University of Nevada, Reno

A path to biodiscovery: isolating Antarctic ascidian-associated bacteria and screening genomes for secondary metabolite-encoded biosynthetic gene clusters and signatures of cold adaptation

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Cellular and Molecular Biology

by

Nathan M. Ernster

Dr. Alison Murray/Thesis Advisor

May 2021

University of Nevada, Reno

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

NATHAN ERNSTER

entitled

A path to biodiscovery: isolating Antarctic ascidian-associated bacteria and screening genomes for secondary metabolite-encoded biosynthetic gene clusters and signatures of cold adaptation

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Alison E. Murray, Ph.D.
Advisor

Subhash Verma, Ph.D.
Committee Member

David Alvarez-Ponce, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May, 2021

ABSTRACT

Natural products synthesized by host-associated microbes from polar regions pose tremendous potential for addressing human health challenges. One prominent example is the ascidian *Synoicum adareanum*, found on the seafloor of coastal regions of Antarctica, of interest in part due to one compound, palmerolide A. Palmerolide A is a cytotoxic macrolide with specific activity towards cancer cells which occurs in high concentration in its tissues. It has since been determined that one core bacterial member of the moderately diverse *Synoicum adareanum* microbiome is responsible for the biosynthesis of palmerolide A. This finding reinforces the significance of investigating natural product potential in host-associated microbes from high-latitude ecosystems. One goal of the *Synoicum adareanum* research program is establishing a sustainable biological source of palmerolide A either through cultivation or bioengineering solutions. Secondly, exploring the microbiome for additional cold-water biosynthesized natural products is of interest given recent metagenome results that suggest substantial natural product biosynthetic potential. To advance this research requires investigating cultivation strategies for the Verrucomicrobia-affiliated palmerolide A producer, establishing a more comprehensive understanding of the biosynthetic potential of other microbes in the *Synoicum adareanum* microbiome, and ascertaining the cold-adaptive characteristics of the genome and encoded proteome evolution of host-associated high latitude marine microorganisms.

Here, we implemented a minimal media cultivation and dereplication strategy to isolate bacteria, rapidly screen for palmerolide A-encoding biosynthetic gene clusters using PCR and determine the identity of the novel isolates with long read 16S rRNA phylogenetic analyses. This work added phylogenetic breadth to previously cultivated

bacterial classes Gammaproteobacteria and Alphaproteobacteria from the *Synoicum adareanum* microbiome, to now include Bacilli, Acidimicrobiia, Actinobacteria and Bacteroidia. Through sequencing representative genomes of 22 isolates, 135 biosynthetic gene cluster types across eight compound categories were identified within this expanded culture collection using bioinformatic approaches.

Comparative genomic analyses between orthologous proteins predicted between our high-latitude, Antarctic isolates and representative low/mid latitude marine microorganisms of the same genera revealed signatures of cold adaptation in the Antarctic genomes. We found an overall decrease in proline residues that form rigid kinks in protein sequences. Likewise, charged amino acids and arginine/(arginine+lysine) content, which form protein-stabilizing salt bridges believed to hinder protein function at low temperatures were reduced. The predicted Antarctic proteins also followed the established trend of increased glycine and serine content in cold adapted proteins, understood to increase protein dynamics in cold through their small size and reduced ionic charge. Interestingly, significant numbers of coding sequences with lower GC content were found in the Antarctic genomes even if genome-wide GC contents were similar, or even higher, for the Antarctic organism. Also surprising, more predicted proteins with higher intrinsic disorder were observed in the low/mid latitude genomes, perhaps indicative of a possible environmental temperature threshold for disorder which balances protein function and cold denaturation resistance. These cultivation, phylogenomic and bioinformatic results provide a deeper view into this Antarctic marine ascidian, the diversity and potential for secondary metabolite production within its

microbial community, and trends observed in protein and genome sequences of cold adapted organisms.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Alison Murray, for her excellent guidance, support, and knowledge as a mentor throughout my research project.

I also wish to thank my committee members Dr. Subhash Verma for encouragement and advising of the Cell and Molecular Biology program, and Dr. David Alvarez-Ponce of the Department of Biology for guidance and assistance with genomic analyses.

I take this opportunity to express gratitude to the Division of Earth and Ecosystem Science at Desert Research Institute and the NV NASA Space Grant for Graduate Research Assistantship funding, and the IDeA Network of Biomedical Research Excellence (INBRE) for a whole genome sequencing and bioinformatics training service award.

I would like to thank Dr. Bill Baker and Dr. Nicole Avalon at University of South Florida for their help and advice regarding biosynthetic gene clusters and enzyme analyses, Dr. Agusto Luzuriaga at University of Nevada, Reno for his assistance with comparative genomic analyses, and Dr. Hans Vasquez-Gross at University of Nevada, Reno for his guidance during my INBRE bioinformatics training award.

I would also like to acknowledge the following principal investigators for their permission to use genome assemblies in this study:

- Dr. Jim Haseloff at University of Cambridge for use of the *Photobacterium leiognathi* CUB1 genome.
- Dr. Mary Ann Moran at University of Georgia for use of the *Sulfitobacter* sp. NAS-14.1 genome.

- Dr. Chris Greening at Monash University for use of the *Anderseniella* sp. isolate Site_C17 genome.
- Dr. R. Eric Collins at University of Alaska for use of the *Psychromonas* sp. SR45-C genome.

I also thank Dr. Jorge Rodrigues at University of California, Davis for consultation while planning the initial microbial cultivation.

Lastly, I thank my fellow Murray lab students Miranda Seixas, Eric Lundin, Lauren Siao and Jaiden Christopher for their assistance with cultivation and molecular and genomic analyses.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Marine invertebrates are known sources of bioactive natural products of which roughly 8% of these secondary metabolites are produced by microorganisms associated with the hosts (McCauley et al., 2020; Schmidt, 2015). From ascidians, a globally widespread class of tunicates in the phylum Chordata, more than 1,000 natural products have been isolated of which at least 150 are known to be synthesized by ascidian-associated microorganisms (Chen et al., 2018). Some tunicate-derived secondary metabolites are being utilized as FDA approved therapeutics, while others are currently undergoing clinical trials (McCauley et al., 2020). Expanding knowledge of symbiotic relationships between invertebrates and their microbiome constituents offers tremendous potential in natural product research through combining culture-independent and dependent techniques for novel biosynthetic discovery (Ayuningrum et al., 2019).

This study was motivated by previous work investigating the microbiome of an Antarctic ascidian, *Synoicum adareanum,* found on the sea floor surrounding Anvers Island Archipelago near Palmer Station, Antarctica, in waters between -1.5℃ and +1.0℃ (Murray et al., 2020; Riesenfeld et al., 2008). Using cultivation independent analyses (16S rRNA gene sequence data), Riesenfeld et al. (2008) and Murray et al. (2020) found the *Synoicum adareanum* microbiome to be moderately diverse across all ascidian samples collected around the archipelago, with the microbial community most highly represented by the phyla Proteobacteria, Bacteroidetes and Verrucomicrobia.

Palmerolide A, a macrolide with cytotoxicity toward melanoma, has been extracted from tissue of *Synoicum adareanum* (Diyabalanage et al., 2006), at amounts between 0.49-4.06 mg palmerolide A x g^{-1} of *Synoicum adareanum* dry weight (Murray

et al., 2020). While total synthesis of palmerolide A has been accomplished, the inefficiencies of these procedures suggest synthesis is not a viable production pathway (Jiang et al., 2007). Likewise, extracting palmerolide A from naturally harvested *Synoicum adareanum* is impractical and unsustainable and transportation and cultivation of these Antarctic invertebrates would pose significant logistic challenges, not to mention the fact that we know little of their life histories.

Metagenome analysis corresponding to palmerolide A distribution between ascidian lobes and across multiple sample sites suggested a novel Verrucomicrobium, *Candidatus* Synoicihabitans palmerolidicus, is responsible for palmerolide A biosynthesis (Avalon et al., 2021; Murray et al., 2021). The same data set revealed an additional 74 biosynthetic gene clusters (BGCs) in the microbiome, spanning a diverse suite of secondary metabolites including terpenes, nonribosomal peptide synthases (NRPS), ribosomally synthesized and post-translationally modified peptides (RiPPs) and polyketide synthases (PKS) (Murray et al., *In prep*). Many secondary metabolites are naturally synthesized as defense mechanisms within ecosystems but have uses in pharmaceutical and industrial applications, either directly or as precursor compounds in natural product engineering, yet natural product potential from polar host-associated bacteria remains understudied. Developing a greater understanding of the role secondary metabolites play in marine polar holobiont partnerships may help identify their value in natural product applications.

While cultivation-independent techniques, including metagenomic analyses, are able to identify probable candidates with interesting biosynthesis, cultivation allows for direct analysis (e.g. of genomes, metabolites, and bioassay-directed screens), but is not

always straightforward. Reproducing the subtleties of environmental conditions limits cultivation of a large percentage of natural environmental microbiota (Bodor et al., 2020; Stevenson et al., 2004). Complex symbiotic relationships in marine invertebrate host-associated systems can also present difficulties in microbial isolation (Selvin et al., 2009). Likewise, it is common that secondary metabolite production is diminished or nonexistent in pure culture where chemical cross-talk between organisms is not induced, leading to "silent" BGCs (Rutledge & Challis, 2015). These are primary examples of challenges in cultivation-dependent work and secondary metabolite detection in host-associated systems.

Expanding knowledge of evolution to cold environments may guide future engineering of microbes for production of secondary metabolites which are naturally found in polar bacteria. This study has aimed to bring previously uncultivated microbial members of the *Synoicum adareanum* microbiome into pure culture, allowing for direct analysis. While we have yet to isolate *Candidatus* Synoicihabitans palmerolidicus, our new culture collection has expanded the known diversity of the microbiome community, and through whole genome sequencing of select isolates we have identified additional secondary metabolite biosynthetic pathways and evidence of cold adaptation.

MATERIALS AND METHODS

Sample collection and bacterial cultivation.

Synoicum adareanum specimens were collected in 2011 at seven sites (three lobes of three colonies each) around the Anvers Island Archipelago at depths between 24 and 31 meters and have been stored at -80℃ (Murray et al., 2020). A sample from Norsel Point (Nor-2C-2011) with relatively high levels of palmerolide A gene targets (AT-1: 440804±37491 and HCS 185837±57040 gene copies x ng^{-1} host tissue wet wt., Murray et al. 2021) was used for microbial cultivation.

A minimal salts medium was prepared following Stevenson et al. (2004). Carbon sources were chosen based on preliminary analysis of biodegradative pathways observed in the *Candidatus* Synoicihabitans palmerolidicus metagenome. The media was amended with five defined carbon sources each at 2.5 mM concentration: chitin (Sigma-Aldrich), D–(+)- xylose (Sigma-Aldrich), cellobiose (MP Biomedicals), taurine (Sigma-Aldrich), and dextrose (VWR) (Table 1) (Murray et al., 2021; Stevenson et al., 2004). A replicate set of media was supplemented with yeast extract at 5 µg/mL concentration. Another replicate set was supplemented with catalase to aid in growth of oxygen stressed microbes (Stevenson et al., 2004). Agar plates (100mm x 15mm, VWR) and liquid media were inoculated with microbial cells enriched from a homogenized Nor-2C-2011 tissue using a differential centrifugation protocol followed by serial 1:10 dilutions (Murray et al., 2020). Two growth conditions, atmospheric and microaerophilic, were used and all inoculated plates and liquid media were incubated at 10 ℃. To achieve microaerophilic conditions, 1) agar plate containers were purged with nitrogen for five minutes while

Table 1a and 1b. Media components: Basal medium and minimal salts medium

a.

5X Basal Medium	
Compound	g/L
KH_2PO_4	0.0085
NH_4Cl	0.0027
KNO_3	0.063
KCl	3.8
$CaCl_2 \cdot 2H_2O$	7.57
NaCl	137.33
$MgCl_2 \cdot 6H_2O$	53.88

b.

Modified Minimal Salts Medium for Isolates	
5X Basal Medium (Table 1a)	200 mL/L
Na_2SO_4 solution (1.25 M)	16 mL/L
ATCC® Trace Mineral Supplement™	1 mL/L
ATCC® Vitamin Supplement™	1 mL/L
MOPS Buffer (pH 7.2, 1 M)	10 mL/L
Carbon source (0.5 M)	5 mL/L
Yeast Extract (5 µg/mL final concentration)	1 mL/L
H_2O	To vol. of 1L

partially open to vent, and 2) septum-capped liquid hungate tubes received four cycles of 5 seconds vacuum and 5 seconds filling with nitrogen, with a final addition of 4.2mL of air (4.2 mL of 20% O_2-containing air into 16.5 mL volume = 5% O_2 tube). Negative control agar plates and liquid media tubes were prepared to monitor for contamination and incubated alongside the cultures.

Isolation and DNA extraction.

Plates were screened for colony growth with use of a dissecting microscope, generally twice-monthly, over an 11-month period and were selected for picking and further purification based on colony morphology and time of first appearance using a dissecting scope. Hungate tubes of liquid cultures were monitored at similar intervals for visible turbidity. Following initial plating, colonies were picked and transferred on agar plates at fourfold higher carbon-source and yeast extract concentrations to encourage isolate growth. Isolates were then assigned logical identifiers using the first three letters of the carbon source and sequential isolate number (ex. CEL 1, CEL 2, etc.) Biomass was

harvested from plates with sufficient growth and the biomass was split; one part was archived at -80 ℃ in 15% glycerol and the other part was pelleted for DNA extraction and purification, following Massana et al. (Massana et al., 1997). DNA extract concentrations were determined using Quant-iT™ Picogreen™ dsDNA quantitation assay kit (Thermo-Fisher Scientific) following manufacturer's instructions.

Palmerolide A screening.

To rapidly screen for presence of the palmerolide A-producing bacteria, we implemented PCR amplification targeting two palmerolide A genes (acyltransferase subunit 1 (AT-1) and a hydroxymethylglutaryl-CoA synthase, HCS; (Murray et al., 2021). As the purpose was to detect positive reactions, we used PCR to screen DNA extracts, blank negative controls, synthetic positive control (PalA-4CDSCtrl), plus a natural *Synoicum adareanum* microbiome DNA extract (Nor2P7AM16C). Each 20 µL PCR reaction included 1 ng of DNA, 10X buffer, 2.5mM dNTPs, 5µg/µL Taq polymerase (New England Biosciences), one of two primer sets (ATIF/ATIR and HCSF/HCSR, 5µM each) and PCR water (Murray et al., 2021). Thermocycling for PCR utilized an Applied Biosystems GeneAmp™, with initial denaturation at 95 ℃ for 5 minutes, then 30 cycles of denaturing, annealing, and extension as follows: 95 ℃ for 30 seconds, 60 ℃ for 1 minute, 72 ℃ for 1 minute. A final 5-minute extension was held at 72 ℃ followed by a 4 ℃ hold.

All PCR products were screened using agarose gel electrophoresis. 1 µL ladder (Promega® 50bp & Bioline® Hyperladder II™) + 1.5 µL gel loading dye + 7.5 µL water were used in reference wells. 5 µL sample PCR product + 1.5 µL gel loading dye + 3.5

µL water were used in unknown wells. Gels were stained with GelRed™ stain (Biotium) and photographed using a UVP® BioDoc-It® UV transilluminator.

Isolate dereplication and 16S rRNA gene sequencing.

Given the unknown levels of redundancy in the culture collection, we utilized a PCR-denaturing gradient gel electrophoresis (PCR-DGGE) de-replication strategy. DGGE is effective as it can separate PCR amplicons with a 40 bp GC-clamp at one end as they migrate through an increasingly denaturing gel, whereupon the amplicons denature at sequence-specific melting points (and not the clamp) and are immobilized (Murray et al., 1996).

In addition to the isolates cultivated in the defined media approach described here, 19 isolates from a former culture collection established in 2007 in the Murray lab were added to the dereplication screening effort to establish a suite of unique taxa for genome sequencing and BGC analysis (Murray et al., 2020). PCR was performed targeting the variable 3 region of the 16S rRNA gene using primers GC358F/517R. The forward primer included a 40 bp GC-clamp to facilitate DGGE separation (Murray et al., 1996). 1ng working stock template DNA for samples, prepared as above, were used for PCR. Applied Biosystems PCR reagents were used, so adjustments to the PCR protocol above included supplemental Applied Biosystems $MgCl_2$ solution following manufacturers protocol for the Applied Biosystems AmpliTaq Gold™ and Buffer II used. Thermocycling parameters were as follows: initial denaturation at 95 ℃ for 5 minutes. Then 10 cycles of: 95 ℃ for 45 seconds, touchdown annealing (starting at 65 ℃) for 30 seconds, 72 ℃ for 30 seconds. Then 22 cycles as follows: 94 ℃ for 30 seconds, 55 ℃ for

30 seconds, 72 ℃ for 30 seconds. A final 10-minute extension was held at 72 ℃. Lastly, the temperature was held at 4 ℃ indefinitely. Early PCR reaction volumes were made at 20 µL, but later volumes were made at 25 µL to generate additional PCR product for subsequent DGGE comparisons. 5 µL of PCR products were ran using gel electrophoresis, as described above, to confirm successful amplification.

Ten µL of PCR products with 2.5 µL gel loading dye were then run for 600 volt*hrs on a 30-65% denaturing gradient gel followed by iterative gel compositions to screen for isolates with unique migration (melting) points using a BioRad DCode™ Universal Mutation Detection System (Appendix 2). Positive controls using Nor2-7AM14C (Nor2) were amplified in individual tubes at 50µL reaction volumes to be used as reference bands in DGGE gels, quantified with Quant-iT™ Picogreen™ as described above, then volumes equaling 200 ng of DNA were prepared in individual aliquot tubes with 2.5 µL gel loading dye to be used in reference wells. This ensured equivalent amounts of positive control DNA were used in each gel. Gels were stained with SYBR™ Gold stain (Thermo Fisher) and photographed using a UVP® BioDoc-It® UV transilluminator.

Nearly fully length 16S rRNA genes of isolates with unique melting points were then obtained using PacBio (Pacific Biosciences) sequencing technology (MR DNA, Shallowater TX). Samples at MR DNA laboratory were prepared with primers ill27Fmod(/bac1492R (AGRGTTTGATCMTGGCTCAG/ GGGTTACCTTGTTACGACTT) then PCR cycles were completed as follows: 3 minutes at 94 ºC, 35 cycles of 94 ºC (30 seconds), 53 ºC (40 seconds), then 72 ºC (90 seconds), with a final 72 ºC held for 5 minutes. PCR products were confirmed by running in a 2%

agarose gel. Based on DNA concentrations and molecular weights, equal proportions of samples were pooled and purified with Ampure PB beads (Pacific Biosciences). SMRTbell (Pacific Biosciences) libraries were prepared for Single Molecule, Real-Time sequencing. The Circular Consensus Sequencing algorithm (Pacific Biosciences) aligns reads into consensus sequences, which were then processed using a MR DNA laboratory pipeline.

In addition to these sequence isolates, 16 previously sequenced 16S rRNA genes from the 2007 *Synoicum adareanum* microbiome culture collection (Murray et al., 2020) were classified using Silva ACT tool (Glockner et al., 2017; Murray et al., 2020). Two near neighbor 16S rRNA sequences were collected from the Silva ribosomal RNA database (Glockner et al., 2017). The sequences were aligned with two neighboring sequences per each isolate and phylogenetic analysis was conducted using MEGA 10 (Kumar et al., 2018).

Whole genome sequencing of isolates.

Following phylogenetic analysis, sample DNA extracts (4-1,390 ng/µL DNA concentration) from a subset of the culture collection (22 isolates in total) were selected across a range of the phylogenetic tree representing the diversity of phyla/classes from both the prior and new microbiome culture collection for whole genome sequencing (Illumina® NextSeq 550; paired end, 150 cycle, 2 x 75 reads; 4 lanes per genome) at the Nevada Genomics Center (University of Nevada, Reno). DNA samples were prepared for sequencing at the Nevada Genomics Center using the Illumina® DNA Prep kit.

Genome assembly, annotation, and secondary metabolite mining.

Initial genome assemblies were tested using MaSuRCA genome assembler and SPAdes with "careful" and "isolate" options, as well as SPAdes with one and four lanes of read data to compare quality of outputs (Bankevich et al., 2012; Zimin et al., 2013). Quality of the assemblies was assessed using BUSCO and Quast tools (Gurevich et al., 2013; Seppey et al., 2019). Core bacterial genes were searched for using BUSCO's "bacteria_odb10" database. Quality assessments were visualized using MultiQC (Ewels et al., 2016). All downstream analyses were completed with assemblies from four lanes of DNA sequence reads and the SPAdes (version 3.14.1) "careful" option (Bankevich et al., 2012). In addition, contigs from one assembly, CEL 36, was binned using MaxBin 2.0 (Wu et al., 2016) to be used in downstream analyses. Assembled genomes were then annotated using EDGE platform (Li et al., 2017) with 500 bp minimum contig length, Prokka annotation tool (Seemann, 2014), taxonomy classification, gene family and antibiotic resistance analysis options. Genome assemblies were run through the bacterial antiSMASH secondary metabolite biosynthesis pipeline using default settings (Blin et al., 2019).

Identifying genomic signatures of cold adaptation.

To identify signatures of cold adaptation within our Antarctic genomes, we first searched for genomes of low/mid latitude samples within the same genus of each isolate genome for use in comparisons between predicted orthologous proteins common between the two organisms. A previously sequenced alphaproteobacterium from the *Synoicum adareanum* microbiome, *Pseudovibrio* sp. TunPHSC 04-5.I4, was included in the cold adaptation

analysis as it served as a positive psychrophile control, as it does not grow at temperatures > 16 ºC (Murray, *pers. com.)* and is considered to be a psychrophile. Taxonomic classification of the isolate genomes was performed using a combination of 16S rRNA gene sequences repositories (Ribosomal Database Project (Cole et al., 2014) and Silva rRNA classification (Glockner et al., 2017)) and genome databases (Microbial Genomes Atlas NCBI Prok (Rodriguez-R et al., 2018) and GTDB-Tk (Chaumeil et al., 2020). Using these classifications, the Joint Genome Institute's (JGI) Integrated Microbial Genomes and Microbiomes (IMG) database was searched for sequenced genomes of the same genus from lower latitude marine environments (Chen et al., 2021).

For each of our Antarctic genomes, a representative low/mid latitude genome was selected based on the following criteria: (1) priority was given to isolated organisms sampled from latitudes as close to the equator as possible (maximum +/- 45º latitude), while (2) the genome assembly was listed as "high quality" on IMG, and (3) the organism was sampled from a marine or coastal environment (with priority given to sample depths ≤ 200 meters, the upper limit of the deep ocean). The low/mid latitude organism's genome was annotated using EDGE as described above, then compared to our corresponding genome assembly using the Average Amino acid Identity (AAI) calculator (Rodriguez-R & Konstantinidis, 2016; Rodriguez-R & Konstantinidis, 2014). For genomes above 65% AAI the organisms are considered within the same genus (Konstantinidis et al., 2017). Other genomes matching the above criteria were tested for selection for AAI if initial genomes had <65% AAI with sample genomes.

Working with Dr. David Alvarez-Ponce's Genome Evolution Laboratory at University of Nevada, Reno we used genome assemblies and translated amino acid

FASTA files to identify signatures of cold adaptation. For each pair of genomes (23 in total), pairs of orthologs were identified using a best reciprocal hit approach (BLASTP, E-value < 10e-10) (Altschul et al., 1990). Orthologous protein pairs predicted from CDS in culture collection isolates and representative mid/low latitude organisms (annotated using the same EDGE parameters) were used in Biostrings R package to determine amino acid percentages and arginine/(arginine + lysine) ratios (R Core Team, 2018; Pagès et al., 2020). IUPred2A was used with the long intrinsic disorder option to score individual amino acids based on their tendency to contribute to disorder (from 0 to 1, with increasing disorder) (Meszaros et al., 2018). Percentages of disordered amino acids were calculated at two thresholds: ≥0.5 and ≥0.4. G-C content was also calculated for the coding sequence of each of the orthologous pairs. For each gene ortholog or predicted protein ortholog, the percentage of each parameter (A.A., ratio, GC content) was compared between the Antarctic and low/mid latitude ortholog to compile a list quantifying how many orthologs were higher in one genome versus the other. The sign test and paired Wilcoxon signed-rank test (Wilcoxon, 1945) were used to determine if the parameter tested was significantly higher for the Antarctic or low/mid latitude genome, and p-values ≤ 0.05 were used as a cut-off. Results from the paired Wilcoxon signed-rank test were used for the cold-adaptation analysis. Mean, median and standard deviation values were calculated across all orthologs within each genome for each parameter.

RESULTS

Cultivation and pal A screening

Over 11 months, 1,576 colonies were observed across agar plates containing individual carbon sources (chitin, xylose, taurine, cellobiose and dextrose). More colonies were observed under microaerophilic conditions overall (2.67X), with xylose and cellobiose media displaying the highest growth of any medium (Figure 1). Taurine plates had the highest growth of all atmospheric condition plates, and with the exception of one colony on chitin+yeast, chitin-containing media did not contain any colony growth (Figure 2). Of all colonies observed, 133 isolates were selected based on unique morphologies and time of first observation (e.g. Figure 3). After 11 months, 61 isolates had sufficient mass on plates to harvest for DNA extraction. Based on visual inspection, no significant growth in liquid media was observed. This set of 61 isolates was PCR screened for PalA-encoded acyltransferase subunit 1 (AT-1) and an hydroxymethylglutaryl-CoA synthase (HCS) to identify if any were the palmerolide A producer, but all results of screening thus far were negative.

PCR-DGGE dereplication and 16S rRNA sequencing

To identify replicates in the 61 isolate collection, 16S rRNA gene regions were amplified and screened by PCR-DGGE. Fifty of the 61 amplified successfully. We used an environmental microbiome DNA extract, Nor2-7AM14C (Nor2), as a positive control and band migration reference on all DGGE gels for the *Synoicum adareanum* microbiome isolates.

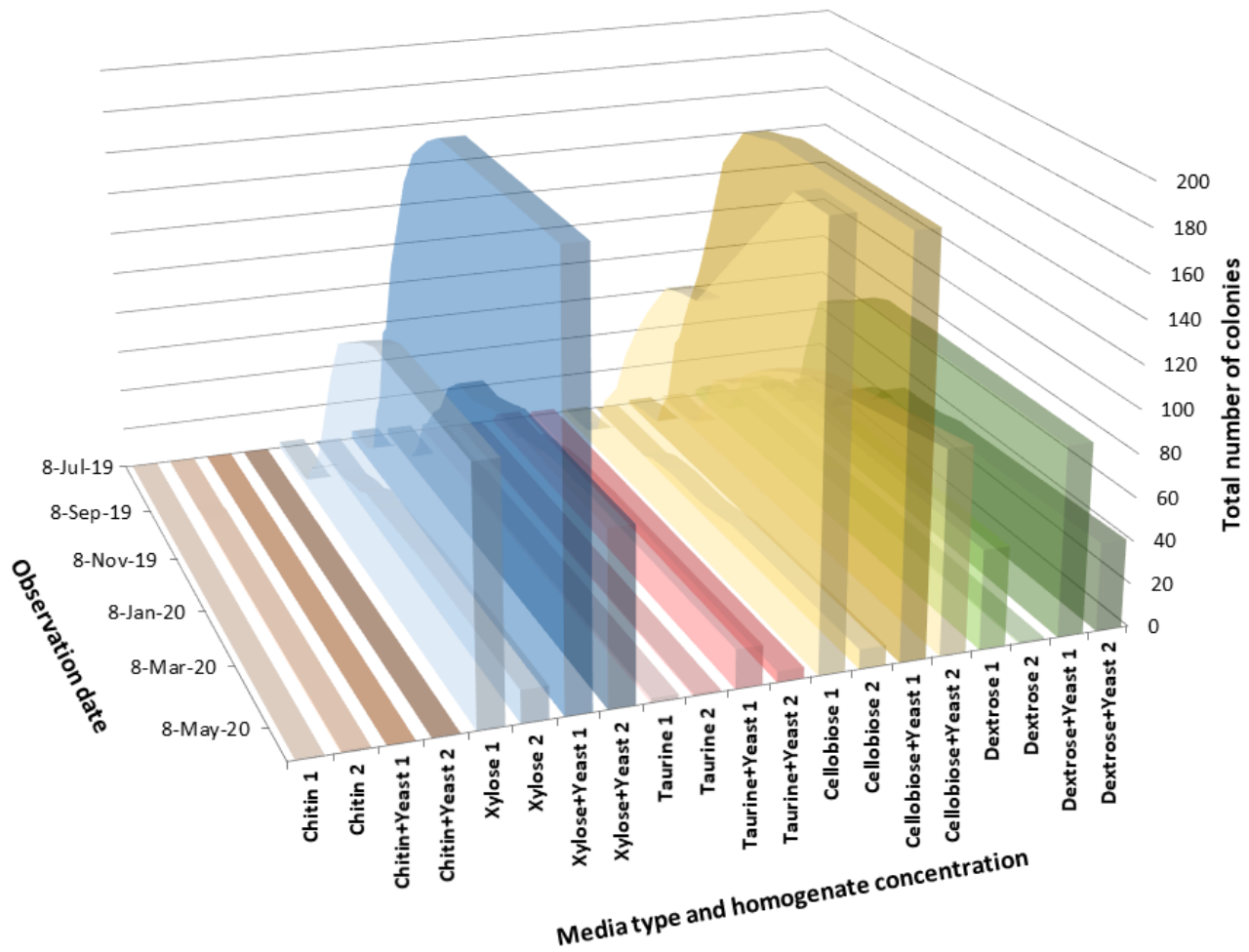


Figure 1. Total colonies observed on agar plates cultivated in 10°C under microaerophilic conditions 8July19-8Jun20, by media type and homogenate concentration (1 and 2 indicate the initial *Synoicum adareanum* tissue homogenate concentration and subsequent 1:10 dilution, respectively).

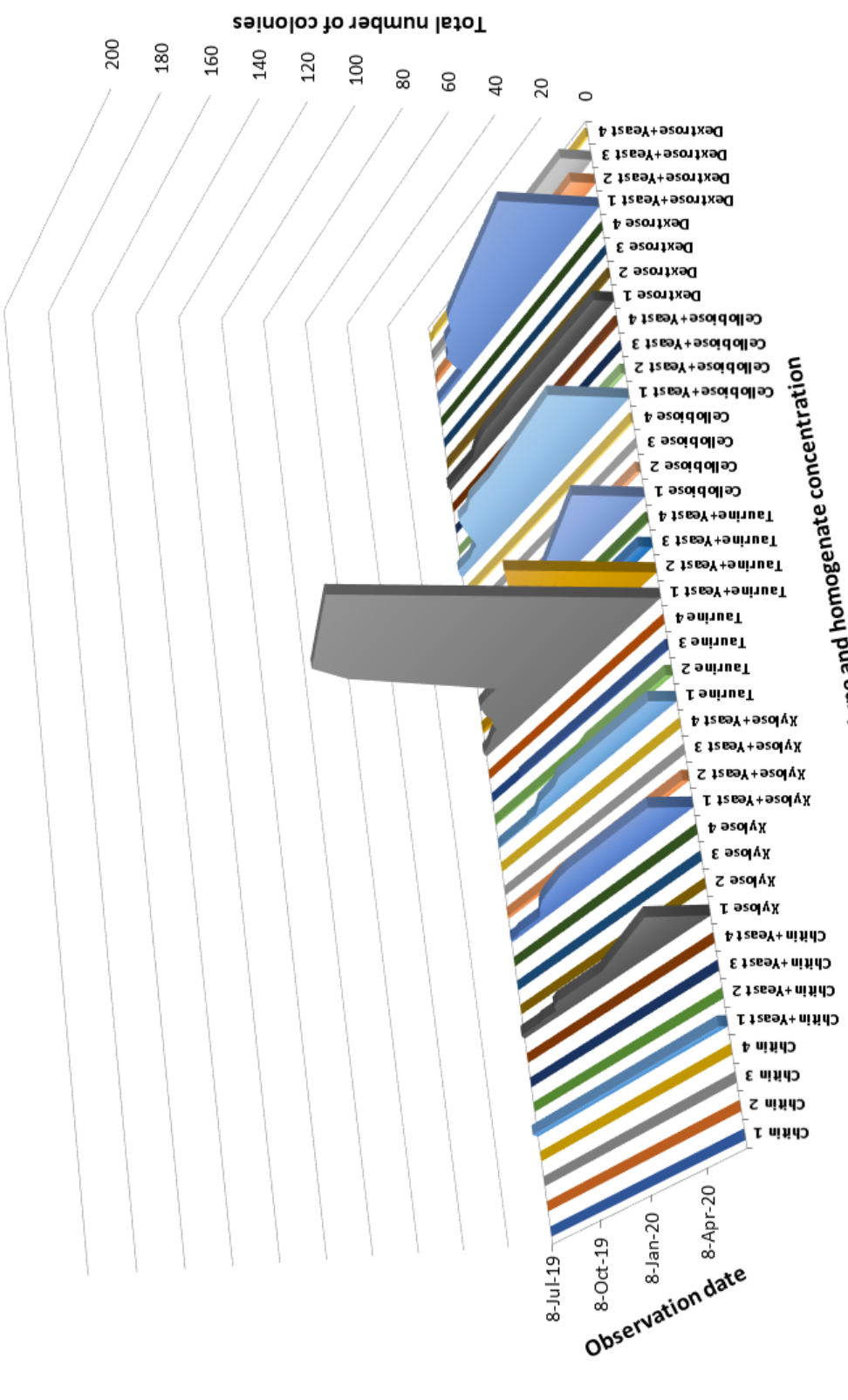


Figure 2. Total colonies observed on agar plates cultivated in 10°C under atmospheric conditions 8July19-8Jun20, by media type and homogenate concentration (1-4 indicate the initial *Synoicum adareanum* tissue homogenate concentration and three subsequent 1:10, 1:100, and 1:1000 dilutions).

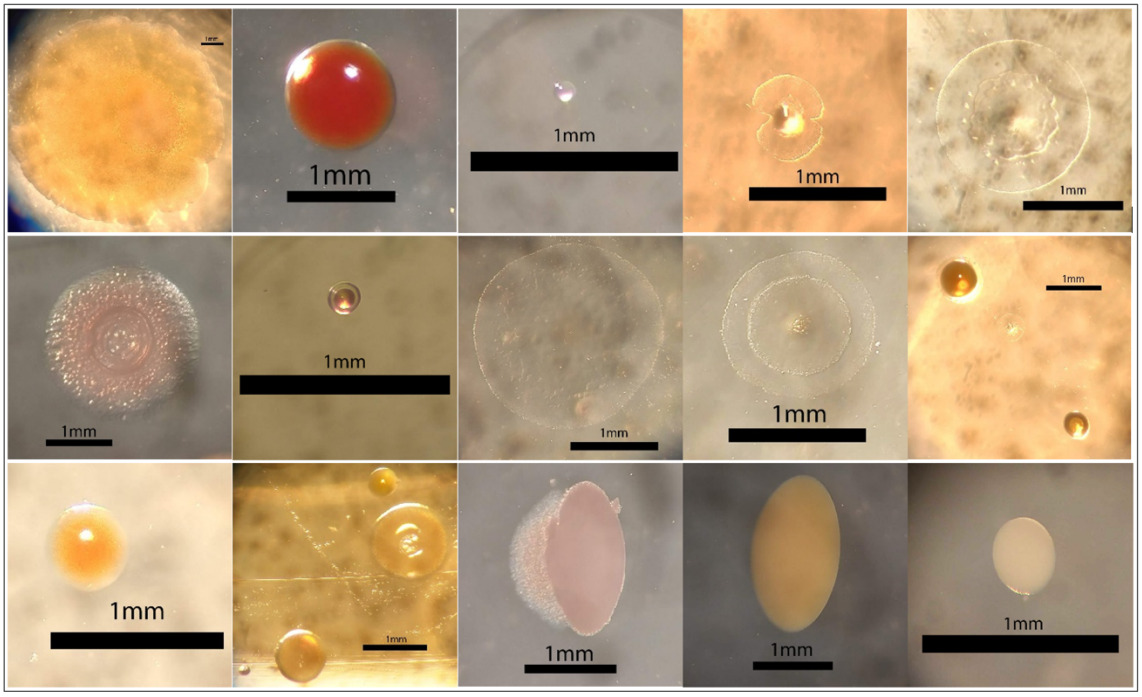


Figure 3. Examples of colony morphologies on agar plates. All scale bars are 1 mm. Media type and date first observed, left to right:

Top row: (Cellobiose + Yeast 08Jul2019), (Xylose + Yeast 31Oct19), (Dextrose + Yeast 31Oct19), (Cellobiose + Yeast 14Nov19), (Cellobiose + Yeast 14Nov19)

Middle row: (Dextrose + Yeast 19Sep19), (Dextrose + Yeast 31Oct19), (Xylose + Yeast 08Dec19), (Xylose + Yeast 14Nov19), (Cellobiose + Yeast, top 27Aug19, middle 14Nov19, bottom 13Sep19)

Bottom row: (Cellobiose + Yeast 14Nov19), (Cellobiose + Yeast, far left 14Nov19, 13Sep19, far right 13Sep19, top 27Aug19), (Dextrose + Yeast, Atmospheric* 22Jul19), (Xylose + Yeast, Atmospheric* 22Jul19), (Cellobiose + Yeast, Atmospheric* 08Dec19)

*Note the rightmost three colonies in the bottom row were initially growing diagonally within the agar and are on three different carbon sources, but all under atmospheric conditions. All other colonies pictured were growing under microaerobic conditions. Selections were isolated and grown under microaerophilic conditions.

DGGE gels were used to screen 69 isolates (50 from the current defined media collection and 19 from the 2007 collection) which resulted in 30 dereplicated 16S rRNA sequence types: 14 unique isolates within the new defined-media culture collection and 16 from the 2007 *Synoicum adareanum* microbiome culture collection. Final PCR-DGGE gels were prepared for displaying solely unique samples (Figure 4). Total numbers of replicate organisms for each isolate are displayed in Table 2, with the highest number of replicates matching the sample XYL 8 (genus *Polaribacter*), 11 of which were also isolated from xylose plates. Of the taurine-isolated organisms, 6 were replicates of the class Flavobacteriales (and replicates of XYL 8), 3 were Alphaproteobacteria, and 1 was an Actinobacterium.

Phylogenetic analyses of 16S rRNA gene sequences from 46 *S. adareanum* isolates (our 30 dereplicated samples in addition to 16 previously-sequenced isolates from the 2007 *Synoicum adareanum* microbiome culture collection; Murray et al. 2020) indicated our isolates were affiliated within classes Bacilli, Acidimicrobiia in the Firmicutes phylum, Actinobacteria (Actinobacteria phylum) and Flavobacteriia (Bacteroidetes phylum) (Figure 5). Twenty two isolates selected for whole genome sequencing were chosen across a range of the phylogenetic diversity represented.

Whole genome sequencing and genome assembly

Compared to MaSuRCA, SPAdes produced higher quality assemblies and the SPAdes "careful" option produced longer contigs and scaffolds. Assembling with all four files gave approximately 4-fold coverage relative to assembling with one file. Therefore, the SPAdes "careful" option and all four files were used for the final assemblies and

downstream analyses. One isolate, CEL 46, was ultimately unsuccessful in sequencing, likely due to low DNA amounts provided for sequencing. This sample is currently being cultivated for more biomass to perform another DNA extraction and whole genome sequencing.

Most genome lengths were between ~3-6 Mbp, with the largest contigs ranging between 72.7-1,916.6 Kbp long. Quast results showed the assemblies were dominated with short contigs between 0-5,000 bp in length (Figure 6). An exception to this observation, CEL 36 Bin 1, had ~22% of its genome assembly comprised of these shorter contig lengths.

High duplication in BUSCO gene hits (using 124 core bacterial genes found in single copy) for sample CEL 36, a relatively large genome size (9.7 Mbp) and bimodal guanine-cytosine content distribution suggested the presence of more than one organism's sequence in the CEL 36 assembly (Figure 7) (Seppey et al., 2019). The CEL 36 contigs were separated into two bins using MaxBin 2.0 (Wu et al., 2016), which uses read depth and guanine-cytosine content to bin contigs. Bin 1 and Bin 2 were identified as an *Illumatobacter* and *Pseudoalteromonas* and had genome sizes of 5.4 Mbp and 4.3 Mbp, respectively (Table 2). Considering contigs of at least 1,000bp in length, CEL 36 Bin 2 had a genome coverage of 3.54X, whereas all other genomes had coverages between 47.64-147.83X, with an average coverage of 69.48X (Table 3).

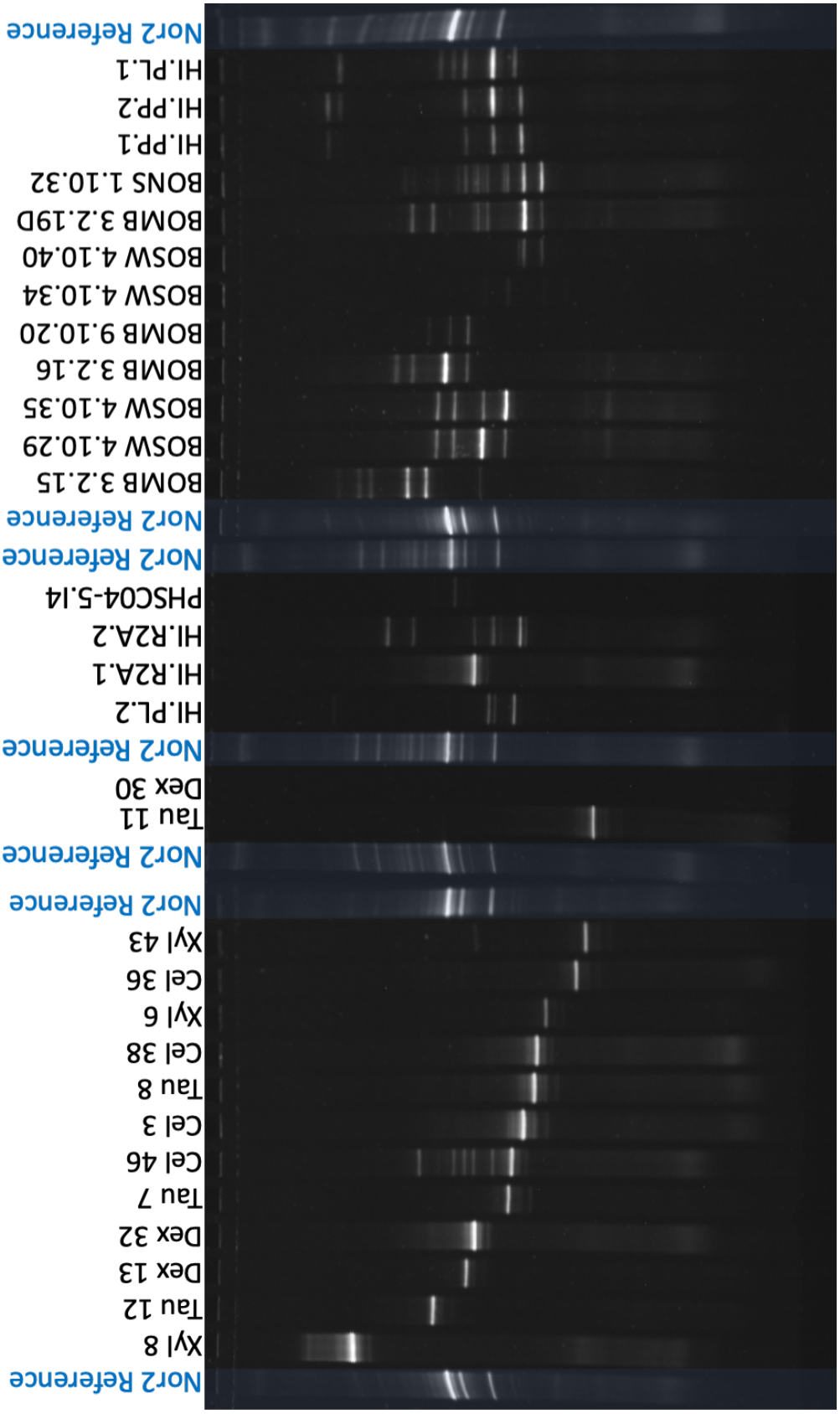


Figure 4. Composite image of 30-65% denaturant PCR-DGGE gels of dereplicated organisms using variable 3 regions of 16s rRNA genes. "Nor2" is a natural microbiome DNA extract used as reference for band patterns between gels and are highlighted in blue. In prior gels, the band for Dex 30 has been observed migrating the furthest of all samples and is presumed to have ran below the gel in this image.

Table 2. Culture collection replicates (total # identical isolates before dereplication), Quast and MultiQC genome assembly statistics.

*500 bp minimum contig length used during binning

Sample Name	# of repli-cates	N50 (Kbp)	N75 (Kbp)	L50 (K)	L75 (K)	GC (%)	Largest contig (Kbp)	Genome length (Mbp)	Total # of contigs	Total # of scaffolds
BOMB 10_10_5	-	96.9	54.7	0.0K	28.0K	39.72	207.7	4.0	456	445
BOMB 10_10_6	-	120.7	55.9	0.0K	23.0K	39.91	331.3	4.0	430	420
BOMB 3_2_15	-	135.7	85.7	0.0K	24.0K	40.24	359.0	5.0	703	696
BOMB 3_2_19D	-	121.9	73.5	0.0K	27.0K	43.63	364.0	5.0	534	518
BOMB 9_10_16	-	119.0	65.2	0.0K	32.0K	44.46	344.5	5.8	636	610
BOMB 9_10_20	-	277.5	163.5	0.0K	14.0K	37.78	578.5	5.4	288	277
BOMB 9_10_21	-	130.0	61.3	0.0K	27.0K	36.84	296.0	4.7	666	650
BOSW 4_10_32	-	154.1	82.1	0.0K	23.0K	41.38	663.0	6.0	552	547
CEL_36 Bin1	-	319.8	117.0	0.0K	15.0K	63.41	569.7	5.4	49*	-
CEL_36 Bin2	-	40.5	19.0	0.0K	72.0K	41.2	199.4	4.3	553*	-
CEL_36	1	166.8	52.5	0.0K	37.0K	53.63	1191.5	9.7	5152	4990
CEL_38	1	321.4	130.3	0.0K	12.0K	53.47	636.4	4.9	130	122
CEL_3	1	438.2	186.2	0.0K	5.0K	61.05	847.2	3.1	35	31
DEX_13	3	1916.6	333.7	0.0K	2.0K	57.7	1916.6	2.9	25	17
DEX_30	1	8.6	3.1	0.2K	488.0K	65.08	72.7	6.3	7244	6969
DEX_32	1	469.5	157.7	0.0K	8.0K	44.72	1243.6	5.5	275	269
TAU_11	2	85.1	50.0	0.0K	30.0K	72.72	420.2	4.0	252	170
TAU_12	1	359.6	261.9	0.0K	7.0K	52.23	979.5	4.4	76	72
TAU_7	4	151.6	62.1	0.0K	23.0K	54.56	617.5	4.9	699	689
TAU_8	2	1076.0	518.9	0.0K	4.0K	55.82	1740.4	5.0	77	76
XYL_43	3	108.2	65.7	0.0K	28.0K	64.29	336.5	4.8	247	214
XYL_6	5	670.7	222.9	0.0K	7.0K	54.81	885.1	4.8	76	70
XYL_8	23	558.1	304.2	0.0K	4.0K	29.46	1573.9	3.6	554	549

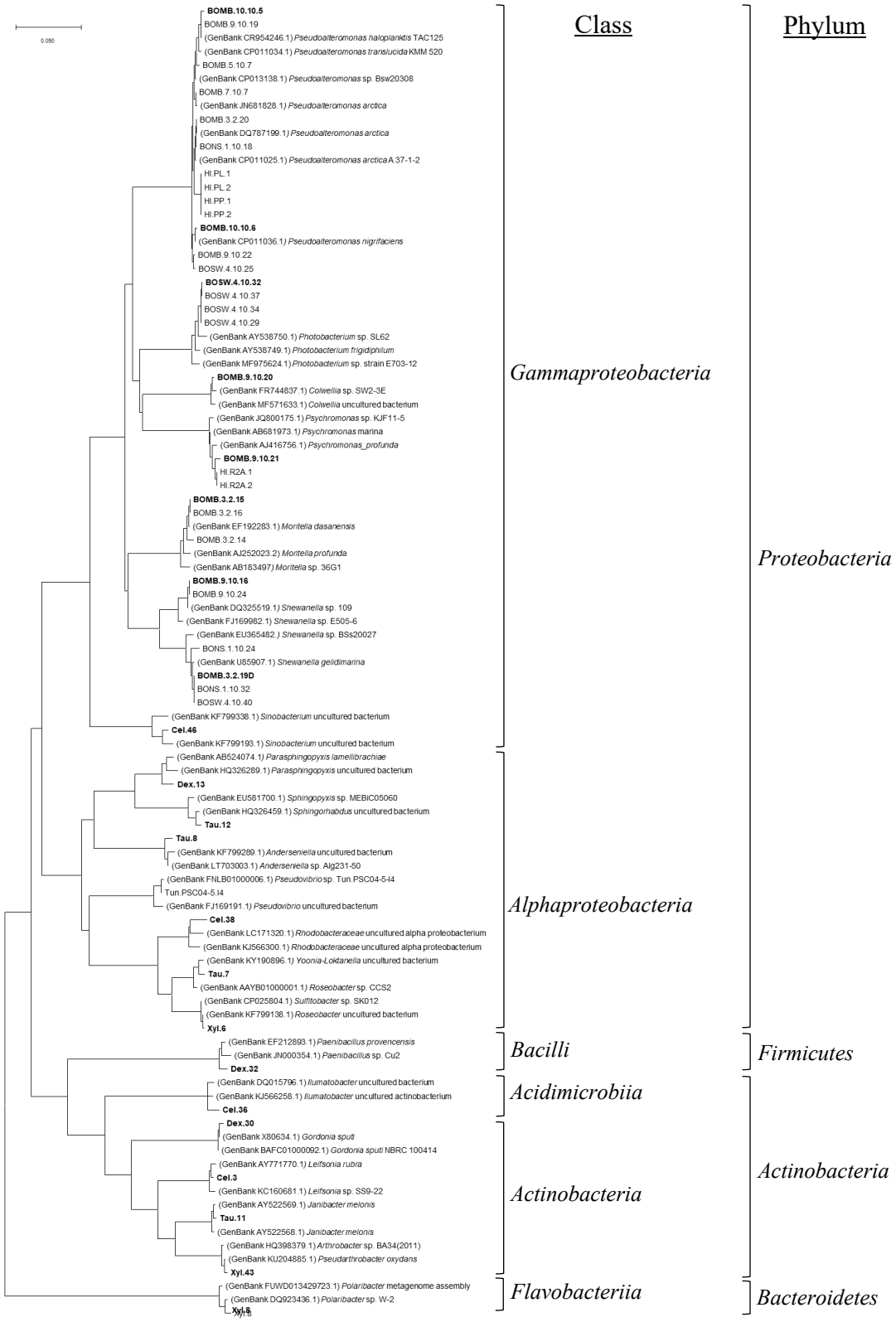


Figure 5. 16S rRNA neighbor-joining phylogenetic tree of isolates and two near neighbors each. Isolates in bold were selected for whole genome sequencing. Note: Previously sequenced isolate *Pseudovibrio* sp. Tun.PSC04-5-I4 was also used in cold adaption analyses.

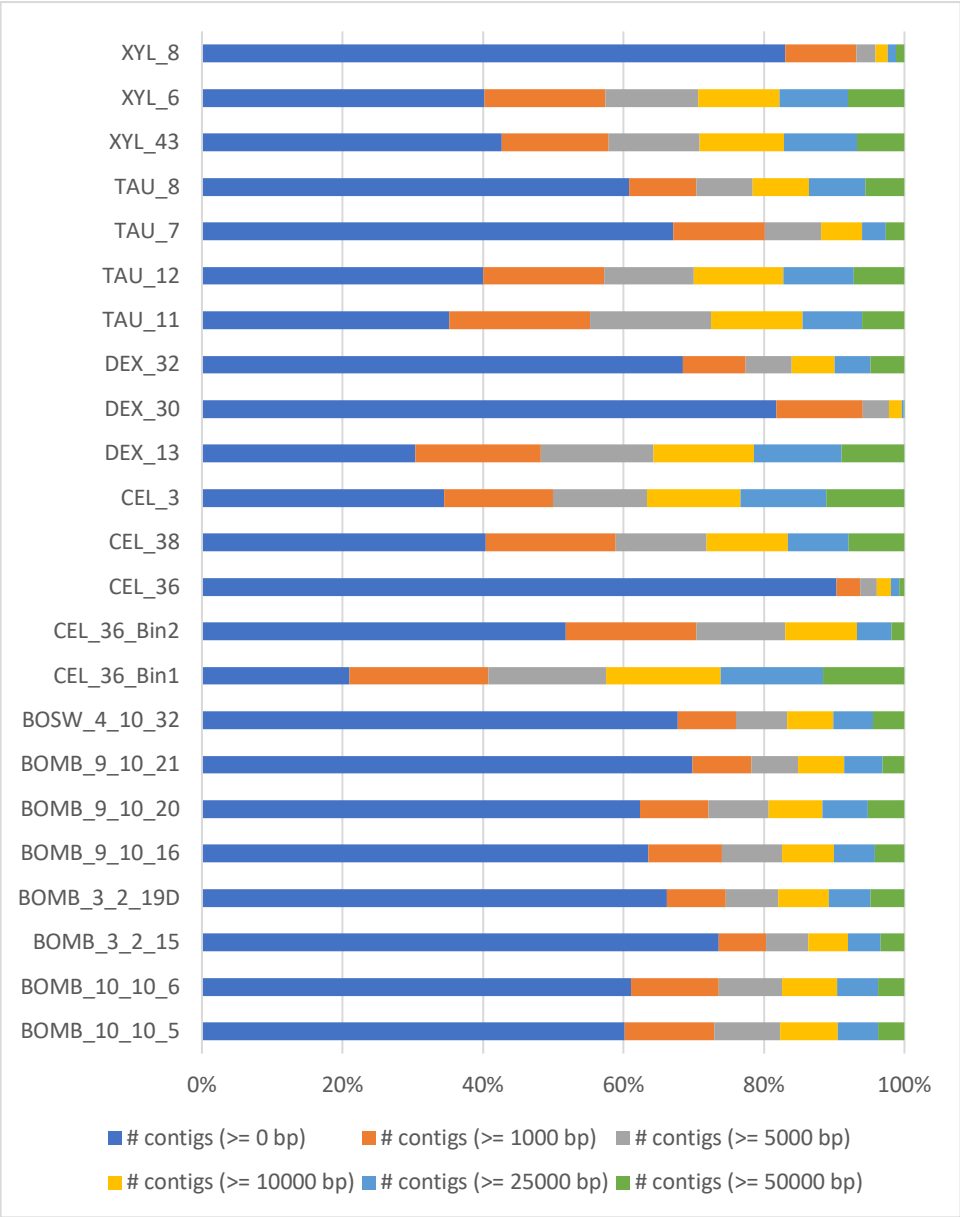


Figure 6. Quast results: percent of contigs by length for each isolate genome assembled in SPAdes assembler using "careful" option.

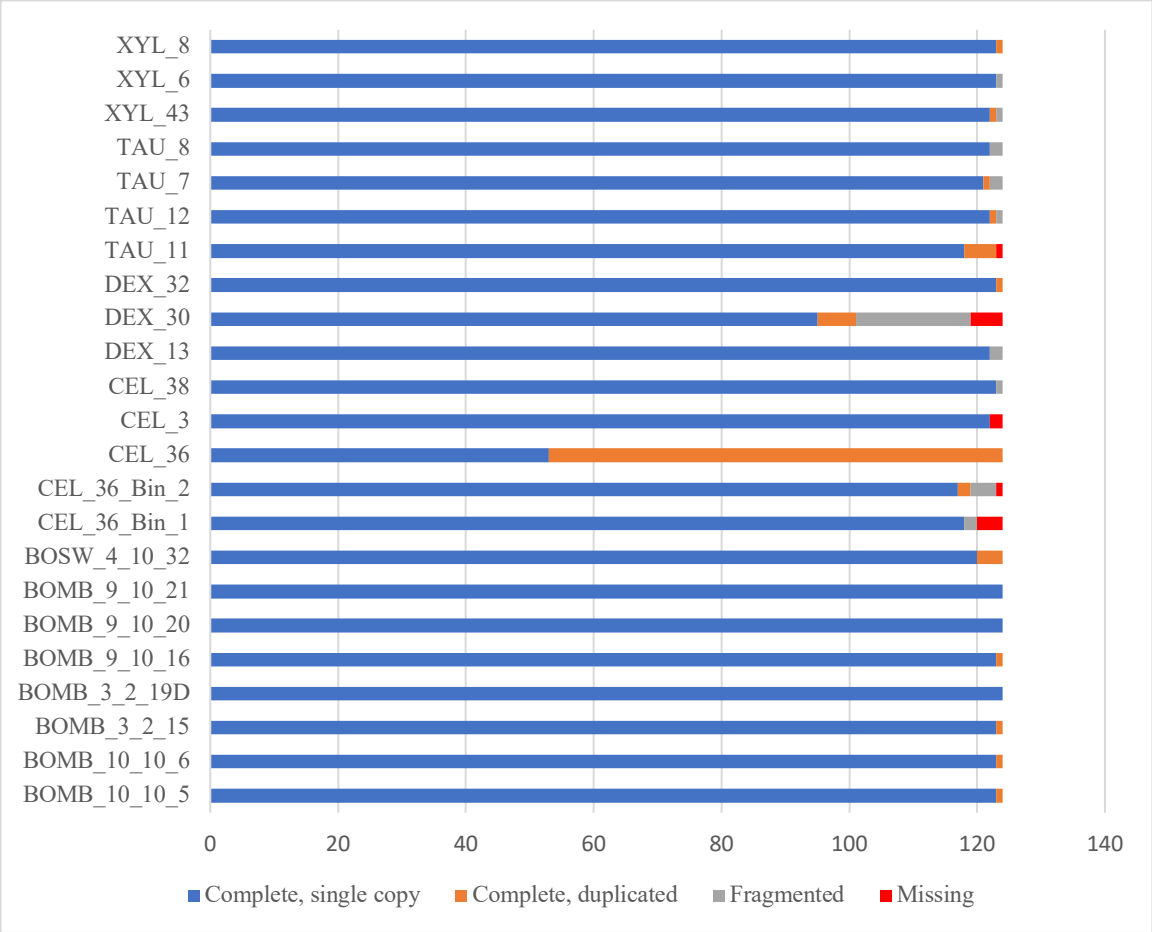


Figure 7. BUSCO results for the presence of 124 core bacterial genes detected within each genome using the BUSCO "bacteria_odb10" database. High duplicate complete genes were observed for Cel 36, one factor leading to suspicions of an impure assembly. Cel 36 was binned into 2 assemblies using MaxBin 2.0 and ran through BUSCO analyses to confirm duplicate BUSCO gene copies were almost entirely removed.

Table 3. Isolate genome coverage of contigs greater than 1,000 bp.

Isolate genome	Contig coverage (contig length>1000bp)
DEX 13	89.07
BOMB 9.10.16	61.6
CEL 36 Bin 1	65.52
CEL 36 Bin 2	3.54
CEL 38	61.42
CEL 3	147.83
DEX 30	47.64
DEX 32	71.18
TAU 11	71.85
TAU 12	63.88
TAU 7	53.85
TAU 8	50.83
XYL 43	80.54
XYL 6	49.81
XYL 8	120.46
BOMB 10.10.5	75.73
BOMB 10.10.6	76.49
BOMB 3.2.15	65.21
BOMB 3.2.19D	50.95
BOMB 9.10.20	79.67
BOMB 9.10.21	69.05
BOSW 4.10.32	72.48

Table 4. Genome novelty predicted using MiGA NCBI Prok tool. ANI % and AAI % were determined and shown here are the closest genome matches. Note: through subsequent searching during cold adaptation comparisons, a related genome of the same genus (*Illumatobacter*) was found for CEL 36 Bin 1 with AAI <65%, indicating this is likely not a new family.

Antarctic isolate	Taxonomic novelty	Evidence: MiGA nearest genome	% ANI	% AAI
CEL 36 Bin 1	New family	*Enterobacter hormaechei* NZ CP041054	-	42.33
TAU 8	New genus	*Nordella* sp. HKS 07 NZ CP049258	-	51.95
DEX 13	New genus	*Sphingomonas* sp. MM 1 NC 020561	-	54.76
CEL 38	New genus	*Sulfitobacter* sp. THAF37 NZ CP045372	-	62.94
DEX 32	New species	*Paenibacillus xylanilyticus* NZ CP044310	-	66.2
BOMB 9.10.21	New species	*Psychromonas* sp. CNPT3 NC 020802	-	67.21
TAU 7	New species	*Yoonia vestfoldensis* NZ CP021431	-	67.22
TAU 12	New species	*Sphingorhabdus* sp. JK6 NZ CP051217	-	68.12
DEX 30	New species	*Gordonia bronchialis* DSM 43247 NC 013441T	-	68.53
XYL 8	New species	*Polaribacter* sp. ALD11 NZ CP025119	85.54	84.98
BOMB 3.2.19D	New species	*Shewanella piezotolerans* WP3 NC 011566T	81.35	85.42
XYL 43	New species	*Arthrobacter* sp. PGP41 NZ CP026514	84.38	85.55
CEL 3	New species	*Salinibacterium* sp. UTAS2018 NZ CP035375	85.37	88.48
BOMB 9.10.16	New species	*Shewanella sediminis* HAW EB3 NC 009831T	87.24	90.06
BOMB 9.10.20	New species	*Colwellia psychrerythraea* 34H NC 003910	89.1	90.66
BOMB 3.2.15	New species	*Moritella marina* ATCC 15381 NZ CP044399T	90.18	92.44
BOSW 4.10.32	New species	*Photobacterium profundum* SS9 NC 006370	92.67	94.46
TAU 11	Same species	*Janibacter melonis* NZ CP044548	97.19	96.01
XYL 6	Same species	*Sulfitobacter* sp. SK012 NZ CP025804	96.03	96.93
CEL 36 Bin 2	Same species	*Pseudoalteromonas haloplanktis* TAC125 NC 007481	98.71	98.58
BOMB 10.10.6	Same species	*Pseudoalteromonas haloplanktis* TAC125 NC 007481	98.76	98.6
BOMB 10.10.5	Same species	*Pseudoalteromonas haloplanktis* TAC125 NC 007481	98.8	98.61

Taxonomic novelty of new *S. adareanum* microbiome isolates

Results from searching our genomes against the NCBI database using the MiGA NCBI Prok tool showed a high level of novelty among most of our culture collection using ANI and AAI % taxonomic criteria (Table 4; (Konstantinidis et al., 2017). Five samples matched a closely related genome > 95% ANI; TAU 11 matched *Janibacter melonis* NZ CP044548 at 97.19% ANI; XYL 6 matched *Sulfitobacter* sp. SK012 NZ CP025804 at 96.03%; and CEL 36 Bin 2, BOMB 10.10.5 and BOMB 10.10.6 all matched *Pseudoalteromonas haloplanktis* TAC125 NC 007481 at 98.71% or higher ANI. Thirteen Antarctic genomes matched closely related genomes between 65-95% AAI, suggesting they are new species within each respective genus. Three Antarctic isolates matched NCBI genomes at 45-65% AAI, indicating these 3 may be new genera. CEL 36 Bin 1 only matched a closely related genome at 42.33% AAI, suggesting this may be a new family.

Genome annotation and secondary metabolite mining

Interesting biosynthetic potential was identified in all isolates by antiSMASH secondary metabolite analyses, with a total of 135 biosynthetic gene cluster types (Table 5). DEX 30 (genus *Gordonia*) had the highest BGC abundance at 21, with NRPS and NRPS-like the predominant types. TAU 8 (genus *Anderseniella*), with 10 types, had the second highest number found, followed by BOMB 3_2_15 (genus *Moritella*), XYL 43 (genus *Pseudarthrobacter*) and CEL 36 Bin 1 (genus *Illumatobacter*) all with 9 types.

Of the 135 biosynthetic gene cluster types identified, the antiSMASH results indicated 49 similar known clusters based on BLAST hits with the MIBiG repository in

Table 5. AntiSMASH results for each culture collection genome. Shading by sample ID. Cluster type glossary: https://docs.antismash.secondarymetabolites.org/glossary/

Sample ID	Region	Type	From	To	Most similar known cluster	Similarity/%	
BOMB 10_10_5	7.1	arylpolyene	31,907	75,493	APE Vf	Other	45%
BOMB 10_10_5	4.1	RiPP-like	87,670	98,527			
BOMB 10_10_6	33.1	arylpolyene	1	30,038	APE Vf	Other	45%
BOMB 10_10_6	2.1	RiPP-like	28,302	39,159			
BOMB 3_2_15	5.1	arylpolyene	145,512	189,142	APE Vf	Other	50%
BOMB 3_2_15	18.1	betalactone	56,229	80,290			
BOMB 3_2_15	11.1	hglE-KS,T1PKS,PUFA	66,555	123,433	eicosapentaenoic acid-like compound	Other	63%
BOMB 3_2_15	8.1	hserlactone	112,566	133,120			
BOMB 3_2_15	3.1	NRPS	44,439	104,844	acinetobactin	NRP	43%
BOMB 3_2_15	35.1	NRPS-like	16,706	48,421			
BOMB 3_2_15	9.1	NRPS-like,betalactone	60,790	106,343			
BOMB 3_2_15	4.1	RiPP-like	5,109	15,951			
BOMB 3_2_15	33.1	RiPP-like	35,623	46,585			
BOMB 3_2_19D	4.1	arylpolyene	31,960	75,722	APE Vf	Other	45%
BOMB 3_2_19D	7.1	betalactone	83,372	112,946	fengycin	NRP	13%
BOMB 3_2_19D	8.1	hglE-KS,PUFA	62,408	119,039	eicosapentaenoic acid	Other	88%
BOMB 3_2_19D	10.1	RiPP-like	100,239	111,075			
BOMB 3_2_19D	12.1	RiPP-like	73,284	84,138			
BOMB 9_10_16	15.1	arylpolyene	1	39,586	APE Vf	Other	45%
BOMB 9_10_16	8.1	betalactone	1	22,471	fengycin	NRP	13%
BOMB 9_10_16	5.1	PUFA,hglE-KS	114,919	171,023	eicosapentaenoic acid	Other	55%
BOMB 9_10_16	10.1	RiPP-like	13,969	24,850			
BOMB 9_10_16	28.1	RiPP-like	17,215	28,051			
BOMB 9_10_20	1.2	hglE-KS,PUFA	443,428	500,742	eicosapentaenoic acid-like compound	Other	18%
BOMB 9_10_20	2.1	hserlactone	103,742	124,566			
BOMB 9_10_20	12.1	redox-cofactor,RiPP-like	111,431	139,119			

Sample ID	Region	Type	From	To	Most similar known cluster	Similarity/%	
BOMB 9_10_20	5.1	RiPP-like	234,000	244,932			
BOMB 9_10_20	12.2	RiPP-like	162,569	173,510			
BOMB 9_10_20	18.1	RRE-containing	89,150	109,458			
BOMB 9_10_20	1.1	thiopeptide	8,773	39,560			
BOMB 9_10_21	36.1	arylpolyene	7,567	39,447	APE Vf	Other	85%
BOMB 9_10_21	28.1	PUFA,hglE-KS	1	51,422	eicosapentaenoic acid-like compound	Other	13%
BOMB 9_10_21	5.1	RiPP-like	49,711	60,562			
BOMB 9_10_21	20.1	RiPP-like	46,660	56,815			
BOSW 4_10_32	29.1	arylpolyene	41,341	71,605	APE Vf	Other	95%
BOSW 4_10_32	26.1	betalactone	48,493	77,791			
BOSW 4_10_32	3.1	ectoine	36,708	47,094	ectoine	Other	66%
BOSW 4_10_32	6.1	hglE-KS,PUFA	123,720	178,917	eicoseicosapentaenoic acid	Other	47%
BOSW 4_10_32	4.1	lassopeptide	216,424	241,100			
BOSW 4_10_32	39.1	NRPS	1	41,514	amonabactin P 750	NRP	100%
BOSW 4_10_32	18.1	RiPP-like	28,314	39,174			
BOSW 4_10_32	37.1	thiopeptide	12,128	39,524			
CEL 3	4.1	betalactone	253,104	280,402	sisomicin	Saccharide	5%
CEL 3	1.1	T3PKS	171,805	212,863	alkylresorcinol	Polyketide	100%
CEL 3	1.2	terpene	373,217	394,083	carotenoid	Terpene	50%
CEL 38	4.1	hserlactone	310,237	330,881			
CEL 38	8.1	hserlactone	67,806	88,414			
CEL 38	9.1	hserlactone	147,747	168,337			
CEL 38	36.1	hserlactone	1	9,780			
CEL 38	22.1	NRPS	1	38,597			
CEL 38	7.1	RiPP-like	151,575	162,435			
CEL 38	15.1	RRE-containing	57,952	79,202			
DEX 13	1.1	RiPP-like	1,835,928	1,846,782			
DEX 13	5.1	T3PKS	1	33,047			
DEX 13	5.2	terpene	114,108	137,687	carotenoid	Terpene	42%
DEX 30	84.1	ectoine	1	9,126	ectoine	Other	75%
DEX 30	109.1	NAPAA	1	12,245			
DEX 30	6.1	NRPS	1	54,541	ishigamide	NRP + Polyketide	11%
DEX 30	32.1	NRPS	1	24,188			

Sample ID	Region	Type	From	To	Most similar known cluster	Similarity/%	
DEX 30	60.1	NRPS	1	17,333			
DEX 30	166.1	NRPS	1	9,480	bacillibactin	NRP	15%
DEX 30	183.1	NRPS	1	8,422			
DEX 30	187.1	NRPS	1	8,235	heterobactin A & S2	NRP	27%
DEX 30	293.1	NRPS	1	5,265			
DEX 30	369.1	NRPS	1	4,443			
DEX 30	558.1	NRPS	1	2,538			
DEX 30	614.1	NRPS	1	2,246			
DEX 30	276.1	NRPS-like	1	5,569			
DEX 30	318.1	NRPS-like	1	4,976			
DEX 30	426.1	NRPS-like	1	3,637	ishigamide	NRP + Polyketide	11%
DEX 30	124.1	redox-cofactor	1	11,457			
DEX 30	81.1	RiPP-like	1	10,206	pimaricin	Polyketide	11%
DEX 30	270.1	RiPP-like	1	5,662			
DEX 30	467.1	T1PKS	1	3,273			
DEX 30	105.1	terpene	1	10,878	carotenoid	Terpene	33%
DEX 30	174.1	terpene	1	8,586			
DEX 32	6.1	arylpolyene,ladderane	78,621	121,873			
DEX 32	7.1	lassopeptide	63,580	87,541			
DEX 32	2.2	NRPS-like	776,872	820,582			
DEX 32	1.1	T3PKS,NRPS	361,004	455,266	ulleungmycin	NRP	5%
DEX 32	2.1	terpene	180,386	202,263			
DEX 32	3.1	terpene	476,069	497,994			
TAU 11	10.1	redox-cofactor	53,572	76,408			
TAU 11	1.1	siderophore	60,414	72,615			
TAU 11	1.2	siderophore	347,480	359,864			
TAU 11	31.1	terpene	4,933	25,820	carotenoid	Terpene	66%
TAU 12	12.1	RiPP-like	22,105	32,989	N-tetradecanoyl tyrosine	Other	6%
TAU 12	6.1	T3PKS	14,741	55,790	berninamycin A	RiPP	26%
TAU 12	4.1	terpene	235,053	258,726	zeaxanthin	Terpene	66%
TAU 7	3.1	hserlactone	132,861	153,538			
TAU 7	17.1	NRPS	30,484	83,854	serobactin A, B & C	NRP:NRP siderophore	15%
TAU 7	7.1	NRPS-like,ectoine	1	29,623	ectoine	Other	100 %
TAU 7	12.1	NRPS,T1PKS	22,771	75,284			
TAU 7	6.1	RiPP-like	13,536	24,390			
TAU 7	9.1	terpene	126,606	147,373			
TAU 8	2.3	arylpolyene	550,739	591,848			
TAU 8	4.1	ectoine	349,379	359,762			
TAU 8	6.1	NAPAA	1	18,200			
TAU 8	1.3	NRPS	610,974	671,027	chejuenolide A & B	Polyketide	7%
TAU 8	1.2	RRE-containing	308,189	328,479			
TAU 8	3.2	RRE-containing	308,621	329,055			
TAU 8	2.1	T3PKS	18,662	59,852			
TAU 8	2.2	terpene	317,408	344,802	hopene	Terpene	15%
TAU 8	3.1	terpene	184,825	205,682			

Sample ID	Region	Type	From	To	Most similar known cluster	Similarity/%	
TAU 8	1.1	thioamitides	52,500	74,478			
XYL 43	19.1	betalactone	26,898	52,312	microansamycin	Polyketide	7%
XYL 43	3.1	butyrolactone	31,224	42,300			
XYL 43	1.1	NAPAA	258,445	292,521	valinomycin / montanastatin	NRP + Saccharide: Hybrid/tailoring	17%
XYL 43	4.1	NAPAA	40,611	74,687			
XYL 43	13.2	NAPAA	65,853	101,554	stenothricin	NRP:Cyclic depsipeptide	13%
XYL 43	23.1	NAPAA	40,239	74,219	stenothricin	NRP:Cyclic depsipeptide	31%
XYL 43	7.1	RRE-containing	87,448	107,729	JBIR-06	NRP + Polyketide	16%
XYL 43	30.1	siderophore	28,132	40,015	desferrioxamine E	Other	100 %
XYL 43	13.1	T3PKS	1	33,222			
XYL 6	4.2	betalactone	223,549	251,066			
XYL 6	2.1	hserlactone	750,438	771,073			
XYL 6	4.3	hserlactone	394,150	414,875			
XYL 6	1.1	NRPS	406,416	450,330	ravidomycin	Polyketide	5%
XYL 6	6.1	RiPP-like	161,676	172,539			
XYL 6	4.1	T1PKS,NRPS-like	22,831	74,710			
XYL 8	11.1	RiPP-like	1	9,478			
XYL 8	1.2	T3PKS	242,618	283,670			
XYL 8	1.1	terpene	24,158	45,306			
XYL 8	2.1	terpene	236,112	256,954	carotenoid	Terpene	28%
CEL 36 Bin.002	24.1	RiPP-like	11,918	22,775			
CEL 36 Bin.002	107.1	arylpolyene	1	10,248	APE Vf	Other	15%
CEL 36 Bin.002	148.1	NRPS	1	3,157	rhizomide A, B & C	NRP	100 %
CEL 36 Bin.001	1.1	RRE-containing	395,414	415,845			
CEL 36 Bin.001	3.1	ranthipeptide	188,053	209,440			
CEL 36 Bin.001	5.1	terpene	141,597	162,823			
CEL 36 Bin.001	6.1	T1PKS,NRPS-like	153,428	204,491			
CEL 36 Bin.001	6.2	T1PKS,hglE-KS	293,093	322,710			
CEL 36 Bin.001	12.1	NRPS	54,406	107,773			
CEL 36 Bin.001	14.1	hglE-KS	1	23,433			
CEL 36 Bin.001	14.2	betalactone	25,569	57,603	quartromicin A1	Polyketide	5%
CEL 36 Bin.001	19.1	terpene	14,749	38,347			

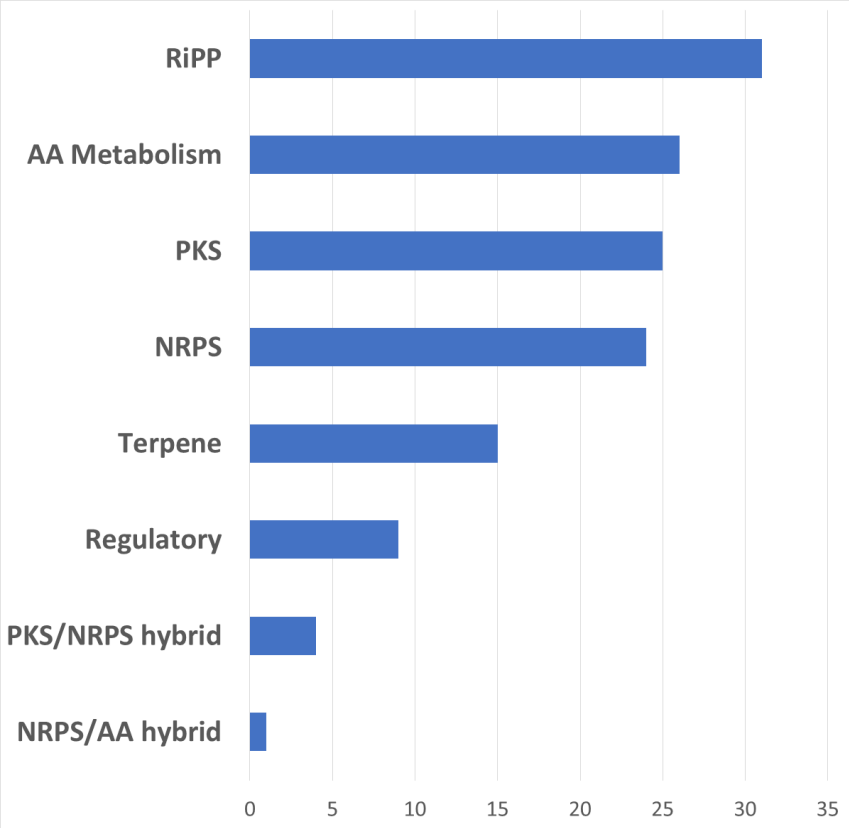


Figure 8. Frequency of biosynthetic gene cluster categories detected by antiSMASH for *Synoicum adareanum* microbiome cultures. Categories by type were structured as follows:

RiPP: included RiPP-like, siderophore, thiopeptide, lassopeptide, thioamitides, and ranthipeptide.
AA metabolism: included hserlactone, betalactone, NAPAA, and ectoine.
PKS: included arylpolyene, all PK types, hgIE-KS, PUFA, butyrolactone, arylpolyene, and ladderane.
NRPS: included NRPS, NRPS-like, "NRPS-like, ectoine."
Terpene: included terpene.
Regulatory: included RRE-containing, redox-cofactor, "redox-cofactor, RiPP-like."
PKS/NRPS hybrid: included "T1PKS, NRPS-like", "T3PKS, NRPS", "NRPS, T1PKS"
NRPS/AA hybrid: include "NRPS-like, betalactone"

Cluster type glossary: https://docs.antismash.secondarymetabolites.org/glossary/

which putative homologs were identified (Kautsar et al., 2020). Of these 49 known clusters, only 16 show similarities of 50% or higher (of which 3 show similarities between 80-99%, with 5 being 100% similar). The 5 that have a 100% similar known cluster include: a NRPS in the BOSW 4.10.32 (genus *Photobacterium*) genome which matches a NRP amonabactin P750; a type 3 PKS found in the CEL 3 (genus *Leifsonia*) genome which matches a polyketide alkylresorcinol; a "NRPS-like, ectoine" found in the TAU 7 (genus *Yoonia*) genome which matches an ectoine; a siderophore in the XYL 43 (genus *Pseudarthrobacter*) genome which matches a desferrioxamine E; and a NRPS in the CEL 36 Bin 2 (genus *Pseudoalteromonas*) genome which matches NRP rhizomide A, B & C. These BGC types were placed into 7 biosynthetic categories; ribosomally synthesized and post-translationally modified peptide (RiPPs, 31 total), amino acid metabolism (26 total), polyketide synthase (PKS, 25 total)), nonribosomal peptide synthetase (NRPS, 24 total), terpene (15 total), regulatory (9 total), and PKS/NRPS and NRPRS/amino acid hybrid (4 and 1, respectively) (Figure 8).

Cold adaptation analysis

Comparisons of orthologs shared between the Antarctic bacterial genomes and low/mid latitude genomes from the same genus (Table 6) were analyzed in two ways, 1) the total number of orthologs for which higher content was calculated for each parameter, by Antarctic or low/mid latitude genome, and 2) mean values for each parameter with standard deviation bars, by both Antarctic and low/mid latitude genome (Figure 9). Mean values typically followed the general trend of the total orthologs found in the genomes tested.

The direction of higher content between each comparison revealed more orthologs from low/mid latitude organisms which displayed higher GC content, disorder, and arginine/(arginine + lysine) (Figures 10 & 11). In three cases, GC content was higher in low/mid open reading frame orthologs even when genome-wide GC content was higher for the Antarctic genome (Table 7). The Antarctic genomes had a greater number of predicted protein orthologs displaying higher asparagine, serine and threonine within the polar amino acids, whereas the low/mid latitude genomes had a greater number of orthologs with higher glutamine and tyrosine (Figures 9 and 11). Of the non-polar amino acids, the Antarctic genomes contained more orthologs with higher glycine, isoleucine, leucine, and phenylalanine (Figures 9 and 11), and low/mid latitude genomes contained more orthologs with higher alanine, proline, tryptophan and valine (Figures 9 and 11). Across all comparisons, a greater number of orthologs from the low/mid latitude genomes had higher charged (+ and -) amino acids (with the exception of lysine) but also had a higher arginine/(arginine+lysine) ratio (Figures 9 and 10).

The heat map (Figure 12) shows the results from Figures 9-11 compiled by comparison number across all parameters tested. The heatmap (Figure 13) displays the parameters considered in the cold adaptation analyses and assigned the directional trend of expected cold-adapted trait for each comparison. The Antarctic genomes presented a higher number of cold adapted traits for all parameters overall, except disorder. For disorder, which is understood to increase in cold-adapted organisms, the low/mid latitude genomes had a greater number of orthologs with higher disorder at both IUPred score thresholds.

Table 6. Culture collection genomes, selected low/mid latitude genome data and Average Amino acid Identity (AAI) shared between each genome. *Strain AIST1 isolated from seawater, but only 16S rRNA sequenced. For analyses, whole genome of *Gordonia terrae* isolate from soil used.

Com-paris-on #	*S. adareanum* microbiome isolate	Low/mid latitude representative NCBI organism name	AAI	Sample location	Latitude	Longitude	GenBank assembly accession	References
1	BOMB 3.2.15	*Moritella* sp. F1	92.71	Japan: Kanagawa, 500m depth	34.4	139.48	GCF_015082235.1	-
2	BOMB 3.2.19D	*Shewanella maritima*	66.18	South Korea: Dokdo Island, surface seawater	37.14	131.52	GCF_004295345.1	(Bae et al., 2020)
3	BOMB 9.10.16	*Shewanella maritima*	65.67	South Korea: Dokdo Island, surface seawater	37.14	131.52	GCF_004295345.1	(Bae et al., 2020)
4	BOMB 9.10.20	*Litorilituus lipolyticus*	70.31	China: Yellow Sea, intertidal sand	35.47	119.6	GCF_006439335.1	(Liu et al., 2020)
5	BOMB 9.10.21	*Psychromonas* sp. SR45-3	68.88	USA: Chesapeake Bay, Maryland, surface seawater	38.988883	-76.481891	GCF_014163575.1	-
6	BOMB 10.10.5	*Pseudoalteromonas atlantica*	82.11	China: East China Sea, coastal seawater, 0.5-1 m depth	30.43	122.46	GCF_002374815.1	(Wang et al., 2019)
7	BOMB 10.10.6	*Pseudoalteromonas atlantica*	81.92	China: East China Sea, coastal seawater, 0.5-1 m depth	30.43	122.46	GCF_002374815.1	(Wang et al., 2019)
8	BOSW 4.10.32	*Photobacterium leiognathi*	70.95	Cuba: Varadero, shore seawater	23.199664	-81.159001	GCF_001557755.1	-
9	CEL 3	*Salinibacterium amurskyense*	89.98	East Sea, Amursky Bay, seawater, 5 m depth	-	-	GCF_003668745.1	(Han et al., 2003)
10	CEL 36 Bin1	*Ilumatobacter coccineus* YM16-304	66.41	Japan: Nonami Beach, seashore sand	35.578	133.096	GCF_000348785.1	(Matsumoto et al., 2013)
11	CEL 36 Bin2	*Pseudoalteromonas atlantica*	81.82	China: East China Sea, coastal seawater, 0.5-1 m depth	30.43	122.46	GCF_002374815.1	(Wang et al., 2019)
12	CEL 38	*Marinosulfonomonas* sp. NORP110 [MAG]	66.46	>4,500 m depth (subseafloor)	-	-	GCA_002732955.1	(Tully et al., 2018)

Com-paris-on #	*S. adareanum* microbiome isolate	Low/mid latitude representative NCBI organism name	AAI	Sample location	Latitude	Longitude	GenBank assembly accession	References
13	DEX 13	*Parasphingopyxis* sp. CP4	78.32	South Korea: Chujado island, seawater, 20-30m depth	-	-	GCF_013378055.1	(Kwon et al., 2020)
14	DEX 30	*Gordonia terrae*	68.8	Strain AIST1: Inland Sea of Japan, surface seawater*.	39.13	117.2	GCF_901542405.1	(Takaichi et al., 2008; Tsukamur.M, 1971)
15	DEX 32	*Paenibacillus* sp. CAA11	61.05	South Korea: Dongmak foreshore, tidal flat sediment	-	-	GCF_003060825.1	(Kim et al., 2018)
16	TAU 7	*Yoonia rosea*	76.27	Sea of Japan, Gulf of Peter the Great, sediments, 13 m maximum depth	-	-	GCF_900156505.1	(Ivanova et al., 2005; Wirth & Whitman, 2018)
17	TAU 8	*Anderseniella* sp. [MAG]	89.94	Australia: Port Phillip Bay, intertidal zone, beach sand	37.863 S	144.971 E	GCA_013041995.1	-
18	TAU 11	*Janibacter cremeus*	65.17	Japan: Rishiri Island foreshore, marine sediment	-	-	GCF_013409205.1	(Hamada et al., 2013)
19	TAU 12	*Parasphingorhabdus marina* DSM 22363	74.19	Korea: Coast of Jeju Island, surface seawater	33.51	126.52	GCF_900128895.1	(Feng et al., 2020; Kim et al., 2008)
20	XYL 6	*Sulfitobacter* sp. NAS-14.1	70.29	N. Atlantic Ocean, Sargasso Sea, surface seawater	33°3′N	74°3′W	GCF_000152645.1	(Slightom & Buchan, 2009)
21	XYL 8	*Polaribacter* sp. MED152	74.36	NW Mediterranean Sea, 1 m depth	41.66667	2.8	GCF_000152945.2	(Gonzalez et al., 2008)
22	XYL 43	*Pseudarthrobacter enclensis*	81.97	India: Chorao island, marine sediments	15°32′34″	73°55′15″	GCF_001457025.1	(Busse & Schumann, 2019; Dastager et al., 2014)
23	*Pseudovibrio* sp. TunPHSC 04-5.I4	*Pseudovibrio denitrificans* JCM 12308	81.45	Taiwan: Nanwan Bay, Kenting National Park, shallow seawater	21.942622	120.76492	GCF_001310815.1	(Shieh et al., 2004)

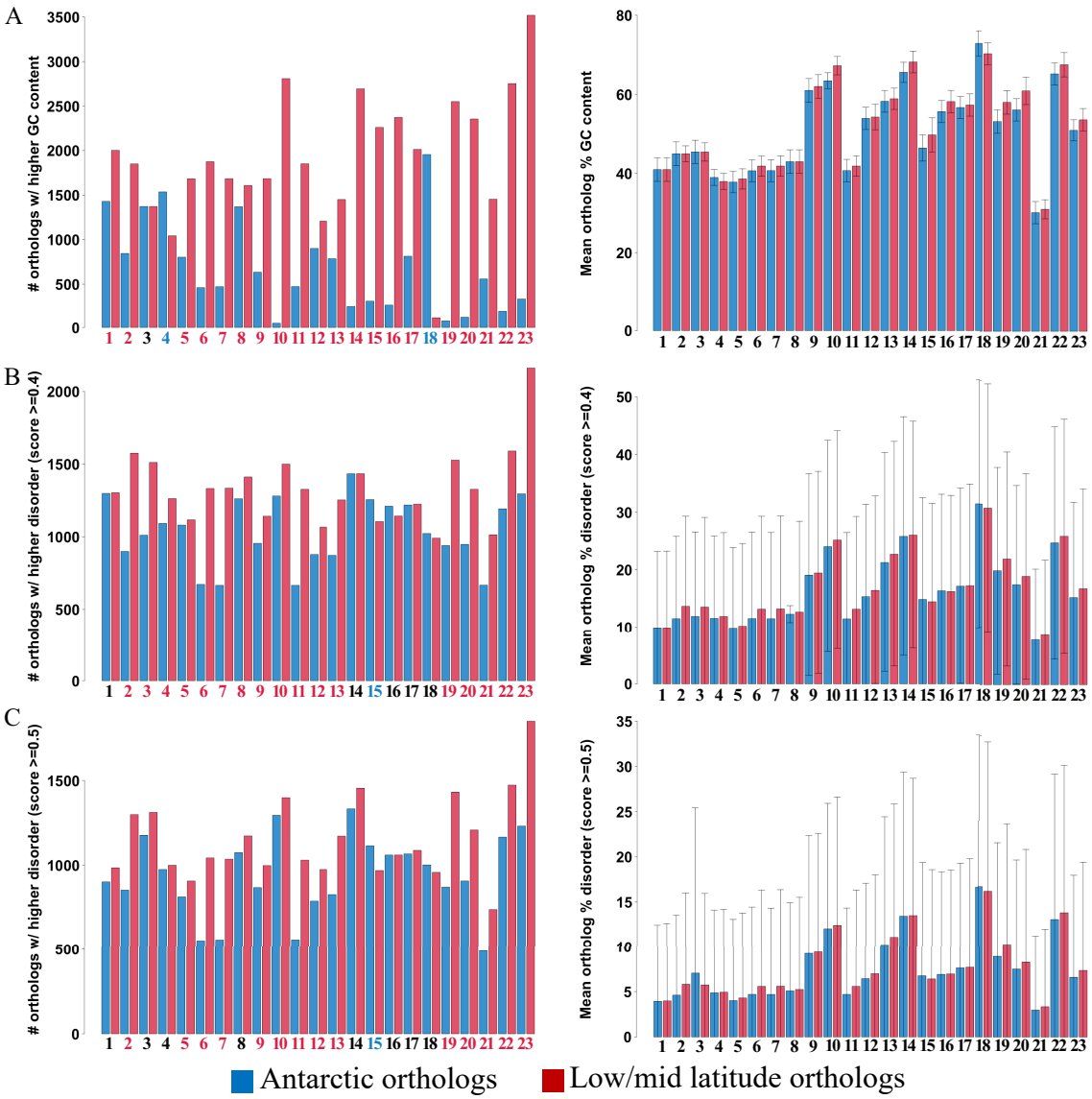
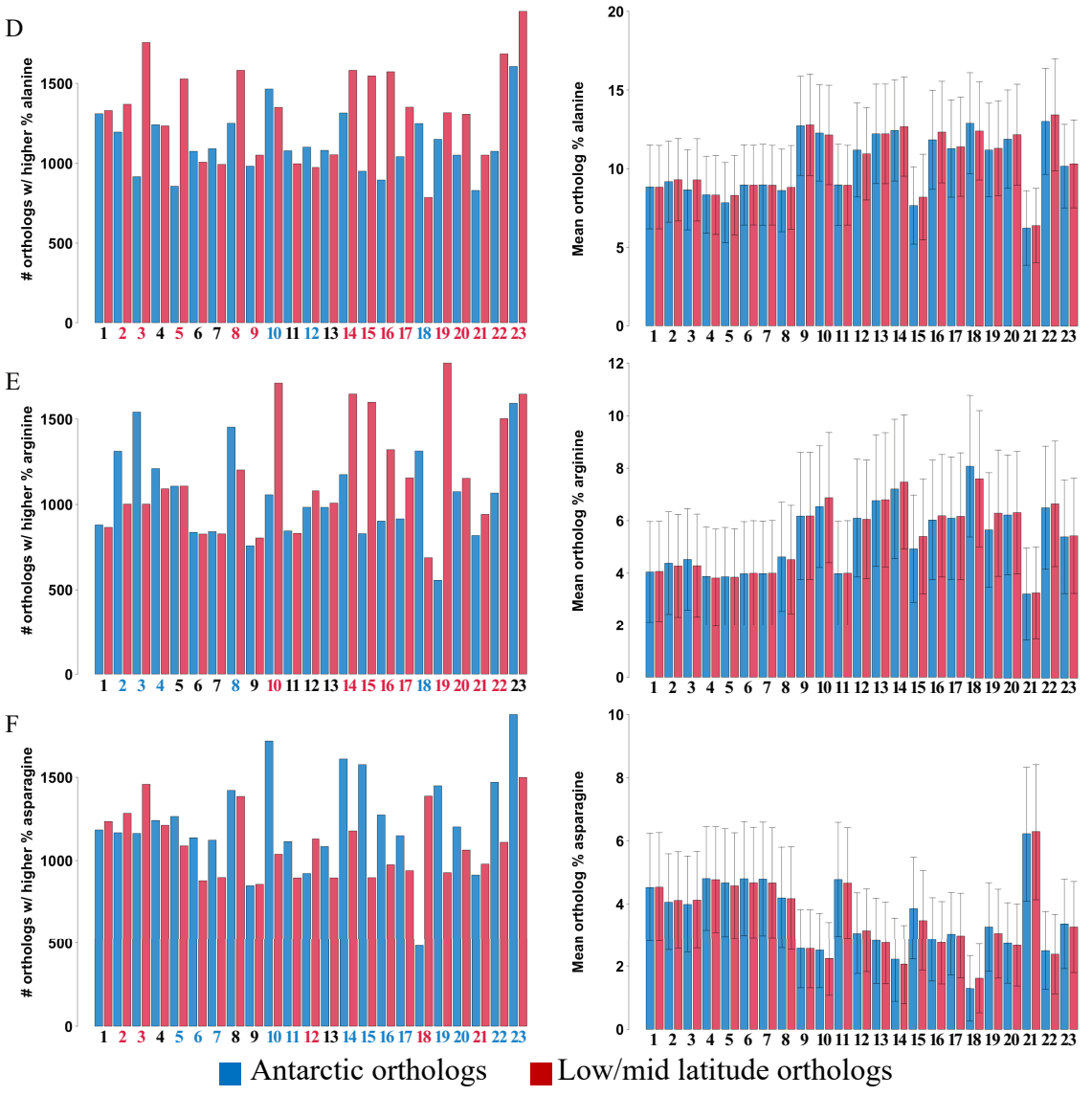
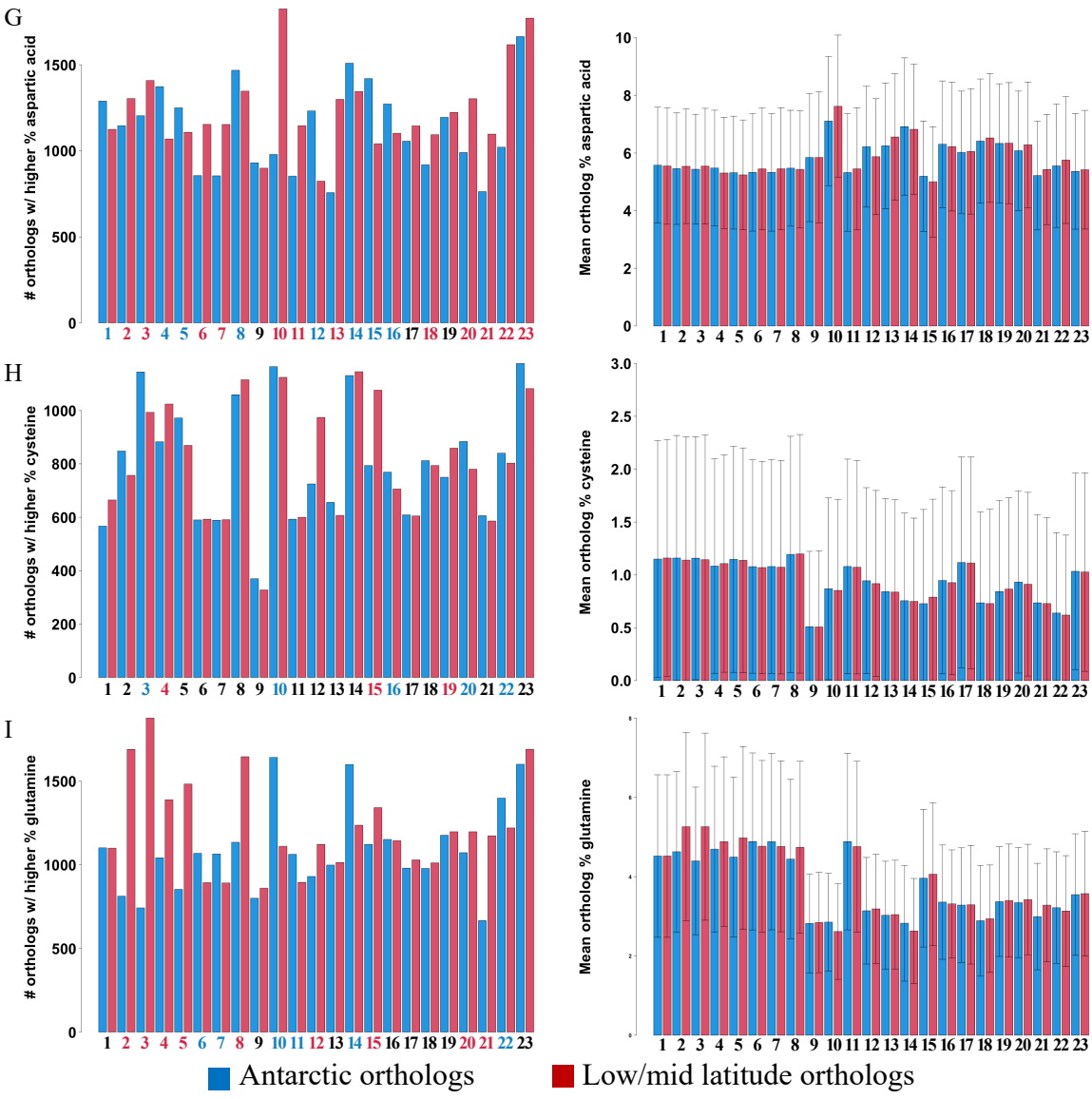
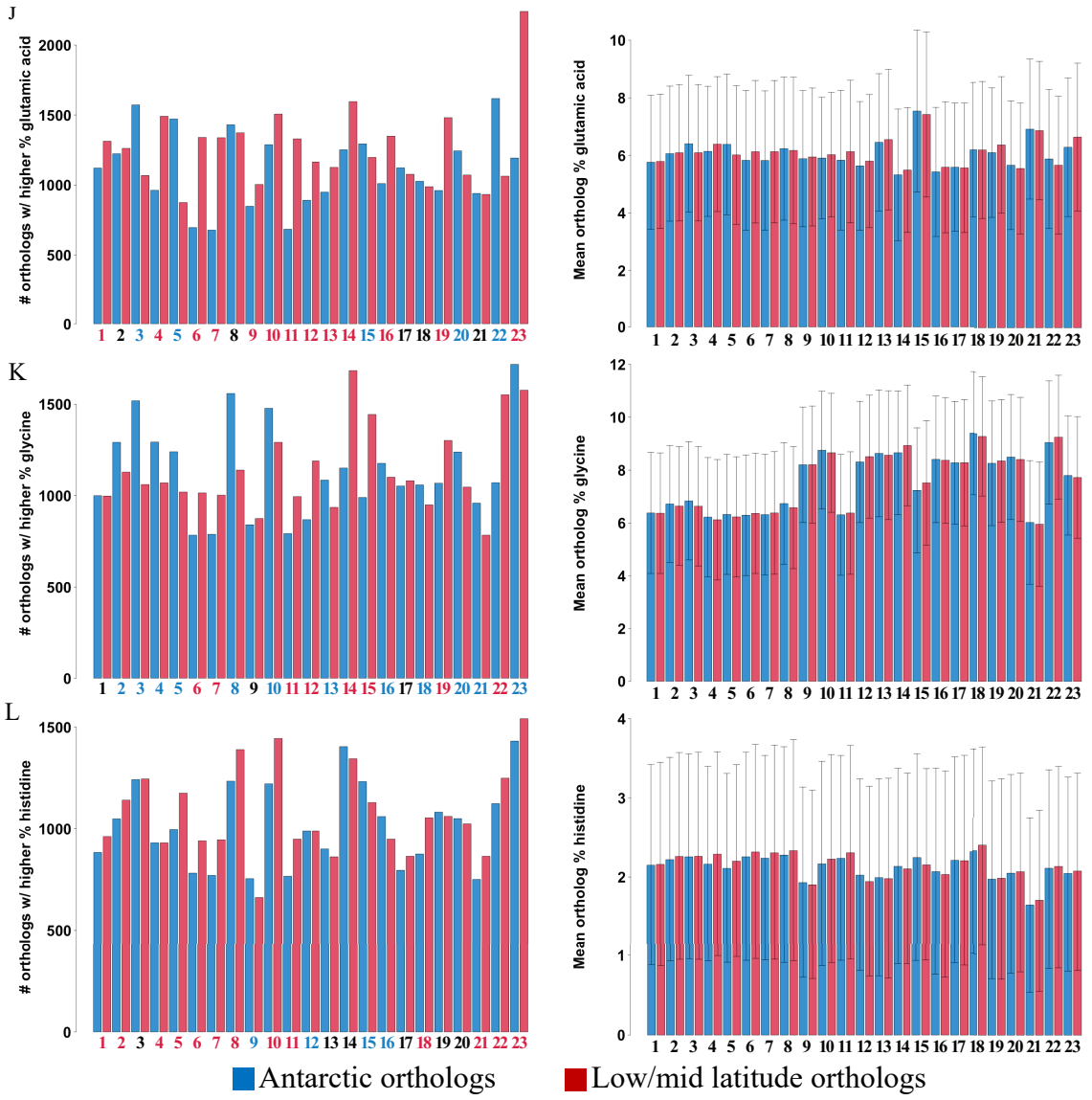


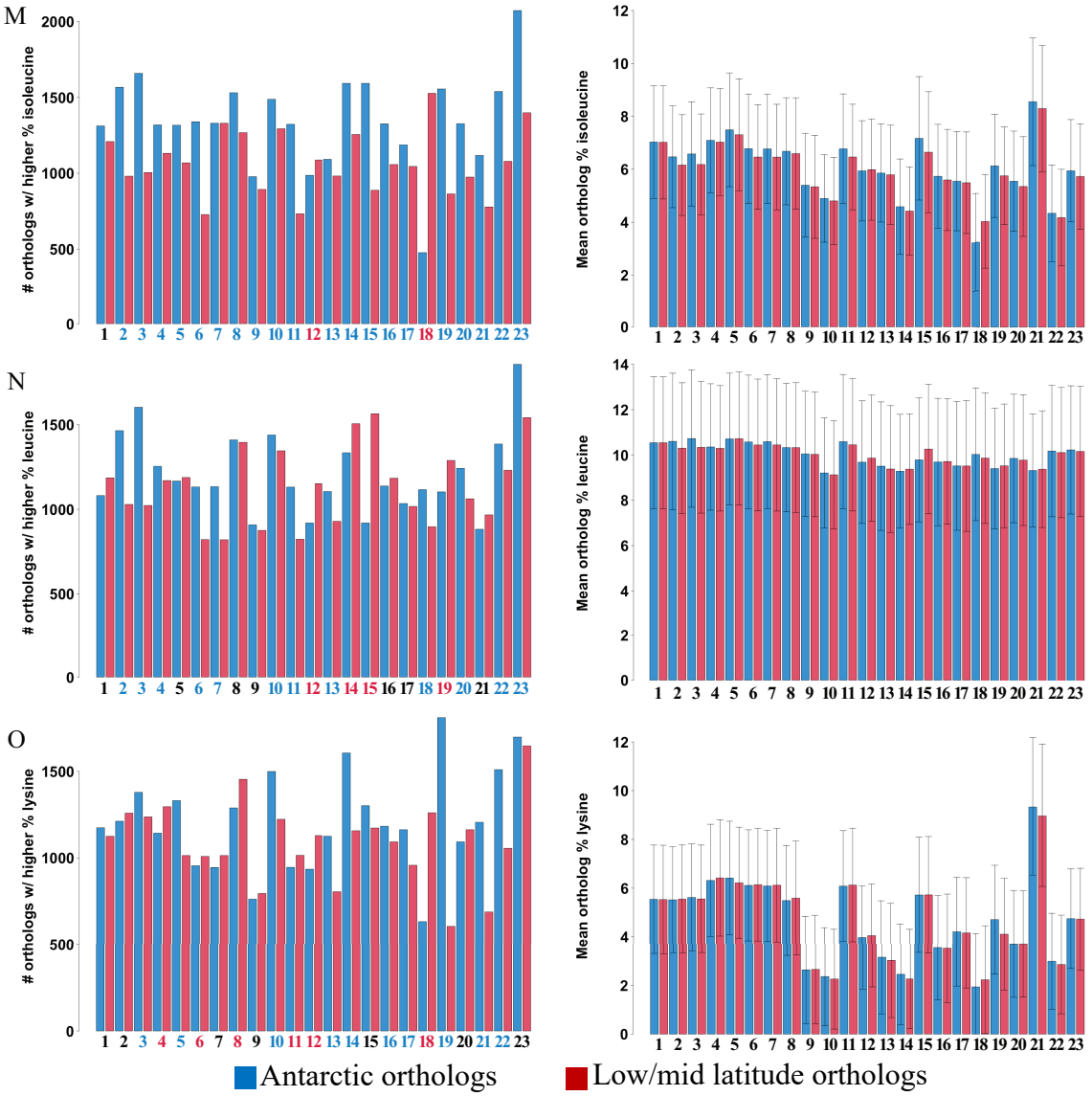
Figure 9. Paired orthologous protein analysis for G-C content, disorder and amino acid content across all 23 isolate genomes. For each panel (A-X), left bar plots show number of orthologs for which each parameter tested was found to be greater in Antarctic genome or representative low/mid latitude genome of the same genus. Colored comparison pair numbers are those found to be significantly different (colored had p-values ≤ 0.05; black numbers had p-values ≥ 0.05) using paired Wilcoxon signed-rank tests. Blue and red numbers show the number of orthologs which had significantly higher amounts or ratios of the parameter tested for the Antarctic genome or low/mid latitude genome, respectively.

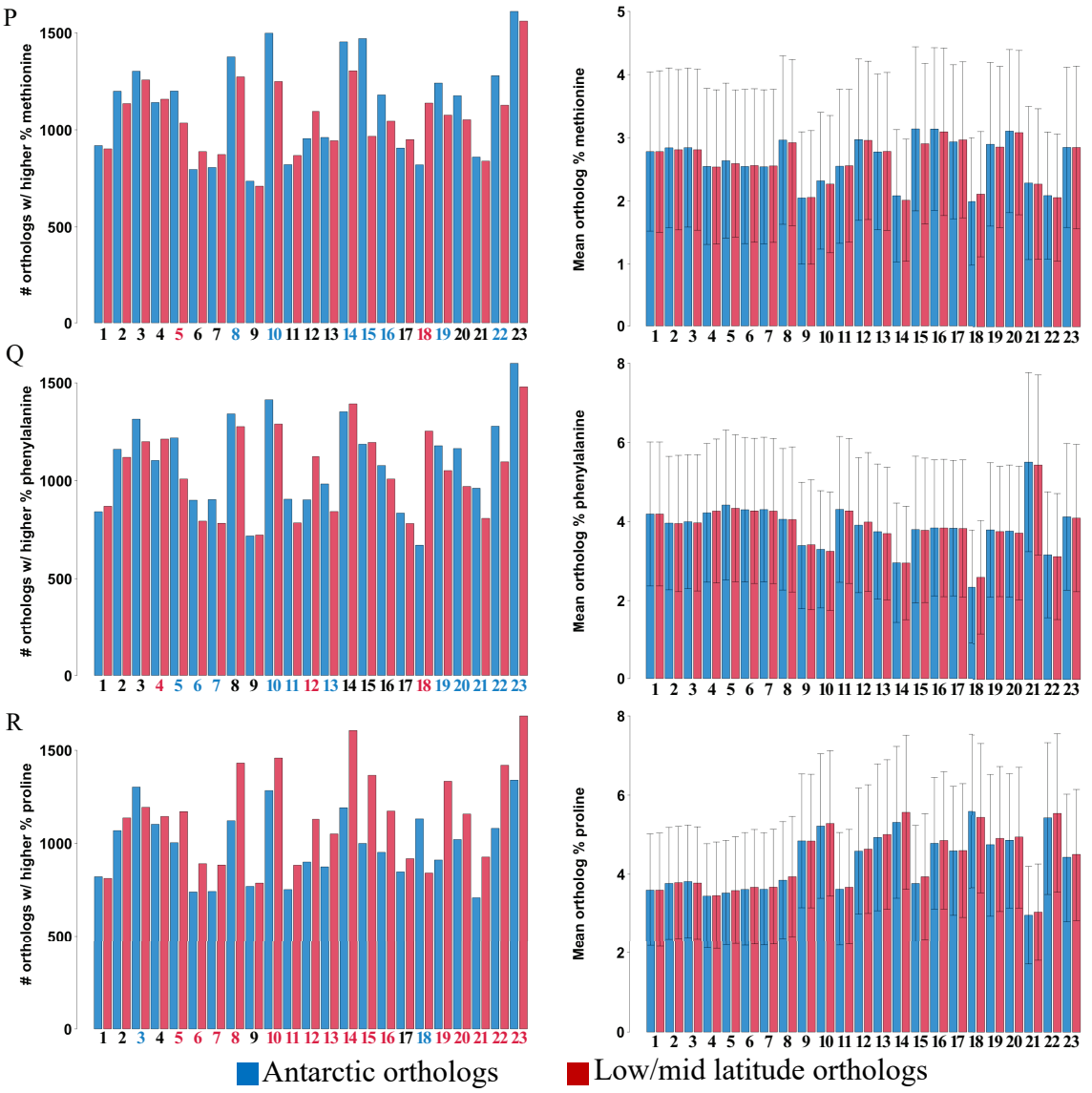
Right bar plots show mean percentage of each parameter found across all orthologs in the Antarctic genome (blue) or low/mid latitude genome (red), with standard deviation bars.

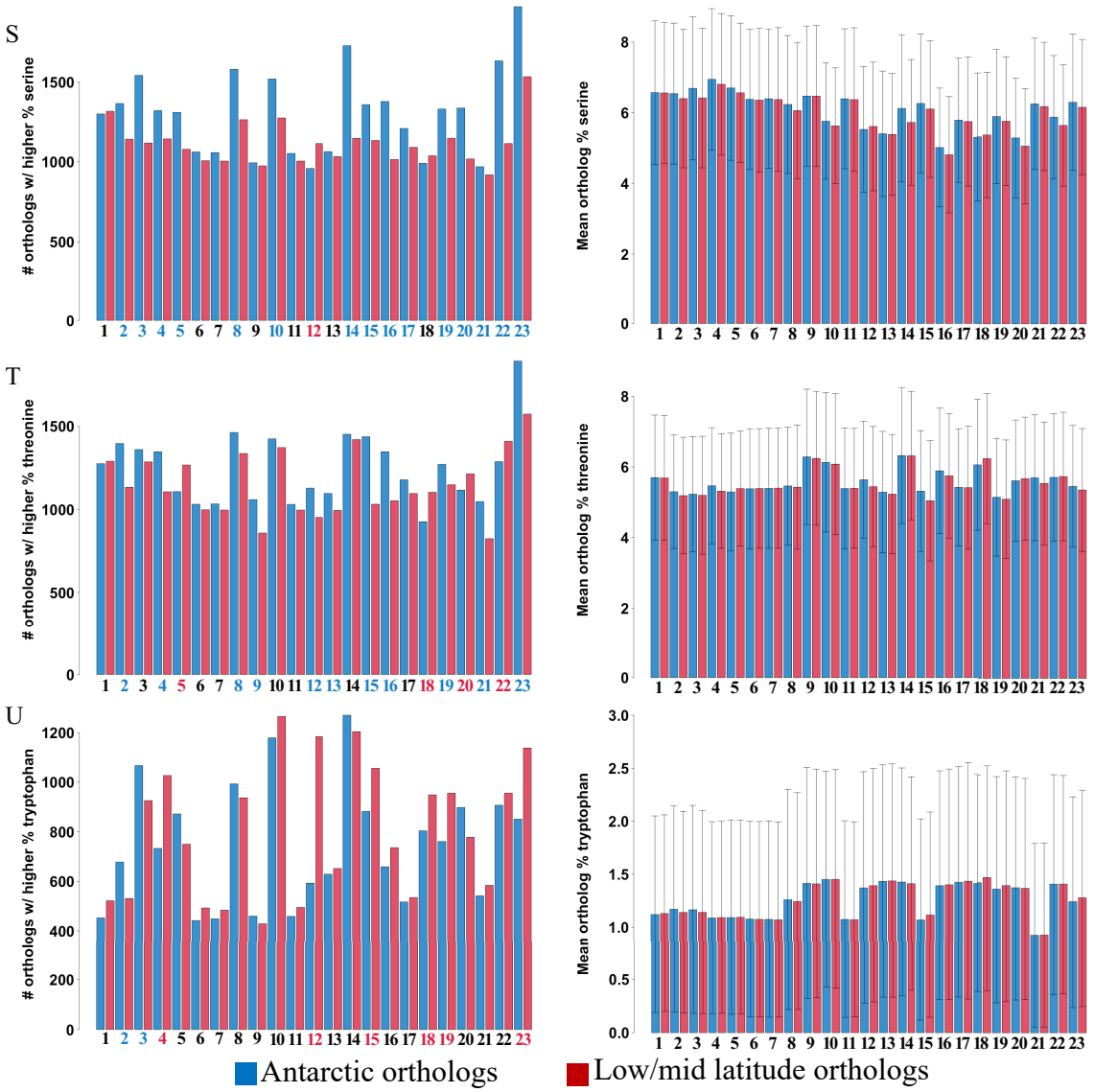

D
orthologs w/ higher % alanine
Mean ortholog % alanine
E
orthologs w/ higher % arginine
Mean ortholog % arginine
F
orthologs w/ higher % asparagine
Mean ortholog % asparagine
Antarctic orthologs
Low/mid latitude orthologs

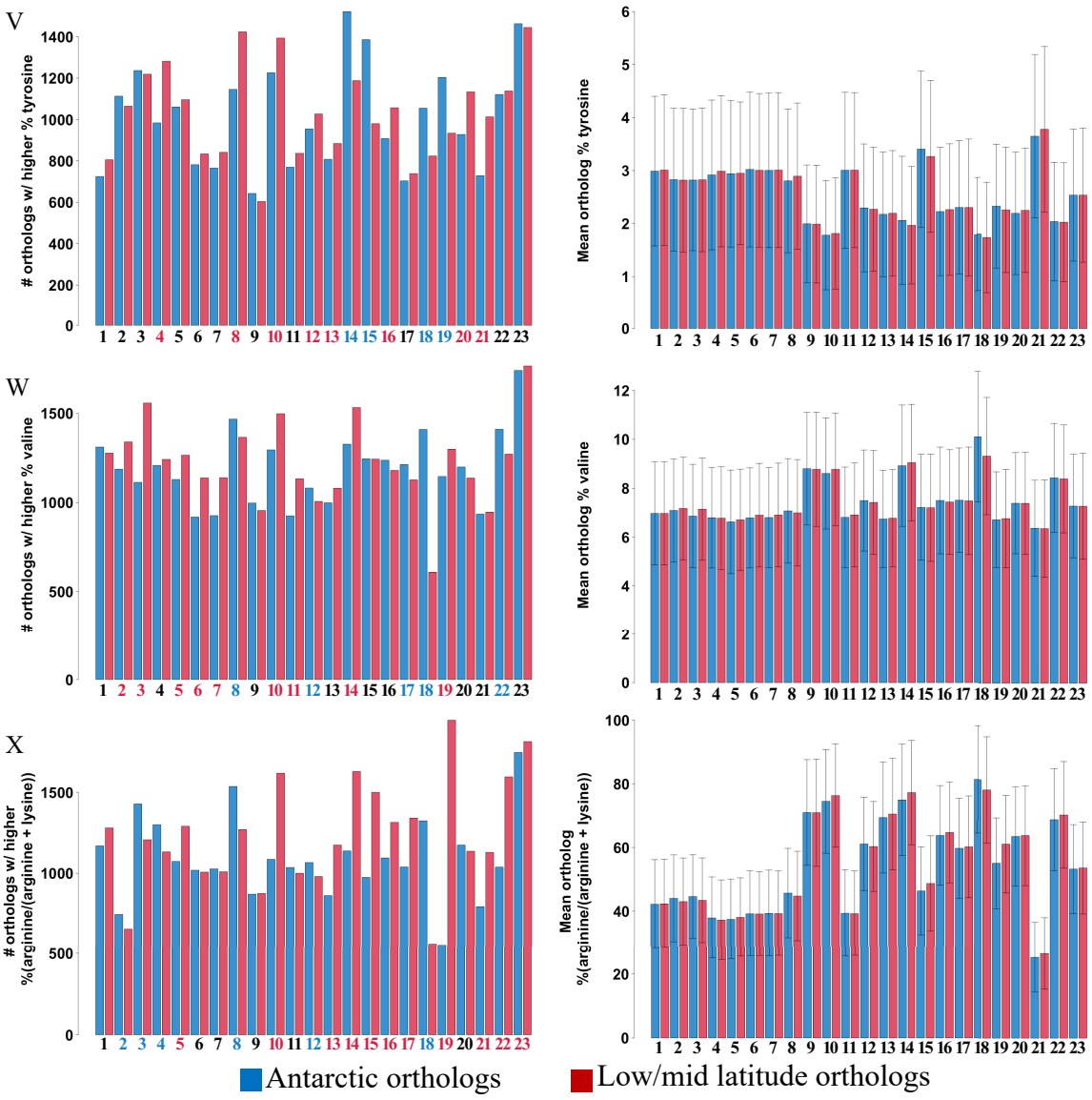

G
orthologs w/ higher % aspartic acid
Mean ortholog % aspartic acid
H
orthologs w/ higher % cysteine
Mean ortholog % cysteine
I
orthologs w/ higher % glutamine
Mean ortholog % glutamine
Antarctic orthologs
Low/mid latitude orthologs


J
orthologs w/ higher % glutamic acid
Mean ortholog % glutamic acid
K
orthologs w/ higher % glycine
Mean ortholog % glycine
L
orthologs w/ higher % histidine
Mean ortholog % histidine
Antarctic orthologs
Low/mid latitude orthologs


M
orthologs w/ higher % isoleucine
Mean ortholog % isoleucine
N
orthologs w/ higher % leucine
Mean ortholog % leucine
O
orthologs w/ higher % lysine
Mean ortholog % lysine
Antarctic orthologs
Low/mid latitude orthologs


P
orthologs w/ higher % methionine
Mean ortholog % methionine
Q
orthologs w/ higher % phenylalanine
Mean ortholog % phenylalanine
R
orthologs w/ higher % proline
Mean ortholog % proline
Antarctic orthologs
Low/mid latitude orthologs


S
orthologs w/ higher % serine
Mean ortholog % serine
T
orthologs w/ higher % threonine
Mean ortholog % threonine
U
orthologs w/ higher % tryptophan
Mean ortholog % tryptophan
Antarctic orthologs
Low/mid latitude orthologs


V
orthologs w/ higher % tyrosine
Mean ortholog % tyrosine
W
orthologs w/ higher % valine
Mean ortholog % valine
X
orthologs w/ higher %(arginine/(arginine + lysine))
Mean ortholog %(arginine/(arginine + lysine))
Antarctic orthologs
Low/mid latitude orthologs

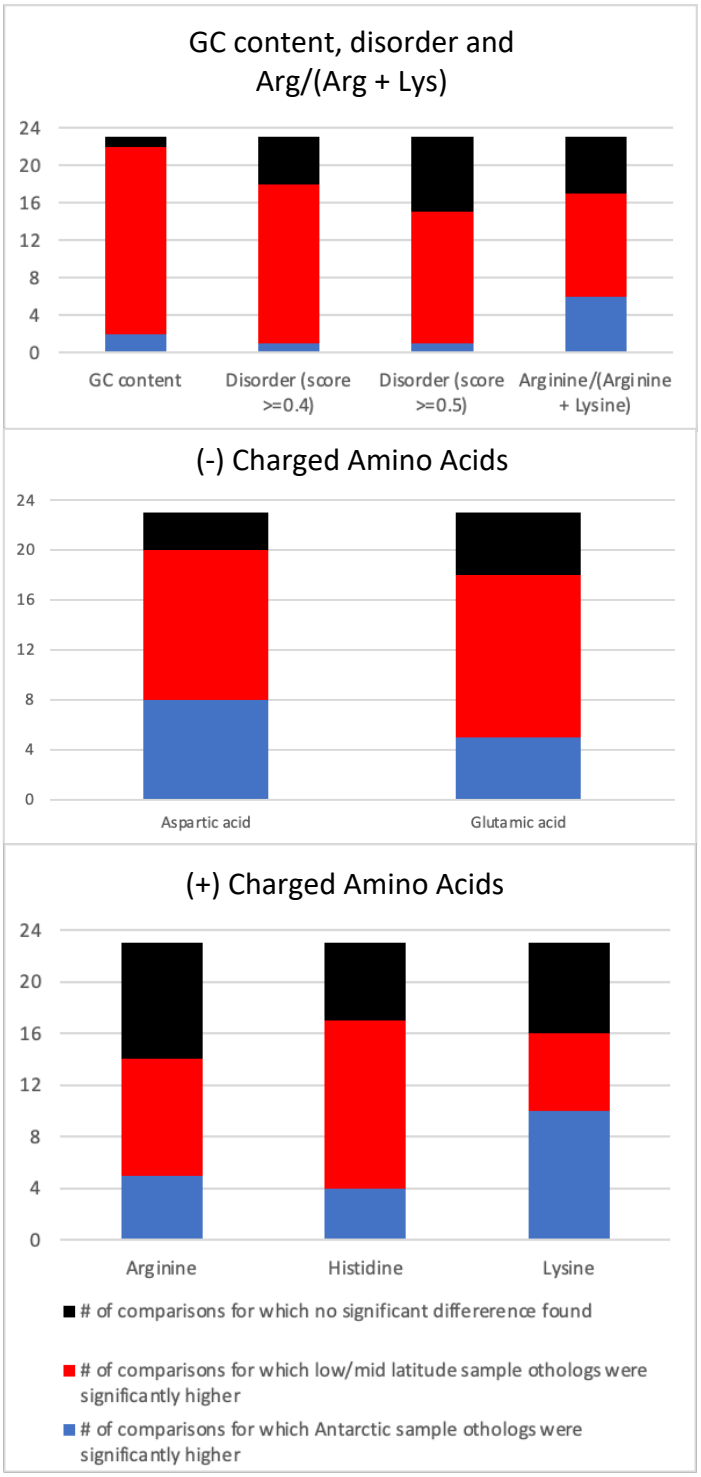


Figure 10. Summary of GC content, intrinsic disorder, Arg/(Arg + Lys) ratio, and charged amino acid content comparison results by comparison #. Blue and red indicate those comparisons for which more orthologs had higher content, in Antarctic or low/mid latitude, respectively. Black indicates no significant difference was found.

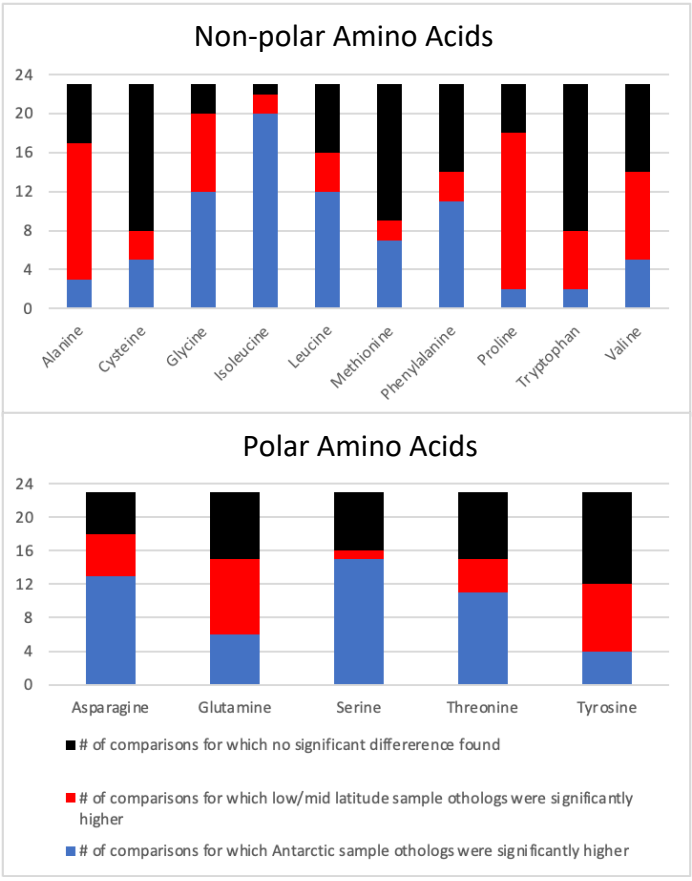


Figure 11. Summary of non-polar and polar amino acid content comparison results by comparison #. Blue and red indicate those comparisons for which more orthologs had higher content, in Antarctic or low/mid latitude, respectively. Black indicates no significant difference was found.

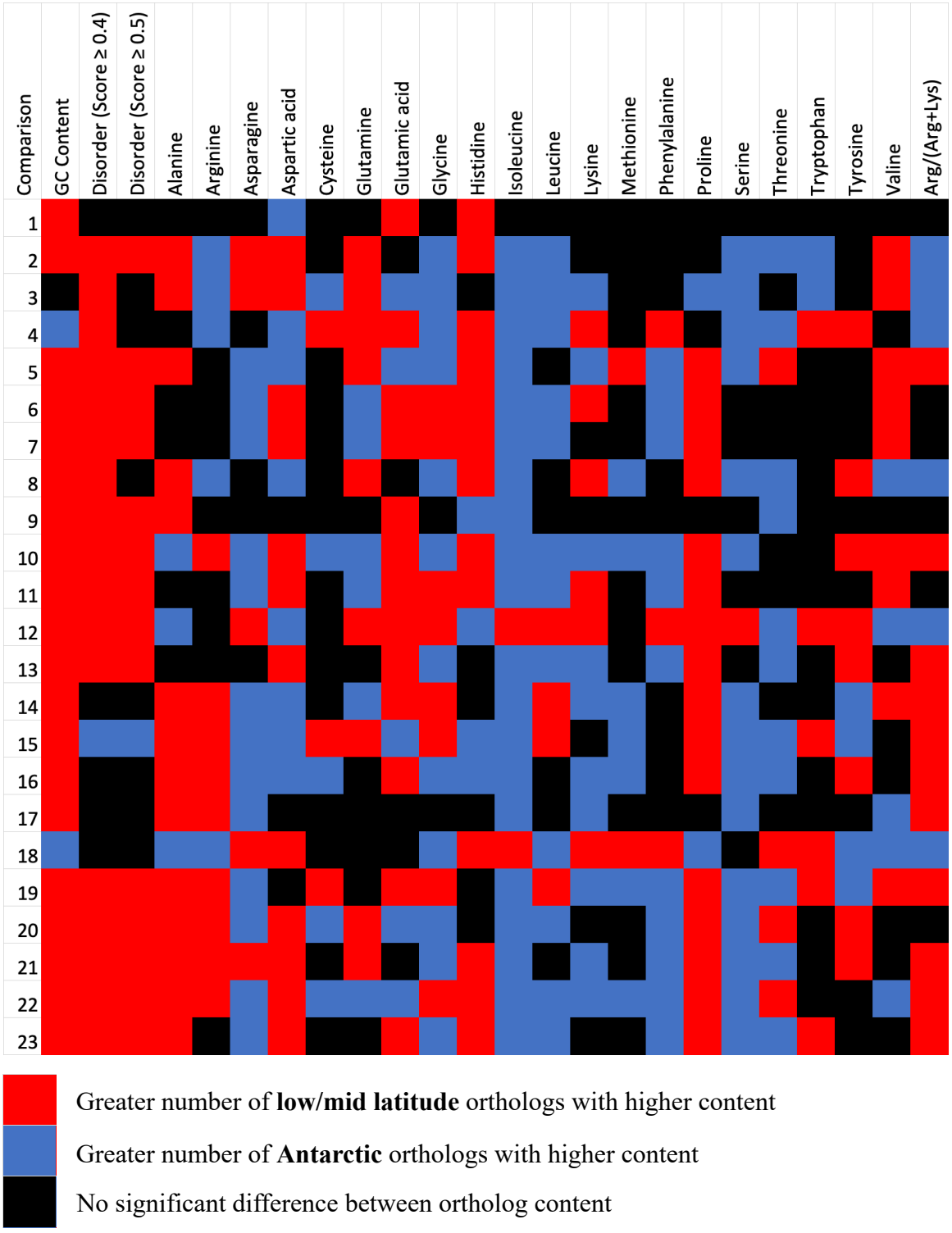


Figure 12. Heat map of comparison results by parameter tested. Blue and red indicate those comparisons for which more orthologs had higher content, in Antarctic or low/mid latitude, respectively. Black indicates no significant difference was found.

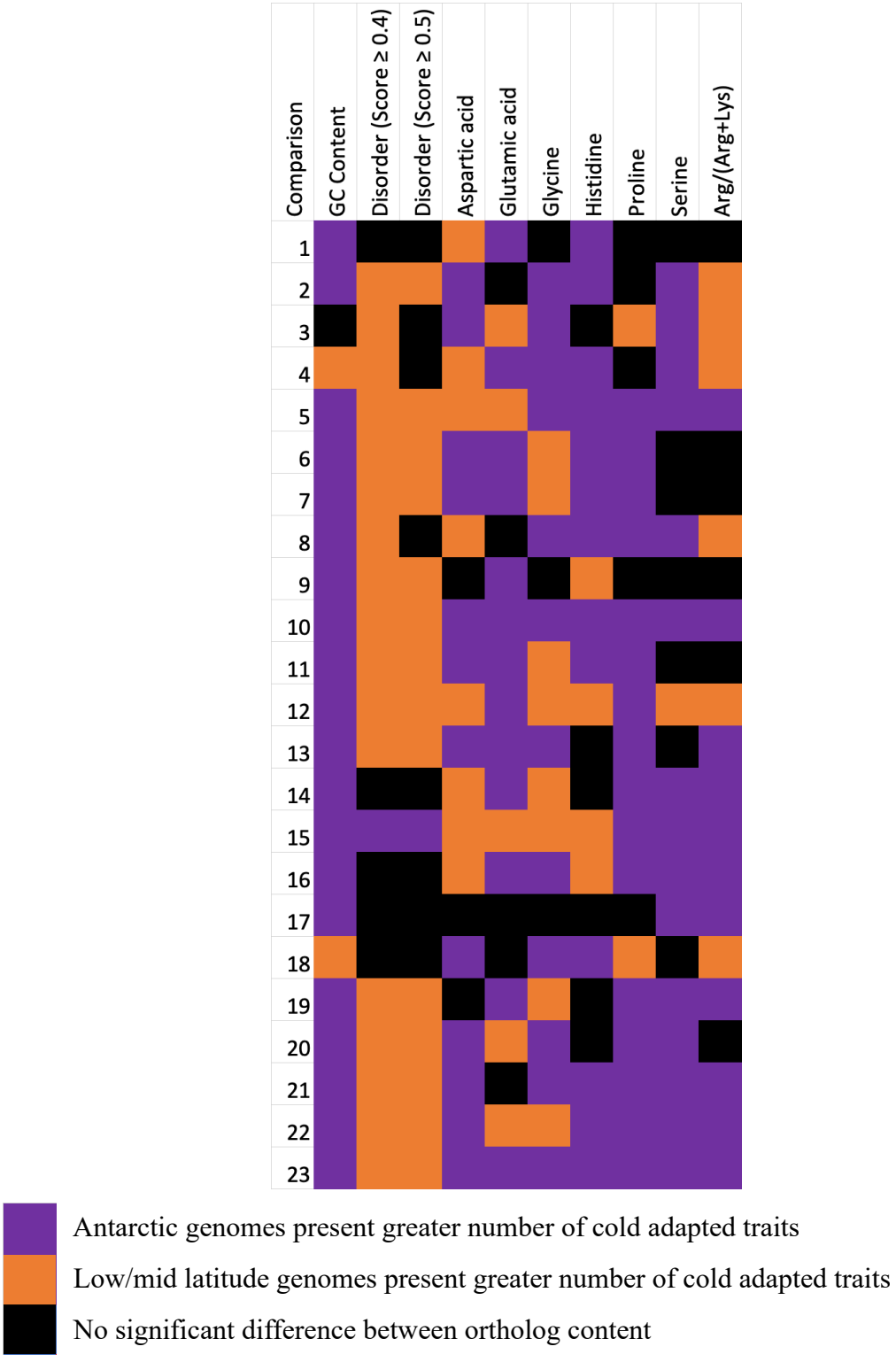


Figure 13. Heat map of comparisons and cold adapted trait. Purple and orange indicate those comparisons for which genomes had more of the cold adapted trend, in Antarctic or low/mid latitude, respectively. Black indicates no significant difference.

DISCUSSION

Phylogenetic diversity and novelty of isolates and their genomes as representatives of the *Synoicum adareanum* microbiome

For our cultivation effort, five carbon sources were selected based on transport or metabolite degradation pathways observed in the *Candidatus* Synoicohabitans palerolidicus metagenome (Murray et al. 2021). This metagenome assembled genome (MAG) encoded five copies of the palmerolide BGC (*pal* BGC). Using a retrobiosynthetic strategy one of the copies was predicted to be responsible for the biosynthesis of palmerolide A (Avalon et al., 2021). This uncultivated species is classified in the Opitutaceae family under the phylum Verrucomicrobia, prompting our use of a minimal salts medium protocol by Stevenson et al. who targeted uncultivated Acidobacteria and Verrucomicrobia strains several of which were Opitutaceae-related (Stevenson et al., 2004).

Previous microbial cultivation using *Synoicum adareanum* tissue utilized complex media which promoted robust growth at the cost of isolating slow growing organisms (Murray et al. 2020, and pers. com.). To maximize diverse growth from the microbiome, duplicate media preparations included these individual carbon sources alone and then a set supplemented with yeast extract, and furthermore, replicate plates of all media types allowed use of both atmospheric and microaerophilic growth conditions. Other than one colony growing on chitin+yeast under atmospheric conditions, no other growth was observed on chitin media. The greater growth under the initial microaerophilic conditions suggests that most bacteria cultivated are either microaerophiles or where under oxygen stress at the outset.

Previously, Riesenfeld et al. (2008) characterized the *Synoicum adareanum* microbial community of a single specimen using 16S rRNA gene sequences and identified five phyla: Actinobacteria, Bacteroidetes (class Flavobacteria), Proteobacteria (classes Alpha-, Beta-, and Gamma- proteobacteria), Verrucomicrobia, and candidate phylum TM7. Prior analyses of the proportions between phyla found within the microbiome community were completed by Murray et al. (2020) through surveying 63 *Synoicum adareanum* samples. Using 16S rRNA gene amplicon sequence variants, fractions of phyla were placed into categories Core80 (bacteria present in >80% of samples studied), Dynamic50 (in 50%-79% of samples) and Variable (in <50% of samples). Gammaproteobacteria were found in the highest proportions of all catergories, followed by Alphaproteobacteria in both Core80 and Variable, and Bateroidetes in Dynamic50. Community wide composition followed this trend of Gammaproteobacteria > Alphaproteobacteria > Bacteroidetes proportions. These three phyla/classes comprised 97.72% of the total community. Of the microbiome members previously brought into culture (2007 culture collection), all but one (*Pseudovibrio* sp. TunPHSC 04-5.I4, an Alphaproteobacteria) have been of the class Gammaproteobacteria. This strain is also the only cultivated member core microbiome, thus far.

16S rRNA sequence analyses of 46 isolates demonstrate the merits of minimal media for systematic isolation of slow-growing, host-associated bacteria; the new culture collection broadened the diversity to now include Bacilli, Acidimicrobiia, Actinobacteria and Flavobacteriia. The phylum Bacterioidetes, which includes the class Flavobacteriia, was identified as the second-highest community-wide phylum by Murray et al. (2020), but at 2.830 ($\pm$ 2.140), was still far lower than the proportion of Proteobacteria

(Gammaproteobacteria, 71.990 ± 6.640; Alphaproteobacteria, 22.90 ± 5.390). Similarly, Acidimicrobiia and Actinobacteria are classified under phylum Actinobacteria, identified in the microbiome analysis by Murray et al. (2020) at even lower proportions (Actinobacteria, 0.100 ± 0.080). The phylum Firmicutes, containing class Bacilli, was not identified in the previous analyses by Murray et al. (2020), although the depth of their study at 10,000 amplicons per sample could miss more rare taxa, especially those with single copy rRNA operons. The novelty and distribution of these new and 2007 cultivars from the *Synoicum adareanum* microbiome guided the selection of twenty-two samples for whole genome sequencing that represented a range of taxa across the phylogenetic tree.

The two organisms binned from sample CEL 36, an *Illumatobacter* and *Pseudoalteromonas*, may both be constituents of the *Synoicum adareanum* microbiome, potentially even mutually essential for the other's growth. The nature of using 2nd generation cultures for our initial pal A screening and dereplication could have also contributed to an impure sample. For these reasons, both CEL 36 Bin 1 and CEL 36 Bin 2 assemblies were used in further genomic analyses.

Assessing the novelty of the Antarctic genomes using the MiGA NCBI Prok tool suggested that there were potentially 13 new species, 3 new genera, and 1 new family. The MiGA tool indicated that the CEL 36 Bin 1, which only had an 42.33% AAI match with Enterobacter hormaechei NZ CP041054 was a new family, however the manual low/mid latitude comparison effort we conducted identified an *Illumatobacter* genome with an AAI of 66.41% to CEL 36 Bin 1. This suggests that CEL 36 Bin 1 is likely a new species within the genus *Illumatobacter*.

Biosynthetic potential of Antarctic isolate genomes

The highest number of biosynthetic gene clusters identified by antiSMASH were found in the DEX 30 isolate (genus *Gordonia*), which is found within the phylum Actinobacteria. Actinobacteria are known producers of interesting metabolites with numerous applications and have been predicted to synthesize roughly 10,000 of ~23,000 known microbially produced antibiotics globally (Manivasagan et al., 2014). The top 5 isolates with the highest number of biosynthetic gene clusters identified fell within the classes, starting with the highest, Actinobacteria (DEX 30), then Alphaproteobacteria (TAU 8), and a tie for third highest between Gammaproteobacteria (BOMB 3.2.15), Acidimicrobiia (CEL 36 Bin 1) and Actinobacteria (XYL 43). In a review of natural products from cold, marine environments, Soldatou and Baker (2017) describe 253 secondary metabolites discovered through 2005, of which 22% were of microbial origin (Soldatou & Baker, 2017). They also describe 352 additional natural products later published between 2006 and 2016, of which 71% are microbially produced. Yet, from this increase in microbially produced cold-water natural products, Soldatou and Baker (2017) point out that numbers of isolated metabolites from psychrophilic bacteria are low, relative to fungi.

The relatively high number of 135 biosynthetic gene clusters found among our isolates, and the relatively low number of 49 known clusters found similar to those found within our isolates (of which only 8 had similarity of 80% or higher), indicate a high degree of novelty across the secondary metabolites encoded in these genomes. More detailed comparisons between sequences of biosynthetic gene clusters identified and their most similar known clusters may allow for more specific quantification of their novelty.

The categories of compound synthesis (RiPPs, PKS, Terpenes, NRPS, etc.) detected in the Antarctic genomes are of significant interest due to numerous applications from their antibacterial, antifungal, and anticancer activity, as well as utility in industrial applications (Manivasagan et al., 2014; Moghaddam et al., 2021; Nunez-Pons et al., 2020). A low proportion of all polar microorganisms have been investigated for synthesis of natural products (Tian et al., 2017), and our results demonstrate the benefits of a screening genomes from a diverse Antarctic bacterial culture collection for identification of novel biosynthesis. Functional demonstration and bioactivity assays are now needed to better understand the biological potential of these newly identified BGCs. The diversity in compound synthesis present in all these isolates presents opportunity for more detailed enzyme and chemical investigations into further explorations of Antarctic microbial natural product potential. Future chemical analyses directly of these cultivars may be a rapid means to determine if the biosynthetic gene clusters encoding these metabolites, identified *in silico*, are actively expressed while in isolation.

Signatures of cold adaptation

Selecting low-latitude, related genomes for cold-adaptation comparative genomic analyses presented challenges due to limited options for some genera and a lack of readily available metadata for assemblies. The IMG database provided helpful parameter options including high quality, location, habitat, latitude/longitude, isolation, depth in meters and subsurface in meters, which helped narrow selections to meet the criteria described. Marine bacteria were selected between +/- 45º latitude and prioritizing those at <200-meter sample depth to reduce chances of selecting a psychrophilic organism.

Likewise, considering genomes of samples from seawater or coastal sediments/sand increased chances that all organisms in our comparative study would be adapted to marine environments. In the case of DEX 32, the top match was *Paenibacillus xylanilyticus* with a 65.95% AAI but is not a marine or coastal organism. Of marine and coastal *Paenibacillus* genomes, *Paenibacillus* sp. CAA11 matched most closely at 61.05% AAI, the only low-latitude genome selected which did not meet the 65% AAI typically considered a threshold for genus level matches.

DEX 30 matched *Gordonia terrae* at 68.8% AAI and one strain, AIST1, has been isolated from the Inland Sea of Japan's surface seawater. For this strain no whole genome sequence data is available, so a genome assembly from a soil derived *Gordonia terrae* was used in the analysis.

The ortholog comparisons between the Antarctic genome and low/mid latitude genome of the same genus follow previously documented trends for cold-adapted genomes to contain lower arginine/(arginine + lysine) or arginine/lysine ratios, which are thought to decrease salt bridges through fewer charged residues in the protein and therefore enhance flexibility (Figure 14) (Ayala-del-Rio et al., 2010; Goordial et al., 2016; Grzymski et al., 2006). Similar increases in protein flexibility in cold-adapted organisms are understood to come from a decrease in proline (its cyclic structure generates rigid kinks in peptides) and acidic and aliphatic residues (which increase salt bridges and hydrogen bonding), but an increase in glycine (for its small size) and serine residues (potentially decreasing overall charge) (Ayala-del-Rio et al., 2010; Goordial et al., 2016; Grzymski et al., 2006; Raymond-Bouchard et al., 2018). Our results follow

these general trends but are noticeably higher for the non-polar amino acids isoleucine and leucine; a surprising increase considering their aliphaticity.

Intrinsic disorder among amino acids comprising a protein sequence is another parameter by which cold-adaption is measured, with an increase in disorder potentially

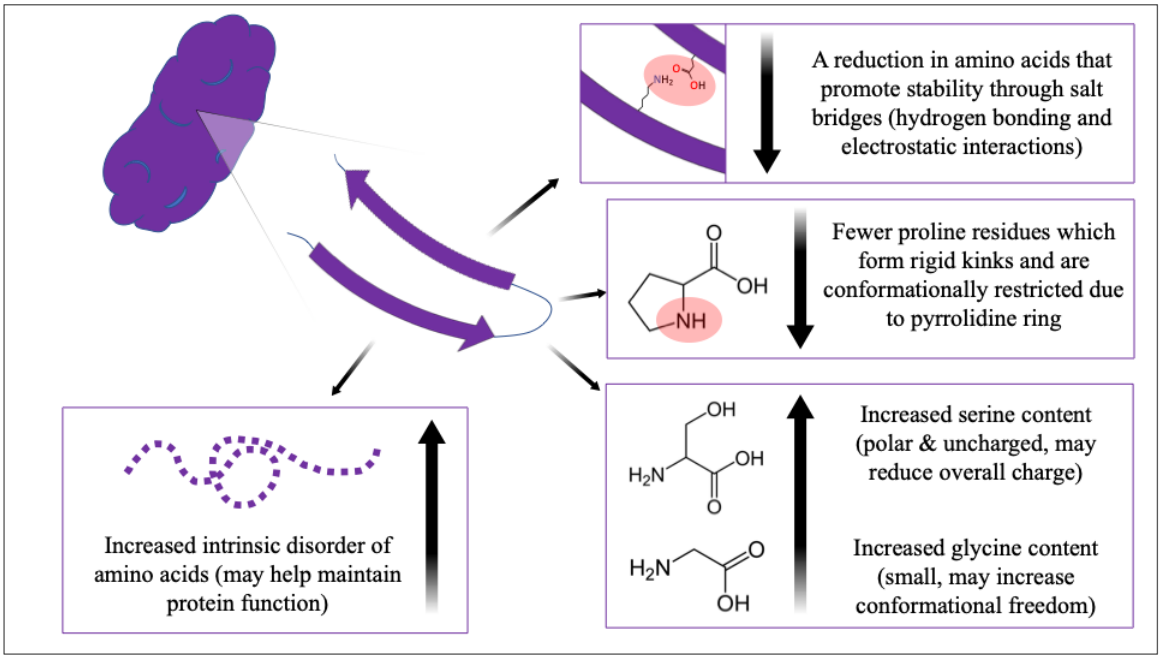


Figure 14. Illustration depicting example cold adaptation trends in protein amino acid content. Residues that form salt bridges or otherwise reduce protein dynamics are decreased, while those that enhance flexibility are increased. This is believed to aid in maintaining protein function at low temperatures.

maintaining protein function through increased flexibility at lower temperatures (Feller, 2007). Surprisingly, our comparisons show the low/mid latitude predicted protein orthologs as far more disordered than the Antarctic orthologs overall (only comparison #15 showed greater disorder for the Antarctic genome). Proteins are subject to misfolding and denaturation at not only high but also low temperatures (Feller, 2007; Raymond-Bouchard et al., 2018), and because our organisms are often found in < 0°C seawater, these results may suggest a disorder threshold which balances protein functionality

through flexibility and cold denaturation resistance through stability. In a study by Grzymski et al. (2006), genome fragments of Antarctic bacterioplankton were compared against mesophilic homologs. While the length of maximum disorder of the predicted proteins was longer in 5 of 6 Antarctic samples, the number of disordered amino acids was only higher in 1 of 6 Antarctic samples relative to the mesophile homologs (Grzymski et al., 2006). These bacterioplankton primarily experience stenothermal < 0℃ seawater temperatures, so the results may also be explained by a potential threshold for decreased disorder of residues in protein-denaturing temperatures.

Considering the trends displayed in the manually constructed heat map (Figure 13), comparisons 23 and 10 showed the greatest cold-adaptation for both Antarctic genomes in these analyses. Only the two disorder parameters showed higher cold-adaptation in the low/mid latitude genomes, which may be explained as described above. The Antarctic genome in comparison 23 was the previously isolated and sequenced bacteria, *Pseudovibrio* sp. TunPHSC04-5.I4, characterized as a psychrophile that does not grow above 16 °C. Plans for future temperature experimentation to confirm psychrophily among our isolates will offer a unique comparison with the results synthesized here.

While we prioritized low/mid latitude genomes that were also from marine samples collected at ≤ 200 m depth, this was not always possible given genome options within a genus and other sample criteria. Comparison #1, between Antarctic sample BOMB 3.2.15 and low/mid latitude sample *Moritella* sp. F1, only shows significant differences in 4 of 24 parameters tested (*Moritella* sp. F1 with higher GC content, glutamic acid and histidine, but lower aspartic acid) (Figure 9). The 500 m sample depth

for *Moritella* sp. F1 may explain the lack of differences as this deep ocean organism may be similarly cold adapted as BOMB 3.2.15.

Interestingly, the three comparisons which had the highest number of parameters with no differences between the Antarctic and low/mid latitude genomes, # 1, 9 and 17, and had the highest % AAI shared between the two genomes. (92.71%, 89.98% and 89.94%, respectively) (Figure 12 and table 6). These may indicate that using high AAI between genomes for amino acid comparisons will result in a reduction of observed differences as the amino acid composition is intrinsically similar between two closely related genomes. Further analysis across a range of related taxa may help to sort this observation out.

In contrast to comparison #1 above, comparison #12 used the low/latitude genome *Marinosulfonomonas* sp. NORP110 which was sampled from > 4,500 m depth (subseafloor) and compared with CEL 38, yet was found to contain orthologs with significantly different parameter contents, and typically followed the predicted trends for amino acid type, GC content, etc. One interesting measurement, serine content, is only higher for orthologs from the low/mid latitude genome in the *Marinosulfonomonas* sp. NORP110 (Figure 9S). According to Raymond-Bouchard et al. (2018), increased serine content is a common phenomenon in cold-adapted organisms, so potentially, the deep ocean *Marinosulfonomonas* sp. NORP110 genome contains certain signatures of cold adaptation as well.

Table 7. Genome GC content for organisms used in cold adaptation analyses, listed by comparison number. See Table 3 for details on genomes used.

Comparison no.	Antarctic genome GC content (%)	Low/mid latitude genome GC content (%)
1	41.58	46.8
2	44.03	44.54
3	44.05	44.54
4	39.47	37.85
5	38.31	39.09
6	40.40	41.02
7	40.55	41.02
8	43.28	43.28
9	58.97	61.87
10	62.20	67.29
11	41.20	41.02
12	53.22	53.38
13	58.24	58.55
14	65.53	57.98
15	46.18	48.68
16	53.33	56.45
17	55.83	56.76
18	72.32	69.69
19	52.45	57.32
20	54.01	57.38
21	28.74	30.59
22	64.24	66.28
23	50.48	50.34

GC content presents another interesting factor in cold adaptation analyses. In our case, high GC content is observed predominantly in low/mid latitude orthologs, but genome-wide GC content is also generally higher, albeit to a lesser degree overall, and typically correlates with directionality in the ortholog comparisons. Three interesting exceptions are observed in comparisons # 11, 14, and 23. In each of these comparisons the Antarctic genomes (CEL 36 Bin 2, DEX 30, and *Pseudovibrio* sp. TunPHSC04-5.I4)

all had higher genome-wide GC content than their low/mid latitude neighbors, yet open reading frame orthologs with higher GC content were observed for the low/mid latitude genomes (Figure 9A and Table 7). This suggests that, in some cases, open reading frames may disproportionately contain lower GC content in cold adapted genomes. One hypothesis for this effect is that selective pressures maintain transcriptional efficiency for coding regions of the genome. Further analysis of coding vs. non-coding sequence GC content within and between known psychrophile and mesophile genomes may uncover if definitive GC variation exists due to cold-adapted genomes.

Conclusion and future directions

Through this work, the *Synoicum adareanum*-derived culture collection has been expanded with respect to diversity, encompassing new classes and phyla using a defined media cultivation effort that spanned 11-months for initial growth. These new cultivars likely include novel genera and species as observed in their average nucleotide identity and average amino acid identity between closely related genomes, showing the merits of using defined media to allow for slow-growing organisms to be isolated. The diversity of secondary metabolite categories identified through genome mining for biosynthetic gene clusters, and the relatively few known similar clusters, supports the metagenomic conclusion that novel biosynthetic pathways exist within the *Synoicum adareanum* microbiome members. Isolates subsequently grown in marine broth are currently being analyzed for production of the secondary metabolites through mass spectrometry of both cell pellets and supernatant media. This is the next step towards understanding if the biosynthetic gene clusters are being actively expressed in culture.

Cold adaptation experimentation identified parameters in amino acid usage that corresponded with known residue content trends in proteins from cold adapted organisms. Disorder, which is typically understood to increase in cold adapted protein sequences, was observed higher in more orthologs from the low/mid latitude genomes, potentially indicative of a threshold for disorder to maintain stability in cold-denaturing (< 0 ℃) temperatures. Future analyses between known psychrophiles (specifically, those physiologically tested below 0 ℃) and known mesophiles will provide more insight into this hypothesis.

Due to the slow nature of culture growth at 10 ℃ and with one goal to rapidly screen for palmerolide A biosynthetic genes, we initially assayed 2nd generation cultures grown only on defined media; the DNA extrants from which were used for 16S rRNA gene sequencing. Later, we revisited the 2nd generation isolate glycerol stocks to prepare 3X purified isolates; these were cultivated on complex media (marine broth). The 3X purified isolates will undergo psychrophily tests through a temperature-dependent cultivation experiment spanning 20 ℃, the operational definition of psychrophily, to determine if these isolates are psychophiles or psychrotrophs. Revisiting our disorder analysis with these psychophily results will offer a clearer view of trends, if they exist, at the limits of protein denaturation.

GC content was typically lower for the Antarctic genome coding regions even when genome-wide GC content was higher. One hypothesis for this observation is that lower GC content may exist specifically within coding sequences to increase the overall transcriptional efficiency in cold adapted organisms. When testing this hypothesis, it may be beneficial to investigate ratios of GC content in coding vs. non-coding sequence (i.e.

no difference equalling 1), then compare these ratios among sets of known psychrophilic genomes and known mesophilic genomes.

Understanding the relationships between natural product capability, expression and production while in isolation, as well as cold adaptation in genome and protein sequences may benefit future sustainable secondary metabolite production. This study has not only expanded a marine invertebrate hosted culture collection, but has identified likely novel secondary metabolites, and offers continued work on biosynthetic gene cluster expression experimentation and direct chemical analysis. This new culture collection also allows for discrete temperature experimentation to provide greater insight to the novel disorder and GC content hypotheses presented in our comparitive genomic analyses. We can then integrate this knowledge with results from compound screening for a more comprehensive understanding of interesting microbial natural product biosynthesis within this polar, host-associated marine system.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, *215*(3), 403-410. https://doi.org/10.1016/s0022-2836(05)80360-2

Avalon, N., Murray, A., Daligault, H., CLo, C.-C., Davenport, K., Dichosa, A., Chain, P., & Baker, B. (2021). Palmerolide PKS-NRPS gene clusters characterized from the microbiome of an Antarctic ascidian. *Proc. Natl. Acad. Sci. Submitted.* https://doi.org/https://doi.org/10.1101/2021.04.05.438531

Ayala-del-Rio, H. L., Chain, P. S., Grzymski, J. J., Ponder, M. A., Ivanova, N., Bergholz, P. W., Di Bartolo, G., Hauser, L., Land, M., Bakermans, C., Rodrigues, D., Klappenbach, J., Zarka, D., Larimer, F., Richardson, P., Murray, A., Thomashow, M., & Tiedje, J. M. (2010). The Genome Sequence of *Psychrobacter arcticus* 273-4, a Psychroactive Siberian Permafrost Bacterium, Reveals Mechanisms for Adaptation to Low-Temperature Growth. *Applied and Environmental Microbiology*, *76*(7), 2304-2312. https://doi.org/10.1128/aem.02101-09

Ayuningrum, D., Liu, Y., Riyanti, Sibero, M. T., Kristiana, R., Asagabaldan, M. A., Wuisan, Z. G., Trianto, A., Radjasa, O. K., Sabdono, A., & Schaberle, T. F. (2019). Tunicate-associated bacteria show a great potential for the discovery of antimicrobial compounds. *Plos One*, *14*(3), 14, Article e0213797. https://doi.org/10.1371/journal.pone.0213797

Bae, S. S., Jung, Y. H., & Baek, K. (2020). *Shewanella maritima* sp. nov., a facultative anaerobic marine bacterium isolated from seawater, and emended description of Shewanella intestini. *International Journal of Systematic and Evolutionary Microbiology*, *70*(2), 1288-1293. https://doi.org/10.1099/ijsem.0.003916

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455-477. https://doi.org/10.1089/cmb.2012.0021

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1), W81-W87. https://doi.org/10.1093/nar/gkz310

Bodor, A., Bounedjoum, N., Vincze, G. E., Kis, A. E., Laczi, K., Bende, G., Szilagyi, A., Kovacs, T., Perei, K., & Rakhely, G. (2020). Challenges of unculturable bacteria: environmental perspectives. *Reviews in Environmental Science and Bio-Technology*, *19*(1), 1-22. https://doi.org/10.1007/s11157-020-09522-4

Busse, H. J., & Schumann, P. (2019). Reclassification of *Arthrobacter enclensis* as *Pseudarthrobacter enclensis* comb. nov., and emended descriptions of the genus *Pseudarthrobacter*, and the species *Pseudarthrobacter phenanthrenivorans* and *Pseudarthrobacter scleromae*. *International Journal of Systematic and Evolutionary Microbiology*, *69*(11), 3508-3511. https://doi.org/10.1099/ijsem.0.003652

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925-1927. https://doi.org/10.1093/bioinformatics/btz848

Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R., Roux, S., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Research*, *49*(D1), D751-D763. https://doi.org/10.1093/nar/gkaa939

Chen, L., Hu, J.-S., Xu, J.-L., Shao, C.-L., & Wang, G.-Y. (2018). Biological and Chemical Diversity of Ascidian-Associated Microorganisms. *Marine Drugs*, *16*(10), Article 362. https://doi.org/10.3390/md16100362

Cole, J. R., Wang, Q., Fish, J. A., Chai, B. L., McGarrell, D. M., Sun, Y. N., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), D633-D642. https://doi.org/10.1093/nar/gkt1244

Dastager, S. G., Qin, L., Tang, S.-K., Krishnamurthi, S., Lee, J.-C., & Li, W.-J. (2014). *Arthrobacter enclensis* sp. nov., isolated from sediment sample. *Archives of Microbiology*, *196*(11), 775-782. https://doi.org/10.1007/s00203-014-1016-9

Diyabalanage, T., Amsler, C. D., McClintock, J. B., & Baker, B. J. (2006). Palmerolide A, a cytotoxic macrolide from the Antarctic tunicate *Synoicum adareanum*. *Journal of the American Chemical Society*, *128*(17), 5630-5631. https://doi.org/10.1021/ja0588508

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

Feller, G. (2007). Life at low temperatures: is disorder the driving force? *Extremophiles*, *11*(2), 211-216. https://doi.org/10.1007/s00792-006-0050-1

Feng, G.-D., Zhang, X.-J., Yang, S.-Z., Li, A.-Z., Yao, Q., & Zhu, H. (2020). Transfer of *Sphingorhabdus marina*, *Sphingorhabdus litoris*, *Sphingorhabdus flavimaris* and *Sphingorhabdus pacifica* corrig. into the novel genus *Parasphingorhabdus* gen. nov. and *Sphingopyxis baekryungensis* into the novel genus *Novosphingopyxis* gen. nov. within the family Sphingomonadaceae. *International Journal of Systematic and Evolutionary Microbiology*, *70*(3), 2147-2154. https://doi.org/10.1099/ijsem.0.004033

Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., & Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, *261*, 169-176. https://doi.org/10.1016/j.jbiotec.2017.06.1198

Gonzalez, J. M., Fernandez-Gomez, B., Fernandez-Guerra, A., Gomez-Consarnau, L., Sanchez, O., Coll-Llado, M., del Campo, J., Escudero, L., Rodriguez-Martinez, R., Alonso-Saez, L., Latasa, M., Paulsen, I., Nedashkovskaya, O., Lekunberri, I., Pinhassi, J., & Pedros-Alio, C. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8724-8729. https://doi.org/10.1073/pnas.0712027105

Goordial, J., Raymond-Bouchard, I., Zolotarov, Y., de Bethencourt, L., Ronholm, J., Shapiro, N., Woyke, T., Stromvik, M., Greer, C. W., Bakermans, C., & Whyte, L. (2016). Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile *Rhodococcus* sp JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. *FEMS Microbiology Ecology*, *92*(2), Article fiv154. https://doi.org/10.1093/femsec/fiv154

Grzymski, J. J., Carter, B. J., DeLong, E. F., Feldman, R. A., Ghadiri, A., & Murray, A. E. (2006). Comparative genomics of DNA fragments from six antarctic marine planktonic bacteria. *Applied and Environmental Microbiology*, *72*(2), 1532-1541. https://doi.org/10.1128/aem.72.2.1532-1541.2006

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075. https://doi.org/10.1093/bioinformatics/btt086

Hamada, M., Shibata, C., Tamura, T., Yamamura, H., Hayakawa, M., & Suzuki, K. (2013). *Janibacter cremeus* sp nov., an actinobacterium isolated from sea sediment. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 3687-3690. https://doi.org/10.1099/ijs.0.051532-0

Han, S. K., Nedashkovskaya, O. I., Mikhailov, V. V., Kim, S. B., & Bae, K. S. (2003). *Salinibacterium amurskyense* gen. nov., sp nov., a novel genus of the family Microbacteriaceae from the marine environment. *International Journal of Systematic and Evolutionary Microbiology*, *53*, 2061-2066. https://doi.org/10.1099/ijs.0.02627-0

Ivanova, E. P., Zhukova, N. V., Lysenko, A. M., Gorshkova, N. M., Sergeev, A. F., Mikhailov, V. V., & Bowman, J. P. (2005). *Loktanella agnita* sp nov and *Loktanella rosea* sp nov., from the north-west Pacific Ocean. *International Journal of Systematic and Evolutionary Microbiology*, *55*, 2203-2207. https://doi.org/10.1099/ijs.0.63461-0

Jiang, X., Liu, B., Lebreton, S., & De Brabander, J. K. (2007). Total synthesis and structure revision of the marine metabolite palmerolide A. *Journal of the American Chemical Society*, *129*(20), 6386-+. https://doi.org/10.1021/ja0715142

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Munoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., van Santen, J. A., Tracanna, V., Duran, H. G. S., Andreu, V. P., Selem-Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian, L., Collemare, J., Linington, R. G., Weber, T., & Medema, M. H. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*, *48*(D1), D454-D458. https://doi.org/10.1093/nar/gkz882

Kim, B.-S., Lim, Y. W., & Chun, J. (2008). *Sphingopyxis marina* sp nov and *Sphingopyxis litioris* sp nov., isolated from seawater. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 2415-2419. https://doi.org/10.1099/ijs.0.65614-0

Kim, E. S., Kim, B. S., Kim, K. Y., Woo, H. M., Lee, S. M., & Um, Y. (2018). Aerobic and anaerobic cellulose utilization by *Paenibacillus* sp CAA11 and enhancement of its cellulolytic ability by expressing a heterologous endoglucanase. *Journal of Biotechnology*, *268*, 21-27. https://doi.org/10.1016/j.jbiotec.2018.01.007

Konstantinidis, K. T., Rossello-Mora, R., & Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME Journal*, *11*(11), 2399-2406. https://doi.org/10.1038/ISMEj.2017.113

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Kwon, Y. M., Chung, D., Cho, E., & Yang, Y. (2020). Complete genome sequence of *Parasphingopyxis* sp. CP4,an asparaginase-producing marine bacterium. *Korean Journal of Microbiology*, *56*(4), 426-429. https://doi.org/https://doi.org/10.7845/kjm.2020.0124

Li, P. E., Lo, C. C., Anderson, J. J., Davenport, K. W., Bishop-Lilly, K. A., Xu, Y., Ahmed, S., Feng, S. H., Mokashi, V. P., & Chain, P. S. G. (2017). Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Research*, *45*(1), 67-80. https://doi.org/10.1093/nar/gkw1027

Liu, A., Zhang, Y. J., Xue, Q. J., Wang, H., Yang, Y. Y., Du, F., Zhao, L. Y., Zhang, H. H., Li, Y. Q., & Li, X. Z. (2020). *Litorilituus lipolyticus* sp. nov., isolated from intertidal sand of the Yellow Sea in China, and emended description of *Colwellia asteriadis*. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, *113*(4), 449-458. https://doi.org/10.1007/s10482-019-01355-8

Manivasagan, P., Venkatesan, J., Sivakumar, K., & Kim, S. K. (2014). Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiological Research*, *169*(4), 262-278. https://doi.org/10.1016/j.micres.2013.07.014

Massana, R., Murray, A. E., Preston, C. M., & DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Applied and Environmental Microbiology*, *63*(1), 50-56. https://doi.org/10.1128/aem.63.1.50-56.1997

Matsumoto, A., Kasai, H., Matsuo, Y., Shizuri, Y., Ichikawa, N., Fujita, N., Omura, S., & Takahashi, Y. (2013*). Ilumatobacter nonamiense* sp nov and *Ilumatobacter coccineum* sp nov., isolated from seashore sand. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 3404-3408. https://doi.org/10.1099/ijs.0.047316-0

McCauley, E. P., Pina, I. C., Thompson, A. D., Bashir, K., Weinberg, M., Kurz, S. L., & Crews, P. (2020). Highlights of marine natural products having parallel scaffolds found from marine-derived bacteria, sponges, and tunicates. *Journal of Antibiotics*, *73*(8), 504-525. https://doi.org/10.1038/s41429-020-0330-5

Meszaros, B., Erdos, G., & Dosztanyi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, *46*(W1), W329-W337. https://doi.org/10.1093/nar/gky384

Moghaddam, J. A., Jautzus, T., Alanjary, M., & Beemelmanns, C. (2021). Recent highlights of biosynthetic studies on marine natural products. *Organic & Biomolecular Chemistry*, *19*(1), 123-140. https://doi.org/10.1039/d0ob01677b

Murray, A. E., Avalon, N. E., Bishop, L., Davenport, K. W., Delage, E., Dichosa, A. E. K., Eveillard, D., Higham, M. L., Kokkaliari, S., Lo, C.-C., Riesenfeld, C. S., Young, R. M., Chain, P. S. G., & Baker, B. J. (2020). Uncovering the Core Microbiome and Distribution of Palmerolide in *Synoicum adareanum* Across the Anvers Island Archipelago, Antarctica. *Marine Drugs*, *18*(6), Article 298. https://doi.org/10.3390/md18060298

Murray, A.E., Chain, P., Lo, C., Daligault, H., NE Avalon, K. D., AKE Dichosa, ML, & Higham, R. R., BJ Baker. (*In prep*). Metagenome of an Antarctic host-associated verrucomicrobium with the biosynthetic capability for anticancer palmerolides.

Murray, A. E., Hollibaugh, J. T., & Orrego, C. (1996). Phylogenetic compositions of bacterioplankton from two California estuaries compared by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Applied and Environmental Microbiology*, *62*(7), 2676-2680. https://doi.org/10.1128/aem.62.7.2676-2680.1996

Murray, A. E., Lo, C.-C., Daligault, H. E., Avalon, N. E., Read, R. W., Davenport, K. W., Higham, M. L., Kunde, Y., Dichosa, A. E. K., Baker, B. J., & Chain, P. S. G. (2021). Discovery of an Antarctic ascidian-associated uncultivated Verrucomicrobia that encodes antimelanoma palmerolide biosynthetic capacity. *Proc. Natl. Acad. Sci. U.S.A.* , *Submitted.*

Nunez-Pons, L., Shilling, A., Verde, C., Baker, B. J., & Giordano, D. (2020). Marine Terpenoids from Polar Latitudes and Their Potential Applications in Biotechnology. *Marine Drugs*, *18*(8), 45, Article 401. https://doi.org/10.3390/md18080401

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2020). *Biostrings: Efficient manipulation of biological strings* . R package version 2.58.0. *https://bioconductor.org/packages/Biostrings*.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Raymond-Bouchard, I., Goordial, J., Zolotarov, Y., Ronholm, J., Stromvik, M., Bakermans, C., & Whyte, L. G. (2018). Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. *FEMS Microbiology Ecology*, *94*(4), Article fiy023. https://doi.org/10.1093/femsec/fiy023

Riesenfeld, C. S., Murray, A. E., & Baker, B. J. (2008). Characterization of the Microbial Community and Polyketide Biosynthetic Potential in the Palmerolide-Producing Tunicate *Synoicum adareanum*. *Journal of Natural Products*, *71*(11), 1812-1818. https://doi.org/10.1021/np800287n

Rodriguez-R, L. M., & Konstantinidis, K. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *4:e1900v1*. https://doi.org/https://doi.org/10.7287/peerj.preprints.1900v1

Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rossello-Mora, R., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the

whole genome level. *Nucleic Acids Research*, *46*(W1), W282-W288. https://doi.org/10.1093/nar/gky467

Rodriguez-R, L. M., & Konstantinidis, K. (2014). Bypassing Cultivation To Identify Bacterial Species: Culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. *Microbe Magazine*, *9*, 111-118.

Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, *13*(8), 509-523. https://doi.org/10.1038/nrmicro3496

Schmidt, E. W. (2015). The secret to a successful relationship: lasting chemistry between ascidians and their symbiotic bacteria. *Invertebrate Biology*, *134*(1), 88-102. https://doi.org/10.1111/ivb.12071

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068-2069. https://doi.org/10.1093/bioinformatics/btu153

Selvin, J., Gandhimathi, R., Kiran, G. S., Priya, S. S., Ravji, T. R., & Hema, T. A. (2009). Culturable heterotrophic bacteria from the marine sponge *Dendrilla nigra*: isolation and phylogenetic diversity of actinobacteria. *Helgoland Marine Research*, *63*(3), 239-247. https://doi.org/10.1007/s10152-009-0153-z

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In M. Kollmar (Ed.), *Gene Prediction: Methods and Protocols* (Vol. 1962, pp. 227-245). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9173-0_14

Shieh, W. Y., Lin, Y. T., & Jean, W. D. (2004). *Pseudovibrio denitrificans* gen. nov., sp nov., a marine, facultatively anaerobic, fermentative bacterium capable of denitrification. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 2307-2312. https://doi.org/10.1099/ijs.0.63107-0

Slightom, R. N., & Buchan, A. (2009). Surface Colonization by Marine *Roseobacters*: Integrating Genotype and Phenotype. *Applied and Environmental Microbiology*, *75*(19), 6027-6037. https://doi.org/10.1128/aem.01508-09

Soldatou, S., & Baker, B. J. (2017). Cold-water marine natural products, 2006 to 2016. *Natural Product Reports*, *34*(6), 585-626. https://doi.org/10.1039/c6np00127k

Stevenson, B. S., Eichorst, S. A., Wertz, J. T., Schmidt, T. M., & Breznak, J. A. (2004). New strategies for cultivation and detection of previously uncultured microbes. *Applied and Environmental Microbiology*, *70*(8), 4748-4755. https://doi.org/10.1128/aem.70.8.4748-4755.2004

Takaichi, S., Maoka, T., Akimoto, N., Carmona, M. L., & Yamaoka, Y. (2008). Carotenoids in a Corynebacterineae, *Gordonia terrae* AIST-1: Carotenoid Glucosyl Mycoloyl Esters. *Bioscience Biotechnology and Biochemistry*, *72*(10), 2615-2622. https://doi.org/10.1271/bbb.80299

Tian, Y. A., Li, Y. L., & Zhao, F. C. (2017). Secondary Metabolites from Polar Organisms. *Marine Drugs*, *15*(3), 30, Article 28. https://doi.org/10.3390/md15030028

Tsukamur.M. (1971). Proposal of a new genus, *Gordona*, for slightly acid-fast organisms occurring in sputa of patients with pulmonary disease and in soil. *Journal of General Microbiology*, *68*(SEP), 15-&. https://doi.org/10.1099/00221287-68-1-15

Tully, B. J., Wheat, C. G., Glazer, B. T., & Huber, J. A. (2018). A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *ISME Journal*, *12*(1), 1-16. https://doi.org/10.1038/ISMEj.2017.187

Wang, J. S., Peng, L. H., Guo, X. P., Yoshida, A., Osatomi, K., Li, Y. F., Yang, J. L., & Liang, X. (2019). Complete genome of *Pseudoalteromonas atlantica* ECSMB14104, a Gammaproteobacterium inducing mussel settlement. *Marine Genomics*, *46*, 54-57. https://doi.org/10.1016/j.margen.2018.11.005

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80-83. https://doi.org/10.2307/3001968

Wirth, J. S., & Whitman, W. B. (2018). Phylogenomic analyses of a clade within the *roseobacter* group suggest taxonomic reassignments of species of the genera *Aestuariivita, Citreicella, Loktanella, Nautella, Pelagibaca, Ruegeria, Thalassobius, Thiobacimonas* and *Tropicibacter*, and the proposal of six novel genera. *International Journal of Systematic and Evolutionary Microbiology*, *68*(7), 2393-2411. https://doi.org/10.1099/ijsem.0.002833

Wu, Y. W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605-607. https://doi.org/10.1093/bioinformatics/btv638

Zimin, A. V., Marcais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669-2677. https://doi.org/10.1093/bioinformatics/btt476

Appendix 1 Detailed cultivation procedures

Isolation and cultivation

Agar plates and liquid media tubes were inoculated with a microbial sample prepared from a tissue homogenate of *Synoicum adareanum* sample "Nor2C-2011". Five carbon sources (chitin, xylose, cellobiose, taurine, and dextrose) were used at 2.5 mM and a basal saltwater media (adapted from (Stevenson et al., 2004) for both liquid media and agar plates during the initial cultivation from Nor2C-2011. Agar and liquid media were also prepared at 2.5 mM with the addition of yeast extract (5 µg/mL final concentration). Both yeast and non-yeast media were inoculated with concentrations of sample at the initial sample concentration, and three consecutive 1:10 dilutions (concentration #s 1-4). A second series of plates and Hungate tubes were prepared and maintained under microaerobic conditions by purging the plate container with nitrogen for 5 minutes, and four cycles of vacuum and nitrogen filling (5 seconds each) into the Hungate tubes. These microaerobic plates and tubes were inoculated with the initial sample concentration and only the first dilution (concentration #s 1&2).

Observations for counting new colonies were made with an Olympus SZ-CTV 10X-63X microscope and an AmScope Trinocular 7X-90X microscope. For each media type and sample concentration, new and potential colonies were documented and labelled on the plate cover. These new colonies were indicated on the plate cover with a unique marker color and shape corresponding to the date they were first observed. Potential colonies were marked with a temporary tape "arrow" to reference during the next count.

Individual colonies were considered unique by a combination of their morphologies and timing when first observed. Unique colonies were selected for isolation and cultivation in 10mM liquid media and agar plates. During the selection and picking process, the plate was visually scanned to identify different colonies of interest, by simultaneously comparing their morphologies, size, and their corresponding date-indication mark(s). A selected colony was documented by its first-observed date, current morphology, diameter, and its location within the plate was circled and numbered on the plate cover. Each media type required its own corresponding isolate number sequence (i.e. all five carbon sources had selected colony numbers beginning at "1"). While under sterile laminar air-flow, the colonies were picked with a sterile inoculating loop from the original plate and suspended in 100 µL of media containing the same carbon type, but at a higher carbon-source concentration and yeast extract concentration (10mM and 20 µg/mL, respectively). After gentle pipette mixing, this 100 µL was split; 50 µL was inoculated into 2 mL of media in a rubber-capped Hungate tube, and the other 50 µL was spread onto an agar plate. The proximity of colonies made some isolates difficult to collect, and any concerns during collection of accidentally touching neighboring colonies were mitigated by directly streaking the questionable sample onto the new plate. These streak plates were then monitored and selected for unique isolates following the same process.

Growth on initial homogenate plates was monitored and isolates were cultivated microaerobically. For plates this was achieved by purging the plate container (either an anaerobic jar or sealed plastic bags) with nitrogen for 5 minutes. In the case of plastic bag containers, bags were individually purged with nitrogen, sealed, and taped closed. For

the isolates in liquid media, the septum cap of the hungate tube was pierced with a needle to allow for four consecutive cycles of 5 seconds of vacuum, followed by 5 seconds of nitrogen filling.

Positive pressure was maintained in the tube after the final nitrogen fill by turning the manifold valve off, and a syringe containing 4.2 mL of air was quickly swapped onto the stationary needle and injected into the tube. The use of one sterile, disposable needle per hungate tube for the microoxygenation process allowed the positive pressure from the nitrogen filling to be released through the needle during the exchange for the syringe, reducing any potential contamination, and minimizing the only overall positive pressure from the introduction of 4.2 mL air. All tubes and plates were incubated at 10º C Hungate tube microaerobic calculations: 4.2 mL of 20% O_2 containing air, into 16.5 mL volume = 5% O2 tube environment.

Negative controls were created in the form of blank samples, inoculated on plates and tubes of different media throughout the collection period. These followed the same naming protocols, with the addition of "Blank" to the end of the plate/tube number.

Preparations for DNA extractions

New isolate Hungate tubes and agar plates were monitored for cultures, and no significant growth has been consistently visible among any liquid media tubes. Plates were scraped of their colonies after being determined to have a significant, collectable mass. The use of sterile, disposable inoculating loops has been a convenient way to collect the mass without gouging the agar. This mass was then deposited into a 2 mL tube containing 1.5 mL of the same growth medium, and lightly vortexed to suspend.

Then, 750 µL was transferred into a labelled Cryovial containing 250 µL 60% glycerol and stored in -80 ºC for preservation. The remaining 750 µL was kept in the 2mL tube and centrifuged in 4 ºC at 13,200 RPM for 5 minutes. The supernatant was decanted, and the pellet stored at -20 ºC for DNA extraction.

Appendix 2 Detailed molecular biology procedures

DNA extractions and quantification

DNA extraction followed the method used by Massana, et al. (1997) (Massana et al., 1997). To each tube containing a colony pellet, 450 µL of sucrose lysis buffer was added, then 10 µL of Roche™ lysozyme (50 mg/mL). The tubes were then rotated in 37 ºC for 30 minutes. 25 µL of proteinase K (10 mg/mL) was added to each tube, quickly shaken, then 25 µL 20% SDS was added. The tubes were then placed back into rotation, but at 55 ºC for 1 hour. 1µL of Ambion™ RNAse A was added to each, then placed into a 37 ºC water bath for 20 minutes. While working in the fume hood, an equal volume (511 µL) of penol:chloroform:isoamyl alcohol (25:24:1) was added to each tube, vortexed for 10 seconds, then centrifuged at 13,200 RPM for 5 minutes. The aqueous layer of each was pipetted and transferred to a new tube. SaTau_y4s had a yellow-tinted supernatant. Any tubes that had a protein layer that created problematic pipetting of the supernatant had an additional 450 µL of penol:chloroform:isoamyl alcohol added, vortexed, and centrifuged again. If necessary, a third series of penol:chloroform:isoamyl alcohol, vortexing, and centrifugation was performed.

The supernatant was pipetted into new tubes, and to this 400 µL chloroform:isoamyl alcohol (24:1) was added, vortexed for 10 seconds, and centrifuged for 5 minutes at 13,200 RPM. As above, some tubes received an additional 400 µL of chloroform:isoamyl alcohol, vortexed and centrifuged to ensure collection of pure supernatants. Volumes of supernatants for each of the new tubes were measured. To each tube, equivalents of 1/10 of the volume of 3M sodium acetate, and equivalents of 2x

the volume of 100% ethanol were added, and the tubes were inverted to mix during the reagent additions. Tubes were stored at -20 ºC for a 2-hour minimum. Tubes were then centrifuged in 4 ºC for 30 minutes at 13,200 RPM. The ethanol was pipetted off, while being careful to not disturb the pellet. To each tube 700 µL 70% ethanol was added, and the pellet was flicked into suspension. The tubes were centrifuged in 4 ºC at 13,200 RPM for 5 minutes. Ethanol was pipetted off, and the tubes were left open, covered under a Kimwipe, for 2+ hours to allow any excess ethanol to evaporate. Once dry, 100 µL of 10mM PCR water was added to each tube with gentle pipette mixing to suspend the DNA pellet. Fluorometry and Quant-iT™ PicoGreen DNA quantitative assay reagents were used to estimate DNA concentrations. Using 20X stock Tris-EDTA (TE) buffer, 1X TE buffer was prepared in quantities sufficient for 2 blanks (100 µL TE buffer each), and as many sample unknowns, in duplicate (98 mL TE buffer each).

Quantities of PicoGreen™ (1:200) were prepared from stock Quant-iT™ for the addition of 100µL to each of the above, as well as to 9 DNA standards in duplicate. Using a black 96 well plate, duplicate 100 µL TE buffer blanks were added, as well as nine 100 µL DNA standards in duplicate (0.0078 ng/µL, 0.0156 ng/µL, 0.0313 ng/µL, 0.0625 ng/µL, 0.125 ng/µL, 0.25 ng/µL, 0.5 ng/µL, 1.0 ng/µL, and 2.0 ng/µL). Into wells containing 98µL TE buffer, 2 µL of unknown sample DNA was added. To all wells 100 µL PicoGreen™ (1:200) was added, with a single mix of the pipette. New DNA standards were made through serial dilutions which covered both high and low range concentrations, as described in protocols.

Using Softmax Pro software and a SpectraMax Gemini fluorometer (excitation: 485 nm, emission: 515 nm), the well plate template was created by assigning to the

program which wells contained blanks, standards, and unknowns. No corrections for the unknown concentrations were applied within the program, and the true concentration was calculated after (to account for the 2µL of sample added).

Screening for palmerolide A using PCR and agarose gel electrophoresis

Using the measured DNA concentrations, 25 µL of working stock was created of DNA suspended in 10mM Tris pH8 (made at 1 ng/µL or kept at initial concentration if < 1ng/µL). From this working stock, 1ng was used in PCR "master mix" (water, NEB 10X buffer, 2.5 mM dNTPs, primers, and Taq polymerase). Two primer sets were used (ATIF/ATIR and HCSF/HCSR), targeting a first acyltransferase and HMG-CoA synthetase regions unique to the palmerolide A gene cluster. Thermocycling for PCR utilized an Applied Biosystems GeneAmp™, with initial denaturation at 95 ℃ for 5 minutes, then 30 cycles of denaturing, annealing, and extension as follows: 95 ℃ for 30 seconds, 60 ℃ for 1 minute, 72 ℃ for 1 minute. A final 5-minute extension was held at 72 ℃. Lastly, the temperature was held at 4 ℃ indefinitely.

Using both primer sets, the DNA samples, two positive controls and one 'no-DNA' blank were amplified. The two positive controls included a natural microbiome-wide DNA extract (Nor2P7AM16C) and a synthetic palmerolide A gene region DNA sequence (PalA-4CDSCtrl).

PCR tubes were kept at 4ºC after amplification. 75 mL of 2% agarose was poured into a gel electrophoresis tray. For the Promega™ 50bp DNA step ladder, 1.5 µL loading dye, 7.5 µL water, and 1 µL ladder were mixed and placed into the first well. 5 µL of the PCR products were mixed with 3.5 µL water and 1.5 µL loading dye.

Early gels were run at 65 volts until loading dye bands migrated ~4 cm, but subsequent gels were run at 100 volts for 1 hour, as this produced clear ladder and sample migration. The gels were then transferred to a glass pan and immersed in 3X Gel Red and agitated for 30 minutes. The gels were photographed using a UVP® BioDoc-It® UV transilluminator and samples were compared to the positive control DNA ladder migration.

Isolate dereplication by 16S rRNA PCR-DGGE

Fifty *Synoicum adareanum* microbiome DNA extracts from the microaerophilic culture collection and 25 microbiome DNA extracts from a previous aerophilic culture collection have been amplified for 16S rRNA encoding regions using Bact GC358F(GC-clamp)/517R primers for screening in denaturing gradient gel electrophoresis (DGGE) using a BioRad DCode™ Universal Mutation Detection System. One ng working stock template DNA for samples, prepared as above, were used for PCR. Applied Biosystems PCR reagents were used, so adjustments to the PCR protocol above included supplemental Applied Biosystems $MgCl_2$ solution following manufacturers protocol for the Applied Biosystems AmpliTaq Gold™ and Buffer II used. Thermocycling parameters were as follows: initial denaturation at 95℃ for 5 minutes. Then 10 cycles of: 95 ℃ for 45 seconds, touchdown annealing (starting at 65 ℃) for 30 seconds, 72 ℃ for 30 seconds. Then 22 cycles as follows: 94 ℃ for 30 seconds, 55 ℃ for 30 seconds, 72 ℃ for 30 seconds. A final 10-minute extension was held at 72 ℃. Lastly, the temperature was held at 4 ℃ indefinitely.

Early PCR reaction volumes were made at 20 µL, but later volumes were made at 25 µL to generate additional PCR product for subsequent DGGE comparisons. Positive controls using Nor2-7AM14C (Nor2) were amplified in individual tubes at 50 µL reaction volumes to be used as reference bands in DGGE gels. PCR products from this Nor2 positive control were combined and purified using sodium acetate and ethanol as described in the extraction protocol above. Nor2 DNA was then pelleted, resuspended in Tris or PCR water, and quantified with PicoGreen™ Quant-iT™ as describe above. Standardized stock Nor2 DNA references were made by aliquoting PCR tubes at volumes equaling 200 ng of DNA per tube, to which 2.5 µL of gel loading dye was then added. Tubes were stored at -20 ℃.

Two % agarose gels were prepared and run as above to confirm successful amplification of PCR products against co-amplified Bioline Hyper Ladder II, positive and negative controls.

We made 80mL of 100% denaturant 8% gels as follows:

-16 mL 40% acrylamide (BioRad)

-1.6 mL 50X TAE buffer

-32 mL formamide (Sigma)

-33.6 g urea (Fisher Sci.)

-4 mL Milli-Q water,

The above were placed into a jar on ~55 ℃ heat with stirring, then Milli-Q water was added up to 80mL once dissolved. The jar was cooled under running water for ~30

seconds and/or allowed to equilibrate with room temperature to avoid accelerated polymerization time.

We made 100 mL of 0% denaturant 8% gel as follows:

-20 mL 40% acrylamide (BioRad)

-2 mL 50X TAE buffer

-78 mL Milli-Q water

Preparations were made for running either one or two gels at a time, so reagent volumes were adjusted for varying total gel volumes accordingly.

DGGE gels were made at a 30%-65% denaturing gradient. To make the 30% denaturant, we used 11.2 mL and 4.8 mL of the 0% and 100% denaturant solution, respectively. The 65% denaturant was made with 5.6 mL and 10.5 mL of the 0% and 100% denaturant solution, respectively. To each of these, 144 µL of 10% APS and 14.4 µL of TEMED was added to catalyze polymerization. Gradient gel was poured to a level ~3 mm below the well comb. A 0% denaturant, 8% acrylamide stacking gel was prepared using concentrations as above and loaded with a glass pipette to fill the remaining area around the comb.

The DGGE tank was filled with 1X TAE buffer and temp was initially preheated to 67 ºC. 2.5 µL of gel loading dye was added to each tube containing 10 µL of PCR product. The outermost wells of each gel were not used due to known problems with distorted band patters near DGGE gel edges. The second-outermost wells were loaded with Nor2 control references. Negative control (no DNA) PCR products were checked with agarose gels described above, and thus were not included in DGGE lanes.

After loading gel wells, the temperature was set to 60 ºC and voltage was set to allow for 1,000V/hours (ex. 15 hours x 66.7V = 1,000V/hours) to run overnight. The change-in-resistance feature of the DGGE power supply system was disabled to avoid power issues.

The initial DGGE gels were stained with GelRed™ for 20 minutes, and briefly washed with 300 mL of 0.5 X TAE as a de-stain. Subsequent gels were stained with 200 mL of 1X SYBR™ Gold nucleic acid gel stain. The gels were photographed using a UVP® BioDoc-It® UV transilluminator over a range of exposures. Images were both printed and saved digitally.

A series of dereplicating efforts required a total of 20 PCR amplifications of GC-clamped 16S rRNA regions as described above, followed by cross-comparisons of DGGE gel band migrations to deduce replicate organisms. For bands which appeared to have similar migration patterns, multiple rounds of DGGE were necessary to run candidate replicates in adjacent lanes to confirm or refute their uniqueness. Template DNA which was not amplified when 1ng DNA was used in PCR reactions was adjusted to both higher volumes and 1µL of 1/10 dilution DNA concentration. A final set of three gels was prepared to display band migrations of unique samples.