The depletion of carbohydrate and energy metabolism pathways in sea ice relative to pelagic communities, together with increased light energy dissipation (via LHCSR proteins), photorespiration, and NO<sub>3</sub><sup>-</sup> uptake and utilization all pointed to irradiance stress and/or inorganic carbon limitation within sea ice. Ice-binding proteins and cold-shock transcription factors were also enriched in sea ice diatoms. Surprisingly, the abundance of gene transcripts for the translational machinery tracked decreasing environmental temperature across only a 4°C range, possibly reflecting constraints on translational efficiency and protein production in cold environments.**

*The ISME Journal* (2015) 9, 2275–2289; doi:10.1038/ismej.2015.40; published online 14 April 2015

## Introduction

The Antarctic is among the most isolated and dynamic regions on Earth, because of the Antarctic circumpolar current, coupled with massive and seasonally variable sea-ice cover (Arrigo *et al.*, 1997; Arrigo and Thomas, 2004; Arrigo, 2014). The Southern Ocean plays a disproportionately large role in the regulation of Earth's climate and biogeochemical cycles (Boyd, 2002). As a repository of macronutrients, it is unmatched among the world's oceans, and although high nutrient-low chlorophyll conditions prevail in the permanently open-ocean area between the Polar Front and Seasonal Ice Zones, the Seasonal Ice Zones and coastal areas are more productive and support intense diatom-dominated

blooms (Tréguer and Jacques, 1992). Consequently, the ecology of the Southern Ocean plays a key role in the global biological carbon pump and nutrient cycling (Murphy *et al.*, 2012), mediated primarily by autotrophic phytoplankton, particularly diatoms.

Over the last half century, dramatic changes associated with rapid warming occurred in the western Antarctic Peninsula (wAP), with implications for global biogeochemical cycling (Meredith and King, 2005; Clarke *et al.*, 2007; Montes-Hugo *et al.*, 2009; Ducklow *et al.*, 2007, 2013). Regional warming rates in the wAP are among the most rapid globally, causing dramatic sea ice decline, increased glacial input and massive loss of permanent ice shelf. Primary production is intimately linked to sea ice dynamics; meltwater from receding ice edges stabilizes the water column allowing massive phytoplankton blooms. Warming effects in the wAP are regional; ice retreat in northern areas reduces primary production because of wind-induced mixing of the water column, whereas to the south, the increased extent of summer ice-free areas has

Correspondence: GA Pearson, Centre of Marine Sciences, University of the Algarve, Faro, 8005-139, Portugal.

E-mail: gpearson@ualg.pt

Received 19 April 2014; revised 29 January 2015; accepted 17 February 2015; published online 14 April 2015

improved bloom-forming conditions (Montes-Hugo *et al.*, 2009).

Diatoms contribute up to 25% of global primary carbon fixation (Field *et al.*, 1998), making them the single most important phytoplankton group. As phytoplankton stocks (Boyce *et al.*, 2010) and diatom production (Bopp *et al.*, 2005) decline globally under climate change, carbon export to deeper layers *via* diatom blooms will be affected (Korb *et al.*, 2010). Additionally, in the Southern Ocean, intense grazing on diatoms by krill efficiently transfers biomass to higher trophic levels (Smetacek *et al.*, 2004).

Owing to the global importance of diatoms, genome sequencing projects (Armbrust *et al.*, 2004; Bowler *et al.*, 2008), genome-wide expression analyses (Allen *et al.*, 2008; Mock *et al.*, 2008; Thamatrakoln *et al.*, 2012) and, most recently, environmental metatranscriptomics (Marchetti *et al.*, 2012; Toseland *et al.*, 2013), have provided the most extensive genomic baseline for any phytoplankton group. However, to date, no environmental genomic comparison between Southern Ocean communities has been reported. Metatranscriptomes are a powerful comparative and experimental tool to gain detailed insight into the metabolic and physiological status of communities (Marchetti *et al.*, 2012; Toseland *et al.*, 2013). Here, we present comparative metatranscriptomes from diverse diatom-rich communities in contrasting coastal habitats of the wAP and western Weddell Sea (WDS). Our data illustrate the flexibility of diatom transcriptomes in natural communities, revealing functional differentiation across habitats, particularly in nutrient acquisition, irradiance and low temperature responses.

## Materials and methods

### *Sampling and environmental data*

Samples were collected in the Antarctic Peninsula–Bellingshausen Sea–Weddell Sea, in the austral summer of 2009 (Figure 1). Water from the Bransfield Strait (BFS; 30 m) and WDS (6–45 m) was collected from a conductivity, temperature and depth instrument array using 12 l Niskin bottles, together with environmental data. From the Wilkins Ice Shelf (WKI), sea ice collected from floes was melted with cold seawater (0 °C; 0.2 µm—filtered). Both WKI (sunrise 06:30, sunset 22:30 h) and WDS (sunrise 05:00, sunset 23:00 h) samples were collected just after midday, whereas BFS was collected near dusk at 21:30 h (sunrise 04:00, sunset 22:00 h). Within 30–60 min after collection, samples were filtered (5 µm pore-etched polycarbonate filters) and frozen in liquid nitrogen for storage and transportation.

Phosphorus, nitrate + nitrite and silicate concentrations were analysed using standard methods (Hansen and Koroleff, 1999) in a Bran Luebbe AA3 autoanalyser (Norderstedt, Germany). Ammonium was measured spectrophotometrically, based on the reaction with ortho-phthalaldehyde and sulphite

following the procedure and recommendations described by Kerouel and Aminot (1997). Dissolved iron concentrations in surface waters were sampled and analysed using trace-metal clean techniques (cf. Tovar-Sánchez *et al.*, 2010).

For cell counts, 2–3 l of seawater was concentrated by gentle pressure for <30 min (Lasternas and Agustí, 2010). Duplicate 10 ml aliquots were filtered onto 2 µm pore size black filters, fixed with glutaraldehyde (1% final concentration), and counted under the epifluorescence microscope (Zeiss Axioplan Imaging, Carl Zeiss Iberia S.L., Madrid, Spain).

### *Library preparation and sequencing*

Filters were vortexed in buffer RTL plus (1200 µl per 47 mm filter) containing 100 µl of glass beads to disrupt the cells. Lysates were transferred onto QIAshredder columns for homogenization. The centrifuged lysate was pipetted onto an AllPrep spin column and total RNA was purified with AllPrep DNA/RNA Mini Kit (Qiagen, Izasa Portugal, LDA, Carnaxide, Portugal), including DNase digestion. RNA quantity and quality were assessed by spectrophotometry (Nanodrop; Fisher Scientific, Porto Salvo, Portugal) and electrophoresis, respectively.

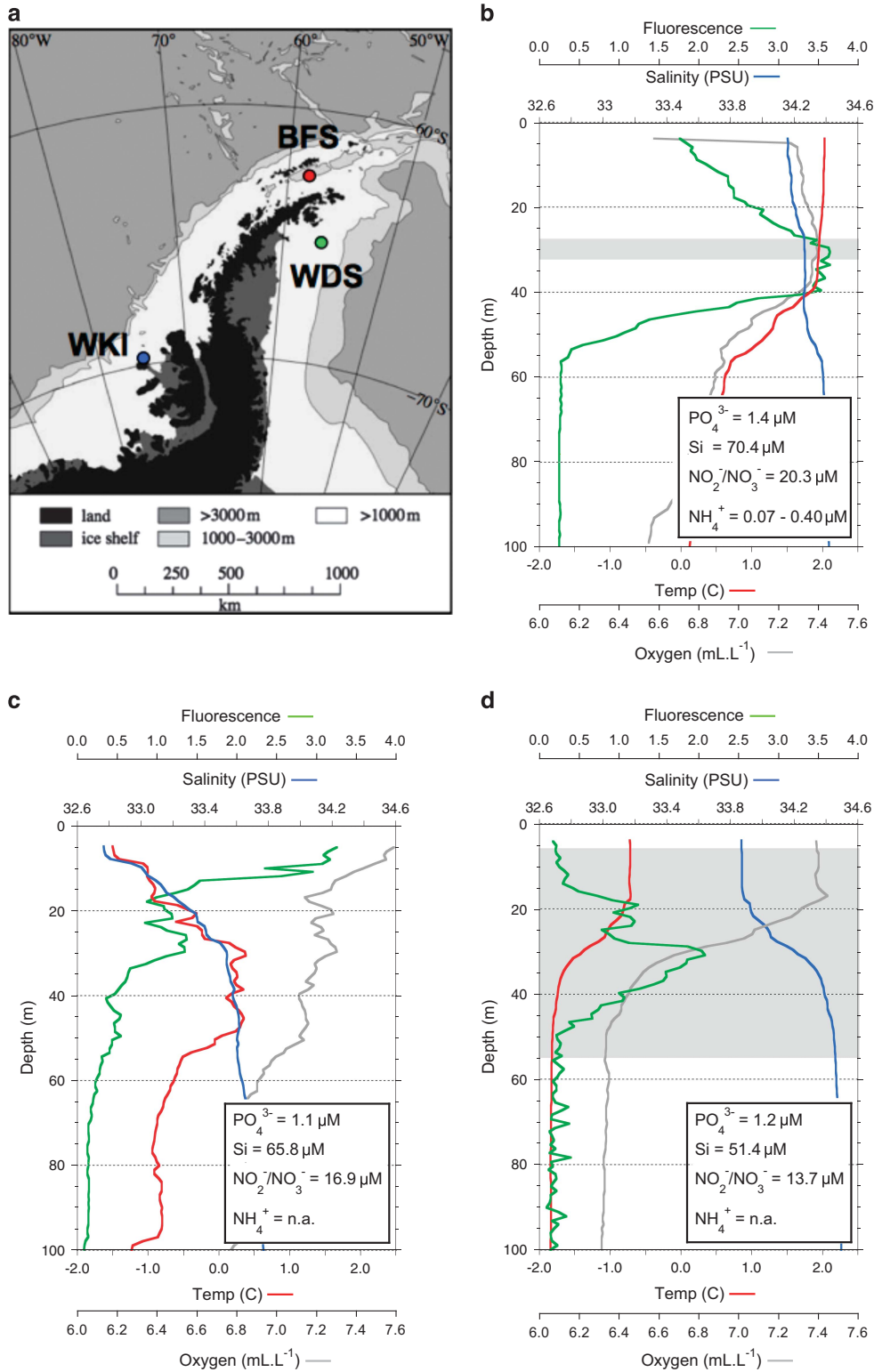
Full-length double-stranded cDNA (ds-cDNA) was synthesized from 250 ng of total RNA (SMARTer PCR cDNA Synthesis Kit; Clontech, Enzifarma S.A., Lisbon, Portugal). First strand cDNA templates were amplified by long-distance PCR (Advantage 2 PCR Kit; Clontech), optimized to avoid over-amplification (20 cycles). Replicate PCR reactions per library were pooled and purified (MiniElute PCR Purification kit; Qiagen) for sequencing.

A total of 560K high-quality GS-FLX Titanium 454 reads were sequenced at Biocant (Cantanhede, Portugal), resulting in 532K reads after cleaning and removal of SMART adaptors (TagCleaner; Schmieder *et al.*, 2010). Reads were screened for rRNA (Meta-RNA 1.0; Huang *et al.*, 2009) using HMMER 3.0 (Eddy, 1998); significant hits were removed for taxonomy binning and biodiversity analyses (Table 1).

### *Sequence assembly and annotation*

Sequences were assembled using MIRA3.0 (Chevreux *et al.*, 2004). BLASTx (E-value  $\leq 10^{-4}$ ) was performed against a custom protein database (NCBI RefSeq plus additional collections from the Joint Genome Institute (JGI: <http://genome.jgi-psf.org/>) for the diatom *Fragilariopsis cylindrus*, the pelagophyte *Aureococcus anophagefferens*, the haptophyte *Emiliania huxleyi* and the crustacean *Daphnia pulex*). Taxonomic assignment of contigs with significant BLASTx scores was performed using MEGAN (v.4.6: Huson *et al.*, 2011) with minimum bit score = 50, min support = 5 contigs, winning bit score = 200.

Predicted protein-coding sequences were determined from significant top BLASTx matches. To



**Figure 1** Sampling locations and physico-chemical conditions. **(a)** Map of the Antarctic Peninsula showing the locations for metatranscriptome sampling in the BFS (red), the WDS (green), and the WKI/Bellingshausen Sea (blue). **(b)** Depth profile for salinity, temperature, oxygen and fluorescence at BFS, shaded area at ca. 30 m indicates the depth for metatranscriptome sampling, and the major nutrient concentrations at this depth are given in the insert box. **(c)** Depth profile for WKI at the site where sea ice samples for metatranscriptome analysis were taken. Near-surface water column nutrient concentrations are shown in the insert box. **(d)** Depth profile at WDS showing broad depth range for metatranscriptome sample; average nutrient concentrations across this range are given in the insert box.

**Table 1** Sequence, assembly and BLASTx annotation information for three western Antarctic Peninsula phytoplankton community metatranscriptomes from the Bransfield Strait (BFS), Wilkins sea ice (WKI) and Weddell Sea (WDS) samples

A							
Source	Total reads		Quality-filtered reads (% of total)		rRNA (% of high quality)		
Total	559 911		532 846 (95.2)		80 152 (15.0)		
BFS	114 323		109 521 (95.8)		8143 (7.4)		
WKI	191 057		182 555 (95.6)		26 889 (14.7)		
WDS	253 631		240 770 (94.9)		45 120 (18.7)		

B							
Source	Contigs	N50	Putative mRNA reads	Reads with BLASTx <sup>a</sup> support (% total)	mRNA reads: diatom (% total)	Contigs: diatom	rRNA reads: diatom (%)
Total	58 101	398	251 415	158 074 (62.9)	89 581 (35.6)	18 104	16 702 (15.7)
BFS			64 620	48 295 (74.7)	35 035 (54.2)	5883	2171 (5.8)
WKI			79 311	49 248 (62.1)	22 240 (28.0)	6177	3535 (13.7)
WDS			107 484	60 531 (56.3)	32 306 (30.1)	7369	10 996 (25.4)

(A) Total number of reads obtained and the number (and percentage) passing quality-filtering. High-quality reads were analysed by rRNA HMMER (Huang *et al.*, 2009) to identify the number and proportion of rRNA read contamination in the metatranscriptomes.

(B) Assembly and BLASTx annotation data, showing total number of contigs assembled and median length (N50), with the number of putative mRNA reads assembled globally and per community. The total numbers of reads coming from contigs with BLASTx support ( $E \leq 10^{-4}$ ), and those with top BLASTx hits to diatoms are shown (as a percentage of the total putative mRNA reads in parentheses). The total number of diatom-assigned contigs is shown, together with the number containing (at least one) read(s) from each of the 3 communities. The final column shows the number of diatom rRNA reads (and percentage of total diatom reads) in the overall, and community-specific, metatranscriptomes.

<sup>a</sup> $E \leq 10^{-4}$

account for reading frame shifts introduced by NGS, multiple high-scoring segment pairs in top BLASTx matches were 'annealed' according to maximum and minimum overlap criteria ( $\pm 5$  nucleotides) to improve annotation success against protein databases. Functional annotation was performed against the KEGG (Ogata *et al.*, 1999), Pfam and KOG databases. Read counts, taxonomy assignment and functional annotation data were stored in a relational database (MySQL, <http://www.mysql.com>) for further analysis.

Total reads assigned to diatoms (Bacillariophyta) by MEGAN were used to define this taxonomic subset. Library size normalization between communities was then performed prior to enrichment analyses for diatoms. Significant differences in normalized read counts for annotation terms or pathways were identified using the *R* statistic (Stekel, 2000), with an empirically determined cut-off of  $R \geq 8$  ( $\approx 98\%$  believability).

#### Community structure, diversity and taxonomy

Large and small subunit rRNA sequence reads were compared by BLASTn against the Silva database (release 108, reference small subunit and large subunit rRNAs; <http://www.arb-silva.de/>). After taxonomic assignment and binning (MEGAN), the results were represented graphically (KronaTools; <http://sourceforge.net/p/krona/wiki/KronaTools>).

Bacillariophyta (diatom) diversity was studied within a phylogenetic framework (Kembel *et al.*, 2011). In total, 20 ribosomal protein-coding genes were selected (Supplementary Table S2) based on

evolutionary conservation, high expression levels and low copy number in sequenced diatoms. Contigs were examined and edited when necessary (Codon-Code Aligner v. 1.6, Dedham, MA, USA) and those with  $\geq 98\%$  nucleotide identity merged into a single operational taxonomic unit. Homologous protein sequences were identified by BLASTp and/or genome annotations from four diatom genomes (*Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Pseudonitzschia multiseriata* and *F. cylindrus*) and the outgroups *Ectocarpus siliculosus* (brown alga) and *A. anophagefferens* (pelagophyte). Each protein data set was then aligned (MAFFT v.7, E-INS-i algorithm; Katoh and Standley, 2013).

Phylogenetic relationships for each protein were inferred by maximum likelihood (PhyML v3.0.1; Guindon and Gascuel, 2003), with node support estimated by approximate likelihood ratio tests under the LG evolutionary model (Le and Gascuel, 2008). These alignments were then concatenated into a single 2781 amino acid alignment stored in 20 partitions. Phylogenetic relationships were then inferred by maximum likelihood for the concatenated dataset. Node support was estimated by approximate likelihood ratio test and bootstrap analyses (1000 replicates). Phylogenetic reconstructions were represented using the ape-package v.2.5-1 (Paradis *et al.*, 2004) in R.

#### Metatranscriptome assembly quality

The accuracy of the assembly was tested by RT-PCR amplification of 12 predicted community-specific

**Table 2** Differential-enrichment of diatom transporters in metatranscriptomes of western Antarctic Peninsula communities

Family	Substrate	Hit accession	Description	BFS	WKI	WDS
ATP-binding cassette (ABC)	Fe	K02016	iron complex transport system, substrate-binding protein (ABC.FEV.S)	<b>35.40</b>	11.17	0
ATP-binding cassette (ABC)	unknown	K05643	subfamily A (ABC1), member 3	0	<b>53.05</b>	2.60
ATP-binding cassette (ABC)	unknown	K05655	subfamily B (MDR/TAP), member	0	<b>20.94</b>	0
ATP-binding cassette (ABC)	unknown	K05658	subfamily B (MDR/TAP), member	<b>23.89</b>	<b>76.79</b>	6.93
ATP-binding cassette (ABC)	unknown	K05665	subfamily C (CFTR/MRP), member 1	0	<b>47.47</b>	0.87
ATP-binding cassette (ABC)	unknown	K05673	subfamily C (CFTR/MRP), member 4	1.77	<b>33.51</b>	3.47
Solute carrier (SLC)	Glutamate	K05614	SCF 1 (glial high affinity glutamate transporter), member 3	0	<b>32.11</b>	0
Solute carrier (SLC)	Cation	K13749	SCF 24 (sodium/potassium/calcium exchanger), member 1	<b>31.86</b>	0.00	3.47
Solute carrier (SLC)	ATP/ADP	K05863	SCF 25 (mitochondrial carrier; adenine nucleotide translocator)	38.05	68.41	<b>157.75</b>
Solute carrier (SLC)	Amino acid	K13576	Sodium-coupled neutral amino acid transporter 3	<b>32.74</b>	0	6.07
Major Facilitator; SP family	Sugar	K08137	MFS transporter, SP family, galactose:H+ symporter	0	<b>79.58</b>	0
Major Facilitator; NNP family	Nitrate/Nitrite	K02575	MFS transporter, NNP family, nitrate/nitrite transporter	6.19	<b>79.58</b>	7.80
Amt	NH <sub>4</sub> <sup>+</sup>	K03320	Ammonium transporter	<b>131.85</b>	8.38	7.80
SSS	Urea	K03307	Solute:Na+ symporter, (urea transporter)	<b>260.16</b>	<b>54.45</b>	9.53
SIT	Si	PF03842	Silicon transporter	<b>252.20</b>	<b>62.83</b>	5.20
PNaS	PO <sub>4</sub> <sup>3-</sup>	K03324	Phosphate:Na+ symporter	<b>34.51</b>	<b>16.75</b>	0
ZIP	Fe	IPR003689	Zinc/iron permease	<b>45.00</b>	19.00	<b>71.00</b>
P-ATPase	Cation	K01552	Arsenite-transporting ATPase	<b>134.51</b>	18.15	24.27

Normalized read counts per hit accession are shown for BFS, WDS and WKI. Enrichment is indicated in bold (*R* statistic of Stekel, 2000).

Abbreviations: BFS, Bransfield Strait; WDS, Weddell Sea; WKI, Wilkins Ice Shelf.

gene targets. Total RNA was prepared from replicate filters as for NGS. Integrity of DNAase-treated RNA was checked on an Experion RNA chip (Bio-Rad Laboratories, LDA, Amadora, Portugal), and oligo-dT primed cDNA was synthesized with SuperScriptIII. PCR with primers designed using BatchPrimer3 (<http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>) amplified small (ca. 400 bp) transcript fragments, using 2 µl of 1/10 diluted cDNA in 25 µl using GoTaq (Promega), 0.2 µM each primer and 2.5 mM Mg. To check for gDNA contamination, 2 µl of 1/100 diluted total RNA were used in control PCRs. The PCR program was: 2' at 94 °C, 34 cycles of (30" at 94 °C, 30" at 58 °C, 30" at 72 °C), final extension 2' at 72 °C. The PCR reaction was run on 2% agarose gel alongside the 'GeneRuler 100bp' ladder (Fisher Scientific).

## Results and discussion

### Sample diversity and community features across habitats

After removing rRNA, 452 694 sequence reads were assembled into 58 101 contigs and 5448 singletons, with a N50 contig length of 398 bp (Table 1). Ribosomal reads accounted for 7.4–18.7% of the total reads, depending on the community (Table 1). Sequences were deposited in the NCBI short read archive (Accessions SRX727358, SRX727361, SRX727362). The resulting 'global' metatranscriptomic assembly has the advantages of (i) maximizing read numbers for contig assembly and transcriptome coverage and (ii) allowing co-assembly of homologous transcripts from different samples (subsequent analysis of normalized reads/sample yields relative enrichment estimates).

Community composition based on rRNA reflected the ecosystems from which they originated (Figure 1). The western BFS, situated on the continental shelf, was the warmest (about 2.0 °C at the surface) and most 'oceanic' (Figure 1a), with a mixed layer extending to 40 m depth. The converging water masses and complex hydrology support moderate primary production, often dominated by small flagellates (*Phaeocystis*, *Cryptomonas*) rather than diatoms (Mura et al., 1995, Varela et al., 2002). Sea ice floes adjacent to the WKI were sampled immediately following massive fragmentation and loss of ≈425 km<sup>2</sup> of the WKI (February 2009, [http://www.esa.int/esaCP/SEMD07EH1TF\\_index\\_0.html](http://www.esa.int/esaCP/SEMD07EH1TF_index_0.html)). Here a cold low salinity surface layer dominated, typical of the seasonal ice zone (Figure 1b). In the WDS, the water column at the post-bloom sampling time displayed a stable mixed layer and strongly developed pycnocline (Figure 1c), but had been more stratified near the bloom peak 3 weeks earlier (Supplementary Figure S1). The fluorescence signal clearly indicated that the algal planktonic community was beginning to sink below the pycnocline (Figure 1c).

Bacillariophyta (diatoms) dominated eukaryotic rRNA reads in BFS (67%) and WDS (73%) communities (Supplementary Figures S2a and c), while heterokont silicoflagellates (Dictyochophyceae; 51%) replaced diatoms (26%) as the dominant taxa sequenced from WKI sea ice (Supplementary Figure 2b). The latter were overwhelmingly Pedinellales; mixotrophic algae (Sekiguchi *et al.*, 2003) associated with ice-covered habitats (Thomsen 1988), which constituted <1% in the other communities. Dinoflagellates accounted for 7% and 8% of sequences from BFS and WDS, respectively, and only 2% in sea ice (WKI). Cryptophyta were largely restricted to BFS (2%). Heterokont algae (chiefly, diatoms and silicoflagellates), with 70–75% of rRNA reads, therefore appear to contribute overwhelmingly to primary productivity in these communities. Interestingly, fungal sequences featured in both BFS and WKI (10% and 5%, respectively), but not in WDS (<1%). Chytrids represent 78% of the fungal sequences in BFS, while reads from WKI were mainly assigned to Basidiomycetes (ca. 40%) and only 10% to chytrids. Fungi have rarely been reported from Antarctic waters, and deserve future attention (López-García *et al.*, 2001), particularly given the potential role of chytrids as parasites on phytoplankton, likely affecting turnover, succession and nutrient recycling (Kagami *et al.*, 2014). Ciliates in WDS (4% rRNA) may play a significant role, perhaps feeding on the abundant bacteria present, and on other algae.

#### *Contrasting phylogenetic and functional diversity within diatoms*

Taxonomic assignment of rRNA reads from the metatranscriptome provided useful information. Bacillariophyceae (= pennate diatoms) dominated BFS and WKI with 60–65% of assigned reads, while in the WDS 50% were Coscinodiscophyceae (= centric diatoms) with only 29% Bacillariophyceae (Supplementary Figure S2).

The phylogenetic method used to describe diatom diversity (based on Kembel *et al.*, 2011) circumvents single gene amplicon approaches (for example, 18S-rRNA). Alignment of individual ribosomal protein contigs against concatenated sequence scaffolds derived from whole genomes revealed divergent communities with distinct taxonomic compositions (Figure 2a). Phylogenetic assignment is limited by the availability of only four diatom genomes, but is consistent with diversity estimates by microscopy (Supplementary Figure S3). *Pseudo-nitzschia* sp. dominated the BFS sample, and most sequences from BFS clustered near *Pseudo-nitzschia multiseries* on the tree. Sequences from sea ice (WKI) were more dispersed over the tree, but many clustered near the sea ice diatom *F. cylindrus*, an abundant genus in our sample. Higher diversity was found in WDS, but most sequences were concentrated near the centric diatom *T. pseudonana* (Figure 2a).

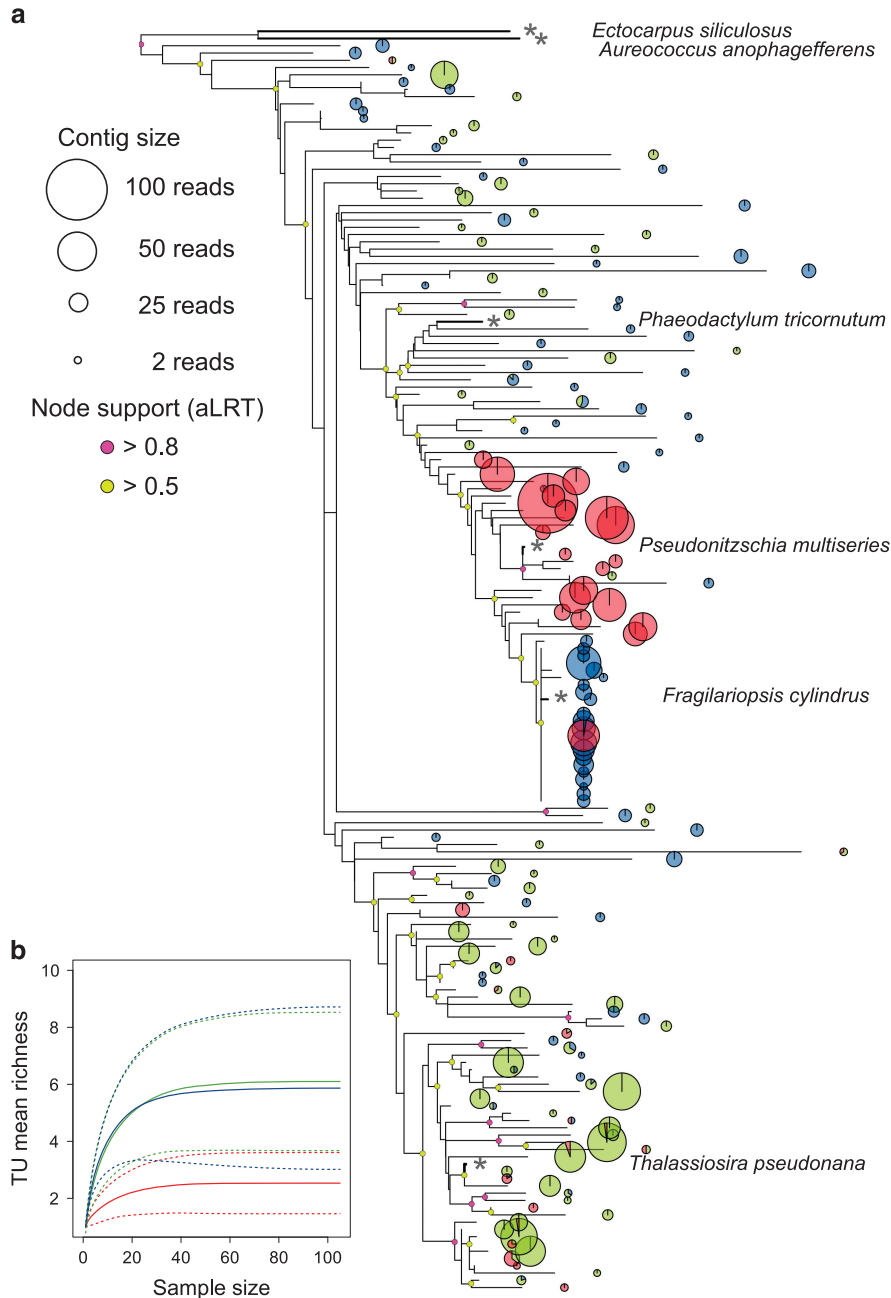
Whole-assembly metrics of sequence similarity support the phylogenetic results; very few contigs contained reads from all three communities (0.6% of contigs, 3.2% of reads), with the large majority uniquely assigned to a single community (93% of contigs, 82% of reads) (Figure 3a). The accuracy of the assembly was checked using RT-PCR, in which 8 out of 12 community-specific contigs tested showed community-specific amplification (Supplementary Figures S4 and S5).

Unsurprisingly, functional similarity between the three diatom communities appeared to be much greater than their taxonomic relatedness (cf. Venn diagrams; Figures 3a and b). Overall, 54% of KEGG and Pfam annotations (91% of reads) were shared between  $\geq 2$  communities, a core set of which (27% of terms, 77% of reads) were common to all three communities (Figure 3b). Nevertheless, 46% of functional terms were unique to one community (ca. 15% unique terms per community), indicating considerable habitat specificity in functional processes and constraints.

#### *Variation in diatom metabolic pathways across communities*

The number of reads mapped to central metabolic pathways varied across communities (Figure 4a). Most importantly, key pathways for carbohydrate (glycolysis/gluconeogenesis and pentose phosphate cycle) and energy metabolism (oxidative phosphorylation, photosynthesis, carbon fixation) were under-represented in sea ice (WKI) relative to pelagic diatoms (BFS and WDS). Hence, a potentially depleted supply of reductant *via* electron transport pathways in sea ice diatoms was matched by down-regulation of both carbon fixation (Calvin/Benson cycle) and the oxidative pentose phosphate cycle, indicating restricted flow of energy into carbohydrates compared with pelagic communities. Reduced carbohydrate assimilation due to low temperature and irradiance may be common in diatoms inhabiting brine channels on the underside of sea ice (reviewed by Thomas and Dieckmann, 2002; Mock and Thomas, 2008; Arrigo, 2014).

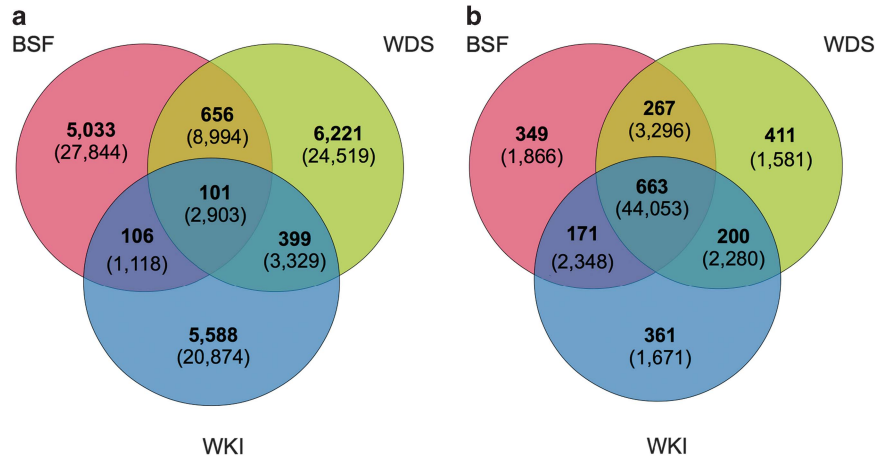
Read counts for carbohydrate, lipid and energy metabolic pathway orthologues confirmed the depletion of general metabolic functions in WKI sea ice; most orthologues lie along the BFS–WDS axis (ternary plot, Figure 4b). Two highly expressed orthologues were transketolase (TKT; a key Calvin cycle and pentose phosphate pathway enzyme, and link to other carbohydrate pathways) and glyceraldehyde 3-phosphate dehydrogenase (GADPH; key enzyme in glycolysis), each had four to fivefold more reads in BFS and WDS compared with WKI. Other photosynthesis and Calvin cycle genes were over-expressed in either or both pelagic communities compared with sea ice (Figure 4b; Supplementary Table S1).



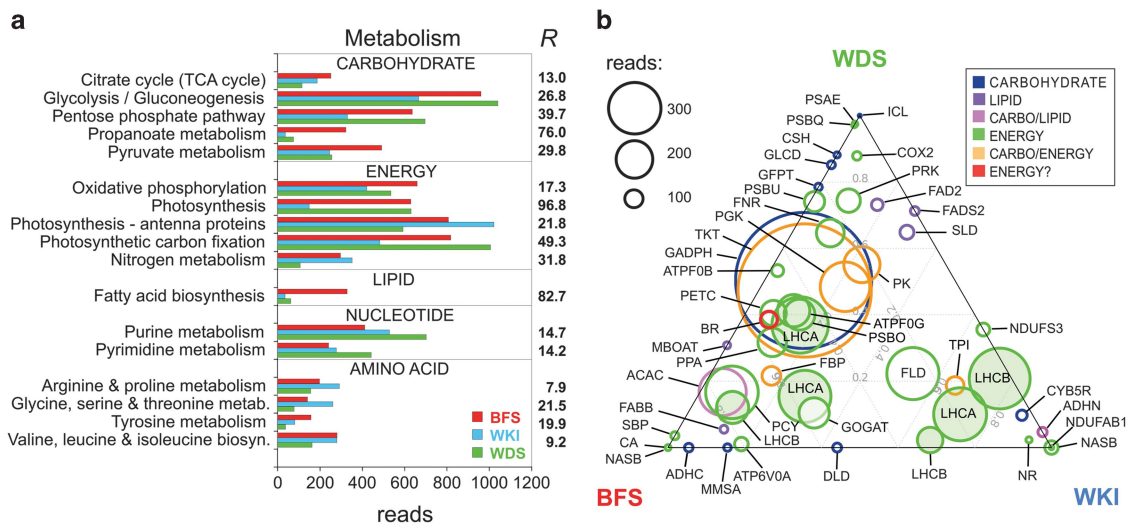
**Figure 2** (a) Maximum likelihood (ML) phylogenetic analysis of diatoms from the western Antarctic Peninsula/WDS communities; Bransfield Strait (BFS; red), Weddell Sea (WDS; green) and Wilkins sea-ice (WKI; blue). The ML tree is based on protein-coding sequences for 20 ribosomal protein genes identified in the metatranscriptome assembly, aligned against the concatenated reference proteins from the genomes of the stramenopiles *E. siliculosus* (brown alga), *A. anophagefferens* (pelagophyte), *P. tricornutum*, *Pseudo-nitzschia multiseriata*, *F. cylindrus* and *T. pseudonana* (diatoms). The reference species are indicated by name on the tree, and their position is shown by an asterisk. The size of the pie at each node is proportional to the number of reads identified in the contig, and the colour(s) indicates the community from which the sequences originate. (b) Rarefaction analysis of the same ribosomal protein dataset used for tree construction. Contigs for each gene were considered independent taxonomic units (TU) at  $\leq 98\%$  identity. Solid lines show the expected mean TU richness, dotted lines show the standard deviation.

Photosynthesis-related transcripts were under-, whereas photosynthetic antenna proteins were over-represented in WKI sea ice diatoms (Figures 4a and b). Further annotation revealed that most (ca. 65%) of the latter encoded stress-related light-harvesting complex proteins (LHCSR or LI818 clade). Psychrophilic diatoms adjust their capacity for non-

photochemical quenching at low temperatures (Mock and Hoch, 2005) and LHCSR proteins are thought to function in photoprotective energy dissipation, analogous to PsbS in higher plants—a gene missing from diatom genomes (Zhu and Green, 2008). However, although some LHCSR family members in *T. pseudonana* respond to high light



**Figure 3** Venn diagrams illustrating the relationship between sequence similarity and functional assignment in the metatranscriptome assembly for diatom communities from the West Antarctic Peninsula; BFS, WDS and WKI. (a) Numbers of putative protein-coding contigs (bold) and sum of the total reads (in parentheses) assigned to Bacillariophyta (using MEGAN and based on NCBI taxonomy), and their distribution between the three communities. (b) Functional annotation of Bacillariophyta. In bold are shown the number of annotation terms (KEGG and PFAM non-redundant set), and in parentheses the corresponding total number of reads.



**Figure 4** Functional assignment of diatom metatranscriptome sequences to metabolic pathways. (a) Total numbers of sequence reads (normalized by library size) assigned to KEGG Carbohydrate, Energy, Lipid, Nucleotide and Amino acid metabolic pathways. Each of the pathways shown differs in expression between the three communities ( $R$ -statistics shown to the right of the plot; Stekel, 2000). (b) Ternary plot showing differentially expressed KEGG orthologues in carbohydrate, lipid and energy metabolic pathways. Orthologues are represented by circles whose size is proportional to read number. The position of an orthologue relative to the vertices of the triangular plot indicates its relative expression by each community sample. ACAC, acetyl-CoA carboxylase/biotin carboxylase; ADHC, alcohol dehydrogenase (cytochrome c); ADHN, alcohol dehydrogenase (NADP+); ATP6V0A, V-type H+-transporting ATPase subunit I; ATPF0B, F-type H+-transporting ATPase subunit b; ATPF0G, F-type H+-transporting ATPase subunit gamma; BR, bacteriorhodopsin; CA, carbonic anhydrase; COX2, cytochrome c oxidase subunit II; CSH, chitin synthase; CYB5R, cytochrome-b5 reductase; DLD, dihydrolipoamide dehydrogenase; FABB, 3-oxoacyl-[acyl-carrier-protein] synthase I; FAD2, omega-6 fatty acid desaturase (delta-12 desaturase); FADS2, fatty acid desaturase 2 (delta-6 desaturase); FBP, fructose-1,6-bisphosphatase I; FLD, flavodoxin I; FNR, ferredoxin-NADP+ reductase; GADPH, glyceraldehyde 3-phosphate dehydrogenase; GFPT, glucosamine-fructose-6-phosphate aminotransferase (isomerizing); GLCD, glycolate oxidase; GOGAT, glutamate synthase (ferredoxin); ICL, isocitrate lyase; LHCA, light-harvesting complex I chlorophyll a/b binding protein 1; LHCA, light-harvesting complex I chlorophyll a/b binding protein 2; LHCA, light-harvesting complex I chlorophyll a/b binding protein 3; LHCB, light-harvesting complex II chlorophyll a/b binding protein 1; LHCB, light-harvesting complex II chlorophyll a/b binding protein 4; LHCB, light-harvesting complex II chlorophyll a/b binding protein 5; MBOAT, lysophospholipid acyltransferase 1/2; MMSA, methylmalonate-semialdehyde dehydrogenase; NASB, nitrite reductase (NAD(P)H) large subunit; NASB, nitrite reductase (NAD(P)H) small subunit; NDUFAB1, NADH dehydrogenase (ubiquinone) 1 alpha/beta subcomplex 1; NDUFS3, NADH dehydrogenase (ubiquinone) Fe-S protein 3; NR, nitrate reductase (NADH); PCY, plastocyanin; PETC, cytochrome b6-f complex iron-sulfur subunit; PGK, phosphoglycerate kinase; PK, pyruvate kinase; PPA, inorganic pyrophosphatase; PRK, phosphoribulokinase; PSAE, photosystem I subunit IV; PSBO, photosystem II oxygen-evolving enhancer protein 1; PSBQ, photosystem II oxygen-evolving enhancer protein 3; PSBU, photosystem II PsbU protein; SBP, sedoheptulose-bisphosphatase; SLD, delta8-fatty-acid desaturase; TKT, transketolase; TPI, triosephosphate isomerase.



conditions (Zhu and Green, 2008), not all are stress-induced (Bailleul *et al.*, 2010), and they probably play other important and integral roles in mediating diatom responses to light. LHCSR-like proteins are also important in WDS, accounting for ca. 50% of antenna protein reads (Supplementary Table S2), but for only ca. 1% in BFS (where diurnal expression changes cannot be excluded). BLASTx analysis suggests over-representation of centric (*Thalassiosira*-like) sequences in the WDS, and *Fragilariopsis*-like sequences in the WKI (Supplementary Table S2). Furthermore, using RT-PCR, we confirmed the specific expression of a WKI LHCSR transcript in sea ice (Supplementary Figure S5).

Although diatom genomes encode the enzymes necessary for classical C4 metabolism (Reinfelder, 2011), subcellular localizations predict futile inorganic carbon cycling and ATP consumption, and a possible role in excess energy dissipation and/or cellular pH homeostasis (Kroth *et al.*, 2008; Haimovich-Dayana *et al.*, 2012). Although not abundant, the key C4 enzyme pyruvate-orthophosphate dikinase (PPDK) was uniquely expressed in WKI, and phosphoenol pyruvate carboxylase (PEPC) was also over-represented (Supplementary Table S1). Sea ice diatoms therefore seem to activate several distinct (but interdependent) mechanisms to deal with excitation pressure arising from high irradiance/low temperature and/or carbon limitation.

Bacteriorhodopsin-like proteins (Pfam PF01036; KEGG orthologue K04641) related to *F. cylindrus* (Protein id 267528; <http://genome.jgi-psf.org/Fracy1/Fracy1.home.html>; see also Marchetti *et al.*, 2012) were identified in 29 contigs. Read numbers were highest in BFS and WDS (BR; Figure 4b), with greater sequence diversity in the latter (Supplementary Figure S6). Enrichment was much lower (ca. five-fold) in WKI. Bacteriorhodopsins are light-dependent proton pumps in archaea and eubacteria, and related proteins occur in some eukaryotes (Waschuk *et al.*, 2005; Wada *et al.*, 2011). Their presence in marine eukaryotic phytoplankton genomes is thought to result from lateral gene transfer from bacteria (Slamovits *et al.*, 2011). Although still unclear, bacteriorhodopsins could function under prevailing low-Fe conditions as an alternative mechanism for ATP generation (Marchetti *et al.*, 2012).

Fatty acid biosynthesis (lipid metabolism) was strongly overexpressed in BFS (Figure 4a), and only this community expressed the full KEGG complement of genes reported for diatoms (Supplementary Table S1). Particularly striking was the 5–10-fold enrichment of acetyl-CoA carboxylase (ACAC, Figure 4b), which catalyzes one of the initial steps in fatty acid biosynthesis. Diatoms accumulate lipids under nutrient-limited conditions (N, Si or Fe; Roessler, 1988; Mock and Kroon, 2002a; Allen *et al.*, 2008), particularly relevant in light of enriched transport functions (see following section).

Several fatty acid desaturase transcripts (delta-12 desaturase (FAD2), delta-6 desaturase (FADS2) and delta-8 desaturase (SLD)) were over-represented in WDS > WKI > BFS (Figure 4b). The biosynthesis of polyunsaturated fatty acids and membrane remodelling by desaturases may be critical for regulating membrane fluidity at low temperatures (Wada *et al.*, 1990; Nishida and Murata, 1996). Polyunsaturated fatty acids or their aldehyde derivatives are also important in signalling and grazer defence (Pohnert, 2002; Vardi *et al.*, 2008), which might be especially important during late bloom conditions.

#### *Transporter and environmental sensing genes as indicators of physiological status*

The chimeric genomes of diatoms confer an unrivalled flexibility among phytoplankton groups to environmental conditions, such as spatio-temporal variability in nutrient levels (Armbrust *et al.*, 2004; Bowler *et al.*, 2008). Here, we identified a wide variety of diatom nutrient acquisition and transport genes, of which 34% (18 of 53 functional terms) showed differential enrichment between communities (Table 2). Ammonium and urea transporters were enriched in the *Pseudo-nitzschia*-dominated BFS community, suggesting that reduced forms of N are important in this habitat, despite the apparent sufficiency of nitrate in the water column (Figure 1b). Silicon transporters were between 4-fold (versus WKI) and 50-fold (versus WDS) enriched in BFS. The energy required for silicic acid uptake appears to derive from respiratory rather than photosynthetic activity (Martin-Jézéquel *et al.*, 2000), consistent with enrichment of both TCA cycle genes (BFS > WKI > WDS; Figure 4a) and vacuolar (V-type) ATPase subunits that appear closely involved in silica biomineralization processes (Vartanian *et al.*, 2009; Supplementary Table S1). Although transcript and protein levels of silicon transporters are loosely correlated (Thamatrakoln and Hildebrand, 2007), silicic acid uptake is closely coordinated with cell cycle activity and division. The *Pseudo-nitzschia* spp. in the BFS community were mitotically active, consistent with enhanced silicic acid uptake relative to species in WKI and particularly WDS communities.

Overall, ATP-binding cassette (ABC) superfamily members were enriched in WKI. ABC transporters are multicomponent active (ATP-dependent) permeases that uptake a broad range of substrates. The specific affinities of ABC members identified here are largely unknown, but their relative enrichment suggests a reliance on energy-dependent uptake of certain nutrients by sea ice diatoms. The major facilitator superfamily and solute carrier (SLC) membrane transporters are largely solute exchangers operating in response to chemiosmotic gradients. Of these, a nitrate/nitrite transporter was ca. 10-fold enriched in WKI, while glutamate and sugar transporters were also uniquely identified in this

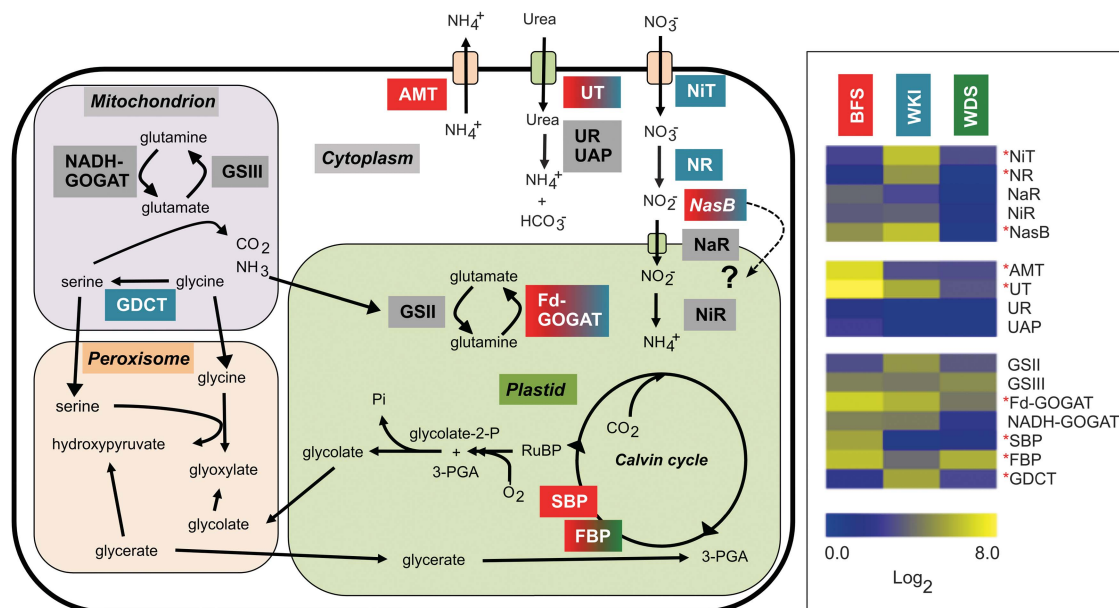
community (Table 2). The WDS community was enriched for only two transporters; a mitochondrial ATP/ADP transporter, responsible for ATP export to the cytosol, and (together with BFS) a ZIP family permease putatively involved in Fe uptake.

#### *Diatom nitrogen metabolism is flexible among habitats*

Expression of primary N-uptake systems differed considerably in the three areas sampled (Figure 5 and Table 2). Subcellular localization data were inferred from annotated diatom gene sequences and the available literature (Armbrust *et al.*, 2004; Bowler *et al.*, 2008; Brown *et al.*, 2009; Kinoshita *et al.*, 2009; Allen *et al.*, 2011). A caveat to the following discussion is that variation in diurnal timing (for BFS) could have transcriptome-wide effects on relative enrichment (Ashworth *et al.*, 2013; Chauton *et al.*, 2013). In contrast to ammonium and urea transporters discussed in the previous section, nitrate transporters were less abundant in the more oceanic BFS community, despite high  $\text{NO}_3^-$  levels in the water column (Figure 1b). The Calvin cycle marker genes sedoheptulose biphosphatase (SBP) and fructose biphosphatase (FBP) (Parker and Armbrust, 2005; Kroth *et al.*, 2008) were both most abundant in BFS (Figure 5) indicating that N availability is not affecting these processes. These

enrichment patterns therefore appear to reflect a physiological stance towards the uptake of reduced N sources, consistent with experimental evidence that  $\text{NH}_4^+$  limits phytoplankton growth in Fe-sufficient areas of the Southern Ocean (Agustí *et al.*, 2009), such as the BFS and the Antarctic Peninsula (Martin *et al.*, 1990; Sañudo-Wilhelmy *et al.*, 2002). Given the enrichment of Fe-transport systems in this community (Table 2), such patterns of N use may minimize the use of Fe-rich N-acquisition proteins, despite similar median values of  $1.8 \text{ nmol Fe l}^{-1}$  across the regions sampled in this study (Tovar-Sanchez, unpublished results).

In contrast, the WKI sea ice diatom community seems to primarily assimilate nitrate. Nitrate transporters, assimilatory nitrate and nitrite reductases (NR and NasB) were most enriched in this community (Figure 5). Transcript profiles may not reflect protein levels (or activity), because although nitrate is required for NR protein accumulation (Poulsen and Kröger, 2005), mRNA levels can remain high under  $\text{NO}_3^-$  limitation. Nevertheless, enrichment of plastid glutamine synthetase (GSII), which is transcriptionally regulated by external  $\text{NO}_3^-$ , but not  $\text{NH}_4^+$ , supports  $\text{NO}_3^-$  as the main N source in the WKI (Takabayashi *et al.*, 2005). Moreover, although sea ice environments are generally considered N-limited, owing to reduced exchange with surrounding



**Figure 5** Simplified schematic representing N acquisition and N-based photorespiratory pathways in a diatom cell (left). The subcellular localization of gene products was determined by reference to annotation data and the literature (see text for details). The uncertain localization of NasB is indicated by dashed arrow and "?". The relative expression (as  $\text{Log}_2$  reads) of the gene products highlighted in colour are shown on the right for BFS, Wilkins sea-ice (WKI) and WDS communities. Asterisks in red indicate significant expression differences between the communities ( $R$ -values (where  $> 8 \approx 98\%$  believability) and annotation codes (KEGG/Pfam/BLASTx) are given in parentheses after the gene acronyms). AMT, ammonium transporter (44.7, K03320); FBP, fructose-1,6-bisphosphatase (9.3, K03841); GDCT, glycine decarboxylase (57.9, K00281); Fd-GOGAT, ferredoxin-glutamate synthase (plastidial) (13.8, K00284); NADH-GOGAT, NADH-glutamate synthase (cytosolic) (2.3/3.1, K00264/K00266); GSII, glutamine synthetase (plastid) (1.1, K01915); GSIII, glutamine synthetase (cytosolic) (1.1, K01915); NaR, nitrite transporter (4.1, PF01226); NasB, NAD(P)H-nitrite reductase (15.9, K00363); NiR, ferredoxin-nitrite reductase (plastid) (2.1, K00366); NiT, nitrate transporter (21.7, K02575); NR, nitrate reductase (12.3, K00360); SBP, sedoheptulose biphosphatase (20.4, K01100); UAP, urease accessory protein (2.6, K03189); UR, urease (BLASTx); UT, urea transporter (74.4, K03307).

seawater (Mock and Kroon, 2002a, 2002b), we sampled ice near the ice–water interface (surface to 5–10 cm depth), where somewhat higher nutrient levels are expected (Lyon *et al.*, 2011). Interestingly, urea transporters were relatively enriched in WKI, suggesting that reduced organic N from heterotrophic sources may provide additional N within sea ice brine channels (Table 2).

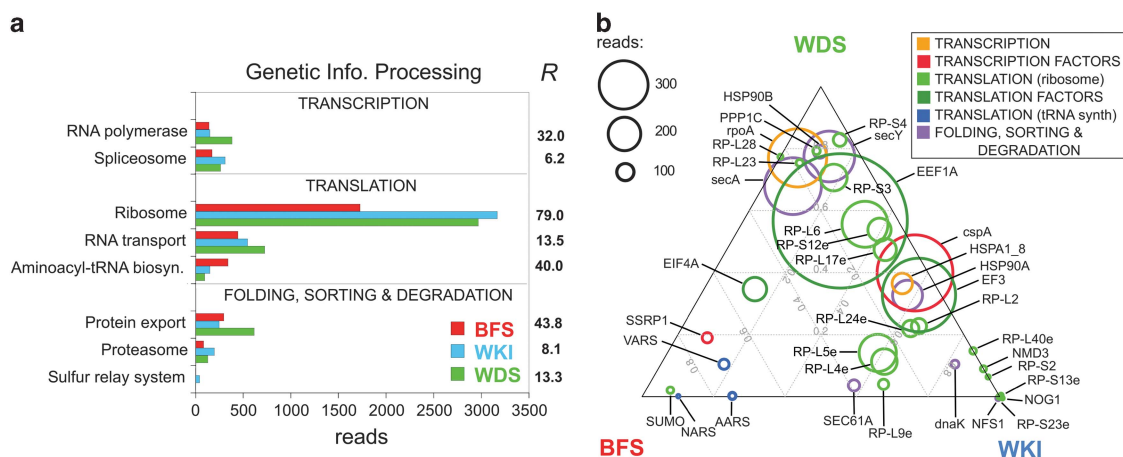
The assimilation of nitrate by sea ice diatoms may help in controlling cellular energy balance under photorespiratory conditions (Lomas and Glibert, 1999). Photorespiration in WKI is suggested by abundant photoprotective LHCSR proteins, reduced photosynthetic pathway enrichment (Figure 4) and by enrichment of mitochondrial glycine decarboxylase (GDCT), a key enzyme and marker for photorespiration (Douce *et al.*, 2001). Photorespiration consumes ATP and reductant 'wastefully', but may prevent overreduction of the photosynthetic electron transport chain (Wingler *et al.*, 2000), while photorespiratory glycolate may contribute to DOC production in sea ice (Arrigo, 2014). Under high light and/or low CO<sub>2</sub> conditions that lead to photoinhibition of photosynthesis, particularly at low temperatures where carbon fixation is enzymatically inefficient, reduction of intracellular NO<sub>3</sub><sup>-</sup> pools by NR may buffer excess

electron transport capacity and act as an effective photoprotective mechanism (Lomas and Glibert, 1999).

In comparison with both BFS and WKI, transporters for NO<sub>3</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup> and urea were all under-represented in WDS (Table 2). Assimilatory pathways were also depleted, (Figure 5), which could indicate severe N limitation in the post-bloom mixed layer. Three weeks before sampling, at the bloom peak, NO<sub>3</sub><sup>-</sup>/NO<sub>2</sub><sup>-</sup> was < 2 μM, and NH<sub>4</sub><sup>+</sup> was 0.2 μM at the same station (Supplementary Figure S1).

#### Enrichment patterns at low temperature

Several contrasts between diatom communities are consistent with strong temperature effects on gene expression and metabolism, even across a relatively small gradient within the wAP (from < -1.8 to 2.0 °C). Communities differed greatly in pathway enrichment for 'Genetic Information Processing', most strikingly in the ribosome pathway, which was enriched in the WDS and WKI relative to the warmer BFS (Figure 6a). In addition to ribosomal proteins, the levels of several other orthologues including translation factors (EEF1A, EF3) are biased towards these two colder communities (Figure 6b). Overexpression



**Figure 6** Functional assignment of diatom metatranscriptome sequences to KEGG Genetic Information Processing pathways. Each of the pathways shown differs in expression between the three communities (*R*-statistics shown to the right of the plot; Stekel, 2000). (a) Total numbers of sequence reads (normalized by library size) assigned to KEGG Transcription, Translation, and Folding, sorting and degradation pathways. (b) Ternary plot showing differentially expressed KEGG orthologues in Transcription, Translation, and Folding, sorting and degradation pathways. Orthologues are represented by circles whose size is proportional to read number. The position of an orthologue relative to the vertices of the triangular plot indicates its relative expression by each community sample. AARS, alanyl-tRNA synthetase; cspA, cold shock protein (beta-ribbon, CspA family); dnaK, molecular chaperone DnaK; EEF1A, elongation factor EF-1 alpha subunit; EF3, elongation factor EF-3; EIF4A, translation initiation factor eIF-4A; HSP90A, molecular chaperone Hsp90; HSP90B, heat shock protein 90 kDa beta; HSPA1\_8, heat shock 70 kDa protein 1/8; NARS, asparaginyl-tRNA synthetase; NFS1, cysteine desulfurase; NMD3, nonsense-mediated mRNA decay protein 3; NOG1, ribosomal RNA large subunit methyltransferase N; PPP1C, protein phosphatase 1, catalytic subunit; RP-L17e, large subunit ribosomal protein L17e; RP-L2, large subunit ribosomal protein L2; RP-L23, large subunit ribosomal protein L23; RP-L24e, large subunit ribosomal protein L24e; RP-L28, large subunit ribosomal protein L28; RP-L40e, large subunit ribosomal protein L40e; RP-L4e, large subunit ribosomal protein L4e; RP-L5e, large subunit ribosomal protein L5e; RP-L6, large subunit ribosomal protein L6; RP-L9e, large subunit ribosomal protein L9e; RP-S12e, small subunit ribosomal protein S12e; RP-S13e, small subunit ribosomal protein S13e; RP-S20, small subunit ribosomal protein S20; RP-S23e, small subunit ribosomal protein S23e; RP-S3, small subunit ribosomal protein S3; RP-S4, small subunit ribosomal protein S4; rpoA, DNA-directed RNA polymerase subunit alpha; SEC61A, protein transport protein SEC61 alpha; secA, preprotein translocase subunit SecA; secY, preprotein translocase subunit SecY; SSRP1, structure-specific recognition protein 1; SUMO, small ubiquitin-related modifier; VARS, valyl-tRNA synthetase.

of the translational machinery is a response to compensate rates of protein synthesis under low temperatures. The capacity for thermal acclimation of RNA translation or protein synthesis appears limited, largely relying on production of additional ribosomes (Nomura *et al.*, 1984; Fraser *et al.*, 2002; Storch and Pörtner, 2003; Kim *et al.*, 2004). Such a large difference in ribosomal protein enrichment implies large increases in cellular RNA at lower temperatures, as RNA:protein in ribosomes is constant. Over a 3–4 °C range, this could have important consequences for population growth rates and nutrient stoichiometry, particularly N:P, as ribosomes are relatively P-rich (Toseland *et al.*, 2013).

The levels of several other transcript classes was also correlated with low temperatures: (i) CSP family cold-shock transcription factors (possible 'RNA chaperones'; Phadtare, 2004) (Figure 6b). (ii) Fatty acid desaturases ( $\Delta -6$ ,  $-8$  and  $-12$ ), implicated in the control of membrane fluidity during acclimation to cold (D'Amico *et al.*, 2006; Figure 4b; Supplementary Table S1). (iii) Ice-binding proteins were prominent in the sea ice community, but absent in BFS.

Some 65 ice-binding protein contigs were identified (ca. 490 reads) from bacteria, archaea and fungi, but diatoms were the most abundant source, accounting for ca. 390 reads in WKI, followed by copepod crustaceans (top BLASTx hit *Stephos longipes*).

## Conclusions

Metatranscriptomes revealed the physiological and metabolic flexibility of diatom genomes facing the challenges of different habitats in the Antarctic Peninsula. The metatranscriptome of a sea ice community reflected its stressful environment: carbohydrate and energy metabolism pathways were depleted relative to pelagic communities, while direct and indirect light energy dissipation pointed to irradiance stress and/or inorganic carbon limitation within the ice.

Transcript abundances for N acquisition and transport varied widely across habitats. Their expression was reduced or absent under post-bloom conditions in the WDS community, suggesting dependence on internal stores for continued growth, in marked contrast with the BFS community where transporters for N sources were highly enriched. The relative enrichment of fatty acid biosynthesis in BFS likely reflects nutrient (possibly including Fe) stress in this community, which consists of species (mainly *Pseudo-nitzschia* spp.) with coping strategies.

Most remarkable was the evidence for an influence of temperature over the narrow 4 °C range separating the communities. Investment in transcription and translation functions (mainly ribosomal proteins) decreased with increasing temperature (Supplementary Figure S7). A similar phenomenon was recently reported over a much broader thermal range and taxon diversity (Toseland *et al.*, 2013). In summary, this

initial metatranscriptomic survey revealed a range of functional responses in diatom communities that reflect the ecological diversity of habitats in waters around the unique Antarctic region.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgements

This work was supported by grants 'SOPA' from the Portuguese Science Foundation (FCT; PTDC/MAR/72630) to GAP and is a contribution to the ATOS project (POL2006-00550/CTM) funded by the Spanish Ministry of Economy and Competitiveness (under the scope of the International Polar Year (IPY)). We are grateful to the crew of the R/V *Hesperides* for technical and logistical support. We thank A. Tovar-Sanchez for information on iron concentrations in the waters sampled.

## References

- Agustí S, Duarte CM, Llabrés M, Agawin NSR, Kennedy H. (2009). Response of coastal Antarctic phytoplankton to solar radiation and ammonium manipulation: An in situ mesocosm experiment. *J Geophys Res* **114**: G01009.
- Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP *et al.* (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**: 203–207.
- Allen AE, Laroche J, Maheswari U, Lommer M, Schauer N, Lopez PJ *et al.* (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc Natl Acad Sci USA* **105**: 10438–10443.
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH *et al.* (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**: 79–86.
- Arrigo KR. (2014). Sea Ice Ecosystems. *Ann Rev Mar Sci* **6**: 439–467.
- Arrigo KR, Thomas DN. (2004). Large scale importance of sea ice biology in the Southern Ocean. *Antarct Sci* **16**: 471–486.
- Arrigo KR, Worthen DL, Lizotte MP, Dixon P, Dieckmann G. (1997). Primary production in Antarctic sea ice. *Science* **276**: 394–397.
- Ashworth J, Coesel S, Lee A, Armbrust EV, Orellana M V, Baliga NS. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. *Proc Natl Acad Sci USA* **110**: 7518–7523.
- Bailleul B, Rogato A, Martino ADE, Coesel S, Cardol P, Bowler C. (2010). An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc Natl Acad Sci USA* **107**: 18214–18219.
- Bopp L, Aumont O, Cadule P, Alvain S, Gehlen M. (2005). Response of diatoms distribution to global warming and potential implications: A global model study. *Geophys Res Lett* **32**: L19606.

- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A *et al.* (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–244.
- Boyce DG, Lewis MR, Worm B. (2010). Global phytoplankton decline over the past century. *Nature* **466**: 591–596.
- Boyd PW. (2002). Environmental factors controlling phytoplankton processes in the Southern Ocean. *J Phycol* **38**: 844–861.
- Brown KL, Twing KI, Robertson DL. (2009). Unraveling the regulation of nitrogen assimilation in the marine diatom *Thalassiosira pseudonana* (Bacillariophyceae): diurnal variations in transcript levels for five genes involved in nitrogen assimilation. *J Phycol* **45**: 413–426.
- Chauton MS, Winge P, Brembu T, Vadstein O, Bones AM. (2013). Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. *Plant Physiol* **161**: 1034–1048.
- Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T *et al.* (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* **14**: 1147–1159.
- Clarke A, Murphy EJ, Meredith MP, King JC, Peck LS, Barnes DKA *et al.* (2007). Climate change and the marine ecosystem of the western Antarctic Peninsula. *Philos Trans R Soc B Biol Sci* **362**: 149–166.
- D'Amico S, Collins T, Marx J-C, Feller G, Gerday C. (2006). Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**: 385–389.
- Douce R, Bourguignon J, Neuburger M, Rébeillé F. (2001). The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci* **6**: 167–176.
- Ducklow HW, Baker K, Martinson DG, Quetin LB, Ross RM, Smith RC *et al.* (2007). Marine pelagic ecosystems: the West Antarctic Peninsula. *Philos. Trans R Soc B Biol Sci* **362**: 67–94.
- Ducklow HW, Fraser WR, Meredith MP, Stammerjohn SE, Doney SC, Martinson DG *et al.* (2013). West Antarctic Peninsula: An ice-dependent coastal marine ecosystem in transition. *Oceanography* **26**: 190–203.
- Eddy SR. (1998). Profile hidden Markov models. *Bioinformatics* **14**: 755–763.
- Feld CB, Behrenfeld MJ, Randerson JT, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.
- Fraser KPP, Clarke A, Peck LS. (2002). Low-temperature protein metabolism: seasonal changes in protein synthesis and RNA dynamics in the Antarctic limpet *Nacella concinna* Strebel 1908. *J Exp Biol* **205**: 3077–3086.
- Guindon S, Gascuel O. (2003). A Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
- Haimovich-Dayan M, Garfinkel N, Ewe D, Marcus Y, Gruber A, Wagner H *et al.* (2012). The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytol* **197**: 177–185.
- Hansen K, Koroleff FF. (1999). Determination of nutrients. In: Grasshoff K, Kremling K, Ehrhardt M (eds). *Methods of Seawater Analysis*. Wiley-VCH Verlag GmbH, Weinheim: Germany, pp 159–228.
- Huang Y, Gilna P, Li W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**: 1338–1340.
- Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552–1560.
- Kagami M, Miki T, Takimoto G. (2014). Mycoloop: chytrids in aquatic food webs. *Front. Microbiol* **5**: 166.
- Katoh K, Standley DM. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.
- Kemmel SW, Eisen J a, Pollard KS, Green JL. (2011). The phylogenetic diversity of metagenomes. *PLoS One* **6**: e23214.
- Kerouel R, Aminot A. (1997). Fluorometric determination of ammonia in sea and estuarine waters by direct segmented flow analysis. *Mar Chem* **57**: 265–275.
- Kim K, Park S, Chung Y, Chung C, Kim J, Lee J. (2004). Molecular cloning of low-temperature-inducible ribosomal proteins from soybean. *J Exp Bot* **55**: 1153–1155.
- Kinoshita S, Isu S, Kaneko G, Yamada H, Hara T, Itoh Y *et al.* (2009). The occurrence of eukaryotic type III glutamine synthetase in the marine diatom *Chaetoceros compressum*. *Mar Genomics* **2**: 103–111.
- Korb RE, Whitehouse MJ, Gordon M, Ward P, Poulton AJ. (2010). Summer microplankton community structure across the Scotia Sea: implications for biological carbon export. *Biogeosciences* **7**: 343–356.
- Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS *et al.* (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* **3**: e1426.
- Lasternas S, Agustí S. (2010). Phytoplankton community structure during the record arctic ice-melting of summer 2007. *Polar Biol* **12**: 1709–1717.
- Le SQ, Gascuel O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307–1320.
- Lomas MW, Glibert PM. (1999). Temperature regulation of nitrate uptake: A novel hypothesis about nitrate uptake and reduction in cool-water diatoms. *Limnol Oceanogr* **44**: 556–572.
- López-García P, Rodríguez-Valera F, Pedrós-Alió C, Moreira D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603–607.
- Lyon BR, Lee PA, Bennett JM, DiTullio GR, Janech MG, Johnson RH. (2011). Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. *Plant Physiol* **157**: 1926–1941.
- Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT *et al.* (2012). Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci USA* **109**: E317–E325.
- Martin JH, Fitzwater SE, Gordon RM. (1990). Iron deficiency limits phytoplankton growth in Antarctic waters. *Global Biogeochem. Cycles* **4**: 5–12.
- Martin-Jézéquel V, Hildebrand M, Brzezinski MA. (2000). Silicon metabolism in diatoms: implications for growth. *J Phycol* **36**: 821–840.
- Meredith MP, King JC. (2005). Rapid climate change in the ocean west of the Antarctic Peninsula during the

- second half of the 20th century. *Geophys Res Lett* **32**: L19604.
- Mock T, Hoch N. (2005). Long-term temperature acclimation of photosynthesis in steady-state cultures of the polar diatom *Fragilariopsis cylindrus*. *Photosynth Res* **85**: 307–317.
- Mock T, Kroon BMA. (2002a). Photosynthetic energy conversion under extreme conditions I: important role of lipids as structural modulators and energy sink under N-limited growth in Antarctic sea ice diatoms. *Phytochemistry* **61**: 41–51.
- Mock T, Kroon BMA. (2002b). Photosynthetic energy conversion under extreme conditions—II: the significance of lipids under light limited growth in Antarctic sea ice diatoms. *Phytochemistry* **61**: 53–60.
- Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K et al. (2008). Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proc Natl Acad Sci USA* **105**: 1579–1584.
- Mock T, Thomas DN. (2008). Microalgae in polar regions: linking functional genomics and physiology with environmental conditions. In: Margesin R, Al E (eds). *Psychrophiles: from Biodiversity to Biotechnology*. Springer-Verlag: Berlin, Heidelberg, pp 285–312.
- Montes-Hugo M, Doney SC, Ducklow HW, Fraser W, Martinson D, Stammerjohn SE et al. (2009). Recent changes in phytoplankton communities associated with rapid regional climate change along the western Antarctic Peninsula. *Science* **323**: 1470–1473.
- Mura MP, Satta MP, Agusti S. (1995). Water-mass influences on summer Antarctic phytoplankton biomass and community structure. *Polar Biol* **15**: 15–20.
- Murphy EJ, Hofmann EE, Watkins JL, Johnston NM, Piñones A, Ballerini T et al. (2012). Comparison of the structure and function of Southern Ocean regional ecosystems: The Antarctic Peninsula and South Georgia. *J Mar Syst* **109–110**: 22–42.
- Nishida I, Murata N. (1996). Chilling sensitivity in plants and cyanobacteria: The crucial contribution of membrane lipids. *Annu Rev Plant Physiol Plant Mol Biol* **47**: 541–568.
- Nomura M, Gourse R, Baughman G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53**: 75–117.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**: 29–34.
- Paradis E, Claude J, Strimmer K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**: 289–290.
- Parker MS, Armbrust EV. (2005). Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J Phycol* **41**: 1142–1153.
- Phadtare S. (2004). Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol* **6**: 125–136.
- Pohnert G. (2002). Phospholipase A2 activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. *Plant Physiol* **129**: 103–111.
- Poulsen N, Kröger N. (2005). A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS J* **272**: 3413–3423.
- Reinfelder JR. (2011). Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Ann Rev Mar Sci* **3**: 291–315.
- Roessler PG. (1988). Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency. *Arch Biochem Biophys* **267**: 521–528.
- Sañudo-Wilhelmy S, Olsen K, Scelfo J, Foster T, Flegal A. (2002). Trace metal distributions off the Antarctic Peninsula in the Weddell Sea. *Mar Chem* **77**: 157–170.
- Schmieder R, Lim YW, Rohwer F, Edwards R. (2010). TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**: 341.
- Sekiguchi H, Kawachi M, Nakayama T, Inouye I. (2003). A taxonomic re-evaluation of the Pedinellales (Dictyochophyceae), based on morphological, behavioural and molecular data. *Phycologia* **42**: 165–182.
- Slamovits CH, Okamoto N, Burri L, James ER, Keeling PJ. (2011). A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat Commun* **2**: 183.
- Smetacek V, Assmy P, Henjes J. (2004). The role of grazing in structuring Southern Ocean pelagic ecosystems and biogeochemical cycles. *Antarct Sci* **16**: 541–558.
- Stekel DJ. (2000). The comparison of gene expression from multiple cDNA libraries. *Genome Res* **10**: 2055–2061.
- Storch D, Pörtner HO. (2003). The protein synthesis machinery operates at the same expense in eurythermal and cold stenothermal pectinids. *Physiol Biochem Zool* **76**: 28–40.
- Takabayashi M, Wilkerson FP, Robertson D. (2005). Response of glutamine synthetase gene transcription and enzyme activity to external nitrogen sources in the diatom *Skeletonema costatum* (Bacillariophyceae). *J Phycol* **41**: 84–94.
- Thamatrakoln K, Hildebrand M. (2007). Analysis of *Thalassiosira pseudonana* silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle. *Eukaryot Cell* **6**: 271–279.
- Thamatrakoln K, Korenovska O, Niheu AK, Bidle KD. (2012). Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom *Thalassiosira pseudonana*. *Environ Microbiol* **14**: 67–81.
- Thomas DN, Dieckmann GS. (2002). Biogeochemistry of Antarctic sea ice. *Oceanogr Mar Biol an Annu Rev* **40**: 143–170.
- Thomsen HA. (1988). Ultrastructural studies of the flagellate and cyst stages of *Pseudopedinella tricostata* (Pedinellales, Chrysophyceae). *Br Phycol J* **23**: 1–16.
- Toseland A, Daines S, Clark J, Kirkham A, Strauss J, Uhlig C et al. (2013). The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nat Clim Chang* **3**: 979–984.
- Tovar-Sánchez A, Duarte CM, Alonso JC, Lacorte S, Tauler R, Galbán-Malagón C. (2010). Impacts of metals and nutrients released from melting multiyear Arctic sea ice. *J Geophys Res: Oceans (1978–2012)* **115**: C7.
- Tréguer P, Jacques G. (1992). Dynamics of nutrients and phytoplankton, and fluxes of carbon, nitrogen and silicon in the Antarctic Ocean. *Polar Biol* **12**: 149–162.
- Vardi A, Bidle KD, Kwityn C, Hirsh DJ, Thompson SM, Callow JA et al. (2008). A diatom gene regulating nitric-

- oxide signaling and susceptibility to diatom-derived aldehydes. *Curr Biol* **18**: 895–899.
- Varela M, Fernandez E, Serret P. (2002). Size-fractionated phytoplankton biomass and primary production in the Gerlache and south Bransfield Straits (Antarctic Peninsula) in Austral summer 1995 – 1996. *Deep Sea Reserch Part II* **49**: 749–768.
- Vartanian M, Desclés J, Quinet M, Douady S, Lopez PJ. (2009). Plasticity and robustness of pattern formation in the model diatom *Phaeodactylum tricornutum*. *New Phytol* **182**: 429–442.
- Wada H, Combos Z, Murata N. (1990). Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. *Nature* **347**: 200–203.
- Wada T, Shimono K, Kikukawa T, Hato M, Shinya N, Kim SY *et al*. (2011). Crystal Structure of the Eukaryotic Light-Driven Proton-Pumping Rhodopsin, *Acetabularia* Rhodopsin II, from Marine Alga. *J Mol Biol* **411**: 986–998.
- Waschuk S a, Bezerra AG, Shi L, Brown LS. (2005). Leptosphaeria rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote. *Proc Natl Acad Sci USA* **102**: 6879–6883.
- Wingler A, Lea PJ, Quick WP, Leegood RC. (2000). Photorespiration: metabolic pathways and their role in stress protection. *Philos Trans R Soc Lond B Biol Sci* **355**: 1517–1529.
- Zhu S, Green BR. (2008). Light-Harvesting and Photoprotection in Diatoms: Identification and Expression of L818-Like Proteins. In: Allen JF, Gantt E, Golbeck JH, Osmond B (eds). *Photosynthesis Energy from the Sun*. Springer: Netherlands, pp 261–264.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)