An inter-Order horizontal gene transfer event enables the catabolism of compatible solutes by *Colwellia psychrerythraea* 34H

R. Eric Collins[1,2] and Jody W. Deming[3]
[1]Department of Earth & Planetary Sciences, McGill University, Montréal, Canada
[2]Department of Natural Resource Sciences, McGill University, Montréal, Canada
[2]School of Oceanography, University of Washington, Seattle, USA

January 21, 2013

Abstract

Colwellia is a genus of mostly psychrophilic halophilic Gammaproteobacteria frequently isolated from polar marine sediments and sea ice. In exploring the capacity of *Colwellia psychrerythraea* 34H to survive and grow in the liquid brines of sea ice, we detected a duplicated 37 kbp genomic island in its genome based on abnormally high G+C content. This island contains an operon encoding for heterotetrameric sarcosine oxidase and is located adjacent to several genes used in the serial demethylation of glycine betaine, a compatible solute commonly used for osmoregulation, to dimethylglycine, sarcosine, and glycine. Molecular clock inferences of important events in the adaptation of *C. psychrerythraea* 34H to compatible solute utilization reflect the geological evolution of the polar regions. Validating genomic predictions, *C. psychrerythraea* 34H was shown to grow on defined media containing either choline or glycine betaine, and on a medium with sarcosine as the sole organic source of carbon and nitrogen. Growth by 8 of 9 tested *Colwellia* species on a newly developed sarcosine-based defined medium indicated that the ability to catabolize compatible solutes is widespread in the genus *Colwellia*. This capacity likely provides a selective advantage to *Colwellia* species in cold, salty environments like sea ice, and may have contributed to the ability of *Colwellia* to invade these extreme niches.

keywords:

compatible solutes, halophiles, psychrophiles, genomics

Introduction

Sea ice is an extreme environment characterized by the low temperature (−2 to −35 ℃) and high salinity (35 to 270‰) of its brine inclusions, where the microbes reside (Junge et al., 2001; Collins et al., 2008). High salinity is known to induce several mechanisms for osmotolerance in Bacteria, Archaea, and Eukarya, with one of the most common being the intracellular accumulation of up to molar quantities of compatible solutes like glycine betaine (Bremer and Kramer, 2000; Roberts, 2005). These compounds may also be released into sea ice brine channels as organisms adjust to seasonal or diurnal changes in brine salinity or following viral lysis or cell death. The ability to transport and metabolize compatible solutes for growth would provide a selective advantage in such an environment. Horizontal gene transfer, an important component of microbial evolution, allows bacteria to rapidly adapt to new environments by importing large fragments of DNA from other microorganisms. Although difficult to identify in situ, recent investigations of sea ice as a hotspot for horizontal gene transfer have focused on its high concentrations of bacteria, viruses, and extracellular free DNA, among other contributing factors (Collins and Deming, 2011a,b).

Colwellia is a genus of heterotrophic, halophilic marine Gammaproteobacteria containing 12 psychrophilic or psychrotolerant species (Deming et al., 1988), two of which are also barophilic. Isolates belonging to this genus have been cultured from sea ice, marine sediments, estuarine waters, and both epipelagic and hadalpelagic marine environments. The complete genome sequence of *Colwellia psychrerythraea* 34H, isolated from Arctic marine sediments (Huston et al., 2000), has been published (Methé et al., 2005). Prior work with *C. psychrerythraea* 34H describing cold-active enzymes (Huston et al., 2000, 2004), motility, amino acid incorporation, and increased production of EPS at subzero temperatures (Junge et al., 2003, 2006; Marx et al., 2009), and a psychrophilic phage-host system (Wells and Deming, 2006a,b; Colangelo-Lillis and Deming, 2013) is complemented by plentiful genomic evidence of adaptation to cold salty environments by this microorganism. In addition to genes involved in maintaining membrane fluidity at low temperature and the synthesis of extracellular polysaccharides and enzymes, the genome also encodes for the production, transport, and degradation of compatible solutes (Methé et al., 2005).

Compatible solutes are small, chemically inert neutral or zwitterionic molecules accumulated intracellularly by microorganisms to maintain turgor pressure and stabilize proteins in high osmolarity environments (Bremer and Kramer, 2000; Roberts, 2005). One of the most common compatible solutes is glycine betaine (*N,N,N*-trimethylglycine), which is used as an osmoprotectant by eukaryotes (including marine diatoms; Keller et al., 1999; Armbrust et al., 2004) and numerous bacteria (Imhoff and Rodriguez-Valera, 1984; Landfald and Strøm, 1986), which can accumulate intracellular concentrations of this compound on the order of 1 M. Glycine betaine can be produced from choline, a component of the common membrane lipid phosphotidylcholine (Sohlenkamp et al., 2000). The transport and conversion of choline to glycine betaine in bacteria is encoded by a widespread operon, *betTABI* (Bremer and Kramer, 2000).

While many bacteria are capable of importing and converting choline to glycine betaine for use as a compatible solute, few utilize it as an energy source. Glycine betaine degradation

proceeds by serial demethylation to dimethylglycine, sarcosine, and glycine, which can be further catabolized to serine and pyruvate. Although the metabolic pathway is conserved, there are multiple non-homologous proteins capable of performing these reactions. *Pseudomonas aeruginosa* PAO1 produces two demethylases that generate dimethylglycine (GbcAB) and sarcosine (DgcAB; Wargo et al., 2008), while alternate genes have been reported from *Sinorhizobium meliloti* (Smith et al., 1988) and *Arthrobacter globiformis* (Meskys et al., 2001).

Sarcosine can be demethylated to glycine by a monomeric sarcosine oxidase (as in *Bacillus spp.*; Trickey et al., 1999), or by a heterotetrameric sarcosine oxidase (TSOX), as described in *Corynebacterium* (Chlumsky et al., 1995). The genes encoding TSOX in *Corynebacterium* are found closely packed with serine hydroxymethyltransferase (*glyA*) and formyltetrahydrofolate deformylase (*purU*) in the order *glyA-soxBDAG-purU*. In the reaction of TSOX, sarcosine is demethylated to glycine while O_2 is reduced to H_2O_2 and formaldehyde is released, except in the presence of tetrahydrofolate, when 5,10-methylenetetrahydrofolate is released instead. The resulting compounds can be utilized by the cell for energetic or biosynthetic purposes: i.e. PurU catalyzes the recycling of 5,10-methylenetetrahydrofolate to tetrahydrofolate; GlyA interconverts glycine and L-serine. Energy in the form of reducing equivalents (i.e. NADH) can also be obtained by the conversion of formaldehyde to formate and then to CO_2 by formaldehyde dehydrogenase (FdhGBAD) and formate dehydrogenase, respectively.

Here we describe the horizontal transfer and partial duplication in *C. psychrerythraea* 34H of an operon responsible for the catabolism of sarcosine to L-serine. Genes enabling the importation and degradation of choline to glycine betaine and then sarcosine were also examined by similarity with experimentally verified genes in other microorganisms. A defined medium for the growth of *C. psychrerythraea* 34H, based on sarcosine, lactate, and vitamins (SLV), was developed on the basis of this genomic evidence. Numerous strains of *Colwellia* grew on this medium, indicating that compatible solute degradation is widespread in psychrophilic halophilic members of the genus *Colwellia* and may provide a selective advantage to these extremophiles in sea ice.

Methods

Phylogenomics

Microbial genomes encoding the glycine betaine catabolic pathway were investigated using tools available at the Joint Genome Institute Integrated Microbial Genomes (JGI-IMG) database v3.5 (`http://img.jgi.doe.gov`; Markowitz and Kyrpides, 2007), containing 3183 complete or partial genomes for Bacteria and Archaea. Genomic islands were detected using IslandPath (Hsiao et al., 2003).

Small subunit 16S ribosomal RNA gene sequences from each bacterial genome in JGI-IMG were imported into ARB (Ludwig et al., 2004) from version 111 of the SILVA database of quality-checked, pre-aligned rRNA gene sequences. Additional bacterial reference sequences were added, including those from *Colwellia* species. 16S rRNA gene sequences not yet avail-

able in SILVA were exported from JGI-IMG and aligned using the SINA Webaligner before being imported into ARB. For each subset of interest, highly variable base positions were trimmed from the nucleotide alignment (SILVA filter `pos_var_ssuref:bacteria`; mask '.') and a maximum likelihood phylogenetic tree was inferred with PhyML v20120412 (options: `model=TN93`, `base frequencies=ML`, `ts/tv=estimated`, `Pinv=estimated`). To reduce computation time, several representative sequences were selected from the large monophyletic genera (*Pseudomonas*, *Shewanella*, and *Vibrio*).

The IMG Homolog Toolkit was used to identify homologs of *C. psychrerythraea* 34H sarcosine oxidase subunit alpha (SoxA). A total of 426 proteins were identified with bit scores greater than 400 and lengths of at least 900 amino acids. The amino acid sequences were exported from IMG and aligned using Clustal Omega v1.0.3 (options: `--iter=5`). A maximum likelihood phylogenetic tree was inferred with PhyML v20120412 (options: `-d aa`) and imported into FigTree v1.3.1 for viewing. The multiple sequence alignment and phylogenetic inference were repeated for the Gammaproteobacterial subtree containing *C. psychrerythraea* 34H SoxA. The final tree included 47 Gammaproteobacterial SoxA sequences and 5 outgroup sequences. The genomic context of each homolog was examined at JGI-IMG; all 52 *soxA* homologs were in conserved gene arrangements of the form *soxBDAG*. Gene cassettes as determined by JGI-IMG are defined as consecutive ORFs separated by less than 300 bp each.

Molecular Evolution

To predict the time of introduction of the horizontally transferred sarcosine oxidase operon, calculations of reverse G+C content amelioration were performed using the techniques of Muto and Osawa (1987) and Lawrence and Ochman (1997), with the following changes. First, to avoid biases associated with non-independence of variables, GC3 was used as the independent variable rather than total G+C content. Second, non-linear functions relating GC3 to GC1 and GC2 were used to predict codon-position-specific G+C content. The functions were selected and their coefficients determined empirically using least-squares regression at the website `http://zunzun.com`, using a dataset of 3980 bacterial and archaeal genomes and plasmids available from JGI-IMG. To obtain 90% confidence intervals, 1000 bootstrap replicates of each sequence (resampled by codon) were reverse-ameliorated for 1000 time steps of 1 My each. Software to perform these calculations is freely available at `http://github.com/rec3141/gcamel`.

Divergence time estimates of paralogous genes were calculated using an empirical substitution rate of 0.91% per synonymous site per My or 0.045% per non-synonymous site per My (Lawrence and Ochman, 1997). Synonymous and non-synonymous substitution rates were calculated by the method of Goldman and Yang (1994) using the program PAL2NAL (Suyama et al., 2006).

Growth on compatible solutes

Defined medium containing sarcosine (N-methylglycine; 2 g L^{-1}) was tested for its ability to support growth of *C. psychrerythraea* 34H as the sole source of organic carbon and nitrogen. *C. psychrerythraea* 34H was first grown at −1 ℃ from duplicate glycerol stocks in 50 ml of $\frac{1}{2}\times$ Marine Broth 2216. Growth was monitored by optical density at 600 nm; 15 mL aliquots were taken when an OD_{600} of 0.4 was reached. To rinse the inoculum of undefined medium these aliquots were centrifuged at 1000× g for 10 min at 8 ℃; the supernatant was discarded, 15 mL of ASW (NaCl 23.4 g L^{-1}, $MgSO_4{\cdot}7H_2O$ 4.9 g L^{-1}, $CaCl_2{\cdot}2H_2O$ 1.1 g L^{-1}, KBr 0.2 g L^{-1}, KCl 0.75 g L^{-1}, $MgCl_2{\cdot}6H_2O$ 4.1 g L^{-1}) was added without re-suspending the pellet and the tubes were centrifuged again at 1000× g for 2 min. The supernatant was again discarded, a final 15 mL of ASW was added and the pellet re-suspended. To three 50 mL aliquots of each experimental medium (and a positive control of $\frac{1}{2}\times$ Marine Broth 2216) was added 0.5 mL of rinsed inoculum, of which two were incubated at 4 ℃ and one at room temperature as a negative control.

Additional media were developed to increase the growth rate of *C. psychrerythraea* 34H on compatible solutes (choline, glycine betaine, and sarcosine) with the addition of lactate and vitamins (media denoted CLV, GBLV, and SLV, respectively). The growth of *C. psychrerythraea* 34H was tested on all three media; 9 additional *Colwellia* isolates were tested for growth on SLV. The recipe is as follows: choline *or* glycine betaine *or* sarcosine (2 g L^{-1}) and calcium lactate (0.5 g L^{-1}) are added to 1 L of TAPSO-buffered ASW and autoclaved. On cooling, 100× RPMI-1640 vitamin solution (D-biotin 0.02 g L^{-1}, choline chloride 0.3 g L^{-1}, folic acid 0.1 g L^{-1}, myo-inositol 3.5 g L^{-1}, niacinamide 0.1 g L^{-1}, p-amino benzoic acid 0.1 g L^{-1}, D-pantothenic acid 0.025 g L^{-1}, riboflavin 0.02 g L^{-1}, thiamine·HCl 0.1 g L^{-1}, vitamin B-12 0.0005 g L^{-1}, KCl 0.2 g L^{-1}, KH_2PO_4 0.2 g L^{-1}, NaCl 8 g L^{-1}, Na_2HPO_4 1.15 g L^{-1}) is added to a final concentration of 0.5 mL L^{-1}. To make solid medium 4 g L^{-1} ultra-pure agarose may be added before autoclaving.

Results and Discussion

Comparative Genomics

Compatible solute catabolic pathway

The genome of *C. psychrerythraea* 34H contained genes encoding for importation and serial demethylation of choline to glycine betaine (*betTABI*; CPS_4009–CPS_4012, another copy was found at CPS_1332–CPS_1335), then to dimethylglycine (*gbcAB*; CPS_4029–CPS_4030), and finally to sarcosine (*soxBDAG*, CPS_4016–CPS_4017, another copy was found at CPS_2476–CPS_2473; Fig. 1). Directly flanking both sets of TSOX genes were homologs of *glyA* and *purU*, each of which was predicted to belong to the *soxBDAG* operon in the arrangement *glyA-soxBDAG-purU*. Both of these enzymes act directly on the products of TSOX: serine hydroxymethyltransferase (GlyA) catalyzes the reversible conversion of glycine to L-serine, while formyltetrahydrofolate deformylase (PurU) catalyzes the conversion of

5,10-methylenetetrahydrofolate to tetrahydrofolate and formate (Fig. 1). Genes encoding formaldehyde dehydrogenase (CPS_4039) and formate dehydrogenase (*fdhGBAD*; CPS_4022–CPS_4026, another copy was found at CPS_2056–CPS_2060) were also co-located on the chromosome, and perform further reactions on the products of glycine betaine degradation. Thus, the net reaction encoded by *C. psychrerythraea* 34H is the importation and complete degradation of choline, and potentially any intermediates, to L-serine (via glycine) and carbon dioxide (via formate).

Compatible solute transporters

Several transporters implicated in the importation of quaternary amine compounds were previously identified in the genome of *C. psychrerythraea* 34H (Methé et al., 2005), none of which were highly similar to previously characterized compatible solute transporters like ProU1 from *Vibrio parahaemolyticus*, OpuABC from *Pseudomonas aeruginosa* PAO1, or OpuAC from *Listeria monocytogenes*. Of the six predicted transporters, one was annotated as a primary transporter of the ATP-binding cassette (ABC) family and five were annotated as secondary transporters of the betaine-choline-carnitine transporter (BCCT) family. Two additional transporters were located in the vicinity of compatible solute degradation genes, but their substrates are as yet unknown: CPS_2472, an outer membrane porin, and CPS_2476, an antiporter of the NhaC family.

While the putative proline/glycine betaine ABC transporter (CPS_4933–CPS_4935) was highly similar to a pair of ABC transporters from *P. aeruginosa* PAO1 (PA5096, PA5103), neither of these transporters were detected in this pseudomonad after induction by growth on glycine betaine (Diab et al., 2006) or by osmotic shock (Aspedon et al., 2006), nor were they required for growth on glycine betaine (Wargo et al., 2008). The substrates for these transporters are not yet known.

The five putative BCCT transporters fell into three distinct classes based on sequence similarity. Two members of the *betT* class (CPS_1335 and CPS_4009) were 80% identical at the amino acid level, located within *betTABI* operons, which convert choline to glycine betaine. These transporters are predicted to import choline. The second class included two genes, CPS_2003 and CPS_4027, which were 43% identical at the amino acid level. Several homologs (52–57% amino acid identity) in marine Gammaproteobacteria belonged to BCCT transporters implicated in the metabolism of dimethylsulphoniopropionate (DMSP; Todd et al., 2007; Johnston et al., 2008). In their respective genomes, these homologs are found as part of the *ddd* operon, which converts DMSP to dimethylsulfide (DMS), but in *C. psychrerythraea* 34H each of these genes stands alone. There are no other indications that *C. psychrerythraea* 34H metabolizes DMSP or DMS (e.g. it lacks the *ddd* operon), so the substrate for this class of BCCT transporters in *C. psychrerythraea* 34H remains speculative at this time. The final class of BCCT transporter in *C. psychrerythraea* 34H was represented only by CPS_3860. Homologs of this gene were found in other marine bacterial genomes, but no specific association with compatible solutes has been demonstrated.

Regulation of compatible solute metabolism

The expression of a niche-specific pathway, like compatible solute metabolism, in a generalist like *C. psychrerythraea* 34H can be expected to be well regulated. A number of putative regulators of choline and sarcosine metabolism were detected in the genome of *C. psychrerythraea* 34H, including a putative homolog of *gdbR* (CPS_4021; 59% identity), an AraC-type transcriptional regulator known from *Pseudomonas aeruginosa* PAO1. This protein has been shown to control the expression of *gbcAB* and *dgcAB* (Wargo et al., 2009). Immediately following each *soxBDAG* operon were found DNA-binding proteins (CPS_2482 and CPS_4036) that could also be involved in differential regulation of sarcosine oxidase activity.

Another class of regulators, acting via cyclic diguanylate (c-di-GMP), have been shown to control motility, attachment, EPS production and biofilm formation in a number of Gammaproteobacteria, including various species of *Pseudomonas* (Gjermansen et al., 2006), *Vibrio* (Beyhan et al., 2008; Ferreira et al., 2008), and *Shewanella* (Thormann et al., 2006). *C. psychrerythraea* 34H encodes 65 regulatory elements associated with c-di-GMP — more than 90% of all sequenced Gammaproteobacteria. The duplicated *glyA-soxBDAG-purU* operon may be in part regulated by a GGDEF domain protein (CPS_2484), based on the presence of the gene within the *glyA-soxBDAG-purU* gene cassette. Most of the known proteins containing GGDEF/EAL domains, including CPS_2484, contain transmembrane domains and signal reception and transduction domains with which they sense the external chemical environment.

Phylogenomics

Phylogenetic distribution of glycine betaine catabolic genes

The use of glycine betaine as a compatible solute is widespread in Bacteria. More than 40% of Gammaproteobacteria in JGI-IMG encode homologs of BetABI (responsible for the importation and conversion of choline to glycine betaine) including many members of the Orders Alteromonadales, Pseudomonadales, Oceanospirillales, and Vibrionales. In this study, we determined that the capacity to degrade this compatible solute is rarer — less than 25% of glycine betaine utilizers also encode homologs of the catabolic pathway: glycine betaine demethylase (GbcAB), dimethylglycine demethylase (DgcAB), and sarcosine oxidase (SoxBDAG). Most of these glycine betaine degraders are Pseudomonadales and Oceanospirillales. We found 46 genomes of Gammaproteobacteria encoding SoxBDAG homologs: 34 members of the genus *Pseudomonas*; 10 marine halophiles of the Oceanospirillales; 1 *Vibrio*, and *C. psychrerythraea* 34H, an Alteromonadales.

Several lines of evidence indicated that *soxBDAG* and other genes in the *C. psychrerythraea* 34H genome were horizontally transferred, probably from marine Gammaproteobacteria of the Order Oceanospirillales. The complete absence of the glycine betaine degradation pathway in existing genomes of most genera belonging to the Alteromonadales suggests that the genes were not simply retained in *C. psychrerythraea* 34H while being lost in all of the rest. Multiprotein phylogenetic trees consistently show an ancient divergence between one group comprised of Pseudomonadales and Oceanospirillales and a second

group of Vibrionales, Alteromonadales, Aeromonadales, and Pastuerellales (Williams et al., 2010). We suggest that glycine betaine degradation is an ancient trait of the Pseudomonadales/Oceanospirillales clade, which has been horizontally transferred into the genomes of (at least) two members of the Vibrionales/Alteromonadales/Aeromonadales/Pastuerellales group since the divergence. The mechanism of horizontal gene transfer by which *C. psychrerythraea* 34H acquired its exogenous genes was not determined in this study, but the likelihood that *C. psychrerythraea* 34H is a lysogen has been explored by Wells (2006), suggesting transduction as a possible mechanism.

Phylogenetic analysis of SoxA homologs demonstrated that the nearest sequenced homolog of *C. psychrerythraea* 34H SoxA was in *Vibrio carribbenthicus*, though the complete absence of SoxA in dozens of other *Vibrio* genomes suggests that this homolog was also acquired by horizontal gene transfer (Fig. 2). Poor phylogenetic resolution at interior nodes of the SoxA tree make deeper relationships murky. The most taxonomically similar genome to contain *soxBDAG* was *Marinobacter* sp. ELB17, formally described as Order Alteromonadales but phylogenetically more similar to Order Oceanospirillales (Fig. 2 and Williams et al., 2010). The remainder of the SoxA homologs belonged to members of the Pseudomonadales and Oceanospirillales.

Evolutionary history of compatible solute utilization in *C. psychrerythraea* 34H

The acquisition of compatible solute utilization genes in *C. psychrerythraea* 34H appears to have been a stepwise process occurring over the course of the last several hundred million years of Earth history. While much of the Phanerozoic Eon (541 Mya to today) was warmer than the present day and lacked polar ice caps, the inferred timing of significant events in the evolution of *C. psychrerythraea* 34H reflect geological periods when polar ice existed.

Two copies of the *betTABI* operon were located on the *C. psychrerythraea* 34H chromosome. The gene clusters (CPS_1337–CPS_1332 and CPS_4007–CPS_4012) appeared as ancient paralogs in phylogenetic analyses (not shown). Divergence time estimates of individual genes in the cluster ranged from 236 to 1879 Mya, with a median age of 298 My (Table 1). This timing coincides with that of the Karoo glaciation, which had a maximal extent around 300 My ago (Crowley and Baum, 1992).

A 37 kbp genomic island (base pairs 4,236,340–4,273,184; loci CPS_4030–CPS_4060) was identified based on abnormal dinucleotide bias, and G+C content (45–49%) exceeding 2 standard deviations from the mean genomic G+C content (38.0±3.5%). The genomic island contained the *glyA-soxBDAG-purU* operon and other genes involved in 1-carbon metabolism. An analysis of the amelioration of the G+C content of the 2435 codons encoding the *glyA-soxBDAG-purU* operon predicted that the G+C content of the donor microorganism was about 55% (90% CI: 53–57%), similar to that of *Marinobacter* sp. ELB17 (54%). Additionally, the G+C amelioration analysis inferred an estimated time since introduction of 150 My (90% CI: 70–220 My). This timing approximates a period of cooling in the Late-Jurassic/Early-Cretaceous (about 160 My ago) that may have precipitated a brief ice age in an otherwise warm era (Dromart et al., 2003).

Two copies of the *glyA-soxBDAG-purU* operon were located on the *C. psychrerythraea* 34H

chromosome. The first set was located within the previously described genomic island; the second set was duplicated from the first in the more recent past, along with a DNA-binding protein (CPS_2483 and CPS_4037). Of the 70 nucleotide substitutions observed in the 7561 bp region encoding *glyA-soxBDAG-purU*, most occurred at the edges of the operon, within *glyA* and *purU*. Only two substitutions were observed throughout *soxBDAG*, both synonymous substitutions in *soxG*. The DNA-binding protein exhibited a greater substitution rate (56 synonymous and 16 non-synonymous changes). Based on the observed substitution rates, this duplication event likely occurred within the last few millions years (Table 1) and thus since the re-establishment of polar ice caps in the Miocene (Zachos et al., 2001).

Any method of molecular clock dating is subject to error and the ages described here are no exception. However, the order of the described events is much more reliable, is internally consistent, and is logical. The reconstructed order of events is as follows: originally, *C. psychrerythraea* 34H, like most extant Alteromonadales, utilized glycine betaine as an osmoprotectant but was not able to degrade it. At some point, the *betTABI* operon was duplicated, perhaps allowing increased or fine-tuned expression of glycine betaine use, or neofunctionalization to allow the importation of related compounds. Next, the genomic island encoding sarcosine oxidase was horizontally transferred into the genome. We did not attempt to date the origin of the glycine betaine demethylase (GbcAB) or dimethylglycine demethylase (DgcAB), but their co-location on the genome suggests they may have arrived coincident with the genomic island. At this point, *C. psychrerythraea* 34H would have gained the ability to degrade glycine betaine. Finally, much later, the sarcosine oxidase operon was duplicated; it has not yet had time to functionally diverge. A future line of research will be to determine the role these paralogs play in the adaptation of *C. psychrerythraea* 34H to cold, salty environments.

Phenotypes

A defined medium for *C. psychrerythraea* 34H

Perhaps the most important qualification to define a successful horizontal gene transfer event is endogenous expression of the transferred genes in the recipient. Sarcosine was tested as a sole source of organic carbon and nitrogen for *C. psychrerythraea* 34H growth. Turbid growth ($OD600 \gtrsim 0.1$) was observed in duplicate aliquots of the sarcosine medium after 5 weeks. Successful growth after two consecutive transfers in the sarcosine medium demonstrated that growth was not dependent on co-factors remaining from prior growth in undefined medium. In an effort to increase the growth rate of *C. psychrerythraea* 34H on sarcosine we amended the medium with calcium lactate, vitamin solution, or a combination of both. We observed growth on the combined medium, denoted SLV (for sarcosine, lactate, and vitamins) after one week of incubation at 4 ℃. After 5 weeks no growth of *C. psychrerythraea* 34H was detected in ASW containing only lactate or only vitamin solution without sarcosine. Time required for growth to turbidity of *C. psychrerythraea* 34H in SLV at temperatures between -1 ℃ and 8 ℃ was comparable to growth on $\frac{1}{2}\times$ Marine Broth 2216 at the same temperatures. Defined media containing lactate, vitamins, and either choline (CLV) or glycine betaine (GBLV) were

also tested for their ability to support growth of *C. psychrerythraea* 34H; growth to turbidity was observed in each medium within ten days at 4 ℃.

Degradation of compatible solutes by *Colwellia* spp.

To determine the taxonomic range of growth on sarcosine among *Colwellia spp.*, we tested the ability of 9 isolates to grow in liquid SLV at 8 ℃. Growth on SLV was widespread throughout *Colwellia* (Fig. 3). All but one (*C. rossensis* ACAM 608) grew on SLV and became turbid within one week. Three strains had no available sequence data but were phenotypically identified as *Colwellia spp.*: *C. demingiae* strain ICP10 (3 days to turbidity; J. Bowman, UT), Arctic sea ice isolate 21C (3 days to turbidity; Borriss et al., 2003), and Arctic nepheloid layer isolate 75C3 (7 days to turbidity; Wells, 2006). Four *Colwellia* strains tested for growth on SLV were originally isolated from Antarctic sea ice diatom assemblages (Bowman et al., 1998). Three of these strains grew on SLV: *C. demingiae* ACAM 459T, *C. psychrotropica* ACAM 179T, and *C. hornerae* ACAM 607T. Sources of compatible solutes in sea ice are likely plentiful, including compounds released by diatoms and the lysis of bacteria by phage, which are well-known from sea ice (Maranger et al., 1994; Wells and Deming, 2006a). The ability of *Colwellia spp.* to grow on compatible solutes as a sole source of organic carbon and nitrogen suggests the possibility of bringing into culture as-yet-uncultured microorganisms from environments expected to contain high levels of compatible solutes, including sea ice and other frozen saline environments.

Conclusions

In a previous analysis of the genome of *C. psychrerythraea* 34H, a psychrophilic, halophilic marine bacterium, Methé et al. (2005) identified a duplicated operon encoding for heterotetrameric sarcosine oxidase (SoxBDAG), an enzyme involved in the catabolism of glycine betaine, a common osmoprotectant molecule. In the present genomic analysis we have identified this operon as the result of a horizontal gene transfer event. Molecular evolutionary estimates of important events in the adaptation of *C. psychrerythraea* 34H to compatible solute utilization appear to recapitulate the geological evolution of the polar regions. We have also demonstrated the genetic potential for *C. psychrerythraea* 34H to use both choline and glycine betaine as substrates for growth. Furthermore, we successfully grew *C. psychrerythraea* 34H on a medium containing sarcosine as the sole source of carbon and energy, which was then developed into SLV, a defined medium for the rapid growth of many *Colwellia* species. With these developments, *Colwellia* species, which previously could only be grown rapidly in Marine Broth, can be included in more defined and sophisticated studies of its strategies to survive and compete in extreme environments.

Acknowledgments

We thank S. Carpenter for laboratory analyses and G. Rocap for insightful discussions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada. Portions of this work were supported by an NSF-IGERT award to the University of Washington, by NSF-OPP award #0520297 (JWD) and by the Walters Endowed Professorship (JWD).

References

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Krger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86

Aspedon A, Palmer K, Whiteley M (2006) Microarray analysis of the osmotic stress response in *Pseudomonas aeruginosa*. J Bacteriol 188:2721–2725

Beyhan S, Odell LS, Yildiz FH (2008) Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in *Vibrio cholerae*. J Bacteriol 190:7392–7405

Borriss M, Helmke E, Hanschke R, Schweder T (2003) Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice. Extremophiles 7:377–384

Bowman JP, Gosink JJ, McCammon SA, Lewis TE, Nichols DS, Nichols PD, Skerratt JH, Staley JT, McMeekin TA (1998) *Colwellia demingiae sp. nov.*, *Colwellia hornerae sp. nov.*, *Colwellia rossensis sp. nov.* and *Colwellia psychrotropica sp. nov.*: psychrophilic Antarctic species with the ability to synthesize docosahexaenoic acid (22:6 omega 3). Int J Syst Bacteriol 48:1171–1180

Bremer E, Kramer R (2000) Coping with osmotic challenges: Osmoregulation through accumulation and release of compatible solutes in Bacteria. In: Storz G, Hengge-Aronis R (eds.) Bacterial Stress Responses, ASM Press. 79–97

Chlumsky LJ, Zhang L, Jorns MS (1995) Sequence analysis of sarcosine oxidase and nearby genes reveals homologies with key enzymes of folate one-carbon metabolism. J Biol Chem 270:18252–18259

Colangelo-Lillis J, Deming J (2013) Genomic analysis of cold-active Colwelliaphage 9A and psychrophilic phage-host interactions. Extremophiles 17:99–114

Collins R, Deming J (2011a) Abundant dissolved genetic material in Arctic sea ice Part I: Extracellular DNA. Polar Biol 34:1819–1830

Collins R, Deming J (2011b) Abundant dissolved genetic material in Arctic sea ice Part II: Viral dynamics during autumn freeze-up. Polar Biol 34:1831–1841

Collins RE, Carpenter SD, Deming JW (2008) Spatial heterogeneity and temporal dynamics of particles, bacteria, and pEPS in Arctic winter sea ice. J Mar Sys 74:902–917

Crowley T, Baum S (1992) Modeling late Paleozoic glaciation. Geology 20:507–510

Deming JW, Somers LK, Straube WL, Swartz DG, Macdonell MT (1988) Isolation of an obligately barophilic bacterium and description of a new genus, *Colwellia gen. nov.* Syst Appl Microb 10:152–160

Diab F, Bernard T, Bazire A, Haras D, Blanco C, Jebbar M (2006) Succinate-mediated catabolite repression control on the production of glycine betaine catabolic enzymes in *Pseudomonas aeruginosa* PAO1 under low and elevated salinities. Microbiology 152:1395–1406

Dromart G, Garcia JP, Picard S, Atrops F, Lecuyer C, Sheppard S (2003) Ice age at the Middle–Late Jurassic transition? Earth Planet Sci Lett 213:205–220

Ferreira RBR, Antunes LCM, Greenberg EP, McCarter LL (2008) *Vibrio parahaemolyticus* ScrC modulates cyclic dimeric GMP regulation of gene expression relevant to growth on surfaces. J Bacteriol 190:851–860

Gjermansen M, Ragas P, Tolker-Nielsen T (2006) Proteins with GGDEF and EAL domains regulate *Pseudomonas putida* biofilm formation and dispersal. FEMS Microbiol Lett 265:215–224

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Hsiao W, Wan I, Jones S, Brinkman F (2003) IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics 19:418–420

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

Huston AL, Methé B, Deming JW (2004) Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. Appl Environ Microbiol 70:3321–3328

Imhoff JF, Rodriguez-Valera F (1984) Betaine is the main compatible solute of halophilic eubacteria. J Bacteriol 160:478–479

Johnston AWB, Todd JD, Sun L, Nikolaidou-Katsaridou MN, Curson ARJ, Rogers R (2008) Molecular diversity of bacterial production of the climate-changing gas, dimethyl sulphide, a molecule that impinges on local and global symbioses. J Exp Bot 59:1059–1067

Junge K, Krembs C, Deming J, Stierle A, Eicken H (2001) A microscopic approach to investigate bacteria under in situ conditions in sea-ice samples. Annal Glaciol 33:304–310

Junge K, Eicken H, Deming JW (2003) Motility of *Colwellia psychrerythraea* strain 34H at subzero temperatures. Appl Environ Microbiol 69:4282–4284

Junge K, Eicken H, Swanson BD, Deming JW (2006) Bacterial incorporation of leucine into protein down to −20 degrees C with evidence for potential activity in sub-eutectic saline ice formations. Cryobiology 52:417–429

Keller MD, Kiene RP, Matrai PA, Bellows WK (1999) Production of glycine betaine and dimethylsulfoniopropionate in marine phytoplankton. I. Batch cultures. Mar Biol 135:237–248

Landfald B, Strøm AR (1986) Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. J Bacteriol 165:849–855

Lawrence J, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 44:383–397

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32:1363–1371

Maranger R, Bird DF, Juniper SK (1994) Viral and bacterial dynamics in Arctic sea-ice during the spring algal bloom near Resolute, NWT, Canada. Mar Ecol Prog Ser 111:121–127

Markowitz VM, Kyrpides NC (2007) Comparative genome analysis in the integrated microbial genomes (IMG) system. Methods Mol Biol 395:35–56

Marx JG, Carpenter SD, Deming JW (2009) Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. Can J Microbiol 55:63–72

Meskys R, Harris RJ, Casaite V, Basran J, Scrutton NS (2001) Organization of the genes involved in dimethylglycine and sarcosine degradation in *Arthrobacter spp.*: implications for glycine betaine catabolism. Eur J Biochem 268:3390–3398

Methé BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X, Moult J, Madupu R, Nelson WC, Dodson RJ, Brinkac LM, Daugherty SC, Durkin AS, DeBoy RT, Kolonay JF, Sullivan SA, Zhou L, Davidsen TM, Wu M, Huston AL, Lewis M, Weaver B, Weidman JF, Khouri H, Utterback TR, Feldblyum TV, Fraser CM (2005) The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. Proc Natl Acad Sci U S A 102:10913–10918

Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci U S A 84:166–169

Roberts MF (2005) Organic compatible solutes of halotolerant and halophilic microorganisms. Saline Systems 1:5

Smith LT, Pocard JA, Bernard T, Rudulier DL (1988) Osmotic control of glycine betaine biosynthesis and degradation in *Rhizobium meliloti*. J Bacteriol 170:3142–3149

Sohlenkamp C, de Rudder KE, Rohrs V, Lopez-Lara IM, Geiger O (2000) Cloning and characterization of the gene for phosphatidylcholine synthase. J Biol Chem 275:18919–18925

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612

Thormann KM, Duttler S, Saville RM, Hyodo M, Shukla S, Hayakawa Y, Spormann AM (2006) Control of formation and cellular detachment from *Shewanella oneidensis* MR-1 biofilms by cyclic di-GMP. J Bacteriol 188:2681–2691

Todd JD, Rogers R, Li YG, Wexler M, Bond PL, Sun L, Curson ARJ, Malin G, Steinke M, Johnston AWB (2007) Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. Science 315:666–669

Trickey P, Wagner MA, Jorns MS, Mathews FS (1999) Monomeric sarcosine oxidase: structure of a covalently flavinylated amine oxidizing enzyme. Structure 7:331–345

Wargo MJ, Szwergold BS, Hogan DA (2008) Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. J Bacteriol 190:2690–2699

Wargo MJ, Ho TC, Gross MJ, Whittaker LA, Hogan DA (2009) GbdR regulates *Pseudomonas aeruginosa* plcH and pchP transcription in response to choline catabolites. Infect Immun 77:1103–1111

Wells LE (2006) Viral adaptations to life in the cold. Ph.D. thesis, University of Washington

Wells LE, Deming JW (2006a) Characterization of a cold-active bacteriophage on two psychrophilic marine hosts. Aquat Microb Ecol 45:15–29

Wells LE, Deming JW (2006b) Effects of temperature, salinity and clay particles on inactivation and decay of cold-active marine Bacteriophage 9A. Aquat Microb Ecol 45:31–39

Williams K, Gillespie J, Sobral B, Nordberg E, Snyder E, Shallom J, Dickerman A (2010) Phylogeny of gammaproteobacteria. J Bacteriol 192:2305–2314

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693

Table 1: Divergence rates and times of paralogous gene clusters involved in compatible solute utilization in *C. psychrerythraea* 34H. Abbreviations: S, number of non-synonymous sites; N, number of synonymous sites; dS, synonymous substitution rate; dN, non-synonymous substitution rate; t_S, synonymous site divergence (Myr); t_N, non-synonymous site divergence (Myr).

gene	locus 1	locus 2	S	N	dS	dN	dN/dS	t_S	t_N
MarR regulator	CPS_4007	CPS_1337	110	335	2.68	0.27	0.10	294	605
hypothetical	CPS_4008	CPS_1336	159	540	2.58	0.26	0.10	283	572
betT	CPS_4009	CPS_1335	328	887	3.72	0.13	0.03	409	281
betA	CPS_4010	CPS_1334	385	1295	2.18	0.11	0.05	239	236
betB	CPS_4011	CPS_1333	349	1112	17.09	0.11	0.01	1879	241
betI	CPS_4012	CPS_1332	139	449	2.75	0.26	0.09	302	567
glyA	CPS_4031	CPS_2477	313	938	0.18	0.00	0.02	19	7
soxB	CPS_4032	CPS_2478	249	999	0.00	0.00	0.00	0	0
soxD	CPS_4033	CPS_2479	49	248	0.00	0.00	0.00	0	0
soxA	CPS_4034	CPS_2480	675	2346	0.00	0.00	0.00	0	0
soxG	CPS_4035	CPS_2481	213	381	0.01	0.00	0.00	1	0
purU	CPS_4036	CPS_2482	210	666	0.08	0.00	0.00	9	0.2
DNA binding protein	CPS_4037	CPS_2483	155	439	0.36	0.04	0.10	40	80

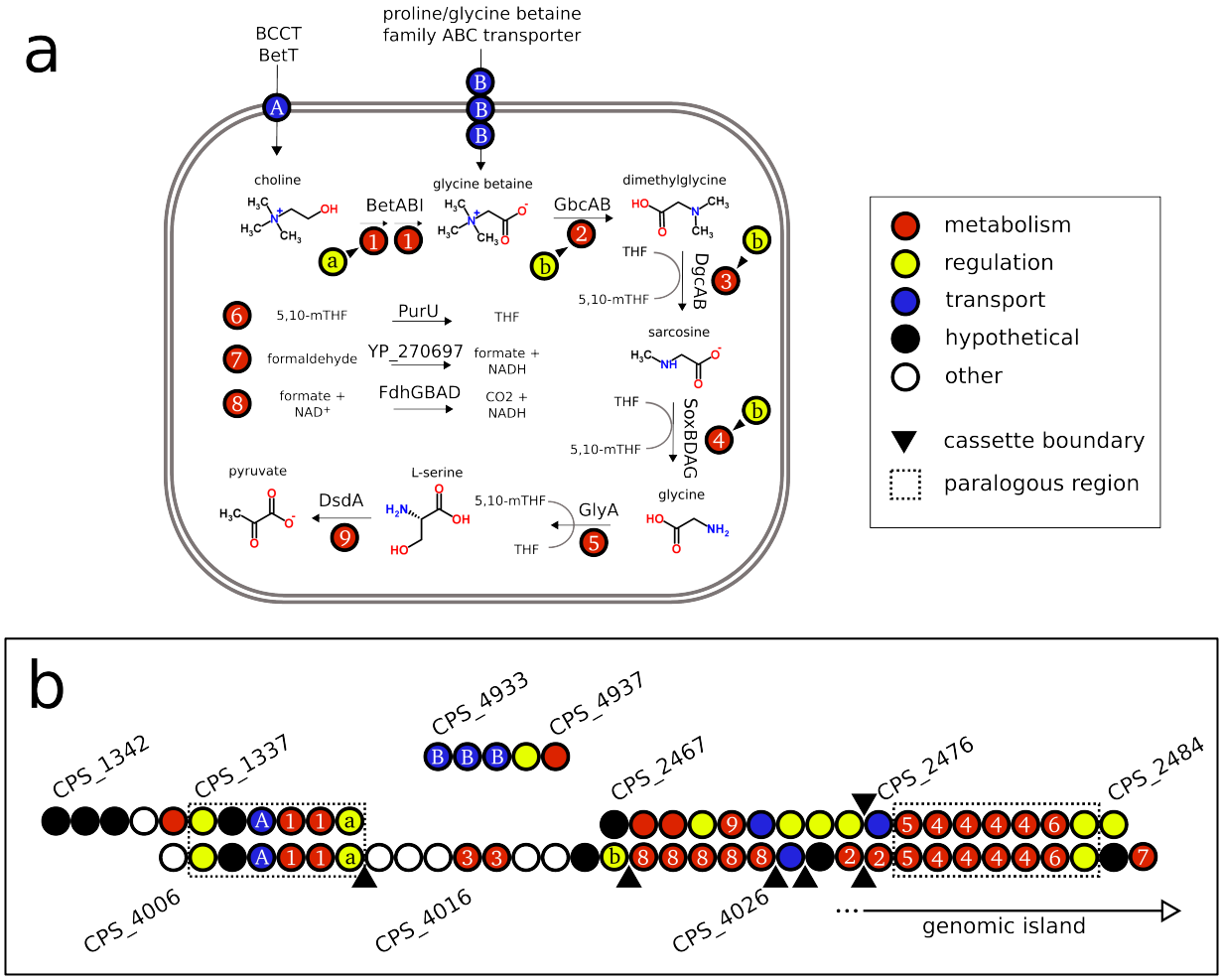

Figure 1: (a) The pathway of choline degradation to L-serine encoded by *Colwellia psychrerythraea* 34H. (b) Genes involved in the degradation process: *betABI* (CPS_4010–CPS_4012, CPS_1332–CPS_1334), *gbcAB* (CPS_4029–CPS_4030), *dgcAB* (CPS_4016–CPS_4017), *soxBDAG* (CPS_2478–CPS_2481, CPS_4032–CPS_4035), *glyA* (CPS_2477, CPS_4031, CPS_0728, CPS_3844), and *dsdA*, a truncated serine dehydratase (CPS_2471). Genes involved in one-carbon cyling and folate metabolism are: *purU* (CPS_2482, CPS_4036, CPS_4357, CPS_3620), *folD* (CPS_3133, CPS_3791), formaldehyde dehydrogenase (CPS_4039), and *fdhGBAD* (CPS_4022-4026, CPS_2056-2060). Genes for putative transporters for quarternary amines are: choline BCCT *betT* (CPS_4009, CPS_1335), putative sarcosine BCCT (CPS_3860), and a putative glycine betaine ABC transporter (CPS_4932–CPS_4934). Abbreviations: tetrahydrofolate, THF; 5,10-mTHF, 5,10-methylenetetrahydrofolate. Dotted boxes indicate regions of paralogy.

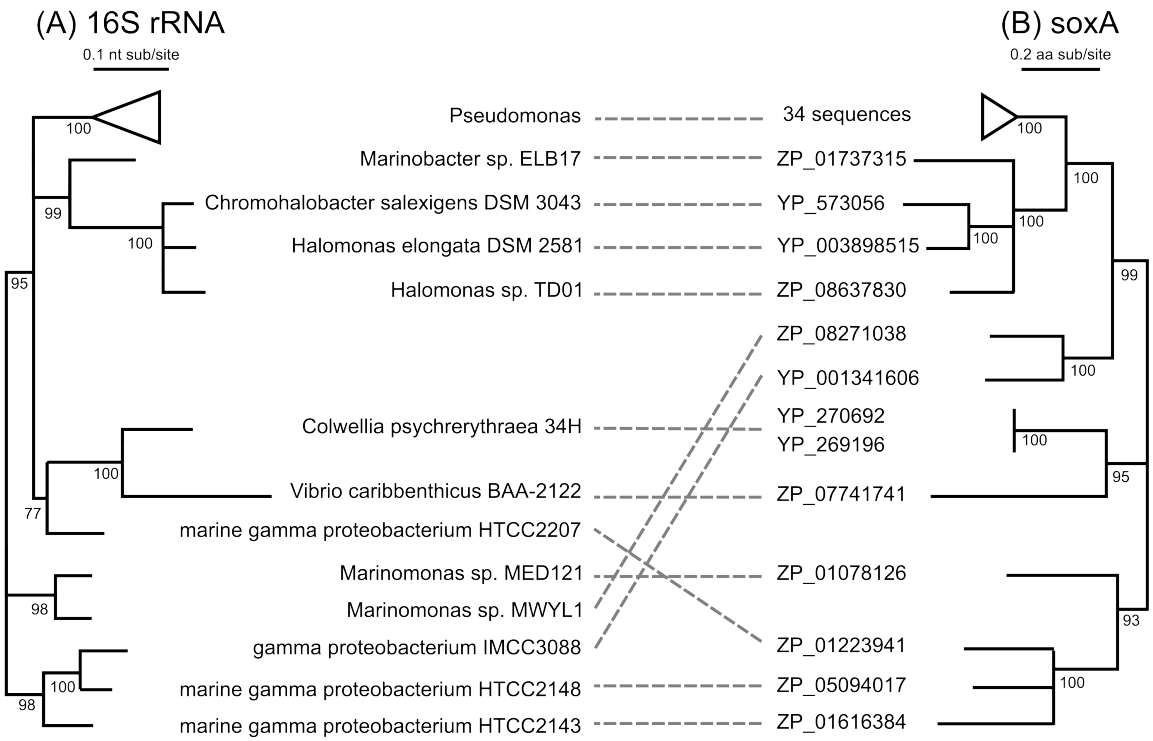

Figure 2: (a) Maximum likelihood phylogenetic tree of 16S rRNA gene sequences from gammaproteobacteria genomes in the JGI-IMG database containing SoxA homologs. (b) Maximum likelihood phylogenetic tree of amino acid sequences encoding heterotetrameric sarcosine oxidase subunit alpha (SoxA) present in the JGI-IMG database. Node labels indicate approximate likelihood ratio test (aLRT) support values as percentages; nodes with aLRT support values less than 70% were collapsed into hard polytomies. Dashed lines match the 16S rRNA and SoxA sequences from the same organism; crossed lines indicate topographic mismatches, possibly due to horizontal gene transfer.

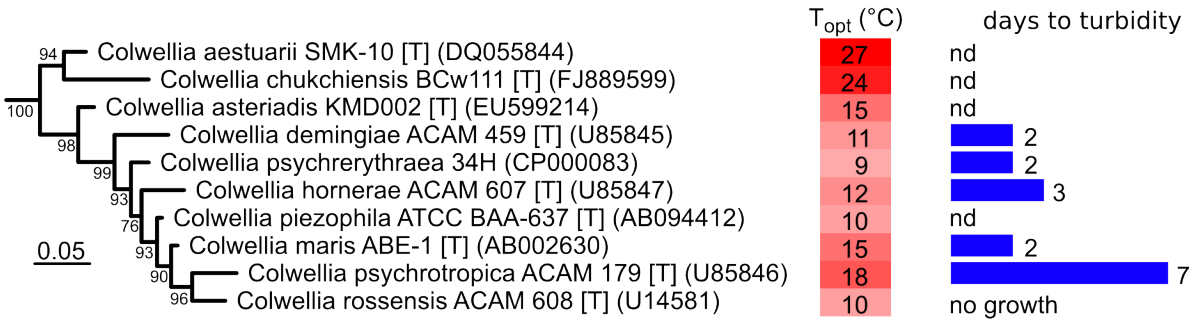

Figure 3: Maximum likelihood phylogenetic tree of 16S rRNA gene sequences from *Colwellia* isolates. Node labels indicate approximate likelihood ratio test (aLRT) support values as percentages. Background shading indicates optimum growth temperature (℃) for each strain. Growth on defined medium SLV at 8℃ is indicated as 'days to turbidity'; nd, not determined.