

Extensive Gene Acquisition in the Extremely Psychrophilic Bacterial Species *Psychroflexus torquis* and the Link to Sea-Ice Ecosystem Specialism

Shi Feng¹, Shane M. Powell¹, Richard Wilson², and John P. Bowman^{1,*}

¹Food Safety Centre, Tasmanian Institute of Agriculture, University of Tasmania, Australia

²Central Science Laboratory, University of Tasmania, Australia

*Corresponding author: E-mail: john.bowman@utas.edu.au.

Accepted: December 20, 2013

Data deposition: This project has been deposited at NCBI database under the accession CP003879 (*Psychroflexus torquis*) and APLF00000000 (*Psychroflexus gondwanensis*).

Abstract

Sea ice is a highly dynamic and productive environment that includes a diverse array of psychrophilic prokaryotic and eukaryotic taxa distinct from the underlying water column. Because sea ice has only been extensive on Earth since the mid-Eocene, it has been hypothesized that bacteria highly adapted to inhabit sea ice have traits that have been acquired through horizontal gene transfer (HGT). Here we compared the genomes of the psychrophilic bacterium *Psychroflexus torquis* ATCC 700755^T, associated with both Antarctic and Arctic sea ice, and its closely related nonpsychrophilic sister species, *P. gondwanensis* ACAM 44^T. Results show that HGT has occurred much more extensively in *P. torquis* in comparison to *P. gondwanensis*. Genetic features that can be linked to the psychrophilic and sea ice-specific lifestyle of *P. torquis* include genes for exopolysaccharide (EPS) and polyunsaturated fatty acid (PUFA) biosynthesis, numerous specific modes of nutrient acquisition, and proteins putatively associated with ice-binding, light-sensing (bacteriophytochromes), and programmed cell death (metacaspases). Proteomic analysis showed that several genes associated with these traits are highly translated, especially those involved with EPS and PUFA production. Because most of the genes relating to the ability of *P. torquis* to dwell in sea-ice ecosystems occur on genomic islands that are absent in closely related *P. gondwanensis*, its adaptation to the sea-ice environment appears driven mainly by HGT. The genomic islands are rich in pseudo-genes, insertional elements, and addiction modules, suggesting that gene acquisition is being followed by a process of genome reduction potentially indicative of evolving ecosystem specialism.

Key words: psychrophilic bacteria, horizontal gene transfer, proteomics, sea ice, ecosystem specialism.

Introduction

Sea ice is a major feature of the surface of high-latitude oceans. It is relatively biologically productive due to extensive blooms of sea-ice algae, embedded in the ice floes as a band of growth or associated with the basal section of the ice floe that contacts the underlying seawater. Sea-ice algae and their epiphytic bacteria form the foundation of an active microbial loop comprised taxa distinct from the underlying seawater (Bowman et al. 1997; Brown and Bowman 2001; Brinkmeyer et al. 2003; Bowman et al. 2012). Tightly coupled to algal-driven primary production, sympagic bacterial populations increase 1–2 weeks after the phytoplankton bloom peaks in late summer and become increasingly dominant

when solar irradiance levels decline and algae subsequently senesce and/or become dormant (Kottmeier et al. 1987; Kottmeier and Sullivan 1987; Grossman and Dieckmann 1994; McMinn and Martin 2013). As sea ice forms, brine is ejected from the ice crystal matrix and collects in cracks and channels (referred to as brine channels), making up 5–20% of the sea-ice volume. Though a very cold and saline environment, sea-ice microbial communities (SIMCO) thrive in sea-ice brines at temperatures of -10°C and at salinities three or more times the concentration of seawater (Thomas and Dieckmann 2002).

The extent of most sea ice changes by more than an order of magnitude between summer and winter. Long term, stable

ice tends to only occur connected to land at high latitudes. Recent trends in Arctic Ocean sea ice decline and a simultaneous increase in Antarctic sea ice suggests that climate change may have an impact on sea ice-associated taxa, though the extent of this impact is far difficult to predict (Berge et al. 2012). Polar sea ice is geologically modern. Based on detection of sea-ice diatom fossils and geological signatures indicating iceberg rafting, ice formation in high latitude oceans has been extensive since the mid-late Eocene ~35–47 Ma. However, multiyear ice that would act as a more stable sympagic platform than seasonal sea ice may not have appeared until as late as the Pliocene or Pleistocene (2.5–3 Ma) with the advent of sustained polar glaciation (Polyak 2010). In that time, microbial life associated with sea ice may have had the opportunity to specialize. Currently, our knowledge of the diversity of microbial sea-ice communities and their obligate sympagy remains limited (Bluhm et al. 2011; Poulin et al. 2011).

The genome sequence of the psychrophilic marine species *Colwellia psychrelythraea* (strain 34H) provided the first genome-based perspective on the traits that allow not only for psychrophilic growth but also the possible means to grow and persist in sympagic ecosystems (Methé et al. 2005). The main traits examined included amino acid composition of proteins and their relation to tertiary structure, secreted and nonsecreted cold-active enzymes, omega-3 polyunsaturated fatty acids (PUFA), compatible solute synthesis, and secreted exopolysaccharides. Ice-active proteins that act to modify ice crystal structure have also been studied (Raymond et al. 2007; Bayer-Giraldi et al. 2010). The important sea ice-dwelling diatom *Fragilariopsis cylindrus* has several ice-active proteins orthologous to generally uncharacterized proteins in cold-adapted bacteria, suggesting that interdomain horizontal gene transfer (HGT) of these proteins may have occurred (Bayer-Giraldi et al. 2010). This raises the question of whether other traits allowing inhabitation and successful competition in sea ice have also been acquired by HGT processes.

In this study, we investigated the genomic properties of the extremely psychrophilic bacterial species *P. torquis*, an unusual member of the family Flavobacteriaceae (phylum Bacteroidetes) that has several traits linked to sea-ice inhabitation and dependence on algae via epiphytism. *Psychroflexus torquis* was originally isolated from algal assemblages in Antarctic multiyear sea ice. It differs from all other related species, including its closest relative *P. gondwanensis*, in being filamentous at an early stage of growth, extremely psychrophilic, able to synthesize omega-3 and omega-6 PUFA, and prolifically secreting soluble exopolysaccharides (EPS) (Bowman et al. 1998). The species, though chemoorganotrophic, can also harness energy from light via proteorhodopsin-driven proton pumping, a feature enhanced under osmotic stress (Feng et al. 2013). The genus *Psychroflexus* is found within moderately hypersaline ecosystems across the world; however, the combined traits of psychrophily and

PUFA synthesis in *P. torquis* make this species stand out among other members of phylum Bacteroidetes. To explore these and other ecologically relevant genomic aspects of *P. torquis* that may provide insight into the relatively recent evolution of psychrophily, we compared the genome of the type strain ATCC 700755^T to that of its closest relative *P. gondwanensis* ACAM 44^T. To better discern important genes, we also performed comprehensive proteomics on ATCC 700755^T grown under a range of conditions. In particular, we searched for mobile genetic elements and pseudogenes as evidence for HGT and its possible role in sea-ice ecosystem specialism.

Materials and Methods

Genome Sequence Determination

Psychroflexus torquis ATCC 700755^T (T = type strain) and *P. gondwanensis* ACAM 44^T (ATCC 51278^T) were cultivated on modified marine agar (0.5% w/v proteose peptone, 0.2% w/v yeast extract, 1.5% w/v agar, and 3.5% w/v sea salts) at 4 and 25 °C, respectively. High molecular weight DNA was extracted and purified from biomass using the Marmur method. DNA was sequenced using the 454 GS-FLX/Plus (454 Life Sciences, Branford, CT.) platform following the manufacturer's de novo sequencing protocol. For ATCC 700755^T and ACAM 44^T 146.5 and 149.1 Mb of sequence data (430–440 bp average length) was assembled using Newbler v. 2.6 (454 Life Sciences). The Sanger sequence draft already available for ATCC 700755^T (generated through the Gordon and Betty Moore Foundation Genome Sequencing Project at the J. Craig Venter Institute) was compared with the pyrosequenced contigs using Artemis (Carver et al. 2012) with the number of contigs reduced from 39 to 9. Gaps between contigs were closed using polymerase chain reaction (PCR) analysis. Apparent misassemblies and regions with sequence inconsistencies were corrected in Artemis after PCR and sequencing confirmation of the regions.

Postsequence Analysis

Gene annotation for the complete ATCC 700755^T sequence was carried out in Artemis and also compared with annotations generated via the Prodigal server (Hyatt et al. 2010) and Glimmer v. 3.02 (Delcher et al. 1999). Transfer RNAs were predicted using tRNAscan-SE (Lowe and Eddy 1997). Predicted CDSs were compared against the National Center for Biotechnology information (NCBI) database. Annotation utilized the NCBI Prokaryotic Genomes Automatic Annotation Pipeline. Pseudogenes in both genomes were defined as described by Lerat and Ochman (2005) based on comparisons with highly similar sequences in ATCC 700755^T, ACAM 44^T and with highly similar orthologs in related taxa. The presence of protein signal peptides and transmembrane helices were predicted using the SignalP 4.1

(Petersen et al. 2011) and THMM v. 2.0 servers (Centre for Biological Sequence Analysis, Technical University of Denmark), respectively. The genome of ATCC 700755^T was visualized using DNAPlotter (Carver et al. 2009). CRISPR palindromic repeats were detected using CRISPR recognition tool (CRT) (Bland et al. 2007).

Protein Extraction and Posttreatment

Psychroflexus torquis ATCC 700755^T was grown on modified marine agar at different sea salt salinities and light intensities at 4 °C to examine the broadest possible set of proteins produced by *P. torquis*. The different salinities were achieved by adding 17.5, 35, 52.5, and 70 g/l of sea salt (Red Sea) to the marine agar and three levels of illumination (0, 3–4, 20–30 $\mu\text{mol photon s}^{-1} \text{m}^{-2}$) were used. Cells were lysed in 1 ml 50 mM Tris–HCl buffer (pH 7.0) by sonication in an ice bath, with 10 s of sonication with a 10 s wait period, cycled 15 times until the opaque cell suspension became translucent. The suspensions were then centrifuged at 16,000 $\times g$ for 25 min at 4 °C. The supernatant protein was precipitated using trichloroacetic acid and the protein pellets were then treated with 0.2 M NaOH to improve subsequent solubilization (Nandakumar et al. 2003). The pellets were then solubilized using 100 μl 50 mM Tris–HCl buffer (pH 7.0), and protein concentration was determined using the Bradford assay (Bio-Rad). Volumes of samples containing 50 μg of soluble protein extract were then reduced in a solution of 50 mM dithiothreitol, 100 mM ammonium bicarbonate for 1 h at room temperature. The samples were then alkylated with 200 mM iodoacetamide in 100 mM ammonium bicarbonate for 1 h at room temperature. After alkylation, the reduction of proteins was repeated, and they were digested in a buffer (50 mM ammonium bicarbonate, 1 mM calcium chloride) that contained sequencing grade modified trypsin (Promega), at a sample protein to trypsin ratio of 25:1, at 37 °C with gentle shaking overnight. Digestion was stopped by acidification with 10 μl of 10% (v/v) formic acid. The samples were then centrifuged for 5 min at 14,000 $\times g$ to remove any insoluble material, and an aliquot (100–200 μl) of peptides was transferred to high-performance liquid chromatography (HPLC) vials for mass spectrometric analysis. Samples were prepared with three biological replicates, and for each biological replicate, two technical replicates were performed.

NanoLC-Orbitrap Tandem Mass Spectrometry

The separation of peptides utilized a Surveyor Plus HPLC system fitted in line with an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific). Aliquots of peptide samples were loaded at 0.05 ml/min onto a C₁₈ capillary trapping column (Peptide CapTrap, Michrom BioResources) controlled by an Alliance 2690 separations module (Waters). Peptides were then separated on an analytical HPLC column packed with 5- μm C18 media (PicoFrit Column, 15 μm i.d. pulled tip,

10 cm, New Objective) using four linear gradient segments controlled using a Surveyor MS Pump Plus (ThermoFisher Scientific) at 200 nl/min. The solvent series included initially 5% acetonitrile in 0.2% formic acid (solvent A), shifting up to 90% acetonitrile in 0.2% formic acid (solvent B). The four-stage process where solvent B gradually replaced solvent A comprised: 0–10% solvent B over 7.5 min; 10–25% solvent B over 50 min, 25–55% B over 20 min, and then 55–100% solvent B over 5 min. This process was followed by reequilibration of the column with solvent A for 15 min. The LTQ-Orbitrap XL was controlled using Xcalibur 2.0 software (ThermoFisher Scientific) and operated in a data-dependent acquisition mode whereby the survey scan was acquired in the Orbitrap with a resolving power set to 60,000 (at 400 m/z). MS/MS spectra were concurrently acquired in the LTQ mass analyser on the seven most intense ions from the Fourier Transform (FT) survey scan. Charge state filtering, where unassigned and singly charged precursor ions were not selected for fragmentation, and dynamic exclusion (repeat count, 1; repeat duration, 30 s; exclusion list size, 500) were used. Fragmentation conditions in the LTQ were 35% normalized collision energy, activation q of 0.25, 30 ms activation time, and minimum ion selection intensity of 500 counts.

Protein Identification and Bioinformatics

The acquired MS/MS data were converted to .mzXML peak list files from .RAW files using the msConvert command in Proteowizard. The MS/MS data were searched against the proteome of *P. torquis* ATCC 700755^T (NCBI accession code CP003879) employing X!Tandem running in the open-source Computational Proteomics Analysis System environment (Rauch et al. 2006). For searches, parent ion tolerance of 20 ppm and fragment ion mass tolerance of 0.5 Da were used and enzyme cleavage was set to trypsin, allowing for a maximum of two missed cleavages. Amino acid residue alterations were also accounted for including S-carboxamidomethylation of cysteine residues specified as a fixed modification, and cyclization of N-terminal glutamine to pyrroglutamic acid, deamidation of asparagine, hydroxylation of proline, and oxidation of methionine specified as variable modifications. Protein identifications were assessed by the Peptide Prophet and Protein Prophet algorithms (Nesvizhskii et al. 2003). Protein identifications were filtered by assigning a Protein Prophet probability >0.7. This filtration constrains the protein false discovery rate to <1%. Spectral counting was used to determine relative protein abundance, taking into account the number of amino acid residues (Liu et al. 2004). To acquire the maximum coverage of proteins, spectra obtained from all samples were pooled together in this study. Filtration removed 6.4% peptides with a total of 743,361 peptide spectra matched to 2,598 protein identification, with 1,936 proteins possessing two or more unique peptides.

Results and Discussion

A Highly Conserved Core Genome Is Shared by *P. torquis* and *P. gondwanensis*

Based on a meta-analysis of the latest metagenome and NCBI database records, *P. torquis* is restricted to sea ice or seawater around ice floes and has a bipolar distribution, occurring in both the Arctic and the Antarctic (fig. 1). *Psychroflexus gondwanensis*, the closest cultivated relative of *P. torquis*, so far is only known to reside in Antarctic hypersaline lakes, where it can be a dominant member of the microbial community (Yau et al. 2013). An equidistant lineage detected in the ice cover of Lake Vida, Antarctica (Mosier et al. 2007), and in salinated Yellow River delta soil from China, is so far uncultured but suggests a broader distribution of closely related *Psychroflexus* genotypes. The 4.32 Mb genome of the *P. torquis* type strain ATCC 700755^T was obtained in two stages, first as an 8× coverage Sanger-sequenced draft, and then closed in this study by a combination of pyrosequencing, gap filling, and subsequent PCR-based checks of potentially misassembled

regions. The 3.32 Mb draft genome of *P. gondwanensis* type strain ACAM 44^T (NCBI accession code APLF00000000) was obtained at a coverage level of 44-fold. Details for the genomes are summarized in table 1.

The 16S rRNA gene sequence similarity between *P. torquis* and *P. gondwanensis* strains and clones averages 99.0%, higher than the empirical 98.5% similarity cut-off that has often been used as a preliminary evidence for defining distinct prokaryotic species (Stackebrandt et al. 2006). However, a full genome comparison between the two genomes yielded an average nucleotide similarity (ANI; Goris et al. 2007) of only 68%, consistent with an overall low DNA:DNA hybridization level of <20% (Bowman et al. 1998). Despite the differences in size and ANI score, both strains share 2,308 genes out of an effective pan genome of 4,225 protein-coding genes. The genes that are in common have a high mean nucleotide similarity (92±4%) and extensive stretches of synteny (supplementary table S1, Supplementary Material online).

A dissection of genes by functional classification demonstrates that the conserved gene overlap includes virtually all

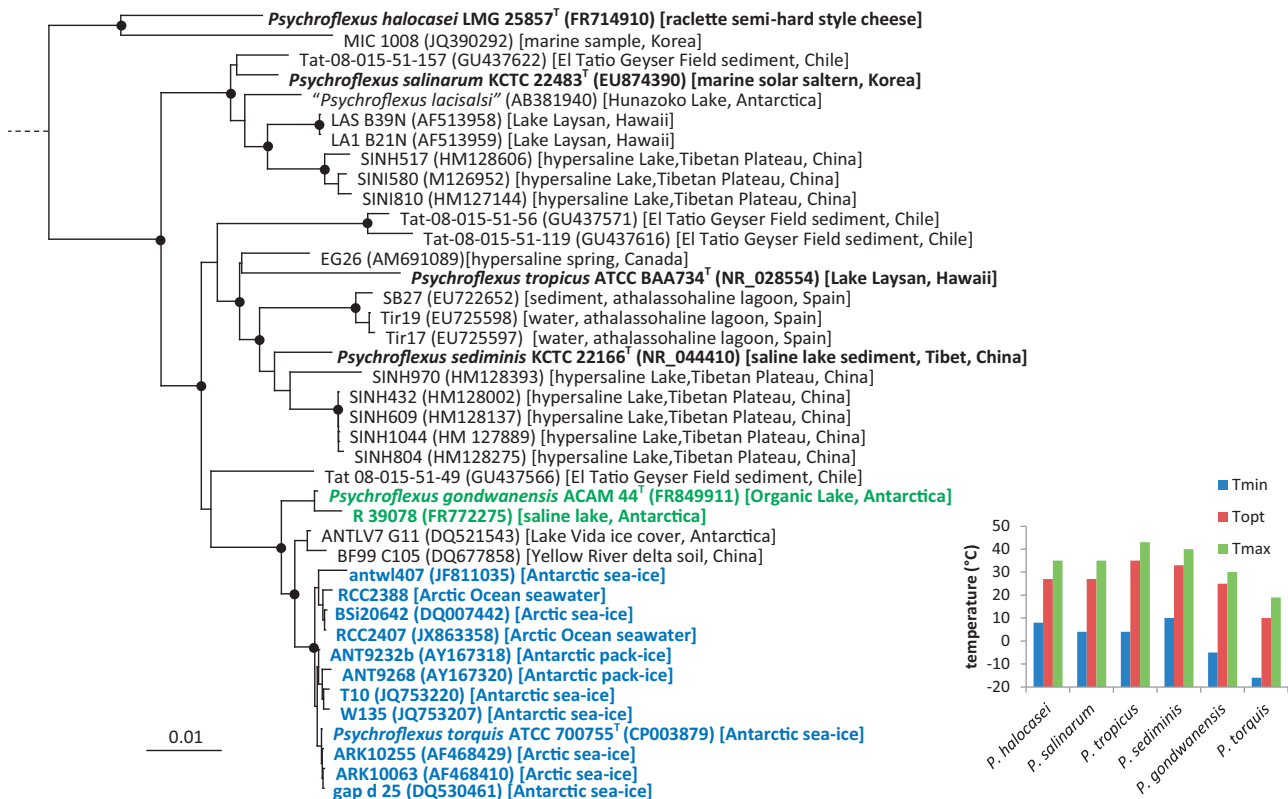


Fig. 1.—The 16S rRNA gene-based phylogenetic tree of the genus *Psychroflexus* (family Flavobacteriaceae, phylum Bacteroidetes). The tree was constructed with complete or near-complete sequences, aligned with ClustalW, clustered with the maximum likelihood algorithm, and created using Neighbor-Joining. Black circles indicate that bootstrapping support for nodes is >80%. The outgroup used was *Capnocytophaga ochracea*. The bar indicates relative sequence distance. NCBI accession codes are given in parentheses followed by location of isolation. For type strains of described species, cardinal temperatures for growth rate are indicated in the inset bar graph including values for temperature: minimum (T_{min}), optimal (T_{opt}), and maximum (T_{max}) temperature values. Values were determined in liquid media, with the minimum temperature a theoretical value calculated from the square root model of Ratkowsky et al. (1983).

Table 1

Genome Data for *P. torquis* ATCC 700755^T and *P. gondwanensis* ACAM 44^T

Species	<i>P. torquis</i>	<i>P. gondwanensis</i> ^a
Strain	ATCC 700755 ^T (= ACAM 623 ^T)	ACAM 44 ^T (=ATCC 52178 ^T , DSM 5423 ^T)
Taxonomic hierarchy	Flavobacteriaceae, Flavobacteriales, Flavobacteria, and Bacteroidetes	
Genome status	Finished	Noncontiguous draft
Platform	Sanger (JCVI), 454 GX FLX	454 GS FLX
Coverage	8×, 33×	44× (62 contigs)
Number of replicons	1	1
Extrachromosomal elements	0	0
GenBank ID	CP003879, NC_018721	APLF00000000
Genome size (bp)	4,321,832	3,325,075
DNA coding region (bp)	3,503,343 (81.06%)	2,852,146 (85.77%)
DNA G + C content (bp)	34.51% (34.9/33.5)	35.72% (36.0/34.4)
Total genes	3,951	3,007
RNA genes	45 (1.14%)	47 (1.61%)
rRNA operons	3	3
Protein coding genes	3,526	2,912
Pseudogenes	379 (9.57%)	48 (1.60%)
Genes with a predicted function	2,278 (64.58%) ^a	2,006 (68.32%)
Proteins with signal peptides or POR secretion system sorting domains	380 (10.78%)	299 (10.18%)
Proteins with transmembrane domains	761 (21.58%)	658 (22.41%)

^aEstimates based on available noncontiguous sequence data.

RNA-coding genes and genes involved in fundamental cellular processes, including DNA-related processes, protein translation, ribosome structure and biogenesis, tRNA processing, protein secretion, cytokinesis, nucleic acid transport/metabolism, cofactor transport/metabolism, and gliding motility (fig. 2). Many of these genes have similarly colocated and syntenic arrangements among related taxa within the family Flavobacteriaceae, demonstrating an apparent ancestral nature. Other functional classes more moderately conserved between the two species include metabolism, protein modification, folding, and turnover, and defence/detoxification systems. Overall, this pattern of gene sharing is consistent with both *P. torquis* and *P. gondwanensis* being inhabitants of cold saline ecosystems and having broadly similar though not identical morphological and metabolic phenotypes (Bowman et al. 1998).

Evidence of Extensive HGT in the Species-Dependent Section of the *P. torquis* ATCC 700755^T Genome

Large tracts of the ATCC 700755^T and ACAM 44^T genomes share no significant nucleotide similarity with 1,265 and 653 strain distinct genes, respectively. The structural and functional categorization of the species-dependent genome regions reveal that in general, *P. torquis* ATCC 700755^T has a wide variety of functional genes in this set compared with *P. gondwanensis* ACAM 44^T (fig. 2, supplementary tables S1 and S2, Supplementary Material online). When directly compared against the ACAM 44^T genome (fig. 3), these regions occur in distinct genomic islands (GIs), of which 44 could be defined

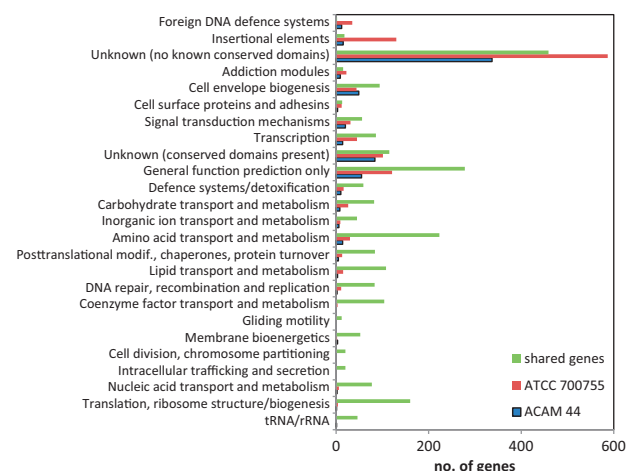


Fig. 2.—Numbers of genes shared or strain dependent in *P. torquis* ATCC 700755^T and *P. gondwanensis* ACAM 44^T organized by functional class.

(size range 3.7–153.2 kb, supplementary table S1, Supplementary Material online). Fourteen of these GIs are flanked by tRNA genes, and the majority are characterized by low gene density (average 74.5%) and are relatively rich in pseudogene, insertional element, and/or addiction modules. The presence of tRNA genes flanking GIs is consistent with tRNA being a well-known hotspot for site-specific recombination (Ou et al. 2006). One tRNA gene (tRNA-Met between P700755_00961/00963) has been disrupted in ATCC

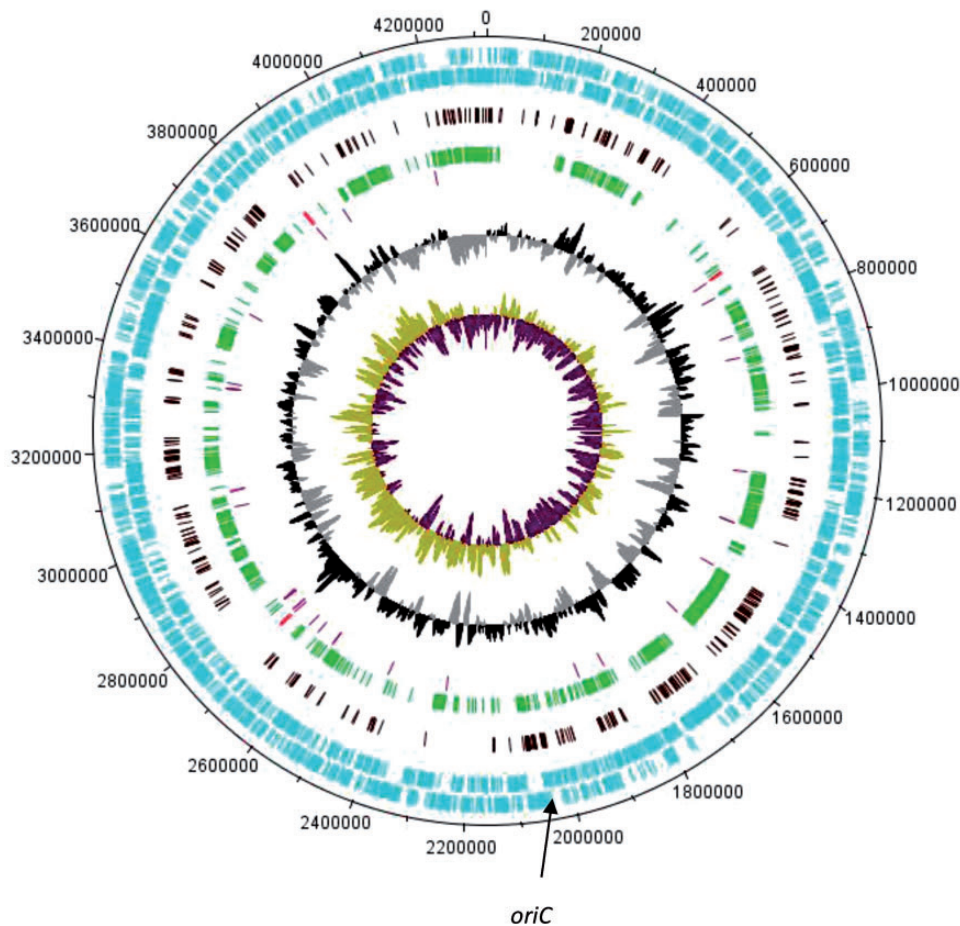


Fig. 3.—Genome map of *P. torquis* ATCC 700755^T drawn using DNAPlotter (Carver et al. 2009). The first and second rings (blue) show gene annotations for the sense and antisense strands, respectively. The third ring (brown) shows the position of annotated pseudogenes. The fourth ring (green) shows genes that occur in *P. torquis* ATCC 700755^T but not *P. gondwanensis* ACAM 44^T. On this ring (in red) are the positions of the three rRNA operons, which are located in syntenic regions of the ACAM 44^T genome. The fifth ring (purple) shows the position of tRNA coding regions. The sixth ring demonstrates GC-bias (black, positive; gray, negative) across the genome calculated from 1,000-bp segments. The inner ring shows GC-skew with the leading strand commencing at the predicted *oriC*.

700755^T by an integrase, suggesting other tRNAs may have also been lost by GI insertions. An example of such may include a tRNA-Ala-GGC present as a pair in ACAM 44^T but only singly in ATCC 700755^T (between P700755_01366/01368), located directly adjacent to a 153.2 kb GI (GI no. 17; [supplementary table S1, Supplementary Material online](#)).

It is suggested that the variation in G + C content can be an important evidence for HGT in regions of a genome (Ochman et al. 2000). Thus, we plotted the overall GC% of *P. torquis* using DNAMAN. The overall GC% of *P. torquis* is 34.5%, and we were able to locate several regions that deviated from this mean ranging between 26% and 50%, with 15 GIs located within these areas ([supplementary fig. S1, Supplementary Material online](#)) suggestive of an external origin and ancestry of these genomic regions. Furthermore, comparisons of GI-associated genes with the NCBI database were made, and the

highest match was noted on the basis of amino acid similarity. A high proportion of genes (27%) did not match anything on the NCBI database, while 45% of matches occurred with proteins from members of the family Flavobacteriaceae that mainly occur in marine, especially polar, ecosystems (i.e., *Polaribacter* spp., *Cellulophaga algicola*, *Gillisia* spp.). The remainder of matches were with a wide variety of bacteria within the phylum Bacteroidetes and in other phyla ([supplementary table S1, Supplementary Material online](#)). A similar pattern was observed for ACAM 44^T, though more muted in terms of the sheer scale of gene acquisition. Sea ice has been proposed as a hotspot for genetic recombination due to its high density of bacteriophage, a result of the concentration of brines during sea-ice formation (Wells and Deming 2006). Despite no genes of definitive phage origin being found on the ATCC 700755^T genome, extensive phage defence

systems were detected, including four large regions containing 7–28 CRISPR repeats, with the main concentration located immediately after a cluster of clustered regularly interspaced short palindromic repeats (CRISPR) genes (P700755_00291 to 00295). In addition, six restriction–modification systems and a variety of proteins in the phage infection resistance family were detected (supplementary table S1, Supplementary Material online). Approximately 4% of genes were insertion sequence (IS) elements belonging to 40 families of retron-associated reverse transcriptase, integrase, and transposase proteins. Ten families of addiction (toxin/antitoxin) modules were also present. These gene types and their distribution suggest that mobile genetic elements and, potentially, phages were involved in building GI-associated genomic content. Overall, the data suggest that there has been a high level of gene acquisition in *P. torquis* ATCC 700755^T. Based on the collective nature of the GIs as explained above, we hypothesize that HGT processes drove this acquisition. Because direct evidence such as prophage and conjugative systems are no longer evident on the genome it would be ideal to collect data from a wider range of strains as well as other sea-ice-derived bacteria to assess the degree of HGT and whether there is a pool of shared genetic homologs within SIMCO.

Modern Evolution of Psychrophily in *P. torquis* ATCC 700755^T and Its Relation to Genome Sequence-Derived Criteria

Psychrophilic prokaryotes unlike thermophilic taxa are neither phylogenetically deep-branching nor tend to cluster together. In terms of phylogenetic relationships, psychrophilic strains tend to occur almost without exception in genera with higher temperature adapted relatives and are usually located at the tips of branches. This aspect of psychrophilic prokaryote phylogeny was first noted by Franzmann (Franzmann 1996). Though it is conceivable that various cold-adapted microbes evolved during earlier cold times on Earth, for example during the purported Cryogenian period (MacDonald et al. 2010) and Snowball Earth periods (Hoffman and Schrag 2002), psychrophily amongst prokaryotes is a modern phenomenon based on the temperature history of Earth and underlying prokaryotic evolution (Schwartzman and Lineweaver 2005). As seen in figure 1 the deeper branching *Psychroflexus* species are classic mesophiles and derive from widespread locations, such as geyser field soils, marine sites, saline lakes, and even cheese. The tip position of *P. torquis* in the *Psychroflexus* phylogenetic tree (fig. 1) and its apparent restriction to polar marine locations, specifically multiyear sea ice, fits the concept of a modern evolution of psychrophily. The cardinal growth temperatures of *P. gondwanensis* and *P. torquis* differ by more than 10°C and this value is even greater for other *Psychroflexus* spp (fig. 1).

Typically amino acid composition in general terms is correlated to growth temperature preference, but in the case

of the strains studied here overall amino acid content did not vary significantly. Specific amino acids Ile, Val, Tyr, Trp, Arg, Glu, Leu (IVYWREL) have been found to be strong determinants of thermostability (Zeldovich et al. 2007); however, predictions based on the levels of these amino acids for ATCC 700755^T and ACAM 44^T overestimated their optimal growth temperatures as 29 and 34°C, respectively. This result is not surprising in that predictions of thermoadaptation are relatively insensitive at the mesophilic/psychrophilic end of the biokinetic spectrum, suggesting an inherent limit to thermostability of proteins and thus growth (Corkrey et al. 2012). Codon usage analysis was also performed to determine whether any significant trends occurred between the strains based on highly translated gene products. This analysis was done by comparing the top ten percentile of the most abundant proteins of ATCC 700755^T that had highly similar orthologs in ACAM 44^T ($n=306$) determined from the protein spectral count data set (supplementary table S1, Supplementary Material online). Analysis of the expected codon adaptation index (Puigbò et al. 2008) revealed significant bias, mainly due to differences in synonymous third-base positions being more AT-rich (average % GC3: 27 vs. 35). Overall, this suggests that amino acid content and codon usage criteria can discriminate between the two species examined; however, the trends may not necessarily relate only to psychrophily. A similar situation was found with the extreme psychrophile *Psychromonas ingrahamii*, which did not exhibit unusual codon usage trends (Riley et al. 2008) nor was its level of IVYWREL amino acid content informative. Another explanation for temperature sensitivity is that *P. torquis* is inherently unstable at mesophilic temperatures and that, because proteins do not seem obviously involved, it is possible that another part of the cell is the temperature “Achilles heel.” A logical candidate is the cell membrane of ATCC 700755^T, which is quite different from that of ACAM 44^T as it is rich in PUFA and anteiso-branched fatty acids (Bowman et al. 1998). This combination creates membrane fluidity compatible with very low temperature even though it has the disadvantage in being potentially more thermolabile (D’Aoust and Kushner 1971).

Evidence of Significant Gene Decay in the *P. torquis* ATCC 700755^T Genome

The genome of ATCC 700755^T has an unusually large number of pseudogenes ($n=379$), making up 9.5% of the total number of protein coding genes. This percentage represents a conservative 8-fold increase over that of the ACAM 44^T genome for which pseudogene numbers have likely been overestimated due to its draft status. The appearance of pseudogenes is believed to be associated with recent mutational processes because they seem to be rapidly deleted from genomes (Kuo and Ochman 2010). In ATCC

700755^T, the pseudogenes primarily occurred as truncated fragments or segmented open reading frames (ORFs) due to one or more nonsense mutations and/or indels, while more rarely, pseudogenes were derived from direct transposon or retron disruptions. In the case of the more overtly truncated ORFs, most have been affected by subsequent frameshifts and partial deletion because the coding region remnants were usually less than 40% of the full length version, with the remaining degenerate region sometimes still adjacent to the pseudogene. Indeed, numerous examples of fragmentary relics lacking both stop and start codons were detectable in intergenic regions. Insertional (IN) element types are very diverse in *P. torquis* ATCC 700755^T, with a high proportion being pseudogenes (supplementary table S3, Supplementary Material online). Given that many IS elements and other mobile genetic elements are concentrated in GIs, insertion and recombination appears to have shaped the genome of ATCC 700755^T extensively. Such high proportions of pseudogenes essentially indicate a process of both gene decay and adaptation, as has been observed in bacteria transiting to a lifestyle of obligate parasitism or symbiotism (Burke and Moran 2011).

The large number of pseudogenes in ATCC 700755^T relative to ACAM 44^T suggest that HGT gene acquisition may have been both advantageous and deleterious. Unnecessary genes, copies of genes involved in the HGT processes themselves, as well as those accidentally disrupted via integration events have been and are being progressively deleted. This process may be due to selection against pseudogenes themselves or selection of processes that actually remove them from genomes (Kuo and Ochman 2010). Other strains will need to be examined to determine whether this pseudogene decay is consistent at a species level and if the “burden” of pseudogenes correlates to fitness. The predicted location of the point of origin of replication (*oriC*), detected using DoriC v. 5.0 (Gao et al. 2012), is of interest in this regard. Surprisingly, two putative *oriC* sites were found, both located near each other within a 31.2 kb GI (no. 22) immediately adjacent to retron-type reverse transcriptases (between P700755_01930/01931; P700755_01949/01953), the latter of which is a pseudogene.

Pseudogenes are generally assumed not to be expressed or translated, although exceptions have been detected (Rusk 2011). Based on our proteomic spectral count data, the vast majority of pseudogenes were not detected after filtration (supplementary table S3, Supplementary Material online). However, some pseudogenes have a substantial number of collated spectral counts that had high confidence of identification. In all cases, these peptides occurred on IS elements that occurred multiple times in the genome either as full length genes or truncated pseudogene versions. It is possible that truncated protein products are still translated but in general appear to have low cellular abundance based on the spectral count data.

Functional Prioritization Suggested by Abundance of Protein Products in *P. torquis* ATCC 700755^T Genome

We assessed to what degree the genes of ATCC 700755^T are translated using proteomic analysis. It is assumed that the more abundant proteins are inherently fundamental to the system biology of ATCC 700755^T. Above an arbitrary threshold set at 0.005% of the filtered and normalized total spectral count, proteins were regarded as being abundant. At this threshold, spectral counts for multiple peptides were observed in most sample replicates, thus suggesting sustained production under the range of growth conditions tested. The proteomic data set is, however, logistically limited due to huge dynamic ranges, with a natural bias against low abundance and transmembrane proteins (Borg et al. 2011). Many proteins belonging to the overlapping proteomes of ATCC 700755^T and ACAM 44^T were readily detected, as expected, as were high proportions of functionally conserved, mainly cytosolic proteins (fig. 4). Among these, the least detected proteins were associated with DNA-related processes, including repair and recombination (only 20% of proteins observed), while the highest proportion (63%) was associated with electron transport. This difference may indicate the prioritization of proteins in terms of cellular processes, where certain functional proteins such as DNA repair are only upregulated to high levels when needed. Other proteins, such as those involved in central metabolism and energy production, are constantly required by the organism. At the other extreme, some functional groups of proteins from the ATCC 700755^T genome were rarely detected, including addition modules, foreign defence proteins, IN elements, and a high proportion

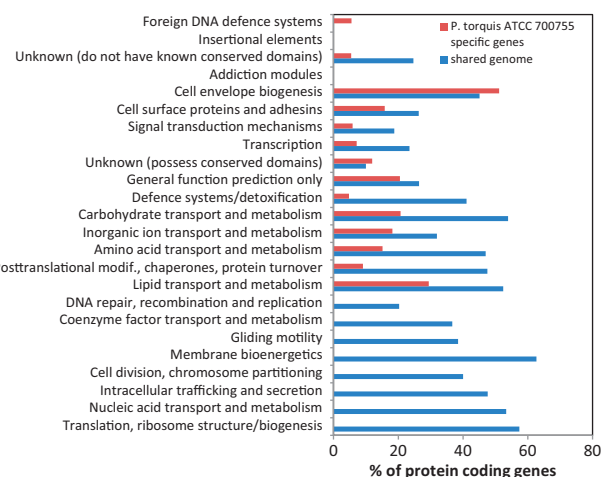


Fig. 4.—Distribution of most abundant proteins of *P. torquis* ATCC 700755^T and whether they are strain-specific or shared with the genome of *P. gondwanensis* ACAM 44^T, grouped by functional classes. Proteomic data were pooled from all treatment samples, an abundant protein was defined as 0.005% of the total spectral counts, and each was detected in most replicates and represented by multiple peptide sequences.

of proteins with seemingly no function at all based on their lack of conserved domains. The weak translation of these latter proteins suggests that their presence could be largely strain dependent (Ou et al. 2005).

Genes That Code Abundant Proteins in *P. torquis* ATCC 700755 Are Potentially Important for Sea-Ice Inhabitation

We assume that species and/or strain-specific features (as summarized in fig. 2 in terms of gene distribution) greatly contribute to the inherent uniqueness of bacterial strains at a phenotypic level and subsequently determine their ecological nature. Features that are strongly expressed likely have important roles in defining this identity. Based on protein abundance analysis, the most prominent products of the unique genome regions of *P. torquis* ATCC 700755^T (fig. 4) are proteins associated with cell envelope biogenesis, cell surface proteins/adhesions, proteins involved in the transport and metabolism of carbohydrates, lipid and inorganic ion transport and metabolism, and those so far with only generalized functions. Many strain-dependent features are likely not well observed due to the inability of laboratory-based growth conditions to adequately capture this information. Given that sea-ice ecosystem conditions are highly complex in terms of resource availability, physicochemical pressures, and biological interactions, many other more transiently expressed proteins potentially play important roles. Nevertheless, these experiments provide a preliminary view of the functionality of ATCC 700755^T in the context of the ecosystem to which it is specialized.

A list of proteins found to be relatively abundant (as defined above) and produced by genes only on GIs (supplementary table S1, Supplementary Material online) summarize as far as possible the potential unique biology of ATCC 700755^T and provide a suite of relevant functions for its sea-ice ecosystem associations. In particular, two GIs (no. 18 and 43) have a large number of moderately to highly abundant proteins while several GIs include no abundant proteins at all. The selective pressure enforced in a sea-ice ecosystem may lead to the retention of some GIs as stable sections of the genome, while others could be eventually lost.

The EPS production by ATCC 700755^T, which includes multiple sulphated, uronic acid-containing, and N-acetylated polysaccharides, is prodigious substantially increasing medium viscosity, with production levels and viscosity increasing with decreasing growth temperature (Mancuso Nichols 2005; Bowman 2008). GI no. 18 includes a 60-kb EPS biosynthesis locus (fig. 5). Though the exact structure of the EPS and its specific functional benefits remain to be elucidated, EPS production has been found to be a common feature of sea-ice bacteria and a crucial factor for sea-ice inhabitation. It provides cryoprotection (Marx et al. 2009), encourages ice crystal modification, retains liquid brine in brine channels, thereby enhancing recruitment into forming sea ice (Ewert and

Deming 2011; Kremps et al. 2011). It also has a role in nutrient acquisition, especially trace elements such as iron and cobalt, because anionic EPS can act as powerful ligands (Hassler et al. 2011). The highly translated level of EPS biosynthesis gene products (fig. 5) suggests that EPS may be crucial for the low temperature growth and activity of ATCC 700755^T even outside of the sea-ice environment. The EPS cluster contains genes that match those of bacterial relatives within and outside of the phylum Bacteroidetes. Intermingled with these genes are intact and remnant transposase genes, as well as an MazEF family addiction module and duplicated genes coding putative UDP-glucose 6-dehydrogenases, suggesting recent acquisition via HGT (fig. 5).

The second region of highly translated proteins is located on GI no. 43 and includes a 42-kb cluster of transporter-like proteins mostly of the protein family referred to as acidobacterial duplicated orphan permeases (ADOP; P700755_03736 to 03759) first observed in the genomes of acidobacteria (Ward et al. 2009). The ADOP proteins include a cluster of ten paralogs associated with other transporter proteins (fig. 6). The ADOPs all contain a MacB-periplasmic domain, suggesting they could be involved in efflux. The relatively high protein abundance and clustered nature of numerous ADOP paralogs and surrounding transporters (fig. 6) is intriguing and suggests that they have an important role in the biology of *P. torquis* ATCC 700755^T. Whether this role is for export of toxic products of metabolism or deliberate release of products that can influence its interactions with other bacteria or algae is unknown.

The genes coding the PUFA biosynthesis (*pfa*) cluster (P700755_01456 to _01462) (fig. 7) are located on a third GI (no.17). The ability to synthesize omega-3 and omega-6-type PUFA is a rare trait among bacteria; it is restricted to class Gammaproteobacteria and phylum Bacteroidetes and, within those groups, is largely restricted to marine psychrophiles (Shulse and Allen 2011). The *pfa* cluster of ATCC 700755^T is similar to previously described clusters but has an altered structural arrangement of conserved domains compared with those found in Gammaproteobacteria, suggesting a different evolutionary process of acquisition. ATCC 700755^T, which can synthesize eicosapentaenoic acid (EPA) via this cluster, has higher levels of EPA in its cytoplasmic membrane at low temperatures (Nichols et al. 1997). Protein abundances were found to be substantial for *pfa* cluster gene products (fig. 7). Unlike Gammaproteobacteria, which form either EPA or docosahexaenoic acid, ATCC 700755^T also forms arachidonic acid, though its levels do not increase with temperature (Nichols et al. 1997) so it may have another role. We assume arachidonic acid is another by-product from the *pfa* cluster. PUFA is capable of maintaining homeoviscosity of membranes at very low temperatures (Russel and Nichols 1999; Usui et al. 2012) and, due to increasing cell hydrophobicity, also potentially shields cells against hydrophilic toxic substances such as peroxides (Nishida et al. 2010). Two

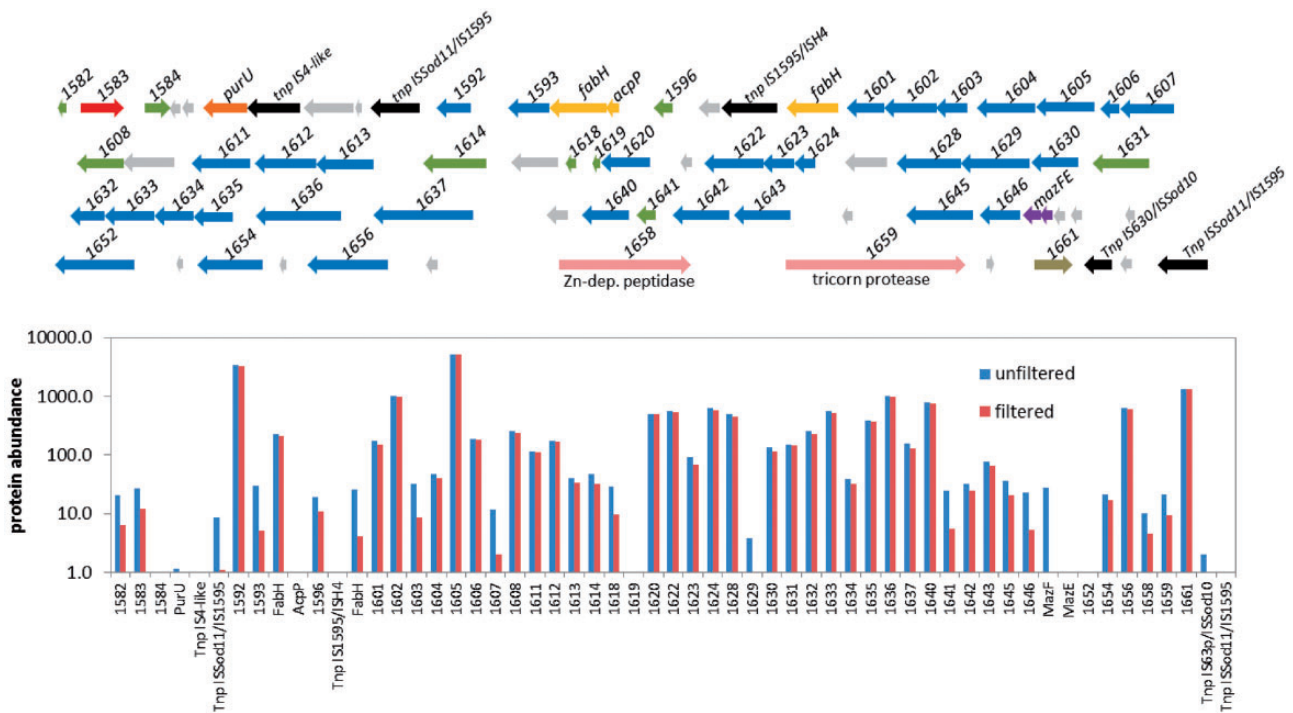


Fig. 5.—The GI region 18 of *P. torquis* ATCC 700755^T containing a large EPS biosynthesis cluster that is highly translated and likely involved in the organism’s manufacture of complex EPS. Relative abundance of gene products for this region is shown in the lower graph. Gray genes denote pseudogenes. Black genes denote intact transposases. Green genes are hypothetical proteins with conserved domain regions. Blue genes are enzymes involved in polysaccharide biosynthesis including synthesis of modified sugars, glycosylation, polymerases, and flippases; P700755_001654 and 1656 are near identical copies of putative UDP-glucose 6-dehydrogenase genes. Yellow genes are associated with lipid metabolism and include two FabH homologs. A dark red gene denotes a NusG-like transcriptional elongation/antitermination factor. Pink genes separated by large intergenic regions include putative exported proteases. Purple genes include an MazEF family addition module. Further details are shown in [supplementary tables S1 and S3, Supplementary Material](#) online.

genome-sequenced relatives of *P. torquis*, strain SCB49 (related to the genus *Ulvibacter*) isolated from the Arctic Ocean and a strain of the genus *Dokdonia* isolated from Arctic Ocean marine sediment (classified as *Krokinobacter* sp. 4H-3-7-5) possess *pfa* gene clusters very similar to that of *P. torquis*. This suggests that this type of *pfa* gene cluster could be prevalent in other psychrophilic members of the family Flavobacteriaceae. The *pfa* cluster in ATCC 700755^T is flanked by a number of intact and remnant transposases and a DNA-binding excisionase is located immediately upstream, suggesting that the *pfa* cluster may have been mobilized into the ATCC 700755^T genome via a phage insertion or conjugative transposon. PUFA production is also very sensitive to lipid oxidation (Imlay 2003). The sea-ice environment is generally saturated with oxygen due to photosynthetic activity and low temperature (D’Amico et al. 2006), which may partly explain why ATCC 700755^T possesses a wide array of enzymes that provide immediate protection against reactive oxygen species and organic peroxides. The array of defences includes genes for two catalases (P700755_00288, _02059), seven peroxidases (P700755_00120, _0196, _01102, _01308, _03056, _03338, _03478), and three superoxide dismutases (P700755_00728, _00729, 01787). Some of these have

homologs in ACAM 44^T that likely also experiences photooxidative stress in its lake environment. However, several are located on GIs in ATCC 700755^T, including genes coding a diheme cytochrome peroxidase (GI no. 41-P700755_03478), nickel- and iron-based superoxide dismutases (GI no. 8-P700755_00728, _00729, GI no. 19-_01787), and a putatively secreted catalase (P700755_00288).

Occurrence of the Proteorhodopsin Gene of *P. torquis* ATCC 700755^T at the Edge of a GI That Contains Putative Ice-Binding Proteins

Some *P. torquis* genes are suspected of having a role in sea-ice inhabitation, but the functions remain tentative and the coded proteins were generally only weakly abundant in the proteomic survey. The suppositions are based on the likelihood they could be advantageous in a sea-ice ecosystem setting with translation requiring specific conditions. One such trait is the ability to bind and/or interact with ice crystalline surfaces, aiding recruitment and persistence within sea ice (Raymond et al. 2007). Genes that could have this role in ATCC 700755^T include a cluster of secreted proteins that may act as adhesins and that have a C-terminal domain homologous to other

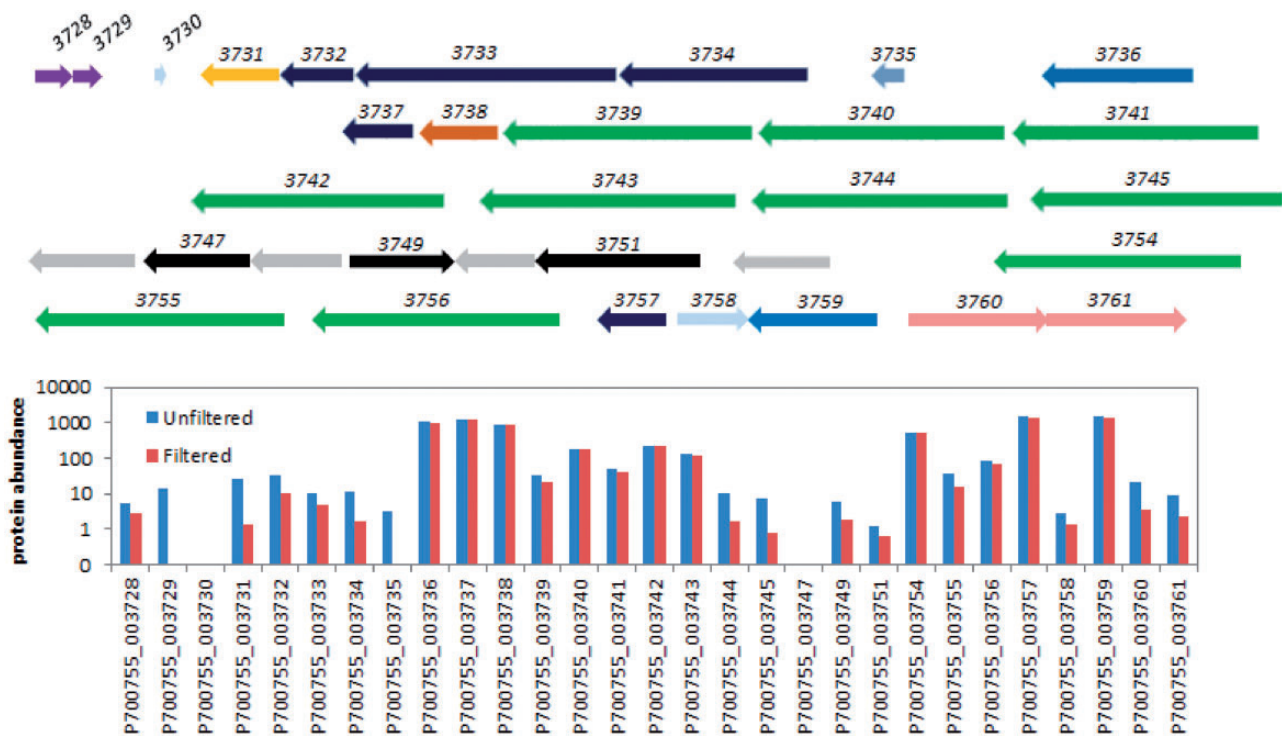


FIG. 6.—The GI region 43 of *P. torquis* ATCC 700755^T containing a highly translated cluster of transporter proteins including eight ADOP family permease-like proteins. Relative protein abundance of gene products for this region is shown in the lower graph. Gray genes denote pseudogenes. Black genes denote intact transposases. Purple genes comprise an addiction module. Green genes denote the ADOP family permease-coding genes. Dark blue and indigo genes show different families of other transporters including those with ATP-binding regions. The orange gene codes a putative multifunctional acyl-CoA thioesterase. Other genes code hypothetical proteins. Further details are shown in [supplementary tables S1 and S3, Supplementary Material](#) online.

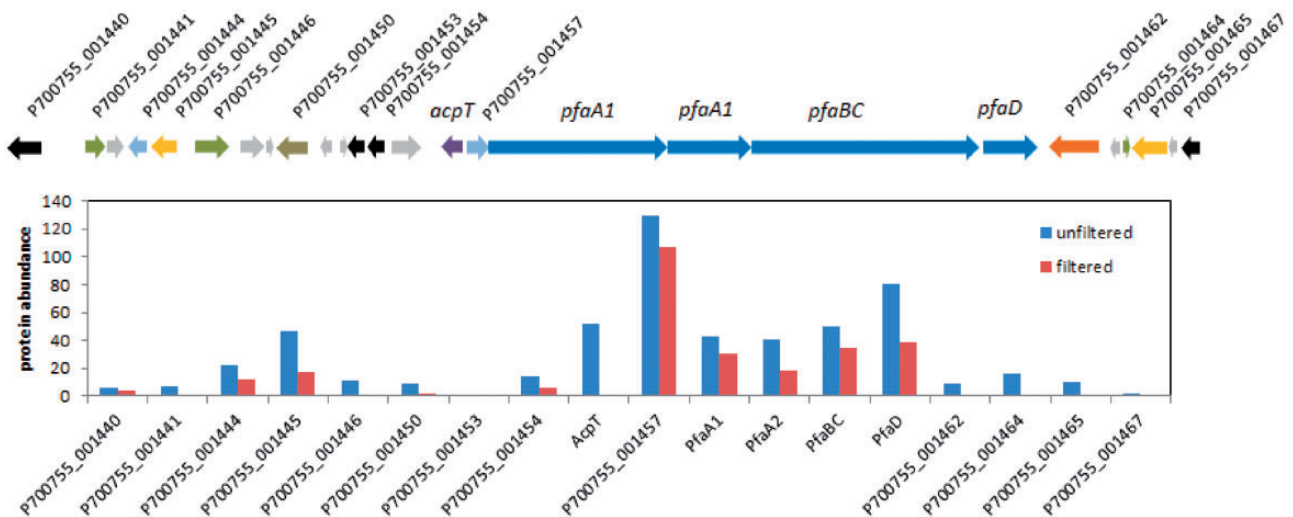


FIG. 7.—The gene content of the GI region 17 of *P. torquis* ATCC 700755^T (upper graph) containing the omega-3 polyunsaturated fatty acid (*pfa*) gene cluster and the relative protein abundance of gene products for this region (lower graph). The *acpT*/P700755_1457 (TetR family protein)/*pfaA1A2BCD* region is conserved in two other genome sequenced bacteria: strain SCB49 and *Krokinobacter* sp. 4H-3-7-5. Gray genes denote pseudogenes. Black genes denote transposases. Green genes are hypothetical proteins. Putative TetR family transcriptional regulators are shown in pale blue. Other genes shown are likely enzymes but function is only generally predicted from conserved domain data.

ice binding proteins, including those of the sea-ice diatom *F. cylindrus* (Bayer-Giraldi et al. 2010). Interestingly, the putative ice-binding/adhesin proteins are located in a GI (no. 2), which has at its edge a proteorhodopsin gene and its cognate carotenoid monooxygenase and, immediately upstream, ABC-type transporters for complexed iron or cobalamin. The genetic arrangement of proteorhodopsin is largely conserved in ACAM 44^T; however, the putative ice-binding protein cluster is mostly absent (fig. 8). The adjacent ice-binding proteins can be readily detected but are far-less abundant, perhaps pointing to a more generalized and important role for proteorhodopsin. The putative ice-active proteins found in *P. torquis* are related to other proteins found in bacteria, algae, and yeast (fig. 8), and several with confirmed ice-binding and anti-freeze functions; nevertheless, substantial work is required to substantiate their functionality. Located on GI no. 17, adjacent to yet another putative ice-binding protein, lies a series of two-component sensor systems clustered over a 16-kb region,

which are weakly orthologous to bacteriophytochromes and contain multiple Per-Arnt-Sim (PAS) and cGMP-specific phosphodiesterases, adenylyl cyclases, and FhIA (GAF) domains that have been suggested to be light sensors in the genome of the proteorhodopsin-possessing strain MED152 (Gonzalez et al. 2008). These proteins are only weakly to moderately abundant and a photobiological function remains unconfirmed. *Psychroflexus torquis*, however, responded significantly to light, increasing its growth yield by two to three times depending on the salinity and light level (Feng et al. 2013), which suggests the presence of sensing and regulatory systems that must involve an ability to respond to changing light and salinity conditions.

Aspects of the *P. torquis* ATCC 700755 Genome Linked to Nutrient Acquisition and Algal Colonization in Sea Ice

A critical aspect for sea-ice adaptation is nutrient acquisition. The nature of this ability can be partly surmised by the already

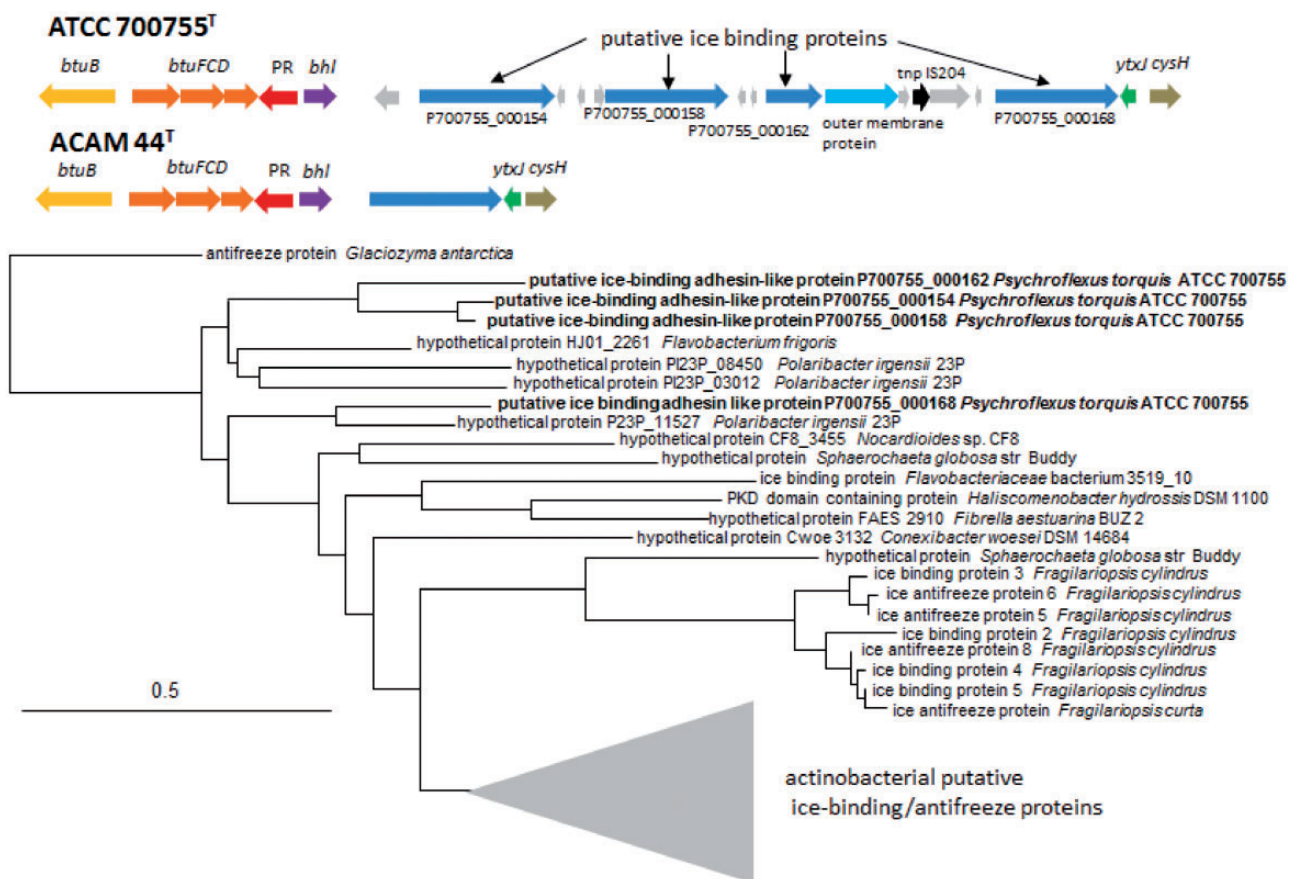


Fig. 8.—The proteorhodopsin and putative ice-binding protein gene cluster associated with the GI region 2 of *P. torquis* ATCC 700755^T genome compared with equivalent region from *P. gondwanensis* ACAM 44^T. Genes shown in gray are pseudogenes (see [supplementary tables S1–S3, Supplementary Material](#) online, for more details). Known and putative ice binding proteins within GI no. 2 are compared with other equivalent proteins from bacteria, diatoms (*Fragilariopsis* spp.), and yeast (*Glaciozyma antarctica*) in a protein sequence-based tree, where distances were calculated with the Grishin algorithm and clustering was via Neighbor-Joining, calculated using the constraint-based multiple alignment tool (www.ncbi.nlm.nih.gov, last accessed January 8, 2014) with default parameters.

known phenotype of strains of *P. torquis*. The species uses an eclectic range of carbon and energy sources including limited ranges of carbohydrates, amino acids, organic acids, and odd chain length lipid oxidation products (Bowman et al 1998). In sea ice, levels of dissolved organic compounds are at much higher levels than in the water column; they include large amounts of exopolymeric material, carbohydrates, free amino acids, and lipids (Norman et al 2011). ATCC 700755^T was thus expected to have a complement of genes specifically geared to access abundant substrates formed during algal primary production, which it may share with ACAM 44^T. Both strains are relatively fastidious and require growth factors, including vitamin B6 and cobalamin, consistent with their gene content for cofactor metabolism. Several nutrient acquisition-related proteins coded by GI-associated genes were found to be comparatively abundant based on proteomic analysis (supplementary table S1, Supplementary Material online). These included: a number of secreted and nonsecreted peptidases (GI no. 4-glutamate carboxypeptidase II, L-carnosine dipeptidase; GI no. 13 two subtilisin-like proteases); enzymes for catabolism of D,L-hydroxyproline (GI no. 34, HCP deaminase, 4-hydroxyproline epimerase); a putative SGLT1-like family glucose/Na⁺ ion symporter; and several TonB-dependent outer membrane receptor and binding proteins, mainly of the RagA/SusCD families (GI no. 5, 13, 16), which typically take up carbohydrates, polypeptides, and/or chelated metallic cations. Because sea ice is a subzero temperature environment, these proteins are almost certainly cold-active (Huston et al. 2000). The genome of *C. psychrerythraea* 34H was noted to possess a large number of extracellular enzyme coding genes (Méthé et al. 2005), and dwelling at extreme cold seems to require the production of large amounts of extracellular enzymes to overcome the mass transfer limitations caused by low-temperature impairment of enzyme function and transport (Struvay and Feller 2012).

Another aspect of sea-ice inhabitation is the link *P. torquis* has to algae as an epiphyte. Given that *P. torquis* has substantial growth factor requirements and is slow growing (fastest doubling time ~1 day), persisting in a dynamic system may require close interaction with sources of nutrients. Details of bacterial/algal associations derivable from the genome data are so far effectively limited to inferences; however, they are compelling and provide several possible lines of research on competitive and mutualistic interactions in sea ice. Both ATCC 700755^T and ACAM 44^T possess a conserved set of surface-gliding motility-associated genes (McBride and Zhu 2013) and the associated Por secretion system (Sato et al 2010); however, they strongly differ in terms of cell-surface proteins and the presence of several adhesin-like proteins (fig. 2). ATCC 700755^T is able to perform a form of slow gliding motility (Bowman et al. 1998) that has not been demonstrated in ACAM 44^T.

Putative adhesins present in ATCC 700755^T include several types such as the aforementioned putative ice-binding

adhesin-like proteins, VCBS, and fasciclin repeat domain-containing proteins and autotransporter adhesins. ATCC 700755^T also possesses a large surface protein of 468 kDa (P700755_00663) homologous to colossin A protein from the slime mold *Dictyostelium discoideum* (Whitney et al. 2010) that could also be involved with surface interactions. The diverse range of these proteins along with ice-active proteins seems to suggest a flexible attachment ability potentially necessary in the highly dynamic sea-ice ecosystem.

The possibility emerges that the algal interactions of ATCC 700755 run deeper than simple commensalism. Both ATCC 700755^T and ACAM 44^T strains possess several signaling proteins that putatively respond to the presence of plant hormones, including GH3 auxin promoter proteins that may allow them to coordinate growth and activity with that of algal hosts (Lambrecht et al. 2000). Unusually, ATCC 700755^T secretes 2-phenylethylamine (PEA) in substantial levels (Hamana and Niitsu 2001). The physiological function of PEA in relation to algae is unclear, but PEA has been found to induce production of oxidative bursts in tobacco that has relatively high levels of PEA in its leaves and could be linked to triggering defence systems (Kawano et al. 2000). An intriguing 35-kb cluster of proteins (P700755_01235 to _01257) on GI no. 15, neighboring a glycogen synthesis cluster (P700755_01229 to _01232) could also be involved in algal interactions (fig. 9). The cluster includes signaling proteins with cyclases/histidine kinase associated sensory extracellular (CHASE) domains as well as several large WD40 repeat domains containing four metacaspase family proteins. Metacaspases have been linked to programmed cell death functions in lower eukaryotes including phytoplankton (Madeo et al. 2012; Choi and Berges 2013). At this stage, few details on the functionality of bacterial metacaspases are available (Vercammen et al. 2007). Cooperative and noncooperative interactive mechanisms between algae and bacteria, whose populations are functionally tightly coupled, are a critical aspect of sea-ice ecology, yet remain largely unknown.

Conclusions

Staley and Gosink (1999) indicated that a number of bacterial genera exist in both Arctic and Antarctic sea ice, but whether this finding can be extended to the species level was unknown (Staley and Gosink 1999). They argued that the global distribution of psychrophiles was essentially blocked by a tropical marine barrier, providing the possibility that ecosystem-linked endemism could emerge. Recently, evidence based on high-throughput sequencing of marine microbial communities has been presented that, like macroscopic organisms, marine bacteria seem to be subject to biogeographic limitations affecting their current and presumably long-term distribution. Not only do these proposed limitations suggest the potential for localized speciation in certain ecosystems (Whitaker 2006), but it

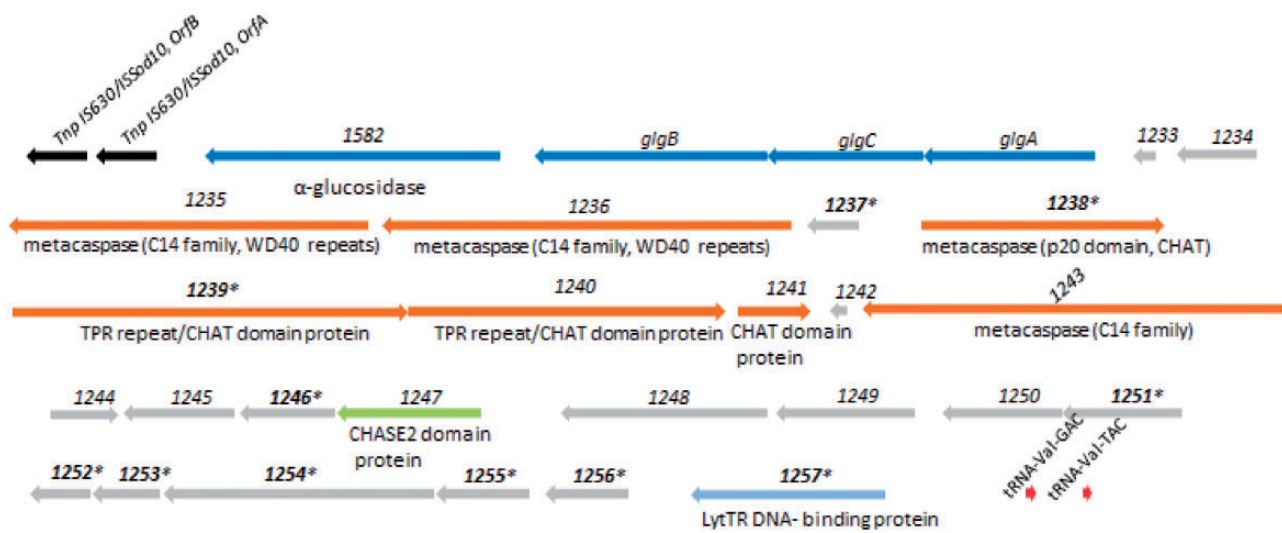


Fig. 9.—The GI region 15 of *P. torquis* ATCC 700755^T that contains a cluster of putative metacaspase coding genes (orange) following a glycogen-synthesis/utilization cluster (blue) with a possible role linked to programmed cell death. The domain structure of metacaspase genes and associated genes are indicated. Genes flagged with asterisks are those with signal peptide regions and thus putatively secreted. A putative LytTR family regulatory protein is coded by gene P700755_001257, while a CHASE2 family protein (usually associated with extracellular sensory proteins) is shown in green. The region is flanked by transposases (black genes) and two tRNA-Val genes.

also suggests that some communities of bacteria are vulnerable to extinction brought about by external disruption such as climate change, habitat destruction, or invasion by other organisms (Sul et al. 2013). Here we demonstrate the comparative genomic features of the bipolar bacterial species *P. torquis* that could be an excellent example of evolving endemism in a bacterial species. A next step in genome-level analysis would be to compare Arctic and Antarctic strains to determine genetic similarity and degree of change, especially in the number and content of GIs, relative state of gene decay and overall occurrence of HGT gene regions, which make up ~35% of the genome of strain ATCC 700755^T. This major finding of our study is consistent with the suggestion, based on abundance of phage and extracellular DNA in sea-ice brines, that sea ice is a hotspot for HGT (Collins and Deming 2013). High levels of HGT in other sea ice-associated bacteria can also be expected. Sea ice or polar seawater-associated *Octadecabacter* species, which possess the light-driven proton pump xanthorhodopsin, also show evidence of high levels of HGT and have genomes rich in IS elements, pseudogenes, and plasmids (Vollmers et al. 2013). Sea-ice specialism apparent in *P. torquis* appears to be linked to its GI-associated genes, including those for EPS and PUFA synthesis, modes of nutrient acquisition and potential ice-binding and algal interactions. Proteomic data provided evidence that many of the associated genes are being actively translated and are thus important to the biology of *P. torquis*. Overall, this study suggests *P. torquis* could be an excellent model to study sea-ice functional biology and evolutionary processes linked with psychrophily, endemism, and algal interactions.

Supplementary Material

Supplementary figure S1 and tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The work was supported by a grant from the Department of Sustainability, Environment, Water, Population, and Communities and the Australian Antarctic Division (project no. 3127) and a grant from the University of Tasmania to S.F. The authors thank Prof. Tom McMeekin for critical discussions related to the manuscript.

Literature Cited

- Bayer-Giraldi M, Uhlig C, John U, Mock T, Valentin K. 2010. Antifreeze proteins in polar sea ice diatoms: diversity and gene expression in the genus *Fragilariopsis*. *Environ Microbiol*. 12:1041–1052.
- Berge J, et al. 2012. Retention of ice-associated amphipods: possible consequences for an ice-free Arctic Ocean. *Biol Lett*. 8:1012–1015.
- Bland C, et al. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209.
- Bluhm BA, et al. 2011. Arctic marine biodiversity: an update of species richness and examples of biodiversity change. *Oceanography* 24: 232–248.
- Borg J, et al. 2011. Spectral counting assessment of protein dynamic range in cerebrospinal fluid following depletion with plasma-designed immunoaffinity columns. *Clin Proteomics*. 8:6.
- Bowman JP. 2008. Genomic analysis of psychrophilic prokaryotes. In: Margesin R, Schinner F, Marx JC, Gerday C, editors. *Psychrophiles: from biodiversity to biotechnology*. New York: Springer. p. 265–284.

- Bowman JP, McCammon SA, Brown MV, Nichols DS, McMeekin TA. 1997. Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl Environ Microbiol.* 63:3068–3078.
- Bowman JP, et al. 1998. *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson, et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* 144:1601–1609.
- Bowman JS, et al. 2012. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S rRNA gene. *ISME J.* 6:11–20.
- Brinkmeyer R, et al. 2003. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl Environ Microbiol.* 69:6610–6619.
- Brown MV, Bowman JP. 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol.* 35:267–275.
- Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol.* 3:195–208.
- Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119–120.
- Choi CJ, Berges JA. 2013. New types of metacaspases in phytoplankton reveal diverse origins of cell death proteases. *Cell Death Dis.* 4:e490.
- Collins RE, Deming JW. 2013. An inter-order horizontal gene transfer event enables the catabolism of compatible solutes by *Colwellia psycherythraea* 34H. *Extremophiles* 17:601–610.
- Corkrey R, Olley J, Ratkowsky D, McMeekin T, Ross T. 2012. Universality of thermodynamic constants governing biological growth rates. *PLoS One* 7:e32003.
- D'Amico S, Collins T, Marx JC, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep.* 7:385–389.
- D'Aoust DY, Kushner DJ. 1971. Structural changes during lysis of a psychrophilic marine bacterium. *J Bacteriol.* 108:916–927.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636–4641.
- Ewert M, Deming JW. 2011. Selective retention in saline ice of extracellular polysaccharides produced by the cold-adapted marine bacterium *Colwellia psycherythraea* strain 34H. *Ann Glaciol.* 52:111–117.
- Fanzmann PD. 1996. Examination of Antarctic prokaryotic diversity through molecular comparisons. *Biodivers Conserv.* 5:1295–1305.
- Feng S, Powell SM, Wilson R, Bowman JP. 2013. Light stimulated growth of proteorhodopsin bearing sea-ice psychrophile *Psychroflexus torquis* is salinity-dependent. *ISME J.* 7:2206–2213.
- Gao F, Luo H, Zhang CT. 2012. DoriC 5.0: an updated database of *oriC* regions in both bacterial and archaeal genomes. *Nucleic Acids Res.* 41:D90–D93.
- Gonzalez JM, et al. 2008. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp MED152 (Flavobacteria). *Proc Natl Acad Sci U S A.* 105:8724–8729.
- Goris J, et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol.* 57:81–91.
- Grossmann S, Dieckmann GS. 1994. Bacterial standing stock, activity and carbon production of sea ice in the Weddell Sea, Antarctica. *Appl Environ Microbiol.* 60:2746–2753.
- Hamana K, Niitsu M. 2001. Large production of an aromatic amine, 2-phenylethylamine, in a psychrophilic marine bacterium, *Psychroflexus torquis*. *J Gen Appl Microbiol.* 47:103–105.
- Hassler C, Schoemann V, Mancuso-Nichols CA, Butler EC, Boyd P. 2011. Saccharides enhance iron bioavailability to Southern Ocean phytoplankton. *Proc Natl Acad Sci U S A.* 108:1076–1081.
- Hoffman PF, Schrag DP. 2002. The snowball Earth hypothesis testing the limits of global change. *Terra Nova* 14:129–155.
- Huston AL, Krieger-Brockett BB, Deming JW. 2000. Remarkably low temperature optima for extracellular enzyme activity from Arctic bacteria and sea ice. *Environ Microbiol.* 2:383–388.
- Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
- Imlay JA. 2003. Pathways of oxidative damage. *Annu Rev Microbiol.* 57:395–418.
- Kawano T, et al. 2000. Aromatic monoamine-induced immediate oxidative burst leading to an increase in cytosolic Ca²⁺ concentration in tobacco suspension culture. *Plant Cell Physiol.* 41:1251–1258.
- Kottmeier ST, Grossi SM, Sullivan CW. 1987. Sea ice microbial communities.VIII. Bacterial production in annual sea ice of McMurdo Sound, Antarctica. *Mar Ecol Prog Ser.* 35:175–186.
- Kottmeier ST, Sullivan CW. 1987. Late winter primary production and bacterial production in sea ice and seawater west of the Antarctic Peninsula. *Mar Ecol Prog Ser.* 36:287–298.
- Krembs C, Eicken H, Deming JW. 2011. Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic. *Proc Natl Acad Sci U S A.* 108:3653–3658.
- Kuo CH, Ochman H. 2010. The extinction dynamics of bacterial pseudogenes. *PLoS Biol.* 6:e1001050.
- Lambrecht M, Okon Y, Vande Broek A, Vanderleyden J. 2000. Indole-3-acetic acid: a reciprocal signalling molecule in bacteria–plant interactions. *Trends Microbiol.* 8:298–300.
- Lerat E, Ochman H. 2005. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res.* 33:3125–3132.
- Liu HB, Sadygov RG, Yates JR. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem.* 76:4193–4201.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
- Macdonald FA, et al. 2010. Calibrating the Cryogenian. *Science* 327:1241–1243.
- Madeo F, et al. 2002. A caspase-related protease regulates apoptosis in yeast. *Mol Cell.* 9:911–917.
- Mancuso Nichols CA. 2005. Exopolysaccharide production by Antarctic marine bacteria [PhD thesis]. [Australia]: University of Tasmania.
- Marx JG, Carpenter SD, Deming JW. 2009. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psycherythraea* strain 34H under extreme conditions. *Can J Microbiol.* 55:63–72.
- McBride MJ, Zhu Y. 2013. Gliding motility and Por secretion system genes are widespread among members of the phylum Bacteroidetes. *J Bacteriol.* 195:270–278.
- McMinn A, Martin A. 2013. Dark survival in a warming world. *Proc Biol Sci.* 280:20122909.
- Méthé BA, et al. 2005. The psychrophilic lifestyle as revealed by the genome sequencing of *Colwellia psycherythraea* 34H through genome and proteomic analyses. *Proc Natl Acad Sci U S A.* 102:10913–10918.
- Mosier AC, Murray AE, Fritsen CH. 2007. Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS Microbiol Ecol.* 59:274–288.
- Nandakumar MP, Shen J, Raman B, Marten MR. 2003. Solubilization of trichloroacetic acid (TCA) precipitated microbial proteins via NaOH for two-dimensional electrophoresis. *J Proteome Res.* 2:89–93.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 75:4646–4658.
- Nichols DS, Brown JL, Nichols PD, McMeekin TA. 1997. Production of eicosapentaenoic acid and arachidonic acids by an Antarctic

- bacterium: response to growth temperature. *FEMS Microbiol Lett.* 152:349–354.
- Nishida T, Hori R, Morita N, Okuyama H. 2010. Membrane eicosapentaenoic acid is involved in the hydrophobicity of bacterial cells and affects the entry of hydrophilic and hydrophobic compounds. *FEMS Microbiol Lett.* 306:91–96.
- Norman L, et al. 2011. The characteristics of dissolved organic matter (DOM) and chromophoric dissolved organic matter (CDOM) in Antarctic sea ice. *Deep-Sea Res Pt II.* 59:1075–1091.
- Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 40:299–304.
- Ou HY, et al. 2006. A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res.* 34:e3.
- Ou K, et al. 2005. Integrative genomic, transcriptional, and proteomic diversity in natural isolates of the human pathogen *Burkholderia pseudomallei*. *J Bacteriol.* 187:4276–4285.
- Petersen TN, Brunak S, Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8: 785–786.
- Polyak L. 2010. History of sea ice in the Arctic. *Quat Sci Rev.* 29: 1757–1778.
- Poulin M, et al. 2011. The pan-Arctic biodiversity of marine phytoplankton and sea-ice unicellular eukaryotes: A first-attempt assessment. *Mar Biodivers.* 41:13–28.
- Puigbò P, Bravo IG, Garcia-Vallvé S. 2008. PE-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 9:65.
- Ratkowsky DA, Lowry RK, McMeekin TA, Stokes AN, Chandler RE. 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J Bacteriol.* 154:1222–1226.
- Rauch A, et al. 2006. Computational proteomics analysis system (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiment. *J Proteome Res.* 5:112–121.
- Raymond JA, Fritsen C, Shen K. 2007. An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol Ecol.* 61:214–221.
- Riley M, et al. 2008. Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genomics* 9:210.
- Rusk N. 2011. From pseudogenes to proteins. *Nat Methods* 8: 448–449.
- Russell NJ, Nichols DS. 1999. Polyunsaturated fatty acids in marine bacteria—a dogma rewritten. *Microbiology* 145:767–779.
- Sato K, et al. 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A.* 107: 276–281.
- Schwartzman D, Lineweaver CH. 2005. Temperature, biogenesis and biospheric self-organization. In: Kleidon A, Lorenz R, editors. *Non-equilibrium thermodynamics and the production of entropy: life, Earth and beyond.* New York: Springer. p. 207–217.
- Shulse CN, Allen EE. 2011. Diversity and distribution of microbial long-chain fatty acid biosynthetic genes in the marine environment. *Environ Microbiol.* 13:684–695.
- Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today.* 33:152–155.
- Staley JT, Gosink JJ. 1999. Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol.* 53:189–215.
- Struvay C, Feller G. 2012. Optimization to low temperature activity in psychrophilic enzymes. *Int J Mol Sci.* 13:11643–11665.
- Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, Sogin ML. 2013. Marine bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci U S A.* 110:2342–2347.
- Thomas DN, Dieckmann GS. 2002. Antarctic sea ice—a habitat for extremophiles. *Science* 295:641–644.
- Usui K, et al. 2012. Eicosapentaenoic acid plays a role in stabilizing dynamic membrane structure in the deep-sea piezophile *Shewanella violacea*: a study employing high-pressure time-resolved fluorescence anisotropy measurement. *Biochim Biophys Acta.* 1818:574–583.
- Vercammen D, Declercq W, Vandenaabeele P, Van Breusegem F. 2007. Are metacaspases caspases? *J Cell Biol.* 179:375–380.
- Vollmers J, et al. 2013. Poles apart: Arctic and Antarctic *Octadecabacter* strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS One* 8:e63422.
- Ward NL, et al. 2009. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol.* 75:2046–2056.
- Wells LE, Deming JW. 2006. Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ Microbiol.* 8:1115–1121.
- Whitaker RJ. 2006. Allopatric origins of microbial species. *Philos Trans R Soc Lond B Biol Sci.* 361:1975–1984.
- Whitney TJ, Gardner DG, Mott ML, Brandon M. 2010. Identifying the molecular basis of functions in the transcriptome of the social amoeba *Dictyostelium discoideum*. *Genet Mol Res.* 9: 394–415.
- Yau S, et al. 2013. Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME J.* 7:1944–1961.
- Zeldovich KB, Berezovsky IN, Shakhnovich EI. 2007. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput Biol.* 3:e5.

Associate editor: John McCutcheon