FEMS Microbiology Ecology, 94, 2018, fiy023

doi: 10.1093/femsec/fiy023
Advance Access Publication Date: 8 March 2018
Research Article

RESEARCH ARTICLE

Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria

Isabelle Raymond-Bouchard[1], Jacqueline Goordial[1,2,†], Yevgen Zolotarov[1], Jennifer Ronholm[1], Martina Stromvik[1], Corien Bakermans[3] and Lyle G. Whyte[1,*]

[1]McGill University, Macdonald Campus, 21,111 Lakeshore Rd, Ste.-Anne-de-Bellevue, QC, H9X 3V9, Canada, [2]Bigelow laboratory for Ocean Sciences, 60 Bigelow Dr, East Boothbay, ME, 04544, USA and [3]Altoona College, Pennsylvania State University, 3000 Ivyside Park, Altoona, PA, 16601, USA

*Corresponding author: Department of Natural Resource Sciences, McGill University, 21111 Lakeshore Rd. Ste-Anne-de-Bellevue, Montreal, Quebec H9X 3V9 Canada. Tel: +1-514-398-7823; Fax: +1-514-398-7990; E-mail: lyle.whyte@mcgill.ca

One sentence summary: This is the first large scale study focused on identifying genomic and molecular traits associated with cold adaptation in subzero-growing bacteria from permafrost.

Editor: Max Haggblom
†Jacqueline Goordial, http://orcid.org/0000-0001-5005-673X

ABSTRACT

Permafrost accounts for 27% of all soil ecosystems and harbors diverse microbial communities. Our understanding of microorganisms in permafrost, their activities and adaptations, remains limited. Using five subzero-growing (cryophilic) permafrost bacteria, we examined features of cold adaptation through comparative genomic analyses with mesophilic relatives. The cryophiles possess genes associated with cold adaptation, including cold shock proteins, RNA helicases, and oxidative stress and carotenoid synthesis enzymes. Higher abundances of genes associated with compatible solutes were observed, important for osmoregulation in permafrost brine veins. Most cryophiles in our study have higher transposase copy numbers than mesophiles. We investigated amino acid (AA) modifications in the cryophiles favoring increased protein flexibility at cold temperatures. Although overall there were few differences with the mesophiles, we found evidence of cold adaptation, with significant differences in proline, serine, glycine and aromaticity, in several cryophiles. The use of cold/hot AA ratios of >1, used in previous studies to indicate cold adaptation, was found to be inadequate on its own. Comparing the average of all cryophiles to all mesophiles, we found that overall cryophiles had a higher ratio of cold adapted proteins for serine (more serine), and to a lesser extent, proline and acidic residues (fewer prolines/acidic residues).

Keywords: psychrophile; psychrotroph; cold adaptation; subzero; permafrost; amino acid composition

INTRODUCTION

Permafrost accounts for 27% of all soil ecosystems on earth and is home to active and diverse microbial communities, despite extreme conditions characterized by permanent sub-zero temperatures, low water activity and often oligotrophic conditions (Goordial *et al.* 2012). The identification of active microbial communities in permafrost has fueled interest in understanding the adaptations that allow these organisms to sustain their growth and survival in such harsh environments. Reduced enzyme

Received: 8 August 2017; **Accepted:** 7 March 2018

activity, protein denaturation and misfolding, inhibitory secondary RNA/DNA structures and increases in membrane rigidity are challenges that must be overcome at low temperatures (Bakermans *et al.* 2012; Mykytczuk *et al.* 2013; De Maayer *et al.* 2014). Noted adaptations to cold found in psychrophiles include, but are not limited to, using cold shock proteins and helicases to stabilize DNA/RNA processes, membrane modifications to increase unsaturated and branched chain fatty acids, synthesis of carotenoids and antifreeze proteins, stimulation of osmotic and oxidative responses, differential expression of isozymes with different temperature optima and broad changes in translational and ribosomal processes (Bakermans *et al.* 2012). In this paper, we use the term psychrophile to encompass both eurypsychrophiles and stenopsychrophiles (Feller and Gerday 2003; Cavicchioli 2016); eurypsychrophiles are defined as organisms that possess a broad growth range (Tmax >20°C), and who, while capable of growth at very low temperatures (−15°C), usually retain optimum growth rates at temperatures above 20°C, while stenopsychrophiles exhibit narrower growth ranges (Tmax <20°C), and usually grow best at temperatures ranging between 5°C and 15°C (Cavicchioli 2006). The term cryophile is used here to denote specifically those psychrophiles that can grow at subzero temperatures (Seckbach 2013; Vester, Glaring and Stougaard 2015; Goordial *et al.* 2016).

Ensuring that adequate flexibility is retained is a crucial adaptation of cold active proteins to preserve function at low temperatures. Cold adapted proteins have been shown to possess a number of amino acid (AA) changes that impart increased flexibility, including fewer salt bridges, a lower content of proline residues, fewer hydrogen bonds, a reduced Arg/(Arg + Lys) ratio and increased serine and glycine content (Aghajari *et al.* 1998; Russell *et al.* 1998; Georlette *et al.* 2000; Huston, Methe and Deming 2004; Collins, Gerday and Feller 2005; Ræder *et al.* 2008; Metpally and Reddy 2009). For example, the elongation factor 2 protein of the eurypsychrophilic archaea *Methanococcoides burtonii*, isolated from the permanently cold Ace Lake in Antarctica, exhibits greater structural flexibility as a result of fewer salt bridges, less densely packed hydrophobic cores and a reduction in proline residues in surface loops (Thomas and Cavicchioli 2000). A lipase from the Antarctic bacterium *Psychrobacter immobilis* B10 also contains a low number of salt bridges and proline residues, along with a reduced proportion of arginine to lysine residues, a small hydrophobic core and a very small number of aromatic-aromatic interactions (Arpigny, Lamotte and Gerday 1997). Differences in k_{cat} and K_m values, corresponding to increased catalytic efficiency at lower temperatures, have been measured in certain psychrophilic enzymes when compared to their meso- or thermophilic homologs (Feller *et al.* 1992; Georlette *et al.* 2000; Thomas, Kumar and Cavicchioli 2001).

Several studies have carried out comparative genomic analyses between psychrophiles and mesophiles and/or thermophiles in order to examine differences in AA composition, which may impart greater flexibility of proteins in cold adapted microorganisms. Of these, only four focused on microorganisms that exhibit growth at subzero temperatures and in each case focused on a single organism (Methé *et al.* 2005; Ayala-del-Río *et al.* 2010; Mykytczuk *et al.* 2013; Goordial *et al.* 2016). Previous genomic analyses include those focused on psychrophilic protein comparisons against large nonredundant protein databases and comparisons against select specific mesophiles (Saunders *et al.* 2003; Rabus *et al.* 2004; Methé *et al.* 2005; Grzymski *et al.* 2006; Metpally and Reddy 2009; Ayala-del-Río *et al.* 2010; Zhao *et al.* 2010; Mykytczuk *et al.* 2013; Goordial *et al.* 2016). Methods for comparison have included those that compare total AA counts or composition across an entire genome between mesophiles and psychrophiles, to those that compare AA counts only between previously matched homologous proteins. Results from these studies have been inconsistent. While some identified a number of changes in AA composition or proportions in psychrophilic genomes, consistent with those identified in psychrophilic proteins (Saunders *et al.* 2003; Ayala-del-Río *et al.* 2010; Zhao *et al.* 2010), others have found limited changes (Methé *et al.* 2005; Mykytczuk *et al.* 2013; Goordial *et al.* 2016), and in some cases, none (Rabus *et al.* 2004; Yang *et al.* 2015). It remains unclear whether these AA substitutions are universal adaptations that can be conclusively identified on a large scale across genomes in psychrophiles. It is also of interest to determine if there are different adaptations associated with psychrophiles capable of growth at sub-zero temperatures (cryophiles), compared with those psychrophiles that have narrower growth ranges and that may not be adapted to the challenges associated with subfreezing temperatures. To our knowledge, very few studies on AA cold adaptive traits in psychrophiles have looked specifically at cryophilic psychrophiles. Furthermore, in those studies that perform comparisons using distantly related organisms, it is not always apparent if the differences observed are the result of cold adaptation or simply phylogenic divergence. Comparisons using organisms in the same genus, or neighbor genus, provide a more stringent analysis, as we expect these members to share close common ancestry, and therefore observed AA differences between psychrophiles and their close mesophilic relatives are more likely to be the result of cold adaptation.

In the present study, we selected five subzero-growing (cryophilic) strains (Table 1), previously isolated from the Canadian high Arctic (Steven *et al.* 2007a, 2008; Lacelle *et al.* 2011) and spanning several phyla (ia, Proteobacteria and Firmicutes), for genome sequencing and annotation, which were used to perform in-depth genomic and AA comparative analyses. These organisms were isolated from permafrost environments that are harsh, often oligotrophic, with low water activity (aw = ∼0.85) and ambient *in situ* temperatures of ∼−17°C (Gilichinsky, Soina and Petrova 1993; Steven *et al.* 2007b, 2008). Survival of microbes in these environments is hypothesized to be possible due to the presence of very thin brine films (Gilichinsky *et al.* 2003; Steven *et al.* 2006), where higher solute and salt concentrations act as freezing point depressants, allowing water to remain liquid. It is expected that microorganisms adapted to these environments would have mechanisms that allow them to grow under both cold (subzero) and high salt conditions. Canadian Arctic permafrost salt concentrations have been estimated at ∼14.6 g/kg (Steven *et al.* 2007a), while salt concentrations in Siberian cryopeg brine lenses have been measured at 170–300 g/L (17–30%) (Gilichinsky *et al.* 2003). We compared the genomes of our cryophilic strains from these environments to closely related mesophilic relatives in the same genus. With these new genome sequences, we aim to increase our potential for understanding and identifying the AA basis of cold adaptation in sub-zero growing organisms, including potentially conserved or important adaptive traits. In addition, we also investigate the presence of known cold adaptation genes as well as genomic redundancy, which is the presence of multiple gene copies, in each cryophile.

MATERIALS AND METHODS

Organism selection and identity

Actinotalea sp. KRMCY2, *Paenisporosarcina* sp. Eur1 9.01.10, *Methylobacterium* sp. AL-11, *Kocuria* sp. KROCY2 and *Polaromonas* sp.

Table 1. Permafrost and ice wedge strains selected for sequencing and cold adaptation analysis.

Strain	Location isolated	Sample type	Temp °C	Salinity (%)	Phylum	Reference
Actinotalea sp. KRMCY2[a]	Moose Lake, Yukon	Ice Wedge	−5–25	0–8	Actinobacteria	This study & Lacelle *et al.* (2011)
Polaromonas sp. Eur3 1.2.1	Eureka, Nunavut	Permafrost	−5–22	0–3	Proteobacteria	This study & Steven *et al.* (2008)
Paenisporosarcina sp. Eur1 9.01.10[a]	Eureka, Nunavut	Permafrost	−5–25	0–8	Firmicute	This study & Steven *et al.* (2007a)
Methylobacterium sp. AL-11	Eureka, Nunavut	Permafrost	−5–25	0–5	Proteobacteria	This study & Steven *et al.* (2008)
Kocuria sp. KROCY2[a]	Old Crow, Yukon	Ice Wedge	−5–25	0–19	Actinobacteria	This study & Lacelle *et al.* (2011)

[a]*Actinotalea* sp. KRMCY2, *Paenisporosarcina* sp. Eur1 9.01.10 and *Kocuria* sp. KROCY2 were originally identified as *Cellulomonas* sp. KRMCY2, *Sporosarcina* sp. Eur1 9.01.10 and *Microbacterium* sp. KROCY2, respectively, but were renamed following genome sequencing and 16S rRNA gene analysis (see text).

Eur3 1.2.1. were isolated from permafrost cores from Eureka, Nunavut or from ice samples from the Yukon, as previously described (Steven *et al.* 2007a, 2008; Lacelle *et al.* 2011). These strains were selected for genomic sequencing due to their ability to grow at sub-zero temperatures under laboratory conditions (Table 1). Based on 16S rRNA gene sequencing *Kocuria* sp. KROCY2, *Paenisporosarcina* sp. Eur1 9.01.10 and *Actinotalea* sp. KRMCY2 were initially identified as belonging to the genus *Microbacterium*, *Sporosarcina* and *Cellulomonas*, respectively (Steven *et al.* 2007a; Lacelle *et al.* 2011). However, recent 16S rRNA gene analyses using the EzBioCloud quality-controlled database (Yoon *et al.* 2017) has identified the closest taxonomically valid type strains for *Kocuria* sp. KROCY2, *Paenisporosarcina* sp. Eur1 9.01.10 and *Actinotalea* sp. KRMCY2 to be *Kocuria palustris* DSM 11 925(T) (100% similarity), *Paenisporosarcina indica* PN2T(T) (99.16% similarity) and *Actinotalea ferrariae* CF5–4(T) (98.13% similarity), respectively (Table S1, Supporting Information). Thus, we have renamed our strains accordingly. The closest taxonomically classified type strains for *Polaromonas* sp. Eur3 1.2.1 and *Methylobacterium* sp. EUR3 AL-11 were *Polaromonas jejuensis* NBRC 106 434(T) (99.04% similarity) and *Methylobacterium radiotolerans* JCM 2831(T) (99.93% similarity), respectively.

Growth conditions

Cultures were maintained on R2A agar or liquid (BD Difco, Sparks, Maryland) for *Actinotalea* sp. KRMCY2, *Paenisporosarcina* sp. Eur1 9.01.10, *Methylobacterium* sp. AL-11 and *Kocuria* sp. KROCY2 and $\frac{1}{2}$ R2A for *Polaromonas* sp. Eur3 1.2.1. Growth temperatures were described previously (Steven *et al.* 2007a, 2008; Lacelle *et al.* 2011) on agar only and tested in this study using liquid cultures at −5°C, 0°C, 5°C, 10°C, 22°C and 25°C in Tryptic Soy Broth (TSB; BD Difco, Sparks, Maryland), 1/10 TSB, $\frac{1}{4}$ TSB, R2A or $\frac{1}{2}$ R2A and supplemented with varying concentrations (2%–7%) of sucrose, glycerol or NaCl at subzero temperatures. Salinity tolerance was determined from liquid cultures grown on R2A amended with NaCl in 1%–2% increments from 0% to 20%. Previous phylogenic classification was verified by comparing the 16S rRNA gene sequence of each organism to the EzBioCloud quality-controlled database for comparison to valid taxonomically classified type strains (Yoon *et al.* 2017).

DNA extraction

Strains were grown to stationary state at room temperature (~23°C) on TSB (BD Difco). Genomic DNA was isolated using the Epicentre MasterPure Gram Positive DNA Purification Kit (Epicentre, Madison, Wisconsin) as per the manufacturer's instructions. Extracted DNA was checked for purity and quantified with the NanoDrop 1000 (Thermoscientific, Wilmington, Delaware) and sent for whole genome sequencing according to DOE Joint Genome Institute (JGI) standards.

Genome sequencing and assembly

The draft genomes of *Kocuria* sp. KROCY2, *Actinotalea* sp. KRMCY2, *Polaromonas sp.* EUR3 1.2.1 and *Methylobacterium* sp. EUR3 AL–11 were generated at the DOE JGI using the Pacific Biosciences (PacBio) technology. Pacbio SMRTbell libraries were constructed for each isolate and sequenced on the PacBio RS platform, which generated 166 364 filtered subreads totaling 486.7 Mbp for *Kocuria* sp. KROCY2, 253 745 filtered subreads totaling 785.8 Mbp for *Actinotalea* sp. KRMCY2, 150 954 filtered subreads totaling 668.5 Mbp for *Polaromonas sp.* EUR3 1.2.1, and 369 217 filtered subreads totaling 928.2 Mbp for *Methylobacterium* sp. EUR3 Al-11. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads were assembled using HGAP (version: 2.3.0) (Chin *et al.* 2013). The final draft assembly contained one contig in one scaffold for *Kocuria* sp. KROCY2, *Actinotalea* sp. KRMCY2 and *Polaromonas sp.* EUR3 1.2.1 and produced an average input read coverage of 186.4X, 179.6X and 108.4X, respectively, and seven contigs in seven scaffolds for *Methylobacterium* sp. EUR3 AL-11 with input read coverage of 158.9X.

The draft genome of *Paenisporosarcina* sp. EUR3 2.2.2 was generated at the DOE JGI using the Illumina technology. The Illumina shotgun library and long insert mate pair library were constructed and sequenced using the Illumina HiSeq 2000 platform (Bennett 2004). 22 752 872 reads totalling 3412.9 Mb were generated from the shotgun and 40 430 662 reads totalling 3679.2 Mb were generated from the long insert mate pair library. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. All raw Illumina sequence data were passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (Li, Copeland and Han 2011). Filtered Illumina reads were assembled using AllpathsLG (PrepareAllpathsInputs: PHRED 64 = 1 PLOIDY = 1 FRAG COVERAGE = 150 JUMP COVERAGE = 25; RunAllpathsLG: RUN = std pairs TARGETS = standard VAPI WARN ONLY = True OVERWRITE = True) (Gnerre *et al.* 2011). The final draft assembly contained 18 contigs in 5 scaffolds. The final assembly is based on 3412.4 Mb of Illumina Std PE, 3679.1 Mb of Illumina CLIP PE post filtered

data, which provides an average 2026.1X Illumina coverage of the genome.

Genome annotation

Genes were identified using Prodigal (Hyatt *et al.* 2010), followed by a round of manual curation using GenePRIMP (Pati *et al.* 2010) for finished genomes and draft genomes in fewer than 10 scaffolds. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG and InterPro databases. The tRNAScanSE tool (Lowe and Eddy 1996) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse *et al.* 2007). Other noncoding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (Nawrocki and Eddy 2013). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes platform (http://img.jgi.doe.gov) developed by the Joint Genome Institute, Walnut Creek, CA, USA (Markowitz *et al.* 2009).

Genome and AA analyses

The genomes of close relatives (genus or neighboring genus) to the cryophilic strains, which have been annotated as mesophiles, were collected from publicly available databases (JGI), and were used for comparative analyses to the cryophilic genomes (Table S2, Supporting Information). Functional gene assignments, including COG and KEGG categories and pathways, were obtained from the JGI website for all the genomes, cross-referenced and used for comparative analyses. AA cold adaptation analyses was carried out as described previously (Goordial *et al.* 2016), with some modifications. Briefly, custom databases were created using the mesophilic genomes, and each cryophilic genome was compared to the mesophilic databases using stand-alone BLASTP (Camacho *et al.* 2009) with a cut-off e-value of 1e–15. The top BLASTP match protein sequence from each mesophile genome for each cryophilic protein was used for further analyses. Only non-hypothetical proteins and their mesophilic matches (full proteins) were included and assessed for cold adaptation at the AA level for the following indices previously associated with cold adaptation: arginine-to-lysine ratio; frequency of acidic, polar uncharged, and charged residues; proline, glycine, serine residues; aromaticity; aliphacity; and grand average of hydropathicity (GRAVY). The values for each AA trait were averaged from the selected mesophilic proteins, and this average was compared to the cryophilic protein using a one-sample t-test. Cryophilic proteins that were found to be significantly different from the average of the mesophilic proteins ($P < 0.05$) for each index were then assigned as 'cold adapted' or 'hot adapted' depending on the direction of change. We considered proteins cold adapted for each index if the direction of change was significantly lower for proline, R/K (arginine/lysine) ratio, acidic, charged, aliphatic, aromatic and hydrophobicity (GRAVY) indices and significantly higher for serine, glycine and polar uncharged indices (Table S3, Supporting Information). This analysis was repeated using the mesophile genomes, in the place of the cryophile genome, as controls and for comparison purposes. The proportions, in the total genome, of cold and hot adapted genes for the cryophiles and their mesophile controls were compared using a chi-square test with Bonferroni correction ($P \leq 0.0167$) to determine if the differences seen were significant. All AA analyses were carried out using in-house python scripts publicly available at https://github.com/ColdAdaptationScripts.

RESULTS AND DISCUSSION

Genome and strain properties

High-quality permanent drafts were obtained for all five cryophilic genomes (Table 2). *Kocuria* sp. KROCY2 had the smallest genome at 2.8 Mbp, while the largest genome (7.2 Mbp) belonged to *Methylobacterium* sp. EUR3 AL-11 (Table 2). *Polaromonas* sp. Eur3 1.2.1 and *Actinotalea* sp. KRMCY2 had genomes of similar size at 4.4 Mbp and 4.5 Mbp, respectively, while *Paenisporosarcina* sp. Eur1 9.01.10 has a slightly smaller genome (3.5 Mbp). Consistent with the genome sizes *Methylobacterium* sp. EUR3 AL-11 had the highest number of predicted genes with 7001, followed by *Polaromonas* sp. Eur3 1.2.1, *Actinotalea* sp. KRMCY2 and *Paenisporosarcina* sp. Eur1 9.01.10 with 4238, 4145 and 3438 genes, respectively, and lastly *Kocuria* sp. KROCY2 with 2480. Close to 80% of all protein coding genes could be assigned function predictions for all five organisms except for *Methylobacterium* sp. EUR3 AL-11, which was closer to 70% (Table 2). The GC content also varied greatly from 38% in *Paenisporosarcina* sp. Eur1 9.01.10 to 72% for *Actinotalea* sp. KRMCY2. *Methylobacterium* sp. EUR3 AL-11 and *Kocuria* sp. KROCY2 both had 71% GC content, very similar to *Actinotalea* sp. KRMCY2. *Polaromonas* sp. Eur3 1.2.1 GC content was slightly lower at 60%. While GC rich regions have been noted in some psychrophiles (De Maayer *et al.* 2014), overall, we found GC content in our cryophiles to be consistent with other members of their respective genera (Table S2, Supporting Information). Interestingly, *Polaromonas* sp. Eur3 1.2.1 had a fairly high % (~10%) of horizontally transferred genes compared to the other cryophiles (1%–4%) (Table 2). However, we found similar levels to that of *Polaromonas* sp. Eur3 1.2.1 in other members of this genus (Table S2, Supporting Information).

We attempted to determine subzero growth rates for all the cryophilic strains; however, although chosen because of their ability to grow on solid R2A agar supplement with 7% sucrose at −5°C (Steven *et al.* 2007a, 2008; Lacelle *et al.* 2011), we struggled to find liquid media adequate for growth at low temperatures in several strains. R2A media supplemented with 7% sucrose is susceptible to freezing in liquid form, and alternative liquid media tests have yet to be as successful for growth. In addition, we found both *Methylobacterium* sp. EUR3 AL-11 and *Actinotalea* sp. KRMCY2 susceptible to forming large clumps of cells in liquid cultures, further complicating growth curve measurements. Figure S1 (Supporting Information) shows growth on agar plates at subzero temperatures for those organisms for which obtaining subzero growth curves proved problematic. *Paenisporosarcina* sp. Eur1 9.01.10 grew well at −5°C with a generation time of 14.6 days, although it grew optimally at 25°C (Table S4, Supporting Information). This growth rate is similar to that observed for other cryophiles at −5°C (Mykytczuk *et al.* 2013; Goordial *et al.* 2016). Only *Polaromonas* sp. Eur3 1.2.1. exhibited characteristics of a stenopsychrophile with a faster generation time at 10°C (3.1 days) than at 22°C (3.4 days). It also showed similar generation times at 0°C (3.6 days) and was incapable of growth above 22°C (Table S4, Supporting Information). Several members of the genus *Sporosarcina* and *Polaromonas* are known to be psychrophilic, having been isolated from numerous polar

Table 2. Genome properties of sequenced permafrost strains.

Organism	*Polaromonas* sp. Eur3 1.2.1	*Actinotalea* sp. KRMCY2	*Kocuria* sp. KROCY2	*Methylobacterium* sp. EUR3 AL-11	*Paenisporosarcina* sp. Eur1 9.01.10
Name in JGI database	*Polaromonas* sp. Eur3 1.2.1	*Cellulomonas* sp. KRMCY2	*Microbacterium* sp. KROCY2	*Methylobacterium* sp. EUR3 AL-11	*Sporosarcina* sp. Eur3 2.2.2
Genome size (Mbp)	4.4	4.5	2.8	7.2	3.5
%GC	60	72	71	71	38
Gene count	4303	4207	2480	7001	3503
Protein coding genes	4238	4145	2415	6907	3438
16S rRNA	2	1	3	6	3
Protein coding genes with function prediction	3471	3205	1931	4913	2711
with enzymes	1234	1000	786	1393	905
with KEGG pathways	1281	1121	817	1653	962
with KEGG orthology	2235	1907	1305	2798	1821
with COG	2915	2687	1710	4206	2314
with Pfam	3582	3359	2014	5100	2850
Horizontally transferred count (percent %)	417 (9.69)	185 (4.4)	122 (1.74)	25 (1.01)	57 (1.63)
JGI Taxon ID	2619618817	2545824564	2540341240	2546826724	2528768230
NCBI taxon ID	1305734	1304865	1305732	1305730	1305836
Sequencing status	Permanent draft	Permanent draft	Permanent draft	Permanent draft	Permanent draft
Release date	30/06/2015	13/11/2013	24/09/2013	02/12/2013	08/08/2013
High quality	Yes	Yes	Yes	Yes	Yes
Bioproject accession	PRJNA195644	PRJNA195883	PRJNA195886	PRJNA195879	PRJNA195884

environments (Irgens, Gosink and Staley 1996; Yu *et al.* 2008; Margesin *et al.* 2012; Wang *et al.*, 2015, 2016; Yan, Xiao and Zhang 2016).

Genome comparisons and cold adaptation genes

All five organisms possess genes known to be important in cold adaptation and cold growth including cold shock proteins, RNA helicases, osmotic and oxidative stress, carotenoid synthesis, translation factors and membrane and peptidoglycan modifications (Table 3). Although the presence of these genes is not unique to psychrophiles, we also found them to be present in mesophiles, their abundance was higher in the cryophiles in many cases. *Polaromonas* Eur3 1.2.1 possesses 13 copies of universal stress protein (UspA family) genes, produced in response to numerous environmental stressors (Kvint *et al.* 2003), and 12 copies of Na+/H+ antiporters. While also present in mesophilic *Polaromonas* strains, we found the MnhD subunit of Na+/H+ antiporters to be higher in copy number in *Polaromonas* sp. Eur3 1.2.1 (six) than in mesophilic relatives (one/two) (Fig. 1A). Similarly, a Na+/melibiose symporter was higher in gene abundance in *Paenisporosarcina* sp. Eur1 9.01.10 than mesophiles (Fig. 1C). *Kocuria* sp. KROCY2, able to grow in up to 19% salt, possesses nine copies of choline-glycine betaine transporter (Table 3), for uptake of the compatible solute glycine betaine, several more than the mesophiles we used in comparative analyses (Fig. 1B). Four of the five cryophiles also possess the genes for synthesis of trehalose, another well-established compatible solute; 10 copies of the first enzyme in trehalose synthesis, trehalose-6-phosphate synthase, are found in the genome of *Actinotalea* sp. KRMCY2 (Table 3). Trehalose and glycine betaine are important compatible solutes in cold adapted organisms, and along with sodium transporters, have roles in osmoregulation and stability (Thomas, Kumar and Cavicchioli 2001; Doyle *et al.* 2012). Salt tolerance would be necessary for survival in permafrost habitats consisting of thin films of salty water. The permafrost cryophilic and halotolerant *Planococcus halocryophilus* is also known to have multiple copies of genes involved in glycine betaine transport (Mykytczuk *et al.* 2013).

All five strains have multiple copies of superfamily II DNA/RNA helicases, which includes the DEAD box helicases, shown to be crucial for stabilizing RNA at lower temperatures (Kuhn 2012). In addition, most cryophiles possessed toxin/antitoxin (TA) modules, with both *Actinotalea* sp. KRMCY2 and *Methylobacterium* sp. EUR3 AL-11 having copies of genes for TA systems not seen in their mesophilic counterparts (Fig. 1D and E). TA systems are important regulators of persistence and stress responses in certain bacteria, including *Mycobacterium tuberculosis* and *Escherichia coli* (Tiwari *et al.* 2015; Sauert *et al.* 2016), and one such system was found to be strongly induced at −10°C in *Planococcus halocryophilus* (Raymond-Bouchard *et al.* 2017). *Paenisporosarcina* sp. Eur1 9.01.10 and *Actinotalea* sp. KRMCY2 both have higher copy numbers of the DNA binding transcriptional regulators of the LacI/PurR family then the mesophiles we analyzed (Fig. 1C and E). This family contains several members and includes repressors of the lactose operon, purine nucleotide synthesis and trehalose operon. Interestingly, transposase gene numbers were higher in abundance in four of the five strains, *Polaromonas* sp. Eur3 1.2.1, *Kocuria* sp. KROCY2, *Actinotalea* sp. KRMCY2 and *Methylobacterium* sp. EUR3 AL-11 compared to their respective mesophilic relatives (Fig. 1A, B, D and E). Transposase activity has been shown to be linked to the stress response, and therefore, may be a mechanism in which these organisms gain genomic plasticity and a selective advantage necessary to survive and colonize extreme cold environments (de Saro *et al.* 2013; Shen *et al.* 2014). Increases in gene

Table 3. Gene functions and COG or Enzyme Commission (EC) categories present in psychrophilic genomes with known or predicted roles in cold adaptation and growth.

Category and Gene name	COG/EC ID	*Polaromonas* Eur3 1.2.1	*Actinotalea* KRMCY2	*Kocuria* KROCY2	*Methylobacterium* EUR3 AL-11	*Paenisporosarcina* Eur1 9.01.10
Cold shock, stress and HSP proteins						
Cold shock protein, CspA family	COG1278	1	1	2	6	4
Universal stress protein, UspA family	COG0589	13	0	1	4	3
Co-chaperonin GroES (HSP10)	COG0234	1	2	1	1	1
Chaperonin GroEL (HSP60 family)	COG0459	2	1	2	1	1
Molecular chaperone DnaK (HSP70)	COG0443	3	0	1	2	1
Molecular chaperone IbpA, HSP20 family	COG0071	5	1	0	4	0
Molecular chaperone, HSP90 family	COG0326	1	1	0	1	0
Molecular chaperone GrpE (heat shock protein)	COG0576	1	3	1	1	1
Ribosomal 50S subunit-recycling heat shock protein	COG1188	1	2	1	1	1
DNA replication and repair						
DNA gyrase/topoisomerase I, subunits A&B	COG0187; 0188	4	4	3	4	4
Recombinational DNA repair protein RecR	COG0353	1	0	1	1	1
RecA-superfamily ATPase, KaiC/GvpD/RAD55 family	COG0467	1	1	0	2	0
RecA/RadA recombinase	COG0468	1	2	1	1	1
DNA repair ATPase RecN	COG0497	1	6	1	1	1
Superfamily II DNA helicase RecQ	COG0514	1	0	2	2	2
Rad3-related DNA helicase	COG1199	3	4	1	0	2
Membrane and peptidoglycan alteration						
3-oxoacyl-(acyl-carrier-protein) synthase	COG0304	2	0	4	4	1
3-oxoacyl-[acyl-carrier-protein] synthase III (KASIII)	COG0332	1	1	3	2	1
3-oxoacyl-[acyl-carrier-protein] reductase	EC:1.1.1.100	12	3	3	6	5
Glycosyltransferase involved in cell wall biosynthesis	COG0438; 0463	17	10	19	39	7
3-hydroxyacyl-CoA dehydrogenase	COG1250	8	1	1	5	4
Fatty-acid desaturase	COG1398; 3239	1	1	1	0	1
D-alanyl-D-alanine carboxypeptidase	COG1686; 2027	2	1	1	4	3
Carotenoid biosynthesis						
Phytoene dehydrogenase-related protein	COG1233	0	2	3	4	1
Phytoene/squalene synthetase	COG1562	2	1	1	4	1
Polysaccharide capsule						
Capsular polysaccharide biosyn. protein EpsC	COG1086	1	0	1	1	1
Capsule polysaccharide export protein	COG3524	0	0	0	3	0
Capsular polysaccharide biosynthesis protein	COG3944; 4421	0	0	3	1	1
Exopolysaccharide biosynthesis protein	COG4632	0	1	0	0	0
Osmotic stress						
ABC proline/glycine betaine transport, ATPase component	COG1125; 4175	1	1	2	2	1
ABC proline/glycine betaine transport, permease	COG1174; 4176	2	0	4	3	1
Choline-glycine betaine transporter	COG1292	0	0	9	0	1
Osmoprotectant binding protein	COG1732	1	0	3	1	1
Choline dehydrogenase or related flavoprotein	COG2303	1	1	1	7	1
Trehalose-6-phosphate synthase	COG0380	2	10	1	1	0
Trehalose-6-phosphatase	COG1877	1	1	1	1	0
Maltooligosyltrehalose synthase	COG3280	0	1	1	1	0
Na+/proline symporter	COG0591	1	1	1	2	3
Na+/H+ antiporters	COGs: 3004; 1055; 0025; 0651; 1009; 1320	12	5	7	11	5
Oxidative stress						
Catalase	COG0753	2	0	2	1	3
Peroxiredoxin	COG0678; 1225	3	3	2	4	3
Glutathione peroxidase	COG0386	2	1	1	2	1
Spermidine synthase	COG0421	1	1	0	0	1
Thioredoxin reductase	COG0492	2	1	2	4	5
Glyoxylase or related hydrolase, β-lactamase superfam II	COG0491	6	4	4	9	7
Toxin/Antitoxin modules						
Ser/Thr kinase RdoA, MazF antagonist	COG2334	2	1	0	2	1

Table 3. Continued

Category and Gene name	COG/EC ID	*Polaromonas* Eur3 1.2.1	*Actinotalea* KRMCY2	*Kocuria* KROCY2	*Methylobacterium* EUR3 AL-11	*Paenisporosarcina* Eur1 9.01.10
Antitoxin component of MazEF	COG2336	1	1	0	0	0
mRNA-interferase, toxin component of MazEF	COG2337	2	1	0	1	1
mRNA interferase YafQ, toxin component of YafQ-DinJ	COG3041	2	1	0	2	0
Antitoxin component of RelBE or YafQ-DinJ	COG3077	1	0	0	2	0
Translation and transcription factors						
tRNA-dihydrouridine synthase	COG0042	3	1	1	2	2
tRNA A37 threonylcarbamoyladenosine dehydratase	COG1179	1	0	0	0	1
Translation elongation factor EF-Tu, GTPase	COG0050	2	1	1	2	1
Translation elongation factor EF-G, GTPase	COG0480	3	1	1	1	2
Translation initiation factor IF-2, GTPase	COG0532	1	1	1	1	1
Translation initiation factor IF-3	COG0290	1	1	1	1	1
Transcription antitermination factor NusA	COG0195	1	1	1	1	1
Transcription termination factor NusB	COG0781	1	2	1	1	1
Superfamily II DNA and RNA helicase	COG0513; 1061	7	4	5	6	3
Superfamily II DNA or RNA helicase, SNF2	COG0553	0	0	1	0	2
Superfamily II RNA helicase	COG4581	0	2	1	0	0

copy numbers have also been theorized to be an adaptive feature in and of itself by offering an advantage through selective expression of cold or hot adapted gene copies at different temperatures (Maki, Yoneta and Takada 2006; Mykytczuk *et al.* 2013). This is believed to confer advantages during cold growth given that selectively upregulated gene copies are likely to function better at their induced temperature, and thus allows organisms to have an expanded growth temperature range, such as *P. halocryophilus*, which grows from −15°C to 37°C.

AA adaptations

To examine AA substitutions that could favor increased flexibility, we performed genome wide comparative AA analyses using the predicted protein sequences for each cryophile. We determined the number of significantly cold or hot adapted genes in each cryophilic genome for specific AA indices associated with cold adaptation (Fig. 2 and Table S3, Supporting Information). Previous studies have used the ratio of cold/hot proteins as a standard to determine cold adaptation for a given trait and deemed those with cold/hot ratios >1, signifying enrichment of cold adapted proteins, as 'cold adapted' (Ayala-del-Río *et al.* 2010; Mykytczuk *et al.* 2013). For comparative purposes and as a control to check if these cold/hot ratios are specific to cold adaptation in cryophiles, we also determined the number of hot and cold adapted genes and cold/hot ratios for two mesophiles for each cryophile (Fig. 2 and Fig. S2, Supporting Information). As these results highlight, cold/hot ratios of >1 can be seen in all of the mesophiles for certain traits, and we therefore argue that using a ratio cutoff of 1 is not enough on its own to indicate cold adaptation for a given trait in psychrophilic genomes. We instead focused on determining significant differences in proportions of hot and cold genes in the cryophiles vs the mesophiles as determined by chi-square analysis with Bonferroni correction ($P \leq 0.0167$) used to correct for multiple comparisons (Fig. 2; Figs S2 and S3, Supporting Information). Of note, most of the proteins in both cryophile and mesophile genomes were classified as neutral with no significant changes in AA content (Fig. S3, Supporting Information). Cold/hot ratios were therefore determined for the remaining proteins that showed significant changes and could be classified as hot or cold adapted.

While we did not find the proportions of cold and hot adapted genes in the cryophilic genomes to be significantly different from the mesophiles for many of the amino acid traits, there were noteworthy exceptions. *Polaromonas* sp. Eur3 1.2.1 had significantly higher and lower proportions of cold and hot adapted genes (Table S3, Supporting Information), respectively, for the index proline then the mesophiles, with an overall much higher cold/hot ratio (Fig. 2A), indicative that more cryophilic proteins possess overall fewer proline residues than their mesophilic counterparts. *Actinotalea* sp. KRMCY2 was also significantly cold adapted for proline (Fig. 2C). Since prolines are covalently bonded to the nitrogen atom of the peptide group, they impose constraint on rotations of the peptide backbone and create rigid kinks in peptide chains, thus contributing to overall reduced flexibility of proteins. A reduction in proline residues has also been observed, when compared to mesophiles, in the genomes of the cold marine bacteria *Shewanella halifaxensis* and *Shewanella sediminis* (Zhao *et al.* 2010) and in a number of cold adapted enzymes, including the psychrophilic α-amylase from *Alteromonas haloplanctis* and the 3-isopropylmalate dehydrogenase from the eurypsychrophile bacterium *Vibrio* sp. 15 (Wallon *et al.* 1997; Aghajari *et al.* 1998).

Both *Kocuria* sp. KROCY2 and *Actinotalea* sp. KRMCY2 had higher proportions of cold adapted proteins for serine (i.e. more serine residues) than the mesophiles, while *Actinotalea* sp. KRMCY2 and *Methylobacterium* sp. EUR3 AL-11 had significantly higher proportions of cold adapted proteins for glycine (more glycine) (Fig. 2B, C and D). Due to its small size, glycine provides more conformational freedom, thus increasing flexibility, and serine, as a polar uncharged amino acid, may compensate partly for the general reduction in charged residues in psychrophilic proteins (Saunders *et al.* 2003). Serine and glycine have been shown to be higher in a number of psychrophilic proteins, including an Antarctic citrate synthase (Russell *et al.* 1998) and a DEAD-box RNA helicase (Linding *et al.* 2003). Increases in serine residues is one of the more commonly observed traits

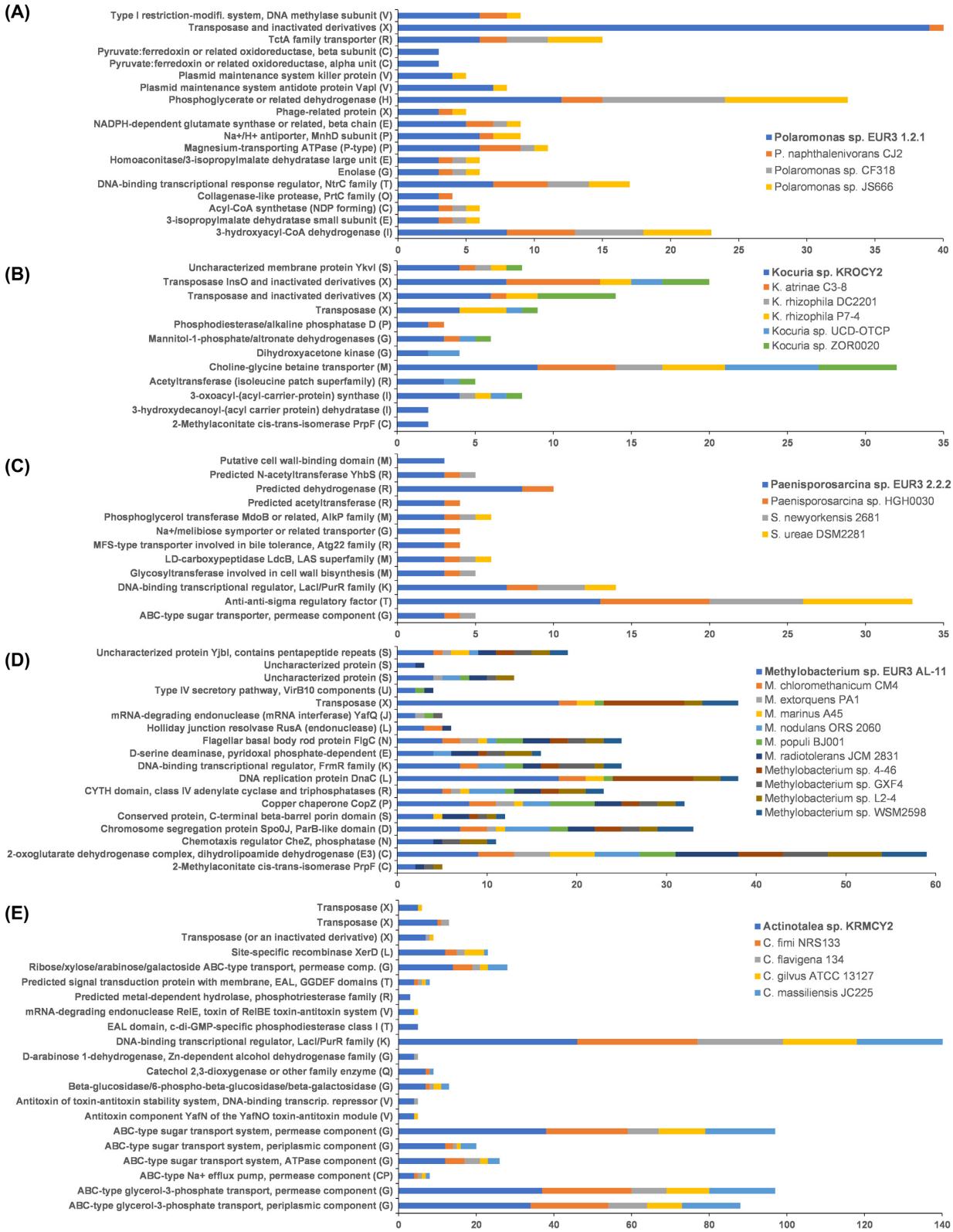

Figure 1. COG categories and gene functions with higher abundances of copies in the cryophilic genomes when compared to mesophiles. **A**) *Polaromonas* sp. Eur3 1.2.1; **B**) *Kocuria* sp. KROCY2; **C**) *Paenisporosarcina* sp. Eur1 9.01.10; **D**) *Methylobacterium* sp. EUR3 AL-11; and **E**) *Actinotalea* sp. KRMCY2. The blue column closest to the Y-axis represents copy numbers for the cryophiles in each case. Letters for the COG categories are given in parenthesis after the gene description: C) Energy production and conversion; D) Cell cycle control, cell division, chromosome partitioning; E) Amino acid transport and metabolism; G) Carbohydrate transport and metabolism; H) Coenzyme transport and metabolism; I) Lipid transport and metabolism; J) Translation, ribosomal structure and biogenesis; K) Transcription; L) Replication, recombination and repair; M) Cell wall/membrane/envelope biogenesis; N) Cell motility O) Post-translational modification, protein turnover, and chaperones; P) Inorganic ion transport and metabolism; R) General function prediction only; S) Function unknown T) Signal transduction mechanisms; U) Intracellular trafficking, secretion, and vesicular transport; V) Defense mechanisms; X) Mobilome: prophages, transposons.

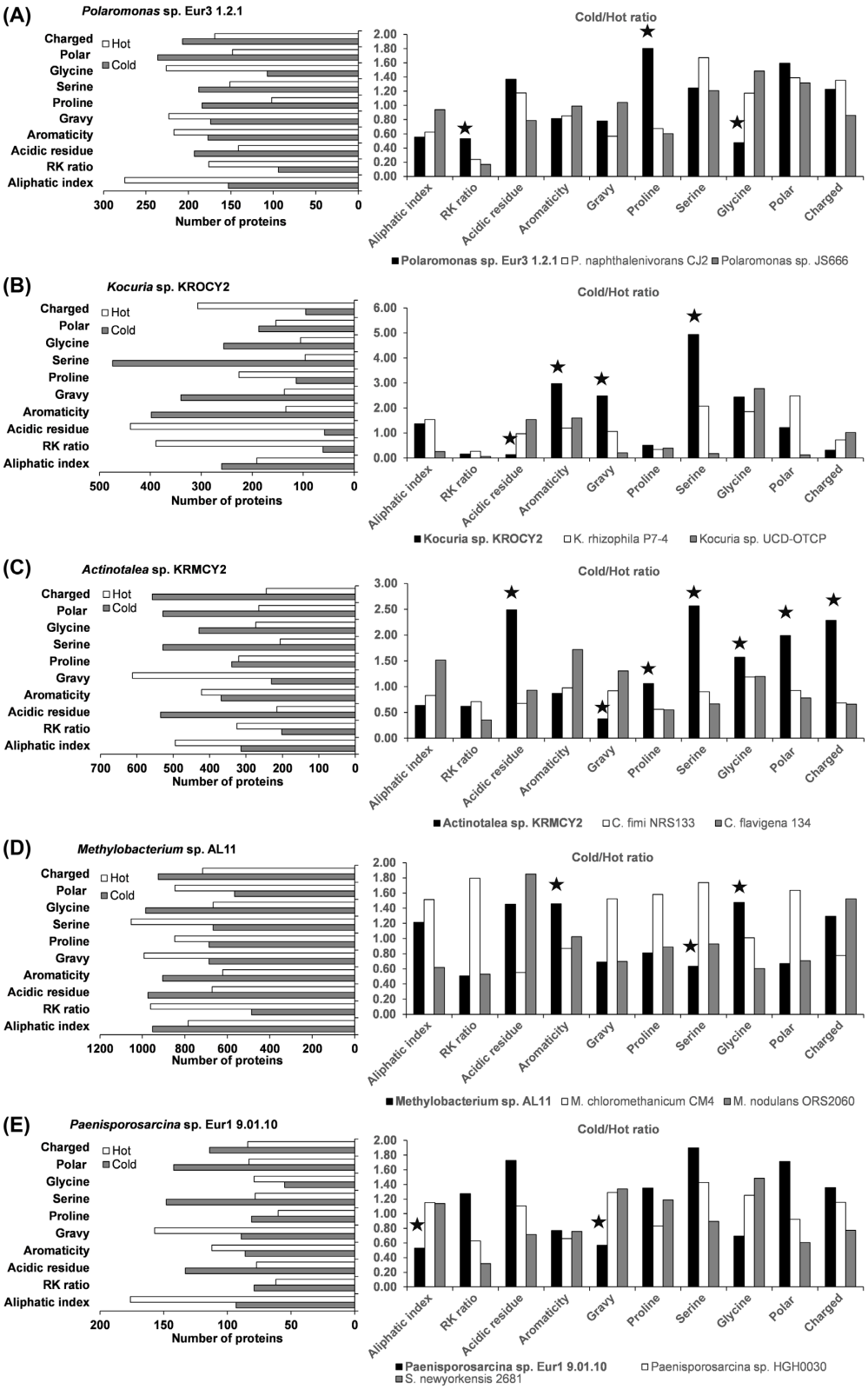

Figure 2. Amino acid traits of cold adaptation in cryophilic proteins. Left side graphs: significantly ($P < 0.05$) hot and cold adapted proteins identified in each cryophile (A-E) for each amino acid trait measured: aliphatic index, R/K ratio, acidic residues, aromaticity, GRAVY and proline, serine, glycine, polar and charged residues. Right side graphs: ratio of cold/hot adapted proteins identified in each cryophile (name bolded; black bar) and in two mesophilic relatives of each cryophile. Proportions of cold adapted proteins that were significantly different in the cryophile genome when compared to both mesophilic genomes are starred.

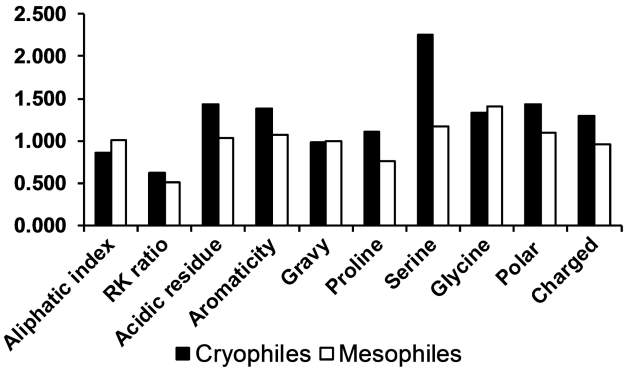

Figure 3. Average ratio of cold/hot adapted proteins for all cryophiles and all mesophiles for each amino acid trait measured.

and has also been found in a uracil-DNA-N-glycosylate (Ræder *et al.* 2008), a thermolysin (Adekoya *et al.* 2006), and the genomes of *Colwellia psychrerythraea* and marine *Glaciecola* species (Methé *et al.* 2005; Qin *et al.* 2013). Significant cold adaptation was also observed for aromaticity (fewer aromatic residues) in *Kocuria* sp. KROCY2 and *Methylobacterium* sp. EUR3 AL-11, and for the R/K ratio, with an overall lower number of arginines to lysines, in *Polaromonas* sp. Eur3 1.2.1. The bacterial lipase from the Antarctic psychrophile *P. immobilis* B10 features a very small number of aromatic-aromatic interactions (Arpigny, Lamotte and Gerday 1997), and proteins present at the cell surface of *Planococcus halocryophilus* at −10°C have significantly fewer aromatic residues then surface proteins present at higher temperatures (Ronholm *et al.* 2015).

In addition to serine and glycine, *Actinotalea* sp. KRMCY2 showed the greatest amount of significantly cold adapted indices with six out of the 10 being significantly higher for cold adaption from the mesophiles. However, for the other cryophiles, only a few of the indices showed evidence of cold adaptation, and in the case of *Paenisporosarcina* sp. Eur1 9.01.10, none showed significant cold adaptation (Fig. 2E). Many members of the genera *Paenisporosarcina* and *Sporosarcina* are psychrophilic, as discussed above, and therefore it is possible that most members of these genera possess characteristics of cold adaption, even in those characterized as mesophiles in the literature. This would explain why little to no significant cold adaptation is detected in these organisms.

By comparing the average cryophile ratios for each amino acid adaptive trait with the average of the mesophile ratios for those traits, we were able to highlight some differences in several amino acid indices. A comparison of all cryophiles with all mesophiles revealed higher average ratios of cold/hot adapted genes in the cryophiles for serine, and slightly higher for acidic, proline, polar, charged and R/K indices (Fig. 3 and Table S3, Supporting Information). We did not, however, find these results to be significantly different by t-test ($P \leq 0.05$), though the results were close for proline ($P = 0.07$), serine ($P = 0.1$) and acidic residues ($P = 0.1$). This lack of significant difference is in large part the result of noticeable variation in cold/hot ratios between the different genera and phyla for the specific amino acid indices we studied. Amino acid cold adaptive strategies may be taxa specific and vary greatly between individual proteins. Indeed, to date, not every amino acid feature discussed here has been found in a single genome or cold-active protein, rather most tend to possess only a few of these traits (Rodrigues and Tiedje 2008). Contradictory findings have also been reported with certain amino acid changes, such as glycine, appearing to be increased in some psychrophilic proteins (Russell *et al.* 1998; Linding *et al.* 2003) but decreased in certain psychrophilic genomes (Zhao *et al.* 2010). Thus, the adaptive strategies employed by psychrophiles may vary from organism to organism. These adaptations may not be present in all psychrophiles, as is the case with *Desulfotalea psychrophila*, which did not show detectable amino acid substitutions indicative of cold adaptation in its genome (Rabus *et al.* 2004), and in our case with *Paenisporosarcina* sp. Eur1 9.01.10. The cold-active isocitrate dehydrogenase from psychrophilic *Colwellia maris* also did not appear to show any characteristic amino acid changes associated with psychrophilic proteins (Watanabe *et al.* 2005).

Proteins with only a single or very few amino acid changes, which is enough to have significant impacts on function, would not be as easily detectable in these large-scale analyses since a single change is less likely to result in a protein begin detected as significantly cold adapted when compared to mesophiles. In addition, changes may, in some cases, only occur in a subset of proteins and not result of large-scale changes across the entire genome/proteome. The overall large number of proteins in this study that were neither cold or hot adapted would add weight to this theory. In *Psychrobacter arcticus*, noted adaptations appeared to be more prevalent in proteins involved in cell growth and reproduction (Ayala-del-Río *et al.* 2010). Changes may also be conformational in nature, affecting secondary, tertiary and quaternary structure, either through changes in location of the amino acids in the sequence, which would not affect overall amino acid counts, or through one or two amino acid changes, which would be difficult to detect in this analysis.

Lastly, and perhaps most importantly, the specific location of the amino acids, not simply total numbers, may be especially important. The citrate synthase from Antarctic strain DS2-3R showed a reduction in salt bridges and proline residues specifically in the loop regions of the protein, while increases in glycine and serine occurred close to catalytic sites (Russell *et al.* 1998). In the psychrophilic *Alteromonas haloplanctis* α-amylase, it is the core of the protein that shows reduced numbers of proline and arginine (Aghajari *et al.* 1998). Nevertheless, while we did not find significantly conserved traits across the cryophiles, nor observed noticeable differences in cold adaptation between cryophiles for many amino acid features, our results do highlight noteworthy trends in amino acid properties of cryophilic proteins, and did identify traits that are significantly different from mesophilic relatives for several of the cryophiles, especially with regards to glycine, proline and serine. Changes in these amino acids are amongst the most consistently observed cold adaptive traits in studies focusing on cold adaption of psychrophilic proteins, and these changes seem to span broad phyla.

CONCLUSION

We found evidence for the presence of many cold adaptation genes in the genomes of cryophiles reported in this study, as well as genomic redundancy (higher gene copy numbers), theorized to be an important feature of cold adapted organisms. We identified significant cold adaptation in the genomes, to maintain protein flexibility at low temperatures, in four of the five cryophiles for a subset of the amino acid traits we studied, including proline, glycine, serine and aromaticity. Comparing all cryophile genomes from our study to all mesophile genomes in our database, we found trends favoring lower proline and acidic residues, and higher serine content in cryophilic proteins. It is possible that changes in content of these specific amino acids may be especially important to increase protein flexibility during

subzero growth. Further comparisons between cryophiles, psychrophiles and mesophiles would need to be done to determine if this is the case. On the whole, however, we could not easily detect cold adaptations on a genome-wide scale for many of the amino acid indices we investigated, and we found a significant degree of variability between the genera we studied. This may indicate that, while some overlap exists, overall, cold-adapted microbes use different approaches to adapt to low temperatures. In addition, since amino acid adaptations may be specific to certain subsets or categories of proteins, we believe that individual comparative analyses focusing on single proteins and taking into account not only total amino acid numbers, but the location of specific amino acids within the sequence, the overall potential conformation of the protein, and, if possible, functional assays, are more likely to provide an accurate overall representation of cold adaption in psychrophilic proteins. Overall, this study has increased our understanding of the genetic basis of cold adaptation in permafrost microbes, and thus helps up better understand how these organisms behave and grow at subzero temperatures.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGEMENTS

Sequencing was performed as part of a U.S. Department of Energy Joint Genome Institute Community Sequencing Program Project (Quartely; ID 1015712). Thanks to Nicole Shapiro, Tanja Woyke, Marcel Huntemann, James Han, Amy Chen, Nikos Kyrpides, Victor Markowitz, Krishna Palaniappan, Natalia Ivanova, Natalia Mikhailova, Galina Ovchinnikova, Andrew Schaumberg, Amrita Pati, Dimitrios Stamatis, Tatiparthi Reddy, Henrik P. Nordberg, Michael N. Cantor, and Susan X. Hua of JGI.

FUNDING

This work was supported by fellowships from the Natural Sciences and Engineering Research Council (NSERC) CREATE Canadian Astrobiology Training Program (CATP) and NSERC Canadian Graduate Scholarship to IRB, and NSERC Discover Grant Program, Northern Research Supplement grant, and a Polar Continental Shelf Project grant to LGW. The work conducted by JGI is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Conflict of interest. None declared.

REFERENCES

Adekoya OA, Helland R, Willassen NP *et al.* Comparative sequence and structure analysis reveal features of cold adaptation of an enzyme in the thermolysin family. *Proteins Struct Funct Genet* 2006;**62**:435–49.

Aghajari N, Feller G, Gerday C *et al.* Structures of the psychrophilic alteromonas haloplanctis alpha-amylase give insights into cold adaptation at a molecular level. *Structure* 1998;**6**:1503–16.

Arpigny JL, Lamotte J, Gerday C. Molecular adaptation to cold of an antarctic bacterial lipase. *J Mol Catal B Enzym* 1997;**3**:29–35.

Ayala-del-Río HL, Chain PS, Grzymski JJ *et al.* The genome sequence of *Psychrobacter arcticus* 273–4, a psychroactive Siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. *Appl Environ Microbiol* 2010;**76**:2304–12.

Bakermans C, Bergholz PW, Rodrigues DF *et al.* Genomic and expression analyses of cold-adapted microorganisms. In: Miller RV , Whyte LG (eds). *Polar Microbiology: Life in a Deep Freeze*. Washington, DC, USA.: ASM Press, 2012, 126–55.

Bennett S. Solexa Ltd. *Pharmacogenomics* 2004;**5**:433–8.

Camacho C, Coulouris G, Avagyan V *et al.* BLAST plus: architecture and applications. *BMC Bioinformatics* 2009;**10**:1.

Cavicchioli R. Cold-adapted archaea. *Nat Rev Microbiol* 2006;**4**:331–43.

Cavicchioli R. On the concept of a psychrophile. *ISME J* 2016;**10**:793–5.

Chin C-S, Alexander DH, Marks P *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;**10**:563–9.

Collins T, Gerday C, Feller G. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev* 2005;**29**:3–23.

De Maayer P, Anderson D, Cary C *et al.* Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* 2014;**15**:508–17.

De Saro FJL, Gómez MJ, González-Tortuero E *et al.* The Dynamic Genomes of Acidophiles. In: Seckbach J , Oren A, Stan-Lotter H (eds). *Polyextremophiles : Life Under Multiple Forms of Stress*. Dordrecht, The Netherlands: Springer, 2013, 83–97.

Doyle S, Dieser M, Broemsen E *et al.* General characteristics of cold-adapted microorganisms. In: Miller RV, Whyte LG (eds). *Polar Microbiology: Life in a Deep Freeze*. Washington, DC, USA: ASM Press, 2012, 103–25.

Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* 2003;**1**:200–8.

Feller G, Lonhienne T, Deroanne C *et al.* Purification, characterization, and nucleotide sequence of the thermolabile alpha amylase from the antarctic psychrotroph alteromonas haloplanctis A23. *J Biol Chem* 1992;**267**:5217–21.

Georlette D, Jónsson ZO, Van Petegem F *et al.* A DNA ligase from the psychrophile *Pseudoalteromonas haloplanktis* gives insights into the adaptation of proteins to low temperatures. *Eur J Biochem* 2000;**267**:3502–12.

Gilichinsky D, Rivkina E, Shcherbakova V *et al.* Supercooled water brines within permafrost—an unknown ecological niche for microorganisms: a model for astrobiology. *Astrobiology* 2003;**3**:331–41.

Gilichinsky DA, Soina VS, Petrova MA. Cryoprotective properties of water in the Earth cryolithosphere and its role in exobiology. *Orig Life Evol Biosph* 1993;**23**:65–75.

Gnerre S, Maccallum I, Przybylski D *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 2011;**108**:1513–8.

Goordial J, Lamarche- G, Lay C *et al.* Left out in the cold: life in cryoenvironments. In: Seckbach J, Oren A, Stan-Lotter H (eds). *Polyextremophiles*. The Netherlands: Springer, 2012, 335–63.

Goordial J, Raymond-Bouchard I, Zolotarov Y *et al.* Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile *Rhodococcus* sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. *FEMS Microbiol Ecol* 2016;**92**:1–11.

Grzymski JJ, Carter BJ, Delong EF *et al.* Comparative genomics of DNA fragments from six antarctic marine planktonic bacteria comparative genomics of DNA fragments from six antarctic marine planktonic bacteria. *Appl Environ Microbiol* 2006;**72**:1532–41.

Huston AL, Methe B, Deming JW. Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. *Appl Environ Microbiol* 2004;**70**:3321–8.

Hyatt D, Chen G-L, Locascio PF *et al*. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119.

Irgens RL, Gosink JJ, Staley JT. *Polaromonas vacuolata* gen. nov., sp. nov., a psychrophilic, marine, gas vacuolate bacterium from antarctica. *Int J Syst Bacteriol* 1996;**46**:822–6.

Kuhn E. Toward understanding life under subzero conditions: the significance of exploring psychrophilic "cold-shock" proteins. *Astrobiology* 2012;**12**:1078–86.

Kvint K, Nachin L, Diez A *et al*. The bacterial universal stress protein: function and regulation. *Curr Opin Microbiol* 2003;**6**:140–5.

Lacelle D, Radtke K, Clark ID *et al*. Geomicrobiology and occluded O2–CO2–Ar gas analyses provide evidence of microbial respiration in ancient terrestrial ground ice. *Earth Planet Sci Lett* 2011;**306**:46–54.

Li M, Copeland A, Han J. DUK – a fast and efficient Kmer matching tool. *Lawrence Berkeley Natl Lab LBNL Pap LBNL-4516E-Poster p*. 2011, 1–3.

Linding R, Jensen LJ, Diella F *et al*. Protein disorder prediction: implications for structural proteomics. *Structure* 2003;**11**:1453–9.

Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1996;**25**:955–64.

Maki S, Yoneta M, Takada Y. Two isocitrate dehydrogenases from a psychrophilic bacterium, *Colwellia psychrerythraea*. *Extremophiles* 2006;**10**:237–49.

Margesin R, Spröer C, Zhang DC *et al*. *Polaromonas glacialis* sp. nov. and *Polaromonas cryoconiti* sp. nov., isolated from alpine glacier cryoconite. *Int J Syst Evol Microbiol* 2012;**62**:2662–8.

Markowitz VM, Mavromatis K, Ivanova NN *et al*. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009;**25**:2271–8.

Methé BA, Nelson KE, Deming JW *et al*. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci USA* 2005;**102**:10913–8.

Metpally RPR, Reddy BVB. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: insights into the molecular basis of cold adaptation of proteins. *BMC Genomics* 2009;**10**:11.

Mykytczuk NC, Foote SJ, Omelon CR *et al*. Bacterial growth at -15°C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J* 2013;**7**:1211–26.

Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 2013;**29**:2933–5.

Pati A, Ivanova NN, Mikhailova N *et al*. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010;**7**:455–7.

Pruesse E, Quast C, Knittel K *et al*. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007;**35**:7188–96.

Qin QL, Xie BB, Yu Y *et al*. Comparative genomics of the marine bacterial genus Glaciecola reveals the high degree of genomic diversity and genomic characteristic for cold adaptation. *Environ Microbiol* 2013;**16**:1642–53.

Rabus R, Ruepp A, Frickey T *et al*. The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* 2004;**6**:887–902.

Raymond-Bouchard I, Chourey K, Altshuler I *et al*. Mechanisms of subzero growth in the cryophile Planococcus halocryophilus determined through proteomic analysis. *Environ Microbiol* 2017;**19**:4460–4479.

Rodrigues DF, Tiedje JM. Coping with our cold planet. *Appl Environ Microbiol* 2008;**74**:1677–86.

Ronholm J, Raymond-Bouchard I, Creskey M *et al*. Characterizing the surface-exposed proteome of *Planococcus halocryophilus* during cryophilic growth. *Extremophiles* 2015;**19**:619–29.

Russell RJ, Gerike U, Danson MJ *et al*. Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure* 1998;**6**:351–61.

Ræder ILU, Leiros I, Willassen NP *et al*. Uracil-DNA N-glycosylase (UNG) from the marine, psychrophilic bacterium *Vibrio salmonicida* shows cold adapted features. *Enzyme Microb Technol* 2008;**42**:594–600.

Sauert M, Wolfinger MT, Vesper O *et al*. The MazF-regulon: a toolbox for the post-transcriptional stress response in *Escherichia coli*. *Nucleic Acids Res* 2016;**44**:6660–75.

Saunders NFW, Thomas T, Curmi PMG *et al*. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Res* 2003;**13**:1580–8.

Seckbach J. Life on the edge and astrobiology: who is who in the polyextremophiles world? In: Seckbach J, Oren A, Stan-Lotter H (eds). *Polyextremophiles: Life Under Multiple Forms of Stress*. Dordrecht, The Netherlands: Springer, 2013, 61–79.

Shen L, Yao T, Liu Y *et al*. Downward-shifting temperature range for the growth of snow-bacteria on glaciers of the Tibetan plateau. *Geomicrobiol J* 2014;**31**:779–87.

Steven B, Briggs G, McKay CP *et al*. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol Ecol* 2007a;**59**:513–23.

Steven B, Léveillé R, Pollard WH *et al*. Microbial ecology and biodiversity in permafrost. *Extremophiles* 2006;**10**:259–67.

Steven B, Niederberger TD, Bottos EM *et al*. Development of a sensitive radiorespiration method for detecting microbial activity at subzero temperatures. *J Microbiol Methods* 2007b;**71**:275–80.

Steven B, Pollard WH, Greer CW *et al*. Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high Arctic. *Environ Microbiol* 2008;**10**:3388–403.

Thomas T, Cavicchioli R. Effect of temperature on stability and activity of elongation factor 2 proteins from antarctic and thermophilic methanogens. *J Bacteriol* 2000;**182**:1328–32.

Thomas T, Kumar N, Cavicchioli R. Effects of ribosomes and intracellular solutes on activities and stabilities of elongation factor 2 proteins from psychrotolerant and thermophilic methanogens. *J Bacteriol* 2001;**183**:1974–82.

Tiwari P, Arora G, Singh M *et al*. MazF ribonucleases promote *Mycobacterium tuberculosis* drug tolerance and virulence in guinea pigs. *Nat Commun* 2015;**6**:6059.

Vester JK, Glaring MA, Stougaard P. Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. *Extremophiles* 2015;**19**:17–29.

Wallon G, Lovett ST, Magyar C *et al*. Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium *Vibrio* sp. I5 suggest reasons for thermal instability. *Protein Eng* 1997;**10**:665–72.

Wang J, Liu B, Liu G *et al*. Draft genome sequence of *Sporosarcina globispora* W 25 T (DSM 4), a psychrophilic bacterium isolated from soil and river water. *Genome Announc* 2015;**3**:2002–3.

Wang N, Xu B, Liu Y *et al*. *Polaromonas eurypsychrophila* sp. nov., isolated from an ice core. *Int J Syst Evol Microbiol* 2016;**66**:2497–501.

Watanabe S, Yasutake Y, Tanaka I *et al*. Elucidation of stability determinants of cold-adapted monomeric isocitrate dehydrogenase from a psychrophilic bacterium, *Colwellia maris*, by construction of chimeric enzymes. *Microbiology* 2005;**151**:1083–94.

Yan W, Xiao X, Zhang Y. Complete genome sequence of the *Sporosarcina psychrophila* DSM 6497, a psychrophilic Bacillus strain that mediates the calcium carbonate precipitation. *J Biotechnol* 2016;**226**:14–5.

Yang L-L, Tang S-K, Huang Y *et al*. Low temperature adaptation is not the opposite process of high temperature adaptation in terms of changes in amino acid composition. *Genome Biol Evol* 2015;**7**:3426–33.

Yoon SH, Ha SM, Kwon S *et al*. Introducing EzBioCloud: a taxonomically united database of 16S rRNA and whole genome assemblies. *Int J Syst Evol Microbiol* 2017;**67**:1613–7.

Yu Y, Xin YH, Liu HC *et al*. *Sporosarcina antarctica* sp. nov., a psychrophilic bacterium isolated from the Antarctic. *Int J Syst Evol Microbiol* 2008;**58**:2114–7.

Zhao JS, Deng Y, Manno D *et al*. *Shewanella* spp. genomic evolution for a cold marine lifestyle and in-situ explosive biodegradation. *PLoS One* 2010;**5**:e9109.