Insights into the ecology and evolutionary history of the bacterial genus *Psychrobacter*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Daphne Kuha Welter

aus Minneapolis, die Vereinigten Staaten von Amerika

Tübingen

2021

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	13.04.2021
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Ruth Ley, Ph.D.
2. Berichterstatter:	Prof. Dr. Nico Michiels

To my Grandma K.
I wish you could have read this.

Acknowledgements

I would like to thank,

Ruth Ley, for giving me the opportunity to join her lab at Cornell and to move to the MPI with her, as well as for her support and mentorship. Her leadership of the department during the pandemic has been inspiring.

Nico Michiels, for agreeing to be my secondary supervisor, and for his patient and steady mentorship, especially in the terrifying last few months of my studies.

Thesis Advisory Committee members Felicity Jones and Talia Karasov for their support and discussions.

Albane Ruaud. Without her encouragement and help throughout the process of writing my manuscript and thesis, I would not have finished.

Lab members, especially Jacobo de la Cuesta, Emily Davenport, Hagay Enav, Zach Henseler, Albane Ruaud, Jess Sutter, Taichi Suzuki, Tony Walters, and Nick Youngblut, for help with everything from coding to writing to moral support.

Technician Silke Dauser, for her help and technical support in the lab, and also for practicing German with me. Lab manager Ursula Schach for keeping our department running, and department secretary Karin Klein for her endless patience and kindness in helping me navigate life in Germany.

Intern Ariadne Papatheodoru for her enthusiasm for my project, which motivated me when I was feeling most discouraged.

Athina Iliopoulou and Katharina Vernali for keeping our department clean and greeting me cheerfully every morning. Mario Pezzuti for his boisterous greetings and conversations in German, French, and Italian, and occasionally making me coffee or tea.

The creators of my favorite podcasts, My Brother My Brother and Me, The Flophouse, The Worst Idea of All Time, and The Last Podcast on the Left, for keeping me company during long nights in the lab by myself, and for making me laugh whenever I was feeling down.

My former teachers, mentors, and advisors, including but not limited to Tony Borgerding, Sara Di Rienzi, Jayna Ditty, Justin Donato, Sarah Howard, Tom Marsh, Angela Poole, Rose Rinder, and Jim Robinson, for their infectious enthusiasm for their respective subjects, and their faith in me to excel in those subjects. Their support put me on the path I am on today.

My friends in Germany, especially Jacobo de la Cuesta, little Franka Erben, Sylvie Erben, Frederik Harder, Sarah Keim, Patrick Putzky, Jess Sutter, Taichi Suzuki, and Tony Walters, for traveling, cooking, singing, hiking, playing board games/role playing games, and

watching lots of movies with me, and in general, for enriching my life with the diversity of their experiences, and for making the lonely adventure of moving to a new country as positive as possible. Nothing I could write here would properly express my gratitude.

My friends abroad, especially J Capecchi-Nguyen, Sara Jaye, Satavee Kijsanayotin, Jingqiu Liao, Madeleine Marrin, Ashley Skwiera, and Darcy Walter, for remaining my boon companions through thick and thin, for loving the person I have become as well as the person I used to be, for reminding me to have fun, and that there is life outside of and after graduate school.

My past and present pets, but especially my dog Rhubarb, whose profound anxiety once forced me to get my own anxiety under control.

My significant other, Zach Henseler. My artistic and personal project collaborator, co-conspirator, and traveling companion, who continuously reminded me that I was not alone, that “finished” is better than “perfect”, and that I have done my best. I strive to achieve his level of optimism.

My family, especially my mother Lee, my brother Roland, my father Samuel, my half-brother Malcolm, and my grandmother Kathleen, for loving me, believing in me, and supporting me unconditionally, and for reminding me that no matter how far I travel or how long I am gone, I always have a home to return to.

Table of Contents

Abstract

The genus *Psychrobacter* presents an unusually broad range of isolation sources, ranging from the bodies of mammals to extremely cold and saline permafrost soils. In this thesis, I present my investigation of the diversity of *Psychrobacter* with the goal of identifying differences within the genus which may contribute to its surprising ecological distribution. I collected 85 *Psychrobacter* strains from diverse locations and isolation sources, and approached this problem using genomics, including phylogenomics and pan-genome analysis, as well as ancestral character estimation and analysis of protein coding sequences. I also performed extensive phenotyping, collecting growth curves in 24 different conditions in a gradient of temperature, salt, and nutritional complexity, as well as information on strain bile resistance, oxygen utilization, and for a subset of strains, ability to colonize germ-free mice. I show that *Psychrobacter* most likely evolved from a *Moraxella*-like ancestor, that is, a mesophilic pathobiont of mammals, and subsequently diverged into strains that maintain the ability to grow at high temperatures (*ie* 37 ^{o}C) versus those that only grow at cooler temperatures (25 ^{o}C or lower). There are some genomic and phenotypic differences amongst *Psychrobacter* strains based on source of isolation, but ultimately, dividing the genus by their behavior may be more useful. Overall, *Psychrobacter* promises to be an interesting model group for disentangling niche diversification among close relatives.

Zusammenfassung

Innerhalb der Bakterien weist die Gattung *Psychrobacter* ein ungewöhnlich breites Spektrum an Isolationsquellen auf, welche von den warmen Säugetierkörpern bis hin zum extrem kalten und salzhaltigen Permafrostböden reichen. In dieser Arbeit präsentiere ich meine Untersuchungen zu der Diversität von *Psychrobacter* mit dem Ziel, Unterschiede innerhalb der Gattung zu identifizieren, die zu ihrer überraschenden ökologischen Verbreitung beitragen. Um diese Fragestellung zu untersuchen, sammelte ich 85 *Psychrobacter*-Stämme von verschiedenen Standorten und Isolationsquellen und analysierte diese mit Hilfe von genomischer Methoden, einschließlich von Phylogenomik- und Pan-Genom-Analysen, sowie einer Abschätzung abstammender Merkmale und der Analyse von Protein-kodierenden Sequenzen. Eine umfangreiche Phänotypisierung führte ich mittels der Erfassung von Wachstumskurven unter 24 verschiedenen Bedingungen mit variierenden Salzkonzentrationen, Temperaturen und Nährstoffkomplexitäten, sowie durch Verträglichkeitstest unterschiedlicher Gallen- und Sauerstoffkonzentrationen durch. Für eine Untergruppe von Stämmen wurde ebenfalls die Fähigkeit, keimfreie Mäuse zu besiedeln, untersucht. Ich zeige, dass sich Psychrobacter mit großer Wahrscheinlichkeit aus einem Moraxella-ähnlichen Vorfahren entwickelt hat, d.h. einem mesophilen Pathobionten von Säugetieren. Anschließend divergierte dieser in Stämme, welche die Fähigkeit beibehalten, bei hohen Temperaturen (d.h. 37 °C) zu wachsen, oder zu Stämmen, die nur bei kühleren Temperaturen (25 °C oder niedriger) wachsen. Es zeigen sich einige genomische und phänotypische Unterschiede zwischen den *Psychrobacter*-Stämmen, welche auf der Isolierungsquelle beruhen, jedoch ist es letztendlich sinnvoller, die Gattung nach ihrem Wachstumsverhalten zu unterteilen. Insgesamt verspricht *Psychrobacter* eine interessante Modellgruppe für die Entflechtung der Nischendiversifizierung innerhalb nah verwandten Organismen zu sein.

Chapter 1. General Introduction

Eukaryotic organisms have evolved in the constant presence of microbial organisms and host microbial communities which are called their microbiomes. The microbiome of an organism frequently provides functions that the host organism would not otherwise have access to; for example, the bioluminescence of many marine animals is dependent upon light-producing symbionts (1). Frequently, the increased function provided by the microbiome is related to host diet; the microbiomes of ruminants are well-adapted to digest complex polysaccharides that would otherwise provide no nutrition for the hosts (2), and there is evidence that the human microbiome can provide lactose tolerance even in genetically lactose intolerant people (3). With microbial communities having such pronounced fitness effects on their hosts, there is growing interest in characterizing the effect microbes may have on the evolutionary trajectory of their hosts and vice versa (4).

It has been shown that hosts influence the evolutionary trajectories of their associated microbiomes (5). There is phylogenetic clustering of microbial communities based on the host organisms (6), as well as phylogenetic clustering of particular bacteria based on host specificity (7). These are likely influenced by physiological and ecological differences between categories of host (8). Selective pressures exerted by the host on its microbiome include the specificity of the host diet, resulting in differential nutritional availability (9), the time since host feeding resulting in a “feast or famine” nutrient availability scheme (10), relatively rapid movement of the luminal liquid and mucosa resulting in microbial “wash out” (11), extremely high density of other microbial competitors (12), predation by bacteriophage (13), and selection by the host immune system (14).

All of these host-specific selection factors are expected to contribute to the fact that there is little taxonomical overlap between microbes that are found in the mammalian gut versus those found outside of the gut (15), with the notable exception of gastrointestinal pathogens, which frequently have significant reservoirs in water and soil (16, 17). A microbe

able to thrive both within and outside of the mammalian gut could provide insights into microbial lifestyle transitions, which are logically common, but have only been predicted based on phylogenetic data (18).

The genus *Psychrobacter,* from the *Gammaproteobacteria* class*,* has been observed in a huge range of environments. *Psychrobacter* spp. were first included as members of the closely related genus *Acinetobacter*, until the 1980s when *P. immobilis*, first isolated from frozen food, was proposed as the *gen. nov. sp. nov.* strain (19). *Psychrobacter* has been isolated from frozen food, glacial ice, permafrost soil, seawater, invertebrates, fish, birds, and mammals (marine as well as terrestrial) (20). It has also been observed in culture-independent surveys of microbial diversity including marine mammals (21, 22), raw dairy (23), and Antarctic soil (24). There are additionally a few cases of documented *Psychrobacter* infections in lambs (25) and humans (26–28), but it does not cause gastrointestinal disease, but rather meningitis, septicemia, or respiratory distress. The environments in which *Psychrobacter* can be found are extremely different from one another, making its widespread presence surprising and intriguing.

Because of its widespread cultivation from extremely cold environments, much of the interest in the genus *Psychrobacter* is related to exploiting its cold adaptation for cold active enzymes and potentially unique bioactive compounds. *Psychrobacter* strains have been known to produce cold active lipases (29) and antibiotics active against the *Burkholderia cepacia* complex (30), and have been associated with intense violet pigment production in the spoilage of cheese (31). One study to date has recognized the potential for *Psychrobacter* as a model for comparing strains colonizing warm-bodied hosts versus colder environments, but was limited to 26 genomes, only 4 of which were from warm hosts (32).

In this work, I aim to identify differences within the genus *Psychrobacter* which may contribute to its ecological distribution using phylogenetics, genomics and phenotyping. In Chapter 2, I present evidence that *Psychrobacter* evolved from a mammalian-associated, sometimes pathogenic organism, and that some strains subsequently diverged into free-living psychrotolerant lifestyles while others maintained their ability to associate with

mammalian hosts. In Chapter 3, I show that while there are no significant differences in predicted protein cold adaptation based on strain source of isolation, *Psychrobacter* strains from fish, invertebrates, and food-processing environments show significant and potentially ecologically relevant genomic differences compared to other strains. Finally, in Chapter 4, I show that *Psychrobacter* strains isolated from mammals and birds exhibit significant differences in their abilities to grow in gradients of salt and temperature compared to strains from other sources, and while there is no significant difference in oxygen usage, strains from warm hosts also tend to be less oxygen-dependent. By contrast, I found few meaningful differences in growth rate or maximum optical density based on isolation source. Overall, *Psychrobacter* evolution and ecology cannot be explained solely based on populations diverging between different isolation sources. Further population genomic studies are needed to further clarify the diversification of this genus.

Chapter 2. Free-living psychrotrophic bacteria of the genus *Psychrobacter* are descendants of pathobionts.

Manuscript contributions: DKW and REL designed the project. DKW maintained the strain collection and performed genomic sequencing, assembly, and pan-genome analysis, with the aid of NDY. DKW and ZMH performed the phenotypic screen. DKW and AR analyzed phenotypic data. PvCdG and LG collected the polar bear fecal samples, and JM performed dietary analysis. HND performed the polar bear sample sequencing, and HND and DKW analyzed the polar bear sequence data. DKW and JLW performed the mouse studies. DKW, AR, and REL wrote the manuscript.

The following manuscript is the accepted form of an article published by mSystems. The final version was published on April 13, 2021, and can be viewed at https://msystems.asm.org/content/6/2/e00258-21 (DOI: 10.1128/mSystems.00258-21).

Title: Free-living, psychrotrophic bacteria of the genus *Psychrobacter* are descendents of pathobionts

Running Title: psychrotrophic bacteria descended from pathobionts

Daphne K. Welter[1], Albane Ruaud[1], Zachariah M. Henseler[1], Hannah N. De Jong[1], Peter van Coeverden de Groot[2], Johan Michaux[3,4], Linda Gormezano[5¥], Jillian L. Waters[1], Nicholas D. Youngblut[1], Ruth E. Ley[1]*

1. Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen, Germany.

2. Department of Biology, Queen's University, Kingston, Ontario, Canada.

3. Conservation Genetics Laboratory, University of Liège, Liège, Belgium.

4. Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR ASTRE, Montpellier, France.

5. Department of Vertebrate Zoology, American Museum of Natural History, New York, NY, USA.

¥deceased

*Correspondence: rley@tuebingen.mpg.de

Abstract. Host-adapted microorganisms are generally assumed to have evolved from free-living, environmental microorganisms, but examples of the reverse process are lacking. In the phylum *Gamma-Proteobacteria*, family *Moraxellaceae*, the genus *Psychrobacter* includes strains from a broad ecological distribution including animal bodies as well as sea ice and other nonhost environments. To elucidate the relationship between these ecological niches and *Psychrobacter*'s evolutionary history, we performed tandem genomic analyses

with phenotyping of 85 *Psychrobacter* accessions. Phylogenomic analysis of the family *Moraxellaceae* reveals that basal members of the *Psychrobacter* clade are *Moraxella* spp., a group of often-pathogenic organisms. *Psychrobacter* exhibited two broad growth patterns in our phenotypic screen: one group we called the “flexible ecotype” (FE) had the ability to grow between 4 and 37 °C, and the other we called the “restricted ecotype” (RE) could grow between 4 and 25 °C. The FE group includes phylogenetically basal strains, and FE strains exhibit increased transposon copy numbers, smaller genomes, and a higher likelihood to be bile salt resistant. The RE group contains only phylogenetically derived strains and has increased proportions of lipid metabolism and biofilm formation genes: functions that are adaptive to cold stress. In a 16S rRNA gene survey of polar bear fecal samples, we detect both FE and RE strains, but in *in-vivo* colonizations of gnotobiotic mice, only FE strains persist. Our results indicate the ability to grow at 37 °C, seemingly necessary for mammalian gut colonization, is an ancestral trait for *Psychrobacter*, which likely evolved from a pathobiont.

Importance. Bacteria and archaea predate the evolution of eukaryotic cells; logically, it follows that host-associated microbes have evolved from free-living ones. The evolutionary transition of microbes in the opposite direction, from host-associated towards free-living, is theoretically possible and has been predicted based on phylogenetic data, but hasn’t been studied in depth. Here we provide evidence that the genus *Psychrobacter*, particularly well-known for inhabiting low-temperature, high-salt environments such as sea ice, permafrost soils, and frozen foodstuffs, has evolved from a mammalian-associated microbe. We show that some *Psychrobacter* strains retain seemingly ancestral genomic and phenotypic traits that correspond with host association while others have diverged to psychrotrophy or psychrophilic lifestyles.

Introduction

Association with vertebrate hosts was shown to be the largest factor driving differences in the 16S rRNA diversity of microbiomes sampled globally (1, 2). Recent analysis of metagenome-assembled genomes from multiple habitats shows that many of these genomes are either animal host-enriched or environment-enriched, but generally not both (3). Specialization to host-association can also be seen at higher taxonomic levels, indicating that whole lineages may have diverged once animal hosts were first successfully colonized. For instance, within the *Bacteroidetes*, the taxa that are mammal-gut associated are derived from phylogenetically basal clades that include free-living and invertebrate-associated taxa (4). These patterns of distribution imply that specialization to the warm animal host habitat is mostly incompatible with fitness in other environments. The specific diet of the host, chemical environment of the gut, competition with an extremely dense surrounding microbial community, and direct selection by the host immune system all likely contribute to the restricted taxonomy of gut-associated microbes (5).

One line of evidence for the adaptive evolution of gut microbes is the emergence of functions that would not be useful outside of the gut, such as the ability to bind and degrade host-specific nutrient sources like mucin (6), or resistance to bile acids and their salt conjugates (7). Additionally, many gut-associated commensal taxa do not have significant environmental reservoirs, due to the divergence in selective pressures between guts and non-gut environments (8). By contrast, many gastrointestinal pathogens such as *Vibrio cholera (9)* or *Yersinia enterocolitica* (10) have high fitness within the gut and outside of the gut, typically in soil or water. One phylogenetic analysis has indicated that the evolutionary transition from a free-living lifestyle to a host-associated (including pathogenic) lifestyle is more common than the contrary (11).

Species of *Psychrobacter* have been recovered through culture-based and sequenced-based methods from a range of animal microbiomes, including the respiratory blow of marine mammals (12); marine mammal skin (13) and guts (14–16); the throats and

guts of birds (17, 18) and fish (19); and many nonhost environments such as seawater (20), sea ice (21), marine sediment (22), glacial ice (23), and permafrost soil (24). Some *Psychrobacter* species are capable of causing disease in mammalian hosts (25, 26); however, *Psychrobacter* infections are very rare, and the factors leading to infection are unclear. *Psychrobacter*'s wide range of environmental sources and adaptation to cold temperatures could call into question whether *Psychrobacter* is truly associated with mammalian communities, or if it is just an allochthonous member of mammalian microbiomes. A comparative genomics analysis of 26 *Psychrobacter* spp. and metadata gleaned from public sources revealed differences in cold-adaptation of protein coding sequences between warm-host-associated strains versus derived marine and terrestrial strains (27). Furthermore, warm-adapted strains were basal in the *rpoB* gene phylogeny, suggesting that *Psychrobacter* evolved from a mesophilic ancestor. These observations make *Psychrobacter* an interesting candidate to assess how phenotype maps onto phylogeny and source of isolation, and to probe into the evolutionary history of a genus with a wide habitat range.

Here, we investigated the evolutionary history of the genus Psychrobacter, a group of closely related bacteria with a broad environmental distribution. We generated 85 genomes of *Psychrobacter* accessions, which we phenotyped for growth under different temperature and salt conditions. We used a collection of wild polar bear feces collected from across the Canadian Arctic and assessed the presence of *Psychrobacter* ecotypes as a function of bear diet. Finally, we conducted tests with a subset of strains for colonization of the mammal gut using germ-free mice. Our phylogenomic results confirm a mesophilic ancestry for *Psychrobacter* and a common ancestor with the genus *Moraxella*. Our phenotyping revealed that overall, *Psychrobacter* tolerate a wide range of salinity, but growth at 37 ^{o}C divided the accessions into two ecotypes: those that retained the ability to grow at warm temperatures and can colonize mammalian hosts (flexible ecotype, FE), and those that lost the ability to grow at warmer temperatures and are instead restricted to colder temperatures (restricted ecotype, RE). A genomic analysis of the two ecotypes showed

genome reduction in the FE strains with high transposon copy numbers, and increased biofilm formation capability in RE strains, as well as higher proportions of lipid metabolism genes. We showed that *Psychrobacter* that are basal are FE, but that FE are also phylogenetically interspersed with RE, indicating either re-adaptation to the animal host or retention of the basal traits. Although both FE and RE ecotypes were detected in the feces of wild polar bears, of the subset of *Psychrobacter* strains tested, only FE strains could colonize the germfree mouse gut. Together our results indicate the genus *Psychrobacter* evolved from a pathobiont ancestor, with one lineage largely losing its ability to associate with animals in its adaptation to psychrophilic, nonhost environments.

Results

***Psychrobacter* forms a clade whose basal members are of the *Moraxella* genus.** To explore the evolutionary history of *Psychrobacter*, we built a phylogeny based on 400 conserved marker genes using publically available whole genomes derived from cultured isolates of 51 species from the *Moraxellaceae* family. We included 18 species each of *Acinetobacter* and *Moraxella* obtained from NCBI, and 15 *Psychrobacter* genomes that we generated in this study (Table S1). For all genomes, we incorporated phenotypic data collected from previously published type-strain research. The 51 species formed 3 distinct clades, each with robust bootstrap support. The *Acinetobacter* clade consists uniquely of *Acinetobacter* species (labeled A in Fig 1A). The *Acinetobacter* clade is a sister taxon to the *Moraxella* (M) clade, consisting entirely of *Moraxella* species, and to the *Psychrobacter* (P) clade, which contains all of the *Psychrobacter* species, as well as 4 *Moraxella* species - *M. boevrei*, *M. atlantae*, *M. osloensis*, and *M. lincolnii* - that are basal.

Consistent with the topology of the phylogeny, a principal coordinates analysis (PCoA) of gene presence/absence data shows that the greatest variation within the family (*i.e.*, Principle Coordinate [PC] 1) is the separation between the A clade versus the M and P

clades, which are grouped (Fig. 1). Genes annotated with very diverse functions, falling under almost every Cluster of Orthologous Groups (COG) category, strongly contribute to the separation between clades according to envfit analysis (Fig. 1B, Table S2). The P-clade *Moraxella* spp. fall between the P-clade *Psychrobacter* and the M-clade *Moraxella* when visualizing PC1-2 (Fig 1B) as well as PC2-3 (Fig 1C).

Despite their close phylogenetic relationship and high similarity in gene presence/absence, *Psychrobacter* and *Moraxella* have different genomic properties. When examined by genus rather than clade, *Psychrobacter* species have an average genome size of 3.12 ± 0.27 Mb, while the average of *Moraxella* is 2.41 ± 0.28 Mb (pairwise Wilcoxon rank sum test, p-value = 8e-07). *Moraxella* species have an average coding density of 86.4 ± 1.33%, which is significantly higher than the *Psychrobacter* species average of 82.7 ± 1.33% (pairwise Wilcoxon rank sum test (WRS), p-value = 1e-5). Despite being phylogenetically more related to *Psychrobacter* species than to the other *Moraxella* species, the P-clade *Moraxella* are significantly different from *Psychrobacter* species for both genome size and coding density (pairwise WRS, p-values = 0.004) while not significantly different from the M-clade *Moraxella* spp. (pairwise WRS, p-values = 0.9).

***Psychrobacter* spp. ranges of growth temperatures differ from those of *Moraxella*.** To examine the phenotypic behaviors of the *Moraxellaceae* family, we applied continuous trait mapping of the ranges of temperatures at which species from the *Moraxellaceae* family can grow ((16, 17, 20, 21, 26, 28–68)) onto the previously generated phylogeny (Fig 2). The *Psychrobacter* spp. included here exhibit a broad range of growth temperatures (0 - 38 ºC), but several strains, such as *P. frigidicola* and *P. glacincola*, are psychrophilic (restricted to growth below 20 ºC), which is a phenotype that is not seen elsewhere in the family. *Psychrobacter* spp. have lower minimum growth temperatures than the included *Moraxella* spp. from either the P- or M-clades (pairwise WRS, p-value = 5e-06), which have a narrow range of temperatures at which they can grow (between 22 ºC and 40 ºC). In contrast to the minimum growth temperatures of the family, there is little variation in the maximum growth

temperatures, except for the notable exceptions of several *Psychrobacter* spp. that are entirely restricted to low temperatures.

As with their genomic properties, P-clade *Moraxella* have more similar phenotypes compared to M-clade *Moraxella* than to *Psychrobacter* strains. Review of type strain descriptions revealed that *Psychrobacter* are consistently urease positive, nitrate reducing, salt tolerant, and non-fastidious, whereas P-clade *Moraxella* are inconsistent with their urease and nitrate reducing phenotypes, are sensitive to high salt concentrations, and are often nutritionally fastidious - they have complex growth requirements (in particular, often blood or bile for growth) (16, 17, 26, 32, 35, 43, 44, 59).

Few *Psychrobacter* strains can grow at 37 ºC. To explore *Psychrobacter* phenotypic diversity, we established a strain collection of 85 *Psychrobacter* accessions isolated from diverse locations and ecological sources (Table S3). To characterize *Psychrobacter* phenotypes under a number of different conditions, we assessed the strain collection for resistance to bile salts as well as ability to grow under 24 different conditions: two different media combined with four different salt concentrations and with three different temperatures (Methods).

We calculated growth probabilities, or the fraction of growth positive conditions out of total conditions tested, for every strain and given variable of the growth curve screen, and compared them along with the bile resistance data across the phylogeny and by isolation source (Fig 3A). We generated a robust genus-level phylogeny for *Psychrobacter* using 400 conserved marker genes, with *M. lincolnii* as an outgroup. In agreement with single marker gene trees generated using *rpoB* sequences (27) and 16S rRNA gene sequences (69) as well as the P-clade structure of *Moraxellaceae* family tree generated in this study, there is a phylogenetically basal group of strains mostly isolated from mammals, and a phylogenetically derived group isolated from mixed sources. Across the entire phylogeny, closely related strains have similar growth probabilities (Pagel's λ ranging from 0.78 to 0.97, all corrected p-values < 1e-3). We observed that most *Psychrobacter* strains are tolerant of a

wide variety of temperatures between 4 and 25 ^{o}C and of salt concentrations between 0 and 5%; more than 90% of all strains can grow under these conditions. However, only 54% of the tested accessions can grow at 10% added salt, and only 31% at 37 ^{o}C.

Since we were interested in the possibility of *Psychrobacter* interacting with warm-bodied hosts, we divided the strains into two ecotypes: the “flexible ecotype” (FE) corresponds to strains that could grow at 37 ^{o}C, and the “restricted ecotype” (RE) corresponds to strains that could not grow at 37 ^{o}C. FE strains are psychrotrophic (mesophilic organisms with a low minimum growth temperature but an optimal growth temperature above 15 ^{o}C), and RE strains are either psychrotrophs or true psychrophiles (unable to grow at temperatures higher than 20 ^{o}C).

Notably, the basal clade of the *Psychrobacter*-only tree (Fig 3) consists solely of FE strains, while the rest of the phylogeny is made up of a mixture of FE and RE strains. Furthermore, the basal FE strains have higher growth probabilities at 37 ^{o}C compared to other FE strains (Pagel’s $\lambda = 0.89$, p-value = 2e-5). Frequencies of RE and FE strains vary significantly across sources of isolation: the FE group is significantly enriched in strains derived from mammalian sources, and the RE group is significantly enriched in strains derived from other hosts, food, marine and terrestrial sources (χ^2-test, 83% of p-values adjusted for group size < 0.05). Nonetheless, both FE and RE ecotypes contain strains from other environments, including mammalian-derived strains within the RE group.

FE strains show higher growth probabilities in complex media compared to defined media, while RE strains showed little difference between the two (WRS, W = 1.13e3, p-value = 0.0005, phylogenetic p-value = 0.002). FE strains also have higher growth probabilities at low- to mid-salt concentrations (WRS, all p-values < 0.05, phylogenetic p-values < 0.4), though there is no difference between FE and RE strains at 10% salt (WRS, W = 811.5, p-value = 0.7, phylogenetic p-value < 0.4). Finally, FE strains are more likely to be resistant to bile salts than RE strains (χ^2-test, 93% of p-values adjusted for group size < 0.05).

FE and RE *Psychrobacter* spp. have functional and structural genomic differences. When exploring gene presence/absence via PCoA, accessions from the basal FE-only subclade cluster closely together, indicating similar gene content. In agreement with the phylogeny, these basal FE accessions are separated from the other accessions on PC1, while the derived FE and RE accessions are more scattered, indicating more diverse gene content (Fig 4A). According to an envfit analysis, the separation is most strongly driven by genes from the COG categories T, signal transduction (r^2 = 0.65, p-value = 0.001); U, trafficking and secretion (r^2 = 0.63, p-value = 0.001); P, inorganic ion transport and metabolism (r^2 = 0.56, p-value = 0.001), and X, unassigned or no homologs in the COG database (r^2 = 0.51, p-value = 0.001)(Table S2).

The average *Psychrobacter* genome carries 2673 genes, 1598 of which are core (present in between 90 - 100% of strains), 1005 are shell (present in 2 or more strains but fewer than 90%), and 70 are cloud (present in only one strain) . The average FE genome has significantly fewer core (WRS, W = 531.5, p-value = 0.02, phylogenetic p-value = 0.3) and shell genes (WRS, W = 462, p-value = 0.003, phylogenetic p-value = 0.4) but significantly more cloud genes (WRS, W = 1014, p-value = 0.02, phylogenetic p-value = 0.2) compared to the average RE genome (Fig 4B).

In terms of functional differentiation between the ecotypes, RE strains' genomes have a higher proportion of "W" (extracellular structures) genes (WRS, W = 275, adjusted p-value = 0.00006, phylogenetic p-value = 0.04), "I" (lipid transport and metabolism) genes (WRS, W = 437, adjusted p-value = 0.01, phylogenetic p-value = 0.2), and "S" (function unknown) genes (WRS, W = 492, adjusted p-value = 0.04, phylogenetic p-value = 0.4). FE strains' genomes have a higher proportion of "P" (inorganic ion transport and metabolism) genes (WRS, W = 1.09e3, adjusted p-value = 0.01, phylogenetic p-value = 0.8). They also have a significantly higher proportion of COG category "L" (replication-, recombination-, and repair-related) genes (WRS, W = 1.10e3, adjusted p-value = 0.01, phylogenetic p-value = 0.3), and in particular, have higher copy numbers of transposons (WRS, W = 1.08e3, p-value = 0.003, phylogenetic p-value = 0.1) compared to RE strains' genomes (Fig 4C and E).

Increased transposon activity can lead to interruption and decay of functional protein-coding genes, leading to an increase in pseudogenes (70). Given the higher number of transposons in FE strains, we next examined the number of predicted pseudogenes between the ecotypes. FE genomes are predicted to have a higher proportion of pseudogenes than RE strains (WRS, W = 997, p-value = 0.03, phylogenetic p-value = 0.6; Fig. 4C). Finally we compared the average number of genes per genome between the ecotypes, as bacterial genomes are known to strongly select against the accumulation of pseudogenes resulting in gene deletions (71), and found that FE strains have significantly fewer genes than RE strains (WRS, W = 547, p-value = 0.05, phylogenetic p-value = 0.6).

We looked for specific gene clusters differentiating the FE and RE ecotypes while controlling for population structure by performing a microbial pan-genome wide association analysis with the R package treeWAS (72). This returned four gene clusters which significantly correlated with ecotype, each of which is enriched in RE strains compared to FE strains (Table 1). The gene with the most significant enrichment is annotated as a transcriptional regulator in the TetR/AcrR family with the KEGG KOs of K16137 and K19335, which is predicted to either be *nemR* or *bdcR*. Other significant results include *lpdA*, dihydrolipoamine dehydrogenase related to amino acid metabolism (KEGG KO K00382/383), a hypothetical protein with the IG domain, and a hypothetical protein with no homologs in eggNOG database and no predicted annotation. Each of these genes has paralogs that are not significantly enriched in one ecotype over the other.

Polar bear feces collected from the Arctic ice have high abundance of *Psychrobacter*.

Psychrobacter spp. have been reported previously in the skin, respiratory, and gut microbiomes of several marine mammals, including whales, porpoises, seals, and sea lions; it could be argued that *Psychrobacter* presence is due to constant exposure from the surrounding seawater. Polar bears, on the other hand, are marine mammals that do not spend as much time swimming. We surveyed the gut microbial diversity of 86 polar bear fecal samples, 76 wild and 10 captive, by 16S rRNA gene sequencing. Sequence variant

clustering of the 16S rRNA gene sequences reveals that *Psychrobacter* is detectable in 88% of the samples (Fig 5A). The large majority of *Psychrobacter* sequences were assigned to unclassified *Psychrobacter* spp., which is prevalent in 87% of samples with a mean abundance of 22%. We also detected RE strain *P. immobilis* in 49% of samples with a mean abundance of 3%, and FE strain *P. pulmonis* in 8% of samples with a mean abundance of 0.5%. Polar bear diet significantly impacts the abundance of *Psychrobacter* spp. (Kruskal-Wallis χ^2 = 13.5, df = 3, p-value = 0.004); we found that polar bears feeding on mammalian prey, including seals and reindeer, had significantly higher abundances of unclassified *Psychrobacter spp.* than polar bears feeding on avian prey or mixed diets (pairwise Wilcoxon rank-sum test, adjusted p-values < 0.05) (Fig 5B). However, diet data is confounded with other metadata, as location (Kruskal-Wallis χ^2 = 8.6, df = 4, p-value = 0.0009) and year (Kruskal-Wallis χ^2 = 16.6, df = 5, p-value = 0.005) of sample collection also significantly impact unclassified *Psychrobacter* spp. abundance. Captive status does not significantly impact abundance, but there is a trend of wild bear samples having higher unclassified *Psychrobacter* spp. abundance than samples from captive bears (WRS, W = 222, p-value = 0.08).

To elucidate the effect that host dietary nutrition may have on *Psychrobacter* growth, we tested the maximum change in absorbance for 190 different carbon sources by a subset of *Psychrobacter* accessions including 9 FE and 10 RE strains (Fig 5C). All *Psychrobacter* spp. reached significantly higher OD_{600} growing on amino acid carbon sources compared to carbohydrates or sugar alcohols (pairwise WRS, adjusted p-values < 0.05). *Psychrobacter* spp. reach the highest OD_{600} growing on fatty acids, surfactants, and peptides, but the differences are not significant. There was no significant difference between FE and RE strains' changes in absorbance (WRS, W = 1.55e6 , p-value = 0.08), though there is a slight trend of RE strains having overall higher changes in absorbance.

***Psychrobacter* strain survival in gnotobiotic mice.** To assess the survivorship of *Psychrobacter* in a mammalian gut, we tested 8 accessions for persistence in the

gastrointestinal tracts of germ-free mice (Fig. 5D). Chosen for phylogenetic breadth, we tested 4 FE strains, *P. ciconiae*, *P. faecalis PBFP-1*, *P. lutiphocae*, and *P. pacificensis*, and 4 RE strains, *P. cibarius JG-220*, *P. immobilis S3*, *P. namhaensis,* and *P. okhotskensis MD17*. Of the four FE strains tested, three were able to persist in the mice, while the FE strain *P. faecalis PBFP-1* and none of the RE strains were detectable after three weeks. Phylogenetic relatedness does not correlate with ability to colonize, as FE strain *P. pacificensis* was able to persist, while closely related RE strains, *P. namhaensis* and *P. okhotskensis*, were not. Phylogenetic placement does correlate with colonization density however, as the two most basal strains tested, *P. lutiphocae* and *P. ciconiae*, colonize at significantly higher densities than the most derived strain that was successful, *P. pacificensis* (pairwise WRS, both adjusted p-values = 0.0007).

Discussion

The phylogenonic and phenotypic characterization of the *Psychrobacter* genus indicates a common ancestor with *Moraxella*, all of which are restricted to growth at higher temperatures. Furthermore, the most basal members of the *Psychrobacter* clade are *Moraxella* species and species of *Psychrobacter* that can grow at 37 ^{o}C, unlike most of the derived *Psychrobacter* species. Our extensive phenotyping indicated that members of the *Psychrobacter* genus grow at a wide range of salinities and temperatures, but distinguishing strains by their ability to grow at 37 ^{o}C results in two groups of strains with differences in response to rich versus defined media, bile salt resistance, and genomic characteristics. Our analysis of a collection of wild polar bear species indicates both RE and FE strains are present, however tests in germfree mice support the notion that only FE can colonize the mammal gut, whereas RE may be allochthonous members or environmental contaminants. Together with previous reports, this work indicates the genus *Psychrobacter* is a lineage of pathobionts, some of which have evolved to inhabit the colder environments of their warm-bodied hosts.

Our results corroborate those of Bakermans, who used the isolation source of *Psychrobacter* as a proxy for temperature adaptation to conclude the genus has a mesophilic ancestor (27). By assessing growth under the same 24 conditions for 85 strains, we remove any ambiguity that can stem from whether an isolate can indeed grow at the temperature of its source of isolation. This is particularly important given that several strains isolated from mammals proved to be RE, and that many of the FE strains came from seawater or other relatively cold environments. We also improved upon the results of Bakermans by including phylogenetically-aware statistics. In any population, identifying traits that are shared due to ecological convergence versus those that are shared due to common ancestry is a well-recognized problem (73); this problem is particularly striking in the genus *Psychrobacter*, where not only is there phenotypic divergence between the more phylogenetically basal strains and the derived strains, but as evidenced by the gene presence/absence PCoA, there is a huge difference in gene content based on phylogeny.

We found that a transcriptional repressor, either NemR or BdcR, is enriched in RE strains' genomes compared to FE strains. NemR plays a role in electrophile sensing (74), while BdcR plays a role in cyclic-di-GMP sensing (75). Both relate to increased biofilm formation when expressed. Similarly, we found a predicted Ig-domain protein, which are thought to be outer membrane proteins with a role in adhesion (76), enriched in RE strains. Adhesion and biofilm formation can increase during cold stress (77), indicating that it may be adaptive for psychrophiles or psychrotrophs to have increased biofilm formation capability.

In addition to specific genes being enriched in RE strains, we found broader functional differences between the ecotypes. FE strains had higher proportions of COG category "L" and "P" genes. The higher proportion of "P" category genes could point to FE strains' association with hosts, as free iron availability is rather low within host bodies (78). The higher "L" proportion seems to be related to higher transposon copy numbers in FE strains, which could point to population bottlenecks relating to host-association (79). RE strains, on the other hand, showed increased proportions of COG category "W", "S", and "I" genes. The increase in the proportion of "W" genes is unlikely to have a large impact on

phenotypic differences between the ecotypes, since there are very few “W” genes in genomes of either RE or FE strains. It is difficult to comment on the increased proportion of “S” genes, since their function is unknown or poorly understood. The increased proportion of “I” category genes, relating to lipid transport and metabolism, could point to increased flexibility in membrane lipids for RE strains, which is a well-characterized adaptation to cold temperatures (80).

The three of four FE strains that we tested were able to colonize germ-free mice, whereas the RE strains could not be detected, indicating that growth at 37 ^{o}C may be necessary (although not sufficient) to colonize mammals. Opportunistic infections in mammals caused by *Psychrobacter* strains are limited to *P. sanguinis*, *P. phenylpyruvicus*, *P. faecalis*, and *P. pulmonis (25)*, which are all FE strains, but do not all belong to the basal clade. Our results suggest that the FE strains are maintaining an ancestral ability to grow at mammalian body temperatures and colonize mammalian host bodies, while RE strains are adapting towards psychrophilic or psychrotrophic lifestyles.

Psychrobacter’s sister taxon *Moraxella* in particular is commonly isolated from host mucosal tissues, and exhibits the reduced genome size and nutritional fastidiousness common to many host-dependent organisms. *Moraxella* contains species that are frequently associated with human respiratory infections, primarily *M. catarrhalis* (81), as well as livestock conjunctivitis, for example, *M. bovis* or *M. equis* (36). Since they are commonly found in healthy individuals and can cause disease in healthy individuals (82), *Moraxella* are best categorized as pathobionts and not dedicated or opportunistic pathogens. Several species of *Moraxella* appear basally in the P-clade of the *Moraxellaceae* family-level phylogeny, suggesting that *Psychrobacter* evolved from a “*Moraxella*-like” ancestor. This is supported by the fact that both phylogenetically basal and derived *Psychrobacter* strains carry genes related to virulence functions, and that many of the basal *Psychrobacter* strains exhibit growth defects in liquid culture, similar to the fastidiousness of *Moraxella*.

Despite clear phenotypic differentiation, *Psychrobacter* and *Moraxella* have similar genomic content, although *Psychrobacter* genomes are larger. A psychrophile emerging

from an apparently mesophilic background through widespread horizontal gene transfer has been suggested before in the genus *Psychroflexus* [96], though the study was limited to comparing only two genomes, which makes drawing conclusions about evolution difficult. In fact, it has been suggested before that this is *Psychrobacter's* evolutionary trajectory (27), and although many *Psychrobacter* trees are constructed using a *Moraxella* outgroup, *Psychrobacter's* potential pathobiont origin has not been widely discussed. Horizontal gene transfer would explain *Psychrobacter's* larger genome size compared to *Moraxella* despite lower coding density, as many newly acquired horizontally transferred genes are expected to be inactivated and pruned by the recipient genome (83, 84).

The history of the genus *Psychrobacter* is that of an ancestral pathobiont or pathogen, some of the descendants of which broadened their ecological distribution, resulting in an attenuation of pathogenicity. The emergence of a psychrotroph - a remarkable generalist - from a background of a more specialized pathobiont or pathogen showcases the adaptability of bacteria, and particularly *Proteobacteria*, to their environments.

Materials and Methods

***Moraxellaceae* family genomics and phenotypic data.** The *Moraxallaceae* family includes 3 well-characterized genera: *Moraxella*, *Acinetobacter* and *Psychrobacter*. To build a family-level phylogeny, we downloaded genomes of 18 species of *Acinetobacter* and 18 species of *Moraxella* from the National Center for Biotechnology Information (NCBI) genome database (85) (June 2020) (Table S2). We also included 15 *Psychrobacter* genomes generated in this study (described below). We determined the phylogenetic relationship between genomes using the whole-genome marker gene analysis software PhyloPhlAn (86) (v2.0, diversity = "medium", set to "accurate"), and determined genome quality and summary characteristics using checkM (87) (v1.0.18) and Prokka (88) (v1.14.6, kingdom = "bacteria"). The *Moraxellaceae* phylogenetic tree was visualized and annotated using the interactive Tree of Life (iTOL) web interface (89) (v5.6.1).

We analyzed the *Moraxellaceae* pan-genome using the PanX pipeline (90) (v1.6.0, core gene cutoff = 0.9). The input genomes used by PanX were initially annotated using Prokka. After their assignment into orthologous clusters using MCL (v14.137), we re-annotated gene clusters using eggNOG-Mapper (91) (v1.0.3). We explored genome content by calculating a distance matrix using the Jaccard metric through the R package ecodist (92) (v0.3.0) from a binary gene presence-absence table, followed by dimensional reduction of that distance matrix through principle coordinate decomposition (PCoA) with the cmdscale function from the R package stats. We investigated variables contributing to the separation of the PCoA using the envfit function from the R package vegan (93) (v2.5-6); we tested whether genes significantly contributing to separation were “core” (genes present in 90% to 100% of strains), “shell” (genes present in greater than two strains, but in fewer than 90%), or “cloud” (genes present in only one strain), and if general gene function - summarized by Cluster of Orthologous Groups (COG) category (94) - contributed to the separation. We collected growth temperature range data from type strain publications (16, 17, 20, 21, 26, 28–68). We used the R package phytools (95) (v0.7-47) to map the temperature ranges onto the phylogeny.

***Psychrobacter* strains.** We obtained 92 isolates of *Psychrobacter* from strain catalogues for phenotypic and genotypic characterization. These represent 38 validly published species of *Psychrobacter* as well as unclassified strains, all isolated from a wide variety of geographical locations and diverse environmental and host samples. All strains were purchased and maintained in compliance with the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity. For a full list of accessions and their catalogue, isolation, and cultivation information, see Table S3. Unless mentioned, we grew accessions as recommended by the strain catalogue (medium and temperature) from which they were purchased.

***Psychrobacter* carbon utilization assay.** To inform the design of the minimal medium used in the phenotypic screen described below, we chose a subset of 19 *Psychrobacter* strains (*P. adeliensis, P. aestuarii, P. aquaticus, P. arenosus, P. celer, P. cibarius JG-219, P. ciconiae, P. cryohalolentis, P. faecalis PBFP-1, P. fozii, P. fulvigenes, P. glacincola ACAM483, P. luti, P. lutiphocae, P. maritimus Pi2-25, P. okhotskensis MD17, P. pacificensis, P. piscatorii, P. urativorans ACAM534*) to evaluate for their ability to grow on 190 different substrates as their sole carbon source. We utilized Biolog plates PM1 and PM2 (Biolog Incorporated, Hayward, CA, USA) following a slightly modified protocol.

Briefly, we streaked out strains on agar plates of their preferred medium incubated at their preferred temperature (Table S3). We scraped cells from these plates and resuspended them at a final optical density at 600 nm (OD_{600}) of 0.07 in the Biolog inoculation fluid (PM IF-0a GN/GP 1.2x) diluted to 1x with sterile water. We inoculated 100 uL of cell suspension into each well of the PM1 and PM2 plates, mixing well to resuspend the carbon sources. All strains were grown under aerobic conditions at their preferred growth temperature (Table S3). Growth was monitored by measuring OD_{600} for 14 days. We calculated total change in absorbance by subtracting the blank wells from the substrate wells, then taking the difference between the absorbances at t = 14 days and t = 0 (inoculation time point). We then averaged the total change in absorbance for all strains for every carbon source, and assigned each carbon source to a “family” of compounds.

All strains reached a significantly higher max OD_{600} in amino acid substrates compared to other carbon sources. L-glutamate was chosen as the carbon source for the defined medium described in our phenotypic screens below, as none of the strains tested failed to grow using it. Some of the other amino acids allowed strains to grow to a larger change in OD_{600}, but failed to allow all strains to grow.

***Psychrobacter* phenotypic screen.** For the 85 *Psychrobacter* accessions that passed genome quality control (see Table S4), we tested their ability to grow under 24 different

conditions; a growth condition being a combination of a medium (complex or defined), salt concentration (0, 2.5, 5 or 10% NaCl) and incubation temperature (4, 25 or 37 ^{o}C).

We used Lysogeny broth (LB) as a complex medium with high nutrient availability, and a variation of M9 minimal medium (MM) as a defined media with lower nutrient availability. We prepared LB following manufacturer instructions except by varying concentrations of sodium chloride (NaCl) (see below), and autoclaving it at 121 °C for 20 minutes to ensure sterility. We prepared MM by adding components in the following final concentrations, followed by filter sterilization: 33.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 9.35 mM NH_4Cl, 0.4% L-glutamic acid, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1x ATCC trace vitamins solution, and 1x ATCC trace minerals solution, pH 7 (adjusted with KOH). We added NaCl to each 'base medium' to 1 of the 4 following concentrations of NaCl - low, with 0%, medium, with 2.5%, medium-high with 5%, and high with 10%. We grew cultures at 4 ^{o}C, representing cold environments, 25 ^{o}C representing mesophilic environments, and 37 ^{o}C representing mammalian host body temperature.

We randomly assigned *Psychrobacter* accessions to blocks of 10 strains to be tested simultaneously (several accessions were included in multiple blocks, see Table S2). We first grew strains to saturation, washed, diluted to optical density at 600 nm (OD_{600}) = 0.3 in sterile phosphate buffered saline (PBS), and inoculated in 100 µL of medium with a final ratio of 1:1000. We used 96-well plates, with 10 inoculum in 5 replicates and 10 uninoculated media wells per plate, all periphery wells were filled with water to reduce edge effects. Plates were incubated at each temperature to reflect all 24 growth conditions. We measured the OD_{600} (Spark® plate reader, Tecan, Zürich, Switzerland) every 8 hours during the first week of incubation, then every 24 hours, and let cultures grow until stationary phase (3 to 12 weeks).

We used the same culture dilutions from the growth assays as inocula for bile resistance assays.We spotted 10 uL of diluted culture onto both tryptic soy agar (TSA) as well as TSA amended with 0.5% w/v bile salt (Difco™ Bile Salts No. 3 from BD Biosciences, San Jose, CA, USA) and incubated the plates at the strains' preferred temperatures (Table

S2). We incubated the plates for 2 weeks, then scored strains for growth on bile compared to the no-bile control as either “sensitive” if no growth, or “resistant” for growth similar to the control. For several strains, no growth was observed on the control plate, and these were marked as “ND” or “no data.”

P. phenylpyruvicus and *P. sanguinis* exhibited the inability to grow in liquid culture under the conditions tested. Hence, we streaked 5 strains of these 2 species first on Columbian Blood Agar which we incubated for 3 days at 37 ^{o}C, then we washed and diluted them as for the liquid cultures. We spotted dilutions onto LB or MM media with 3% agar, supplemented with the tested salt concentrations as above, and incubated at 4 ^{o}C, 25 ^{o}C, and 37 ^{o}C. *P. sanguinis* strain 1501 was unable to grow on any of the base media, so we repeated the same process with media supplemented with 0.1% Tween80. We scored growth for all plates after 2 weeks.

For 2 accessions, we could not confirm the purity of the cultures used in the phenotypic screen, and subsequently removed them from analysis. They were removed from subsequent genomic analysis as well.

Growth probabilities. We scored each replicate as either ‘growth positive’, meaning the accession grew, or else, ‘growth negative,’ if the replicate never reached a maximum OD_{600} of 0.15 over the course of the experiment. For each strain and condition (medium, salt and temperature), we calculated a growth probability, which corresponds to the median value of growth positivity/negativity of all replicates for that condition.

For the type strains included in the phenotypic screen, we cross-checked growth probabilities at 4 ^{o}C and 37 ^{o}C with the published type strain descriptions where the original publications were clear about conditions tested. For the majority of strains, the type strain data and our data are in agreement; however, there are discrepancies between the type strain publications and our own data for the following strains: *P. ciconiae*, *P. fozii*, *P. marincola*, *P. namhaensis*, and *P. submarinus*. In some cases, we observed growth under temperatures where growth was previously unobserved; *P. ciconiae* and *P. marincola* have

not before been reported to grow at 4 ºC, and *P. fozii* has not been reported to grow at 37 ºC. In other cases, we observed no growth in temperatures that have previously been reported to support growth; *P. submarinus* is expected to grow at 4 ºC and *P. namhaensis* is expected to grow at 37 ºC. The conditions tested in the type strain publications often differ from the conditions used in our phenotypic screen.

Genome sequencing, assembly and annotation. We extracted genomic DNA from cultures grown in their preferred conditions using the Gentra Puregene Tissue Kit (Qiagen, Valencia, CA, USA). We initially sequenced samples using the MiSeq 2x250 bp and the HiSeq 2x150 bp paired-end read technology (Illumina, San Diego, CA, USA) as previously described (96). We constructed libraries using the Nextera DNA Sample Preparation Kit (Illumina) with modifications: we sheared 1 ng of DNA with in-house-generated Tn5 transposase, then amplified and barcoded it with custom primers for 7 to 14 cycles. We pooled samples and size-selected using magnetic beads for MiSeq libraries or BluePippin (Sage Science, Beverly, MA, USA) for HiSeq libraries. After dilution to 4 nM for MiSeq and 2.5 nM for HiSeq, we stored libraries at -20 °C until sequencing.

After sequencing and demultiplexing, we validated raw reads with fqtools (v2.0), and de-duplicated them with Clumpify (v37.78, dedupe = t, dupedist = 40 for HiSeq/2500 for MiSeq, optical = t). bbduk (v37.78) and Skewer (v0.2.2) were used to remove sequencing adapters and filter reads (minimum read length = 100 bp, minimum PHRED quality score = 25). At multiple steps throughout, we used fastqc (v0.11.7) and MultiQC (v1.7) to monitor quality. After quality control, reads were assembled *de novo*. First, we subsampled reads using seqtk (v1.3, number of sub-sampled reads per sample = 1000000), normalized by bbnorm (v37.78, target = 50, k = 31, minkmers = 15, prefilter = t), and then assembled with SPAdes (v3.12.8, cov_cutoff = off, set to “careful”, minimum scaffold length = 500 bp), followed by refinement with Pilon (v1.22, chunksize = 10000000). Finally, we assessed the assemblies for quality using CheckM. We assigned taxonomy using Sourmash (v2.0.0a4,

scaled = 10000, k = 31) and GTDB-Tk (v1.0.2, min_perc_aa = 10), and annotated assemblies using Prokka.

For genomes that were particularly fragmented (having greater than 250 contigs), we performed additional long-read sequencing (Table S4). We constructed Oxford Nanopore libraries using the Ligation Sequencing and Native Barcode Ligation kits (Oxford Nanopore, Oxford, UK). We sequenced the libraries using the MinION® system run with software MinKNOW (Oxford Nanopore). We basecalled and demultiplexed reads using ont-Guppy (v3.2.4), and used Porechop (97) (v0.2.4, adapter threshold - 90, minimum PHRED quality score = 8, minimum read length = 500 bp) to ensure that the adapters were removed, along with poor quality reads. We generated summary statistics, summary plots, and removed lambda phage reads using NanoPack (98) (nanocomp v1.11.3, nanofilt v2.7.1, nanoget v1.14.0, nanolyse v1.1.3, nanomath v0.23.3, nanoplot v1.31.0, nanostat v1.2.1). After quality control, we combined the long reads with the short (generated by the MiSeq and HiSeq libraries described above) for hybrid assembly. We assembled and analyzed the hybrid assemblies in largely the same way as the short-read assemblies described above, however replacing SPAdes with Unicycler (99) (v0.4.8, minimum contig length = 500 bp, mode = normal).

We followed the quality control cutoffs suggested by CheckM, and removed two genomes for having contamination higher than 5%, and one for having completion less than 90%. We removed two additional genomes as the taxonomic classification was outside of the *Psychrobacter* genus. These accessions were removed from genomic and phenotypic analysis. All further analyses use the 85 accessions which passed all QC measures. For a full description of sequencing for each accession, see Table S4.

We annotated genomes with Prokka and eggNOG mapper. A phylogeny of the accessions was generated using PhyloPhlAn, with *Moraxella lincolnii* as an outgroup. Again the phylogeny was visualized and annotated using iTOL. We used PanX to analyze the *Psychrobacter* pan-genome and R to explore gene presence-absence data as described above with the Moraxellaceae family. Pseudogenes were predicted using the DFAST core

workflow (100). We used treeWAS (72) to perform a pan-genome wide association study relating data collected in the phenotypic screen to genomic data.

Microbiome diversity of polar bear feces. We collected 86 polar bear fecal samples from several regions in Canada, including samples from 10 captive bears, fed varying diets or fasted, and 76 samples from an unknown number of wild bears, feeding on unknown diets. The feces from the captive bears were forwarded from institutions within Canada and did not require permitting for their passage to Queen's University. The captive samples comprised: 5 fecal samples from a single bear sequentially fed varying diets of Arctic char, harp seal and a "zoo diet" at the Polar Bear Habitat in Cochrane during 2010; 2 samples from each of 2 bears held at the Metro Toronto Zoo, fed a consistent "zoo diet" during 2010; and a single sample from a bear held at the Churchill Polar Bear Holding Facility in Churchill during 2010, where the bears are given only water until release. We collected all wild bear feces from M'Clintock Channel and Hudson Strait in Nunavut in accordance with permits prior to shipping to Queen's University: we collected 24 samples from M'Clintock Channel under *Wildlife Research* permits issued in 2007, 2008, 2009, 2010 and 2011 to Peter Van Coeverden de Groot, and 9 samples from Hudson Straight collected in 2011 under *Wildlife Research* permit to Grant Gilchrest (Environment Canada). We collected 43 wild samples from the Wapusk National Park in Manitoba from 2007-2010 under a Canada Parks permit to Robert Rockwell (American Museum of Natural History).

We confirmed that samples collected from wild bears originated from bears by sequencing a cytochrome b gene fragment using metabarcoding approaches based on Ion Torrent (Ion Torrent Systems Inc., Gilford, NH, USA) and 454 pyrosequencing (454 Life Sciences, Branford, CT, USA) next-generation sequencing technologies (101). We also used this method to analyze what prey animals the polar bears had been feeding on at the time of sample deposition. We visually inspected the samples for confirmation on the dietary analysis. We extracted DNA from the fecal samples using the DNeasy Blood and Tissue kit (Qiagen), and characterized the gut bacterial community by amplification and sequencing of

the V4 region of the 16S rRNA gene as described previously (102). We used QIIME (103) (v2, DADA2 for quality control, F read trimmed to 200 bp, R read trimmed to 110 bp, samples rarefied to 20000 sequences) for sequence processing and Silva (138 SSURef NR99 515F/806R) to assign taxonomy.

Isolation of *Psychrobacter* sp. from polar bear feces. 2 wild polar bear fecal samples were combined and diluted to 1 mg/mL in PBS. We plated the solution onto LB agar + 6% NaCl and incubated it at 14°C for 10 days. A single colony grew and was identified as *Psychrobacter faecalis* by colony PCR and Sanger sequencing of the full length 16S rRNA gene (described below). This isolate is designated as *P. faecalis PBFP-1* and is included in the phenotypic screen described above and the genomic analysis.

Gnotobiotic mouse colonizations. We selected 8 accessions of *Psychrobacter* - *P. cibarius* JG-220*, P. ciconiae, P. faecalis* PBFP-1*, P. immobilis* S3*, P. lutiphocae, P. okhotskensis* MD17*, P. namhaensis,* and *P. pacificensis* - as inocula for gnotobiotic mouse studies based on their derivation sources and phylogenetic breadth. We grew each accession in its preferred conditions (Table S3) to saturation, spun at 10 °C at 2500 rpm for 25 minutes, washed with sterile PBS, resuspended in 15% (v/v) glycerol in PBS, and flash frozen in liquid N_2. Inoculum samples were stored at -80 ºC until the mouse experiments were performed by animal caretakers as follows. For *P. ciconiae*, *P. faecalis*, *P. namhaensis*, and *P. pacificensis*, experiments were performed by Taconic Biosciences (Rensselaer, NY, USA) staff, while for *P. cibarius*, *P. immobilis*, and *P. okhotskensis*, experiments were performed by Max Planck Institute for Developmental Biology (MPIDB) animal caretakers. *P. lutiphocae* was included in both Taconic and MPIDB experiments.

5- to 6-week old germfree male C57BL/6J mice were orally inoculated with approximately 10^7 cfu (n=4 per *Psychrobacter* accession, n = 8 for *P. lutiphocae*). Mice were co-housed (Taconic) or single-housed (MPIDB) in sterile cages (IsoCage P, Tecniplast) and provided autoclaved water and sterile chow (NIH31M) ad libitum. 3 weeks post-colonization,

mice were sacrificed via CO_2 asphyxiation (Taconic) or CO_2 asphyxiation followed by cervical dislocation (MPIDB), after which cecal contents were immediately collected, flash frozen, and stored at -80 °C prior to use. The Taconic experiments were performed in compliance with Taconic's IACUC, and the MPIDB experiments were approved by and performed in accordance with the local animal welfare authority's legal requirements.

To determine the bacterial colonization density in the mouse gastrointestinal tract, we serially diluted 2 aliquots per mouse of cecal material (50 mg each), incubated them on plates under their preferred conditions (Table S3) for 3 to 5 days. If no colonies were observed, we spread an inoculating loop of the undiluted aliquot onto Brain-Heart Infusion Agar and incubated at 37 °C for 2 weeks. For samples where we again observed no colonies, we categorized these *Psychrobacter* accessions as "non-persistent." For the samples that did show colony growth, we confirmed the identity of the colonies as *Psychrobacter* using Sanger sequencing of the full length 16S rRNA gene as described below.

16S rRNA gene Sanger sequencing. We prepared "colony" polymerase chain reactions following the protocol for Phusion High Fidelity Polymerase (New England Biolabs, Ipswitch, MA, USA), swirling a colony of interest in the reaction mixture as a substitute for the DNA template, and the 27F and 1391R universal full length 16S rRNA gene primers (104) for amplification. The reactions were incubated on the Mastercycler pro S thermocycler (Eppendorf, Hamburg, Germany) following the Phusion protocol, using a touchdown program for the annealing temperature (dropping the annealing temperature at a rate of 1 °C per cycle from 70 °C to 55 °C, then annealing at 55 °C for 15 cycles). We cleaned the products using the DNA Clean & Concentrator -25 kit (Zymo Research, Irvine, CA, USA) and checked the concentration using the DS 11+ Spectrophotometer (DeNovix, Wilmington, DE, USA). We used the cleaned DNA products as the templates for the Sanger sequencing reaction, following the ABI PRISM BigDye Terminator Cycle Sequencing Kit (ThermoFisher Scientific, Waltham, MA, USA) protocol with the 27F primer. The sequencing was performed

using the 3730xl DNA Analyzer (ThermoFisher). Upon sequencing completion, we BLASTed (105) the sample sequences against the NCBI non-redundant nucleotide database. We examined the top ten hits to confirm the sample identity.

Statistical analysis. We performed all data processing and statistical analysis using R (106) (v3.6.2) or Python (v3.6.10). We compared means between groups using a Kruskal-Wallis test followed by a pairwise Wilcoxon rank sum when more than two groups were compared. We tested differences in frequencies between groups with more than 10 observations with a χ^2-test, repeated a 100 times with down-sampling in order to correct for sampling sizes between groups. We measured phylogenetic signal using a log-likelihood ratio test on Pagel's λ (Pagel's λ fitted using the phylosig function from the phytools R package [(95)] [v0.7-47], and the null hypothesis being $\lambda = 0$). When applicable, we tested for the confounding of phylogeny with our groups of interest using the aov.phylo function from the R package geiger (107) (v2.0.7). We adjusted all p-values for multiple comparisons with the Benjamini-Hochberg (BH) correction method.

Data availability. Raw sequences for the *Psychrobacter* genome sequencing and polar bear feces 16S rRNA gene sequencing, as well as assembled *Psychrobacter* genomes, are available in the European Nucleotide Archive under the accession PRJEB40380. Annotated *Psychrobacter* genomes are available at ftp://ftp.tue.mpg.de/pub/ebio/dwelter. Raw data, R notebooks, and Python scripts for the analyses are available at https://github.com/dkwelter/Welter_et_al_2020.

Acknowledgments. This work was supported by the Max Planck Society. We thank Jacobo de la Cuesta-Zuluaga, Sara Di Rienzi, Hagay Enav, Angela Poole, Jessica Sutter, Taichi Suzuki, and William Walters for discussions regarding project design and analysis, and Andrea Belkacemi, Ilja Bezrukov, Pablo Carbonell, Silke Dauser, Julia Hildebrandt, and Christa Lanz for their advice on and assistance with sequencing. We also thank Jürgen

Berger and Katharina Hipp for performing electron microscopy, and Ariadne Papatheodorou for her assistance with the bile sensitivity assays. We thank Markus Dyck and Patricia Morin for providing the Polar Bear Habitat samples, Maria Frank for samples from the Metro Toronto Zoo, Daryll Hedman and Manitoba Conservation for the sample from the Polar Bear Holding Facility in Churchill, and Sam Iverson for samples from Hudson Straight. The collection of Nunavut samples would not have been possible without collaboration of colleagues at the Gjoa Haven Hunters and Trappers Association and their Traditional Ecological Knowledge relating to polar bears. The Nunavut field work was supported by funds from the Nunavut Wildlife Management Board (NWMB), the Nunavut General Monitoring Plan (NGMP), the National Science and Engineering Research Council (NSERC), and Environment Canada (Gov. of Canada). We would also like to thank Marie Pages and Maxime Galan for assistance with cytochrome b barcode sequencing.

Competing interest.

We have no competing interests to declare.

References

1. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6:776–788.

2. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108 Suppl 1:4516–4522.

3. Youngblut ND, de la Cuesta-Zuluaga J, Reischer GH, Dauser S, Schuster N, Walzer C, Stalder G, Farnleitner AH, Ley RE. 2020. Large scale metagenome assembly reveals novel animal-associated microbial genomes, biosynthetic gene clusters, and other genetic diversity. mSystems 5:e01045-20.

4. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-bacterial mutualism in the human intestine. Science 307:1915–1920.

5. Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124:837–848.

6. Salyers AA, Vercellotti JR, West SE, Wilkins TD. 1977. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. Appl Environ Microbiol 33:319–322.

7. Ruiz L, Margolles A, Sánchez B. 2013. Bile resistance mechanisms in Lactobacillus and Bifidobacterium. Front Microbiol 4:396.

8. Browne HP, Neville BA, Forster SC, Lawley TD. 2017. Transmission of the gut microbiota: spreading of health. Nat Rev Microbiol 15:531–543.

9. Sakib SN, Reddi G, Almagro-Moreno S. 2018. Environmental role of pathogenic traits in Vibrio cholerae. J Bacteriol 200:e00795–17.

10. Wren BW. 2003. The yersiniae--a model genus to study the rapid evolution of bacterial pathogens. Nat Rev Microbiol 1:55–64.

11. Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. Proc Natl Acad Sci U S A 108 Suppl 2:10800–10807.

12. Apprill A, Miller CA, Moore MJ, Durban JW, Fearnbach H, Barrett-Lennard LG. 2017. Extensive core microbiome in drone-captured whale blow supports a framework for health monitoring. mSystems 2.

13. Apprill A, Robbins J, Eren AM, Pack AA, Reveillaud J, Mattila D, Moore M, Niemeyer M, Moore KMT, Mincer TJ. 2014. Humpback whale populations share a core skin bacterial community: towards a health index for marine mammals? PLoS One 9:e90785.

14. Kudo T, Kidera A, Kida M, Kawauchi A, Shimizu R, Nakahara T, Zhang X, Yamada A, Amano M, Hamada Y, Taniyama S, Arakawa O, Yoshida A, Oshima K, Suda W, Kuwahara H, Nogi Y, Kitamura K, Yuki M, Iida T, Moriya S, Inoue T, Hongoh Y, Hattori M, Ohkuma M. 2014. Draft genome sequences of Psychrobacter strains JCM 18900, JCM 18901, JCM 18902, and JCM 18903, isolated preferentially from frozen aquatic organisms. Genome Announc 2.

15. Banks JC, Craig Cary S, Hogg ID. 2014. Isolated faecal bacterial communities found for Weddell seals, Leptonychotes weddellii, at White Island, McMurdo Sound, Antarctica. Polar Biol 37:1857–1864.

16. Yassin AF, Busse H-J. 2009. Psychrobacter lutiphocae sp. nov., isolated from the faeces of a seal. Int J Syst Evol Microbiol 59:2049–2053.

17. Kämpfer P, Jerzak L, Wilharm G, Golke J, Busse H-J, Glaeser SP. 2015. Psychrobacter ciconiae sp. nov., isolated from white storks (Ciconia ciconia). Int J Syst Evol Microbiol 65:772–777.

18. Kämpfer P, Glaeser SP, Irgang R, Fernández-Negrete G, Poblete-Morales M, Fuentes-Messina D, Cortez-San Martín M, Avendaño-Herrera R. 2020. Psychrobacter pygoscelis sp. nov. isolated from the penguin Pygoscelis papua. Int J Syst Evol Microbiol 70:211–219.

19. Svanevik CS, Lunestad BT. 2011. Characterisation of the microbiota of Atlantic mackerel (Scomber scombrus). Int J Food Microbiol 151:164–170.

20. Yoon J-H, Lee C-H, Kang S-J, Oh T-K. 2005. Psychrobacter celer sp. nov., isolated from seawater of the South Sea in Korea. Int J Syst Evol Microbiol 55:1885–1890.

21. Bowman JP, Nichols DS, McMeekin TA. 1997. Psychrobacter glacincola sp. nov., a halotolerant, psychrophilic bacterium isolated from Antarctic sea ice. Syst Appl Microbiol 20:209–215.

22. Matsuyama H, Minami H, Sakaki T, Kasahara H, Watanabe A, Onoda T, Hirota K, Yumoto I. 2015. Psychrobacter oceani sp. nov., isolated from marine sediment. Int J Syst Evol Microbiol 65:1450–1455.

23. Zeng Y-X, Yu Y, Liu Y, Li H-R. 2016. Psychrobacter glaciei sp. nov., isolated from the ice core of an Arctic glacier. Int J Syst Evol Microbiol 66:1792–1798.

24. Bakermans C, Ayala-del-Río HL, Ponder MA, Vishnivetskaya T, Gilichinsky D, Thomashow MF, Tiedje JM. 2006. Psychrobacter cryohalolentis sp. nov. and Psychrobacter arcticus sp. nov., isolated from Siberian permafrost. Int J Syst Evol Microbiol 56:1285–1291.

25. Deschaght P, Janssens M, Vaneechoutte M, Wauters G. 2012. Psychrobacter isolates of human origin, other than Psychrobacter phenylpyruvicus, are predominantly Psychrobacter faecalis and Psychrobacter pulmonis, with emended description of P. faecalis. Int J Syst Evol Microbiol 62:671–674.

26. Wirth SE, Ayala-del-Río HL, Cole JA, Kohlerschmidt DJ, Musser KA, Sepúlveda-Torres L del C, Thompson LM, Wolfgang WJ. 2012. Psychrobacter sanguinis sp. nov., recovered from four clinical specimens over a 4-year period. Int J Syst Evol Microbiol 62:49–54.

27. Bakermans C. 2018. Adaptations to marine versus terrestrial low temperature environments as revealed by comparative genomic analyses of the genus Psychrobacter. FEMS Microbiol Ecol 94.

28. Vela AI, Sánchez-Porro C, Aragón V, Olvera A, Domínguez L, Ventosa A, Fernández-Garayzábal JF. 2010. Moraxella porci sp. nov., isolated from pigs. Int J Syst Evol Microbiol 60:2446–2450.

29. Vela AI, Arroyo E, Aragón V, Sánchez-Porro C, Latre MV, Cerdà-Cuéllar M, Ventosa A, Domínguez L, Fernández-Garayzábal JF. 2009. Moraxella pluranimalium sp. nov., isolated from animal specimens. Int J Syst Evol Microbiol 59:671–674.

30. Lindqvist K. March-April 1960. A Neisseria species associated with infections keratoconjunctivitis of sheep Neisseria ovis nov. spec. The Journal of Infections Diseases 106:162–165.

31. Xie C-H, Yokota A. 2005. Transfer of the misnamed [Alysiella] sp. IAM 14971 (=ATCC 29468) to the genus Moraxella as Moraxella oblonga sp. nov. Int J Syst Evol Microbiol 55:331–334.

32. Boevre K, Henriksen SD. 1967. A new Moraxella species, Moraxella osloensis, and a revised description of Moraxella nonliquefaciens. International Journal of Systematic and Evolutionary Microbiology 17.

33. Catlin BW. 1970. Transfer of the organism named Neisseria Catarrhalis to Branhamella gen. nov. Int J Syst Bacteriol 20:155–159.

34. Eyre JW. 1900. A clinical and bacteriological study of diplo-bacillary conjunctivitis. Journal of Pathology and Bacteriology 6:1–13.

35. Vandamme P, Gillis M, Vancanneyt I. M, Hoste I. B, Kersters K, Falsen E. 1993. Moraxella lincolnii sp. nov., Isolated from the human respiratory tract, and reevaluation of the taxonomic position of Moraxella osloensis. International Journal of Systematic Bacteriology 43:474–481.

36. Hughes DE, Pugh GW Jr. 1970. Isolation and description of a Moraxella from horses with conjunctivitis. Am J Vet Res 31:457–462.

37. U. B. 1962. Über das Vorkommen von Neisserien bei einigen Tieren. Zeitschrift f Hygiene 148:445–457.

38. Pelczar MJ Jr. 1953. Neisseria caviae nov spec. J Bacteriol 65:744.

39. Kodjo A, Tønjum T, Richard Y, Bøvre K. 1995. Moraxella caprae sp. nov., a new member of the classical Moraxellae with very close affinity to Moraxella bovis. International Journal of Systematic Bacteriology 45:467–471.

40. Jannes G, Vaneechoutte M, Lannoo M, Gillis M, Vancanneyt M, Vandamme P, Verschraegen G, Van Heuverswyn H, Rossau R. 1993. Polyphasic taxonomy leading to the proposal of Moraxella canis sp. nov. for Moraxella catarrhalis-like strains. International Journal of Systematic Bacteriology 43:438–449.

41. Angelos JA, Spinks PQ, Ball LM, George LW. 2007. Moraxella bovoculi sp. nov., isolated from calves with infectious bovine keratoconjunctivitis. Int J Syst Evol Microbiol 57:789–795.

42. Henriksen SD. 1971. Designation of a neotype strain for Moraxella bovis (Hauduroy et al.) Murray. Int J Syst Evol Microbiol 21:28–28.

43. Kodjo A, Richard Y, Tønjum T. 1997. Moraxella boevrei sp. nov., a new Moraxella

species found in goats. International Journal of Systematic Bacteriology 47:115–121.

44. K. Boevre, J.E. Fuglesang, N. Hagen, E. Jantzen, and L.O. Froeholm. 1976. Moraxella atlantae sp. nov. and its distinction from Moraxella phenylpyrouvica. International Journal of Systematic Bacteriology 26:511–521.

45. Vaneechoutte M, Nemec A, Musílek M, van der Reijden TJK, van den Barselaar M, Tjernberg I, Calame W, Fani R, De Baere T, Dijkshoorn L. 2009. Description of Acinetobacter venetianus ex Di Cello et al. 1997 sp. nov. Int J Syst Evol Microbiol 59:1376–1381.

46. Nemec A, De Baere T, Tjernberg I, Vaneechoutte M, van der Reijden TJ, Dijkshoorn L. 2001. Acinetobacter ursingii sp. nov. and Acinetobacter schindleri sp. nov., isolated from human clinical specimens. Int J Syst Evol Microbiol 51:1891–1899.

47. Nishimura Y, Ino T, Iizuka H. 1988. Acinetobacter radioresistens sp. nov. isolated from cotton and soil. Int J Syst Evol Microbiol 38:209–211.

48. Li Y, Piao C-G, Ma Y-C, He W, Wang H-M, Chang J-P, Guo L-M, Wang X-Z, Xie S-J, Guo M-W. 2013. Acinetobacter puyangensis sp. nov., isolated from the healthy and diseased part of Populus xeuramericana canker bark. Int J Syst Evol Microbiol 63:2963–2969.

49. Yang Liu, Qiuhua Rao, Jiefeng Tu, Jiaonan Zhang, Minmin Huang, Bing Hu, Qiu Lin and Tuyan Luo. 2018. Acinetobacter piscicola sp. nov., isolated from diseased farmed Murray cod (Maccullochella peelii peelii). Int J Syst Evol Microbiol 68:905–910.

50. Choi JY, Ko G, Jheong W, Huys G, Seifert H, Dijkshoorn L, Ko KS. 2013. Acinetobacter kookii sp. nov., isolated from soil. Int J Syst Evol Microbiol 63:4402–4406.

51. Malhotra J, Anand S, Jindal S, Rajagopal R, Lal R. 2012. Acinetobacter indicus sp. nov., isolated from a hexachlorocyclohexane dump site. Int J Syst Evol Microbiol

62:2883–2890.

52. Nemec A, Musílek M, Maixnerová M, De Baere T, van der Reijden TJK, Vaneechoutte M, Dijkshoorn L. 2009. Acinetobacter beijerinckii sp. nov. and Acinetobacter gyllenbergii sp. nov., haemolytic organisms isolated from humans. Int J Syst Evol Microbiol 59:118–124.

53. Smet A, Cools P, Krizova L, Maixnerova M, Sedo O, Haesebrouck F, Kempf M, Nemec A, Vaneechoutte M. 2014. Acinetobacter gandensis sp. nov. isolated from horse and cattle. Int J Syst Evol Microbiol 64:4007–4015.

54. Nemec A, Radolfova-Krizova L, Maixnerova M, Vrestiakova E, Jezek P, Sedo O. 2016. Taxonomy of haemolytic and/or proteolytic strains of the genus Acinetobacter with the proposal of Acinetobacter courvalinii sp. nov. (genomic species 14 sensu Bouvet & Jeanjean), Acinetobacter dispersus sp. nov. (genomic species 17), Acinetobacter modestus sp. nov., Acinetobacter proteolyticus sp. nov. and Acinetobacter vivianii sp. nov. Int J Syst Evol Microbiol 66:1673–1685.

55. Nemec A, Radolfova-Krizova L, Maixnerova M, Sedo O. 2017. Acinetobacter colistiniresistens sp. nov. (formerly genomic species 13 sensu Bouvet and Jeanjean and genomic species 14 sensu Tjernberg and Ursing), isolated from human infections and characterized by intrinsic resistance to polymyxins. Int J Syst Evol Microbiol 67:2134–2141.

56. Radolfova-Krizova L, Maixnerova M, Nemec A. 2016. Acinetobacter celticus sp. nov., a psychrotolerant species widespread in natural soil and water ecosystems. Int J Syst Evol Microbiol 66:5392–5398.

57. Álvarez-Pérez S, Lievens B, Jacquemyn H, Herrera CM. 2013. Acinetobacter nectaris sp. nov. and Acinetobacter boissieri sp. nov., isolated from floral nectar of wild Mediterranean insect-pollinated plants. Int J Syst Evol Microbiol 63:1532–1539.

58. Philippe Bouvet And. 1986. Taxonomy of the genus Acinetobacter with the recognition of Acinetobacter baumannii sp. nov. Acinetobacter haemolyticus sp. nov. Acinetobacter johnsonii sp. nov. and Acinetobacter junii sp. nov. and emended descriptions of Acinetobacter calcoaceticus and Acinetobacter lwoffii. International Journal of Systematic Bacteriology 36:228–240.

59. Baik KS, Park SC, Lim CH, Lee KH, Jeon DY, Kim CM, Seong CN. 2010. Psychrobacter aestuarii sp. nov., isolated from a tidal flat sediment. Int J Syst Evol Microbiol 60:1631–1636.

60. Bowman JP, Cavanagh J, Austin JJ, Sanderson K. 1996. Novel Psychrobacter species from Antarctic ornithogenic soils. Int J Syst Bacteriol 46:841–848.

61. Kämpfer P, Albrecht A, Buczolits S, Busse H-J. 2002. Psychrobacter faecalis sp. nov., a new species from a bioaerosol originating from pigeon faeces. Syst Appl Microbiol 25:31–36.

62. Romanenko LA, Lysenko AM, Rohde M, Mikhailov VV, Stackebrandt E. 2004. Psychrobacter maritimus sp. nov. and Psychrobacter arenosus sp. nov., isolated from coastal sea ice and sediments of the Sea of Japan. Int J Syst Evol Microbiol 54:1741–1745.

63. Yumoto I, Hirota K, Sogabe Y, Nodasaka Y, Yokota Y, Hoshino T. 2003. Psychrobacter okhotskensis sp. nov., a lipase-producing facultative psychrophile isolated from the coast of the Okhotsk Sea. Int J Syst Evol Microbiol 53:1985–1989.

64. Juni E, Heym GA. 1986. Psychrobacter immobilis gen. nov., sp. nov.: genospecies composed of Gram-negative, aerobic, oxidase-positive coccobacilli. Int J Syst Evol Microbiol 36:388–391.

65. Yoon J-H, Lee C-H, Yeo S-H, Oh T-K. 2005. Psychrobacter aquimaris sp. nov. and Psychrobacter namhaensis sp. nov., isolated from seawater of the South Sea in Korea.

Int J Syst Evol Microbiol 55:1007–1013.

66. Maruyama A, Honda D, Yamamoto H, Kitamura K, Higashihara T. 2000. Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species Psychrobacter pacificensis sp. nov. Int J Syst Evol Microbiol 50 Pt 2:835–846.

67. Yumoto I, Hirota K, Kimoto H, Nodasaka Y, Matsuyama H, Yoshimune K. 2010. Psychrobacter piscatorii sp. nov., a psychrotolerant bacterium exhibiting high catalase activity isolated from an oxidative environment. Int J Syst Evol Microbiol 60:205–208.

68. Poppel MT, Skiebe E, Laue M, Bergmann H, Ebersberger I, Garn T, Fruth A, Baumgardt S, Busse H-J, Wilharm G. 2016. Acinetobacter equi sp. nov., isolated from horse faeces. Int J Syst Evol Microbiol 66:881–888.

69. Dziewit L, Cegielski A, Romaniuk K, Uhrynowski W, Szych A, Niesiobedzki P, Zmuda-Baranowska MJ, Zdanowski MK, Bartosik D. 2013. Plasmid diversity in arctic strains of Psychrobacter spp. Extremophiles 17:433–444.

70. Bobay L-M, Ochman H. 2017. The evolution of bacterial genome architecture. Front Genet 8:72.

71. Kuo C-H, Ochman H. 2009. Deletional bias across the three domains of life. Genome Biol Evol 1:145–152.

72. Collins C, Didelot X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 14:e1005958.

73. Stone GN, Nee S, Felsenstein J. 2011. Controlling for non-independence in comparative analysis of patterns across populations within species. Philos Trans R Soc Lond B Biol Sci 366:1410–1424.

74. Subhadra B, Surendran S, Kim DH, Woo K, Oh MH, Choi CH. 2019. The transcription factor NemR is an electrophile-sensing regulator important for the detoxification of reactive electrophiles in Acinetobacter nosocomialis. Res Microbiol 170:123–130.

75. Ma Q, Yang Z, Pu M, Peti W, Wood TK. 2011. Engineering a novel c-di-GMP-binding protein for biofilm dispersal. Environ Microbiol 13:631–642.

76. Bodelón G, Palomino C, Fernández LÁ. 2013. Immunoglobulin domains in Escherichia coli and other enterobacteria: from pathogenesis to applications in antibody technologies. FEMS Microbiol Rev 37:204–250.

77. Lee B-H, Hébraud M, Bernardi T. 2017. Increased adhesion of Listeria monocytogenes strains to abiotic surfaces under cold stress. Front Microbiol 8:2221.

78. Bullen JJ, Rogers HJ, Griffiths E. 1978. Role of iron in bacterial infection, p. 1–35. *In* Arber, W, Henle, W, Hofschneider, PH, Humphrey, JH, Klein, J, Koldovský, P, Koprowski, H, Maaløe, O, Melchers, F, Rott, R, Schweiger, HG, Syruček, L, Vogt, PK (eds.), Current Topics in Microbiology and Immunology: Volume 80. Springer Berlin Heidelberg, Berlin, Heidelberg.

79. Hendry TA, Freed LL, Fader D, Fenolio D, Sutton TT, Lopez JV. 2018. Ongoing transposon-mediated genome reduction in the luminous bacterial symbionts of deep-sea ceratioid anglerfishes. MBio 9.

80. Collins T, Margesin R. 2019. Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. Appl Microbiol Biotechnol 103:2857–2871.

81. Perez Vidakovics ML, Riesbeck K. 2009. Virulence mechanisms of Moraxella in the pathogenesis of infection. Curr Opin Infect Dis 22:279–285.

82. 2015. Moraxella, p. 1–17. *In* Whitman, WB, Rainey, F, Kämpfer, P, Trujillo, M, Chun, J, DeVos, P, Hedlund, B, Dedysh, S (eds.), Bergey's Manual of Systematics of Archaea

and Bacteria. John Wiley & Sons, Ltd, Chichester, UK.

83. Feng S, Powell SM, Wilson R, Bowman JP. 2014. Extensive gene acquisition in the extremely psychrophilic bacterial species Psychroflexus torquis and the link to sea-ice ecosystem specialism. Genome Biol Evol 6:133–148.

84. Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol 5:R64.

85. NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8–D13.

86. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304.

87. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

88. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

89. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

90. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration. Nucleic Acids Res 46:e5.

91. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

92. Goslee SC, Urban DL, Others. 2007. The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw 22:1–19.

93. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930.

94. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36.

95. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223.

96. Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, Neher RA, Kemen E, Weigel D. 2018. Arabidopsis thaliana and Pseudomonas pathogens exhibit stable associations over evolutionary timescales. Cell Host Microbe 24:168–179.e4.

97. Wick R. Porechop. Github. https://github.com/rrwick/Porechop

98. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669.

99. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

100. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039.

101. Galan M, Pagès M, Cosson J-F. 2012. Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. PLoS One 7:e48374.

102. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M,

Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. Cell 159:789–799.

103. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336.

104. 16S ribosomal DNA | Lutzoni Lab. http://lutzonilab.org/16s-ribosomal-dna/

105. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

106. R Development Core Team. 2011. R: a language and environment for statistical computing. R Project for Statistical Computing.

107. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218.

Figures

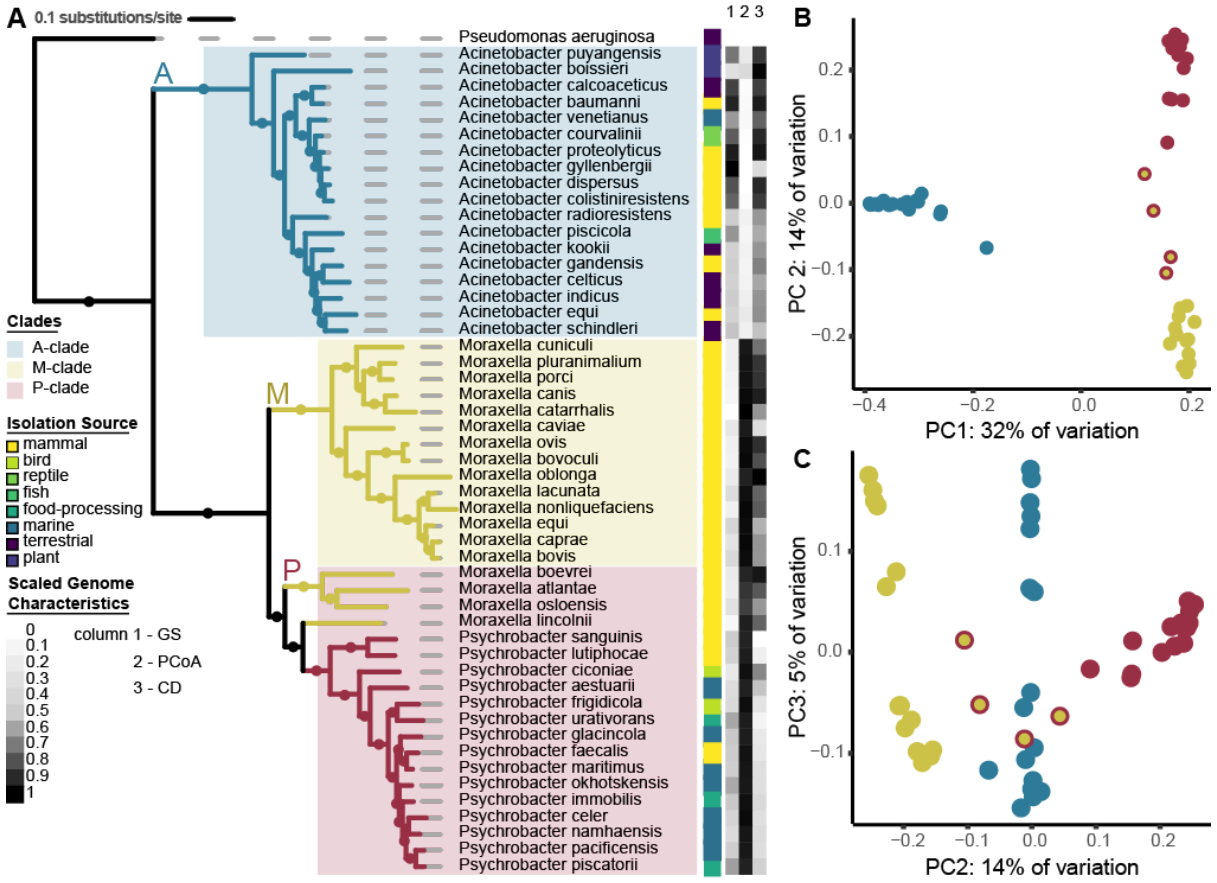

Figure 1. Genomic and phenotypic diversity in the family *Moraxellaceae*. A) The *Moraxellaceae* family phylogeny was constructed with 51 diverse *Moraxellaceae* genomes using the software PhyloPhlan, which constructs a phylogeny using fasttree with 1000 bootstraps, refined by RAxML under the PROTCATG model. Amino acid sequences from 400 marker genes were used in the alignment. Branches with bootstrap support greater than 70% are represented by filled circles. The scale bar represents the average amino acid substitutions per site. *Pseudomonas aeruginosa* was used as an outgroup. Clades are highlighted in colored blocks, and branches are colored by genus. Isolation source is depicted in a color strip, along with a heatmap of scaled notable genome characteristics that differ between the genera, with 0 representing the smallest value present and 1 the largest value (*P. aeruginosa* not included). GS = Genome size, ranging between 1.8 Mb and 4.5 Mb. PCoA = PC1 values from a PCoA based on gene presence/absence data. CD = genome

coding density, ranging from 80% to 89%. Growth temperature range data was collected from type strain publications. B) PC1 and 2 of a PCoA analysis of a binary matrix of gene presence/absence for 51 species of *Moraxellaceae*, explaining 32% and 14% of the variation, respectively. Each genome is represented by one point, colored by genus. The P-clade *Moraxella* spp. are represented by yellow points with red outlines. C) PC2 and 3, explaining 14% and 5% of the variation.

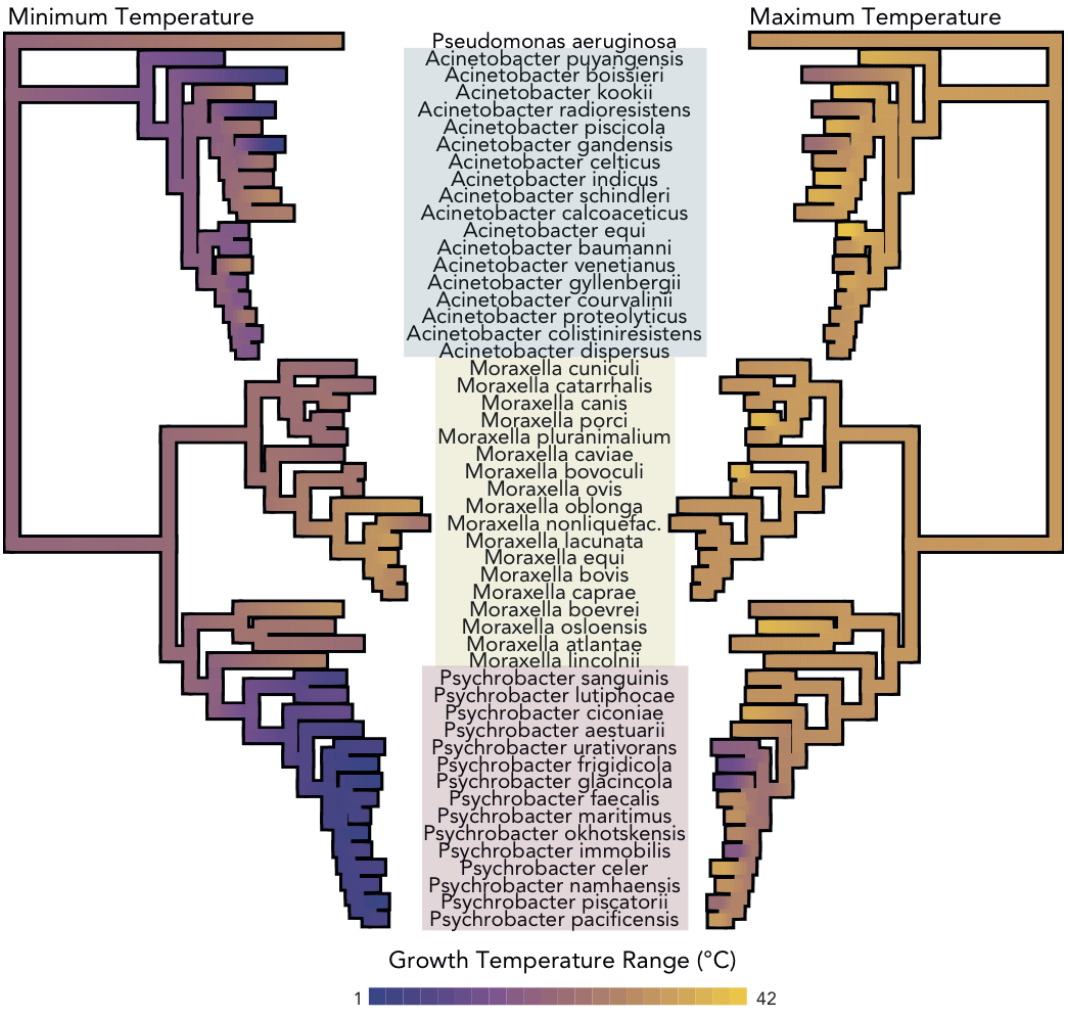

Figure 2. Psychrophily is only observed in *Psychrobacter* spp. among the family *Moraxellaceae*. Continuous trait mapping growth temperature ranges of 51 species from the *Moraxellaceae* family, taken from type strain publications. Values at nodes are imputed by maximum likelihood analysis. The phylogeny was constructed by marker-gene analysis including 400 genes, as in Fig. 1. Genera are indicated with colored boxes.

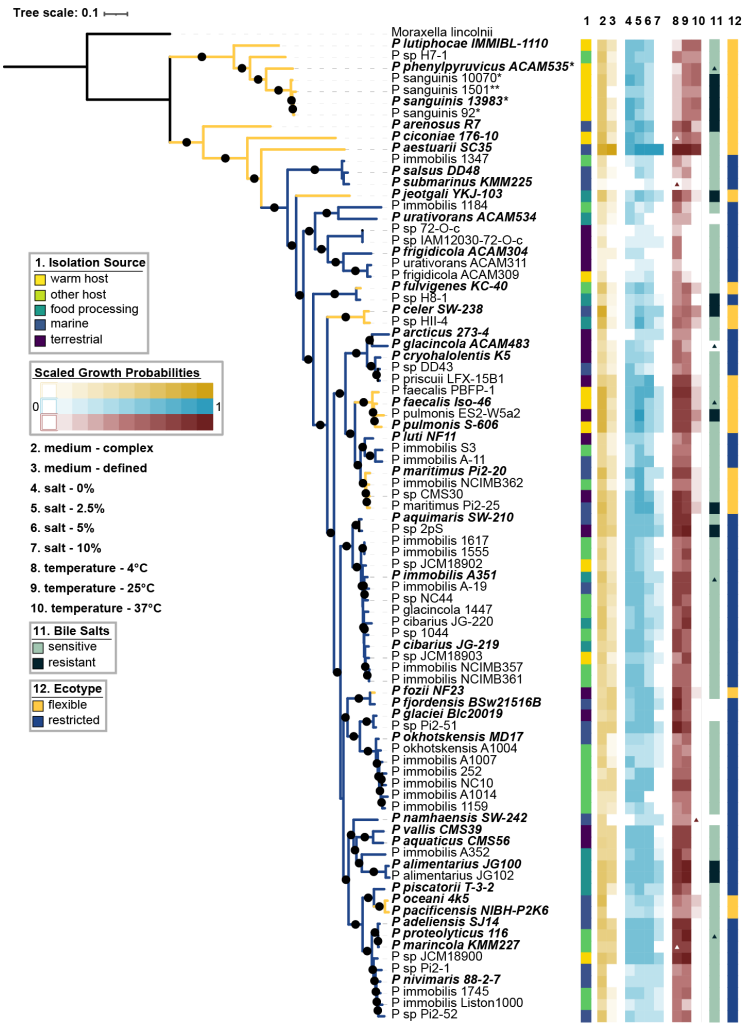

Figure 3. *Psychrobacter* phenotypic and phylogenomic diversity. The *Psychrobacter* phylogeny was constructed using fasttree with 1000 bootstraps, refined by RAxML under the PROTCATG model. Amino acid sequences from 400 marker genes were used in the alignment. Branches with bootstrap support greater than 70% are represented by filled circles. The scale bar represents the average amino acid substitutions per site. *M. lincolnii* was used as an outgroup. Type strain isolate names are indicated in bold and italicized type. Strains indicated with * next to their name exhibited growth defects in liquid media, and were tested on solid agar media instead. Strains indicated with ** exhibited growth defects on solid and liquid media, and were tested on solid media supplemented with 0.1% Tween80. Isolation source is depicted in column 1 as a color strip. Columns 2 – 10 represent the growth probabilities of each strain for each condition; media complexity is represented in yellow, salt concentration is represented in blue, and temperature is represented in red. Type

strain data supports our temperature data except where indicated - colored triangles show conditions in which we expected growth but did not observe it, while white triangles represent conditions in which we observed growth we did not expect. Bile salt resistance is shown in column 11, with black triangles indicating strains where we expected resistance but observed sensitivity, and green triangles representing the opposite. The “ecotype” is shown in column 12 in a colorstrip, and in the color of the branches.

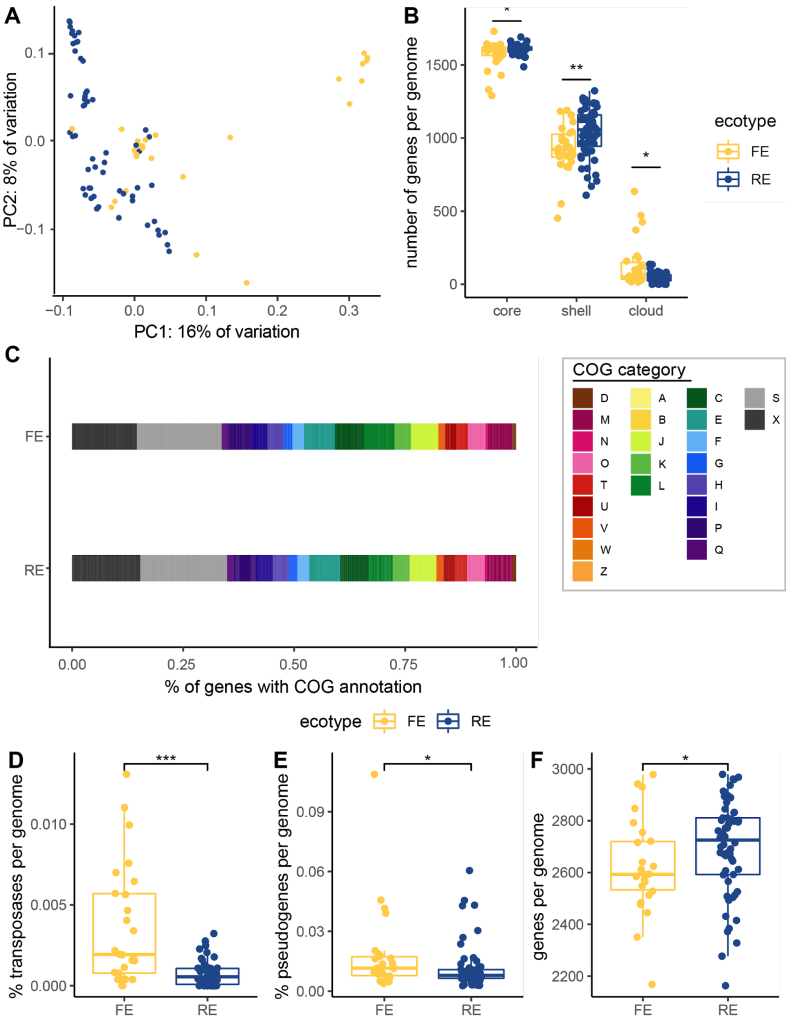

Figure 4. Structural differences between the genomes of FE and RE *Psychrobacter* strains. A) The first two axes of a PCoA of a gene presence-absence matrix of all of the included accessions, colored by ecotype. B) Pan-genome categories. C) Proportions of the average genome of each ecotype devoted to each cluster of orthologous genes (COG) category. Cellular processing and signaling: [D] cell cycle control cell division, chromosome partitioning; [M] cell wall/membrane/envelope biogenesis; [N] cell motility; [O] post-translational modification, protein turnover, and chaperones; [T] signal transduction mechanisms; [U] intracellular trafficking, secretion, and vesicular transport; [V] defense mechanisms; [W] extracellular structures; [Y] nuclear structure; [Z] cytoskeleton. Information storage and processing: [A] RNA processing and modification; [B] chromatin structure and dynamics; [J] translation, ribosomal structure and biogenesis; [K] transcription; [L] replication, recombination, and repair. Metabolism: [C] energy production and conversion; [E]

amino acid transport and metabolism; [F] nucleotide transport and metabolism; [G] carbohydrate transport and metabolism; [H] coenzyme transport and metabolism; [I] lipid transport and metabolism; [P] inorganic ion transport and metabolism; [Q] secondary metabolites biosynthesis, transport, and catabolism. Poorly Characterized: [S] function unknown, [X] not in COG database. D) Percentage of genes per genome that are transposases. E) Percentage of predicted pseudogenes per genome. F) Genes per genome. Means were compared using the Wilcoxon rank sum test. * indicates a p-value less than 0.05, ** indicates $p < 0.005$, *** indicates $p < 0.0005$

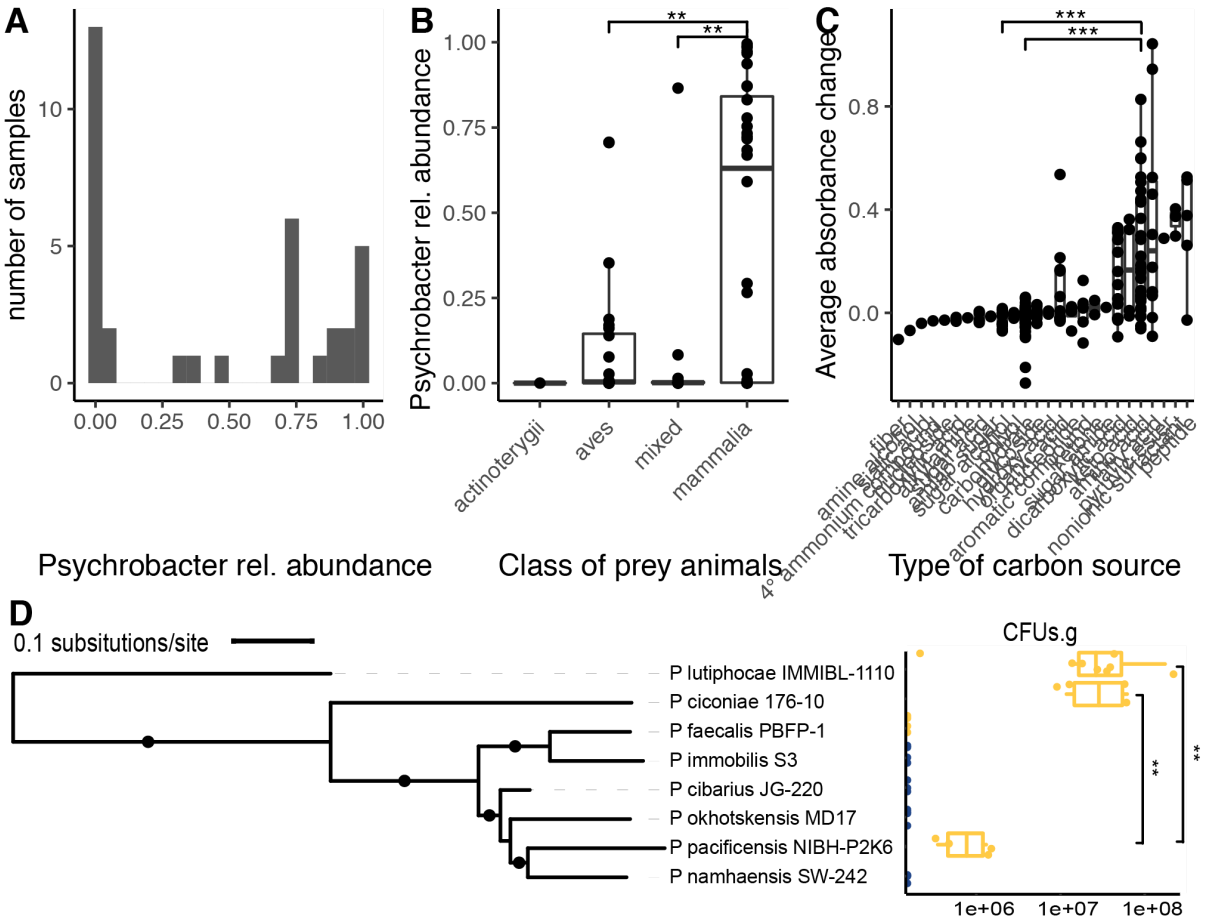

Figure 5. *Psychrobacter* strains occur and persist in the mammalian gut. A) Histogram of relative abundance at the genus level of 16S rRNA gene SVs, clustered at 99% identity, of polar bear fecal samples. B) Psychrobacter relative abundance in comparison to taxonomy of prey consumed by polar bears. C) The average change in OD600 of 19 Psychrobacter accessions grown on 190 different substrates as sole carbon sources, compared across different classes of compounds. D) CFUs per gram of cecal contents of gnotobiotic mice is shown in comparison to accession phylogeny. FE strains are shown in yellow, while RE strains are shown in blue. The phylogeny was constructed as described in Fig. 5. Branches with bootstrap support higher than 70% are indicated with a filled circle.

Means were compared using the Wilcoxon rank-sum test. * indicates a p-value less than 0.05, ** indicates $p < 0.005$, *** indicates $p < 0.0005$

Chapter 3. *Psychrobacter* spp. isolated from food-processing environments, invertebrates, and fish show distinct genomic characteristics.

3.1 Abstract.

Ecological niche can drive the genomic diversification of closely related bacterial species. Here, I apply genomic analyses to the genus *Psychrobacter*, known for its particularly wide ecological distribution. I performed a pan-genome analysis, microbial pan-genome wide association, ancestral character estimations, and analysis of selection on protein sequences of 85 strains of *Psychrobacter* to clarify the interactions between isolation source and *Psychrobacter* genome evolution. There is some evidence that Psychrobacter isolation source correlates with available gene pool; each isolation source - from warm-bodied hosts such as mammals or birds, other hosts like fish or invertebrates, to food processing, marine, or terrestrial environments - is correlated with unique genes from different COG categories. Strains isolated from invertebrate and fish hosts, as well as food processing environments, have higher numbers of total genes, though fewer unique genes than strains from mammals, birds, or marine environments. After accounting for population structure, however, very few genes correlate with isolation source; there is some evidence for increased horizontal gene transfer in strains isolated from fish and invertebrates, and evidence for increased biofilm formation in strains isolated from food-processing environments. Ancestral character estimation supports that *Psychrobacter* are descended from a warm host-associated bacterium. Finally, there is no correlation between isolation source and cold-adaptation of proteins. Overall, my results show that isolation source has some impact on *Psychrobacter* genome evolution, though the population structure of the genus makes it difficult to disentangle its effects.

3.2 Introduction.

With the advent of widespread next-generation sequencing (NGS) technology, microbiologists realized that closely related bacterial strains can carry vastly different genes, acquired through lateral gene transfer, and the pan-genome analysis was born (33). The term “core genome” was coined to describe a set of genes common to all genomes included in the analysis, and is associated with the general lifestyle of that group of organisms. By contrast, there are genes that are not shared by all genomes in the group, which is called the “accessory genome.” It has been shown that despite bacteria having relatively low recombination rates compared to sexually-reproducing eukaryotic organisms, specific regions of the genome which are linked to adaptive advantage in a specific environmental niche can increase in frequency in that niche (34), supporting the “ecotype” theory of bacterial evolution in which accessory genomes can contribute to ecological differentiation within close relatives (35). With their small size (between 0.5 (36) and 14 Megabases (37)) and high coding density (typically between 85 - 90% (38)), bacterial genomes and pan-genome analyses provide the ideal opportunity to explore the link between genomics and ecology.

Of course, one of the first areas of interest in the study of bacterial population structures was to explore population structures between clinical and environmental isolates of pathogens with environmental reservoirs. In some bacterial groups, such as *Escherichia coli* and *Shigella*, this was studied before sequence-based methods were available (39). This work has since been expanded to include studies based on whole genome sequencing rather than serotyping, multi-locus sequence typing, or restriction fragment length polymorphism analysis. For example, whole genome analyses of *Salmonella enterica* serotypes *Dublin*, *Newport*, and *Typhimurium* show accessory genome differentiation by whether they were isolated from human or cattle hosts (40). In a whole genome study of *Vibrio parahaemolyticus* isolates from oysters and hospital patients, the authors found that *V.*

parahaemolyticus is separated phylogenetically into oyster-only clades, human-only clades, and mixed clades, supporting differential adaptation for these environments between closely related isolates; furthermore, the *V. parahaemolyticus* accessory genome differed between human-only and oyster-only clades (41) .

With the NGS revolution and the emergence of the field of metagenomics, or the study of whole communities of microorganisms through direct sequencing of all DNA recovered from an environmental sample (42), microbiologists realized that non-pathogenic organisms greatly outnumber pathogens (43). One of the most well-studied, non-pathogenic groups for comparative genomics across multiple environments is the genus *Lactobacillus* (7, 44–46). There is, however, unusual genomic diversity within the genus (45); in April 2020, one study found using a polyphasic approach the genus could be split into 25 total genera (47). Given the extreme genetic diversity within its members, *Lactobacillus* may not be the most appropriate model group for bacterial ecological diversification.

The genus *Psychrobacter* exhibits a similarly broad ecological distribution as *Lactobacillus*, and because the genus was described relatively recently (19), after the introduction of 16S classification, the members are genomically closely related. To date, several comparative genomics studies have been done with *Psychrobacter* (32, 48–50), focused on elucidating its approaches to cold adaptation or exploring its potential for bioactive compounds. These studies revealed high levels of recombination and rearrangement within *Psychrobacter* genome architecture, and yet a relatively large core genome, indicating potentially little horizontal gene exchange or very recent diversification. A striking feature of the phylogeny of the genus is a deep divide between a smaller basal group which seems more warm-adapted, and a larger derived group which seems more cold-adapted.

While valuable, these studies utilize between just 1 genome (in which the protein sequences were compared to homologous sequences from SwissProt) and 26 genomes, when there are 41 validly published *Psychrobacter* species (51) and 290 *Psychrobacter* genomes available through the NCBI genome database (52) (including the 85 genomes

assembled for this study, as of December 2020). The study that included 26 genomes, and therefore captured the most *Psychrobacter* diversity to date, only separated *Psychrobacter* into three groups: those that were isolated from warm hosts, those that were isolated from marine environments, and those that were isolated from terrestrial environments. For convenience, this study classified one strain isolated from a fish as being warm host associated, and one strain isolated from a porpoise gut as being marine associated. I used several comparative genomics approaches to study the impact of isolation source, and therefore by proxy ecological diversification, on *Psychrobacter* genome evolution in a larger group of isolates with more detailed (and more accurate) information on isolation sources.

3.3 Materials and Methods.

I performed all analysis using R version 3.6.3; my notebooks are available on github at https://github.com/dkwelter/dkw_thesis_notebooks. I used the Benjamini-Hochberg method for all multiple test corrections on p-values.

***Psychrobacter* pan genome summary.** As described in Chapter 2, I performed a pan-genome analysis on 85 *Psychrobacter* genomes.

Briefly, I sequenced the genomes sequenced using a mixture of long read and short read sequencing, assembled them *de novo* using SPAdes (53), and quality-checked them using CheckM (54) recommended cutoffs. I annotated genes using eggNOG-mapper (55), which assigned them to Cluster of Orthologous Groups (COG) categories (56), performed gene clustering with the PanX pipeline (57), and created a phylogeny using the protein sequences of 400 marker genes per genome using PhyloPhlAn (58).

I performed pan-genome rarefaction and a Heaps law analysis using the R package micropan (59) , using 999 permutations for both the rarefaction() and heaps() functions.

***Psychrobacter* isolation source pan-genome-wide association study.** I performed a pan-genome-wide association study using the R package treeWAS (60) with isolation material as the traits of interest. Given the isolation material for each strain, provided by the strain catalogue form which I purchased the strains, I categorized strains as either “warm host”- (mammals or birds, n = 14), “other host”- (fish or invertebrates, n = 23), “food”- (frozen or fermented food products, n = 11), “marine”- (seawater or sea ice, n = 21), or “terrestrial”- (fresh water or soil, n = 16) derived.

***Psychrobacter* and *Moraxella* ancestral state reconstruction.** I generated a phylogeny with PhyloPhlAn using the same 85 *Psychrobacter* genomes described above, as well as genomes of 18 *Moraxella* species and *Acinetobacter puyangensis*, downloaded from the NCBI genome database (63) as described in Chapter 2. I estimated the ancestral state of the source of isolation using the ace function from ape (64) (type = “discrete”, method = “ML”, marginal = TRUE).

I first tested the likelihood of *Psychrobacter’*s ancestor being derived from a “warm host” isolation source versus any other isolation source, meaning there were two possible character states. To determine which model was most appropriate, I performed ace() using each of the three possible models: ER, or equal rates, assuming that the transition between the states occurs at equal rates; SYM, or symmetric rates, assuming that transitions between any two states are equal, but not necessarily equal to the transitions between any other two states; and ARD, or all rates different, which does not assume that any of the transitions between states are equal. In this case, since there are only two states, the ER model and SYM model were identical.

Table 3.1 Two-state ancestral character estimation for *Psychrobacter* isolation source. I performed ancestral character estimation using the ace() function from the R package ape using three different models of rate changes between states, in order to determine which model maximized the log likelihood while not overfitting the data.

Model	**Log Likelihood**	**Standard Error**
equal rates (ER)	-39.5	0.425
symmetric rates (SYM)	-39.5	0.425
all rates different (ARD)	-38.5	0.393, 0.606

The log likelihood was maximized for the ARD model, but when performing a likelihood test between models using the following equation,

p = 1- pchisq(2*abs(log likelihood of ER/SYM model - log likelihood of ARD model), 1)

p = 0.15, which is not significant, so I rejected the more complicated ARD model in favor of the simpler ER model.

I sampled the log likelihood of the isolation source of *Psychrobacter*'s ancestor at two positions along the phylogeny: at the root of the *Psychrobacter* and *Moraxella* clades and the root of the *Psychrobacter* clade. I performed the first using the entire tree described above. I performed the second by dropping *A. puyangensis* and all *Moraxella* tips except *M. lincolnii* from the phylogeny, and building a second model using the same parameters as the first.

I next tested the likelihood of *Psychrobacter*'s ancestor being derived from a warm host isolation source versus other host, food, marine, or terrestrial isolation sources. In this case, there were five possible ancestral states. I again performed ace() with the same parameters as described above using all three types of models, and used the same likelihood test to determine which was most appropriate.

Table 3.2 Five-state ancestral character estimation for *Psychrobacter* isolation source.

Model	Log Likelihood	Standard Error
ER	-155	3.05
SYM	-131	1.32, 0.703, 1.13, 1.09, 8.37, 15.0, 60.1, 43.8, 62.3, NaN
ARD	-123	3.57, 4.38, NaN, 4.22, 2.43, 5.53, NaN, NaN, NaN, NaN, NaN, NaN, 2.15, 11.5, NaN, NaN, NaN, NaN, NaN, NaN

In this case, the ARD model was significantly better than both the ER (p = 1.11e-15) and the SYM (p = 4.26e-05) models. However, both the SYM and ARD models' standard errors contain *NA* values as well as very large values, indicating that the likelihood surface is flat and the values for the transitions at each of the nodes are not necessarily reliable (65, 66). I again used the simplest ER model. I sampled the likelihood of *Psychrobacter'*s ancestral state as described above.

***Psychrobacter* cold adaptive traits.** For every gene cluster of the *Psychrobacter* pan-genome obtained, I used DIAMOND to map the consensus sequence against the UniRef90 database, excluding results from the family *Moraxellaceae*. We thus obtained homologous sequences while ensuring that the analysis would not be biased by self-comparison. I removed gene clusters without UniRef90 homologs from further analysis.

The putative cold adaptive traits of every protein-coding gene from the *Psychrobacter* set and their UniRef90 homologs were evaluated following the methods of Bakermans (32). Briefly, I used the relative abundance of every amino acid per coding sequence (CDS) to calculate the arginine to lysine ratio, acidity, hydrophobicity and KVYWREP ratios (32). I additionally calculated the grand average of hydropathicity (GRAVY) (61) and the isoelectric point (using Protein Analysis method class from Biopython). Next, I compared the distributions of these measures from *Psychrobacter* protein sequences to the distributions of the measures from the UniRef sequences. For each protein sequence, if a majority of measures were in the top 25% of the homologs distribution, I defined the protein sequence

as “highly cold adaptive” (HCA). The top 25% were defined as follows for each trait: highest 25% for glycine relative abundance, acidity, hydrophobicity, and GRAVY; and the lowest 25% for proline relative abundance, arginine relative abundance, arginine-lysine ratio, isoelectric point, and KVYWREP.

I performed a test for phylogenetic clustering of the data using phylosig from the R package phytools (62), using Blomberg’s K, (based on 1000 randomizations). I conducted a chi-square test to determine whether any of the COG categories were enriched or depleted in the HCA proteins compared to the *Psychrobacter* pan-genome as a whole.

3.4 Results and Discussion

The *Psychrobacter* pan-genome is relatively open. In the 85 *Psychrobacter* genomes I included in my analysis, there are a total of 227,212 gene sequences which cluster into 13,223 homologous gene clusters. Of those gene clusters, 1130 fall in the pan-genome category of “core” - that is, present in 99-100% of genomes; 453 gene clusters fall in “soft core” - present in between 90-98% of included genomes; 5744 gene clusters fall in the “shell” - between 2-89% of genomes; and 5896 gene clusters fall in the “cloud” - only present in 1 genome (Figure 3.1A).

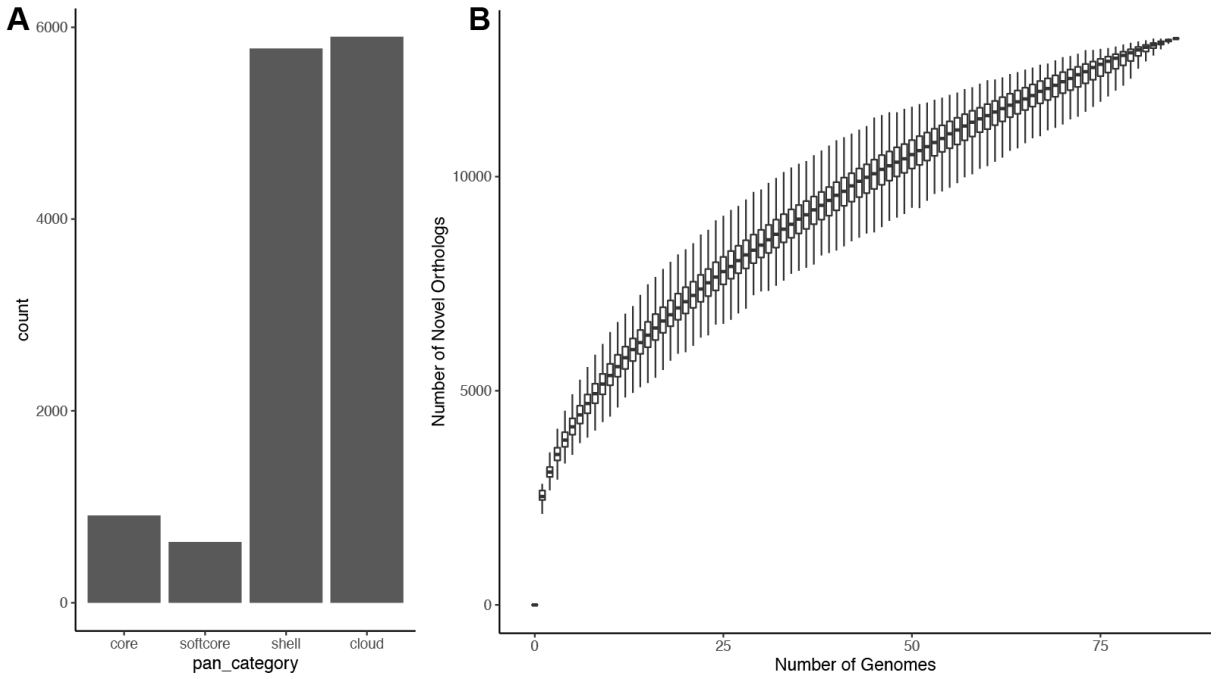

Figure 3.1 The *Psychrobacter* pan-genome is open. A) The PanX pipeline was used to cluster homologous sequences from 85 *Psychrobacter* genomes. Genes were separated into pan-genome categories according to how many strains carried homologous sequences: “core” 99 - 100% of strains, “softcore” 90 - 98% of strains, “shell” 2 - 89% of strains, “cloud” only 1 strain. B) The rarefaction() function from the R package micropan was used to estimate how many novel homologous gene clusters are encountered from the addition of each genome within the 85 genomes included for this analysis, with 999 random permutations of the order in which genomes are added.

I was interested in whether I had fully sampled *Psychrobacter* diversity, so I performed pan-genome rarefaction, which is an analysis of how many novel gene clusters are added to the pan-genome with the addition of each new genome, using the R package micropan (Figure 3.1 B). An analysis with Heaps law, which describes the rate of novel gene discovery within a pan-genome, revealed the decay parameter of the *Psychrobacter* pan-genome $\alpha = 0.552$. Values less than 1 indicate that the pan-genome is still open (33), meaning that with the genomes I included in my analysis, I have not yet sampled the full diversity of the genus. Future analyses should be conducted including many more genomes; this would likely increase the number of shell and cloud genes, and perhaps slightly decrease the size of the core and softcore genomes. For reference, the Bakermans study including 26 *Psychrobacter* genomes (32) found 1,200 gene clusters in the *Psychrobacter*

core genome, which is substantially larger than the number of core gene clusters I found (although that study defined core gene clusters as being contained by 95% of genomes).

I was interested in whether *Psychrobacter* genes differed in function depending on which pan-genome category they fell into (Figure 3.2 A). The COG categories encoded by each of the core, softcore, shell, and cloud genomes are significantly different (χ^2 (69, 96) = 2748, p-value < 2.2e-16). By examining the residuals, I determined that *Psychrobacter's* core genome is significantly enriched in gene clusters from COG category J (translation, ribosomal structure and biogenesis); slightly enriched in O (post-translational modification and protein turnover), C (energy production), E (amino acid transport and metabolism), F (nucleotide transport and metabolism), H (coenzyme transport and metabolism); and substantially depleted in COG category X (no homologs in COG database) (Figure 3.2 B). These categories relate to general lifestyle and housekeeping functions, so it is not surprising to find them enriched in the core genome.

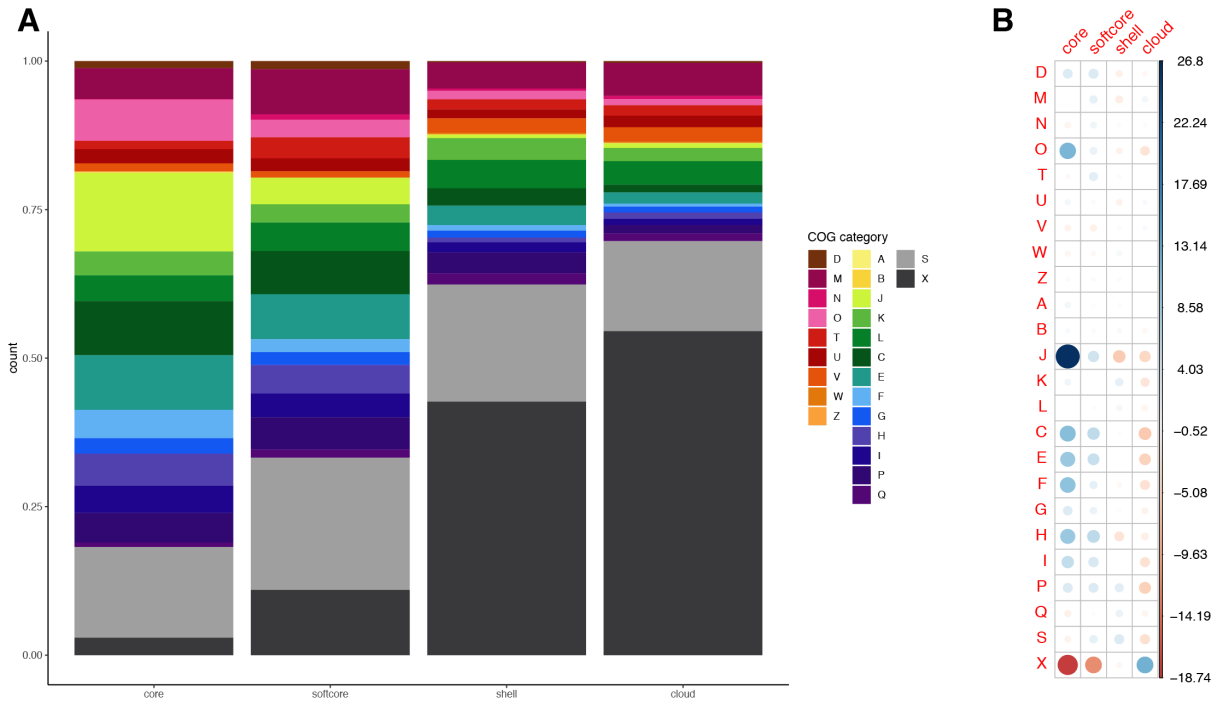

Figure 3.2 The *Psychrobacter* shell- and cloud- genomes are enriched for proteins of unknown function. A) The proportion of each pan-genome category devoted to each cluster of orthologous genes (COG) category. Cellular processing and signaling: [D] cell cycle control cell division, chromosome partitioning; [M] cell wall/membrane/envelope biogenesis; [N] cell motility; [O] post-translational modification, protein turnover, and chaperones; [T] signal transduction mechanisms; [U] intracellular trafficking, secretion, and vesicular transport; [V] defense mechanisms; [W] extracellular structures; [Y] nuclear

structure; [Z] cytoskeleton. Information storage and processing: [A] RNA processing and modification; [B] chromatin structure and dynamics; [J] translation, ribosomal structure and biogenesis; [K] transcription; [L] replication, recombination, and repair. Metabolism: [C] energy production and conversion; [E] amino acid transport and metabolism; [F] nucleotide transport and metabolism; [G] carbohydrate transport and metabolism; [H] coenzyme transport and metabolism; [I] lipid transport and metabolism; [P] inorganic ion transport and metabolism; [Q] secondary metabolites biosynthesis, transport, and catabolism. Poorly Characterized: [S] function unknown, [X] not in COG database. B) residuals from a chi-square test of the COG category representation across each of the pan-genome categories, produced by the R function corrplot(). The size of the circle reflects the difference between the expected and observed values, and the color indicates the direction of the difference, with blue values being positive and red values being negative.

The softcore and shell gene cluster COG categories largely resemble the core, except that there is an ever larger portion of X category clusters as well as S (function unknown) category clusters. The cloud genome is slightly depleted in the core-enriched categories while being extremely enriched in category X. This huge swell in the proportion of genes with no homologs is also unsurprising, given that genes that are necessary for housekeeping functions that are common to all bacteria are more likely to already have been sequenced and characterized compared to genes specific to a particular bacterium.

***Psychrobacter* spp. from different isolation sources have varying genome sizes.** The average *Psychrobacter* genome has 2,673 gene clusters: 1148 or 43% of them are core, 450 or 6% are soft-core, 1005 or 48% are shell, and 70 or 3% are cloud. The number of genes per genome differs significantly by isolation source (Kruskal-Wallis chi-squared = 16.388, df = 4, p-value = 0.00254), with strains isolated from food and other host bodies carrying more genes than strains from other isolation sources (Figure 3.3 A). Additionally, the ‘pan-genome structure’ differs significantly by isolation source. Strains from warm hosts, marine, and terrestrial sources have greater proportions of core (Kruskal-Wallis chi-squared = 18.4, df = 4, p-value = 0.00101) and soft-core (Kruskal-Wallis chi-squared = 12.5, df = 4, p-value = 0.014) genes than strains from other hosts or food. Conversely, strains isolated from other hosts and food have significantly greater proportions of shell (Kruskal-Wallis chi-squared = 18.2, df = 4, p-value = 0.00114) genes than strains from warm hosts, marine, or terrestrial sources. Strains isolated from food trend towards having greater numbers of cloud genes,

although it is not significant (Kruskal-Wallis chi-squared = 7.69, df = 4, p-value = 0.103).

(Figure 3.3 B)

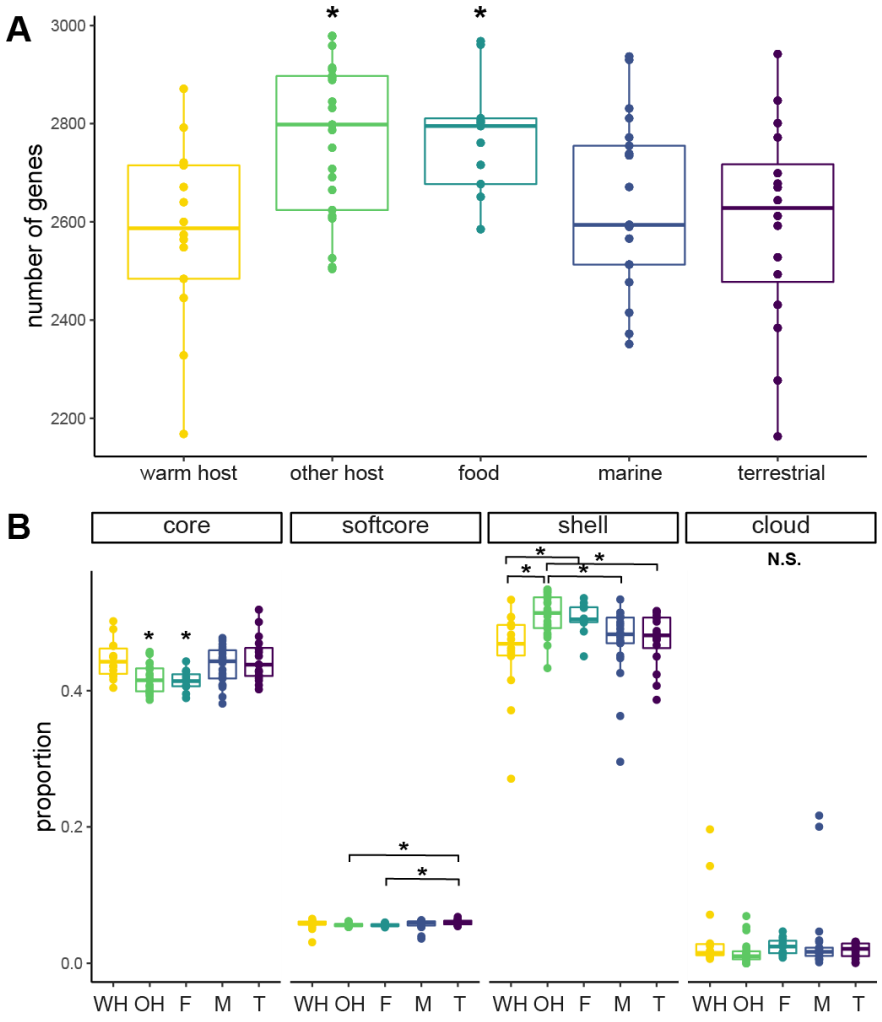

Figure 3.3 The pan-genome structure of *Psychrobacter* differs by isolation source. A) Strains from other hosts and food have significantly higher numbers of genes per genome than the other three isolation sources. B) Core genes are present in 99-100% of genomes, soft core in 90-98%, shell in 2-89%, and cloud only in a single genome. Strains from other hosts and food have significantly lower proportions of core genes and higher proportions of shell genes than the other three isolation sources; WH = strains isolated from warm hosts, OH = other host, F = food-processing, M = marine, T = terrestrial.

These patterns, of course, are unsurprising. Given that strains from other hosts and food have higher numbers of genes, it is by definition that they would have smaller proportions of their genomes devoted to core and soft-core genes and larger proportions of shell and cloud genes. What is more interesting, is that several genomes have drastically higher proportions of cloud genes than other members of that isolation source, since larger

cloud genomes indicate rapid gene gain and likely also indicate niche differentiation compared to relatives (67). The median proportion of cloud genes for warm host isolates is 1.5%, yet *P. ciconiae* has 19%, *P. lutiphocae* has 14%, and *P. phenylpyruvicus* has 7%. The median proportion for other host isolates is 1%, but *P. sp. H7-1* has 7%, and *P. fulvigenes* and *P. immobilis* strain *1184* have 5%. For marine strains, the median proportion is 2%, but *P. arenosus* has 22%, *P. aestuarii* has 20%, and *P. aquimaris* has 5%. Many of these strains are very basal on the tree and occur as outgroups to other, more derived clades, which could mean that these high proportions of cloud genes represent unsampled *Psychrobacter* diversity, rather than true gene gain.

***Psychrobacter* from different isolation sources harbor unique genes.** I further characterized the *Psychrobacter* pan-genome by identifying gene clusters that are unique to each isolation source. There are 1,632 gene clusters unique to strains isolated from warm hosts, 85% of which are cloud genes (only appearing in one strain); 1,214 gene clusters unique to other hosts, 83% of which are cloud; 834 unique to food, 88% of which are cloud; 2,043 unique to marine, 97% of which are cloud; and 955 unique to terrestrial, 81% of which are cloud (Figure 3.4 A). The warm host unique gene clusters are enriched for COG categories N (cell motility), U (intracellular trafficking), L (replication recombination), E (amino acid transport and metabolism), H (coenzyme transport and metabolism), P (inorganic ion transport and metabolism), Q (secondary metabolite synthesis), and S (function unknown). The other host unique gene clusters are enriched for COG categories V (defense mechanisms) and X (no homologs). The food and terrestrial unique gene clusters are both enriched for COG categories X (no homologs). The marine unique gene clusters are enriched for COG categories O (post-translational modification) and T (signal transduction). (Figure 3.4 B)

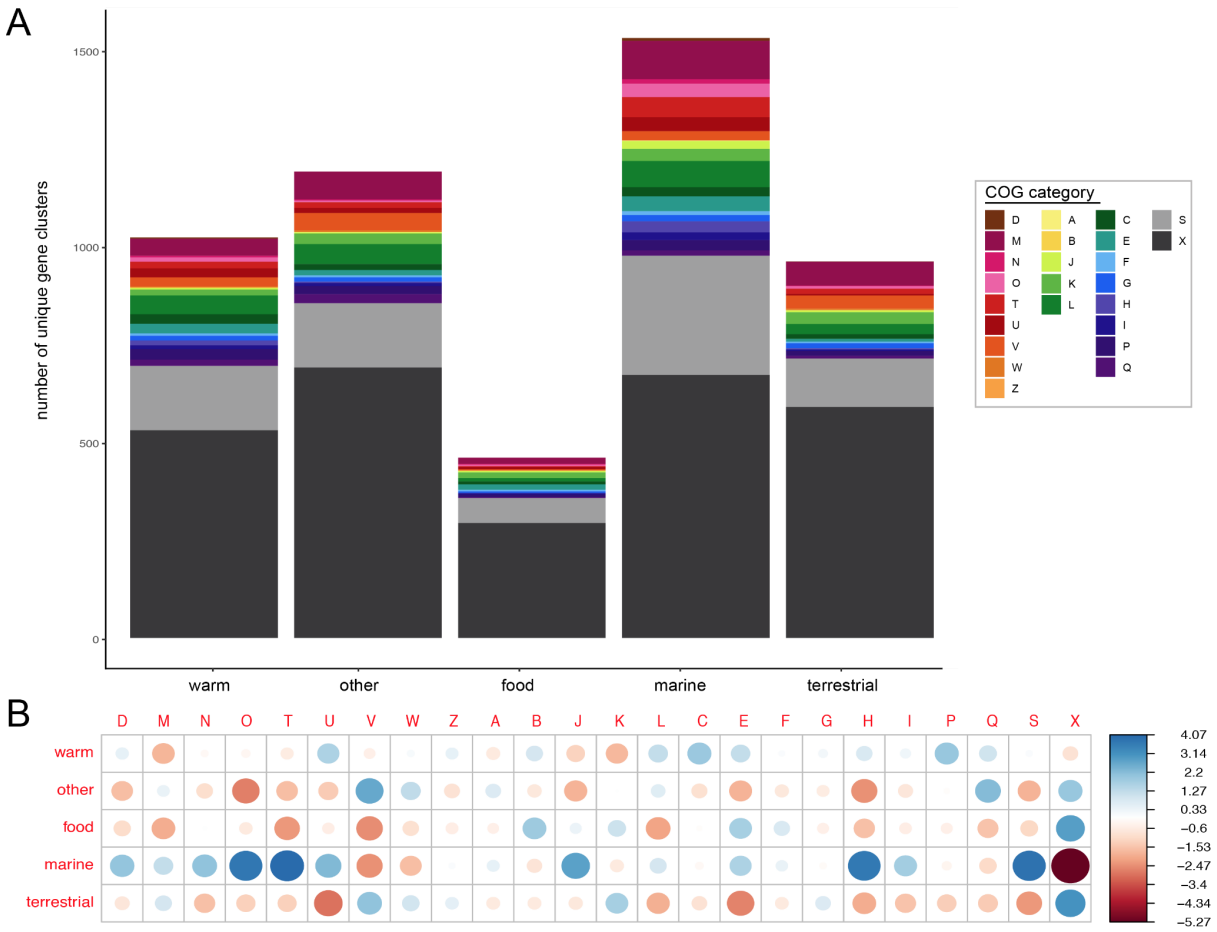

Figure 3.4 *Psychrobacter* isolation sources are associated with unique cloud gene content. A) The COG category counts of genes unique to each isolation source. Note that for genes annotated with multiple COG categories, each category is counted separately. Cellular processing and signaling: [D] cell cycle control cell division, chromosome partitioning; [M] cell wall/membrane/envelope biogenesis; [N] cell motility; [O] post-translational modification, protein turnover, and chaperones; [T] signal transduction mechanisms; [U] intracellular trafficking, secretion, and vesicular transport; [V] defense mechanisms; [W] extracellular structures; [Y] nuclear structure; [Z] cytoskeleton. Information storage and processing: [A] RNA processing and modification; [B] chromatin structure and dynamics; [J] translation, ribosomal structure and biogenesis; [K] transcription; [L] replication, recombination, and repair. Metabolism: [C] energy production and conversion; [E] amino acid transport and metabolism; [F] nucleotide transport and metabolism; [G] carbohydrate transport and metabolism; [H] coenzyme transport and metabolism; [I] lipid transport and metabolism; [P] inorganic ion transport and metabolism; [Q] secondary metabolites biosynthesis, transport, and catabolism. Poorly Characterized: [S] function unknown, [X] not in COG database. B) residuals from a chi-square test of the COG category representation across each isolation source, produced by the R function corrplot(). The size of the circle reflects the difference between the expected and observed values, and the color indicates the direction of the difference, with blue values being positive and red values being negative.

It is difficult to comment on the ecological relevance of the unique genes, since so many of them appear only in one strain. Strains isolated from other hosts carry a unique

number of V- (cellular defense) COG category genes, most of which are restriction-modification endonucleases and methyltransferases. Restriction-modification systems are well-known for deactivating mobile genetic elements, but high numbers can also indicate higher rates of genetic exchange in a particular lineage (71). Higher rates of horizontal gene transfer in *Psychrobacter* spp. isolated from other hosts could help explain why other hosts have carry more genes than strains from other isolation sources.

Strains isolated from marine sources carry many unique O- (post-translational modification) and T- (signal transduction) COG category genes. Most of the O-category genes are related to either protein degradation or oligomerization, though some of them are heat shock proteins or their chaperones. The unique T-category genes include many histidine kinases, genes involved in c-di-GMP formation and sensing, and extra-cytoplasmic function sigma and anti-sigma factors. Unfortunately the annotations for these genes are not specific enough to predict what conditions these genes might be involved in sensing or what the resulting response might be. Previous genomic comparisons of marine oligotrophs to copiotrophs revealed that marine copiotrophs tend to have increased numbers of signal transduction genes to quickly respond to changing environmental stimuli, which could include changing concentrations of dissolved proteins to scavenge (72).

Few genes correlate with isolation source after accounting for phylogeny. To determine whether isolation source impacts genomic content of *Psychrobacter* strains while accounting for population structure, I performed a pan-genome wide association study (pan-GWAS) using the R package treeWAS. TreeWAS performs several phylogenetically-aware tests to determine whether a trait is significantly correlated with the presence or absence of gene clusters in comparison to a random “null distribution” of the traits compared to phylogenetic relation; treeWAS has already been used in several publications to examine bacterial populations for signs of host adaptation (73–75) and ecological diversification in general (75).

Using this approach, I found 5 gene clusters significantly associated with isolation from other host bodies, 5 associated with isolation from food, 4 associated with marine environments, and 1 associated with terrestrial environments. There were no gene clusters significantly associated with isolation from warm host bodies. (Appendix I: Table 3.3)

I expected to find some genes associated with warm host bodies, as the previous 26-genomes *Psychrobacter* comparative genomics study found that warm host isolates had more unique gene content compared to isolates from marine or terrestrial sources (32), and there were a high number of gene clusters unique to warm host strains in my own analysis. Intuitively, the body of a warm host is more different from the other environments *Psychrobacter* can be isolated from than those environments are from each other. However, it must be noted that the previous study did not control for population structure, and was only examining the gene pool differences between the basal and derived *Psychrobacter* clades. After including warm host isolates from outside of the basal clade, and after accounting for the phylogenetic structure of the genus, these gene presence/absence differences are no longer significant.

The gene clusters significantly enriched in strains isolated from other hosts include one gene with no homologs in the eggNOG database, a nucleotide binding protein, a DNA primase, and an adenine specific DNA methyltransferase. Genes relating to K- and L- COG categories support my earlier finding that strains isolated from other hosts carry many unique DNA replication and transcription regulation genes, potentially due to greater rates of gene exchange than other *Psychrobacter* strains.

The gene clusters significantly enriched in strains isolated from food sources include one uncharacterized protein with a KWG leptospira repeat domain, two predicted fimbrial chaperone proteins, and two fimbrial subunit proteins. Fimbriae are typically associated with pathogenesis, especially in mediating cell attachment to the gastrointestinal tract, increasing cell survival inside macrophages, and increasing biofilm formation (76–79). Biofilm formation is a known bacterial response to stressful environments and so could be advantageous for bacteria in frozen or fermented food processing environments (80).

Significantly associated with terrestrial strains is the depletion of a particular resolvase from the Tn3 transposon family. Tn3 transposons are best known for their role in the dissemination of antibiotic resistance genes, but have also been implicated in horizontal gene transfer of novel catabolic gene functions (81), since they are capable of carrying quite long segments of DNA compared to many other transposase families (82). In total, there are 5 homologous gene clusters annotated as Tn3 family genes, including 2 transposases, 2 integrases, and the resolvase. The transposases and integrases are not significantly depleted in terrestrial strains, so the depletion of the resolvase may not be ecologically relevant, just a coincidence.

Finally, strains isolated from marine environments are significantly enriched for one hypothetical protein with no homologs in the eggNOG database and a protein of unknown function with a DUF1810 domain. It is difficult to comment on the potential ecological significance of these genes, since neither has an annotated function.

***Psychrobacter*'s ancestor was likely a bacterium associated with warm-bodied hosts.** As discussed in Chapter 2, one of the things that makes *Psychrobacter* unique is the possibility that its ancestor was a mammalian pathobiont, since evolutionary transitions from environmental to host-associated are more common than the converse. To further explore this *Psychrobacter*'s possible evolutionary history, I performed ancestral character estimation using the R package phytools with the equal rates model, based on a phylogeny constructed from 85 *Psychrobacter* genomes and 18 *Moraxella* genomes, with *Acinetobacter puyangensis* as an outgroup.

In comparing “warm host” versus all other isolation sources (Figure 3.5 A), the likelihood that the ancestor of *Moraxella* and *Psychrobacter* was warm host-associated is 83%. The likelihood of the direct ancestor of *Psychrobacter* being warm host-associated is 55%. In comparing warm host versus all other isolation sources (other host, food, marine, and terrestrial) (Figure 3.5 B), the likelihood of warm host-association for the ancestor of *Psychrobacter* and *Moraxella*, as well as the direct ancestor of *Psychrobacter*, is 22%.

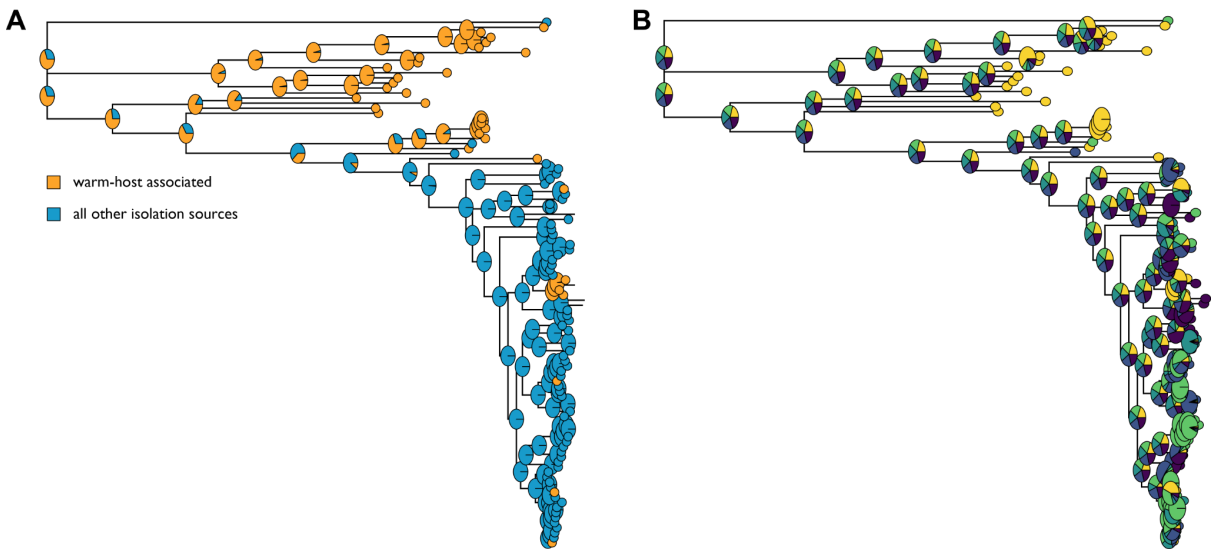

Figure 3.5 *Psychrobacter* ancestral state was more likely to be 'warm host' associated than any other isolation source. A phylogeny of the *Moraxella* and *Psychrobacter* genera, with *Acinetobacter puyangensis* as an outgroup, was constructed using PhyloPhlAn, using amino acid sequences from 400 marker genes in the alignment. Isolation sources were mapped for each strain on the tree tips, and likelihood of each isolation source was calculated for each node on the phylogeny using the ace() function from ape, with marginal estimation, an ER model, and the maximum likelihood method. A) Two-state ancestral character estimation, collapsing strains into "warm host associated" or "not." B) Five-state ancestral character estimation.

This supports that *Psychrobacter* likely evolved from a bacterium that was associated with warm bodied hosts, but there are many pitfalls associated with ancestral character estimation. In particular, my data set is too small to allow the use of a more complicated model than the equal rates model, which is not very realistic, and predicts that all isolation sources are equally likely for the five-state model.

***Psychrobacter* adaptation to cold temperatures is not uniform.** The amino acid content of proteins can be studied to learn about selection pressures on organisms. In particular, properties of proteins adapted to function at cold temperatures are well known (83, 84). I was interested in examining the predicted cold adaptation of *Psychrobacter* protein coding sequences, so I classified proteins as highly cold adaptive (HCA) or not by comparison to homologs from Uniref90. Out of 227,212 total gene sequences, 75,813 (33%) were classified as HCA. This corresponded with 5,238 total gene clusters out of 13,223 (39%).

I was interested in if gene clusters as a whole, or instead only a subset of sequences within a gene cluster, would be classified as HCA. Therefore, I classified the percentage of HCA sequences within a gene cluster, both with (Figure 3.6 A) and without (Figure 3.6 B) gene clusters only containing a single sequence. Most gene clusters do not include any HCA gene sequences. Of the gene clusters that do include HCA sequences, it is more common that every sequence in the gene cluster is HCA, than some sequences being HCA and some not. This is not entirely surprising, as the gene clusters were generated by comparison of their amino acid sequences; for homologous genes, those with more cold adaptive amino acid properties may form separate gene clusters than those without.

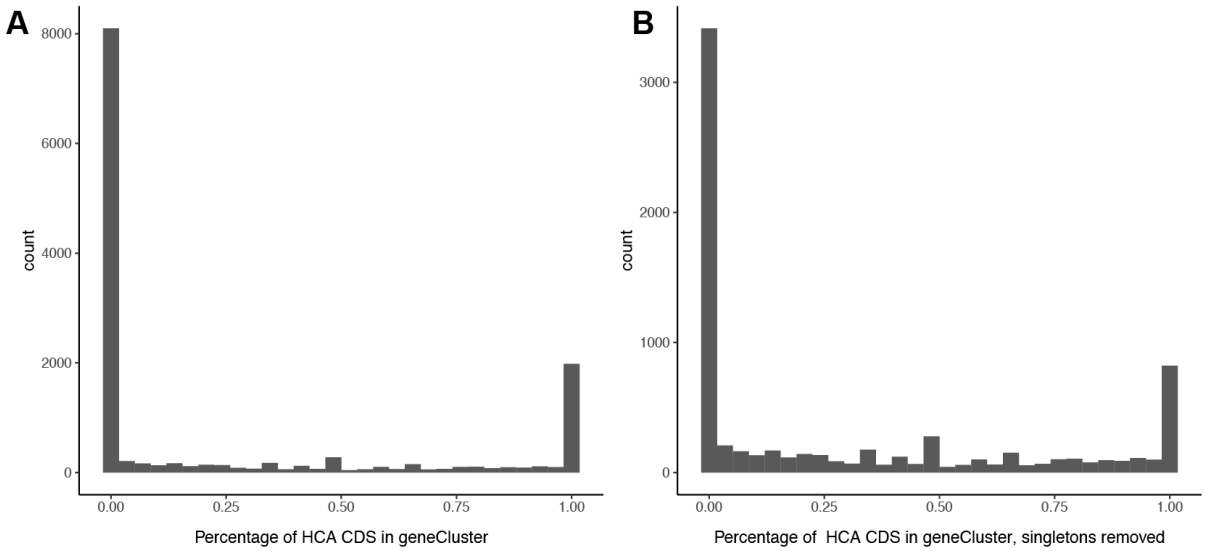

Figure 3.6 Cold adaptation of homologous *Psychrobacter* protein sequences is typically all or nothing. A) The percentage of sequences within each *Psychrobacter* homologous gene cluster defined as highly cold adaptive (HCA) by comparison to Uniref90 sequences. B) The percentage of sequences within gene clusters excluding 'singleton' gene clusters (containing only one sequence) defined as HCA.

I next examined the percentage of HCA genes carried by each strain (Figure 3.7 A). Between 25% and 40% of all genes carried by *Psychrobacter* strains are HCA, with a median of 34%. This value varies significantly by phylogeny (Blomberg's K : 4.50e-05, p-value = 0.04), but not as expected under Brownian evolution, *i.e.* genetic drift (85). Given that the derived clade of the *Psychrobacter* phylogeny exhibits higher likelihood of growth at

cool temperatures compared to the basal clade (Chapter 2 and 4), it is safe to assume that cold adaptation has been differentially selected for across the *Psychrobacter* phylogeny.

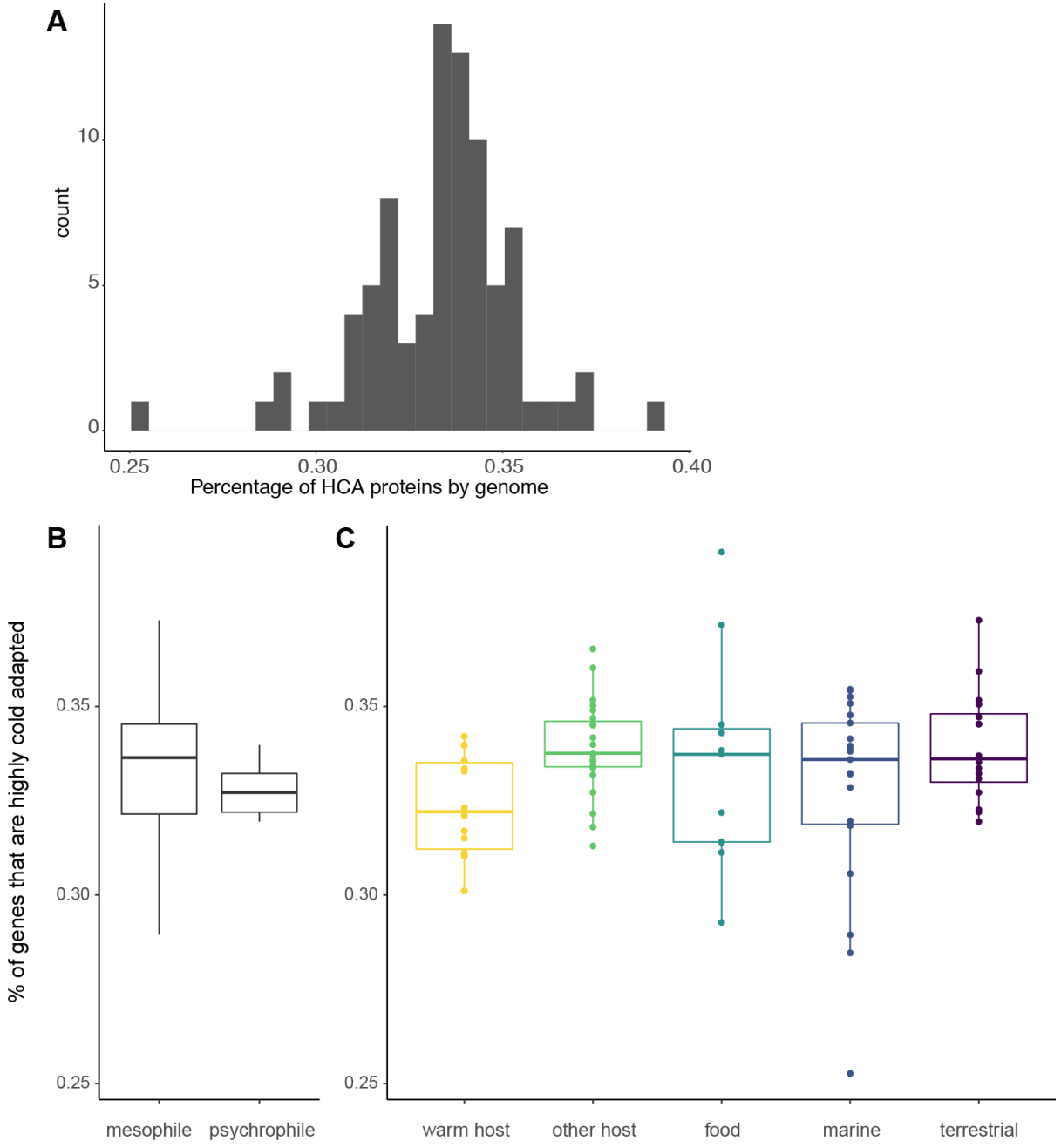

Figure 3.7 The percentage of highly cold-adaptive protein sequences varies widely across Psychrobacter strains, but not by lifestyle or isolation source. A) The percentage of HCA genes within each *Psychrobacter* strain. The percentage of HCA genes divided by B) restriction to temperatures under 25 ºC in my own phenotyping experiment (see Chapter 2) or C) isolation source. No significant differences in either B or C.

I also divided *Psychrobacter* strains by whether or not they exhibited psychrophilic growth behaviors (growth only at 20 ºC or below) in my growth curve screen and checked

the percentage of HCA genes between groups (Figure 3.7 B). There was no significant difference between the psychrophilic strains and mesophilic strains (Wilcoxon rank sum test, W = 256, p-value = 0.30), although only five strains were classified as psychrophilic in this case, which greatly decreases the power of the test. I also tested the percentage of HCA genes by isolation source, where again, there was no significant difference (Kruskal-Wallis χ^2 = 0.853, df = 4, p-value = 0.07) (Figure 3.7 C). In this case, there is a trend of warm host strains having a lower percentage of HCA genes compared to strains from other sources, which is what one might expect.

I next wanted to learn where in the pan-genome the HCA genes occurred: whether they were core genes versus softcore versus shell versus cloud (Table 3.4).

Table 3.4 Psychrobacter shell genes are more likely to be highly cold adapted. Core genes are present in 99-100% of genomes, soft core in 90-98%, shell in 2-89%, and cloud only in a single genome.

pan-genome category	total genes	number of HCA genes	percent of HCA genes
core	78745	23204	0.29
softcore	53746	18211	0.34
shell	88735	33205	0.37
cloud	5986	1193	0.20

The *Psychrobacter* shell genome is enriched for HCA gene sequences (χ^2 (3, 8) = 868, p-value < 2.2e-16). Since the whole genus is psychrotrophic, I expected that the core or softcore genomes may be enriched for HCA protein sequences, but this is not the case. Of course, the classification of “shell genome” includes a huge range of prevalence, between 2% and 89%. It may be more useful and informative to have more resolved categories of gene clusters, or to examine the genus on a clade-by-clade basis rather than as a whole.

***Psychrobacter* genes with high cold-adaptation have well-characterized functions.** I next wanted to determine whether particular functions were enriched in the HCA genes compared to *Psychrobacter* genes as a whole. The COG categories encoded by HCA genes are significantly different from those encoded by non-HCA genes (χ^2 (23, 48) = 3 750,

p-value < 2.2e-16) (Figure 3.8 A). HCA genes are highly enriched for L category genes (replication, recombination, and repair) and slightly enriched for O (post-translational modification and protein turnover), T (signal transduction), and Q (secondary metabolites biosynthesis and transport) category genes. HCA genes are greatly depleted for X (no homolog) category genes compared to non-HCA genes (Figure 3.8 B).

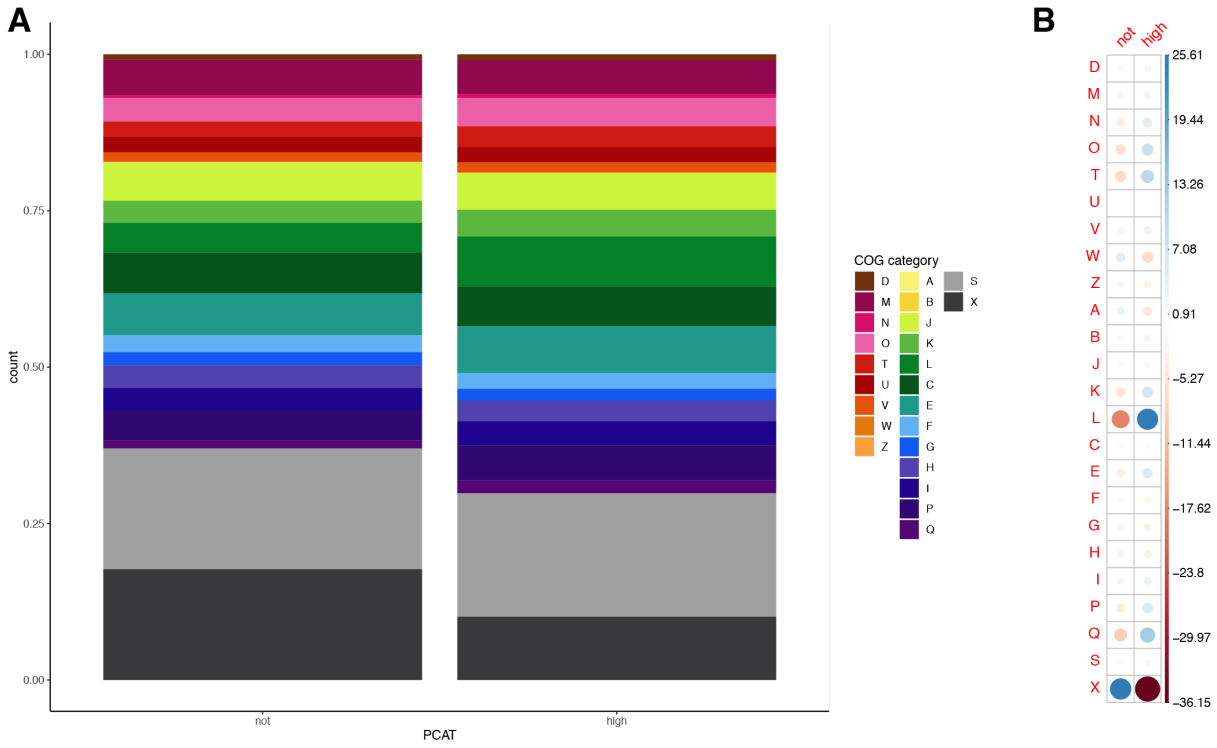

Figure 3.8 *Psychrobacter* proteins classified as cold-adapted are enriched for replication and recombination related functions compared to non-cold adapted proteins. A) The proportion of HCA versus non-HCA genes devoted to each cluster of orthologous genes (COG) category. Cellular processing and signaling: [D] cell cycle control cell division, chromosome partitioning; [M] cell wall/membrane/envelope biogenesis; [N] cell motility; [O] post-translational modification, protein turnover, and chaperones; [T] signal transduction mechanisms; [U] intracellular trafficking, secretion, and vesicular transport; [V] defense mechanisms; [W] extracellular structures; [Y] nuclear structure; [Z] cytoskeleton. Information storage and processing: [A] RNA processing and modification; [B] chromatin structure and dynamics; [J] translation, ribosomal structure and biogenesis; [K] transcription; [L] replication, recombination, and repair. Metabolism: [C] energy production and conversion; [E] amino acid transport and metabolism; [F] nucleotide transport and metabolism; [G] carbohydrate transport and metabolism; [H] coenzyme transport and metabolism; [I] lipid transport and metabolism; [P] inorganic ion transport and metabolism; [Q] secondary metabolites biosynthesis, transport, and catabolism. Poorly Characterized: [S] function unknown, [X] not in COG database. B) residuals from a chi-square test of the COG category representation of HCA versus non-HCA genes, produced by the R function corrplot(). The size of the circle reflects the difference between the expected and observed values, and the color indicates the direction of the difference, with blue values being positive and red values being negative.

Many of the HCA L-category genes are transposases and integrases. Insertion sequences have been reported to be common in bacteria associated with sea ice (86), which could explain why they may have cold-optimized amino acid sequences. Other L-category genes include DNA polymerases and helicases, which have a known role in cold adaptation, as do the T-category signal transduction genes (87). The O-category genes include protein chaperones, which have been shown to have enriched expression during cold stress in cyanobacteria (88). The Q-category genes include many genes involved in siderophore production, which has been shown to be upregulated at cold temperatures in *Rhodococcus* (89). I was initially surprised that there were so few HCA X-category genes, given the high number of X-category genes in the Psychrobacter pan-genome, but since genomic adaptations to cold temperatures is fairly well-studied, many genes involved in cold adaptation have been identified and characterized (84).

3.5 Conclusion

In the analysis presented above, I examined 85 strains from the genus *Psychrobacter* for genomic differences relating to their different sources of isolation: warm-bodied hosts (mammals or birds), other hosts (fish or invertebrates), food processing environments, marine environments, or terrestrial environments. Much of my work recapitulates what was found based on a previous analysis using 26 *Psychrobacter* genomes (32), divided by whether they were isolated from warm hosts, marine, or terrestrial environments. However, my analysis included many more genomes and more detailed information about the source of the strains involved. I also incorporated phylogenetically-aware tests, which was not previously done. The *Psychrobacter* pan-genome is open, but the genomes of individual strains share a large number of core conserved genes. The core and soft-core genes consist of largely regulatory functions, but the shell and cloud genes are dominated by hypothetical proteins with no known homologs in the eggNOG database.

Strains from differing isolation sources have unique gene content, but after controlling for population structure, very few genes are significantly correlated with isolation source. Strains isolated from invertebrates, fish, and food-processing environments have greater numbers of genes. In particular, strains from invertebrates and fish are significantly enriched for genes controlling DNA transcription and replication, which could support increased rates of horizontal gene transfer in their genomes. Strains isolated from food-processing environments are enriched for genes increasing biofilm formation capacity, which is likely advantageous in frozen and fermented food colonization. There are no genes significantly associated with isolation from warm bodied hosts, though an ancestral character estimation supports that *Psychrobacter* are derived from a warm host-associated bacterium.

I also examined *Psychrobacter* protein coding sequences for evidence of adaptation to cold temperatures. Genes with high predicted cold adaptation had predicted functions already known to associate with cold adaptation, such as DNA replication and signal transduction. These highly cold-adapted genes are typically found in the shell genome, not the core or soft-core genomes, even though *Psychrobacter* as a genus is psychrotolerant. There is no difference in proportion of highly cold adaptive protein sequences per genome when examined by isolation source, further supporting my assertion that while isolation source has some impact on *Psychrobacter* evolution, it is not a perfect predictor of genomic or phenotypic traits.

Future analysis of the implications of *Psychrobacter* ecology on genome evolution should include a direct analysis of the population structuring of *Psychrobacter* strains. Reverse ecology approaches have previously been used to define microbial populations with diverging adaptations to temperature (90); perhaps there is a more meaningful population structure within the genus *Psychrobacter* than by defined “isolation source.” It would be interesting to use SNP data from core genes to estimate migration between different isolation sources, which would also improve information about the population structure within the genus. My analysis did not include any direct prediction of ancestral sequences or horizontal gene transfer. Horizontal gene transfer is known to contribute to ecological

differentiation between close relatives (91), so a definitive prediction of gene transfer rates within the genus would be enlightening. Finally, future analysis should include many more genomes. The 85 genomes I included here were not enough to introduce a more detailed isolation source analysis, as evidenced by the difficulty in applying ancestral character estimation. Including additional genomes, especially ones that are more closely related to “outgroup strains” such as those exhibiting extremely high numbers of cloud genes, may also help resolve any issues in disentangling the effects of isolation source on evolution from phylogenetic relatedness of strains.

Chapter 4: *Psychrobacter* spp. from warm-bodied hosts show the most distinct *in vitro* growth characteristics.

4.1 Abstract.

I was interested in how *Psychrobacter* isolated from different ecological niches might exhibit differences in phenotypic behavior, so I performed a large scale phenotype collection on 85 *Psychrobacter* strains. I collected information on strains' oxygen utilization profiles and abilities (or lack thereof) to grow in 24 different conditions, along with growth rate and maximum optical density reached. I found that *Psychrobacter* isolation source is strongly correlated with ability/inability to grow under certain conditions, with strains isolated from mammals and birds more likely to grow in rich media, low salt, and high temperature conditions compared to strains from other isolation sources. Isolation source and isolation location are also very weakly correlated with growth rate and maximum optical density, but explain very little of the variation in the data. Strains from birds and mammals are less likely to be strict aerobes than strains from other isolation sources, but not significantly. Overall, my data shows that *Psychrobacter* strains isolated from warm-bodied hosts behave somewhat differently than *Psychrobacter* strains from other isolation sources, and in particular, this difference is likely driven by more phylogenetically basal strains.

4.2 Introduction.

While metagenomics has revealed much uncultured microbial diversity, a solely sequenced-based approach ultimately cannot give us a complete understanding of microbial

ecology, and cultivating bacteria in the lab remains very important (92). Once cultured, publishing information characterizing novel bacterial species is very challenging; there are 11 strict criteria for publication of strain as “new species” (*species nova*) in the International Journal of Systematic and Evolutionary Microbiology (93). It is expensive and difficult or impossible to perform such thorough phenotyping on every cultured strain, leading to deposition of many uncharacterized strains in catalogues and the genomes of many uncharacterized strains in genomic databases.

The genus *Psychrobacter* has shown interesting phenotypic flexibility, with some strains being restricted to psychrophilic lifestyles, while others are capable of growth between 0 ºC and 42 ºC (20). Type strain papers can be patchy in reporting how phenotypic tests were performed, however. For example, *P. lutiphocae* is reported to grow between 10 and 37 ºC. The relevant portion of the type strain paper methods section simply states, “The growth temperature range of the isolate was 10 - 37 ºC” (94), but does not mention which temperatures were included in the tests, or on which media the tests were performed. It is thus difficult to compare phenotypic data directly taken from type strain papers.

In particular, I was interested in phenotypes that may relate to *Psychrobacter* strains’ abilities to colonize the mammalian gut. One obvious difference between the mammalian gut and other environments in which *Psychrobacter* is found is temperature; mammalian body temperature ranges between 30 - 40 ºC (95), while seawater is typically between -2 - 2 ºC (96). Along with temperature, there is another clear difference in salinity; it is difficult to measure the exact osmolarity of intestinal content, but solutions of 0.16 mM NaCl, or around 0.9%, are considered isotonic for mammalian tissues (97), while seawater has a salinity of around 3.3% (98), and glacial cryopegs can have a salinity of 30% (99). Another potential difference is nutrient availability; nutrient levels in the gut vary depending on location and time since feeding (10), but are relatively high in comparison to an environment such as a cryopeg (100). Finally, the mammalian gut is an anoxic environment in which most of the bacteria are anaerobic (101), whereas psychrophiles are frequently aerobic due to the high oxygen stress associated with low temperatures (102).

Here, I apply the same phenotypic tests consistently across many *Psychrobacter* strains with the aim of identifying broad patterns relating to *Psychrobacter*'s ecological diversification. I collected growth curves under 24 different conditions, including a range of temperatures and salinities with different basal media, from which I estimated growth rate (during exponential phase) and maximum optical density reached, as well as information on strain oxygen utilization. I compare patterns in these data by isolation source, as well as by ecotype (as defined in Chapter 2).

4.3 Materials and Methods.

I performed all analysis using R version 3.6.3; my notebooks are available on github at https://github.com/dkwelter/dkw_thesis_notebooks.

***Psychrobacter* growth screen data collection and modeling.** For a full description of the *Psychrobacter* strains used and the conditions included in my phenotypic screen, see Chapter 2. Briefly, I tested 85 strains of *Psychrobacter* under 24 different conditions, including 2 different media bases (LB and a modified M9 medium), 3 different temperatures (4, 25, and 37C), and 4 different salt concentrations (0, 2.5, 5, and 10% added NaCl). I randomly assigned strains to batches of 10 that were screened together (see Chapter 2 for more information on batches). After growing strains to saturation, I washed and diluted them to optical density at 600 nm (OD600) of 0.3, which I used to inoculate 100 uL of each medium at a ratio of 1:1000. In 96-well plates, I inoculated 5 replicate wells per strain, with 10 wells of uninoculated medium acting as blanks per plate. The wells around the periphery were filled with water to reduce edge effects. I inoculated three sets of 12 plates in this manner, and incubated each set at either 4, 25, or 37 ºC. I measured the OD600 of each plate every 8 hours for 1 week, then every 24 hours, and let cultures grow to stationary phase (between 3 and 12 weeks after inoculation).

At some points, it became necessary to restart a few plates per batch due to random catastrophe destroying the originals. Since no strains were present in multiple complete batches, a batch effect analysis was not possible.

For each separate replicate, I fitted the absorbance data to the logistic equation for growth with the Growth Curver package(103) in R to estimate growth rate r and maximum optical density K:

$$N_t = \frac{K}{1 + (\frac{K - N_0}{N_0}) * e^{-rt}}$$

where N_t is the population size at time t, N_0 is the initial population size, K is the maximum possible population size or carrying capacity, and r is the intrinsic growth rate (Figure 4.1 A).

I manually removed estimated data for growth curves that had not been fitted correctly or had not been fitted at all (Figure 4.1 B). I also binned each replicate as growth positive (meaning the strain grew in that replicate of that condition) or growth negative (meaning the strain never reached an OD600 of 0.15 over the course of the experiment) (Figure 4.1 C).

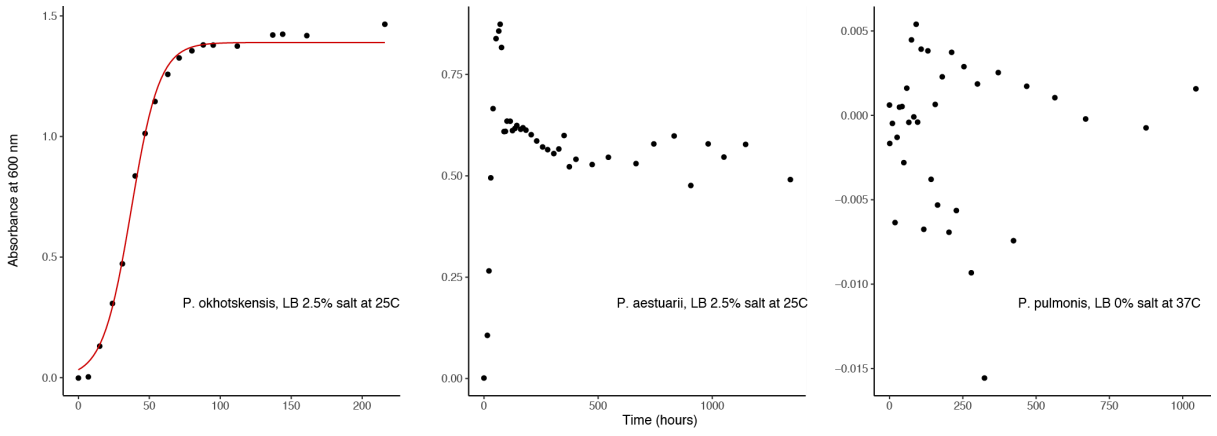

Figure 4.1 *Psychrobacter* growth curve modeling. Growth curves were collected for 85 different strains of Psychrobacter under 5 replicates of 24 different conditions. The time in hours is shown along the x-axis and the optical density at 600 nm is shown along the y-axis. The R package Growthcurver was used to model each growth curve. A) Representative of a well-modeled growth curve. The modeled logistic growth equation is shown in a red line which closely fits the data, with a low estimated error σ = 0.03. B) Representative of a growth curve that could not be modeled by Growthcurver. C) Representative of a condition in which there was no growth.

Analysis of *Psychrobacter* growth bin data. For every strain, for every condition, I collected growth bin (growth positive versus negative) data. However, because not every replicate growth curve could be modeled, some data were missing for K and r. In order to avoid “throwing out” growth bin data for conditions with missing K and r data, I analyzed growth bin data separately.

As the distribution for growth bin data was not normal (Figure 4.2 A), I used the non-parametric approach of non-metric multidimensional scaling (NMDS) to reduce the growth bin data into two dimensions using the metaMDS() function with Bray-Curtis distances from the R package vegan (104) with autotransform set to FALSE and trymax equal to 100, followed by a visualization of the fit between environmental variables with the ordination using the envfit() function - omitting strains with missing metadata - with 1000 permutations. The two-dimensional ordination resulted in a stress value of 0.087, indicating a fair to good fit, with a non-metric R^2 of 0.99 and a linear fit R^2 of 0.97 (Figure 4.2 B). I tested the association of strain and condition metadata with the growth bin data using an analysis of similarities (ANOSIM) with the anosim() function from vegan, using Bray distance and 9999 permutations.

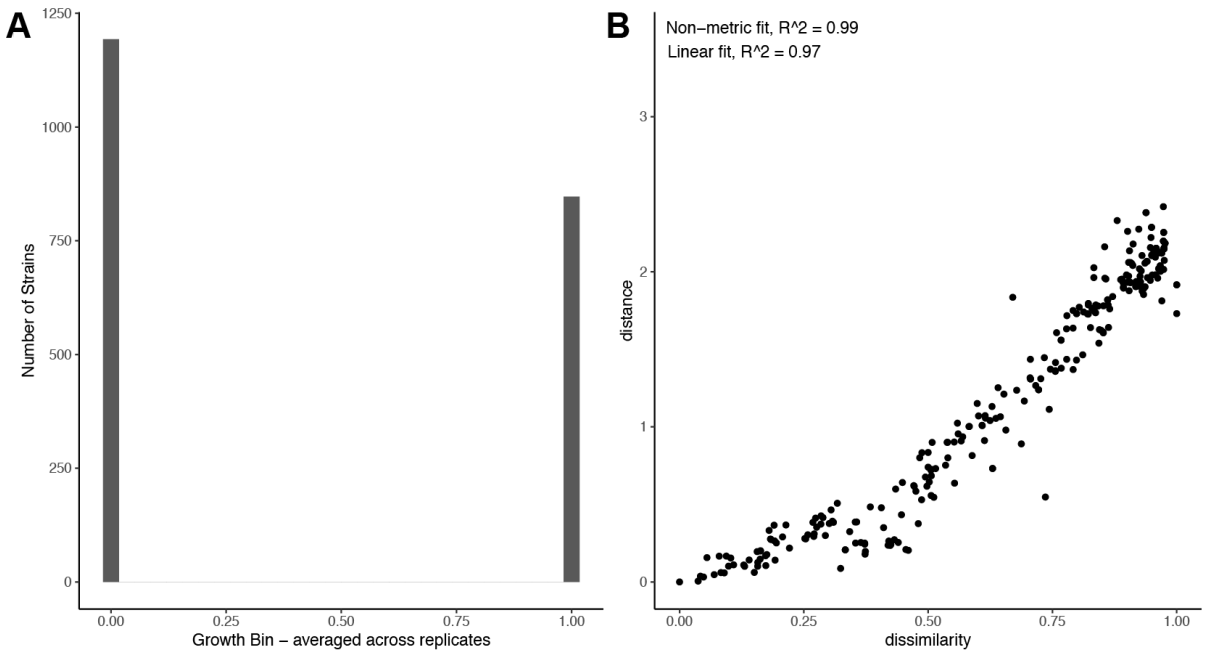

Figure 4.2 Non-parametric analysis of *Psychrobacter* growth bin data. A) Histogram of values of growth bin data. B) Shepard stressplot of the growth bin data, showing on the x axis dissimilarity between data points and on the y axis distances between those points in the NMDS ordination. The R^2 values for the non-metric and linear fits, calculated by stressplot() function from the vegan package in R, are shown in the top left.

After confirming that the isolation source significantly impacted growth bin data, I used chi square to test for differences in growth bin by condition. I filtered data to only the condition being tested, then downsampled such that each group was equal in size, repeating for 100 permutations of randomly included strains, adjusting p-values with the Benjamini-Hochberg correction.

Analysis of *Psychrobacter* growth rate and maximum OD600. The models for *Psychrobacter* growth rate (r) and maximum OD600 (K) were highly zero-inflated, since many strains could not grow in many conditions (Figure 4.3 A). I removed r values for conditions that didn't grow, because there is no way to use a value r = infinity in a model. K = 0 is biologically meaningful, so I included K values for no-growth conditions. I scaled the data by strain such that all values were between 0 and 1, and again used the NMDS approach to examine r and K with the same parameters as used in the growth bin analysis described above. The NMDS ordination resulted in a stress of 0.11, indicating a fair fit, with a

non-metric R2 of 0.99 and a linear R2 of 0.96 (Figure 4.3 B). I again visualized the fit of environmental variables to the ordination using envfit(), tested the association of metadata to k and r using the ANOSIM test.

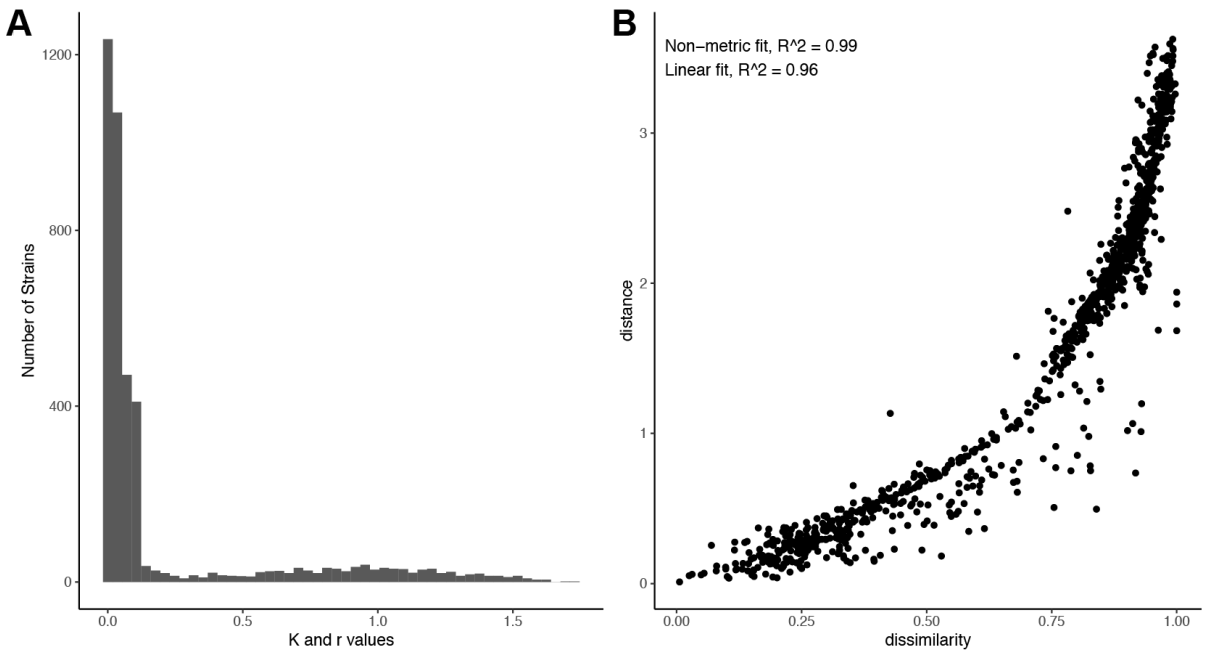

Figure 4.3 Non-parametric analysis of *Psychrobacter* growth rate and maximum optical density data. A) Histogram of values of growth curve modeled K (maximum OD) and r (intrinsic growth rate) data. B) Shepard stressplot of the K and r data, showing on the x axis dissimilarity between data points and on the y axis distances between those points in the NMDS ordination. The R^2 values for the non-metric and linear fits, calculated by stressplot() function from the vegan package in R, are shown in the top left.

***Psychrobacter* oxygen utilization assays.** I used the same growth curve cultures diluted to OD600 = 0.3 as inocula for oxygen requirements assays. To test the oxygen requirements of each strain, I used a modified thioglycollate broth (TGB) recipe. Per liter of TGB, I added 15 g pancreatic digest of casein, 5 g yeast extract, 5.5 g dextrose, 2.5 g sodium chloride, 0.5 g L-cysteine, 0.5 g sodium thioglycollate, 8 g agar, and 500 ug resazurin. I boiled the TGB until the agar dissolved, mixed well, then pipetted 8 mL each into 16 mm test tubes and autoclaved at 121 °C for 20 minutes for sterilization. When the TGB cooled and solidified, I inoculated each strain into a tube using 1 uL disposable inoculating loops, stabbing straight down in the center of the tube. I incubated the tubes at each strains' preferred temperature. When each tube showed visible growth, I scored the strain by comparison between the

growth zone and the oxygenated zone (figure TGB scoring) as either “AR” for aerobic, “MA” for microaerophilic, “AR,MA” for growth in both the aerobic and microaerophilic zones, “FA” for facultative anaerobic growth, or “ND” if there was no visible growth or the growth pattern was not clear. No strains showed solely anaerobic growth. If a strain did not grow before the pink “oxygenation zone” (visible due to the resazurin) expanded to the bottom of the tube, I marked the strain as “ND”.

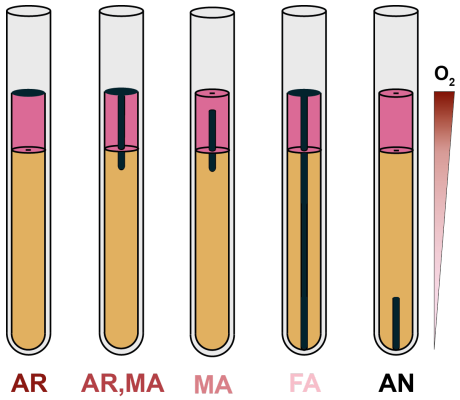

Figure 4.4 oxygen utilization assay scoring. *Psychrobacter* oxygen utilization was tested using a thioglycollate broth recipe containing resazurin as an oxygen indicator; the oxygenation zone is defined as the pink area at the top of slant. Oxygen utilization was defined in the following categories: AR = growth in aerobic zone only (on top of the slant). AR,MA = growth in aerobic zone as well as oxygenated portion of tube. MA = Growth within the oxygenated zone of the slant, but not within the aerobic or anaerobic zones. FA = growth on top of and throughout the slant. AN = growth only in the anaerobic zone of the slant.

To test the significance of differences between the behavior of *Psychrobacter* isolation sources and ecotypes in this assay, I performed a chi-square test with 100 random permutations of subsetted strains such that each of the groups were equal in size, and adjusted the p-values for multiple testing using the Benjamini-Hochberg method.

4.4 Results and Discussion.

***Psychrobacter* isolated from warm hosts are better suited to growing in rich media, low salt, and high temperatures than *Psychrobacter* from other isolation sources.** To explore what conditions were most impactful on *Psychrobacter* strains’ abilities to grow in my

phenotypic screen, I reduced the growth bin data into two dimensions with an NMDS analysis, followed by fitting metadata to the ordination using the envfit function from vegan.

Visualization of the NMDS variables and the fit between environmental factors and variables shows that most conditions cluster at low NMDS axis 1 values and mid NMDS axis 2 values, while some conditions do not (Figure 4.5). Notably, conditions at 37 ºC are significantly separated from the other conditions (temperature ANOSIM R = 0.42, significance = 0.0001). Base medium also has a large effect (medium ANOSIM R = 0.16, significance = 0.02), with base medium LB conditions mostly occurring in lower values of NMDS axis 2 and MM conditions mostly occurring in higher values of NMDS axis 2. Salt does not have a significant effect (ANOSIM R = 0.10, significance = 0.09).

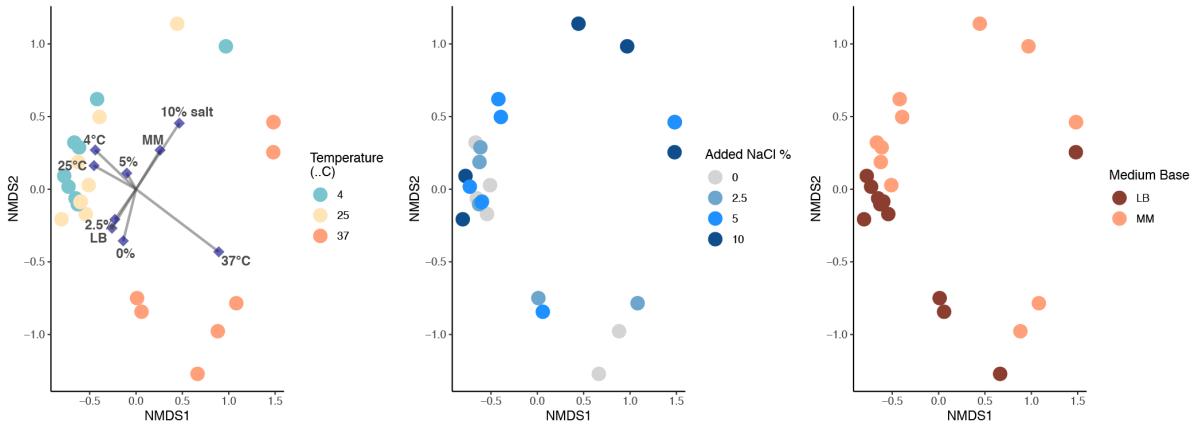

Figure 4.5 NMDS variables plot for growth bin. non-metric multidimensional scaling used to examine overall patterns of growth bin by growth conditions, the metaMDS() function from vegan with Bray-Curtis distances, autotransform = FALSE. Left-most: colored by temperature, with all conditions mapped to ordination using the envfit function. Middle plot: colored by % added sodium chloride. Right-most: colored by basal medium.

Visualization of the NMDS individuals reveals there is a relatively tight cluster of strains with low values on NMDS axis 1 and high values on NMDS axis 2, with a loose group of strains spreading out across NMDS axis 1 (Figure 4.6). The tight group represents strains from many different isolation sources, while the strains in the looser group are largely derived from warm bodied hosts. ANOSIM analysis comparing the growth bin data to strain isolation information confirms that isolation source is significant (ANOSIM R = 0.13, significance = 0.0003), however isolation location (latitude ANOSIM R = 0.27, significance =

0.007, longitude ANOSIM R = 0.26, significance = 0.03) also has a significant impact. Isolation location is confounded with isolation material, since samples from warm hosts were either isolated from high latitudes and low longitudes or have no isolation location data. Essentially, all of the strains from Antarctica, which are all either 'marine' or 'terrestrial' strains, cluster very closely. Ideally, to determine whether there is any impact of isolation location on strains' likely behavior in a phenotypic screen, I would like to have strains isolated from warm hosts located in the Antarctic, such as penguins (105) or seals (106), though to my knowledge, such strains are not (yet) available in strain catalogues.

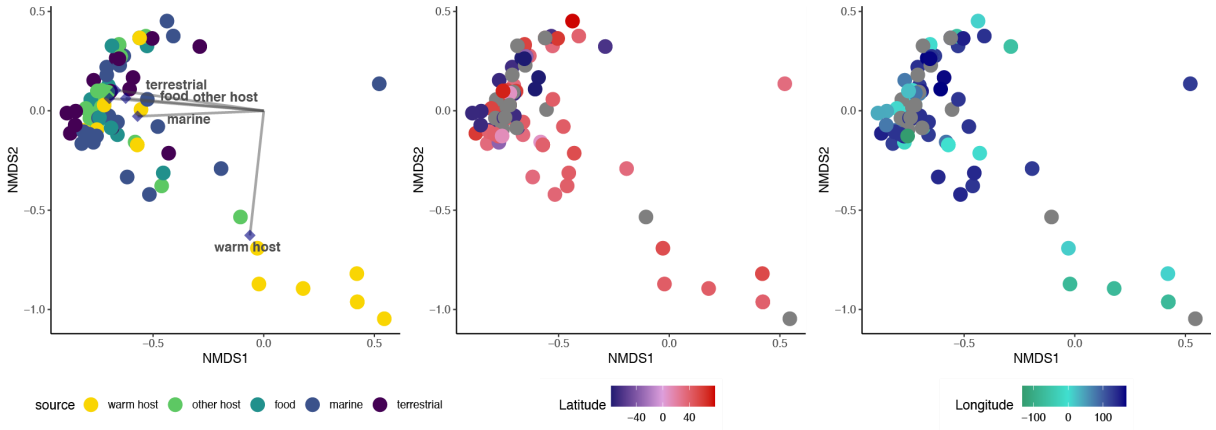

Figure 4.6 NMDS individuals plot for growth bin. non-metric multidimensional scaling used to examine overall patterns of growth bin by strain, the metaMDS() function from vegan with Bray-Curtis distances, autotransform = FALSE. Left-most: colored by isolation source. Isolation source mapped to ordination using the envfit function. Middle plot: colored by isolation latitude, with high latitudes (the Arctic) represented by red, and low latitudes (the Antarctic) represented by purple. Right-most: colored by isolation longitude.

After confirming that isolation source has a significant impact on growth bin values, I examined growth bin by isolation and growth condition; since isolation material and location were confounded, I decided to focus on isolation material (Figure 4.7). Unsurprisingly, *Psychrobacter* strains isolated from warm hosts are significantly more likely to grow in rich media (χ^2 (4, 1020) = 18.9, p-value = 0.0008) compared to other isolation materials, more likely to grow in low salt (χ^2 (4, 510) = 12.5, p-value = 0.01), less likely to grow at 4 ºC (χ^2 (4, 680) = 13.2, p-value = 0.01), and more likely to grow at 37 ºC (χ^2 (4, 680) = 84.9, p-value = 2e-17).

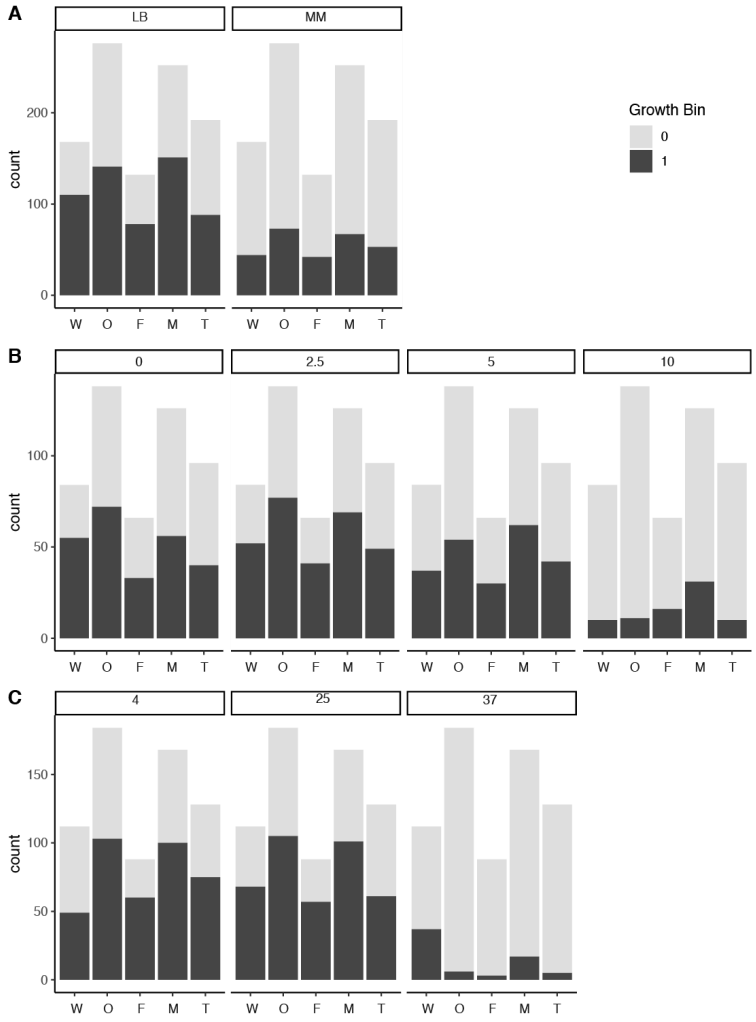

Figure 4.7 *Psychrobacter* isolation source is correlated with changes in growth bin. 85 *Psychrobacter* strains were tested for their ability to grow in 24 conditions with 5 replicates per condition, varying the basal medium, % added sodium chloride, and temperature of incubation. 0 = no growth in the majority of replicates for the condition, 1 = growth in the majority of replicates for the condition. Isolation material abbreviations: W = warm hosts, O = other hosts, F = food, M = marine, T = terrestrial.
A) Growth bin separated by medium: LB = Lysogeny Broth, a complex medium; MM = minimal medium, a defined medium based off of M9 salts. B) Growth bin separated by % sodium chloride added. C) Growth bin separated by temperature of incubation.

Overall, warm host derived strains follow some of the same patterns for growth bin as FE strains - strains capable of growth at 37 ºC - as described in Chapter 2, with improved likelihood of growth in rich media and warm temperatures. Unlike FE strains as a whole, warm host derived strains also exhibit improved likelihood of growth in low salt conditions and decreased likelihood of growth at low temperatures. Many of the warm host isolated strains are present in the basal clade of the *Psychrobacter* phylogeny, supporting the idea

that the *Psychrobacter* ancestor preferred conditions similar to those found in warm host bodies - nutritionally rich, little osmotic stress (relative to ocean water or glacial ice, that is), and warm temperatures. In comparison to even the most closely related *Moraxella* strains, *Psychrobacter* is more tolerant to cold temperatures and higher salt concentrations (107), and between the divergence of the basal and derived clades, *Psychrobacter* became even more halo- and psychrotolerant. Not all warm host derived strains behave alike; 4 strains - *P. frigidicola* strain *ACAM309*, *P. sp. JCM18900*, *P. sp. JCM18902*, and *P. sp. JCM18903* - were isolated from the guts of penguins and porpoises, but are categorized as RE strains, meaning they could not grow at 37 ºC with any combination of medium or salt. These strains are all members of the derived phylogenetic clade. Likely, their presence in the hosts was transient and due to host diet. Overall, though isolation source is a useful predictor of *Psychrobacter* ability to grow under particular conditions, it is not perfect.

***Psychrobacter* isolation source have little impact on growth rate and maximum optical density.** To explore the conditions of the growth screen that were most impactful on *Psychrobacter* strains' growth rate and maximum optical density reached, I performed an NMDS analysis on the modeled growth curve data, followed by envfit and ANOSIM analyses.

Visualization of the variables plot for NMDS shows that K and r are inversely correlated, with K variables falling in lower values for NMDS axis 1 and r variables falling in higher values for NMDS axis 1 (Figure 4.8). This is unsurprising, given that in ideal conditions, you would expect a strain to have a high carrying capacity (high K) and a fast growth rate (low r). However, what is surprising, is that all of the r values are tightly clustered, indicating that there is not much variation in the r data, while the K values are spread out. This suggests that *Psychrobacter* strains tend to grow at similar rates regardless of what conditions they are growing in, while the carrying capacity varies by condition. Similarly to the growth bin data, temperature had the largest effect, and was the only effect that was significant (ANOSIM R: 0.11, significance = 0.006). K values for conditions at 37 ºC

are the most similar to each other and most different from other conditions, with little to no separation between 4 and 25 ºC. While r values clustered tightly, there is a clear gradient from conditions at 4 ºC having lower NMDS axis 2 values to 37 ºC having higher NMDS axis 2 values. Neither base medium (ANOSIM R: 0.039, significance = 0.08) nor salt (ANOSIM R: -0.016, significance = 0.65) had significant effects, and there is very little separation by either variable in the variables plot. It is clear from looking at the NMDS plots there is not much variation in the data, and from the ANOSIM R values that the variation that exists is not attributable to the factors I measured.

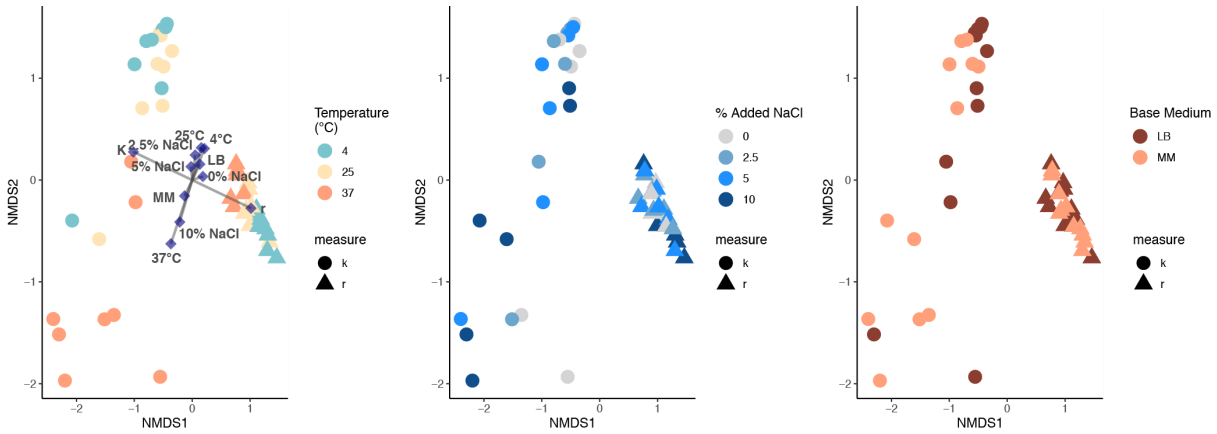

Figure 4.8 NMDS variables plot for growth rate and max OD. non-metric multidimensional scaling used to examine overall patterns of growth rate and maximum optical density reached compared to growth conditions, the metaMDS() function from vegan with Bray-Curtis distances, autotransform = FALSE. Left-most: colored by temperature, with all conditions mapped to ordination using the envfit function. Middle plot: colored by % added sodium chloride. Right-most: colored by basal medium.

Visualization of the individuals plot for NMDS reveals that there is very little structure outside of several strains - *P. sanguinis* strain *13983* and *P. sanguinis* strain *92*, *P. phenylpyruvicus*, *P. lutiphocae*, *P. aestuarii*, and *P. ciconiae* - having quite different behaviors and all other strains clustering tightly (Figure 4.9). The two *P. sanguinis* strains and *P. phenylpyruvicus* did not grow in any liquid culture, so it makes sense that they would be outliers. Conversely, *P. aestuarii* grew in nearly every condition, making it another easy outlier to explain. *P. lutiphocae* and *P. ciconiae* were remarkable only in that they were some of the few strains able to grow at 37 ºC, but did not show anywhere near the flexibility of *P. aestuarii*. Isolation source (ANOSIM R: 0.11, sig = 0.0008) and longitude (ANOSIM R: 0.21,

significance = 0.06) significantly affected the data clustering, while latitude did not (ANOSIM R: 0.12, sig = 0.12). The effect of isolation source seems largely driven by the strains that did not grow, which all happen to be isolated from the Western hemisphere, leading to a significant impact of longitude.

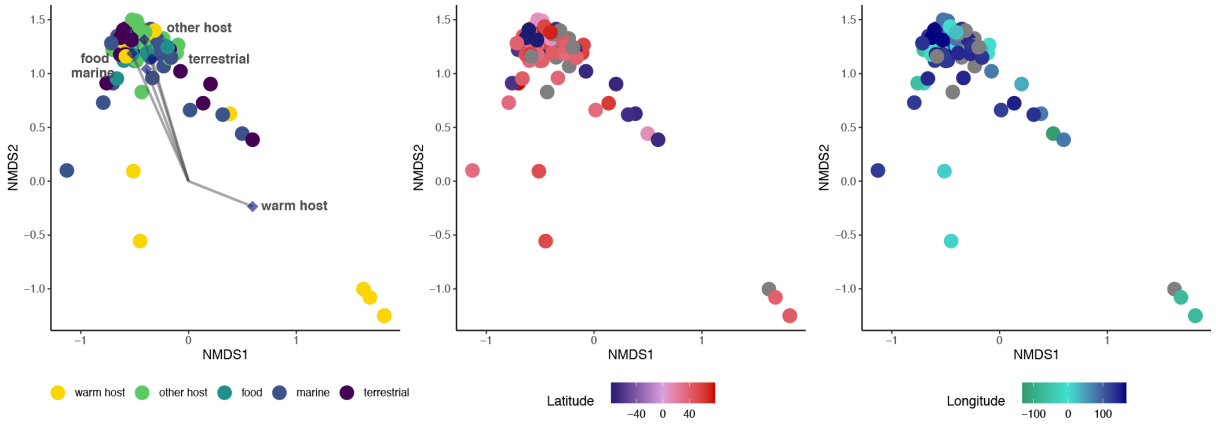

Figure 4.9 NMDS individuals plot for growth rate and max OD. non-metric multidimensional scaling used to examine overall patterns of growth rate and maximum optical density reached compared by strain, the metaMDS() function from vegan with Bray-Curtis distances, autotransform = FALSE. Left-most: colored by isolation source. Isolation source mapped to ordination using the envfit function. Middle plot: colored by isolation latitude, with high latitudes (the Arctic) represented by red, and low latitudes (the Antarctic) represented by purple. Right-most: colored by isolation longitude.

Because so little of the variation in the data was explainable by isolation source, I did not further characterize it through other tests, such as linear mixed modeling. While in theory, growth rate may be a predictor of bacterial behavior in communities (108, 109), I was not able to glean any useful information from the growth rates I collected. If I were to repeat this screen, it would be useful to include one strain in every batch, so that I would have a better idea of batch effect. Even better would be to remove batches entirely by incubating all strains in the same 96-well plates, but logistically, it seems impossible to coordinate the growth of that many strains for the inoculation.

***Psychrobacter* strains exhibit more diverse oxygen utilization than previously reported.** Though *Psychrobacter* has previously been reported as a strict aerobe, I was interested in whether there might be some variation in *Psychrobacter's* oxygen utilization

profile, so I conducted an oxygen utilization assay using thioglycollate broth. I defined isolates as either “aerobic”, “microaerophilic”, “aerobic/microaerophilic”, or “facultative anaerobic”, based on where in the broth they were able to grow (Figure 4.4). Though I defined no strains as strict anaerobes, there was still a surprising amount of variation in the oxygen use of the strains (Figure 4.10). There is no significant difference in the oxygen use categories by either isolation source or by ecotype (χ^2-test, all p-values adjusted for group size > 0.05), though isolation source trends towards significance (p-value = 0.1). In particular, the strains isolated from warm hosts seem enriched for microaerophiles - 43% of warm host isolates were microaerophilic versus 16% of all strains tested - and depleted in strict aerophiles - 7% of warm host isolates versus 26% of all strains tested - compared to other isolation sources. There was no trend for differences between FE and RE strains in oxygen usage.

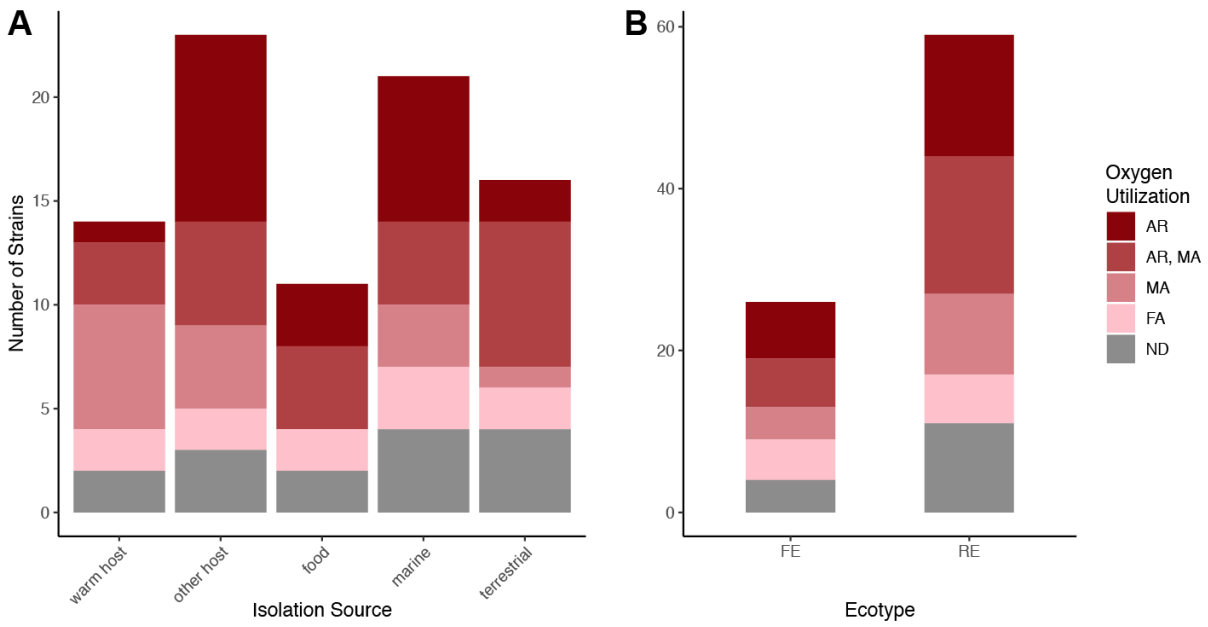

Figure 4.10 No significant differences in *Psychrobacter* oxygen utilization by isolation source or ecotype. Oxygen utilization was assessed using thioglycolate broth.
AR = aerobic growth only; AR,MA = aerobic and microaerophilic growth; MA = microaerophilic growth only; FA = facultative anaerobic growth; ND = no data (no visible growth).
A) oxygen utilization separated by isolation source. B) separated by ecotype.

The implications of oxygen use on *Psychrobacter’s* ecology remain unclear. Likely the apparent differences between warm host isolated strains and other isolation sources are

largely driven by the basal clade *Psychrobacter* strains, particularly *P. sanguinis* and *P. phenylpyruvicus*. *Moraxella* spp. are reported as aerobic, with some strains having weak anaerobic growth; likely the *Psychrobacter* ancestor behaved the same. The propensity of many *Psychrobacter* strains for aerobic growth may have to do with its evolved psychrotolerance; oxygen solubility is higher at low temperatures, increasing the need for oxidative stress responses in psychrophiles and psychrotrophs (110).

4.5 Conclusion

In this chapter I presented the results of a large-scale phenotypic screen for 85 strains of the genus *Psychrobacter*, including detailed information about oxygen usage, as well as growth capability, growth rate, and maximum optical density reached in 24 different conditions. In particular, I was interested in *Psychrobacter* strains isolated from warm-bodied hosts and whether their phenotypes may indicate fitness in a mammalian gut. *Psychrobacter* strains isolated from mammals and birds greatly prefer complex media to defined media, prefer low to mid-salt concentrations over medium-high or high, and prefer high temperatures over low ones. I found little explainable variation in growth rate or max optical density based on differing isolation sources. *Psychrobacter* strains isolated from warm-bodied hosts trend towards enrichment for microaerophilic status, but the effect is not significant. Overall, *Psychrobacter* strains isolated from warm hosts exhibit some distinct behaviors compared to strains from other isolation sources, but these differences seem largely driven by *P. phenylpyruvicus* and *P. sanguinis*, two species from the basal phylogenetic clade of *Psychrobacter* that demonstrate *Moraxella*-like tendencies. Isolation source is a valuable predictor of *Psychrobacter* phenotype and therefore ecological lifestyle, but perhaps not as useful as the ecotypes discussed in Chapter 2.

Because my genomic analysis and phenotyping scale were done in tandem, I did not test what may be the most relevant phenotypes for the genus based on genomic patterns. For example, in Chapter 3 I present evidence that strains isolated from food-processing environments may have increased biofilm formation capabilities; collecting data on biofilm

productivity would therefore have been interesting. For the data I did collect, I did not measure several variables that would have been helpful in my analysis, such as cell counts for the inocula or detailed lag time for the growth curves, or distance from the oxygenation zone in the thioglycollate assay. Further work could be done to characterize the phenotypic behaviors of *Psychrobacter* under diverse conditions.

Chapter 5. General Conclusion

Psychrobacter is a genus of closely related bacteria with a complex ecological distribution. Members can be isolated from environments such as the guts of mammals or marine sediment 6000 meters underwater, which raises questions about the evolutionary history of these microbes. It promises to be a unique model system to unravel the effects of niche differentiation on phenotype, genome content and pan-genome structuring.

In Chapter 2, I introduce the main body of my work in the form of a submitted manuscript, which details a phylogenomic analysis of the family *Moraxellaceae*, a wide-scale screen of *Psychrobacter* ability or inability to grow under diverse conditions, comparative genomics based on strain phenotypic behavior, and *Psychrobacter* ability to colonize mammalian guts. I found that phylogenomically, *Psychrobacter* are derived from *Moraxella*, implying that *Psychrobacter* evolved from a *Moraxella*-like, mammalian pathobiont. Dividing strains based on their ability to grow at 37 ^{o}C - into “flexible” (FE) strains versus “restricted” (RE) strains - reveals some broad genomic differences within the genus, such as genome size, number of pseudogenes, and diversity of iron-aquisition or lipid biosynthesis genes. While strains isolated from warm-bodied hosts like birds and mammals are more likely to be FE, strains from every other isolation source also display this phenotype. Both FE and RE *Psychrobacter* strains were detected in the stool of polar bears, but only FE strains could colonize germfree mice. Overall, my work suggests that the ability to grow at 37 ^{o}C is ancestral for *Psychrobacter*, which subsequently diverged into strains that favor cold environments versus strains that maintain some phenotypic flexibility.

I next present a comparative genomics analysis based on isolation source rather than phenotypic behavior in Chapter 3. I found that strains isolated from invertebrates and fish and food-processing sources have significantly higher gene numbers than other strains. There are also functional genome content differences depending on isolation source, but most of this unique gene content is due to cloud genes rather than a shared isolation source ‘core’ genome, and after controlling for population structure, there are only a few significant

genes correlated with isolation source. The biological and ecological significance of these genes is unclear, except for food-processing strains, which are significantly associated with pili, which could increase adhesion and survival in the food-processing process, and strains from fish and invertebrates, which are significantly associated with DNA replication and transcription regulation, which could point to increased horizontal gene transfer. There were no significant differences in levels of protein cold adaptation between isolation sources.

Finally, in Chapter 4, I present a more detailed analysis of the phenotypic screen I present in Chapter 2. I examined the growth curve data by isolation source, and found that *Psychrobacter* strains isolated from warm hosts share the broad characteristics of FE strains; they are more likely to grow in complex media than defined media, at low salt concentrations than high, and at high temperatures rather than low. There is not much variation in growth rate or maximum optical density reached by isolation source, but isolation source, rather than ecotype, seems to impact strain oxygen use. Overall these differences seem to be driven by the phylogenetically basal clade of *Psychrobacter* spp.

My work is the most comprehensive examination of the genus *Psychrobacter* to date, both in comparative genomics and in *in vitro* and *in vivo* growth phenotypes. I offer a critical examination of the evolution and ecology *Psychrobacter* in the context of its family and its behavior in mammalian colonization capability. Sources of isolation for *Psychrobacter* strains are correlated with some differences genomically and phenotypically, but dividing strains by their behavior may be more meaningful.

There is still much to learn about *Psychrobacter*. The genes or gene regulation responsible for the remarkable generalism of FE strains could be elucidated with transcriptomics studies comparing their growth under cold versus warm temperatures, or even in vitro growth versus their growth in mouse guts. This would be especially interesting using FE strains from both the basal phylogenetic clade versus the derived. The largest genomic and phenotypic differences in the genus are between the basal and derived clades; more detailed analysis of what makes the basal *Psychrobacter* strains different from their very close Moraxella relatives, in particular, what increases their psychro- and halotolerance,

would be be revealing for *Psychrobacter*'s evolutionary history. Finally, there is much to be learned from already published data. Metagenomics datasets available through databases such as MGnify (111) have been used to study the ecological distributions of other microbes (112); the same approach could examine the differences in distributions between *Psychrobacter*, *Moraxella*, and *Acinetobacter* to further clarify the differences in lifestyle between these closely related taxa.

References

1. Wei SL, Young RE. 1989. Development of symbiotic bacterial bioluminescence in a nearshore cephalopod, Euprymna scolopes. Mar Biol 103:541–546.

2. Shabat SKB, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Berg Miller ME, White BA, Shterzer N, Mizrahi I. 2016. Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. ISME J 10:2958–2972.

3. Goodrich JK, Davenport ER, Clark AG, Ley RE. 2017. The Relationship Between the Human Genome and Microbiome Comes into View. Annu Rev Genet 51:413–433.

4. Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, Cryan JF, Gilbert SF, Goodnight CJ, Lloyd EA, Sapp J, Vandenkoornhuyse P, Zilber-Rosenberg I, Rosenberg E, Bordenstein SR. 2016. Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. mSystems 1.

5. Kokou F, Sasson G, Nitzan T, Doron-Faigenboim A, Harpaz S, Cnaani A, Mizrahi I. 2018. Host genetic selection for cold tolerance shapes microbiome composition and modulates its response to temperature. Elife 7.

6. Ochman H, Worobey M, Kuo C-H, Ndjango J-BN, Peeters M, Hahn BH, Hugenholtz P. 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8:e1000546.

7. Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, Walter J. 2010. Diversification of the gut symbiont Lactobacillus reuteri as a result of host-driven evolution. ISME J 4:377–387.

8. Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, Ley RE, Farnleitner AH. 2019. Host diet and evolutionary history explain different aspects of gut

microbiome diversity among vertebrate clades. Nat Commun 10:2200.

9. Kartzinel TR, Hsing JC, Musili PM, Brown BRP, Pringle RM. 2019. Covariation of diet and gut microbiome in African megafauna. Proc Natl Acad Sci U S A 116:23588–23593.

10. Lachnit T, Bosch TCG, Deines P. 2019. Exposure of the Host-Associated Microbiome to Nutrient-Rich Conditions May Lead to Dysbiosis and Disease Development-an Evolutionary Perspective. MBio 10.

11. McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. 2016. Host Selection of Microbiota via Differential Adhesion. Cell Host Microbe 19:550–559.

12. Sender R, Fuchs S, Milo R. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 14:e1002533.

13. Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. 2016. Healthy human gut phageome. Proc Natl Acad Sci U S A 113:10400–10405.

14. Cullender TC, Chassaing B, Janzon A, Kumar K, Muller CE, Werner JJ, Angenent LT, Bell ME, Hay AG, Peterson DA, Walter J, Vijay-Kumar M, Gewirtz AT, Ley RE. 2013. Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. Cell Host Microbe 14:571–581.

15. Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124:837–848.

16. van Overbeek LS, van Doorn J, Wichers JH, van Amerongen A, van Roermund HJW, Willemsen PTJ. 2014. The arable ecosystem as battleground for emergence of new human pathogens. Front Microbiol 5:104.

17. Sakib SN, Reddi G, Almagro-Moreno S. 2018. Environmental role of pathogenic traits in Vibrio cholerae. J Bacteriol 200:e00795–17.

18. Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. Proc Natl Acad Sci U S A 108 Suppl 2:10800–10807.

19. Juni E, Heym GA. 1986. Psychrobacter immobilis gen. nov., sp. nov.: Genospecies Composed of Gram-Negative, Aerobic, Oxidase-Positive Coccobacilli. Int J Syst Evol Microbiol 36:388–391.

20. 2015. Psychrobacter, p. 1–10. *In* Whitman, WB, Rainey, F, Kämpfer, P, Trujillo, M, Chun, J, DeVos, P, Hedlund, B, Dedysh, S (eds.), Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Ltd, Chichester, UK.

21. Nelson TM, Rogers TL, Carlini AR, Brown MV. 2013. Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. Environ Microbiol 15:1132–1145.

22. Apprill A, Miller CA, Moore MJ, Durban JW, Fearnbach H, Barrett-Lennard LG. 2017. Extensive Core Microbiome in Drone-Captured Whale Blow Supports a Framework for Health Monitoring. mSystems 2.

23. Almeida M, Hébert A, Abraham A-L, Rasmussen S, Monnet C, Pons N, Delbès C, Loux V, Batto J-M, Leonard P, Kennedy S, Ehrlich SD, Pop M, Montel M-C, Irlinger F, Renault P. 2014. Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis of dairy fermented products. BMC Genomics 15:1101.

24. Pearce DA, Newsham KK, Thorne MAS, Calvo-Bado L, Krsek M, Laskaris P, Hodson A, Wellington EM. 2012. Metagenomic analysis of a southern maritime antarctic soil. Front Microbiol 3:403.

25. Vela AI, Collins MD, Latre MV, Mateos A, Moreno MA, Hutson R, Domínguez L, Fernández-Garayzábal JF. 2003. Psychrobacter pulmonis sp. nov., isolated from the lungs of lambs. Int J Syst Evol Microbiol 53:415–419.

26. Lloyd-Puryear M, Wallace D, Baldwin T, Hollis DG. 1991. Meningitis caused by Psychrobacter immobilis in an infant. J Clin Microbiol 29:2041–2042.

27. Wirth SE, Ayala-del-Río HL, Cole JA, Kohlerschmidt DJ, Musser KA, Sepúlveda-Torres L del C, Thompson LM, Wolfgang WJ. 2012. Psychrobacter sanguinis sp. nov., recovered from four clinical specimens over a 4-year period. Int J Syst Evol Microbiol 62:49–54.

28. Ortiz-Alcántara JM, Segura-Candelas JM, Garcés-Ayala F, Gonzalez-Durán E, Rodríguez-Castillo A, Alcántara-Pérez P, Wong-Arámbula C, González-Villa M, León-Ávila G, García-Chéquer AJ, Diaz-Quiñonez JA, Méndez-Tenorio A, Ramírez-González JE. 2016. Fatal Psychrobacter sp. infection in a pediatric patient with meningitis identified by metagenomic next-generation sequencing in cerebrospinal fluid. Arch Microbiol 198:129–135.

29. Yumoto I, Hirota K, Sogabe Y, Nodasaka Y, Yokota Y, Hoshino T. 2003. Psychrobacter okhotskensis sp. nov., a lipase-producing facultative psychrophile isolated from the coast of the Okhotsk Sea. Int J Syst Evol Microbiol 53:1985–1989.

30. Fondi M, Orlandini V, Perrin E, Maida I, Bosi E, Papaleo MC, Michaud L, Lo Giudice A, de Pascale D, Tutino ML, Liò P, Fani R. 2014. Draft genomes of three Antarctic Psychrobacter strains producing antimicrobial compounds against Burkholderia cepacia complex, opportunistic human pathogens. Mar Genomics 13:37–38.

31. Kamelamela N, Zalesne M, Morimoto J, Robbat A, Wolfe BE. 2018. Indigo- and indirubin-producing strains of Proteus and Psychrobacter are associated with purple rind defect in a surface-ripened cheese. Food Microbiol 76:543–552.

32. Bakermans C. 2018. Adaptations to marine versus terrestrial low temperature environments as revealed by comparative genomic analyses of the genus Psychrobacter. FEMS Microbiol Ecol 94.

33. Tettelin H, Riley D, Cattuto C, Medini D. 2008. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472–477.

34. Polz MF, Alm EJ, Hanage WP. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet 29:170–175.

35. Lassalle F, Muller D, Nesme X. 2015. Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. Res Microbiol 166:729–741.

36. Antczak M, Michaelis M, Wass MN. 2019. Environmental conditions shape the nature of a minimal bacterial genome. Nat Commun 10:3100.

37. Han K, Li Z-F, Peng R, Zhu L-P, Zhou T, Wang L-G, Li S-G, Zhang X-B, Hu W, Wu Z-H, Qin N, Li Y-Z. 2013. Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. Sci Rep 3:2101.

38. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141–161.

39. Chaudhuri RR, Henderson IR. 2012. The evolution of the Escherichia coli phylogeny. Infect Genet Evol 12:214–226.

40. Liao J, Orsi RH, Carroll LM, Wiedmann M. 2020. Comparative genomics reveals different population structures associated with host and geographic origin in antimicrobial-resistant Salmonella enterica. Environ Microbiol 22:2811–2828.

41. Miller JJ, Weimer BC, Timme R, Lüdeke CHM, Pettengill JB, Bandoy DD, Weis AM, Kaufman J, Huang BC, Payne J, Strain E, Jones JL. 2020. Phylogenetic and Biogeographic Patterns of Vibrio parahaemolyticus from North America Inferred from Whole Genome Sequence Data. Appl Environ Microbiol

https://doi.org/10.1128/AEM.01403-20.

42. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774.

43. 2011. Microbiology by numbers. Nat Rev Microbiol 9:628.

44. Cen S, Yin R, Mao B, Zhao J, Zhang H, Zhai Q, Chen W. 2020. Comparative genomics shows niche-specific variations of Lactobacillus plantarum strains isolated from human, Drosophila melanogaster, vegetable and dairy sources. Food Bioscience 35:100581.

45. Sun Z, Harris HMB, McCann A, Guo C, Argimón S, Zhang W, Yang X, Jeffery IB, Cooney JC, Kagawa TF, Liu W, Song Y, Salvetti E, Wrobel A, Rasinkangas P, Parkhill J, Rea MC, O'Sullivan O, Ritari J, Douillard FP, Paul Ross R, Yang R, Briner AE, Felis GE, de Vos WM, Barrangou R, Klaenhammer TR, Caufield PW, Cui Y, Zhang H, O'Toole PW. 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat Commun 6:8322.

46. Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole PW. 2006. Diversity of the genus Lactobacillus revealed by comparative genomics of five species. Microbiology 152:3185–3196.

47. Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol 70:2782–2858.

48. Ayala-del-Río HL, Chain PS, Grzymski JJ, Ponder MA, Ivanova N, Bergholz PW, Di Bartolo G, Hauser L, Land M, Bakermans C, Rodrigues D, Klappenbach J, Zarka D, Larimer F, Richardson P, Murray A, Thomashow M, Tiedje JM. 2010. The genome sequence of Psychrobacter arcticus 273-4, a psychroactive Siberian permafrost

bacterium, reveals mechanisms for adaptation to low-temperature growth. Appl Environ Microbiol 76:2304–2312.

49. Papaleo MC, Fondi M, Maida I, Perrin E, Lo Giudice A, Michaud L, Mangano S, Bartolucci G, Romoli R, Fani R. 2012. Sponge-associated microbial Antarctic communities exhibiting antimicrobial activity against Burkholderia cepacia complex bacteria. Biotechnol Adv 30:272–293.

50. Zhang S, Song W, Yu M, Lin X. 2017. Comparative genomics analysis of five Psychrobacter strains isolated from world-wide habitats reveal high intra-genus variations. Extremophiles https://doi.org/10.1007/s00792-017-0927-1.

51. Genus: Psychrobacter.

52. Genome.

53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.

54. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

55. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

56. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36.

57. Ding W, Baumdicker F, Neher RA. 2018. panX: pan-genome analysis and exploration.

Nucleic Acids Res 46:e5.

58. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304.

59. Snipen L, Liland KH. 2015. micropan: an R-package for microbial pan-genomics. BMC Bioinformatics 16:79.

60. Collins C, Didelot X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 14:e1005958.

61. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132.

62. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–223.

63. NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8–D13.

64. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

65. Paradis E. 2012. Analysis of Phylogenetics and Evolution with R. Springer, New York, NY.

66. HowTo/Ancestral State Reconstruction.

67. Collins RE, Higgs PG. 2012. Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. Mol Biol Evol 29:3413–3425.

68. Luke NR, Howlett AJ, Shao J, Campagnari AA. 2004. Expression of type IV pili by Moraxella catarrhalis is essential for natural competence and is affected by iron

limitation. Infect Immun 72:6262–6270.

69. Luke NR, Jurcisek JA, Bakaletz LO, Campagnari AA. 2007. Contribution of Moraxella catarrhalis type IV pili to nasopharyngeal colonization and biofilm formation. Infect Immun 75:5559–5564.

70. de Vries SPW, Bootsma HJ, Hays JP, Hermans PWM. 2009. Molecular aspects of Moraxella catarrhalis pathogenesis. Microbiol Mol Biol Rev 73:389–406, Table of Contents.

71. Oliveira PH, Touchon M, Rocha EPC. 2016. Regulation of genetic flux between bacteria by restriction-modification systems. Proc Natl Acad Sci U S A 113:5658–5663.

72. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J, Ferriera S, Lapidus A, Anderson I, Kyrpides N, Munk AC, Detter C, Han CS, Brown MV, Robb FT, Kjelleberg S, Cavicchioli R. 2009. The genomic basis of trophic strategy in marine bacteria. Proc Natl Acad Sci U S A 106:15527–15533.

73. Bradley PH, Nayfach S, Pollard KS. 2018. Phylogeny-corrected identification of microbial gene families relevant to human gut colonization. PLoS Comput Biol 14:e1006242.

74. Melnyk RA, Hossain SS, Haney CH. 2019. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. ISME J 13:1575–1588.

75. Fritsch L, Felten A, Palma F, Mariet J-F, Radomski N, Mistou M-Y, Augustin J-C, Guillier L. 2019. Insights from genome-wide approaches to identify variants associated to phenotypes at pan-genome scale: Application to L. monocytogenes' ability to grow in cold conditions. Int J Food Microbiol 291:181–188.

76. Edwards RA, Puente JL. 1998. Fimbrial expression in enteric bacteria: a critical step in intestinal pathogenesis. Trends Microbiol 6:282–287.

77. Dibb-Fuller MP, Allen-Vercoe E, Thorns CJ, Woodward MJ. 1999. Fimbriae- and flagella-mediated association with and invasion of cultured epithelial cells by Salmonella enteritidis. Microbiology 145 (Pt 5):1023–1031.

78. Avalos Vizcarra I, Hosseini V, Kollmannsberger P, Meier S, Weber SS, Arnoldini M, Ackermann M, Vogel V. 2016. How type 1 fimbriae help Escherichia coli to evade extracellular antibiotics. Sci Rep 6:18109.

79. Schroll C, Barken KB, Krogfelt KA, Struve C. 2010. Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation. BMC Microbiol 10:179.

80. Van Houdt R, Michiels CW. 2010. Biofilm formation and the food industry, a focus on the bacterial outer surface. J Appl Microbiol 109:1117–1131.

81. Nicolas E, Lambin M, Dandoy D, Galloy C, Nguyen N, Oger CA, Hallet B. 2015. The Tn3-family of Replicative Transposons. Microbiol Spectr 3.

82. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38:865–891.

83. Bowman JP. 2017. Genomics of Psychrophilic Bacteria and Archaea, p. 345–387. *In* Margesin, R (ed.), Psychrophiles: From Biodiversity to Biotechnology. Springer International Publishing, Cham.

84. Collins T, Margesin R. 2019. Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. Appl Microbiol Biotechnol 103:2857–2871.

85. Revell LJ, Harmon LJ, Collar DC. 2008. Phylogenetic signal, evolutionary process, and rate. Syst Biol 57:591–601.

86. Feng S, Powell SM, Wilson R, Bowman JP. 2014. Extensive gene acquisition in the extremely psychrophilic bacterial species Psychroflexus torquis and the link to sea-ice ecosystem specialism. Genome Biol Evol 6:133–148.

87. Raymond-Bouchard I, Goordial J, Zolotarov Y, Ronholm J, Stromvik M, Bakermans C, Whyte LG. 2018. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. FEMS Microbiol Ecol 94.

88. Porankiewicz J, Clarke AK. 1997. Induction of the heat shock protein ClpB affects cold acclimation in the cyanobacterium Synechococcus sp. strain PCC 7942. J Bacteriol 179:5111–5117.

89. Raymond-Bouchard I, Tremblay J, Altshuler I, Greer CW, Whyte LG. 2018. Comparative Transcriptomics of Cold Growth and Adaptive Features of a Eury- and Steno-Psychrophile. Front Microbiol 9:1565.

90. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proceedings of the National Academy of Sciences 108:2831–2836.

91. Hehemann J-H, Arevalo P, Datta MS, Yu X, Corzett CH, Henschel A, Preheim SP, Timberlake S, Alm EJ, Polz MF. 2016. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. Nat Commun 7:12860.

92. Pace NR. 2009. Mapping the tree of life: progress and prospects. Microbiol Mol Biol Rev 73:565–576.

93. International Journal of Systematic and Evolutionary Microbiology – About.

94. Yassin AF, Busse H-J. 2009. Psychrobacter lutiphocae sp. nov., isolated from the faeces of a seal. Int J Syst Evol Microbiol 59:2049–2053.

95. Clarke A, Rothery P. 2007. Scaling of body temperature in mammals and birds. Funct Ecol 0:071029083929001–???

96. Duxbury AC, Byrne RH, Mackenzie FT. 2020. Temperature distribution. Encyclopædia Britannica. Brittanica.

97. Waymouth C. 1970. Osmolality of mammalian blood and of media for culture of mammalian cells. In Vitro 6:109–127.

98. Duxbury AC, Byrne RH, Mackenzie FT. 2020. Salinity distribution. Encyclopædia Britannica. Brittanica.

99. Gilichinsky D, Rivkina E, Shcherbakova V, Laurinavichuis K, Tiedje J. 2003. Supercooled Water Brines Within Permafrost—An Unknown Ecological Niche for Microorganisms: A Model for Astrobiology. Astrobiology 3:331–341.

100. Gilichinsky D, Rivkina E, Bakermans C, Shcherbakova V, Petrovskaya L, Ozerskaya S, Ivanushkina N, Kochkina G, Laurinavichuis K, Pecheritsina S, Fattakhova R, Tiedje JM. 2005. Biodiversity of cryopegs in permafrost. FEMS Microbiol Ecol 53:117–128.

101. Friedman ES, Bittinger K, Esipova TV, Hou L, Chau L, Jiang J, Mesaros C, Lund PJ, Liang X, FitzGerald GA, Goulian M, Lee D, Garcia BA, Blair IA, Vinogradov SA, Wu GD. 2018. Microbes vs. chemistry in the origin of the anaerobic gut lumen. Proc Natl Acad Sci U S A 115:4170–4175.

102. D'Amico S, Collins T, Marx J-C, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep 7:385–389.

103. Sprouffske K, Wagner A. 2016. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics 17:172.

104. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930.

105. Kämpfer P, Glaeser SP, Irgang R, Fernández-Negrete G, Poblete-Morales M, Fuentes-Messina D, Cortez-San Martín M, Avendaño-Herrera R. 2020. Psychrobacter pygoscelis sp. nov. isolated from the penguin Pygoscelis papua. Int J Syst Evol Microbiol 70:211–219.

106. Banks JC, Craig Cary S, Hogg ID. 2014. Isolated faecal bacterial communities found for Weddell seals, Leptonychotes weddellii, at White Island, McMurdo Sound, Antarctica. Polar Biol 37:1857–1864.

107. 2015. Moraxella, p. 1–17. *In* Whitman, WB, Rainey, F, Kämpfer, P, Trujillo, M, Chun, J, DeVos, P, Hedlund, B, Dedysh, S (eds.), Bergey's Manual of Systematics of Archaea and Bacteria. John Wiley & Sons, Ltd, Chichester, UK.

108. Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, Matot E, Jona G, Harmelin A, Cohen N, Sirota-Madi A, Thaiss CA, Pevsner-Fischer M, Sorek R, Xavier R, Elinav E, Segal E. 2015. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science 349:1101–1106.

109. Tramontano M, Andrejev S, Pruteanu M, Klünemann M, Kuhn M, Galardini M, Jouhten P, Zelezniak A, Zeller G, Bork P, Typas A, Patil KR. 2018. Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. Nat Microbiol 3:514–522.

110. Tribelli PM, López NI. 2018. Reporting Key Features in Cold-Adapted Bacteria. Life 8.

111. Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, Crusoe MR, Kale V, Potter SC, Richardson LJ, Sakharova E, Scheremetjew M, Korobeynikov A, Shlemov A, Kunyavskaya O, Lapidus A, Finn RD. 2019. MGnify: the microbiome analysis resource in 2020. Nucleic Acids Res 48:D570–D578.

112. de la Cuesta-Zuluaga J, Spector TD, Youngblut ND, Ley RE. 2020. Genomic insights into adaptations of TMA-utilizing methanogens to diverse habitats including the human gut. Cold Spring Harbor Laboratory.

Appendix A:

Supplementary materials for Chapter 2

Tables S1 - S4.

Table S1. Moraxellaceae genomes and their characteristics.

Species	Strain	NCBI Genome	Isolation	Citation	Type Strain Temperature	
					Minimum	Maximum
Acinetobacter_baumanni	AB30	GCF_000746645.1_ASM74664v1_genomic.fna	human surgical infection	69	15	44
Acinetobacter_boissieri	ANC 4422	GCF_900096955.1_IMG-taxon_2671180229_annotated_assembly_genomic.fna	floral nectar	68	4	30
Acinetobacter_calcoaceticus	CA16	GCF_002055515.1_ASM205551v1_genomic.fna	soil	69	15	37
Acinetobacter_celticus	ANC 4603	GCF_001707755.1_ASM170775v1_genomic.fna	fresh water	67	1	30
Acinetobacter_colistiniresistens	NIPH 2036	GCF_000413935.1_Acin_sp_NIPH_2036_V1_genomic.fna	human catheter	66	15	37
Acinetobacter_courvalinii	CCUG 67960	GCF_008802255.1_ASM880225v1_genomic.fna	lizard conjunctiva	65	15	40
Acinetobacter_dispersus	NCCP 16014	GCF_009884975.1_ASM988497v1_genomic.fna	human sputum	65	15	37
Acinetobacter_equi	strain 114	GCF_001307195.1_ASM130719v1_genomic.fna	horse feces	81	27	37
Acinetobacter_gandensis	ANC 4275	GCF_001678755.1_ASM167875v1_genomic.fna	horse	64	25	37
Acinetobacter_gyllenbergii	NIPH 230	GCF_000488195.1_Acin_gyll_NIPH_230_V1_genomic.fna	human vagina	63	25	37
Acinetobacter_indicus	CIP 110367	GCF_000488255.1_Acin_indi_CIP110367_V2_genomic.fna	hexachlorocyclohexane dump site	62	22	42
Acinetobacter_kookii	ANC 4667	GCF_900096895.1_IMG-taxon_2671180219_annotated_assembly_genomic.fna	soil	61	24	41
Acinetobacter_piscicola	LW15	GCF_002233755.1_ASM223375v1_genomic.fna	diseased cod	60	4	32
Acinetobacter_proteolyticus	ANC 3849	GCF_001753605.1_ASM175360v1_genomic.fna	human wound	65	15	37
Acinetobacter_puyangensis	ANC 4466	GCF_900096995.1_IMG-taxon_2671180233_annotated_assembly_genomic.fna	tree bark	59	10	41
Acinetobacter_radioresistens	DSM 6976	GCF_000368905.1_Acin_radi_CIP_103788_V1_genomic.fna	cotton tampon	58	27	42
Acinetobacter_schindleri	strain ACE	GCF_001971565.1_ASM197156v1_genomic.fna	laboratory	57	30	41
Acinetobacter_venetianus	VE-C3	GCF_000308235.1_ASM30823v2_genomic.fna	lagoon water	56	30	37
Moraxella_atlantae	NBRC 14588	GCF_001591265.1_ASM159126v1_genomic.fna	human blood	55	22	37
Moraxella_boevrei	DSM 14165	GCF_000379845.1_ASM37984v1_genomic.fna	goat nasal cavity	54	33	35
Moraxella_bovis	Epp63 -300	GCF_003287015.1_ASM328701v1_genomic.fna	cattle conjunctivitis	53	30	37
Moraxella_bovoculi	strain 58069	GCF_000988605.1_ASM98860v1_genomic.fna	cattle conjunctivitis	52	25	42
Moraxella_canis	CCUG 8415A	GCF_002014965.1_ASM201496v1_genomic.fna	dog bite wound	51	25	37
Moraxella_caprae	DSM 19149	GCF_000426885.1_ASM42688v1_genomic.fna	goat nasal cavity	50	30	37
Moraxella_catarrhalis	BBH18	GCF_000092265.1_ASM9226v1_genomic.fna	human respiratory tract	44	22	37
Moraxella_caviae	CCUG 355	GCF_002014985.1_ASM201498v1_genomic.fna	guinea pig pharynx	49	22	37
Moraxella_cuniculi	DSM 21768	GCF_900156515.1_IMG-taxon_2681812931_annotated_assembly_genomic.fna	rabit oral mucosa	48	22	37
Moraxella_equi	NCTC 11012	GCF_900453335.1_48853_C02_genomic.fna	equine eye	47	35	37
Moraxella_lacunata	NBRC 102154	GCF_001591245.1_ASM159124v1_genomic.fna	human conjunctivitis	45	35	37
Moraxella_lincolnii	CCUG 9405	GCF_002014765.1_ASM201476v1_genomic.fna	human nasal cavity	46	28	37
Moraxella_nonliquefaciens	CCUG 348	GCF_001679005.1_ASM167900v1_genomic.fna	human nasal cavity	43	25	37
Moraxella_oblonga	NBRC 102422	GCF_001598275.1_ASM159827v1_genomic.fna	sheep oral cavity	42	35	37
Moraxella_osloensis	CCUG 350	GCF_001553955.1_ASM155395v1_genomic.fna	human cerebrospinal fluid	43	25	42
Moraxella_ovis	199/55	GCF_001636015.1_ASM163601v1_genomic.fna	conjunctivitis	41	25	37
Moraxella_pluranimalium	CCUG 54913	GCF_002014825.1_ASM201482v1_genomic.fna	pig nasal cavity	40	22	37
Moraxella_porci	DSM 25326	GCF_002014855.1_ASM201485v1_genomic.fna	pig brain, meningitis	39	22	42
Psychrobacter_aestuarii	JCM 16343	-	marine soil	71	4	37
Psychrobacter_celer	DSMZ 23510	-	sea water	24	4	40
Psychrobacter_ciconiae	DSMZ 28412	-	stork throat	21	10	40
Psychrobacter_faecalis	-	-	polar bear feces	73	4	36
Psychrobacter_frigidicola	DSMZ 12411	-	ornithogenic soil	72	0	22
Psychrobacter_glacincola	DSMZ 12194	-	sea water	25	0	22
Psychrobacter_immobilis	DSMZ 7229	-	frozen meat	76	5	25
Psychrobacter_lutiphocae	DSMZ 21542	-	seal feces	20	10	37
Psychrobacter_maritimus	DSMZ 15397	-	sea water	74	4	37
Psychrobacter_namhaensis	DSMZ 16330	-	sea water	77	4	37
Psychrobacter_okhotskensis	DSMZ 18357	-	sea water	75	0	35
Psychrobacter_pacificensis	DSMZ 23406	-	sea water	78	4	38
Psychrobacter_piscatorii	JCM 15603	-	food processing	79	0	30
Psychrobacter_sanguinis	DSMZ 23634	-	human blood	70	4	37
Psychrobacter_urativorans	DSMZ 14009	-	food processing	72	4	27

Table S2. Contributions of COGs to variation in gene presence/absence PCoA.

COG_category	moraxellaceae_r^2	moraxellaceae	moraxellaceae	psych_r^2	psych_p-value	psych_p-adjust
A	0.233	0.005	0.006	0.295	0.001	0.002
B	0.355	0.001	0.001	0.077	0.034	0.037
C	0.66	0.001	0.001	0.438	0.001	0.002
D	0.192	0.005	0.006	0.119	0.009	0.011
E	0.648	0.001	0.001	0.299	0.001	0.002
F	0.748	0.001	0.001	0.179	0.002	0.003
G	0.802	0.001	0.001	0.067	0.061	0.061
H	0.629	0.001	0.001	0.069	0.051	0.053
I	0.854	0.001	0.001	0.283	0.001	0.002
J	0.645	0.001	0.001	0.181	0.002	0.003
K	0.736	0.001	0.001	0.379	0.001	0.002
L	0.028	0.508	0.508	0.136	0.002	0.003
M	0.692	0.001	0.001	0.432	0.001	0.002
N	0.141	0.028	0.031	0.467	0.001	0.002
O	0.462	0.001	0.001	0.388	0.001	0.002
P	0.669	0.001	0.001	0.556	0.001	0.002
Q	0.861	0.001	0.001	0.128	0.005	0.007
S	0.697	0.001	0.001	0.487	0.001	0.002
T	0.791	0.001	0.001	0.646	0.001	0.002
U	0.353	0.001	0.001	0.63	0.001	0.002
V	0.208	0.002	0.003	0.074	0.033	0.037
W	0.339	0.001	0.001	0.292	0.001	0.002
X	0.541	0.001	0.001	0.515	0.001	0.002
Z	0.097	0.083	0.087	0.103	0.014	0.017

Table S3. Psychrobacter strain collection cultivation information.

Full Name	Species	Catalogue	Isolation Source	Isolation Location	Previously Published Genome	Type Strain Publication	Cultivation Temperature	Cultivation Medium	Phenotyping Block(s)	Note
P_adeliensis_SJ14 (T)	adeliensis	DSMZ 15333	sea water	Antarctica		Shivaji et al 2004	25	lysogeny broth	2, 17	
P_aestuarii_SC35 (T)	aestuarii	JCM 16343	marine soil	Suncheon Bay, Korea		Baik et al 2010	30	marine broth	16, 2	
P_alimentarius_JG100 (T)	alimentarius	DSMZ 16065	food	South Korea		Yoon et al 2005	30	marine broth	7, 23	
P_alimentarius_JG102	alimentarius	DSMZ 16066	food	South Korea			25	marine broth	4, 18, 19	
P_aquaticus_CMS56 (T)	aquaticus	DSMZ 15339	terrestrial water	McMurdo Antarctica	GCA_000471625.1	Shivaji et al 2005	20	lysogeny broth	2, 9	
P_aquimaris_SW-210 (T)	aquimaris	DSMZ 16329	sea water	South Sea, Korea		Yoon et al 2005	30	marine broth	3	
P_arcticus_273-4 (T)	arcticus	DSMZ 17307	soil	Kolyma Lowland, Siberia	GCA_000012305.1	Bakermans et al 2006	20	marine broth	3	
P_arenosus_R7 (T)	arenosus	DSMZ 15389	marine soil	Amursky Bay, Sea of Japan		Romanenko et al 2004	25	marine broth	16, 2	
P_celer_SW-238 (T)	celer	DSMZ 23510	sea water	South Sea, Korea		Yoon et al 2005	30	marine broth	2, 17	
P_cibarius_JG-219 (T)	cibarius	DSMZ 16327	food	Korea		Jung et al 2005	30	marine broth	2, 7	
P_cibarius_JG-220	cibarius	DSMZ 16328	food	Korea			30	marine broth	5, 20, 21	
P_ciconiae_176-10 (T)	ciconiae	DSMZ 28412	bird	Klopot, Poland		Kämpfer et al 2015	30	marine broth	2, 7, 23	
P_cryohalolentis_K5 (T)	cryohalolentis	DSMZ 17306	terrestrial water	Kolyma Lowland, Siberia	GCA_000013905.1	Bakermans et al 2006	20	marine broth	16, 1	
P_faecalis_Iso-46 (T)	faecalis	DSMZ 14664	bird	Germany		Kämpfer et al 2002	30	lysogeny broth	16, 15	
P_faecalis_PBFP-1	faecalis	generated in this study	mammal				25	lysogeny broth	11	
P_fjordensis_BSw21516B (T)	fjordensis	KCTC 42279	sea water	glacial fjord in Svalbard		Zeng et al 2015	15	marine broth	3	
P_fozii_NF23 (T)	fozii	BCCM LMG 21272	soil	Antarctica		Bozal et al 2003	15	tryptic soy broth	2, 17, 9	
P_frigidicola_ACAM304 (T)	frigidicola	DSMZ 12411	soil	Magnetic Island, Vest fold Hills	GCA_007997305.1	Bowman et al 1996	10	marine broth	10	
P_frigidicola_ACAM309	frigidicola	NCIMB 13662	bird	Magnetic Island, Antarctica			20	tryptic soy broth	3	
P_fulvigenes_KC-40 (T)	fulvigenes	JCM 15525	invertebrate	Peter the Great Bay, Russia		Romanenko et al 2009	25	marine broth	16, 15	
P_fulvigenes_KC-65	fulvigenes	JCM 15526	invertebrate	Sea of Japan, Russia		Romanenko et al 2009	26	marine broth	16, 15	removed due to poor quality genome
P_glaciei_Blc20019 (T)	glaciei	KCTC 42280	terrestrial water	Svalbard		Zeng et al 2016	20	marine broth	3	
P_glacincola_1447	glacincola	NCIMB 115	fish	North Sea			10	marine broth	5, 20, 21	
P_glacincola_ACAM483 (T)	glacincola	DSMZ 12194	terrestrial water	Amery Ice Shelf Antarctica		Bowman et al 1997	15	marine broth	1, 17	
P_immobilis_1159	immobilis	NCIMB 206	fish	North Sea			20	nutrient broth	6, 22	
P_immobilis_1184	immobilis	NCIMB 208	fish	North Sea			20	nutrient broth	7, 23	
P_immobilis_1347	immobilis	NCIMB 211	fish	North Sea			20	nutrient broth	9	
P_immobilis_1555	immobilis	NCIMB 97	fish	North Sea			20	nutrient broth	5, 20, 21	
P_immobilis_1617	immobilis	NCIMB 99	fish	North Sea			20	nutrient broth	6, 22	
P_immobilis_1745	immobilis	NCIMB 100	fish	North Sea			20	nutrient broth	6, 22	
P_immobilis_252	immobilis	NCIMB 197	fish	North Sea			20	nutrient broth	6, 22	
P_immobilis_63	immobilis	NCIMB 11650	food				25	nutrient broth	8, 24	removed due to culture contamination
P_immobilis_A-11	immobilis	NCIMB 363	sea water				20	nutrient broth	7, 23	
P_immobilis_A-19	immobilis	NCIMB 364	sea water				20	nutrient broth	6, 22	
P_immobilis_A1007	immobilis	NCIMB 29	fish				20	nutrient broth	5, 20, 21	
P_immobilis_A1014	immobilis	NCIMB 309	fish				20	nutrient broth	6, 22	
P_immobilis_A351 (T)	immobilis	DSMZ 7229	food		GCA_003148585.1	Juni and Heym 1986	20	tryptic soy broth	3	
P_immobilis_A352	immobilis	NCIMB 10762	food				20	nutrient broth	8, 24	
P_immobilis_Liston1000	immobilis	NCIMB 98	fish	North Sea			20	nutrient broth	5, 20, 21	
P_immobilis_NC10	immobilis	NCIMB 131	fish	North Cape, Norway			20	nutrient broth	6, 22	
P_immobilis_NCIMB357	immobilis	NCIMB 357	fish	Sri Lanka			25	tryptic soy broth	6, 22	
P_immobilis_NCIMB361	immobilis	NCIMB 361	fish	Sri Lanka			20	nutrient broth	6, 22	
P_immobilis_NCIMB362	immobilis	NCIMB 362	fish	Sri Lanka			20	nutrient broth	6, 22	
P_immobilis_PAY	immobilis	NCIMB 13238	built	Germany			25	nutrient broth	-	removed due to cultue contamination
P_immobilis_S3	immobilis	NCIMB 2060	fish				20	marine broth	8, 24	
P_jeotgali_YKJ-103 (T)	jeotgali	DSMZ 23423	food	South Korea		Yoon et al 2003	30	marine broth	3	
P_luti_NF11 (T)	luti	BCCM LMG 21276	soil	Antarctica		Bozal et al 2003	15	tryptic soy broth	16, 2	
P_lutiphocae_IMMIBL-1110 (T)	lutiphocae	DSMZ 21542	mammal	Schleswig-Holstein, Germany	GCA_000382145.1	Yassin and Busse 2009	37	tryptic soy broth	16, 15	
P_marincola_KMM227 (T)	marincola	DSMZ 14160	invertebrate	Indian Ocean		Romanenko et al 2002	25	marine broth	7, 23	
P_maritimus_Pi2-20 (T)	maritimus	DSMZ 15387	sea water	Amursky Bay, Sea of Japan		Romanenko et al 2004	30	marine broth	16, 2	
P_maritimus_Pi2-25	maritimus	DSMZ 15397	sea water	Sea of Japan, Russia			25	marine broth	4, 18, 19	
P_namhaensis_SW-242 (T)	namhaensis	DSMZ 16330	sea water	South Sea, Korea		Yoon et al 2005	25	marine broth	3	
P_nivimaris_88-2-7 (T)	nivimaris	DSMZ 16093	sea water	Southern Atlantic Ocean	GCA_009812035.1	Heuchert et al 2004	30	marine broth	3	

P_oceani_4k5 (T)	oceani	JCM 30235	marine soil	Pacific Ocean		Matsuyama et al 2015	25	marine broth	3	
P_okhotskensis_76	okhotskensis	NCIMB 178	fish	North Sea			20	marine broth	8, 24	removed - unable to confirm pure culture
P_okhotskensis_A1004	okhotskensis	NCIMB 28	fish				20	nutrient broth	8, 24	
P_okhotskensis_MD17 (T)	okhotskensis	DSMZ 18357	sea water	Okhotsk Sea		Yumoto et al 2003	30	marine broth	1, 17	
P_pacificensis_NIBH-P2K6 (T)	pacificensis	DSMZ 23406	sea water	Japan Trench		Maruyama et al 2000	20	marine broth	15, 8, 24	
P_phenylpyruvicus_ACAM535 (T)	phenylpyruvicus	DSMZ 7000	mammal		GCA_001591185.1	Bowman et al 1996	30	blood agar	13	
P_piscatorii_T-3-2 (T)	piscatorii	JCM 15603	food	Japan		Zhou et al 2016	25	lysogeny broth	1, 15	
P_pisciculus	pisciculus	ACTCC BAA-2286	fish				22	tryptic soy broth	9	removed due to poor quality genome
P_priscuii_LFX-15B1	priscuii	ATCC BAA-1624	terrestrial water	Lake Fryxell, Antarctica			20	tryptic soy broth	4, 18, 19	
P_proteolyticus_116 (T)	proteolyticus	DSMZ 13887	invertebrate	King George Island Antarctica		Denner et al 2001	20	marine broth	4, 18, 19	
P_pulmonis_ES2-W5a2	pulmonis	DSMZ 30630	built	Noordwijk Netherlands			30	lysogeny broth	11	
P_pulmonis_S-606	pulmonis	DSMZ 16214	mammal	Zaragoza, Spain		Vela et al 2003	30	lysogeny broth	11	
P_salsus_DD48 (T)	salsus	DSMZ 15338	sea water	Adelie Island		Shivaji et al 2005	30	lysogeny broth	10	
P_sanguinis_10070	sanguinis	DSMZ 23637	mammal	Erie county NY USA			30	blood agar	13	
P_sanguinis_13983 (T)	sanguinis	DSMZ 23635	mammal	Queens county, NY USA		Wirth et al 2012	30	blood agar	13	
P_sanguinis_1501	sanguinis	DSMZ 23634	mammal	Oneida county, NY USA			37	blood agar	14	
P_sanguinis_92	sanguinis	DSMZ 23636	mammal	Albany county, NY USA			30	blood agar	13	
P_sp_1044	sp. 1044	NCIMB 111	fish	North Sea			20	marine broth	8, 24	
P_sp_2pS	species 2ps	DSMZ 16092	terrestrial water	Kolyma Lowland, Siberia			20	lysogeny broth	4, 18, 19	
P_sp_72-O-c	immobilis	NCIMB 2071	soil	Lang Hovde, Antarctica			20	nutrient broth	9	
P_sp_CMS30	sp. CMS30	DSMZ 15323	terrestrial water	McMurdo Antarctica			20	lysogeny broth	4, 18, 19	
P_sp_DD2	sp. DD2	DSMZ 15331	sea water	Antarctica			22	lysogeny broth	10	removed - unable to confirm pure culture
P_sp_DD43	sp. DD43	DSMZ 15332	sea water	Adelie Island			20	lysogeny broth	7, 23	
P_sp_H7-1	sp H7-1	DSMZ 5682	fish				30	nutrient broth	4, 18, 19	
P_sp_H8-1	sp H8-1	DSMZ 5684	food				30	nutrient broth	5, 20, 21, 8, 24	
P_sp_HII-4	sp HII-4	DSMZ 5683	food	Japan			30	nutrient broth	4, 18, 19	
P_sp_IAM12030-72-O-c	sp. IAM 12030	JCM 20360	soil	Showa Station, Antarctica			30	lysogeny broth	8, 24	
P_sp_JCM18900	sp. 18900	JCM 18900	mammal		GCA_000586415.1		30	marine broth	5, 20, 21, 8, 24	
P_sp_JCM18902	sp. 18902	JCM 18902	mammal		GCA_000586455.1		30	marine broth	7, 23	
P_sp_JCM18903	sp. 18903	JCM 18903	mammal		GCA_000586475.1		30	marine broth	5, 20, 21	
P_sp_NC44	sp. NC44	NCIMB 578	fish	North Cape, Norway			20	nutrient broth	8, 24	
P_sp_Pi2-1	sp. Pi 2-1	DSMZ 15403	sea water	Sea of Japan, Russia			25	marine broth	5, 20, 21	
P_sp_Pi2-51	sp. Pi 2-51	DSMZ 15386	sea water	Sea of Japan, Russia			25	marine broth	4, 18, 19	
P_sp_Pi2-52	sp. Pi 2-52	DSMZ 15405	sea water	Sea of Japan, Russia			25	marine broth	4, 18, 19	
P_spJCM18901	sp. JCM 18901	JCM 18901	mammal		GCA_000586435.1		30	marine broth	5, 20, 21	removed due to poor quality genome
P_submarinus_KMM225 (T)	submarinus	DSMZ 14161	sea water	Pacific Ocean		Romanenko et al 2002	25	marine agar	7, 23	
P_urativorans_ACAM311	urativorans	NCIMB 13663	soil	Magnetic Island, Antarctica			10	tryptic soy broth	9	
P_urativorans_ACAM534 (T)	urativorans	DSMZ 14009	food			Bowman et al 1996	15	tryptic soy broth	1, 17	
P_vallis_CMS39 (T)	vallis	DSMZ 15337	terrestrial water	Miers Valley Antarctica		Shivaji et al 2005	20	lysogeny broth	7, 23	

Table S4. Psychrobacter strain collection sequencing summary.

Full name	MiSeq Reads (post QC)	HiSeq Reads (post QC)	Nanopore Reads (post QC)	Completeness	Contamination	Strain Heterogeneity	Genome Size	Scaffold Number	Scaffold N50	Mean Scaffold Length	Longest Scaffold	Note
P_adeliensis_SJ14	1.48E+05	-	5.91E+04	99.44	1.14	37.5	3.09E+06	10	1.64E+06	3.09E+05	1.64E+06	
P_aestuarii_SC35	-	2.56E+06	1.21E+04	100	0.82	0	2.80E+06	3	2.78E+06	9.32E+05	2.78E+06	
P_alimentarius_JG100	-	2.44E+07	-	99.73	1.7	0	3.38E+06	36	2.05E+05	9.39E+04	4.58E+05	
P_alimentarius_JG102	7.61E+05	-	-	99.73	1.7	0	3.38E+06	39	2.02E+05	8.66E+04	6.77E+05	
P_aquaticus_CMS56	2.35E+06	-	-	99.58	0.55	0	3.22E+06	34	1.88E+05	9.48E+04	7.29E+05	
P_aquimaris_SW-210	1.11E+06	-	1.60E+05	100	0.32	0	3.44E+06	17	7.80E+05	2.02E+05	1.26E+06	
P_arcticus_273-4	-	3.62E+06	-	99.73	0	0	2.60E+06	223	2.22E+04	1.17E+04	1.45E+05	
P_arenosus_R7	7.00E+05	-	4.62E+04	99.95	0.69	0	3.70E+06	1	3.70E+06	3.70E+06	3.70E+06	
P_celer_SW-238	-	7.70E+06	-	99.63	0.62	0	3.02E+06	68	1.11E+05	4.44E+04	3.63E+05	
P_cibarius_JG-219	4.00E+05	-	-	100	0.27	0	3.22E+06	129	7.18E+04	2.50E+04	2.07E+05	
P_cibarius_JG-220	2.15E+05	-	1.17E+05	99.56	0.5	33.33	3.41E+06	35	5.06E+05	9.74E+04	8.64E+05	
P_ciconiae_176-10	3.14E+05	-	2.31E+05	98.83	0.27	0	2.48E+06	2	2.44E+06	1.24E+06	2.44E+06	
P_cryohalolentis_K5	4.48E+05	-	-	100	0	0	3.04E+06	31	2.77E+05	9.79E+04	5.10E+05	
P_faecalis_Iso-46	-	8.55E+06	-	99.44	0.27	0	3.07E+06	101	6.55E+04	3.04E+04	1.56E+05	
P_faecalis_PBFP-1	2.49E+05	-	3.16E+05	99.53	0.27	0	3.21E+06	19	1.11E+06	1.69E+05	1.50E+06	
P_fjordensis_BSw21516B	2.94E+05	-	2.36E+04	98.35	1.5	0	3.44E+06	8	2.04E+06	4.30E+05	2.04E+06	
P_fozii_NF23	2.05E+06	-	-	99.68	1.25	16.67	3.50E+06	24	4.28E+05	1.46E+05	6.77E+05	
P_frigidicola_ACAM304	-	1.07E+07	2.95E+04	99.6	1.28	0	2.86E+06	2	2.86E+06	1.43E+06	2.86E+06	
P_frigidicola_ACAM309	-	6.78E+06	-	97.73	0.69	0	2.78E+06	78	7.74E+04	3.56E+04	3.36E+05	
P_fulvigenes	314632	-	-	98.28	8.93	88.89	3558315	246	29134	14464	117589	removed due to poor quality genome
P_fulvigenes_KC-40	2.11E+05	-	9.45E+02	99.18	1.85	20	3.47E+06	5	3.40E+06	6.93E+05	3.40E+06	
P_glaciei_Blc20019	1.07E+06	-	-	99.86	0.62	0	3.35E+06	20	5.49E+05	1.68E+05	9.21E+05	
P_glacincola_1447	2.09E+05	-	-	99.86	0.32	0	3.30E+06	5	1.17E+06	6.61E+05	1.46E+06	
P_glacincola_ACAM483	8.02E+05	-	-	100	0.37	100	3.25E+06	117	1.26E+05	2.78E+04	3.06E+05	
P_immobilis_1159	1.83E+06	-	-	99.73	1.21	62.5	3.56E+06	198	1.05E+05	1.80E+04	3.66E+05	
P_immobilis_1184	3.02E+05	-	6.78E+03	99.31	1.24	0	3.25E+06	6	3.16E+06	5.41E+05	3.16E+06	
P_immobilis_1347	5.24E+05	-	-	99.91	0.27	0	3.01E+06	124	4.75E+04	2.43E+04	1.25E+05	
P_immobilis_1555	1.29E+06	-	-	100	0.27	0	3.49E+06	68	1.94E+05	5.13E+04	2.64E+05	
P_immobilis_1617	2.01E+06	-	-	100	0.27	0	3.47E+06	74	1.13E+05	4.69E+04	2.20E+05	
P_immobilis_1745	7.68E+05	-	-	99.95	1.24	0	3.17E+06	72	1.25E+05	4.40E+04	3.51E+05	
P_immobilis_252	3.25E+05	-	-	99.28	0.96	0	3.36E+06	178	3.73E+04	1.89E+04	1.17E+05	
P_immobilis_63	377746	213901	-	100	0.77	0	4933919	64	179966	77092	366327	removed - culture contamination
P_immobilis_A-11	9.24E+05	-	-	100	0.82	0	3.12E+06	28	4.48E+05	1.12E+05	6.30E+05	
P_immobilis_A-19	5.13E+05	-	9.04E+03	99.25	1.1	0	3.30E+06	37	3.05E+06	8.92E+04	3.05E+06	
P_immobilis_A1007	5.27E+05	-	-	99.86	0.55	0	3.62E+06	132	9.39E+04	2.74E+04	3.23E+05	
P_immobilis_A1014	1.11E+06	-	-	99.95	0.96	0	3.55E+06	107	7.94E+04	3.32E+04	3.78E+05	
P_immobilis_A351	1.17E+06	-	-	100	0.55	0	3.24E+06	52	1.45E+05	6.24E+04	4.68E+05	
P_immobilis_A352	-	9.64E+06	-	99.73	0.82	0	3.34E+06	63	1.25E+05	5.30E+04	4.00E+05	
P_immobilis_Liston1000	1.09E+06	-	-	99.95	0.82	0	3.22E+06	68	1.78E+05	4.74E+04	2.92E+05	
P_immobilis_NC10	1.75E+05	-	4.24E+03	99.18	1.47	50	3.47E+06	27	5.36E+05	1.29E+05	1.65E+06	
P_immobilis_NCIMB357	1.51E+06	-	-	100	0.27	0	3.39E+06	75	1.62E+05	4.52E+04	2.69E+05	
P_immobilis_NCIMB361	1.09E+06	-	-	100	0.27	0	3.44E+06	60	1.42E+05	5.74E+04	4.73E+05	
P_immobilis_NCIMB362	3.46E+05	-	-	99.4	0	0	3.21E+06	127	6.85E+04	2.53E+04	1.58E+05	
P_immobilis_PAY	336961	-	-	99.66	0.95	0	5356516	195	101235	27469	408453	removed - culture contamination
P_immobilis_S3	-	8.82E+06	-	99.95	0.27	0	3.23E+06	79	1.30E+05	4.09E+04	4.57E+05	
P_jeotgali_YKJ-103	-	8.16E+05	2.75E+04	98.3	0.55	0	3.14E+06	7	3.11E+06	4.48E+05	3.11E+06	
P_luti_NF11	1.07E+06	-	-	99.95	0	0	2.97E+06	20	5.11E+05	1.48E+05	5.97E+05	
P_lutiphocae_IMMIBL-1110	2.49E+05	-	2.00E+05	99.79	0.17	33.33	3.21E+06	7	2.89E+06	4.59E+05	2.89E+06	
P_marincola_KMM227	-	1.26E+07	-	99.85	0.55	0	3.06E+06	43	1.75E+05	7.11E+04	3.99E+05	
P_maritimus_Pi2-20	-	1.84E+06	-	99.95	0	0	3.16E+06	53	1.71E+05	5.97E+04	3.80E+05	
P_maritimus_Pi2-25	7.20E+05	-	-	99.54	0	0	3.18E+06	28	3.53E+05	1.13E+05	7.12E+05	
P_namhaensis_SW-242	3.01E+05	-	3.29E+05	96.98	1.1	75	2.82E+06	12	4.85E+05	2.35E+05	5.82E+05	
P_nivimaris_88-2-7	7.55E+05	-	-	99.95	1.17	0	3.39E+06	84	8.05E+04	4.04E+04	2.50E+05	
P_oceani_4k5	-	7.95E+06	-	100	0.27	0	2.98E+06	81	8.19E+04	3.68E+04	2.15E+05	
P_okhotskensis_76	65468	-	-	100	0.55	0	3156063	49	134244	64409	467980	removed - unable to confirm pure culture
P_okhotskensis_A1004	-	1.07E+07	-	99.86	1.24	0	3.43E+06	90	7.90E+04	3.81E+04	2.33E+05	

P_okhotskensis_MD17	-	2.41E+07	-	99.95	1.17	0	3.45E+06	105	7.47E+04	3.28E+04	3.18E+05	
P_pacificensis_NIBH-P2K6	2.91E+05	-	2.05E+05	100	1.65	71.43	3.17E+06	44	4.43E+05	7.20E+04	8.44E+05	
P_phenylpyruvicus_ACAM535	1.02E+06	-	-	99.89	0.49	0	3.28E+06	9	3.23E+06	3.64E+05	3.23E+06	
P_piscatorii_T-3-2	1.06E+06	-	-	99.91	1.51	12.5	3.51E+06	120	9.30E+04	2.93E+04	1.94E+05	
P_pisciculus_MF2-10b2		736797	-	-	39.79	3.82	56	1302753	1617	790	805	3620	removed due to poor quality genome
P_priscuii_LFX-15B1	1.33E+06	-	-	100	0.55	0	3.08E+06	47	1.89E+05	6.55E+04	3.44E+05	
P_proteolyticus_116	1.62E+06	-	-	99.85	0.55	0	3.04E+06	17	6.58E+05	1.79E+05	8.71E+05	
P_pulmonis_ES2-W5a2	2.23E+05	-	-	97.29	1.73	92.31	3.31E+06	29	3.18E+06	1.14E+05	3.18E+06	
P_pulmonis_S-606	5.24E+05	-	-	99.67	0.64	0	2.98E+06	13	6.37E+05	2.29E+05	8.76E+05	
P_salsus_DD48	1.00E+06	-	-	99.08	0.92	0	2.89E+06	68	1.14E+05	4.26E+04	2.69E+05	
P_sanguinis_10070	1.05E+06	-	-	99.75	1.76	15.38	3.16E+06	127	1.43E+05	2.49E+04	3.36E+05	
P_sanguinis_13983	-	8.89E+06	-	99.79	0.69	57.14	3.18E+06	173	5.03E+04	1.84E+04	1.62E+05	
P_sanguinis_1501	-	1.06E+07	-	99.79	0.6	50	3.07E+06	180	3.40E+04	1.71E+04	1.02E+05	
P_sanguinis_92	1.22E+06	-	-	99.66	0.41	0	3.25E+06	177	4.82E+04	1.84E+04	2.15E+05	
P_sp_1044	-	2.64E+06	-	100	0.32	0	3.35E+06	62	1.73E+05	5.40E+04	3.44E+05	
P_sp_2pS	1.19E+06	-	-	100	0.38	0	3.35E+06	28	4.21E+05	1.20E+05	5.71E+05	
P_sp_72-O-c	-	1.09E+07	-	99.73	0.59	33.33	3.18E+06	82	7.67E+04	3.87E+04	1.74E+05	
P_sp_CMS30	2.24E+05	-	2.39E+04	98.58	0.73	87.5	3.13E+06	52	1.33E+05	6.02E+04	2.49E+05	
P_sp_DD2		343471	-	-	99.08	0.92	0	2900144	82	63280	35367	164171	removed - unable to confirm pure culture
P_sp_DD43	-	2.11E+07	-	99.95	0	0	3.04E+06	49	1.13E+05	6.21E+04	2.74E+05	
P_sp_H7-1	1.20E+06	-	-	99.93	0.37	0	3.15E+06	117	5.37E+04	2.69E+04	1.85E+05	
P_sp_H8-1	2.37E+05	-	-	99.63	1.44	42.86	3.35E+06	7	3.35E+06	4.79E+05	3.35E+06	
P_sp_HII-4	5.96E+05	-	-	99.63	0.55	0	3.24E+06	122	9.02E+04	2.66E+04	4.10E+05	
P_sp_IAM12030-72-O-c	1.06E+06	-	-	99.73	0.59	33.33	3.22E+06	81	7.92E+04	3.98E+04	1.74E+05	
P_sp_JCM18900	1.54E+06	3.26E+05	-	99.67	0.69	0	3.28E+06	18	6.50E+05	1.82E+05	7.41E+05	
P_sp_JCM18902	3.36E+05	-	-	99.86	0.27	0	3.31E+06	5	3.22E+06	6.62E+05	3.22E+06	
P_sp_JCM18903	2.46E+06	-	-	100	0.82	75	3.46E+06	30	3.38E+05	1.15E+05	6.60E+05	
P_sp_NC44	-	1.09E+07	-	100	0.82	0	3.25E+06	46	1.88E+05	7.05E+04	3.46E+05	
P_sp_Pi2-1	1.48E+06	-	-	99.95	0.82	0	3.19E+06	67	1.33E+05	4.77E+04	5.29E+05	
P_sp_Pi2-51	-	1.51E+07	-	99.68	0.6	0	3.42E+06	69	1.39E+05	4.96E+04	2.98E+05	
P_sp_Pi2-52	-	9.64E+06	-	99.95	0.62	0	3.25E+06	69	1.08E+05	4.72E+04	2.94E+05	
P_spJCM18901		93588	-	-	88.58	5.89	82	3192919	89	184819	35875	572067	removed due to poor quality genome
P_submarinus_KMM225	-	2.78E+07	4.95E+03	99.91	0.55	50	3.01E+06	6	7.38E+05	5.02E+05	1.43E+06	
P_urativorans_ACAM311	-	2.25E+06	-	97.73	0.69	0	2.73E+06	63	1.04E+05	4.34E+04	3.10E+05	
P_urativorans_ACAM534	2.03E+06	-	-	99.68	1.61	0	3.46E+06	75	1.28E+05	4.61E+04	5.45E+05	
P_vallis_CMS39	-	9.20E+06	-	99.85	0.6	0	3.22E+06	56	1.51E+05	5.75E+04	3.28E+05	

Appendix B: Table 3.3

Table 3.1 Gene clusters significantly associated with *Psychrobacter* isolation source.

geneCluster	paralogs	predicted gene name	KEGG KO	COG category	eggNOG annotation	isolation source	num. isolation source	% isolation source	num. other isolation sources	% other isolation sources
	GC00002129_1						0	0	3	5
	GC00002129_2						0	0	1	2
	GC00002129_r1_1						0	0	2	3
	GC00002129_r1_2						0	0	1	2
	GC00002129_r1_3* **						**9**	**39**	**6**	**10**
	GC00002129_r1_r1_1						0	0	1	2
GC00002129	GC00002129_r1_r1_r1_1	-	-	X	-	other host	1	4	7	11
	GC00002231_1						0	0	1	2
	GC00002231_2						2	9	2	3
GC00002231	**GC00002231_r1_1*** **	-	-	K	nucleotide binding protein	other host	**11**	**48**	**9**	**15**
GC00002376	GC00002376***	BGR_15850	-	L	DNA primase	other host	6	26	10	16
	GC00002380_1	-	K00590	L	cytosine specific DNA methyl		0	0	1	2
	GC00002380_r1_1	-	K00571,K07319,K13581	L	adenine specific DNA methylt		0	0	1	2
GC00002380	**GC00002380_r1_2*** **	**-**	**K00571,K07319**	**L**	**adenine specific DNA meth**	other host	**10**	**43**	**6**	**10**
	GC00000256_1						0	0	2	3
	GC00000256_2						0	0	1	1
	GC00000256_3						0	0	1	1
	GC00000256_r1_1						0	0	1	1
	GC00000256_r1_r1_1_p1* **						**7**	**64**	**21**	**28**
GC00000256	GC00000256_r1_r1_1_p2	-	-	S	KWG leptospira repeat protein	food	11	100	72	97
GC00002564	GC00002564	fimC/D	K07347	M	outer membrane usher protein	food	5	45	7	9
GC00002565	GC00002565	fimB	K07346,K07357,K15540	N,U	fimbrial chaperone	food	5	45	7	9
GC00002566	GC00002566	fimA	K07345	N,U	major subunit type 1 fimbrin	food	5	45	7	9
GC00002666	GC00002666	fimA	K07345	N,U	major subunit type 1 fimbrin	food	5	45	6	8
GC00002047	GC00002047	-	-	S	DUF1810	marine	15	71	13	20
	GC00002181_1						1	5	0	0
GC00002181	**GC00002181_2*** **	-	-	X	-	marine	**11**	**52**	**12**	**19**
	GC00000111_1						0	0	1	1
	GC00000111_2						0	0	1	1
	GC00000111_3						0	0	1	1
	GC00000111_4						1	6	0	0
	GC00000111_5						0	0	1	1
	GC00000111_6* **						**6**	**38**	**54**	**78**
	GC00000111_r1_1						0	0	1	1
GC00000111	GC00000111_r1_r1_1	tnpR	-	L	resolvase	terrestrial	6	38	47	68

All homologs of the significant gene clusters are listed. Significant genes, as determined by treeWAS, are marked with an "***" and written in bold.