The microbial genomics of glacier-fed streams: adaptations to an extreme ecosystem

Massimo BOURQUIN

*"Nothing in Biology Makes Sense
Except in the Light of Evolution"*

Theodosius Dobzhansky, 1973

Acknowledgements

To my PhD supervisor and director, Tom Battin, for his guidance, support, and insights throughout the course of my doctoral studies. I extend my gratitude to the committee that provided insightful views and comments during and after the defence.

To the RIVER laboratory (ex-SBER), especially Hannes for the countless hours spent in front of a whiteboard, to Nic for maintaining the coffee stock in the lab and for always welcoming people with a hug when they come back from vacation, to Andrew and Martina for being the ones that risked their life to come and watch vipers with me in Lavaux, but also Jade, David, Jonas, the newer PhD students, the interns and master students.

A special mention to my office mates, namely Stelios that taught me the very basics of microbial ecology, but also Leïla for messing with Greg, Tyler for the never-ending motivation to go to Satellite, Grégoire and Aïleen for bearing with me on a daily basis during the final stages of the thesis, and last but not least, a special mention to Susheel for his help and support, with whom we spent hours debugging together almost on a daily basis over Zoom during the pandemic.

To the Vanishing Glaciers project's field team, namely Mike, Matteo, Martina and Vincent, for their effort and dedication in collecting samples from many glacier-fed streams globally.

To all the fixers, helpers, sherpas without whom the expeditions carried out by the field team would not have been possible.

To Paraskevi, Emmy, Flo, and the laboratory team that did a tremendous effort to create the environmental and metagenomic data that was used for the Vanishing glaciers project.

I would like to extend my gratitude to the NOMIS foundation for funding the Vanishing glaciers project and my doctoral studies, and to all the collaborators of the project.

To the great afternoons spent in Satellite with all of them. To the "Alperos" and the community of Alpôle.

To the *Ficus Lyrata* that never deceived us and followed the lab in its journey to Alpôle.

To my family, who have always supported me in my studies.

To my friends, especially Nathan and Marilou for the countless and necessary memes exchanged. And to Margaux with whom I lived for most of this thesis, and the insightful exchanges we had on the philosophy of science, good argumentation and logical reasoning.

Abstract

Glacier-fed streams are the cold, ultra-oligotrophic, and unstable streams that are fed by glacial meltwater. Despite these extreme conditions, they harbour a diverse and abundant microbial diversity that develops into biofilms, covering the boulders and sediments that form the streambed. These biofilms play key roles in ecosystem processes and exert a direct influence on downstream biogeochemistry. Here we aim to define the genomic potential of glacier-fed stream microbial communities using metagenomic analyses. As a reference point, we first present a global inventory of cryospheric microbiomes, and find shared taxonomic, functional and phylogenetic features that shape the bacterial communities globally. However, we also denote how these ecosystems remain understudied, and thus further work is required to characterise fully the microbiome of cryospheric ecosystems. Using the dataset of metagenomes generated by the Vanishing glaciers project, we then unravel strategies that microbes developed to thrive in the harsh environmental conditions of glacier-fed streams, including the importance of biofilm formation and cross-domain interactions. Moreover, using metagenome-assembled genomes, we find a unique phylogenomic diversity that harbours distinct genomic features.

Limited knowledge exists on how glacier influence shapes bacterial communities in glacier-fed streams. However, improving our understanding is crucial to better forecast how climate change will affect this extreme, yet endangered ecosystem. Using the global dataset of metagenomes and environmental parameters collected by the Vanishing glaciers project, we shed new light on the future of the glacier-fed stream microbiome. We first project environmental parameters onto future scenarios of climate change using predicted changes in glaciology and climate. These predictions corroborate conceptual models that forecast the "greening" of glacier-fed streams, and we further link this process with glacier size. Moreover, using a modelling approach and environmental, glaciological, and climatic variables as covariates, we forecast the changes in abundance of 2333 strains at a global scale. These models predict ecological shifts associated with the phylogenetic structure of the microbiome. Additionally, we find an association between these forecasted changes and the functional potential of these genomes, but also their genomic bulk features. This altered microbiome is expected to play a more important role in future glacier-fed streams, particularly in carbon cycling.

To better understand how glacier-fed stream microbial genomes are shaped by glacier influence, we create a new method that identifies phylogenetic clades that drive this relationship. This approach allows us to identify genomic optimisation patterns

Abstract

along the gradient of glacier influence, highlighting the importance of *Gammaproteobacteria* in shaping the genomic landscape of glacier-fed streams. Overall, this work serves as a reference resource for climate change microbiology, by providing a global dataset of cryospheric microbiomes, a modelling framework that allows to forecast the abundance of bacterial strains, and other methods to analyse microbiomes in a changing environment. Owing to human-induced climate change, the cryosphere is rapidly shrinking. Thus, targeted efforts are still required to unravel the threatened biodiversity of cryospheric ecosystems, and anticipate potential changes in ecosystem functioning.

Keywords: cryosphere, glacier-fed streams, microbial ecology, metagenomics, phylogenetics, modelling, climate change

Résumé

Les ruisseaux glaciaires sont les cours d'eau froids, ultra-oligotrophes et instables qui sont alimentés par les eaux de fonte des glaciers. Malgré ces conditions extrêmes, ils abritent une diversité microbienne abondante qui se développe en biofilms, recouvrant les blocs rocheux et les sédiments qui forment le lit du cours d'eau. Ces biofilms participent à d'importants processus écosystémiques qui exercent une influence directe sur la biogéochimie en aval. Nous visons ici à caractériser le répertoire génétique des communautés microbiennes des ruisseaux glaciaires à l'aide d'analyses métagénomiques. Comme point de référence, nous présentons d'abord un inventaire global des microbiomes cryosphériques, et nous trouvons des caractéristiques taxonomiques, fonctionnelles et phylogénétiques communes qui façonnent ces communautés bactériennes. Cependant, nous dénotons également que ces écosystèmes restent peu étudiés et que des travaux supplémentaires sont nécessaires pour caractériser pleinement le microbiome des écosystèmes cryosphériques. Ensuite, utilisant les métagénomes séquencés par le « Vanishing glaciers project », nous dévoilons les stratégies développées par les microbes pour prospérer dans les conditions environnementales difficiles des ruisseaux glaciaires, y compris l'importance de la formation de biofilms et des interactions entre les différents représentants des domaines du vivant. De plus, en utilisant les génomes assemblés par métagénome, nous trouvons une diversité phylogénomique unique qui révèle des caractéristiques génomiques particulières.

L'influence des glaciers sur les communautés bactériennes dans les ruisseaux glaciaires reste méconnue. Or, il est essentiel d'améliorer notre compréhension afin de mieux prévoir comment les changements climatiques affecteront cet écosystème. En utilisant des métagénomes collectés à l'échelle globale, et à l'aide de paramètres environnementaux mesurés par le «Vanishing glaciers project », nous explorons l'avenir du microbiome des ruisseaux glaciaires. Tout d'abord, nous réalisons des projections des paramètres environnementaux sur des scénarios de changement climatiques, en réalisant des modèles basés sur des projections de futurs changements glaciologiques et climatiques. Ces prévisions corroborent les modèles conceptuels qui prédisent le "verdissement" des ruisseaux glaciaires, et nous établissons un lien entre ce processus et la taille des glaciers notamment. De plus, en combinant modélisation et les projections de variables environnementales, glaciologiques et climatiques, nous prévoyons les changements en abondance de 2333 souches bactériennes à l'échelle globale. Ces modèles prédisent des changements écologiques associés à la structure phylogénétique du microbiome, avec un effet sur le potentiel fonctionnel de ces génomes, mais aussi leurs caractéristiques

génomiques. Ce microbiome altéré devrait jouer un rôle plus important dans les futurs cours d'eau alimentés par les glaciers, en particulier dans le cycle du carbone

Pour mieux comprendre comment les communautés bactériennes de ruisseaux glaciaires sont associées à l'influence des glaciers, nous créons une nouvelle méthode qui identifie les clades phylogénétiques à l'origine de cette relation. Cette approche nous permet d'identifier l'optimisation de la taille du génome en fonction de l'influence des glaciers, et souligne l'importance des *Gammaproteobacteria* dans cette relation. Dans l'ensemble, ce travail sert de ressources pour la microbiologie du changement climatique, en assemblant un ensemble de données mondiales sur les microbiomes cryosphériques, un cadre de modélisation qui permet de prévoir l'abondance de souches bactériennes, et d'autres méthodes pour analyser les microbiomes dans un environnement changeant. En raison des changements climatiques induits par l'homme, la cryosphère fond rapidement. Des efforts ciblés sont donc nécessaires pour découvrir la biodiversité menacée des écosystèmes cryosphériques et anticiper les changements potentiels dans leur fonctionnement.

Mots-clés: cryosphère, ruisseaux glaciaires, écologie microbienne, métagénomique, phylogénies, modélisation, changements climatiques

Table of contents

List of figures

List of supplementary figures

List of tables

Chapter 2

List of supplementary tables

Foreword

❖ *List of articles*

1. Fodelianakis, S. et al. Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. ISME J. 16, 666–675 (2022).

2. Busi, S. B. et al. Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams. Nat. Commun. 13, 2168 (2022).

3. Bourquin, M. et al. The microbiome of cryospheric ecosystems. Nat. Commun. 13, 3087 (2022).

4. Kohler, T. J. et al. Glacier shrinkage will accelerate downstream decomposition of organic matter and alters microbiome structure and function. Glob. Change Biol. 28, 3846–3859 (2022).

5. Ezzat, L. et al. Benthic biofilms in glacier-fed streams from Scandinavia to the Himalayas host distinct bacterial communities compared with the streamwater. Appl. Environ. Microbiol. 88, e00421-22 (2022).

6. Brandani, J. et al. Spatial patterns of benthic biofilm diversity among streams draining proglacial floodplains. Front. Microbiol. 13, (2022).

7. Brandani, J. et al. Homogeneous Environmental Selection Structures the Bacterial Communities of Benthic Biofilms in Proglacial Floodplain Streams. Appl. Environ. Microbiol. 89, e02010-22 (2023).

8. Busi, S. B. et al. Glacier-Fed Stream Biofilms Harbor Diverse Resistomes and Biosynthetic Gene Clusters. Microbiol. Spectr. 11, e04069-22 (2023).

9. Michoud, G. et al. The dark side of the moon: first insights into the microbiome structure and function of one of the last glacier-fed streams in Africa. R. Soc. Open Sci. 10, 230329 (2023).

❖ *List of conferences*

1. Predicting the impact of climate change on the glacier-fed streams microbiome
   Poster, ISME18, 2022, Lausanne, Switzerland
2. Leveraging species distribution models to predict the impact of climate change on glacier-fed stream biofilms
   Poster, INTECOL 2022, Geneva, Switzerland
3. Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change
   Poster, EGU23, 2023, Vienna, Austria
4. Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change
   Poster, ENAC Research Day 2023, EPFL, Switzerland
5. Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change
   Poster, EDCE Research Day 2023, EPFL, Switzerland
   Best poster jury award

Chapter 1. Introduction

❖ *1.1 The cryosphere*

The cryosphere, defined as the places on Earth where water is found in its solid state (e.g., glaciers, ice sheets, and permafrost) constitutes up to 20% of our planet's surface area (Fountain et al., 2012). Characterised by cold temperatures and impacted by the physical properties of ice and snow, the cryosphere plays a crucial role in the Earth's planetary systems, regulating the climate and sea levels (Barry and Gan, 2011). At a molecular scale, the cold alters the mechanisms involving proteins and other molecules, and slows down reactions. As ice and snow necessitate cold temperatures to form, cryospheric conditions are predominantly met at high altitudes and latitudes, concentrating glaciers, ice sheets, and other cryospheric environments in polar and high mountain regions.

Despite shared physical constraints, cryospheric ecosystems are diverse (Barry and Gan, 2011). In locations with low temperatures and high precipitation such as high-mountain areas, ice accumulates to form glaciers. At the scale of a single glacier, various habitats are created such as the supraglacial surface where streams and cryoconite holes form when the glacier melts, and the subglacial part that encompasses streams, frozen soil, and fissures in the ice (Anesio et al., 2017). In polar regions, ice accumulates into extensive ice sheets, even at lower elevation, while at the intersection of ice and oceans, sea ice is released into marine waters. Even where snow accumulation does not lead to permanent ice, snow covers substantial land areas, especially during winter. In alpine and high-latitude regions, the soil can be frozen perennially and form permafrost (Dobinski, 2011). Indeed, the cryosphere undergoes seasonal variations associated with temperature cycles, with the ice accumulated during winter melting in summer, releasing significant meltwater. In oceans, the freshwater dissolves, whereas on land, melted ice contribute to the formation of glacier-fed streams and lakes.

Cryospheric environments provide important ecosystem services that affect global hydrological and climatic processes (Barry and Gan, 2011), but also support an abundant and diverse biodiversity (Anesio and Laybourn-Parry, 2012). Foremost, snow and ice have high albedo and thus affect air temperature, regulating the climate at a global scale (Crook and Forster, 2014). Secondly, the insulating properties of snow and ice modulate the physical environments locally, decreasing the variations in temperature over large portions of land and sea that are covered (Barry and Gan, 2011).

Thirdly, the cryosphere represents a primary source of global freshwater, vital for potable water supplies and hydropower generation (Su et al., 2019), but also exerting a notable influence on global sea levels (Pörtner et al., 2019). Additionally, the cryosphere assumes cultural and recreational significance locally, exemplified for instance in the Swiss cultural context by the role of glaciers in the national identity, traditions, and tourism (Kosanic et al., 2023). Last but not least, as this will be the focus of this dissertation, cryospheric ecosystems sustain an abundant and diverse microbial biodiversity.

❖ *1.2 The microbial ecology of the cryosphere*

Often characterised by cold and often oligotrophic conditions, cryospheric ecosystems have for long thought to be deprived of life, and polar regions were typically not listed in any biome (Anesio and Laybourn-Parry, 2012). Indeed, these extreme environmental conditions result in low growth rate for the organisms that manage to survive in the cryosphere, and consequently these ecosystems can show relatively low biomass. Nevertheless, in the past decades, advances in microbial ecology have shown that cryospheric ecosystems support microbial life with representatives of all three domains of life (Anesio et al., 2017, 2009; Boetius et al., 2015). Moreover, despite the harsh nature of these environments, periods of warmer temperatures, such as the melt season of glaciers, represent period where the growth is important for most community members (Boetius et al., 2015). During these warmer intervals, microorganisms thrive as the availability of organic carbon provided mainly by primary producers increases, with an abundance of cells close to freshwater ecosystems. Hence, most cryospheric ecosystems harbour a diversity encompassing all three domains of life, including bacterial communities that are of high importance for ecosystem functioning (Anesio et al., 2009).

In the diverse habitats constituting the cryosphere, bacterial communities exhibit distinct compositions shaped by the unique conditions of each environment. In snow and cryoconite holes, *Cyanobacteria* thrive, capitalising on light availability, while *Alpha-* and *Betaproteobacteria* (taxa now affiliated to *Gammaproteobacteria* since *Burkholderiales* was reclassified) collectively dominate the bacterial communities. Within sea ice, abundance of *Flavobacteriia* and *Gammaproteobacteria* characterises the microbial landscape (Anesio et al., 2017; Brown and Jumpponen, 2019; Carey et al.,

2016). The supraglacial environment witnesses the high abundance of *Chlamydomonaceae,* especially on snow during summer where these algae form large blooms, and *Cyanobacteria* including genera like *Oscillatoria*, *Leptolyngbya*, *Phormidium*, and *Nostoc*, that thrive owing to their photosynthetic capabilities (Anesio et al., 2017; Stibal et al., 2012). In contrast, the subglacial habitat is characterised by the absence of light and a high abundance of *Betaproteobacteria* (these taxa are now classified as *Gammaproteobacteria* since the *Burkholderiales* order has been reclassified recently) (Foght et al., 2004). In permafrost, *Proteobacteria*, *Firmicutes*, *Chloroflexi*, *Acidobacteria*, *Actinobacteria*, and *Bacteroidetes* dominate the communities (Jansson and Taş, 2014). However, four years ago, direct comparisons between cryospheric habitats were restricted to high taxonomic levels (i.e., Phyla, Classes), and it remained unclear how this apparent diversity would appear at increased taxonomic resolution.

Due to the low temperatures characterising the cryosphere, Psychrophily or psychrotolerance, the ability of organisms to survive, adapt and grow at low temperatures, caught naturally the interest of microbial ecologists working on cold environments (De Maayer et al., 2014). While work on psychrophiles relies on the study of cultivated isolates, molecular adaptations in bacteria have been associated with cold temperatures (Bowman, 2017; De Maayer et al., 2014). Most psychrophiles possess alterations in the lipids composing membranes, providing more flexibility at cold temperatures (Konings et al., 2002). Furthermore, both intra- and extracellular proteins and enzymes, possess altered amino acid sequences, forming more flexible tertiary structures that are able to function at cold temperatures (Feller and Gerday, 2003). To maintain the cell turgor and enzymatic functions under cold conditions, these bacteria also use different solutes compared to their meso- and thermophilic counterparts (Feller et al., 1996; Feller and Gerday, 2003). Additional adaptations include the expression of specific molecules on the cell surface or their release outside of the cell, such as ice-binding or anti-freeze proteins (Tribelli and López, 2018). For instance, psychrophiles have been shown to produce a large amount of extracellular polymeric substances (EPS), participating to the formation of biofilms that provide protection against the harsh environmental conditions (Casillo et al., 2017; Marx et al., 2009).

Studies focusing on thermophilic bacteria with streamlined genomes have revealed an association between genomic bulk features and growth temperature (Sabath et al., 2013). Additionally, psychrophiles have been linked to distinctive traits, such as a high abundance of tRNAs and relatively large genomes compared to thermophiles (Dutta and Chaudhuri, 2010; Sabath et al., 2013; Satapathy et al., 2010). However, most

studies on psychrophiles are restricted to cultivable isolates, and thus focus on a minor part of natural communities. Additionally, the amount of data recently generated using metagenomics could help unravel many other psychrophilic adaptations in uncultured bacteria. Thus, unravelling the adaptations of bacteria within cryospheric ecosystems using metagenomics can not only shed light on the microbial communities within ecosystems and their role in its functioning, but could also contribute significantly to a broader comprehension of how bacteria adapt to cryospheric conditions.

❖ *1.3 Climate change, microorganisms and the cryosphere*

The cryosphere is undergoing rapid and unprecedented changes as a consequence of global climate change, and this has profound implications on its ecosystems. Increases in air temperatures and alterations in precipitation patterns at a global scale are changing the balance between the accumulation of ice and snow and their melting, leading to a large decrease in the extent of all cryospheric ecosystems. In the European Alps for example, the decrease in snow cover has been associated with an increase in primary production, a phenomenon commonly referred to as "greening" that has profound implication for terrestrial ecosystems (Rumpf et al., 2022). As climate change continues, understanding how it affects the cryosphere is thus important to forecast future changes in the environmental template that consequently alter ecosystem functioning.

Owing to their importance for the Earth' systems, understanding how microorganisms will be affected by climate change is crucial (Cavicchioli et al., 2019; Huss et al., 2017). In various ecosystems, they have been shown to affect ecosystem processes significantly. For example, the bacterial communities of permafrost have been shown to be relevant for global carbon cycling. The impact of global warming on this ecosystem results in soil thawing, a process that has been shown to induce ecological shifts in the microbial communities (Jansson and Taş, 2014). A direct impact on ecosystem functioning of these shifts is that bacterial communities then release more greenhouse gases up to levels that are relevant for the global carbon budget (Schuur et al., 2015). Hence, understanding how bacterial communities react to climate change is also key to forecast changes in ecosystem functioning. Moreover, owing to climate change, many ecosystems that potentially harbour unique biodiversity are at threat, many of which are still poorly understood (Elser et al., 2020). Currently,

we are in the last window of opportunity to sample and study some of these ecosystems that will be altered irreversibly (Elser et al., 2020). Thus, more sampling efforts, especially at large spatial scales are required to understand their impact on global cycles, and catalogue their biodiversity.

❖ *1.4 Glacier-fed streams*

Formed by glacial meltwater, glacier-fed streams represent important components of the Earth's hydrological systems. Draining the top of many important river basins, they exert a direct influence on downstream ecosystems through flow regimes and biogeochemical cycling (Hood et al., 2015; Horgby et al., 2019; Singer et al., 2012). These streams are characterised by extreme conditions imposed by glacier influence (Milner et al., 2017). The water temperature is ice-cold, and owing to the weathering of the bedrock by the glaciers, is loaded in abrasive and abundant mineral particles creating high turbidity (Jacobsen and Dangles, 2012). Additionally, due to the lack of allochthonous supply of organic matter and the scarcity of nutrients in the streams, these ecosystems are highly oligotrophic. Moreover, the high seasonality of melting cycles creates highly variable flow regimes that shape the constantly evolving streambed formed by boulders and sediments, and subsequently affects most environmental parameters in the streams.

Nevertheless, some organisms survive despite these extreme conditions, and have developed specific adaptations to cope with the environment. Biofilms formed mainly by bacteria and algae dominate the landscape and develop on boulders and in the sediments (Wilhelm et al., 2013). Importantly, these microbes play a crucial role in nutrient cycling within the ecosystem, thus impacting downstream ecological communities and biogeochemistry (Hotaling et al., 2017). Alongside the biofilms, cold-adapted macroinvertebrates are also able to survive these conditions (Becquet et al., 2022; Cauvy-Fraunié and Dangles, 2019; Scotti et al., 2019). Glacier-fed streams are highly seasonal: during winter the streams are snow covered, and during peak meltwater the flow is too turbulent for biofilms to develop (Ren et al., 2017a; Schütz et al., 2001; Scotti et al., 2019). Thus, the conditions are met for organisms to thrive mainly in summer before and after this peak, because during these windows of opportunities, the melting of the glaciers, the sunlight, and temperature become conducive for primary production (Boix Canadell et al., 2021).

Analogously to other alpine ecosystems, glacier-fed streams are predicted to undergo an increase in primary production associated with climate change (Jacobsen et al., 2012). While the reduction in snow cover will allow longer periods of exposition to sunlight, the decrease in turbidity associated with the reduction in glacier size, and

thus weathering, is expected to additionally drive the "greening" of alpine streams (Boix Canadell et al., 2021; Milner et al., 2017). This is because light availability at the stream bottom represents the main limiting factor to the growth for primary producers (Boix Canadell et al., 2021). Additionally, shifts in nutrient supply due to the transition from glacier-fed meltwater to other sources is expected to further impact the ecosystem and associated ecological communities (Milner et al., 2017). However, currently no quantitative predictions exist for these changes to the environmental template. While an overall increase in the abundance of bacteria is expected (Cauvy-Fraunié and Dangles, 2019), macroinvertebrates are predicted to undergo taxon-specific ecological shifts (Wilkes et al., 2023). Nevertheless, the potential effects of climate change on the various taxa that form the bacterial communities inhabiting glacier-fed streams are unknown.

❖ *1.5 Microbial ecology of glacier-fed streams*

Despite the harsh environmental conditions, the bacterial communities of glacier-fed streams thrive in various habitats. Bacterial communities have been shown to differ in community compositions across habitats including the ice, the streamwater, the particle-associated communities, and the biofilms that develop in the sediments, suggesting they represent various niches (Ezzat et al., 2022; Wilhelm et al., 2013). Several drivers have been shown to affect the communities such as light availability, pH and conductivity (Wilhelm et al., 2013). While an increased elevation has been associated with reduced alpha-diversity, the community composition across high-altitude communities differs more than across their low-altitude counterparts (Wilhelm et al., 2015). Moreover, studies performed on three glacial floodplains in the alps have shown that glacier-fed streams have reduced richness and distinct communities compared to tributary streams (that have non-glacial water sources) (Brandani et al., 2022).

In glacier-fed streams biofilms, key taxa dominate the bacterial communities, and exhibit phylogenetic patterns associated with the harsh environmental conditions. While the most abundant phyla are *Proteobacteria*, *Cyanobacteria*, *Bacteroidetes*, and *Actinobacteria*, at family level, representatives of the *Gammaproteobacteria* class (e.g., *Comamonadaceae, Oxalobacteraceae,* etc.) are among the most abundant (Brandani et al., 2022; Fodelianakis et al., 2022; Wilhelm et al., 2013). These

communities are characterized by homogeneous selection, and patterns of microdiversity have been shown in the most prevalent taxa (Brandani et al., 2023; Fodelianakis et al., 2022). Moreover, this phylogenetic structure of the microbiome has been associated with the amount of chlorophyll-*a* in the sediments. Nevertheless, these findings are restricted to glacier-fed streams in the Caucasus mountains, and the European and Southern Alps, and it is remains unclear how they would generalise in other mountain ranges.

While amplicon data (i.e., metabarcoding) of glacier-fed streams existed four years ago and allowed to characterise the taxa forming their communities, the functional potential of their ecological communities was poorly understood. Current knowledge was based on methods that link taxonomic profiles with genomes available in databases, and thus not representative of the environment (Ren et al., 2017b). This represented a missing link between the bacterial communities, and ecosystem processes, and thus investigating the genomic potential was particularly relevant in the context of climate change. Moreover, while conceptual models existed as to how the greening will affect the glacier-fed stream ecosystem, no quantitative predictions were available for the different scenarios of climate change. Furthermore, we did not know how these changes might affect the microbes, including the important and diverse bacterial communities.

❖ *1.6 Thesis objectives and the Vanishing glaciers project*

In this thesis, we aimed to investigate the microbiome of cryospheric ecosystems and glacier-fed streams in order to characterise adaptations to these extreme environments. We tried to answer this question: given the unique microbial communities of glacier-fed streams and their association with environmental parameters which are affected by climate change, what else besides water are we losing as the cryosphere vanishes?

To this end, we used metagenomic data including 16s and 18s rRNA amplicon data and shotgun metagenomes. This works consists in the processing of sequencing data, in the analysis of the microbial communities, and in the creation of models that associate the microbiome with environmental parameters. To this end, we combined methods rooted in ecology, genomics, and use phylogeny to account for the evolutionary history of the community members. While chapter 2 and 3 include work on amplicon data, entire metagenomes, and metagenomes assembled genomes (MAGs), chapter 4 and 5 included only the latter. These two chapters focus on modelling how some properties of the microbiome and its members are associated with the extreme environment of glacier-fed streams.

We first asked the question: what commonalities do the cryospheric ecosystems microbiomes share? And do they harbour unique features? Owing to the shared environmental conditions across the cryosphere, we expected similarities in community composition and shared unique strains and genomic potential that also translate in the genomic content of the metagenomes and genomes. Due to the cold temperatures, and the physical it has on cells, we hypothesised that the genomic potential of the cryosphere is enriched in cold-shock proteins and other functions that have been associated with psychrophily.

In Chapter 2, we investigated the microbiome of cryospheric ecosystems, collecting previously published datasets and performing a meta-analysis at a global scale. While highlighting shared taxonomic, functional and phylogenetic properties of cryospheric microbiome, we also found a unique and still unknown diversity. Thus, we prompted the need to further characterise these ecosystems, especially at large spatial scale, as they are rapidly disappearing. In this context, the vanishing glaciers project sampled successfully 170 glacier-fed streams distributed across all continents between 2019 and 2022, with the aim to characterise their microbiome. To this end, 173 biofilm metagenomes were sequenced. To link the microbiome with the environmental conditions, many physicochemical parameters (e.g., water temperature, nutrients, etc.) were measured, along with biomass measurements (bacterial abundance and chlorophyll-*a* content), and glaciological parameters (e.g., glacier surface area, glacial coverage, etc.). Data generated by the project will be used in all subsequent chapters.

We next focused on the microbiome of glacier-fed streams, using the metagenomic data generated by the project. Due to the oligotrophy and small-time scale at which the windows of opportunities open in these streams, we hypothesised that ecological communities possess metabolic adaptations to utilise resources efficiently. Moreover, due to the low temperatures, we also expected genomic adaptations to cold conditions such as genes involved in cell walls, biofilm formation and membrane biosynthesis.

In Chapter 3, we investigated the diversity of bacteria, eukaryotes and viruses inhabiting glacier-fed stream epilithic biofilms using data sampled in the Caucasus mountains and in the Southern Alps. We characterised their genomic potential and highlighted some genomic adaptations to the cold conditions that were previously associated with psychrophiles. We also found other potential adaptations to glacier-

fed streams such as chemolithoautotrophy and mixotrophy, and also showed the importance of cross-domain interactions using network analysis. Additionally, we investigated the phylogenomic diversity of the genus *Polaromonas* in glacier-fed streams, highlighting the novelty of these metagenome assembled genomes (MAGs).

Due to the tight link between environmental parameters and glacial influence in glacier-fed streams, this ecosystem is anticipated to undergo deep changes owing to climate change. However, few is known on how these changes in the environmental template will affect ecological communities, and thus if biodiversity losses could occur. Moreover, currently, no predictions exist to quantify how key environmental parameters of the ecosystem will change.

Thus, in Chapter 4, using globally distributed data, we projected the environmental template of glacier-fed streams onto future scenarios of climate using glacier recession and future predictions of climate. For this, we took advantage of the time-for-space substitution design of the Vanishing glaciers project, that was set up by sampling close to the glacier (upstream reach), and further downstream (downstream reach) for all glaciers, in order to characterise the gradient of glacier influence. We then used these projections of the environmental template to model the abundance of bacterial strains, forecasting phylogenetically structured ecological shifts under future scenarios of climate change. Moreover, we linked these shifts with the functional potential, suggesting links with ecosystem functioning.

Oligotrophy has been linked with reduced genome and cell size in several ecosystems. Due to the pronounced scarcity in organic carbon in glacier-fed streams, especially at high glacier influence, we hypothesised bacterial communities to optimise the size of their genomes accordingly. We thus investigated the genomic landscape of glacier-fed streams, identifying the optimisation of genomic bulk features as a potential adaptation to glacier influence. This topic is further developed in Chapter 5 where we additionally develop a method that allows us to identify *Gammaproteobacteria* as a main driver of community averages of genomic bulk features. Finally, we analyse the ecological success of this bacterial class in glacier-feds streams globally, and link it with functional adaptations associated with the evolution of *Gammaproteobacteria* through a pangenome analysis.

Chapter 2. The microbiome of cryospheric ecosystems

[1]Massimo Bourquin, [2]Susheel Bhanu Busi, [1]Stilianos Fodelianakis, [1]Hannes Peter, [3]Alex Washburne, [1]Tyler J. Kohler, [1]Leïla Ezzat, [1]Grégoire Michoud, [2]Paul Wilmes & [1]Tom J. Battin

[1]River Ecosystems Laboratory, Centre for Alpine and Polar Environmental Research (ALPOLE), École Polytechnique Fédérale de Lausanne, EPFL, Lausanne, Switzerland

[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, 7, avenue des Hauts-Fourneaux, L-4362, Esch-sur-Alzette, Luxembourg

[3]Selva Analytics LLC, Bozeman, MT, 59718, USA

❖   *Authors' contribution*

M.B.: conceptualisation, methodology, investigation, formal analysis, data curation and writing – original draft preparation, visualisation; S.B.B.: conceptualisation, methodology, investigation, data curation and writing – original draft preparation; S.F.: conceptualisation, methodology, data curation and writing; H.P.: conceptualisation, methodology, data curation and writing; A.W.: conceptualisation and methodology; T.K.: data curation and writing; L.E.: methodology; G.M.: methodology; P.W.: conceptualisation and supervision; T.J.B.: conceptualisation, writing – original draft preparation, data curation, supervision and funding acquisition.

In this work, I conceptualised the study, collected and curated the datasets of published 16s rRNA amplicon and processed them, performed all analyses, created the figures, and wrote the manuscript with the help of all co-authors. SBB lead the collection and curation of the shotgun metagenomes datasets. All co-authors gave advice and feedback on the analyses.

❖   *Publication*

❖ Abstract

The melting of the cryosphere is among the most conspicuous consequences of climate change, with impacts on microbial life and related biogeochemistry. However, we are missing a systematic understanding of microbiome structure and function across cryospheric ecosystems. Here, we present a global inventory of the microbiome from snow, ice, permafrost soils, and both coastal and freshwater ecosystems under glacier influence. Combining phylogenetic and taxonomic approaches, we find that these cryospheric ecosystems, despite their particularities, share a microbiome with representatives across the bacterial tree of life and apparent signatures of early and constrained radiation. In addition, we use metagenomic analyses to define the genetic repertoire of cryospheric bacteria. Our work provides a reference resource for future studies on climate change microbiology.

❖ Introduction

Microorganisms dominate the biosphere, maintain ecosystem processes, and play key roles in global biogeochemical cycles. The microbiome of cryospheric ecosystems, the nearly 20% of Earth's surface where water remains frozen for at least one month of the year (Fountain et al., 2012), currently figures among the least understood microbiomes on Earth (Boetius et al., 2015; Goordial, 2021; Jansson and Taş, 2014; Nayfach et al., 2020; Thompson et al., 2017). This is noteworthy given that the cryosphere is melting at an unprecedented pace owing to climate change. Motivated by the exploration of life in a planetary context (Merino et al., 2019) and the discovery of new biomolecules for biotechnology (Feller and Gerday, 2003), classical microbiology and (more recently) advances in sequencing technologies have unravelled physiological and molecular processes underpinning the adaptation of cold-adapted bacteria (i.e., psychrophiles) to the cryospheric environment (Bowman, 2017; D'Amico et al., 2006). More specifically, metagenomics has provided new insights into the structure and function of complex microbial communities of some cryospheric ecosystems, such as permafrost soils (Jansson and Taş, 2014; Tripathi et al., 2018), leading to a better understanding of the role of these ecosystems in global biogeochemical cycles and their vulnerability to climate change.

However, we are still missing an integrative understanding of the microbiome across the various and often underexplored cryospheric ecosystems on Earth (Goordial, 2021; Nayfach et al., 2020; Thompson et al., 2017). Here we present a global catalogue of microorganisms from various cryospheric ecosystems and at a taxonomic resolution that allows detection of cryosphere-adapted lineages and associated traits. We leverage sequence data from numerous published studies ranging from snow to permafrost ecosystems to shed light on the global cryospheric microbiome. While also illuminating geographical biases and underexplored habitats in the currently available cryospheric data, our study constitutes an important resource for the study of cryospheric life in general and its potential future in a warmer world.

❖ *Results and discussion*

The dataset

We curated and explored 695 published 16S rRNA gene samples from cryospheric ecosystems (Methods section and Supplementary Table 2.7), including polar ice sheets, mountain glaciers and their proglacial lakes, permafrost soils and the coastal ocean under the influence of glacier runoff, and compared these to 3552 published 16S rRNA gene samples from non-cryospheric ecosystems, including temperate and tropical lakes and soils (Supplementary Table 2.7). This approach allowed us to identify and explore features specific to the cryospheric microbiome and compare it to other environmental microbiomes. However, we note a geographical bias towards polar regions in current publicly available repositories, and the paucity of alpine samples specifically highlights the need to further characterise these habitats given that they are among the most endangered cryospheric ecosystems globally. This bias is further compounded by the inconsistent methodologies applied across studies (e.g., primer pairs and sequencers used). To account for potential primer biases, we analysed two 16S rRNA primer pairs (Primer Pair 1, PP1: 341f-785r; Primer Pair 2, PP2: 515f-806r) (Caporaso et al., 2012; Klindworth et al., 2013) commonly used in amplicon high-throughput sequencing. In total, this dataset contains 241,502,708 paired sequence reads, resulting in 530,254 and 410,931 amplicon sequence variants (ASVs) for PP1 and PP2, respectively. Moreover, all taxonomic analyses were performed at the genus level, to account for the limitations of 16s rRNA amplicon data. To gain deeper insights into the functional space of the cryospheric microbiome, we compared 34 published metagenomes from cryospheric ecosystems with 56 metagenomes from similar but non-cryospheric ecosystems (Figure 2.1A). Given the difficulty of obtaining high-quality metagenomes from cryospheric ecosystems, we restricted our analyses to glacier surfaces, ice-covered lakes, and Antarctic soils. Although our analyses were limited to

samples where raw sequence data are available (Methods section), the breadth of habitats covered are representative of the most abundant cryospheric ecosystems, e.g., glacier ice, cryoconites, subglacial lakes and sea ice. On the other hand, several niches such as glacier snow, glacier-fed rivers/streams, and the full-breadth of permafrost may not entirely be represented due to data unavailability. We reanalysed all metagenomes using the same bioinformatic pipeline (IMP3; see Methods section) to avoid analytical biases. Overall, the metagenomic analyses from 2,427,818,072 paired reads yielded 41,068,842 gene sequences. Thus, we here present a catalogue representing a snapshot of the functional diversity in the cryospheric microbiome, integrating across diverse habitats. This represents what we believe to be the first global overview of the functional repertoire of the Earth's cryosphere compared to other ecosystems.

A cryospheric microbiome

Given the communal constraints imposed by the harsh environment of cryospheric ecosystems (e.g., low temperature, oligotrophy), we expected them to harbour a specific microbiome. Accordingly, machine-learning classification (logistic regression models, Methods) based on community composition was able to differentiate between cryospheric and non-cryospheric microbiomes with high accuracy (balanced accuracy >0.96, Supplementary Table 2.1). Both primer pairs consistently yielded a high classification accuracy and especially a high precision. Interestingly, many of the discriminating cryospheric ASVs were spread widely across the bacterial tree of life (Figure 2.1A and Supplementary Figure 2.1).

The notion that a part of the microbiome is specific to the cryosphere is also strongly supported by phylogenetic analyses of the 16 S rRNA gene amplicon dataset. First, we found higher pairwise phylogenetic overlap among cryospheric samples than among cryospheric/non-cryospheric or non-cryospheric samples (Sorensen's index, Figure 2.1C; Wilcoxon test, Holm adj. $p < 0.001$). This points towards a phylogenetic diversity that is specific to the cryosphere. Second, when we examined cross-sample nearest taxon distances (β-NTDs), we found that taxa in cryospheric samples have lower β-NTDs in other cryospheric samples than in non-cryospheric samples (Figure 2.1D; Wilcoxon test, Holm adj. $p < 0.001$). This was less evident for non-cryospheric samples (Supplementary Table 2). Because phylogeny and functional

similarity usually correlate at short phylogenetic distances (Dini-Andreote et al., 2015), this finding suggests higher niche similarity for cryospheric bacteria compared to their non-cryospheric equivalents. This evokes specific selective constraints of cryospheric environments acting on taxa across the entire bacterial tree of life. Interestingly, when we further examined radiation patterns, we found that taxa in a given cryospheric microbial community had on average larger phylogenetic distances (α-MPD) than their counterparts in a non-cryospheric community (linear model, $p < 0.001$). This could suggest early radiation events with subsequent "pruning" of phylogenetic diversity, which would explain the observed patterns (Mazel et al. 2015). However, we cannot exclude nor disentangle the action of contemporary evolutionary and assembly processes that can jointly shape community phylogenies. For example, transduction and genome plasticity have repeatedly been linked with cold adaptation in cryospheric bacteria. Moreover, horizontal gene transfer has also been shown to promote the diffusion of cold-adaptation genes (Dorrell et al., 2021). Nevertheless, given the large number and breadth of included cryospheric ecosystems, we posit that the topologies of the inferred phylogenies are less prone to assembly processes. We rather interpret that the observed patterns are signs of early and constrained radiation in the cryospheric microbiome. Collectively, these results point to similar evolutionary trajectories in cryospheric microbiomes, probably owing to similar environmental conditions across various cryospheric ecosystems, over timescales, relevant for bacterial macroevolution.

The abundance of a given species in an ecosystem generally reflects its fitness and adaptive capacity to the respective environmental conditions. Therefore, we explored patterns of differential abundance (Methods section) and found 589 bacterial genera with higher abundances in cryospheric compared to non-cryospheric samples (Ancom, W statistic > 0.7, CLR mean difference > 0) that hereafter will be referred to as cryospheric genera. These genera were distributed widely across the bacterial tree of life and encompassed 46 different phyla. Despite this wide distribution, we found that 34.8% and 13.4% of the cryospheric genera were affiliated *Proteobacteria* and *Bacteroidota*, respectively (Figure 2.2A). The relevance of *Proteobacteria* is in line with the high prevalence of *Alpha-* and *Gammaproteobacteria* typically reported in the cryospheric literature (Anesio and Laybourn-Parry, 2012; Boetius et al., 2015). Genera belonging to the *Alpha-* and *Gammaproteobacteria* classes displayed the highest differential abundance and included *Sphingomonas*, *Polaromonas*, *Rhodoferax, Brevundimonas, and Acidiphilum* (Figure 2.2B) — some of them with taxa typically reported to be psychrophiles (Comte et al., 2018; D'Amico et al., 2006; Fodelianakis et al., 2022; Sajjad et al., 2021). *Bacteroidota* was the second most important phylum of the cryospheric microbiome with *Hymenobacter*, *Ferruginibacter*, and *Polaribacter* (for

instance) as dominant genera, all of which are known from permafrost soils and ice ecosystems (Collins et al., 2010; Kim et al., 2013). Furthermore, as previously reported (Christner et al., 2003; Kohler et al., 2020b), the cryospheric genera included members of the *Actinobacteria, Chloroflexi* and *Cyanobacteria* phyla, alongside some *Firmicutes*. The former two are particularly common in supraglacial environments (Boetius et al., 2015), and *Cyanobacteria* are important components of cryoconite microbiomes (Anesio et al., 2017). Our global analyses thus corroborate and extend previous reports on microbiome composition in distinct cryospheric ecosystems. Furthermore, our differential abundance analysis unveiled genera (e.g., *Oryzihumus* or *Pseudolabrys*) that have not been previously associated with the cryosphere (Figure 2.2B). More importantly, many of the detected cryospheric genera only have placeholder names due to the lack of cultivated representatives (e.g., CL_500-29_marine_group, hgcl_clade, TRA3-20), underlining unique bacterial groups that are yet to be described. Collectively, these findings unveil an unexpectedly diverse and likely well-adapted microbiome specific to the cryosphere, and supports the notion of the cryosphere as a biome with its distinct association of microorganisms, alongside plants and animals (Anesio and Laybourn-Parry, 2012).

Compositional patterns across cryospheric ecosystems

We next explored how microbial community composition varies across cryospheric ecosystems. Using similarity analyses, we found that the microbiome composition differed significantly between cryospheric ecosystem types (PERMANOVA, r^2 = 0.183, p < 0.001; pairwise.adonis, p < 0.001 for all pairwise comparisons) (Figure 2.3A and Supplementary Table 4). Most conspicuous was the segregation of snow/ice and marine communities, bracketing freshwater and terrestrial cryospheric communities. We also found a significant but relatively small effect of the primer pair (PERMANOVA, r^2 = 0.019, p < 0.001) that could be attributable to primer bias, or inherent differences related to sampling. To further assess these distributions, we explored prevalence patterns to identify a core microbiome across cryospheric ecosystems (Figure 2.3B). We found 37 bacterial genera, including *Pseudomonas*, *Acinetobacter*, and *Flavobacterium*, for instance, to constitute the core microbiome. The disproportionate representation of these core genera in the above-identified cryospheric genera (Fisher's exact test, p < 0.001, odds ratio = 6.93) underlines their high abundance in cryospheric

ecosystems (Supplementary Figure 2.2). It also shows the prevalence and abundance of some cryospheric genera, indicating their potential relevance for ecosystem processes.

Additionally, analysing the relative abundance of the core cryospheric genera for each primer pair and cryospheric ecosystem types, we found that ice and snow microbiomes were associated with the highest proportions of core genera (23.05% and 24.8% for PP1 and PP2, respectively) (Figure 2.3D). In contrast, the marine cryospheric microbiome is only marginally composed of these genera (16.9% and 13.3% for PP1 and PP2, respectively). This pattern is in line with our unconstrained ordination analysis (Figure 2.3B) and suggests that snow and ice represent endmember cryospheric systems, while the cryospheric component of the microbiome dissipates in downstream freshwaters, soils and the coastal ocean. Furthermore, the alpha-diversity was higher in terrestrial (Shannon H = 3.67), marine (H = 3.25) and freshwater (H = 2.99) ecosystems than in snow and ice (H = 2.86), corresponding to increasing contributions of ancillary taxa to their microbiomes (Supplementary Table 5). These differences in diversity are likely attributable to environmental gradients characterised by more diverse energy sources and niches, such as when moving from snow and ice to aquatic and soil ecosystems. Our analyses revealed compositional patterns of the cryospheric microbiome suggesting that snow and ice ecosystems including supraglacial habitats (e.g., mountain glaciers, ice sheets, snow and cryoconites) may serve as a source of cold-adapted bacterial diversity, upon losing which the downstream diversity may become threatened as well.

Functional potential of the cryospheric microbiome

The adaptive and survival strategies of microorganisms to the extreme environmental conditions of the cryosphere have received substantial attention over the last years (Frey et al., 2016; Margesin and Collins, 2019; Tribelli and López, 2018). For example, genomic insights from bacterial cultures have revealed mechanisms of thermal adaptation linked to bulk genomic features, such as GC content and genome size (Wu et al., 2012). Moreover, genome streamlining has been shown to be a relevant evolutionary force in the cryosphere (Margesin and Collins, 2019). Therefore, we analysed the GC content and genome size of 13,414 reference genome sequences from the NCBI Refseq genomes database (Pruitt et al., 2007) to investigate shared properties of cryospheric genera, and to provide a framework to contrast future cryospheric metagenomic results. By comparing these reference genomes representing 660 bacterial genera present in our taxonomic analyses (29.8% of which are cryospheric genera according to our differential abundance analysis), we found that the cryospheric

genera had a significantly higher GC content (Supplementary Figure 2.3B; Wilcoxon test, Holm adj. p = 0.0011, median difference = 8.8%) compared to the other genera; a pattern also supported by an enrichment in sequences that encode GC-rich amino acids (e.g., Alanine, Arginine, Glycine) (Supplementary Figure 2.3A and Supplementary Table 2.6). Therefore, our findings suggest that cryospheric genera indeed share an elevated GC content (Almpanis et al., 2018), in line with reports on cold-adapted *Synechococcus* (SynAce01) (Tang et al., 2019) and *Actinobacteria* (Goordial et al., 2015). We also report that the average genome size of cryospheric genera is closely bracketed by published values for psychrophilic bacteria (Sabath et al., 2013).

Next, using a gene-centric approach, we explored the functional space of the cryospheric metagenomes dataset. Out of 17,191 KEGG orthologues (KO), 980 KO were significantly enriched in cryospheric samples. Cryospheric genera and particularly cryospheric core members (e.g., *Pseudomonas*, *Sphingomonas* and *Novosphingobium*) disproportionately accounted for these gene families (Figure 2.4A). Our analysis highlighted the relevance of chemolithotrophic pathways (e.g., manganese and iron uptake, sulfur, nitrogen and hydrogen metabolism), complementing earlier reports on these particular functional attributes of cryospheric ecosystems (Figure 2.4B) (Boyd et al., 2011, 2010; Christner et al., 2003). The apparent relevance of chemolithotrophic pathways is likely attributable to a relative scarcity of organic carbon in cryospheric ecosystems. Interestingly, we consistently identified chitinase genes, which are involved in permafrost carbon cycling, but may also be an adaptation to freezing (Liu et al., 2016). Finally, genes involved in adhesion, motility and various secretion systems collectively point to biofilm formation as an important strategy for life in cryospheric ecosystems (Smith et al., 2016), which are often characterised by extended periods of oligotrophy and elevated UV-radiation.

Our cross-ecosystem metagenomic analyses not only shed light on potential functions of the cryospheric microbiome across ecosystems, but also unveiled a large uncharacterised functional space with 43.4% of the protein coding genes in cryospheric samples unannotated to a KEGG orthologous group. While this does not seem unusual for environmental metagenomes in general (Nayfach et al., 2020), it is notable that we may lose this functional potential as the cryosphere vanishes. In order to shed light on this uncharacterised functional space, we clustered 41,068,842 gene sequences based on a 30% sequence similarity and 80% sequence coverage

threshold, subsequently mapping representative sequences of the largest clusters (>29 sequences in at least 2 samples, n = 12,125) to the UniProt TrEMBL database (Figure 2.4C). While the KEGG assigned clusters overall had a high percentage of sequences that matched genes in the UniProt database (Table 2.1), we found that cryosphere specific sequences show a large decrease in the clusters assigned to multiple KEGG (i.e., ambiguous) and even more in the ones containing exclusively unassigned sequences, compared to non-cryospheric environmental metagenomes. In addition to the low percentage of gene sequences matching UniProt sequences, we found that the cryosphere specific clusters that align to the database show a largely decreased identity with the matching sequence (Supplementary Figure 2.4). These findings underline the lack of representation of cryospheric sequences in current gene sequences databases, potentially linked to the specificity of certain taxa to the cryosphere, and/or functions. Finally, the large nucleotide similarity within these clusters (Supplementary Figure 2.4) suggests that these are conserved functions of particular importance to microbial life under cryospheric constraints, and corroborates the notion of specific lineages of closely related taxa to dominate microbial life in the cryosphere. Aside from being uncharacterised, 170 of the unassigned gene clusters were only detected in cryospheric metagenomes and could thus represent unknown gene families of importance to understanding the adaptation of bacteria to these extreme ecosystems.

Collectively, our insights both at the taxonomic and functional level reveal key microbiome features that are exclusive to cryospheric ecosystems. Although entire taxonomic lineages are not unique to cryospheric ecosystems, it is evident that specific species and potentially strains are novel and adapted to these environments. Similarly, the emergent functional properties clearly demonstrate the exclusivity of functions, especially those that are yet to be characterised or that can be classified based on existing databases, within the cryosphere. On the contrary, we find that in both the taxonomic and functional complements, several taxa and functions are shared with non-cryospheric ecosystems. This is expected since the underlying genomic content supporting the taxonomic and functional annotations are shared between the cryospheric and non-cryospheric ecosystems. This is evident based on the >50% identity among the shared gene clusters that had matching identities in the KEGG database (Table 2.1).

Here we present what we believe is the first global data-driven approach to unravel specific features of the cryospheric microbiome. Our meta-analysis revealed diverse, distinct and functionally specific bacterial communities that appear to have been shaped by sustained evolutionary forces, suggesting an ancient origin of this

biodiversity. While our study highlights key taxonomic groups such as *Proteobacteria* and *Bacteroidota*, our findings also disclose the importance of yet-uncultured bacteria and an uncharacterised genetic repertoire. In light of the threatened nature of the cryosphere, targeted efforts to unravel the phylogenetic and genomic underpinnings of bacterial adaptation to cryospheric ecosystems, including prospecting for cold-adapted biomolecules as well as the cultivation of cryospheric bacteria, are urgently required.

❖ *Methods*

16S rRNA datasets

Two primer pairs typically used in microbial ecology targeting the prokaryotic 16S rRNA were assessed: 341f-785r and 515f-806r. They will be referred to as Primer Pair 1 (PP1) (Klindworth et al., 2013) and Primer Pair 2 (PP2) (Caporaso et al., 2012), respectively. All articles citing the PP1 and PP2 reference articles were retrieved using Web of Science (All databases, searched on the 7 December 2019, 1727 articles). The first selection based on title and abstract was performed as described herein. Only studies having sequenced environmental samples were kept. Simultaneously, studies assessing pollution or contamination and involving major climatic or ecological events, e.g., storms or blooms, were removed. Thereafter, a second selection was performed based on the whole article, assessing technical criterions. Only studies using the aforementioned primer pairs, Illumina paired-end sequencing and having available data were kept; and the corresponding NCBI bioproject accessions were extracted. At a later stage, a few studies meeting the filtering criteria but not included in the Web of Science search were added.

The raw sequencing (fastq) data were downloaded using the ENA browser (European Nucleotide Archive; www.ebi.ac.uk/ena/browser/). At this stage, only the control samples were downloaded for experimental studies. The read files were filtered as follows: First, Trimmomatic was used to remove low quality reads, truncating the reads at the first instance of a sliding-window (5 bp) having a mean quality below 15 (Bolger et al., 2014). At this stage, the raw data from each BioProject was imported into qiime2 (Bolyen et al., 2019). Denoising was performed with the dada2 plugin using the primers sequences length for the '–p-trim-left-r' and '–p-trim-left-f' parameters (Bokulich et al.,

2018). This step removed integrally two studies in the PP1 dataset ("negative values in quality" and "all samples discarded" errors). Only sequences assigned to bacterial taxa were kept, and chloroplast and mitochondrial sequences were also removed. Finally, all samples with less than 5000 reads after this initial filtering were removed.

Taxonomy classification for PP1 and PP2 ASVs was performed using the qiime2 'feature-classifier' plugin and the Silva 138 nr99 database (Bokulich et al., 2018; Quast et al., 2013). First, reads were extracted from the reference sequences using the extract-reads method. For this, the primer sequences were used for the '–p-r-primer' and '–p-f-primer' arguments. The length of the extracted reads was set to min. 250 and max. 450 for the PP1 dataset and min. 200 and max. 400 for the PP2 dataset. A classifier was then created using the fit-classifier-naïve-bayes method with the extracted reads and the reference taxonomy. Finally, this classifier was run on the dataset's sequences using the 'classify-sklearn' method to get the sequences taxonomy (Bokulich et al., 2018). To keep only high-quality samples, all samples having <75% of their ASVs assigned to the phylum level, and 50% assigned to the genus level were removed. This filtering resulted in 2508 samples and 530,254 ASVs for PP1 and 1739 samples and 410,931ASVs for PP2. The ASV tables and metadata tables for these datasets can be found on Zenodo, under the file names: 'Data/PP1_table.tsv', 'Data/PP2_table.tsv' and 'Metadata/PP1_metadata.tsv' and 'Metadata/PP2_metadata.tsv', respectively.

Metagenomic dataset

To address the functional aspect of identified taxa, accession numbers of studies comprising of the following keywords: metagenomics, whole genome shotgun, and environmental, were queried using NCBi's EDirect (Winter, 2017). The results were manually curated to select studies from a broad Geographic distribution, yielding a total of 382 datasets. The selection of metagenomic samples was further restricted to raw fastq data, thus precluding the use of samples from MG-RAST since only the metagenome assembly files were provided. Additionally, all samples still under embargo in accordance with the Joint Genome Institute (JGI; USA) policy, were excluded. From this collection, samples with fewer than 1 million reads or with a quality of reads less than Q25 were removed for a final collection of 91 samples (Figure 2.1A). Paired reads were processed using the Integrated Meta-omic Pipeline (IMP) (Narayanasamy et al., 2016). The workflow includes pre-processing such as primer/adaptor removal and trimming followed by an iterative assembly. Additionally, functional annotation of genes based on custom databases was performed (described below). The entire workflow is setup in a reproducible *Snakemake* format (Köster and Rahmann, 2012). Briefly, after preprocessing the reads, de novo assembly using

MEGAHIT (v1.2) assembler was performed (Li et al., 2015). All the methods and parameters used are listed on the Github repository, in the 'Preprocessing/IMP_config.yaml' file. The metagenomic dataset KEGG Orthologs (KO) table, taxonomy table, and metadata are available on Zenodo under the 'Data/MTG_KEGG_counts.tsv', 'Data/MTG_table.tsv', and 'Metadata/MTG_metadata.tsv'.

Metagenomic taxonomic classification and functional analyses

Functional potential analyses from contigs were determined by predicting open-reading frames using a modified version of *Prokka* (Seemann, 2014) including *Prodigal* (Hyatt et al., 2010) gene predictions for complete and incomplete open reading frames. Genes identified subsequently were annotated with Hidden Markov Models (HMM) (Eddy, 2011), trained using an in-house database (Heintz-Buschart et al., 2016). The annotations were further annotated with KO (Kanehisa and Goto, 2000) groups using 'hmmsearch' from *HMMER* 3.1 (Eddy, 2011). Upon multiple functional group assignments, the best hits based on bit scores were selected. *FeatureCounts* (Liao et al., 2014) with the '-p' and '-O' arguments were then used to extract the number of reads per functional category.

Logistic regression classification of cryospheric bacterial communities

The Logistic regression implemented in *scikit-learn* python module (version 0.23.2) was trained on presence-absence ASV tables to classify cryospheric samples (Pedregosa et al., 2011). To reduce the amount of ASVs considered, the table was filtered based on relative abundance: presence was defined at a 0.005 relative abundance threshold. A stratified 5-fold cross-validation (CV) was ran and the scores were averaged across the CVs. This process was repeated 40 times and the mean and standard deviations are reported for each metric. To ensure reproducibility, the seed was set as 23 for the classifier, and as the iteration number for the stratified cross validation iterator (from 0 to 39). The C parameter controlling L2 penalisation was turned off using the 'none' argument and the lbfgs solver was used. ROC curves were plotted using the 'plot_roc_curve' function of the scikit-learn python module. Balanced accuracy, precision and recall were computed using the 'accuracy_score', 'precision_score' and 'recall_score' methods, respectively, with sample weights

correcting for the sample size of the cryospheric and non-cryospheric datasets (Supplementary Table 1). The means and standard deviations of scoring metrics for the classifiers can be found in table S1. Odds ratios were calculated using the exponent of the logistics models coefficients. The tables containing the ASVs logistic regressions odds ratios can be found in the Data folder available on Zenodo under the name 'PP1_Logistic_coefs.csv' and 'PP2_Logistic_coefs.csv' for PP1 and PP2, respectively.

Phylogenetic analyses

Phylogenetic trees were built using the set of ASVs found in the dataset used for the logistic regression classification. Due to the different 16 S regions targeted, phylogenies for both PP1 and PP2 datasets were constructed separately. The ASVs sequences were aligned using the *FFT-NS-2* algorithm implemented in the *Mafft* aligner with default parameters (Katoh and Standley, 2013). The alignments were subsequently trimmed using *TrimAl* with the '-gt 0.95' parameter, and the trees built using *IQ-TREE* with the *GTR* model of nucleotide substitution and the '-fast' option (Capella-Gutiérrez et al., 2009; Nguyen et al., 2015). Phylogenetic tree visualisations were created using the *ggtree* and *ggtreeExtra* R packages (Xu et al., 2021; Yu et al., 2018). Only positive coefficients showing enriched presence in cryospheric environments are shown in the phylogenetic barplots (Figure 2.1). The number of ASVs with an odds ratio above 1 was shown for taxonomic summaries (Supplementary Figure 2.1B, C).

ß-diversity phylogenetic metrics (Sorensen's Index and ß-MNTD) were computed using the 'phylosor' and 'comdistnt' functions of the *Picante* R package (Kembel et al., 2010), using custom functions to compute pairwise comparisons. For each metric, 50 iterations were performed where we calculated the pairwise distances between and within 50 cryospheric, and 50 non-cryospheric samples. This random sub-sampling approach was chosen to reduce computing time. Kruskal–Wallis tests were used to determine whether the distribution was different across groups, and Wilcoxon tests were used to calculate pairwise post-hoc comparisons. Wilcoxon tests implemented in the *compare_means* function of the *ggpubr* R package were used, effects sizes (r) were calculated with the *wilcox_effsize* function implemented in the *statix* R package. Sample specific calculations of α-PD (and species richness), α-MPD and α-MNTD were computed using the 'pd', 'mpd' and 'mntd' functions of the *Picante* R package (Kembel et al., 2010). Linear models were used to compare the values of α-PD, α-MPD, and α-MNTD across samples, taking the logarithm of the species richness and the dataset (PP1 and PP2) as a fixed effect.

Differential abundance analysis

Using the Silva Taxonomic information (Quast et al., 2013), ASV raw counts were aggregated to the genus-level using a custom R script, removing the ASVs not assigned taxonomically to the genus-level. *Ancom v2.1* was used on the count data for the differential abundance analysis, using the default W statistic threshold of 0.7 (Lin, 2019). The 'zero-cut' parameter was set to 0.995 to consider all bacterial genera present in at least 21 samples (*n* = 4247), and the primer pair (PP1 and PP2) variable was taken as a random effect with the rand_formula parameter (*"~1|Dataset"*). We considered significantly enriched genera (i.e. cryospheric genera), the ones with a W statistic above the threshold (0.7) and a positive value of CLR mean difference. GGplot2 was used to modify the Ancom v2.1 figure showing the results of the differential abundance analysis. The 'heat_tree' function of the metacoder R package (version 0.3.4) was used to show the number of cryospheric bacterial genera, at various taxonomic level, using taxonomic trees (Foster et al., 2017). The results of this analysis can be found in the Data/ folder available on Zenodo under the name 'Ancom_amplicon_res.csv' file.

NCBI Refseq genomes properties

To assess the genome size and GC content of publicly available prokaryote genomes, a non-redundant list encompassing all the genera in our datasets was compiled. Thereafter, the list of prokaryote genomes (prokaryotes.txt) available on NCBI (Pruitt et al., 2007) was downloaded on March 15th, 2021 from https://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS. The prokaryote list was filtered based on the list of genera found in our dataset, simultaneously retrieving the accession IDs. The accession IDs were used to download the complete bacterial genome sequences using the *ncbi-genome-download* python package (https://github.com/kblin/ncbi-genome-download). The genome sizes for the downloaded genomes were additionally retrieved from the prokaryotes.txt metadata file. Subsequently, *Prodigal* (Hyatt et al., 2010) was used to annotate the open-reading frames per genome obtaining both the general feature format (*gff*) files and aminoacid fasta (*faa*). These were used thereafter as input used to estimate the predicted growth

time (in hours) and their codon usage analyses (CUB) using *gRodon* and *coRdon* (https://github.com/BioinfoHR/coRdon) R package respectively (Elek et al., 2019; Weissman et al., 2021). The amino acid enrichment analysis was performed on by converting the codon counts to amino acids using the R-package *Biostrings* using DEseq2 with default parameters (log-median ratio normalisation across genera). Wilcoxon tests implemented in the *compare_means* function of the *ggpubr* R package were used, effects sizes (*r*) were calculated with the *wilcox_effsize* function implemented in the *statix* R package. The relevant scripts and information for these analyses are openly available and included in the code availability section. The corresponding files used for this analysis can be found in the Data/ folder available on Zenodo under the names 'prokaryotes.txt', 'merged_all_codon_counts.txt' and 'merged_all_growth_prediction.txt'.

Structure of the cryospheric microbiome

Non-metric multidimensional scaling was used to visualise the composition of cryospheric bacterial communities according to the ecosystem types and primer pairs. For this, the 'metaMDS' function implemented in the package vegan was used with Bray-Curtis distances (Oksanen et al., 2013). The stress for the chosen value of $k = 2$ was 0.206. The 'adonis2' function was used to perform a PERMANOVA analysis to test the effect of the ecosystem type and the primer pairs on the composition of bacterial communities (Supplementary Table 4). Pairwise comparisons between ecosystem types were tested using the function 'pairwise.adonis2' (Martinez Arbizu, P., 2020). *P*-values were adjusted using the default Bonferroni method, to account for multiple comparisons.

The prevalence of each genus was modelled as the probability of presence using a logistic binomial regression (with the R stats 'glm' method), using the ecosystem type (snow/Ice, terrestrial, marine and freshwater) and the primer pair as fixed effect. To calculate the probability of occurrence in the cryosphere for each genus, the prediction was calculated for all ecosystem types and primer pair combinations, and averaged. The core microbiome was defined at 0.1% abundance, and 20% prevalence across the cryosphere, for genera present in at least one sample in all four ecosystem types (Supplementary Figure 2.2B). The core microbiome presence in the different ecosystem types was shown using an upset-plot using the *complex-upset* R package (Conway et al., 2017). The taxonomic tree available in Supplementary Figure 2.2A was created using the *Metacoder* R package (Foster et al., 2017). The α-diversity was calculated using Shannon's index with a custom R functions (Shannon and Weaver, 1949). To test the difference across ecosystems and datasets, the Wald-Type statistics implemented

in the 'GFD' function of the R *GFD* package was used (Supplementary Table 5). This test was performed instead of an ANOVA, as the data was not normally distributed. The mean values given by the function were used for the ecosystem comparison.

KEGG enrichment

The standard DESeq2 pipeline with default parameters was used on raw KEGG counts for the enrichment analysis, using the default Wald tests (Kanehisa and Goto, 2000; Love et al., 2014). We considered significantly enriched Kegg Orthologs (KOs) with an FDR adjusted p < 0.01 and a $\log_2$ fold-change >1. To unravel the contribution of these gene families to functional pathways, we ran *KEGGdecoder* (Graham et al., 2018) on the KOs enriched in cryospheric samples, to identify environmental-associated pathways in all samples.

To understand and unravel the origins of the gene families specific to the cryospheric metagenomes, contigs were taxonomically classified following which the specific gene families were mapped to the contigs. We used *Kraken2* to taxonomically assign all the contigs present in the metagenomes followed by custom python scripts (provided) to link the genes belonging to enriched KEGG orthologs (KO and the corresponding NCBI taxon ID (Federhen, 2012; Wood et al., 2019, p. 2). An R script using the NCBI *entrez* package was used to retrieve the taxonomy based on the taxon ID, and to get the genus-level taxonomy (Winter, 2017). To link the Silva genus taxonomies with their NCBI counterparts, the grep function included in R allowing partial matches was used to find Silva genera name matching the NCBI genus name. The DEseq2 results, KEGG-decoder output and taxonomy matches can be found in the Data/ folder of the Zenodo under the names 'KEGG_deseq_results.csv', 'KEGG_decoder_output', and 'KEGG_sign_tax_genera.csv', respectively.

Gene clusters and unassigned protein coding sequences

Predicted gene sequences annotated to the KEGG database and those unassigned were gathered into individual groups based on KEGG ID or Unassigned using a custom python script. 'annotation2gene.py'. The fasta files were subsequently concatenated and clustered to identify functional homologues within the dataset. We used *mmseq2* 'linclust' (Steinegger and Söding, 2017) to cluster the 41,068,842 gene sequences

found in the entire metagenomic dataset. Subsequently, fasta sequences associated with each cluster were retrieved into separated clusters (n = 12,125) and linked to the coverages to estimate abundances. *MAFFT* (Katoh and Standley, 2013) was then used to create a multiple sequence alignment of the sequences per cluster, while the 'cons' method from *EMBOSS* was used to generate a consensus sequence. The generated consensus sequences from each cluster were subsequently annotated and their identity verified against the UniProt TrEMBL (Bairoch et al., 2005) database. The pairwise identity of sequences within each cluster was measured using *CLUSTAL* (Larkin et al., 2007) 'distmat' option with the '–percent-id'. Wilcoxon tests implemented in the *compare_means* function of the *ggpubr* R package were used, effects sizes (r) were calculated with the *wilcox_effsize* function implemented in the *statix* R package. The unassigned clusters summary statistics and Uniprot matches can be retrieved on Zenodo, in the Data/ folder under the names 'Unassigned_clusters_stats.tsv', and 'unassigned_uniprot_matches.txt'.

❖ *Data availability*

The data generated in this study have been deposited in Zenodo, under https://doi.org/10.5281/zenodo.6541278. Source data used for figures are provided with this paper.

❖ *Code availability*

All scripts used for analyses, along with the conda environments, and additional information is provided in a Github repository archived on Zenodo: https://doi.org/10.5281/zenodo.6587400.

❖ *Figures & Tables*
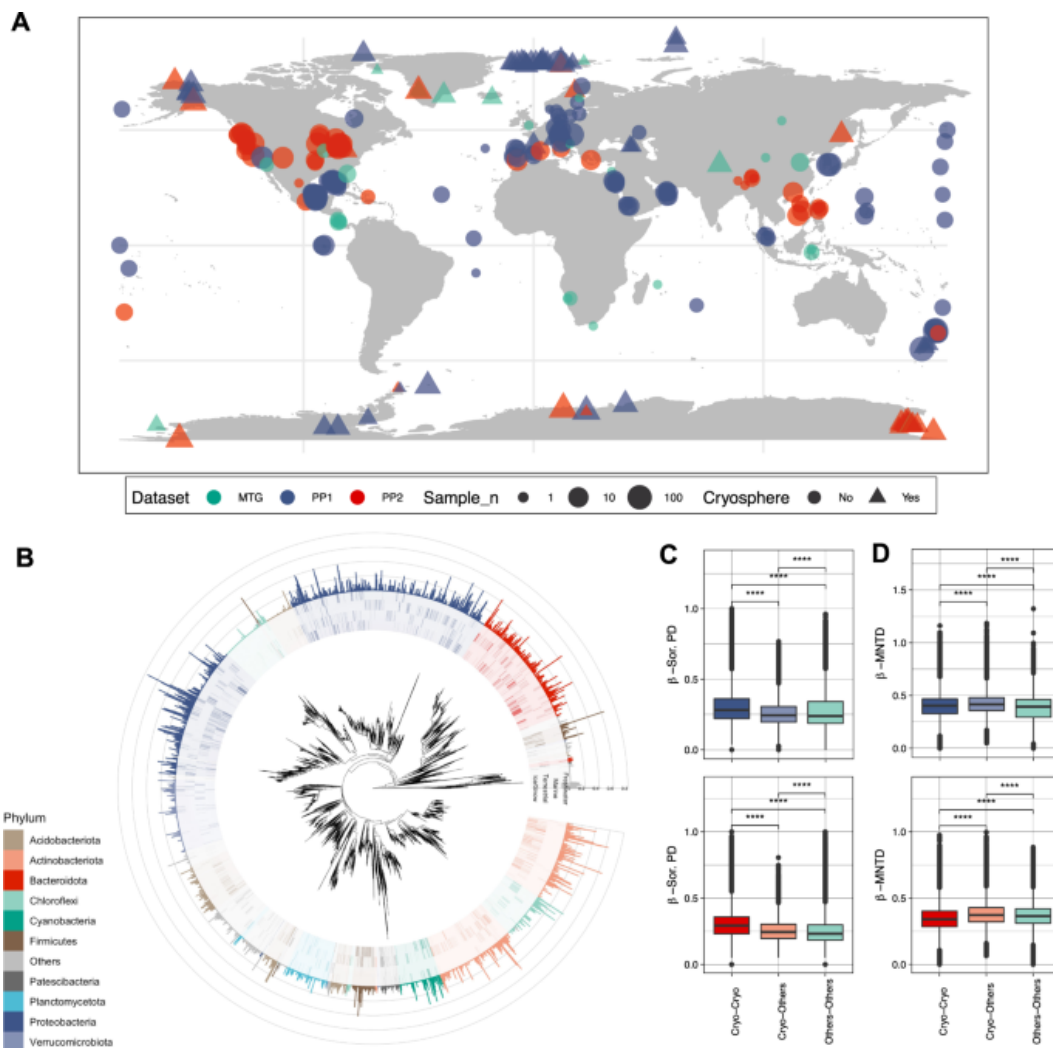
Figure 2. 1 *A unique cryospheric microbiome.*

A Geographic distribution of the 16 S rRNA gene samples for the two primer pairs (PP) and metagenomes for both cryospheric and non-cryospheric ecosystems, where GPS coordinates were available on NCBI. Symbol size denotes the number of samples per site (see Supplementary Table 7). **B** Phylogenetic tree based on abundant ASVs (>0.5% relative abundance in at least one sample) in the PP1 dataset. The heatmap (inner rings) shows the presence (at a > 0.5% relative abundance threshold) of ASVs in the four ecosystem types of the cryosphere (ice and snow, terrestrial, coastal ocean and freshwater). The barplot (outer ring) represents the coefficient for the SVM classifier analysis, highlighting discriminating ASVs. **C** Sorensen's phylogenetic index of β-diversity ($n_1 = n_2 = 84{,}461$ for PP1, and $n_1 = n_2 = 99{,}000$ for PP2) and **D** β-MNTD calculated across pairs of samples in the cryospheric samples (Cryo-Cryo), pairs of cryospheric and non-cryospheric samples (Cryo-Others) and pairs of non-cryospheric (Others-Others) samples (sample sizes are listed in Supplementary Table 2). The top panel (shades of blue) is for PP1, the bottom one (shades of red) for PP2; two-sided Wilcoxon tests were performed to assess significance in panels **C** and **D**; the Holm method was used to correct for multiple testing (****: 0–0.0001). Boxplots depict the median and the 25th and 75th quartiles, whiskers extend to values within 1.5 times the interquartile range, and the remaining points are outliers. Effect sizes and exact p-values are available in Supplementary Table 2. Source data are provided as a Source Data file.
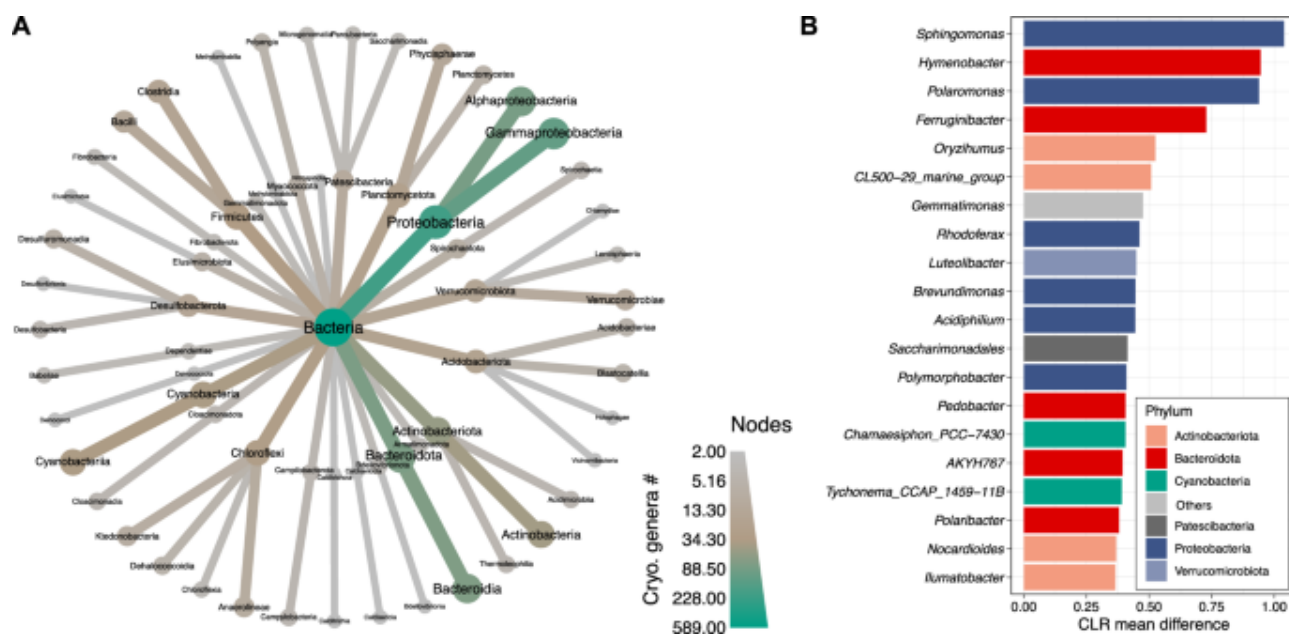
Figure 2. 2 *Cryospheric genera and shared genomic properties.*

A Taxonomic tree representing the number of cryospheric genera per taxon with colours. Only taxa containing at least two cryospheric genera are shown (down to the class level). **B** Bar plot showing the bacterial genera significantly enriched in the cryosphere with the highest centered log-ratio (CLR) mean difference (based on ANCOM analysis). The colours represent the phylum level taxonomic classification. Source data are provided as a Source Data file.
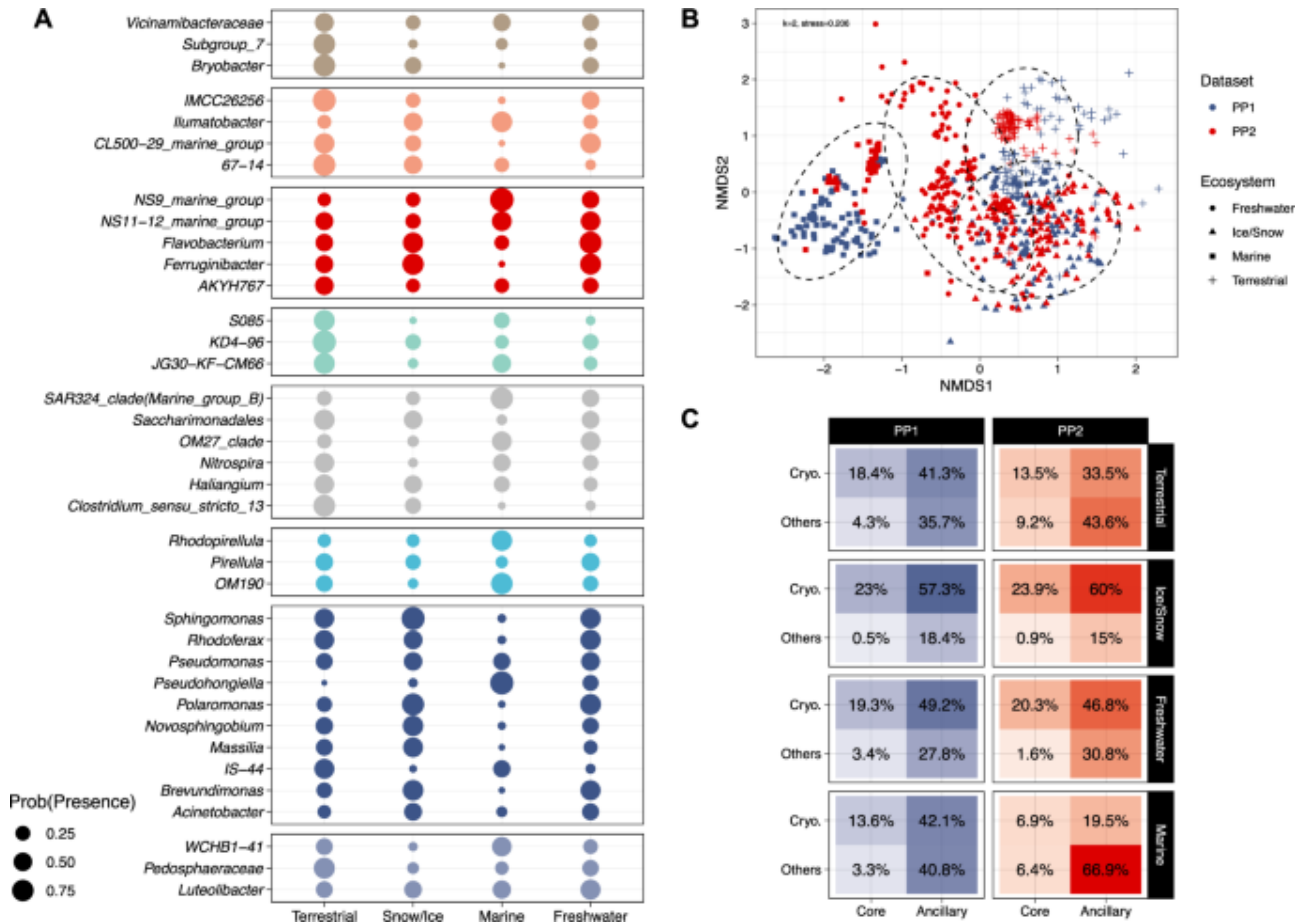
Figure 2. 3 *Microbiome structure across various cryospheric ecosystems.*

A The probability of presence of members of the core microbiome is shown across cryospheric environments. Colours and facets separate phylum-level taxonomic affiliation. **B** Unconstrained ordination showing differences (Bray-Curtis dissimilarity) of bacterial communities among different cryospheric ecosystems ($k = 2$, stress = 0.206). Hulls demark 95% confidence intervals for a multivariate t distribution for the respective ecosystem types. **C** Mean relative abundance (in percentage) of core/ancillary and cryospheric/others bacterial genera across the four ecosystem types and the two primer pairs. Source data are provided as a Source Data file.
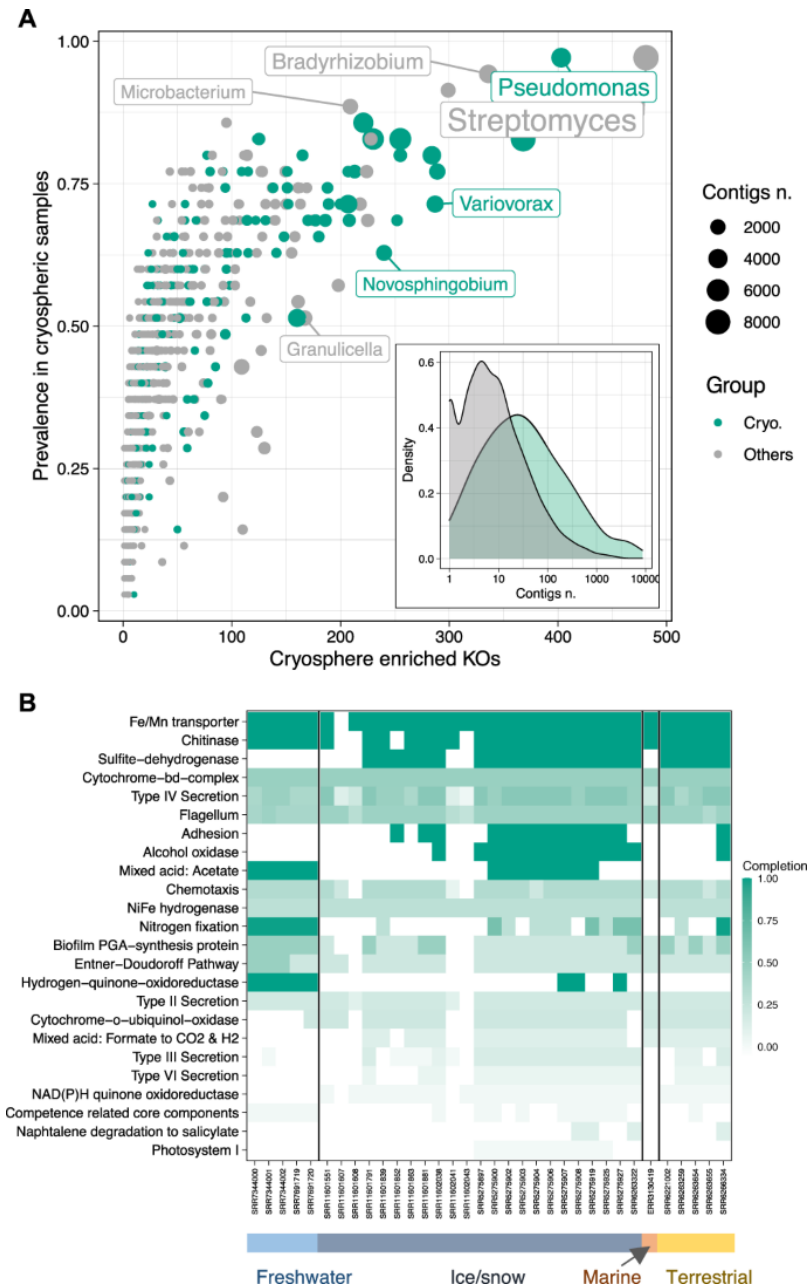
Figure 2. 4 *Functional enrichment analysis and taxonomy of enriched functions.*

A Prevalence represented against the number of enriched KOs in cryospheric samples across bacterial genera. The shading represents the cryospheric and others bacterial genera whereas symbol size represents the number of contigs taxonomically assigned to the respective genus within cryospheric metagenomes. The insert represents the distribution of the number of contigs harbouring cryospheric enriched KOs across the cryospheric and others genera. **B** Heatmap representing the completion of

pathways across cryospheric samples based only on the KOs enriched in the cryosphere. Source data are provided as a Source Data file.

Table 2. 1 *Description of the gene sequences clustering approach.*

Annotation	Category	Number of clusters	Uniprot match (%)
KEGG	Cryosphere	47	61.70
	Shared	1663	54.18
	Non-cryosphere	2325	55.14
Ambiguous	Cryosphere	113	40.71
	Shared	1056	52.65
	Non-cryosphere	3105	54.17
Unassigned	Cryosphere	170	17.65
	Shared	1524	5.18
	Non-cryosphere	2122	46.94

Table summarising the 12'125 largest gene sequence clusters present in at least two samples. The annotation refers to the assignment of the genes to one KEGG Orthologous group (KO), multiple KOs or unassigned (Ambiguous) and only unassigned (Unassigned). Distribution of assigned (KEGG), ambiguous and unassigned functional gene clusters highlighting the bias against cryospheric gene clusters. Shared refers to the representatives of both categories of samples contained gene sequences in the cluster. The number of clusters is shown, along with the proportion of clusters having a consensus sequence matching the UniProt database.

Chapter 3. Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams

Susheel Bhanu Busi[1], Massimo Bourquin[2], Stilianos Fodelianakis[2], Grégoire Michoud[2], Tyler J. Kohler[2], Hannes Peter[2], Paraskevi Pramateftaki[2], Michail Styllas[2], Matteo Tolosano[2], Vincent De Staercke[2], Martina Schön[2], Laura de Nies[1], Ramona Marasco[3], Daniele Daffonchio[3], Leïla Ezzat[1], Paul Wilmes[1, 4] & Tom J. Battin[2]

[1]Systems Ecology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]River Ecosystems Laboratory, Center for Alpine and Polar Environmental Research (ALPOLE), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[3]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[4]Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

❖ Authors' contribution

SBB, MB, SF, HP, PW, and TJB conceived of the project. MiST, MT, VDS, MaSc, and HP conducted the fieldwork. PP, SF and SBB extracted DNA, while SBB and PP prepared the metagenomic and metabarcoding libraries, and RM and DD performed the sequencing. SBB conceptualized the data analyses, while SBB, MB, SF, GM and LE performed the analyses. LdN contributed to the python scripts and Snakemake workflows for the analyses. SBB, MB, TJK, PW and TJB wrote the manuscript with significant input and editing from all coauthors.

These authors contributed equally: Susheel Bhanu Busi, Massimo Bourquin, Stilianos Fodelianakis.

In this work, I processed the 16s rRNA data, participated in the processing of the shotgun metagenomes, lead the MAGs phylogenomic analyses, lead the *Polaromonas* pangenome analysis, and produced several figures.

❖ Publication

❖ *Abstract*

In glacier-fed streams, ecological windows of opportunity allow complex microbial biofilms to develop and transiently form the basis of the food web, thereby controlling key ecosystem processes. Using metagenome-assembled genomes, we unravel strategies that allow biofilms to seize this opportunity in an ecosystem otherwise characterized by harsh environmental conditions. We observe a diverse microbiome spanning the entire tree of life including a rich virome. Various co-existing energy acquisition pathways point to diverse niches and the exploitation of available resources, likely fostering the establishment of complex biofilms during windows of opportunity. The wide occurrence of rhodopsins, besides chlorophyll, highlights the role of solar energy capture in these biofilms while internal carbon and nutrient cycling between photoautotrophs and heterotrophs may help overcome constraints imposed by oligotrophy in these habitats. Mechanisms potentially protecting bacteria against low temperatures and high UV-radiation are also revealed and the selective pressure of this environment is further highlighted by a phylogenomic analysis differentiating important components of the glacier-fed stream microbiome from other ecosystems. Our findings reveal key genomic underpinnings of adaptive traits contributing to the success of complex biofilms to exploit environmental opportunities in glacier-fed streams, which are now rapidly changing owing to global warming.

❖ *Introduction*

Ecosystems and their constituent biota are finely tuned to the seasonal variations of their environment. This phenology is particularly pronounced in glacier-fed streams (hereafter GFSs), which are commonly enveloped by snow cover and darkness in winter, and subject to high flow and sediment mobilization in summer. Yet, ecological 'windows of opportunity' arise in spring and autumn (Battin et al., 2004; Uehlinger et al., 2010) when nutrient (N, P) and light availability is elevated and streamflow is moderate (Battin et al., 2004; Kuhn, 2001; Uehlinger et al., 2010). During the onset of spring snowmelt, inorganic N that has accumulated from atmospheric deposition and concentrated at the snowpack surface is washed into GFSs (Kuhn, 2001; Milner et al., 2017), whereas hydrologic connectivity with various glacial sources (e.g., subglacial) can increase concentrations of phosphorus as the melt season progresses (Milner et

al., 2017; Tockner et al., 2002). Following the height of the melt season in summer, discharge and turbidity decline in autumn, again elevating nutrient concentrations and light availability. These favorable conditions allow algae and cyanobacteria to rapidly develop into 'green oases' of phototrophic biofilms. Partially due to the absence of major terrestrial organic matter subsidies from the catchment, this punctuated exploitation of solar energy in an otherwise energy-limited ecosystem transiently forms the base of the GFS food web and ecosystem energetics (Boix Canadell et al., 2021; Uehlinger et al., 2010). Such windows of opportunity may therefore function as 'ecosystem control points' (Bernhardt et al., 2017) with disproportionately high ecological processing rates affecting ecosystem dynamics relative to longer intervening time periods. These ecosystem control points are widely distributed across ecosystems and vary across spatial and temporal scales (Bernhardt et al., 2017). However, our understanding on the microbiology of the communities that facilitate ecosystem control points remains limited to date.

Owing to climate change, the mass balance and melting dynamics of mountain glaciers are rapidly changing worldwide, altering the annual distribution of runoff in GFSs (Huss and Hock, 2018). Invigorated glacial melt increases discharge and sediment delivery, but after glaciers shrink past a certain point (i.e., 'peak water'), GFSs are likely to become warmer, less turbid, and less hydrologically dynamic (Milner et al., 2017). These changes are almost certain to have substantial impacts on GFS ecosystem structure and function by either contracting or extending the duration of these windows of opportunity. It is therefore critical to understand how benthic biofilms operate during these times in order to predict how these ecosystems are likely to change in the future (Milner et al., 2017).

In streams, biofilms closely interact with the sedimentary environment (Battin et al., 2016). For example, extracellular polymeric substances (EPS) produced by biofilms bind fine sediment grains together, which can locally stabilize substrata, reducing scour and vertical permeability (Roncoroni et al., 2019). Similarly, boulders resist flow-induced disturbance to promote biofilm growth (Hoyle et al., 2017), and if protruding through the water column, may also increase light availability to further facilitate photosynthesis. Therefore, it seems advantageous for phototrophic biofilms to colonize boulders, which can be regarded as islands of stability in otherwise highly unstable GFS channels. These islands may allow biofilm growth to locally persist beyond the typical windows of opportunity (at least until snow cover), drive ecosystem energetics (i.e., gross primary production) (Boix Canadell et al., 2021), and to sustain the GFS food web and related benthic biodiversity (Fell et al., 2017; Milner et al., 2017).

The relationships between photoautotrophs (such as algae and cyanobacteria), prokaryotes and fungi regulate nutrient and carbon cycling, and therefore represent a fundamental ecological interface in aquatic ecosystems. This interface (i.e., the phycosphere) has received substantial attention in pelagic ecosystems over the last decades (Amin et al., 2015; Christie-Oleza et al., 2017; Cole, 1982; Seymour et al., 2017), but less so in stream ecosystems. While early work on phototrophic biofilms colonizing the benthic zone in streams has highlighted the role of algal–bacterial interactions for carbon and nutrient fluxes (Center, 1989; Haack and McFeters, 1982), we do not currently understand the fine-scale mechanisms of such interactions. For example, cyanobacteria produce pigments that protect the biofilm as a whole against harmful UV-radiation (Vincent et al., 1993), while mucilage-rich algal colonies (e.g., *Hydrurus* spp.) provide labile organic matter to heterotrophic microorganisms and facilitate their attachment. Such interactions may foster facultative interactions between photoautotrophs and other microorganisms, which, similarly to the phycosphere, may be particularly beneficial to microbial life in oligotrophic and harsh ecosystems such as GFSs. Unraveling the genomic and metabolic underpinnings of algal–bacterial relationships in biofilms helps to better understand the success of the biofilm mode of life in an extreme ecosystem.

Here we dissect the microbiome of GFSs and describe the genomic underpinnings of the adaptive mechanisms that potentially contribute to the success of complex biofilms. Using 16S rRNA and 18S rRNA gene amplicon sequencing, we assess the microbiome structure of biofilms associated with two sedimentary habitats that are common in GFSs, namely sandy sediments (i.e., epipsammic biofilms) and boulders (i.e., epilithic biofilms). We sampled geographically distant streams, transcending hemispheres (Southern Alps in New Zealand, NZ, and the Caucasus, CC), to draw more generalisable conclusions about microbiome structure and assembly. Furthermore, using genome-resolved metagenomics, we screen twenty-one epilithic biofilm microbiomes for energy pathways and cross-domain metabolic interactions. Our findings suggest the diversification of energy-acquiring pathways and metabolic interactions are relevant for epilithic biofilms to thrive during the ecological windows of opportunity, and beyond, within low-disturbance patches in GFSs. Moreover, our findings shed light on what the future biofilm mode of life in GFSs may look like as

glaciers shrink and GFS ecosystems are predicted to become more autotrophic (Milner et al., 2017).

❖ *Results and discussion*

Sedimentary habitats affect microbiome structure and assembly

We used 16S rRNA and 18S rRNA gene amplicon sequencing to compare the microbiome structure of 48 epipsammic and epilithic biofilm samples from GFSs in NZ and CC collected during spring and autumn, respectively (Methods) (Figure 3.1a; Supplementary Figure 3.1a, b). These seasons broadly align with the windows of opportunity in these GFSs; however, we recognize that epilithic biofilms, in particular, may extend beyond these windows well into summer or until snow coverage. We found that both prokaryotic and eukaryotic communities differed between the two habitat types in terms of community structure and alpha diversity (Figure 3.1b, c). Overall, taxonomic differences were even apparent at the phylum level, despite high inter-sample variability within the categories (Supplementary Figure 3.1c, d). Geography (i.e., NZ versus CC) explained 11.5% and 12.9% of the variability in the prokaryotic and eukaryotic datasets (db-RDA, $p < 0.05$ for both datasets), while sedimentary habitats explained an additional 10% and 8.3% of the variability (db-RDA, $p < 0.05$ for prokaryotes and eukaryotes).

The estimated α-diversity (i.e., richness of amplicon sequence variants; ASVs) was higher for both prokaryotes and eukaryotes in epipsammic biofilms when compared to epilithic biofilms (2–3 fold differences, non-parametric *t*-tests, $p < 0.001$) (Figure 3.1d, e). These observations are in accordance with findings by Tolotti and colleagues (Tolotti et al., 2020) where α-diversity of the epipsammic habitats were higher than the epilithic biofilms in rock glacier- and groundwater/precipitation-fed waters (Tolotti et al., 2020). It is plausible that continuous dispersal and mixing facilitated by the transport of fine sediments from various upstream sources (e.g., the subglacial environment and adjacent soils) leads to the greater diversity of the epipsammic biofilms. Overall, our results unravel distinct microbiome structures for both sediment habitats within the same GFS reaches. This agrees with previous studies (Tolotti et al., 2020), and more generally with the relationship between streambed physical variation and spatial biodiversity dynamics (Besemer et al., 2009; Risse-Buhl et al., 2020). Streambeds, including their biofilms, are understood as landscapes where dispersal among patches can shape biodiversity and resilience (Battin et al., 2007; Dzubakova et al., 2018; Palmer et al., 2000). Therefore, we hypothesized that epilithic communities are partially

structured by dispersal from epipsammic communities that typically dominate the GFS streambeds by area. Using Sloan's neutral community model (Sloan et al., 2006), we instead found that the composition of the epilithic biofilms is not dictated by a source-sink relationship with the epipsammic communities (Supplementary Note). In other words, the epilithic biofilm communities are not determined by epipsammic communities that typically surround the boulders within the complex landscape of the GFS streambed.

Metagenomics unveils the complexity of epilithic biofilms

To unveil the full complexity of the epilithic biofilms, we performed whole genome shotgun metagenomics on 21 epilithic samples from four GFSs each in NZ and CC (Supplementary Figure 3.1a, b); low biomass associated with sandy sediments precluded epipsammic biofilms from metagenomic analysis. Metagenomic sequencing, after quality filtering, yielded on average 1.2×10^8 ($\pm 1.4 \times 10^7$ s.d.) reads per sample which were assembled to obtain an average of 8.7×10^5 contigs per sample that were subsequently binned. Bacteria and eukaryotes dominated the biofilm communities across all samples (Supplementary Figure 3.2a). Seventy-three (70 bacteria and three archaea) medium-to-high quality (>70% completion, < 5% contamination) metagenome-assembled genomes (MAGs) from a total of 662 MAGs formed the pool of the prokaryotes. As seen from the phylogenomic analysis, the high-quality MAGs (n = 49, >90% completion and <5% contamination) span the bacterial tree of life. Based on the phylogenomic analyses along with the taxonomic information (Figure 3.2), we sought to further characterize these MAGs that could represent novel species or species that have not previously been reported (Figure 3.2a). We found that only 30% of these high-quality MAGS were annotated up to the family level, whereas the remaining MAGs could be taxonomically labelled at the genus level. Only high-quality MAGs were used for the phylogenetic analyses to mitigate disparities arising from incomplete MAGs. Aggregated at the genus level, *Polaromonas* was both abundant and prevalent in the biofilms along with representatives of *Flavobacterium*, *Cyanobacteria*, and unclassified MAGs from the Bacteroidota and Candidate Phyla Radiation (CPR; *Patescibacteria*) (Figure 3.2b). These taxa were found in over half of the samples, irrespective of geographic origin. The CPR bacteria have only recently been identified

based on genomic data (Hug et al., 2016), and *Patescibacteria* specifically have been reported from oligotrophic ecosystems, including groundwater (Chaudhari et al., 2021) and thermokarst lakes (Vigneron et al., 2020). Their apparently minimal biosynthetic and metabolic pathways may help them dwell in these ecosystems, which is of equal relevance in GFSs.

Alongside these bacteria, archaea contributed less than 1% to the microbiome of epilithic biofilms, with representatives of Asgardarchaeota, Crenarchaeota and Nanoarchaeota. Intriguingly, the recently discovered lineages of Asgardarchaeota (Cai et al., 2021; Y. Liu et al., 2021) have been reported from freshwater sediments, yet not from cryospheric environments. Algae, mostly diatoms and *Hydrurus* (Ochrophyta phylum), as well as dinoflagellata, were the most important photoautotrophs of the eukaryotic domain (Figure 3.2c). The prevalence of *Hydrurus* (~87% relative abundance) underscores the function of these filamentous algae as a resource to higher trophic levels in GFS (Niedrist and Füreder, 2018). Our metagenomic insights further support the notion that phototrophic biofilms are highly diverse with representatives from all three domains of life (Hug et al., 2016).

In addition to the archaeal, bacterial and eukaryotic community members, we also found a diverse viral community associated with epilithic biofilms (Supplementary Figure 3.2b). Most of the viruses were bacteriophages targeting abundant MAGs such as *Flavobacterium*, *Pseudomonas*, and *Bacillus* genera, but we also identified eukaryotic phages (i.e., *Paramecium bursaria* Chlorella virus). Few have studied viruses in stream biofilms to date (Payne et al., 2020), potentially because it was common wisdom that the biofilm mode of life protects bacteria from viral infection. While viruses have previously been shown to be abundant in glaciers (Anesio et al., 2007; Bellas et al., 2020), our findings provide evidence for a diverse and likely active viral community in GFS biofilms where they may influence bacterial growth and both carbon and nutrient cycling as on the glacier surface (Anesio et al., 2007).

Epilithic biofilms form the basis for a 'green' food web in glacier-fed streams

Cyanobacteria and eukaryotic algae dominated the photoautotrophs in the epilithic biofilms and hence form the basis of the 'green' food web during the windows of opportunity. While these photoautotrophs are well known to use chlorophyll to capture solar energy, little is known on retinal-based phototrophy using rhodopsins in GFSs. Intriguingly, we found that MAGs from sixteen out of twenty phyla in the epilithic

biofilms, including the abundant groups, such as Proteobacteria (*Polaromonas*) and Bacteroidota (*Flavobacterium*), encoded for (bacterio-) rhodopsins (Figure 3.3a). These also included genes encoding for light-harvesting complex 1 (LH1), reaction centre (RC) subunits (*pufBALM*), and transcriptional regulators (*ppsR*) required for aerobic anoxygenic phototrophs along with rhodopsins as a signature of energy-limitation adaptations (Figure 3.3a). Recently, rhodopsins were also reported to serve as a photoprotectant in *Flavobacterium* from glaciers (Q. Liu et al., 2021). Collectively, our findings unveil multiple strategies of photoautotrophy, which may help cyanobacteria and algae to maximize their utilization of solar energy and to thrive on boulders in GFSs.

In order to exploit the favorable habitat provided by boulders during and beyond the windows of opportunity in GFS, rapid growth may be advantageous for primary producers such as cyanobacteria. Moreover, functional independence from other microorganisms could allow them to seize environmental opportunities. To test this hypothesis, we assessed the relationship between projected times of growth (doubling time in hours) with the median KEGG pathway completion within each MAG. Given the partial completeness of the MAGs, including possibly missing metabolic modules, we performed a linear regression between median KEGG pathway completion and projected time of growth, accounting for MAG completion as a fixed effect. Strikingly, 86% of the cyanobacterial MAGs (n = 38 out of 44) exhibited decreased projected times of growth with an increase in median KEGG module completion per MAG (r_s = −0.47, Two-way ANOVA, adj. $p < 0.05$). These observations suggest that when encoding all genes to form a complete KEGG pathway, phototrophic taxa within these epilithic biofilms may indeed grow rapidly and be self-sufficient, putatively autonomously from other microorganisms from other (micro)organisms and fostering growth.

Given the energetic constraints in GFSs, it would be beneficial for bacterial heterotrophs to interact with these photoautotrophic (micro)organisms for meeting their energy and nutrient demands. To investigate such cross-domain relationships, we used network analyses and identified key interacting taxa based on positively co-occurring nodes using all prokaryotic and eukaryotic MAGs (see Methods). Based on a null model assessment (see Methods), our interaction networks showed preferential attachment within the nodes, along with increased centralities (i.e., degree and edge-betweenness, Supplementary Figure 3.3a, b), suggesting that the interactions within

these networks were not random. More importantly, the largest connected component (based on degree and betweenness centralities) of the interaction network contained taxa spanning archaea, bacteria and eukaryotic domains (Figure 3.3b and Supplementary Figure 3.3b). Though *Acidobacteria* had a high degree of centrality, both *Polaromonas* and *Methylotenera* demonstrated strong interactions (>0.6 betweenness centrality) with primary producers (including eukaryotic algae) and fungi. Specifically, *Polaromonas* had a strong interaction with algae, while *Methylotenera* co-occurred with *Chytridiomycetes* (Figure 3.3b). Interestingly, we found similarly connected nodes demonstrating cross-domain interactions within the largest component of the individual regions, i.e., NZ (Supplementary Figure 3.3c, d) and CC (Supplementary Figure 3.3e, f), albeit the two regions had varying numbers of edges (NZ = 205 and CC = 30). This suggests that inherent interactions within these GFS epilithic biofilms are conserved irrespective of geographic origins. These results also support our hypothesis of heterotrophic bacteria co-occurring with eukaryotes, primarily algae, for metabolic cross-feeding, similar to those occurring in the phycosphere (Seymour et al., 2017).

Furthermore, our results hint at the existence of a more cryptic interaction in epilithic biofilms between the parasitic fungi *Chytridiomycetes* and algae (mostly *Ochrophyta*). Fungal parasitism on pelagic algae has been recently reported to be more important than expected, even with consequences for carbon and nutrient cycling as mediated by the fungal shunt (Klawonn et al., 2021; Sánchez Barranco et al., 2020). The possibility of fungal parasitism on algae in epilithic biofilms further supports the notion of photoautotrophs forming the foundation of a complex food web in GFS ecosystems.

Genomic underpinnings of algae–bacteria metabolic interactions

As photoautotrophs grow and senesce, they increasingly exude intracellular material into their ambient environment, where it can be metabolized by heterotrophic bacteria through extracellular enzymes (Chróst, 1991). To explore this metabolic cross-feeding between bacterial heterotrophs and algae, we assessed the MAGs for genes encoding five common extracellular enzymes required for cleaving complex polysaccharides, phosphomonoesters and proteins (Sinsabaugh et al., 2009). Not unexpectedly, these genes were associated with bacterial heterotrophs rather than with the photoautotrophs (Supplementary Figure 3.4), which suggests adapted genomic traits to meet specific metabolic needs of the heterotrophs. However, based on the presence of extracellular enzyme genes among *Cyanobacteria*, we cannot discount the possibility of mixotrophy in the epilithic biofilms (Supplementary Figure 3.4b). Additionally, genes associated with mixotrophy, such as those encoding for auto- and

heterotrophic pathways, were also found in other abundant members of the epilithic microbiome (e.g., *Proteobacteria*). The widespread occurrence of mixotrophy in planktonic communities (Stoecker and Lavrentyev, 2018), including members of the *Cyanobacteria*, and the ensuing food web dichotomy is considered as an adaptive strategy to oligotrophic and cold ecosystems (e.g., the polar sea (Stoecker and Lavrentyev, 2018) and alpine lakes (Waibel et al., 2019)). Therefore, we argue that mixotrophy may also be an important trait of Cyanobacteria within GFS biofilms.

Carbohydrate-active enzymes (CAZymes) are the primary tools used by heterotrophic bacteria to initiate the degradation of polysaccharides, largely algae-derived in the GFS epilithic biofilms. To shed light on this potential trophic interaction identified through specific extracellular enzyme activities (EEAs), we tested if all the CAZymes in the metagenomes covaried with the abundance of eukaryotes. Overall, we found positive correlations between eukaryote abundances and CAZymes, particularly carbohydrate-binding modules (CBM) and glycoside hydrolases (GH) (Supplementary Figure 3.4d). More specifically, these correlations were particularly pronounced for GH and some of the algal groups (e.g., *Ochrophyta*, *Haptophyta*, *Cryptophyta*) that we found at relatively high abundances in the epilithic biofilms (Figure 3.3c and Supplementary Figure 3.4d). As some of these algae are known to copiously produce sulfated carbohydrates (Avcı et al., 2020), we suggest a similar involvement of CAZymes (Supplementary Data 1) in relation to polysaccharide degradation in GFS epilithic biofilms as recently reported from *Verrucomicrobia* isolates[44]. Given that sulfated carbohydrates are more resistant to bacterial degradation than other carbohydrates (Sichert et al., 2020), our findings suggest that they are still relevant to carbon turnover in an ecosystem that is inherently carbon limited.

In order to understand whether functions potentially geared towards cross-domain interactions were enriched in epilithic biofilms in GFSs, we compared the KEGG orthology (KO) annotations from our metagenomes to 105 metagenomes from a wide range of ecosystems (Supplementary Data 2). Strikingly, we found that whole metagenome comparisons revealed that KOs associated with quorum sensing, vitamin B12 (cobalamin) transporters and thiamine biosynthesis were enriched in epilithic GFS biofilms compared to other ecosystems (Supplementary Data 3). The associated pathways and their completion levels were evaluated using KEGGDecoder (Figure 3.3d;

Supplementary Figure 3.5) indicating a high completion of pathways associated with cross-domain interactions. These findings are in line with previous genomic insights into algal–bacterial interactions (Croft et al., 2005; Zhou et al., 2016), specifically with the observed upregulation of vitamin biosynthesis in bacteria (*Halomonas*) growing in the presence of algal extracts.

Furthermore, several MAGs were found to encode genes (e.g., quorum sensing, cobalamin metabolism, tryptophan synthesis) potentially facilitating algal–bacterial interactions (Figure 3.3a). Particularly, cobalamin metabolism may be relevant for nutrient acquisition in algal–bacterial relationships (Grossman, 2016), whereas tryptophan was reported as a key signalling molecule involved in interactions between bacteria and associated phytoplankton (Amin et al., 2015; Segev et al., 2016). Collectively these genomic insights stress cross-domain interactions as an adaptive potential that the epilithic microorganisms have developed to exploit the window of opportunity in GFSs.

Energy acquisition and biogeochemical pathways in epilithic biofilm MAGs

The dominance (~88%) of MAGs encoding for organic carbon metabolism suggests a 'baseline' heterotrophy in GFSs likely supported by organic carbon subsidies from melting glaciers (Boix Canadell et al., 2021; Fellman et al., 2015; Hood et al., 2015) 'green food web' during the windows of opportunity, potentially sustaining metabolic interactions between primary producers and heterotrophs. Given the notoriously low concentrations of dissolved organic carbon in GFSs (Boix Canadell et al., 2019; Hood et al., 2015; Singer et al., 2012), including our study sites in NZ (96.18 ± 21.35 µg C L^{-1}) and CC (221.36 ± 31.01 µg C L^{-1}), we suggest that the 'green food web' dominates over allochthonous subsidies.

Exploring the gene repertoire of the epilithic biofilms, we found that Cyanobacteria were one of the largest bacterial contributors to carbon fixation along with *Bacteroidota* and few *Gammaproteobacteria* (Figure 3.4a). An in-depth analysis across the 662 MAGs revealed that 583 MAGs encoded genes involved in organic carbon oxidation, while 120 MAGs encoded genes involved in CO_2 fixation. In line with the above findings, the majority of these MAGs was identified as *Cyanobacteria* along with few other phyla such as *Proteobacteria*, *Asgardarchaeota*, *Crenarchaeota* and *Huberarchaeota*. We also note that 351 MAGs encoded genes for fermentation (Figure 3.4b) spanning several phyla, including *Actinobacteriota*, *Bacteroidota*, *Patescibacteria*, *Planctomycetota* and *Verrucomicrobiota*.

For biofilms to thrive in GFSs, particularly during the windows of opportunity, it appears opportune to diversify the exploitation of energy sources. Therefore, we performed an in-depth characterisation of chemolithotrophic pathways to explore the potential role of minerals derived from the glacial comminution of bedrock as an energy source for microorganisms (Anesio et al., 2017). The prevalence of the *sox* gene cluster in representatives of the *Bacteriodota* (UBA7662) and *Bdellovibrionota* reveals the potential importance of inorganic sulfur oxidation in epilithic biofilms. This notion is supported by the broad occurrence of sulfur dioxygenases (SDOs) across the various phyla that facilitate sulfur oxidation (Figure 3.4c). Interestingly, Tranter and Raiswell suggested that sulfates derived from sulfide oxidation in comminuted bedrock (Tranter et al., 1989) potentially increase sulfur availability and acquisition in glacial meltwaters (Tranter et al., 1993). Sulfide oxidation can stimulate carbonate weathering with the resulting CO_2 potentially being fixed by algae and cyanobacteria in the epilithic biofilms—a link that appears relevant given that GFSs are often undersaturated in CO_2 (St. Pierre et al., 2019). Furthermore, we found that almost all MAGs encoded for group IV hydrogen dehydrogenases (NiFe_Gp4; Figure 3.4c), which potentially serve as an alternate energy acquisition pathway. Hydrogen dehydrogenases have recently been reported to support primary production in various glacial and other extreme environments (Dunham et al., 2021; Hernández et al., 2020). This suggests that lithogenic hydrogen may also contribute energy to bacteria within the epilithic biofilms.

Genomic insights into the nitrogen cycle revealed the Dissimilatory Nitrate Reduction to Ammonium (DNRA, or nitrite ammonification) and, to a lesser extent, denitrification, as major pathways (Figure 3.4d). Relatively little is known regarding these two competing pathways in stream biofilms or sediments (Quick et al., 2019), particularly in GFSs. This is in line with other ecosystems where DNRA is favoured over denitrification when alternate electron donors prevail over nitrate (Kuypers et al., 2018). For instance, predicting metagenomes from 16S rRNA sequences, Ren et al. (Ren et al., 2017b) found DNRA to be an important pathway in GFSs, suggesting that bacteria use inorganic nitrogen more as an energy source than a source for biosynthesis. Our analyses revealed *Burkholderiales* (*Gammaproteobacteria*) as the largest contributor to nitrate assimilation and ammonia-oxidation genes (Figure 3.4a, c). DNRA, if not conducive to N_2O production, would enhance nitrogen recycling within epilithic

biofilms through ammonia assimilation by algae and cyanobacteria, for instance. Our genomic evidence for nitrogen recycling that potentially overwhelms nitrogen losses through denitrification is corroborated by flux measurements from microbial mats in Antarctic GFSs (Gooseff et al., 2004), and highlights recycling as a strategy to cope with nutrient limitation in glacier ecosystems (Gooseff et al., 2004; Kohler et al., 2020a; Varin et al., 2010).

Strikingly, we found only few MAGs, mostly belonging to *Deinococcota*, *Gammaproteobacteria*, *Beijerinckiaceae* and *Crenarchaeota*, involved in the oxidation of ammonia and nitrite, potentially leading to the accumulation of nitrate. The involvement of archaea would be in line with recent studies showing ammonia oxidation by archaea in Arctic soils (Alves et al., 2019) and with the observation that archaea couple ammonia oxidation with biomass formation (i.e., via CO_2 fixation) (Könneke et al., 2014). Our finding that archaeal MAGs encode for carbon fixation genes (Figure 3.4b) further highlight their role in ammonia oxidation and biomass accrual in epilithic biofilms. Overall, the overlap of metabolic capacities within the MAGs suggests that the epilithic biofilms efficiently recycle carbon and nutrients. Internal recycling in stream biofilms is thought to be facilitated by increased residence times of water and contained solutes within the biofilms compared to the overlying water (Battin et al., 2003), which is certainly an advantage in a losing ecosystem such as GFSs.

Genomic underpinnings of adaptation to the extreme GFS environment

The GFS environment is extreme as illustrated by near-freezing temperatures, high UV-radiation, and high flow velocities. To assess potential adaptive traits of bacteria dwelling in epilithic biofilms, we first performed a phylogenomic analysis of *Polaromonas* spp., one of the most abundant and prevalent genera in the studied GFSs. Our analysis revealed that a few of the GFS *Polaromonas* formed clades that are distinct from *Polaromonas* identified in other environments (Methods), thus potentially comprising novel 'species' (Figure 3.5a). This phylogenomic pattern indicates that *Polaromonas* has evolved traits that facilitates its success in GFS, both in NZ and CC. To identify such traits, we created a pangenome and performed an enrichment analysis for clusters of orthologous genes. We found three categories that were significantly enriched in GFS *Polaromonas* compared to those from other environments (Supplementary Data 4). Two categories are related to defense mechanisms, both general and transcription, and one to energy production (Figure 3.5b). It is plausible that these mechanisms are related to high UV-radiation (Cockell et al., 2002; Sommaruga, 2001) and oxidative stress (Margesin and Collins, 2019), as well as to cold stress

responses as previously reported from other bacteria (De Maayer et al., 2014; Tribelli and López, 2018; Varin et al., 2012). Furthermore, the presence of CRISPR-Cas proteins in the enriched clusters of orthologous genes (COGs) hint at defense mechanisms against phages (Supplementary Data 4), which we showed to be present in the epilithic biofilms. This is in accordance with reports demonstrating that cryospheric bacteria (such as *Janthinobacterium* spp.) develop defense strategies, including biofilm formation (Alonso-Sáez et al., 2014) and extracellular vesicle formation (Hornung et al., 2013) to escape viruses. On the other hand, the transcription of 'defense mechanism' genes have been linked to cold adaptation in psychrophiles (De Maayer et al., 2014). Cold-shock proteins regulate transcription at low temperature, while genes involved in membrane biogenesis (Maillot et al., 2019) and membrane transport proteins (Konings et al., 2002), several of which are also enriched in the GFS *Polaromonas* genomes, are up-regulated. For example, in the psychrophilic *Colwellia psychrerythraea* 34H, adaptation to cold includes the maintenance of the cell membrane in a liquid-crystalline state via the expression of genes involved in polyunsaturated fatty acid synthesis (Methé et al., 2005). Similarly, ATP-driven or proton motive secondary transport systems have been associated with solute transfers across membranes in bacteria and archaea as an adaptation to the cold (Ayala-del-Río et al., 2010).

Our insights into the adaptive potential of *Polaromonas* to the GFS environment prompted us to expand our search for adaptive traits across all MAGs from the epilithic biofilms. Querying for 76 genetic traits spanning nine categories related to cold adaptation (Tribelli and López, 2018), we indeed found distinct patterns of genomic adaptation across MAGs (Figure 3.5c). Several MAGs encoded for genes associated with membrane and peptidoglycan alterations, cold and heat shock proteins, oxidative stress, and transcription/translation factors alongside DNA replication and repair. While all major phyla encoded for adaptive traits related to the outer membrane and cell wall, Proteobacteria were the predominant group with an overall higher copy number of genes (~5 copies/genome), albeit insignificant compared to other phyla, involved in counteracting osmotic and oxidative stress. This was followed by Bacteroidota, Cyanobacteria and Actinobacteriota with three, two and two copies per genome respectively. Interestingly, we found that Patescibacteria MAGs had significantly lower copies of cold adaptation genes, whilst both Actinobacteria and

Asgardarchaeota demonstrated a significantly higher number of osmotic stress genes (Supplementary Data 7). This is in line with metagenomic studies reporting an enrichment of sigma B genes in Antarctic mats, allowing for surviving severe osmotic stress during freezing (Varin et al., 2012). Similarly, *Psychrobacter arcticus* (Ayala-del-Río et al., 2010) and *Planococcus halocryophilus* Or1 (Mykytczuk et al., 2013, p. 15) were shown to have specific genomic modifications, particularly with genes involved in putrescine and spermidine accumulation, both of which are associated with alleviating oxidative stress. Furthermore, MAGs from Proteobacteria were characterized by a high prevalence of genes potentially expressed in response to stressors, such as UV and reactive oxygen species (Figure 3.5c).

Our genomic insights into possible adaptive traits of epilithic microorganisms may also contribute to our understanding of their adaptation beyond the windows of opportunity when the GFS environment is even harsher. In fact, with the onset of winter and during winter, GFSs partially freeze and become snow-covered thereby inhibiting primary production. Mixotrophy as observed within the Cyanobacteria (Supplementary Figure 3.4b) would be advantageous during these periods.

Furthermore, it is recognized that cell membrane alterations and lipid composition allow for withstanding cold conditions (Tribelli and López, 2018). Our observations regarding several MAGs encoding genes associated with membrane and peptidoglycan alterations are concordant with previous reports of increased membrane fluidity in *Psychrobacter arcticus* 273–4 (Ayala-del-Río et al., 2010), *Sphingopyxis alaskensis* (Ting et al., 2010), and *Pseudomonas extremaustralis* (Tribelli et al., 2015). Simultaneously, at low temperatures oxygen solubility increases, potentially generating reactive oxygen species and subsequently leading to oxidative stress (Blagojevic et al., 2011). As reported above, we observed several MAGs encoding genes to counteract this phenomenon which may be even more critical as temperatures may decrease outside of the 'warmer' windows of opportunity. Overall, the diversity of the cold adaptation genes and their potential mechanisms within MAGs support the notion that these taxa are potentially equipped to deal with the even harsher GFS environment outside the windows of opportunity.

In conclusion, our genome-resolved metagenomic analyses have set the stage for a mechanistic understanding of how the diversification of energy and matter acquisition pathways, metabolic interactions, and genomic adaptations to harsh environmental conditions allow GFS biofilms to persist and thrive during windows of opportunity and beyond. We acknowledge that a metagenomic time series outside and throughout

windows of opportunity would be required to substantiate some of our observations. Nevertheless, our findings shed light on boulders as important habitats that confer stability to biofilms even outside the typical windows of opportunity. GFSs count among the ecosystems that are most vulnerable to climate change. Therefore, our findings open a window into the future of how microbial life, with a strong photoautotrophic component, may look like in GFSs when the environmental conditions become more favorable for primary producers as glaciers shrink.

❖ Methods

Sample collection

We sampled a total of eight GFSs from the Southern Alps in New Zealand Southern Alps and the Caucasus in Russia in early- and mid-2019, respectively, for a total of 27 epipsammic samples taken from sandy sediments and 21 epilithic biofilm samples from boulders adjacent to the epipsammic samples (Supplementary Data 5). In order to have comparable samples, the collection was largely constrained to the vernal and autumnal windows of opportunity, respectively. Epipsammic samples were collected from each GFS by first identifying three patches within a reach of ~5–10 m. From each patch, epipsammic samples were taken from the <5 cm surface of the streambed with a flame-sterilized metal scoop and sieved to retain the 250 μm to 3.15 mm size fraction. While three epipsammic samples were taken from each stream, epilithic samples were taken opportunistically from up to three boulders per reach (Supplementary Data 5) due to their heterogeneity within and among the streams due to the unequal presence of boulders in each GFS. Epilithic biofilms were sampled using a sterilized metal spatula. All samples were immediately flash-frozen in liquid nitrogen in the field and transported and stored frozen pending DNA extraction. Streamwater turbidity, conductivity, temperature, and pH were measured in situ during the sampling (Supplementary Data 5). Samples for the determination of streamwater dissolved organic carbon and inorganic nutrient concentrations were filtered through pre-combusted (450 °C) glass microfiber filters (GF/F, Whatman), frozen, and analyzed in the laboratory. DOC concentration was measured with a TOC carbon analyzer (Sievers M9 TOC Analyser, GE). Phosphate, ammonium, nitrite and nitrate were measured with

a continuous flow injection analyzer (Lachat QuikChem 8500, methods 10-115-01-1-M (PO_4), 10-107-04-1-B (NO_3/NO_2) and 10-107-06-3-D (NH_3)) (Supplementary Data 5).

DNA extraction and purification

A previously established protocol (Busi et al., 2020) was used to extract DNA from all samples. Briefly, 5 g of epipsammic and 0.05–0.1 g of epilithic biofilm were subjected to a phenol:chloroform-based extraction and purification method. The differential input volume for the DNA extractions were established to account for the differences in biomass between the epipsammic and epilithic biofilms. The samples were treated with a lysis buffer containing SDS along with 0.1 M Tris-HCl pH 7.5, 0.05 M EDTA pH 8, 1.25% SDS and RNase A (10 µl: 100 mg/ml). The samples were vortexed and incubated at 37 °C for 1 h. Proteinase K (100 µl; 20 mg/ml) was subsequently added and further incubated at 70 °C for 10 min. Samples were purified once with phenol/chloroform/isoamyl alcohol (ratio 25:24:1, pH 8) and the supernatant was subsequently extracted with a 24:1 ratio chloroform/isoamyl alcohol. Linear polyacrylamide (LPA) was used along with sodium acetate and ice-cold isopropanol for precipitating that DNA overnight at −20 °C. For epilithic biofilms, the entire protocol was adapted to a smaller scale due to the availability of higher DNA concentrations compared to sediment. The former was treated with 0.75 ml of lysis buffer (instead of 5 ml for sediment) and all subsequent volumes of reagents were adapted accordingly (see supplementary material). Furthermore, a mechanical lysis step of bead-beating was necessary along with a lysis buffer to facilitate DNA release from the more developed epilithic biofilms. Due to the higher DNA yields, the addition of LPA was omitted from the DNA precipitation step. DNA quantification was performed for all samples with the Qubit dsDNA HS kit (Invitrogen).

Metabarcoding library preparation and sequencing

The prokaryotic 16S rRNA gene metabarcoding library preparation was performed as described in Fodelianakis et al. (Fodelianakis et al., 2022), targeting the V3-V4 hypervariable region of the 16S rRNA gene with the 341 F (5'-CCTACGGGNGGCWGCAG-3') and 785R (5'-GACTACHVGGGTATCTAATCC-3') primers and following Illumina guidelines for 16S metagenomic library preparation for the MiSeq system. The eukaryotic 18 S rRNA gene metabarcoding library preparation was performed likewise but using the TAReuk454F (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukREV3 (5'- CTTTCGTTCTTGATYRA-3') primers to target the 18 S rRNA gene V4 loop (Stoeck et al., 2010). Samples were sequenced using a 300-bp paired-end protocol

partly in the Genomic Technologies Facility of the University of Lausanne (27 epipsammic samples) and partly at the Biological Core Lab of the King Abdullah University of Science and Technology (21 epilithic samples).

Metabarcoding analyses

The 16S rRNA gene metabarcoding data were analysed using a combination of Trimmomatic (Bolger et al., 2014) and QIIME2 (Bolyen et al., 2019) as described in Fodelianakis et al. (Fodelianakis et al., 2022), with the exception that here the latest SILVA database (Quast et al., 2013) v138.1 was used for taxonomic classification of 16S rRNA and 18S rRNA gene amplicons. Non-bacterial ASVs including those affiliated to archaea, chloroplasts and mitochondria were discarded from the 16S rRNA amplicon dataset in all downstream analyses. ASVs observed only once were removed from both 16S rRNA and 18 S rRNA amplicon datasets. Diversity analyses were performed in R using the *vegan* (Dixon, 2003) and *metacoder* (Foster et al., 2017) packages. For non-metric multidimensional scaling (nMDS) and distance-based redundancy (db-RDA) analyses data were *log(x + 1)* transformed and the *capscale* and *ordiR2step* (backwards direction, 200 permutations) functions from *vegan* were used. To test for a source-sink hypothesis from epipsammic to epilithic, the Sloan's Neutral Community Model (Sloan et al., 2006) was used based on the R implementation developed by Burns et al. (Burns et al., 2016).

Whole-genome shotgun libraries and sequencing

All epilithic biofilm DNA samples underwent random shotgun sequencing following library preparation using the NEBNext Ultra II FS library kit (Biolabs, 2020). Briefly, 50 ng of DNA was used for constructing metagenomic libraries under 6 PCR amplification cycles, following enzymatic fragmentation of the input DNA for 12.5 min. The average insert size of the libraries was 450 bp. Qubit (Invitrogen) was used to quantify the libraries followed by quality assessment using the Bioanalyzer from Agilent. Sequencing was performed at the Functional Genomics Centre Zurich on a NovaSeq (Illumina) using a S4 flowcell.

Metagenomic preprocessing, assembly, binning, and analyses

For processing metagenomic sequence data, we used the Integrated Meta-omic Pipeline (IMP) (Narayanasamy et al., 2016) workflow to process paired forward and reverse reads using version 3.0 (commit# 9672c874; available at https://git-r3lab.uni.lu/IMP/imp3) (Heintz-Buschart et al., 2016). IMP's workflow includes preprocessing, assembly, genome reconstructions and additional functional analysis of genes based on custom databases in a reproducible manner. Briefly, adapter trimming is followed by an iterative assembly using MEGAHIT v1.2.9 (Li et al., 2015). Concurrently, MetaBAT2 v2.12.1 (Kang et al., 2019) and MaxBin2 v2.2.7 (Wu et al., 2016) are used for binning in addition to an in-house method, binny (Hickl et al., 2022), for reconstructing metagenome-assembled genomes (MAGs). Binning was completed by selecting a non-redundant set of MAGs using DASTool (Sieber et al., 2018) based on a score threshold of 0.7. The quality of the MAGs was assessed using CheckM v1.1.3 (Parks et al., 2015), while taxonomy was assigned using the GTDB-toolkit v1.4.1 (Chaumeil et al., 2020).

For the downstream analyses including identification of viruses, VIBRANT v1.2.1 (Kieft et al., 2020) was used on the metagenomic assemblies. The output from this was used to identify the viral taxa using vConTACT2 v0.9.22 (Zablocki et al., 2019, p. 2). Independently, the viral contigs were also validated using CheckV v0.7.0 (Nayfach et al., 2021). To estimate the overall abundances of eukaryotes along with prokaryotes including archaea, we used EUKulele v1.0.5 (Krinos et al., 2020) with both the MMETSP and the PhyloDB databases, run separately, to confirm the detected eukaryotic profiles. To understand the overall metabolic and functional potential of the metagenome and reconstructed MAGs we used MANTIS (Queirós et al., 2021). Additionally, we used METABOLIC v4.0 (Zhou et al., 2022), metabolisHMM v2.21 (McDaniel et al., 2019), and Lithogenie from MagicLamp v1.0 (https://github.com/Arkadiy-Garber/MagicLamp) to identify metabolic and biogeochemical pathways relevant for determining nutritional phenotypes of all MAGs along with the '*anvi-estimate-metabolism*' function from anvi'o (Eren et al., 2015). This information was manually validated based on the different tools to identify which MAGs encode for the respective pathways. Subsequently, to determine the growth rates of prokaryotes, we used codon usage statistics for detecting optimization of genes that are highly expressed, as an indicator of maximal growth rates with gRodon v1.0 (Weissman et al., 2021). All the parameters, databases, and relevant code for the analyses described above are openly available at https://git-r3lab.uni.lu/susheel.busi/nomis_pipeline and included in the Code availability section.

Eukaryote assembly and binning

To obtain eukaryotic MAGs, an alternate, custom pipeline (https://github.com/Mass23/NOMIS_ENSEMBLE/tree/coassembly) was established for coassembling the twenty-one epilithic biofilm sequence data with subsequent binning. Individual samples were first preprocessed similar to the workflow used in IMP, i.e., using FastP v0.20.0 (Chen et al., 2018). Subsequently, the reads were deduplicated to avoid overlap and enhance computation efficiency using *clumpify.sh* from the BBmap suite v38.79 ("BBMap: A Fast, Accurate, Splice-Aware Aligner (Conference) | OSTI.GOV"). Thereafter, any reads mapping to bacteria or viruses were removed by filtering the reads against a Kraken2 v2.0.9beta (Wood et al., 2019) maxikraken database available at https://lomanlab.github.io/mockcommunity/mc_databases.html. Only reads that were unknown or mapping to eukaryotes were retained and concatenated. This was followed by another round of deduplication using *clumpify.sh*. The concatenated reads were assembled using MEGAHIT v1.2.7 with the following options: *-kmin-1pass -m 0.9 -k-list 27,37,47,57,67,77,87 -min-contig-len 1000*. Following assembly, EukRep v0.6.7 (West et al., 2018) was used for retrieving eukaryotic contigs with a minimum length of 2000 bp and the '*-m strict*' flag. These contigs were used for binning into MAGs as described herein.

Eukaryotic MAGs were binned using CONCOCT v1.1.0 (Alneberg et al., 2013). To do this, coverages were estimated for the contigs by mapping the reads of all samples against the contigs using the coverm v0.6.1 (https://github.com/wwood/CoverM) to generate bam files. These files were then used to generate a table with coverage depth information per sample. The protein coding genes of the MAGs was predicted with MetaEuk v4.a0f584d (Levy Karin et al., 2020) with their in-house database made with MERC, MMETSP and Uniclust50 (http://wwwuser.gwdg.de/~compbiol/metaeuk/). The annotation was then subsequently done with eggNOG-mapper v2.1.0 (Huerta-Cepas et al., 2019). The completeness and contamination of the MAGs were assessed with Busco v5.0.0 (Simão et al., 2015) and the eukaryotic lineage (255 genes). We determined their taxonomy by comparing the results of the EUKulele v1.0.3 (Krinos et al., 2020) and EukCC v0.3 (Saary et al., 2020) along with homology comparisons with

publicly available genomes not included in the previous tools by protein BLAST v2.10.0 (Altschul et al., 1990).

Co-occurrence interaction networks

Co-occurrence networks between the pro- and eukaryotic MAGs were constructed using an average of the distance matrices created from SparCC (Friedman and Alm, 2012), Spearman's correlation and SpiecEasi (Kurtz et al., 2015), where the networks were constructed using the 'Meinshausen and Bühlmann (mb)' method. Nodes with fewer than two degrees were discarded to identify cliques with three or more interactions, while negative edges were removed to visualize only mutualistic relationships. The matrix was visualised using the *igraph* (Csardi and Nepusz) R package. The largest component from the overall co-occurrence network was determined using the *components* module of the *igraph* package. Null model hypothesis was tested by assessing the distribution of the node degree and the respective probabilities of the occurrence network against those simulating the Erdos-Renyi, Barabasi-Albert, Stochastic-block null models (Dormann et al., 2009). The *igraph* package was also used for plotting the networks.

Phylogenomics and pangenomes

For the pangenome analyses, we collected all the bins taxonomically identified as *Polaromonas* spp. and used the pangenome workflow described by Meren *et al*. (http://merenlab.org/2016/11/08/pangenomics-v2) using anvi'o (Eren et al., 2015), along with NCBI (Pruitt et al., 2007) refseq genomes for comparison and an outgroup from the closely related *Rhodoferax* genus. The choice of *Polaromonas* spp. was based on its high abundance and prevalence within the epilithic biofilms. The accession IDs from the reference genomes obtained from NCBI are provided in the supplementary material. The pangenome was run using the *-min-bit 0.5, -mcl-inflation 10* and *-min-occurence 2* parameters, excluding the partial gene calls. A phylogenomic tree was built using MUSCLE v3.8.1551 (Edgar, 2004) and FastTree2 v2.1.10 (Price et al., 2010) on all single-copy gene clusters in the pangenome that were present in at least 30 genomes and had a functional homogeneity index below 0.9, and geometric homogeneity index above 0.9. The phylogenomic tree was used to order the genomes, the frequency of gene clusters (GC) to order the GC dendrogram. A phylogenomic bacterial tree of life containing the 47 high-quality MAGs along with 264 NCBI bacterial genomes was built based on a set of 74 single-copy genes using the GToTree v1.5.51 (Lee, 2019) pipeline with the -D parameter, allowing to retrieve taxonomic information

for the NCBI accessions. Briefly, HMMER3 v3.3.2 (Eddy, 2011) was used to retrieve the single-copy genes after gene-calling with Prodigal v2.6.3 (Hyatt et al., 2010) and aligned using TrimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009). The entire workflow is based on GNU Parallel v20210222 (Tange, 2018).

Data analyses and figures

Figures for the study including visualizations derived from the taxonomic and functional components, were created using version 3.6 of the R statistical software package (R Core Team, 2023). The maps indicating the collection sites were generated using the *ggmap* (Kahle and Wickham, 2013) package in R. *KEGGDecoder* (Graham et al., 2018) was used to assess enriched KEGG orthology (KO) IDs in comparison to 105 publicly available metagenome sampled in various ecosystems at a global scale (Supplementary Data 3 and 6), which were processed using the IMP workflow. *DESeq2* (Love et al., 2014) with FDR-adjustments for multiple testing were used to assess KOs significantly enriched in the GFS metagenomes compared to this comparison dataset. The volcano plot highlighting the significant KOs was generated using the *EnhancedVolcano* (Blighe, 2023) R package. Figures from metabarcoding data were also generated in Rv3.6 using the *ggplot2* (Kahle and Wickham, 2013) package and were further annotated graphically using Inkscape ("Inkscape") while the network plots were generated using the *igraph* v1.2.2 package.

❖ *Data availability*

Raw sequencing data samples and the MAGs are available at NCBI's sequence read archive under BioProject accession **PRJNA733707**. The Biosample accession IDs and the metadata associated with each sample are listed under Supplementary Data 6. A snippet of the results and source data generated and used in this study have been deposited in Zenodo at https://doi.org/10.5281/zenodo.5545722. Data used to generate the figures are also provided as a 'Source Data' file. Source data are provided with this paper.

❖ *Code availability*

The detailed code used for the downstream functional and growth analyses is available at https://git-r3lab.uni.lu/susheel.busi/nomis_pipeline and https://doi.org/10.5281/zenodo.6372573. The custom pipeline for eukaryote analyses can be found here: https://github.com/Mass23/NOMIS_ENSEMBLE/tree/coassembly. Subsequent binning and manual refinement of eukaryotic MAGs was done as described here: https://git-r3lab.uni.lu/susheel.busi/nomis_pipeline/-/blob/master/workflow/notes/MiscEUKMAGs.md.

❖ *Acknowledgements*

❖ *Figures & Tables*

Figure 3. 1 Sedimentary habitats affect microbiome structure and assembly.

a Representative images of sample collection indicating GFS and adjacent epilithic biofilm (left) with images of epilithic biofilms (right). Photo credits: Martina Schön and Matteo Tolosano. Ordination analyses of the epipsammic (*n* = 27 biologically independent samples) and epilithic (*n* = 21 biologically independent samples) biofilm based on prokaryote (**b**) and eukaryote (**c**) metabarcoding profiles from

Southern Alps and Caucasus. Microbial richness across geographic locations and sample types in (**d**) prokaryotes and (**e**) eukaryotes. The statistical analyses was performed on 27 epipsammic and 21 epilithic samples using a two-sided non-parametric *t* test. Bonferroni-corrected *p* values are indicated by *, i.e., *** represents $p < 0.001$. Boxplots represent the median richness with the 25th and 75th quartiles observed within the samples.
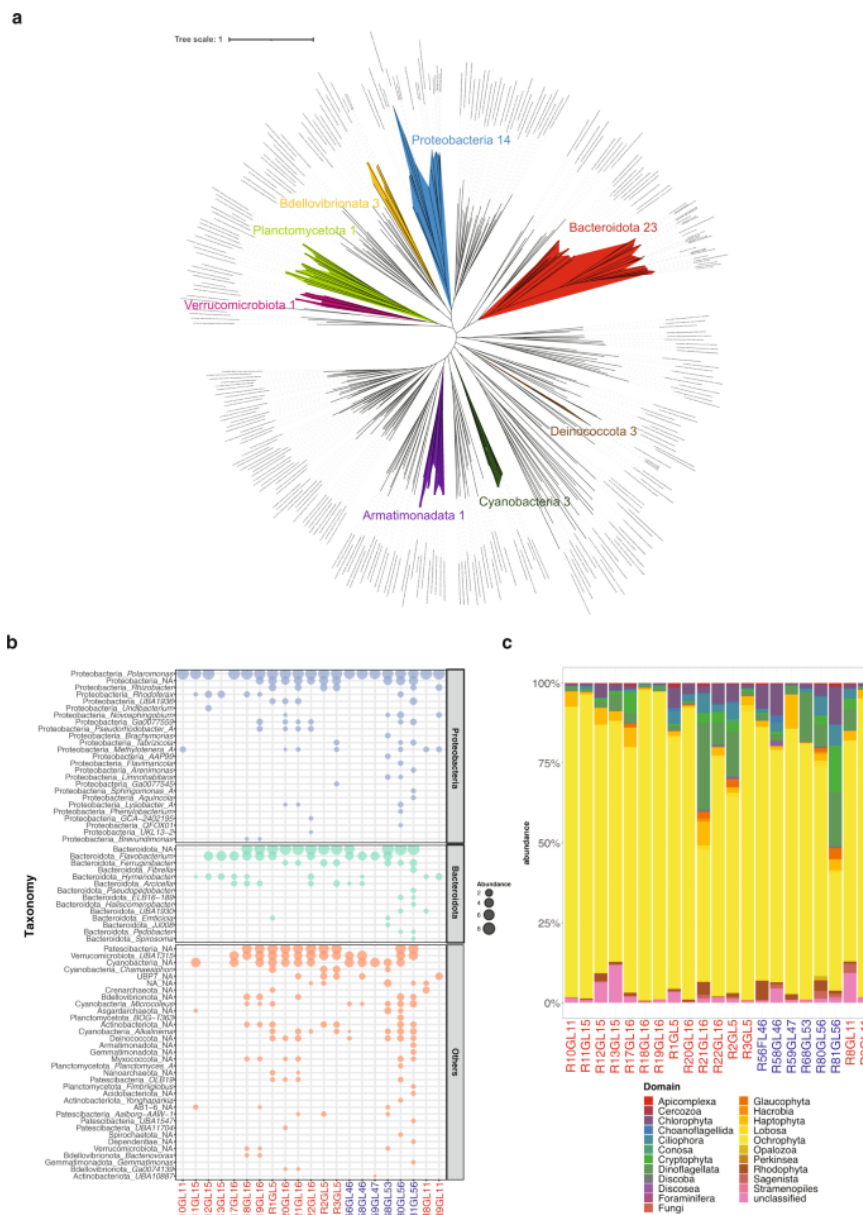
Figure 3. 2 Metagenomics unveils the complexity of epilithic biofilms.

a Bacterial phylogenetic tree constructed using high-quality ($n = 49$, >90% completion and <2% contamination) MAGs reconstructed from the epilithic biofilms. The numbers beside the phylum names indicate the number of high-quality MAGs assigned to the respective phylum. Only high-quality MAGs were used to mitigate phylogenetic disparities from incomplete MAGs. **b** Normalized abundance of

reconstructed prokaryotic genomes, i.e., MAGs, from the epilithic biofilms. Taxonomy at phylum and genus levels is depicted. NA: unclassified genus. Samples from the Southern Alps are indicated in red, while those from Caucasus are shown in blue. Medium-to-high quality MAGs (n = 73) are depicted. **c** Eukaryotic relative abundance profile obtained from metagenomic sequencing across all epilithic biofilms samples.
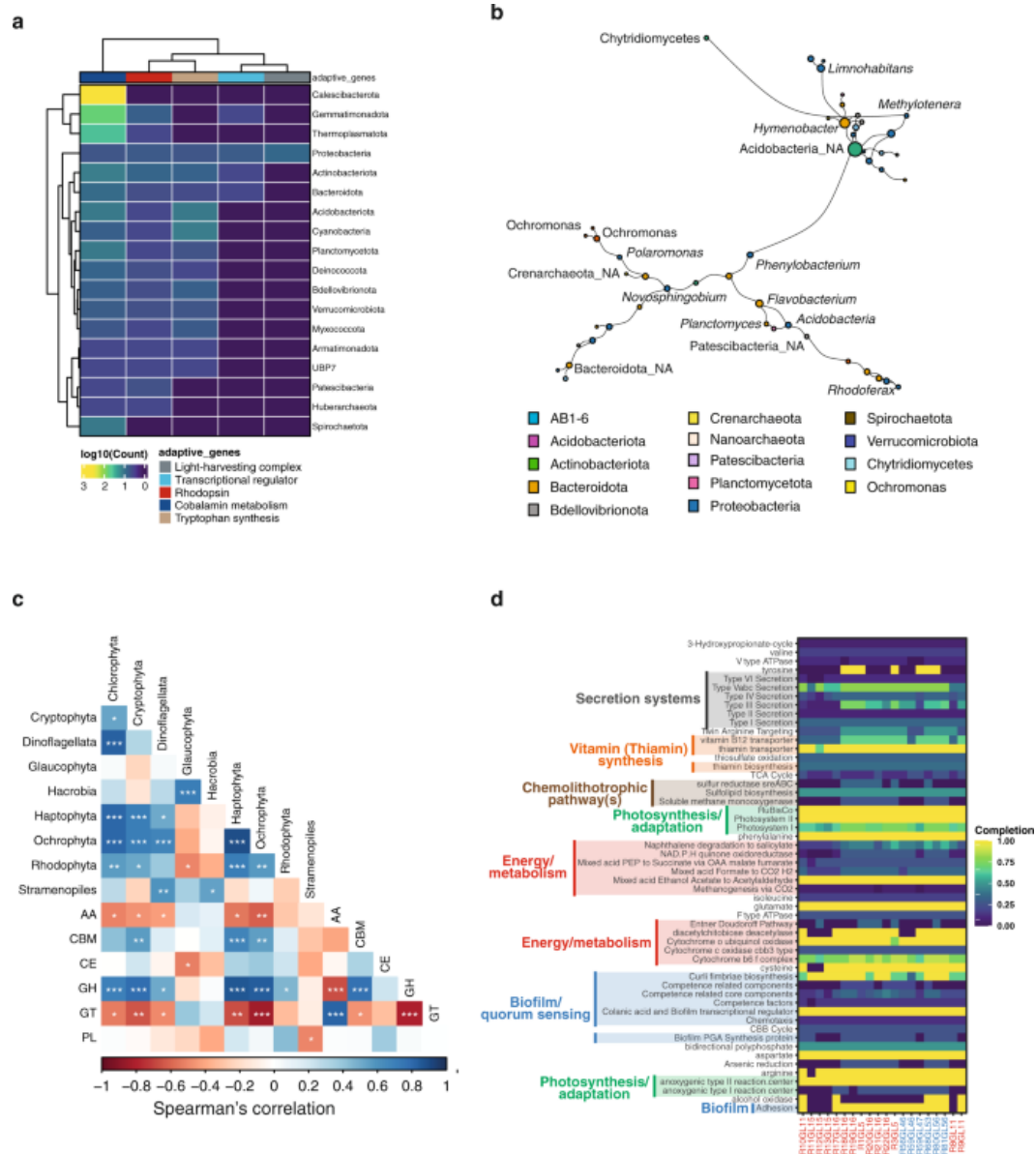
Figure 3. 3 Epilithic biofilms are the basis for a 'green food chain'.

a Abundance of genes involved in energy production (light-harvesting complex, transcriptional regulator for phototrophy, and rhodopsin) and photo-heterotrophic interactions (cobalamin metabolism and tryptophan synthesis), across all prokaryotic phyla are represented in the heatmap. Values indicate the $\log_{10}$ abundance per gene within the phyla. **b** Largest component of the co-occurrence network between pro- and eukaryotic MAGs. Each node corresponds to a MAG (pro- or

eukaryote). Size of the node corresponds to degree centrality and the edges represent the positive coefficients of correlation between each node. Colour of each node represents the phylum annotation. NA: unclassified genus. **c** Spearman's correlation analyses of relative abundances of eukaryotic primary producers with the CAZyme abundances. CAZymes include AA auxilliary activities, CBM carbohydrate-binding module, CE carbohydrate esterases, GH glycoside hydrolases, GT glycosyltransferases, PL polysaccharide lyases. FDR-adjusted p values were estimated using the 'cor.mtest' function from the *corrplot* R package and are indicated by *, i.e., *<0.05, **<0.01, ***<0.001. **d** KEGG orthology (KO) pathways enriched in epilithic biofilms compared to publicly available cryospheric metagenomes were further assessed via KEGGDecoder for pathway completion and are displayed. The completeness of the pathways is indicated in the heatmap, per sample.
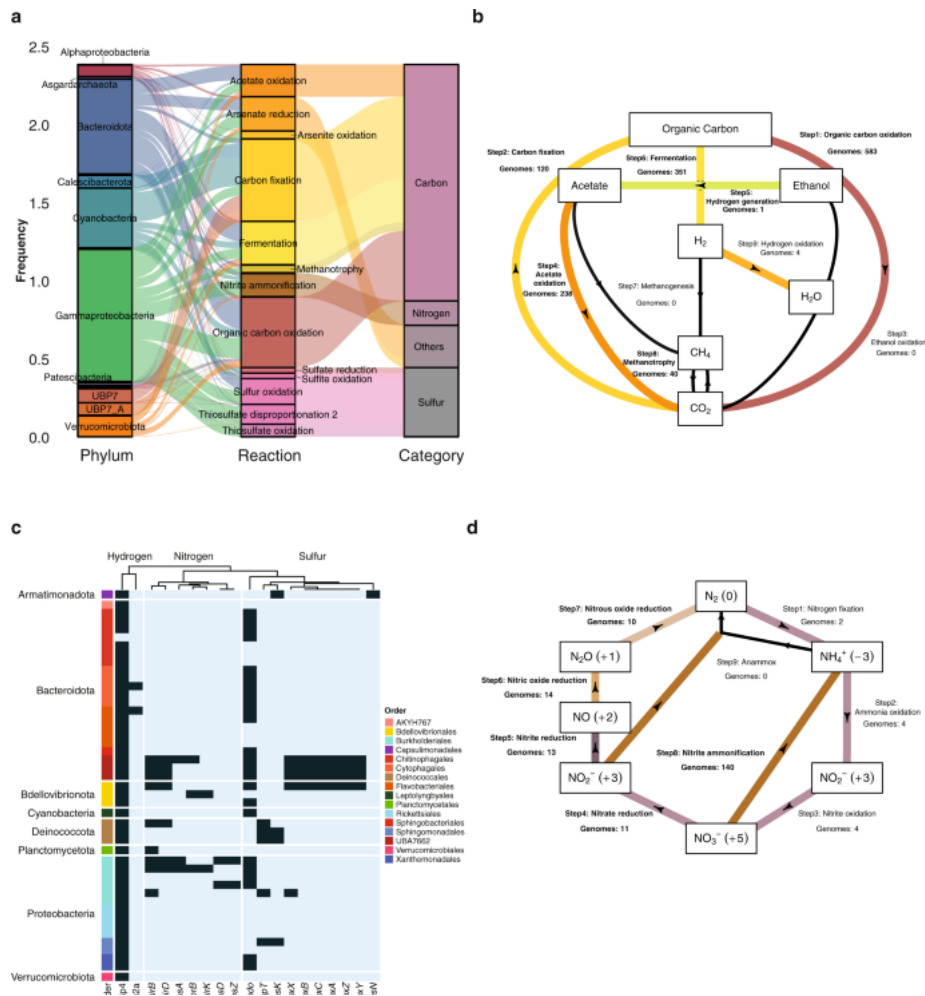
Figure 3. 4 Functional redundancies across MAGs enable diverse energy acquisition and biogeochemical pathways.

a The alluvial plot represents the metabolic pathways identified within all prokaryotic MAGs, with the respective taxonomic classification and category of nutrients. **b** Total number of MAGs encoding genes for and involved in the Carbon cycle (Methods) are depicted in the flow gram created using a modified script from METABOLIC[109]. Each sub-pathway is indicated as a step with the corresponding number of genomes encoding the respective genes. **c** Phylum and order-level distributions of chemolithotrophic (hydrogen, nitrogen and sulfur) pathways with the respective gene copies per pathway are depicted in the heatmap. **d** Flow diagram indicating the MAGs encoding for pathways in the nitrogen cycle (Methods). Each sub-pathway is indicated as a step with the corresponding number of genomes encoding the respective genes.
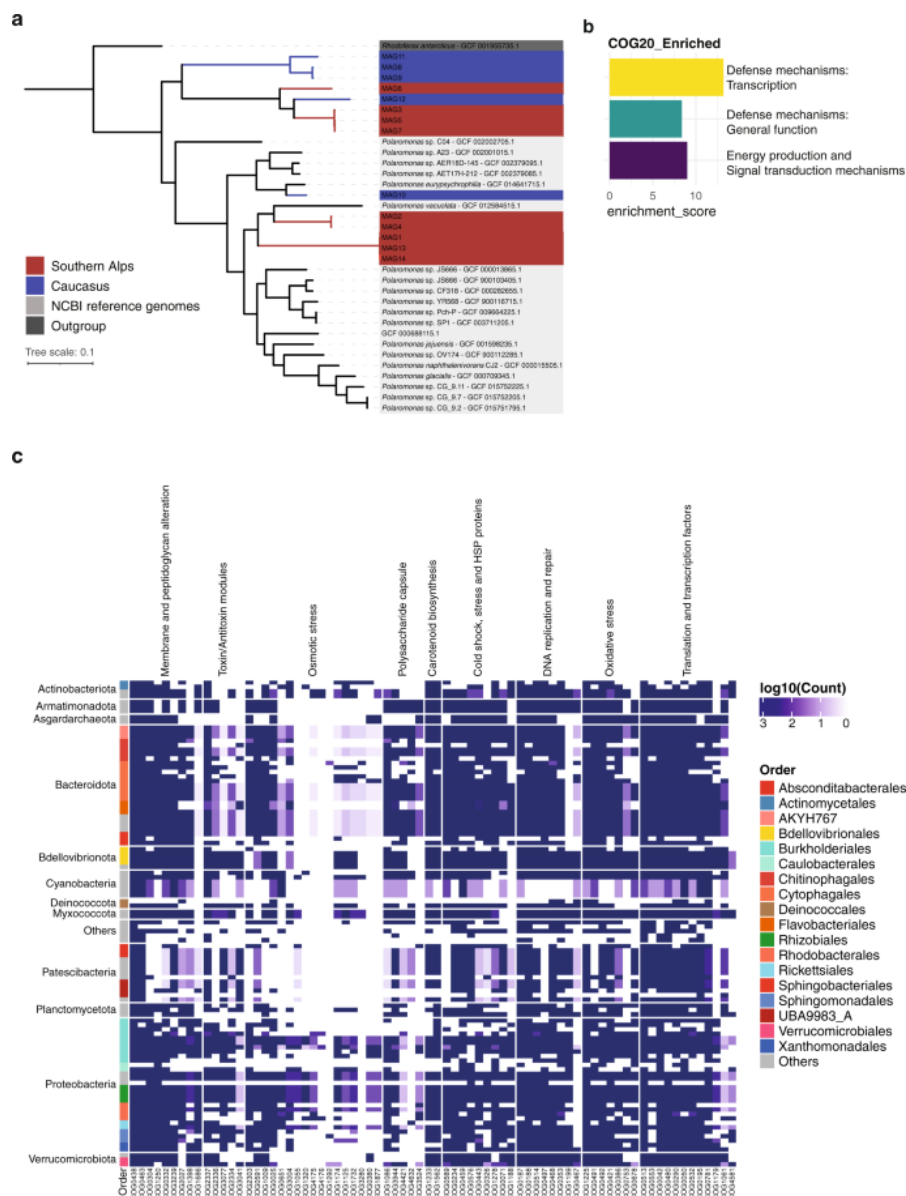
Figure 3. 5 Genomic underpinnings of adaptation to the extreme GFS environment.

a Phylogenomic tree based on *Polaromonas* genomes recovered from Southern Alps (red) and Caucasus (blue) along with publicly available genomes (grey) and an outgroup (*Rhodoferax*, dark grey). **b** Clusters of orthologous (COG20) group pathways enriched in epilithic biofilms MAGs compared to the reference genomes are depicted in the barplot. **c** Heatmap representing the abundance of genes involved in cold adaptation. Taxonomy at phylum and order levels is depicted. Columns indicate clusters of orthologous groups associated with adaptive genes.

Chapter 4. Predicting climate-change impacts on the global glacier-fed stream microbiome

Massimo Bourquin[1], Hannes Peter[1], Grégoire Michoud[1], Susheel Bhanu Busi[3], Tyler J. Kohler[1,4], Andrew L. Robison[1], Leïla Ezzat[1,2], Aileen U. Geers[1], Matthias Huss[5,6,7], Stilianos Fodelianakis[1], The Vanishing Glaciers Field Team, Tom J. Battin[1]

[1]River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne, EPFL, Lausanne, Switzerland

[2]MARBEC, University of Montpellier, CNRS, Ifremer, IRD, Montpellier, France

[3]UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

[4]Department of Ecology, Faculty of Science, Charles University, Prague, Czechia

[5]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland

[6]Department of Geosciences, University of Fribourg, Fribourg, Switzerland

[7]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

❖ *Authors' contribution*

In this chapter, I collected climatic data, curated the climatic, glaciological, environmental data, designed and implemented the modelling framework, performed all statistical analyses, and created all visualisations with advices from co-authors. I also participated in the processing of the metagenomic data used for this study (MAGs dataset of the Vanishing glaciers project).

❖ *Publication*

Preprint version of the manuscript that has been submitted on December 21st 2023.

❖ *Abstract*

Forecasting climate change impacts on complex microbiomes is non-trivial but critical to understand future ecosystem functioning. The worldwide shrinkage of glaciers and vanishing glacier-fed streams are emblematic of climate change, but consequences for the glacier-fed stream microbiome remain poorly understood. Similar environmental constraints across the world's glacier-fed streams make them ideal systems for predictive climate change microbiology. Here, leveraging environmental and metagenomic data sampled by the *Vanishing Glaciers* project from more than 150 glacier-fed streams draining Earth's major mountain ranges, we model responses of the glacier-fed streams environment and microbiome to future climate-induced glacier shrinkage. Using statistical learning approaches and Shared Socioeconomic Pathways projections, we show how key environmental constraints will diminish, thereby promoting the growth of benthic primary producers and bacteria. Furthermore, investigating 2,333 genome-level bacterial strains assembled from metagenomes, we project that the biodiversity of the glacier-fed stream microbiome will change and that entire clades rather than single strains may be at threat under future environmental conditions. We also show that microbiome function will experience shifts related to energy and carbohydrate metabolism, while genome size may increase. Altogether, our results indicate that glacier shrinkage renders the glacier-fed stream environment less extreme, thereby shifting the phylogenetic structure and function of the microbiome. This altered microbiome may play a more important role for carbon cycling in the future glacier-fed streams.

❖ *Main text*

High mountains play crucial roles in global systems, harbor a unique biodiversity and provide water resources for large human populations (Clason et al., 2023; Immerzeel et al., 2020). At the same time, they are particularly vulnerable to climate change (Barnett et al., 2005; Pörtner et al., 2019). Changing precipitation and temperature regimes, diminishing snow pack and shrinking glaciers altogether alter the ecology of high-mountain ecosystems (Gottfried et al., 2012; Rumpf et al., 2022). The rapid shrinkage of glaciers and the vanishing of their streams (glacier-fed streams; GFS) worldwide count among the most emblematic consequences of climate change. Significant

advances have recently been made regarding the impacts of glacier shrinkage on invertebrate ecology in GFSs. For instance, species adapted to the harsh environmental conditions in GFSs become increasingly imperiled as glaciers shrink, and their distribution ranges reduced to cold-water refugia under the influence of residual ice (Cauvy-Fraunié and Dangles, 2019; Giersch et al., 2017; Jacobsen et al., 2012; Wilkes et al., 2023). However, how glacier shrinkage may affect microbial life in GFSs under various climate change scenarios until the end of this century remains unknown to date. Biofilms dominate microbial life in streams, particularly in GFSs, where they regulate nutrient cycling and ecosystem metabolism, and form the basis of the food web (Battin et al., 2016). They host a microbiome that is distinct from the assemblage in the streamwater and locked in the glacier ice, and that spans all domains of life (Ezzat et al., 2022; Wilhelm et al., 2013). The genomic potential of the GFS microbiome allows microorganisms to withstand the harsh environmental conditions (e.g., near-freezing water temperatures, high sediment loads and ultra-oligotrophy) and seize opportunities of more favorable environmental conditions determined by glacier runoff dynamics on an annually recurring basis (Busi et al., 2022; Michoud et al., 2023).

As the distribution of microorganisms integrates past and present environmental conditions, and as microbial metabolism underlies elemental cycling, microbiomes may be used to forecast the future of ecosystem functioning under climate change (Correa-Garcia et al., 2023). However, forecasting how entire microbiomes may change under such scenarios is challenging. Efforts largely remain restricted to soil and marine microbiomes and use amplicon sequencing rather than genomic information (Frémont et al., 2022; Mod et al., 2021). The GFS microbiome is well suited for predictive modeling under climate change scenarios because it is under strong selection by a harsh environment, which deterministically shapes its assembly (Fodelianakis et al., 2022). Moreover, key environmental constraints are similar across the world's GFSs, which are now rapidly changing. Interfacing the cryosphere, GFSs are thus ideal systems for modeling future climate impacts. While modeling glacier shrinkage under future climate scenarios has become increasingly sophisticated (Rounce et al., 2023), no quantitative forecasts currently exist for downstream ecosystems and their microbial life.

In this article, we developed a hierarchically structured modeling framework to forecast changes of the environment and both microbiome structure and function of the world's GFSs until the end of the century in the context of three greenhouse gas emissions scenarios (Shared Socioeconomic Pathways; SSP) and using Global Glacier Evolution Model (GloGEM) (Huss and Hock, 2018) projections and climatologies for

Earth's land surface areas (CHELSA)[28] at high resolution (Figure 4.1A). We report prediction averages for the 2070-2100 period under SSP3, representing an intermediate scenario of climate change; predictions for the more extreme SSP1 and SSP5 are reported in supplementary information only. For our forecasts, we leverage environmental data and 2,333 bacterial genomes spanning the entire bacterial tree of life that were assembled from metagenomes (strain-level resolution, 99% average nucleotide identity) from GFSs sampled by the *Vanishing Glaciers* project in the Caucasus Mountains, Chilean and Ecuadorian Andes, Himalayas, Pamir and Tien Shan, Rwenzori Mountains, Scandinavian Mountains, New Zealand Southern Alps, and Southwest Greenland (Figure 4.1B). In each GFS, we captured the recent deglaciation history along a chronosequence (Figure 4.1C), which we used in a machine learning approach to predict how the GFS environmental template (i.e., the physicochemical niches affecting ecological communities) will change by the end of the century and how these shifts affect microbial biomass and diversity. We leveraged these predictions of the GFS's environment to model the abundance of all bacterial strains, and assess shifts in the composition, phylogenetic structure and functional potential of the GFS microbiome under future climate scenarios. Finally, we inferred the putative consequence of an altered GFS environment and microbiome for ecosystem functioning under future projected climate change.

Climate-induced changes of the environmental template of glacier-fed streams

The environmental template and its niches determine the composition of ecological communities in stream ecosystems (Southwood, 1977). In GFSs, the environmental template, defined by flow regime, streamwater turbidity, temperature, pH and nutrients, for instance, is under the influence of the glaciers (Milner et al., 2017; Milner and Petts, 1994). The overall magnitude and temporal dynamics (e.g., seasonal and diel fluctuations) of streamflow, as well as sediment production through subglacial weathering and erosion are influenced by glacier area (Huss and Hock, 2018; Zhang et al., 2022). Streamflow and sediment load drive turbidity and hence light availability in GFSs, thereby affecting primary production (Boix Canadell et al., 2021). Glacier coverage, the fraction (%) of the catchment that is glacierized, serves as proxy for

streamflow contributions from glacial meltwaters and affects streamwater physicochemistry (e.g., nutrients, pH) (Milner et al., 2017). Furthermore, as glacier influence diminishes with downstream distance, streamwater temperature increases and channels consolidate (Jacobsen and Dangles, 2012). Given the entanglement of these processes with glacier dynamics and their fundamental role for life in GFSs, it is of utmost relevance to understand and predict climate-change impacts on the GFS environmental template.

To forecast potential changes of the global GFS environmental template under future climate change scenarios, we first obtained climatological and glaciological parameters and their projections for the mountain ranges from where we sampled GFSs (Figure 4.1B). We sourced climatic parameters (e.g., mean daily temperature, annual snow-covered days, monthly precipitation) from the CHELSA database[28], which contains climatologies downscaled to a 1 km resolution for historical data, as well as future projections under future climate change scenarios. To assess how glacier influence may change in the future, we retrieved glaciological parameters from GloGEM, which we complemented with data from satellite imagery and field survey (*Methods*). Both allowed us to collect data for the 2019-2022 period when GFSs were sampled, (Supplementary table 4.1) and the 2070-2100 period. Values of climatic parameters (e.g., temperature, precipitation and annual snow cover) that were extrapolated for the 2019-2022 period did not exhibit extreme values compared to the distribution of the 1981-2010 values (Supplementary figure 4.1).

CHELSA projections predict that median air temperature will increase by 4.4°C (IQR: 3.6-5.3°C) (Wilcoxon test, p < 0.001), whereas annual snow-covered days will decrease with a median reduction by 66.4 days (IQR: 55 - 99.4 days) (Wilcoxon test, p < 0.001). Monthly precipitation is projected to shift regionally and therefore does not show a clear trend (Supplementary figure 4.1). Consequently, future projections predict size reductions of the glaciers that feed our GFSs. Glacier area is expected to shrink by 78.3% (IQR: 64.4% - 96.0% km^2), reducing coverage by 41.3% (IQR: 29 - 60 %) and reflected in a median glacier snout recession of 1.84 km (IQR: 1.33 - 2.78 km) on average by the 2070-2100 time period (Figure 4.1C). These datasets, encompassing present conditions and future projections according to climate change scenarios, were then used to forecast changes in the GFS environmental template.

To this end, we trained generalized additive models (GAMs) using current climatic, as well as glaciological data and sediment mineralogy (*Methods*) to predict how GFS physicochemical parameters may change under different climate change scenarios.

Our projections reveal that median streamwater temperature in the upstream reaches will increase by 3.0 °C (IQR: 2.2 - 4.0°C) by the end of the century (Figure 4.2A, Supplementary figure 4.2). Future increases of streamwater temperature will be comparatively higher in GFSs with currently low temperatures, such as in the Pamir, Tien Shan and the Himalayas (Figure 4.2B). Median streamwater turbidity will decrease by 44.4% (IQR: 31.6 - 71.7 NTU), while median electrical conductivity will increase by 88.2% (IQR: 51.6% - 130.7% μS cm^{-1}) (Figure 4.2C). Projections also indicate that median streamwater pH, soluble reactive phosphorus and dissolved inorganic nitrogen will significantly decrease by 2.8% (IQR: 1.6 - 4.2 %), 14.1% (9.6 - 27.3 %) and 11.5% (IQR: 5.2 - 16.3 %), respectively, by the end of the century (Figure 4.2, Supplementary table 4.2). The analysis of response curves reveals that streamwater temperature will increase with glacier length recession, while streamwater turbidity and pH will decrease with diminishing glacier area (Supplementary figure 4.3). Shifts in streamwater nutrient concentrations appear more related to geography and streambed sediment mineralogy, hence yielding smaller relative changes associated with climate change and the associated glacier shrinkage overall (Supplementary figure 4.3). Nonetheless, our results suggest that, globally, streamwater soluble reactive phosphorus and dissolved inorganic nitrogen are associated with glacier area. These quantitative forecasts of the GFS environment underpin conceptual model considerations on the impacts of glacier shrinkage on nutrient availability and biogeochemistry in GFSs (Milner et al., 2017; Ren et al., 2019).

Greening of glacier-fed streams and shifting biodiversity

Microbial benthic biomass is a key stream ecosystem attribute as it fuels the food web and regulates ecosystem metabolism and nutrient cycling (Battin et al., 2016). Our projections (*Methods*, Supplementary table 4.2) reveal significant increases for median values of benthic chlorophyll-*a* (339.7%; IQR: 183 - 852.2 %), a proxy for algal biomass, and bacterial abundance (88.5%; IQR: 60.4 - 150.2 %) (Figure 4.3A, B). Interestingly, the projected benthic biomass in the future GFSs is still low compared to other stream ecosystems (Brandani et al., 2022), which points to the persistence of other environmental constraints (e.g., high turbulence) that are not directly linked to climate change. Nonetheless, we suggest that GFS primary production will increase beyond the

sediment peak, when streamwater turbidity decreases because of the fading capacity of glaciers to generate fine sediments (Zhang et al., 2022), hence decreasing light attenuation and physical abrasion of benthic algae. This notion is indeed supported by the negative correlation between projected values of streamwater turbidity and benthic chlorophyll-*a* (rho = -0.96, p < 0.001), which also underlies the relationship between glacier area and GFS benthic chlorophyll-*a* (Supplementary figure 4.3). Changes in resource availability, associated with elevated biomass of primary producers in future GFSs, may have impacts on the diversity, structure and function of the GFS microbiome.

Despite the harsh environment of GFSs, their benthic biofilms host a diverse microbiome (Busi et al., 2022; Wilhelm et al., 2013). Our projection shows that this diversity (alpha-diversity expressed as Shannon *H*; see *Methods*) will increase by 6.2% (IQR: 4.7 - 8.9 %) under SSP3 (Figure 4.3C). Positive correlations of projected bacterial Shannon diversity (rho = 0.69, p < 0.01) and abundance (rho = 0.98, p < 0.001) with chlorophyll-*a* corroborate the notion that increased primary production will sustain higher microbial biodiversity. Our projections of increasing local microbial diversity in future GFSs is in line with reports showing that invertebrate alpha-diversity increases in GFSs as glacier influence diminishes (Brown et al., 2018; Cauvy-Fraunié and Dangles, 2019, 2019).

Strong environmental selection imprints a distinct phylogenetic signature on the GFS microbiome, characterized by microdiverse bacteria that are phylogenetically clustered and putatively well adapted to the GFS environment (Fodelianakis et al., 2022). To test how phylogenetic structuring may change under future climate change, we project mean nearest taxon distance and mean phylogenetic distance, indicating relatedness at shallow and deeper phylogenetic branching, respectively. We find that median values of the mean nearest taxon distance and mean phylogenetic distance will significantly (p < 0.001) increase by 3.5% (IQR: 2.2 - 5.8 %) and 3.2% (IQR: 1.7 - 6.3 %), respectively (Figure 4.3D), suggesting that phylogenetic clustering will diminish in the future GFS microbiome. In combination with the predicted increase in diversity, this implies that phylogenetically distant taxa compared to those currently present in GFSs may augment microbiome diversity. The correlation (rho = 0.97, p < 0.001) between future mean nearest taxon distance and chlorophyll-*a* further suggests that increased resource availability through algal exudates may alleviate current selective constraints, ultimately leading to the creation of new niches. In fact, autochthonous organic compounds derived from algae are readily available to the metabolism of microbial heterotrophs in stream biofilms (Wagner et al., 2017).

Climate change shifts strain-level distribution

Species distribution models are commonly used to forecast climate change impacts on the structure and diversity of ecological communities. To predict the impacts of a changing climate and integrating glaciological and environmental controls on key components of the GFS microbiome, we built separate models for each of the 2,333 strains from our GFSs (except the Alaska Range) in a strain distribution model framework (*Methods*). For this, we used a combination of climatic, glaciological, and mineralogical data, along with the predictions from the previous section. Overall, our models predicted strain abundance with satisfactory accuracy (median $r^2_{prediction}$ = 0.25; IQR = 0.13-0.36) (Supplementary figure 4.4).

First, a feature selection procedure was performed to select the best set of covariates (*Methods*), as well as to assess the predictive power of these covariates to model strain-specific abundance distributions (Supplementary figure 4.5). Across all strains, we found latitude, streamwater electrical conductivity, pH, and temperature were the most powerful predictors of relative abundance. Bioclimatic variables reduced by principal component analysis (*Methods*; Supplementary figure 4.6), as well as distance to the glacier and annual snow cover, were also identified as important predictors (Figure 4.5A). Next, using forecasts of these predictors, we assessed future abundance distributions for each strain. The majority of the 2,333 strains (i.e., 64.7%) is expected to increase in abundance, while 30.0% will decrease in abundance and 5.3% are predicted to not significantly change in abundance (Supplementary table 4.3). This overall gain in abundance aligns with our independent forecast of increasing bacterial biomass.

The classes *Gammaproteobacteria* and *Alphaproteobacteria*, which dominate the present-day GFS microbiome, are projected to experience the most pronounced absolute increases, a result that can be attributed to their high abundance and the high number of strains affiliated to these classes (Figure 4.4C). Moreover, classes predicted to increase the most in abundance compared to their present-day abundance (e.g., *Paceibacteria*, *Gemmatimonadetes*) are known for their parasitic or symbiotic lifestyles (Mujakić et al., 2022; Nelson and Stegen, 2015), potentially promoted by higher biofilm biomass in future GFSs. We also identified groups of strains (e.g., *Bdellovibriona*,

Polyangia, *Actinomycetia*) whose representatives will decrease in abundance the most (Figure 4.4B, C). Strikingly, strains that currently occur at low abundance (i.e., lower half of the abundance distribution) are projected to increase more in abundance compared to strains that are currently abundant in GFS microbiomes (i.e., upper half of the abundance distribution). This points to an internal reorganization of GFS microbial communities, and implies that the future GFS microbiome will be characterized by higher evenness, consistent with our projection of Shannon *H* and *Pielou evenness* (Supplementary figure 4.7). Annual snow cover, distance to the glacier, bioclimatic variables, and streamwater temperature best predict those strains that will decrease in abundance (Figure 4.5B). Our results thus suggest that ecological niches, putatively linked to streamwater temperature and resource stoichiometry in GFSs (Elser et al., 2020), and to which bacterial strains have adapted over evolutionary times, will undergo major transformations owing to climate-induced glacier shrinkage.

We found a strong correlation between the predicted changes in abundance of the strains and the phylogeny (phylogenetic signal measured as the correlation with the $\log_2$ fold-change: lambda = 0.88, p < 0.001), as well as predictability (lambda = 0.77, p < 0.001) (Figure 4.4A, Supplementary table 4.4). The latter reflects how well strain abundance was captured by the models, and was obtained via cross-validation (i.e., building a model on 9 folds, and computing r^2 on the left-out one). Under the assumption of response trait conservatism (i.e., closely related strains share traits that allow them to respond similarly to environmental shifts), these phylogenetic patterns validate our modeling approach. Indeed, we observed more similar environmental predictors being selected for phylogenetically closely related strains (Spearman correlation, rho = -0.15, p < 0.0001, Supplementary figures 4.8, 4.9). Taken together, these results reflect how well strain abundance was captured by our modeling approach (*Methods)*, as we consider that phylogenetic patterns are directly linked to response trait conservatism (i.e., that closely related strains share traits that allow them to respond similarly to environmental shifts).

However, the finding of phylogenetically clustered increases and decreases in abundances also raises concern, as it suggests that climate change could imperil entire clades in GFSs rather than individual strains. In fact, we identify 186 strains – representing 8.0 % of all strains and 26.6 % of those projected to decrease in abundance – as belonging to monophyletic clades of at least three representatives, all projected to significantly decrease under SSP3 (Supplementary table 4.5). Notably, among the largest clades, some belong to taxa (e.g., *Ferruginibacter*, *Lacisediminihabitans*, *Acetobacteraceae*) that are hallmarks of cryospheric

ecosystems and known for their cold adaptation genes (Bourquin et al., 2022; Carey et al., 2016; Liang et al., 2021), for instance. These clades exhibit deep branching, with a median relative phylogenetic depth of 0.25 (IQR: 0.20 - 0.35). Taken together, these findings suggest that the microbiome of the world's GFSs will experience profound phylogenetic restructuring under future climate change scenarios with potential implications for ecosystem functioning.

Changing microbiome functions

Owing to the possible loss of entire phylogenetic clades, it is important to also assess how the functional potential of the GFS microbiome may change under future climate change scenarios. To do so, we constructed random forest classifiers to identify KEGG (Kyoto Encyclopaedia of Genes and Genomes) ortholog groups (KOs) that are important at segregating strains that decrease in abundance from others (*Methods*). Taking phylogenetic structure into account by clustering strains into ten phylogenetic clusters, then building a model on nine and testing the importance of KOs in the left-one out, we identified 408 KOs that significantly characterize strains predicted to decrease in abundance ($p < 0.05$, feature importance quantile > 0.95). Notably, these KOs were found in taxa across the entire phylogeny, indicating that the response of the GFS microbiome to climate change is associated with potentially adaptive functions found throughout the bacterial tree of life. While some of these functions are associated with biofilm formation, many are involved with cold adaptation (Supplementary table 4.6). Further, we found evidence for increased metabolic diversity and reduced genome size in strains that are predicted to decrease in abundance, which may be related to the optimisation of genome content. Specifically, compared to strains that are predicted to increase in abundance or remain invariant, genomes of strains predicted to diminish in the future are smaller (GAM, fixed effect estimate = -0.2 mbp, $p < 0.01$) yet contain an elevated number of KOs (GAM, mean difference = 4.02, $p < 0.001$). This is explained by reduced KO redundancy (GAM, mean difference = 3.44, $p < 0.001$), and suggests that despite a broad metabolic repertoire, smaller genomes are favored under current GFS conditions but that these lifestyle adaptations may not be favored under future conditions. Using enrichment analysis, we found that KEGG categories related to carbohydrate metabolism (Fisher test, OR = 2.20, $p < 0.001$) and energy metabolism

(Fisher test, OR = 1.80, p < 0.001) exhibited a large number of significant KOs in strains predicted to decrease in abundance (Supplementary table 4.7). Interestingly, these categories were overrepresented in the genome of the putatively losing strains (Figure 4.6A). Taken together, reduced genomes with maximized metabolic diversity suggest adaptations to the ultra-oligotrophic, yet temporally highly dynamic, GFS environment. This may be intrinsically linked to the carbon sources and availability, currently varying in response to glacier shrinkage (Hood et al., 2020; Robison et al., 2023). In the future, this may shift towards the more predictable availability of algal-derived organic carbon. Our results thus provide further evidence for the alleviation of selective constraints in future GFS microbiomes, trailed by a reduction in the diversity of the metabolic repertoire at the microbiome level. In combination with increased genome redundancy, this finding further suggests that metabolic specialization becomes promoted by the use of algal exudates as primary energy source in the future GFSs.

In conclusion, our global-scale forecasting provides hitherto missed insights into the future of a rapidly vanishing ecosystem and its microbiome, thereby filling an important knowledge gap in the climate change microbiology of cryospheric ecosystems. We acknowledge potential caveats related to species distribution models, including potentially missing variables and overfitting, for instance. However, we argue that *a priori* knowledge of the GFS environment and microbiome, as well as cross-validation and ensemble modeling make our projection architecture robust. Our findings show how the GFS microbiome, characterized by phylogenetic clustering and lifestyle adaptations (e.g., optimisation of genome size, metabolic flexibility), will restructure in the future as the GFS environment becomes less extreme. We predict that as climate change erodes glacier influence, the world's GFSs will become 'greener' and their microbiome more diverse. "Greening" has the potential to alter carbon and energy supplies in GFSs, which may favor taxa with a metabolic repertoire tailored to life in phototrophic biofilms. Our forecasts suggest major restructuring of the GFS microbiome with entire bacterial clades likely decreasing in abundance and implications for its functional potential. Future research should focus on deciphering the genetic potential of the bacterial clades that are most likely to be lost from the world's GFS microbiome.

❖ *Methods*

Study sites and sample collection

GFSs were sampled by the *Vanishing Glaciers* project between January 2019 and July 2022 (see Supplementary table 4.1). Our global sampling included the European Alps, Scandinavian Mountains, Himalayas, Pamirs and Tian Shan, Ecuadorian and Chilean Andes, Southwest Greenland, Alaska Range, Caucasus, Rwenzori Mountains, and the New Zealand Southern Alps. For the sake of comparability, GFSs were predominantly sampled in spring or autumn during 'windows of opportunities'[17] as this facilitates comparability between GFSs. For each GFS, two reaches were sampled: one as close as possible to the glacier (median = 76 m, IQR = 29-301) and one downstream (median = 706 m, IQR =336-1280). This approach allowed us to capture changes in glacier influence over the two stream reaches, creating a time-for-space substitution design with the idea that sites currently located further downstream from the glacier correspond to future conditions at the current position of the glacier snout. From each reach, we sampled three separate sediment patches with flame-sterilized devices. At each patch, we collected sandy (250 μm to 3.15 mm; Retsch©) sediments from the benthic zone. Sediment samples were transferred into sterile cryovials, immediately flash-frozen in liquid nitrogen and subsequently stored at –80°C prior to and following shipping to Switzerland for DNA extraction and biomass analyses.

Streamwater and sediment physicochemical parameters

In the field, we measured streamwater temperature, pH, specific conductivity and turbidity (Turb® 430 IR, WTW) expressed as nephelometric turbidity units (NTU). We filtered streamwater (pre-combusted GF/F filters, Whatman©, UK) into Nalgene HDPE bottles and froze the samples within 48 h pending nutrient analyses. Nutrient analyses were conducted using a LaChat QuikChem 8500 flow injection analyser for ammonium (N-NH_4^+; QuikChem method 10-107-06-3-D), nitrate (N-NO_3^-, QuikChem method 10-107-05-1-C), and soluble reactive phosphorus (P-SRP; Method 10-115-01-1-M). We combined inorganic forms of nitrogen into dissolved inorganic nitrogen (DIN). Sterile-filtered (0.2 μm, Millipore) samples for major cations and anions were stored in the dark (4°C) and analyzed using a Metrohm 930 Compact IC flex system. Benthic sediment

mineralogy was determined using a X-TRA ThermoARL Diffractometer. Errors varied between 5% and 10% for the phyllosilicates and 5% for grain minerals. Raw data files were generated and transformed into calculated files by WinXRD 2.0-6 (ThermoFischer). Relative abundances of the main mineral groups (clays, quartz, feldspar and calcite) were computed from the raw counts of mica, chlorite, amphibole, feldspars, calcite and quartz divided by the sum of counts.

Climatic dataset

Climatology data at high spatial resolution were collected from the CHELSA database (version 2.1, https://www.nature.com/articles/sdata2017122) ("Envidat" ; Karger et al., 2017). CHELSA provides both climatic and bioclimatic data typically used in species distribution modeling approaches (and has been shown to improve their accuracy), at high spatial resolution (~1 km) based on a global downscaling approach. The database based on an ERA-Interim climatic reanalysis additionally contains future projections of climate changes for Shared Socioeconomic Pathways (SSP, https://link.springer.com/article/10.1007/s10584-013-0905-2). The data was extracted in the python programming language using the database API and processed with the *rasterio* and *gdal* python libraries (v1.3.8; v3.7.0) (GDAL Development Team; Gillies and others, 2013). GPS coordinates of all sampling locations were used to identify the corresponding grid cells of the (bio-)climatic data set. The specific sampling months were used for monthly parameters. For future projections, five different insitutions models were downloaded and combined by averaging (gfdl-esm4, ukesm1-0-ll, mpi-esm1-2-hr, ipsl-cm6a-lr, mri-esm2-0). As the historical values in the database were available only up to 2010, linear extrapolation based on historical values (average for 1981-2010) and projections for the years 2020-2040 was performed to create the 2019-2022 dataset, and the value for the sampling year and month was extracted. In addition to these values, climatic data for the time period 2070-2100 were collected to create projections onto future scenarios of climate change. For this, three SSPs scenarios corresponding to different greenhouse gas emission trajectories were considered. These scenarios included SSP1, SSP3, and SSP5, encompassing a range of potential future climate outcomes. We report median change and interquartile ranges for the site-specific changes for all parameters. To test for a significant shift, we conducted Wilcoxon tests (wilcox.test function) of the future projections minus the present predictions. All statistical analyses were performed using the R programming language and figures were prepared using the *ggplot2* R package (v3.4.2) (Wickham, 2016). Data processing was performed with the help of the *tidyverse* R packages suite (Wickham et al., 2019).

Glaciological dataset

The future evolution of all glaciers was assessed based on the Global Glacier Evolution Model (GloGEM) (Huss and Hock, 2018). The model is initialized with present-day glacier extent and computes changes in snow accumulation and melt, as well as changes in glacier length based on an ensemble of Global Circulation Models using different greenhouse-gas emission scenarios. The model has been calibrated to match observed mass change at the scale of every individual glacier globally (Hugonnet et al., 2021). For this study, we extracted data on distance of sampling locations to the glacier terminus over the time period 2000 to 2100, as well as the area of the glacierized surface. We report median change and interquartile ranges for the site-specific changes for all parameters. To test for a significant shift, we conducted Wilcoxon tests (wilcox.test function) of the future projections minus the present predictions. All statistical analyses were performed using the R programming language (v4.3.1, https://intro2r.com/citing-r.html) and figures were prepared using the *ggplot2* R package (v3.4.2) (Wickham, 2016). Data processing was performed with the help of the *tidyverse* R packages suite (v2.0.0) (Wickham et al., 2019).

Metagenomics

Metagenomes were sequenced for 155 sediment samples covering all mountain ranges except the Alaska Range. DNA extraction, purification, library preparation, sequencing and metagenome assembly steps were performed as described in Busi et al. (2022) (Busi et al., 2022). Briefly, 5 g of GFS sediments were treated using a phenol:chloroform-based extraction method subsequently followed by an ethanol precipitation step (Busi et al., 2020). This protocol yielded on average 50 ng of DNA per sample which was used for library preparation using the NEBNext Ultra II FS library kit, which also included 6 PCR cycles (Busi et al., 2022). The metagenomic sequence data was processed using the *Integrated Meta-omic Pipeline* workflow (version 3.0; commit# 9672c874) (Narayanasamy et al., 2016). Briefly, adapter trimming from reads is followed by an iterative assembly using *MEGAHIT* (Li et al., 2015) and *Flye* (Kolmogorov et al., 2020). Reads assembled into contigs were subjected to binning. For each individual assembly, we mapped the reads of the 5 closest samples (euclidean distances of gps coordinates) using *BWA-mem* (v0.7.17) (Li, 2013). To reduce

computation time, we removed sequences in the assembly of < 1.5 kbp. Subsequently, 10 % of the preprocessed reads were randomly selected before mapping with *seqtk* (v1.3) (Shen et al., 2016). We then used *metabat2* (v2.15)[70], *concoct* (v1.1.0) (Alneberg et al., 2014) and *metabinner* (v1.4.3) (Wang et al., 2023) using default parameters to obtain bins. *DAS Tool* (v1.1.4) (Sieber et al., 2018) was then employed to generate a non-redundant set of bins using a score threshold of 0.3. The quality of the redundant bins was assessed with *CheckM2* (v1.0.1) (Chklovski et al., 2023).

Bins from all samples (including the ones generated by IMP3) with a completeness of more than 50% were then selected for further analyses which accounted for 12,599 bins. We then used *MDMCleaner* (v0.8.3) (Vollmers et al., 2022) to reduce contamination of the bins. Finally, after rerunning *CheckM2* on the bins to get final estimates of completeness and contamination, we used *dRep* (v3.2.2) (Olm et al., 2017) to dereplicate bins using a minimum completeness of 70 % and maximum contamination of 10 % and an ANI of 99 % that arise to 2868 MAGs (strain-level dereplication). Functional annotation of the MAGs was performed with *EggNog-Mapper* (v2.1.9) (Huerta-Cepas et al., 2017) after obtaining coding regions (CDS) with *prodigal* (v2.6.3) (Hyatt et al., 2010). The coverage of the MAGs was estimated by mapping the reads of the samples to the genomic contigs using *CoverM* (v0.6.1, available at https://github.com/wwood/CoverM) using the trimmed_mean parameter. We normalized the coverage by similarly mapping the reads on the recA gene (K03553). After filtering out low-abundance strains (prevalence lower than 20% at a 10x recA coverage abundance threshold), 2,333 were selected for the strain distribution modeling.

Modeling environmental parameters and microbial biomass

Models of streamwater temperature, turbidity, conductivity, pH, soluble reactive phosphorus (SRP), and dissolved inorganic nitrogen (DIN) were built using climatic, bioclimatic, glaciological and geological parameters as covariates, chosen by feature selection (Supplementary figure 4.3). Additionally, biomass and diversity metrics for the bacterial communities were modelled analogously. Response variables and covariates were log-transformed where necessary to improve the normality of residuals, adding a constant equal to the half of the smallest non-zero value. Before model fitting, all variables were scaled by subtracting the mean and dividing by the standard deviation. All models were created using the gaussian() family function. Generalized additive models (GAMs) were used to model stream parameters and biomass in order to account for spatial autocorrelation at the regional scale using a spatial spline (formula:

s(latitude, longitude, bs='sos', m=1, k=-1)), and the models were validated using 10-fold cross validation. GAMs have been successfully used to model and predict future changes in environmental sciences and ecology especially for spatially structured data (Colón-González et al., 2013; Grüss et al., 2016; Mod et al., 2021; Ravindra et al., 2019), and have also been applied to stream ecosystems, including GFSs (Coleman et al., 2021; Jowett et al., 2008; Paillex et al., 2020; Selle et al., 2019). GAMs were created using the *bam* function of the *mgcv* package (v1.9_0) (Wood, 2023). The choice of parameters used, and of performing ensemble modeling, was made to control potential overfitting. Moreover, we selected only three variables for each strain distribution model to further avoid overfitting. The shape of the smoothed splines were then inspected and all had reasonable smoothing (Supplementary figure 4.3).

To further improve the robustness of our approach, we performed ensemble models by creating nine GAMs (each time omitting the data for one of the training cross-validation folds), and then averaging predictions using an elastic net linear regressor. Different alpha values of 0, 0.5 and 1 were tested and the best one retained based on the error reported by the *cv.glmnet* function of the *glmnet* R package (v4.1_7) (Tay et al., 2023). Using 5-fold cross-validation, the function automatically computes the best lambda value. We additionally avoided autocorrelation at the GFS level (as two samples were collected from within each GFS) by randomly sampling one of the two samples before creating each of the submodels. By fitting nine separate models to predict each cross-validation fold, we allowed a fraction to represent local variations in the data, and stacking multiple models allowed for all samples to be included in the final model (since only one sample per GFS is included in each model). Given the number of samples, models and predictions were first computed for all folds, and performance metrics were calculated on the result of all cross-validation folds afterwards (Collart and Guisan, 2023).

Feature selection was performed by building individual generalized additive mixed models with a spline for each covariate (k=3, bs='ts') along with the spatial spline, and were then ranked by -log(p-value) of the covariate spline. The top three variables were then kept to build a final model. This procedure was repeated on each of the nine models for each cross-validation fold. A final model was then built with the spatial spline and a spline for the three top variables with the following parameters: k=3,

bs='ts'. These parameters allowed a non-linear relationship while the low number of knots and the penalisation on the spline were added to control for the smoothness, in order to avoid overfitting.

Response curves were created using predictions for all sampling points and averaging values. The consistency of the stream parameter models created for each SSP scenario was assessed by comparing the selected features and their corresponding response curves. We observed reasonably shaped response curves for all models (Supplementary figure 4.3), and measured model performance as the cross-validation r^2 value computed comparing predicted and observed data for the held-out folds (Supplementary figure 4.2). The statistics comparing present and future conditions were extracted as for the climatic dataset statistics. The correlations between stream parameters' changes were computed by comparing the predicted changes in parameter A with the predicted changes in parameter B across all sites with a Spearman correlation. Plots containing ridges were created using the *ggridges* R package (v0.5.4) ("wilkelab/ggridges: Ridgeline plots in ggplot2").

Strain distribution modeling

Models of the abundance (normalized using the *recA* gene coverage) of each strain were built using climatic, glaciological, mineralogical data as well as environment stream parameters forecasts as covariate, chosen by feature selection (one set of features for each strain). The tested covariates are listed in Supplementary figure 4.3. Bioclimatic variables that describe annual seasonal processes and are potentially limiting factors for the survival and growth of species were included (Busby, 1991). To decrease the number of variables (in an effort to reduce potential overfit and collinearity issues), we created a PCA based on bioclimatic variables using the *prcomp* function of the R stats package *part* or the *r-base* (v4.3.1), and kept the first six axes as they represented more than 95% of the variation (Supplementary figure 4.6). Strain abundances and some covariates were log-transformed where necessary to improve the normality of residuals, adding a constant equal to the half of the smallest non-zero value. Before model fitting, all variables were scaled by subtracting the mean and dividing by the standard deviation. The strain distribution models were created similarly to the stream parameters and biomass models. Generalized additive models (GAMs) were used and stacked using an elastic net as for the stream parameters and biomass models. Differences are that the spatial spline approach was not used, but latitude, absolute latitude, and longitude, along with elevation and the slope of the stream, were added in the tested covariates. Thus, four variables were selected in the feature

selection procedure instead of three. Model performance was assessed using the $r^2_{prediction}$ since our goal was to assess the predictive power of our models (Araújo et al., 2019).

Strain abundance statistical analyses

The phylogenetic signal analysis was conducted by using the "lambda" method of the *phytools* R package (v1.5_1) that allows the correlation of parameters with the phylogeny. As response variables, we used $\log_2$ fold-change for the future changes and the $r^2_{prediction}$ as a measure of predictability. We report median change and interquartile ranges for the site-specific changes for all parameters. To test for a significant shift, we conducted Wilcoxon tests (wilcox.test function) of the future projections minus the present predictions. The evenness was assessed by comparing present and future median predictions of abundance for all strains, and showing that the relationship had a slope smaller than one (Supplementary figure 4.6). The decreasing clades were defined as the largest monophyletic sub-trees composed of only representatives that are predicted to decrease under the SSP3 scenario. The manipulation of the trees including the computation of sub-trees and branch lengths was performed using *phytools* and *ape* R packages (v1.5_1; v5.7_1) (Paradis and Schliep, 2019; Revell, 2012). The phylogenetic tree was plotted using the *ggtree* and *ggtreeExtra* R packages (v3.8.0; v1.10.0) (Xu et al., 2021; Yu et al., 2018).

To identify drivers based on the feature selection procedure results, only strain models with an $r^2_{prediction} > 0.05$ were considered. The bacterial strains were classified as "decreasing" in the future if the Wilcoxon test for future change was significant and the change negative. For these analyses, we considered only the SSP3 scenario, since the models did not differ significantly across scenarios (Pearson correlation coefficients with SSP3 = 0.98 and 1.0, for SSP1 and SSP5 respectively, $p < 0.001$). The relative ranks of the drivers in the strain distribution models were created using the number of times each of the covariates were selected during the feature selection procedures. To compare the importance of variables with the phylogenetic structure, we compared the number of shared top five variables as assessed by the feature selection procedure to the phylogenetic distance between pairs of strains. A Spearman correlation was computed to highlight the significance, *ggplot2* was used to create a scatter plot and fit

a 'loess' moving average curve. To compare classes, we separated taxonomies according to the GTDB-tk taxonomy, and ran Kruskal-Wallis test to test whether variables had a similar importance in all classes, or whether the variables had varying importance across taxa. This was computed within R with the *kruskal.test* function, and plots were created to show the distribution of median relative ranks across taxa with the *ggridges* R package (v0.5.4).

Functional analyses

A random forest analysis was performed to identify KOs that are associated with the strains that decrease in abundance. To this end, we took into account the phylogenetic structure by separating the data into 10 phylogenetic clusters and training models on nine of them while testing the KOs importance on the left-out one. The models were created using the *ranger* function of the *ranger* R package (v0.15.1) , allowing to easily test hyperparameters settings (random grid search, n=50, hyperparameters values tested are available in Supplementary table 4.8), compute feature importance, and also test the significance of the KOs with the method developed by Altmann et al. (2010) (Altmann et al., 2010). We considered as significant KOs with a p-value < 0.05, and in the 95% percentile of the importance values in at least one of the phylogenetic clusters. We considered "top" KOs that were significant in at least eight out of the 10 phylogenetic clusters (n=21), and for these, descriptions were gathered on the KEGG website (https://www.genome.jp/kegg/), and additional information was collected in literature for these genes (Supplementary table 4.6).

An enrichment analysis was carried out at the level of KEGG categories to identify categories overrepresented in the set of significant KOs. This was done with the *fisher.test* in R and p-values were corrected using the Bonferroni method. We only considered positively significant categories with the thresholds: p-value < 0.05, and odds ratio > 1. To compare the number of KOs per categories in the 'Decrease' and 'Others' bacterial genomes, we used GAMs to create models taking into account the completeness, the contamination and the N50 of these genomes and their interactions with a tensor (k=3, bs='cs'). Models were created for the counts of KOs in the genomes corresponding to the category (*gaussian()* family model). Weights were added by multiplying the inverse of the number of genomes for each phylogenetic cluster (so that all phylogenetic clusters weigh equally), the completeness of the genomes (to give less weight to the absences owed to incomplete genomes), and the mean relative abundance in present conditions. These models were fit using the bam function of the *mgcv* R package. Using the same approach, we tested for a difference between the two

groups of bacterial genomes in the total number of KOs, the genome length, the genome redundancy (unique KOs / total KOs), and also regressed the number of KOs of the genomes fitting a linear effect of the genome length (taking into account completeness, contamination and N50 with a tensor; function te with parameters k=3 and bs='cs'). For these models, we tested the difference between the decreasing strains and the others by fitting fixed effects (and an interaction for the KO number ~ genome length model) with ANOVAs as implemented in the *stats* package of the R programming language, and reported the estimated means and standard errors.

❖ *Code and data availability*

The data generated in this study have been deposited in Zenodo, under the DOI https://doi.org/10.5281/zenodo.10409762 (https://zenodo.org/records/10409762), and the code in a Github repository: https://github.com/Mass23/CrystalBall.

❖ *Acknowledgments*

❖ **Figures**

Figure 4. 1 Worldwide distribution of glacier-fed streams sampled, sampling design and modeling framework.

(**A**) Framework of our hierarchically structured model to predict the abundance of 2,333 bacterial genomes assembled from metagenomes (Methods). We combined upstream and downstream streamwater physicochemistry and sediment mineralogy data with present-day climatological and glaciological data to train models that relate environmental shifts to deglaciation. Projections of climate and glaciology were available from the CHELSA database v2.1 and in the output of GloGEM models for SSP1, SSP3 and SSP5. This allowed us to forecast the environmental template under SSP-scenarios for 2070-2100, accounting for the spatial structure of our data. In a second step, we

combined climatic, glaciological, mineralogical and projected environmental data, with abundance data of the bacterial strains in a modeling framework similar to species distribution models. Feature selection was used for each bacterial strain to individually constrain the environmental niche, and we predicted strain-level abundances under future climate change. (**B**) Mountain ranges and number (in circles) of glacier-fed streams sampled per mountain range. (**C**) Conceptual depiction of sampling designed as a time-for-space substitution based on an upstream and downstream reach (median distance between reaches: 515 m; interquartile range, IQR: 236 - 933 m) and simulating the glacier retreat. The models are trained on environmental and metagenomic data from upstream and downstream reaches; predictions of present-day and future conditions in upstream reaches take into account glacier shrinkage.
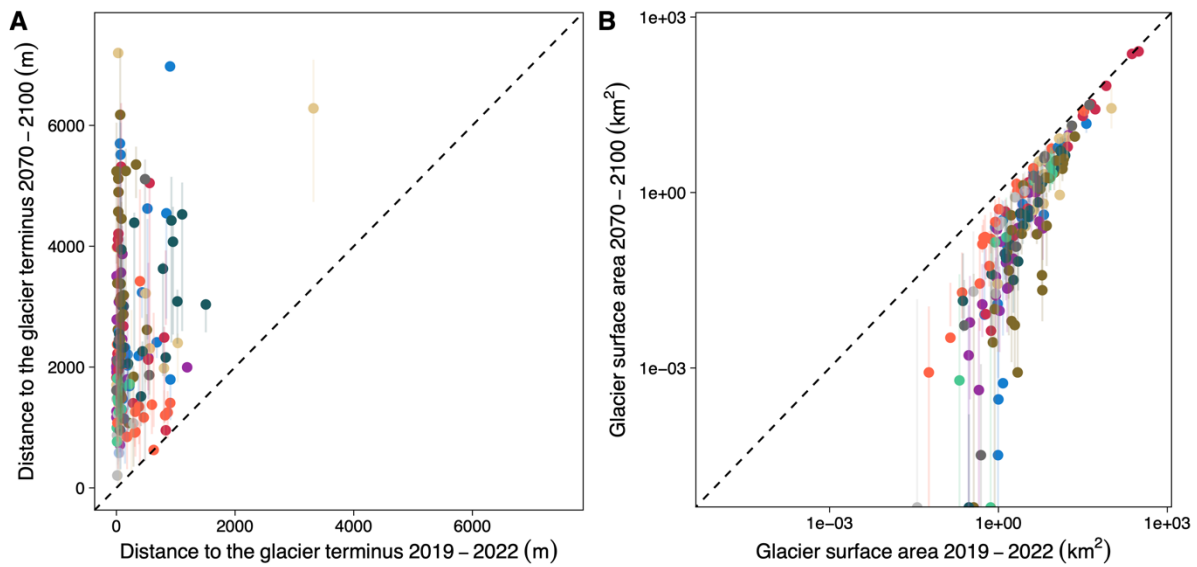
Figure 4. 2 Sampling extent and glacier shrinkage

(**A**) Distance between the upstream sampling point and the glacier snout, and (**B**) the area of the glacier according to the GloGEM models comparing values for the sampling year, and the averages for the 2070-2100 predictions. The dots correspond to the SSP3 scenario, and the error bar displays values for the SSP1 and SSP5 scenarios.
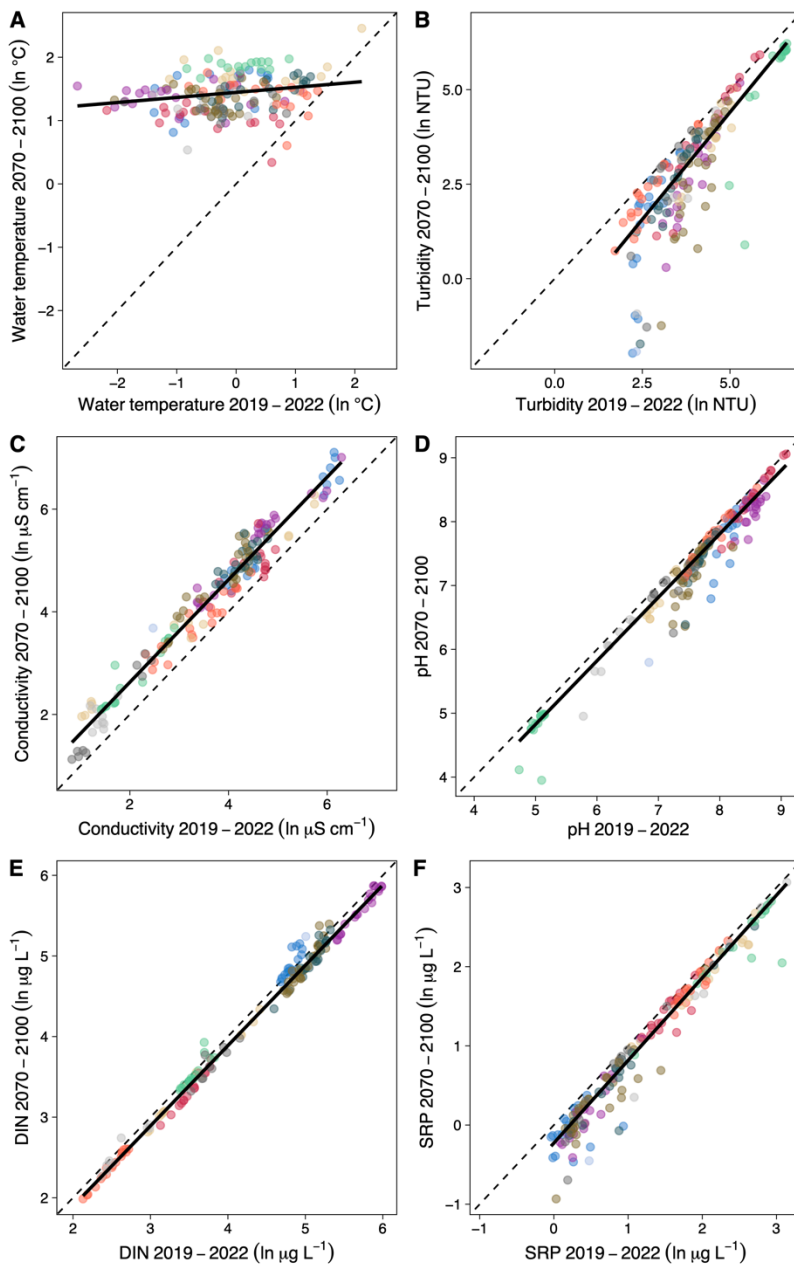
Figure 4. 3 Environmental template predictions

Scatter plots comparing predictions of stream parameters for current conditions and future projections for the 2070-2100 period. These parameters are streamwater temperature (**A**), turbidity (**B**), conductivity (**C**), pH (**D**), dissolved inorganic nitrogen concentration (DIN, **E**), and soluble reactive phosphorus (SRP, **F**). A quantile regression (median) is shown with the solid line, and the one-to-one line is displayed with a dashed line. The colors represent the different mountain ranges with a palette matching the world map in Figure 4.1A.
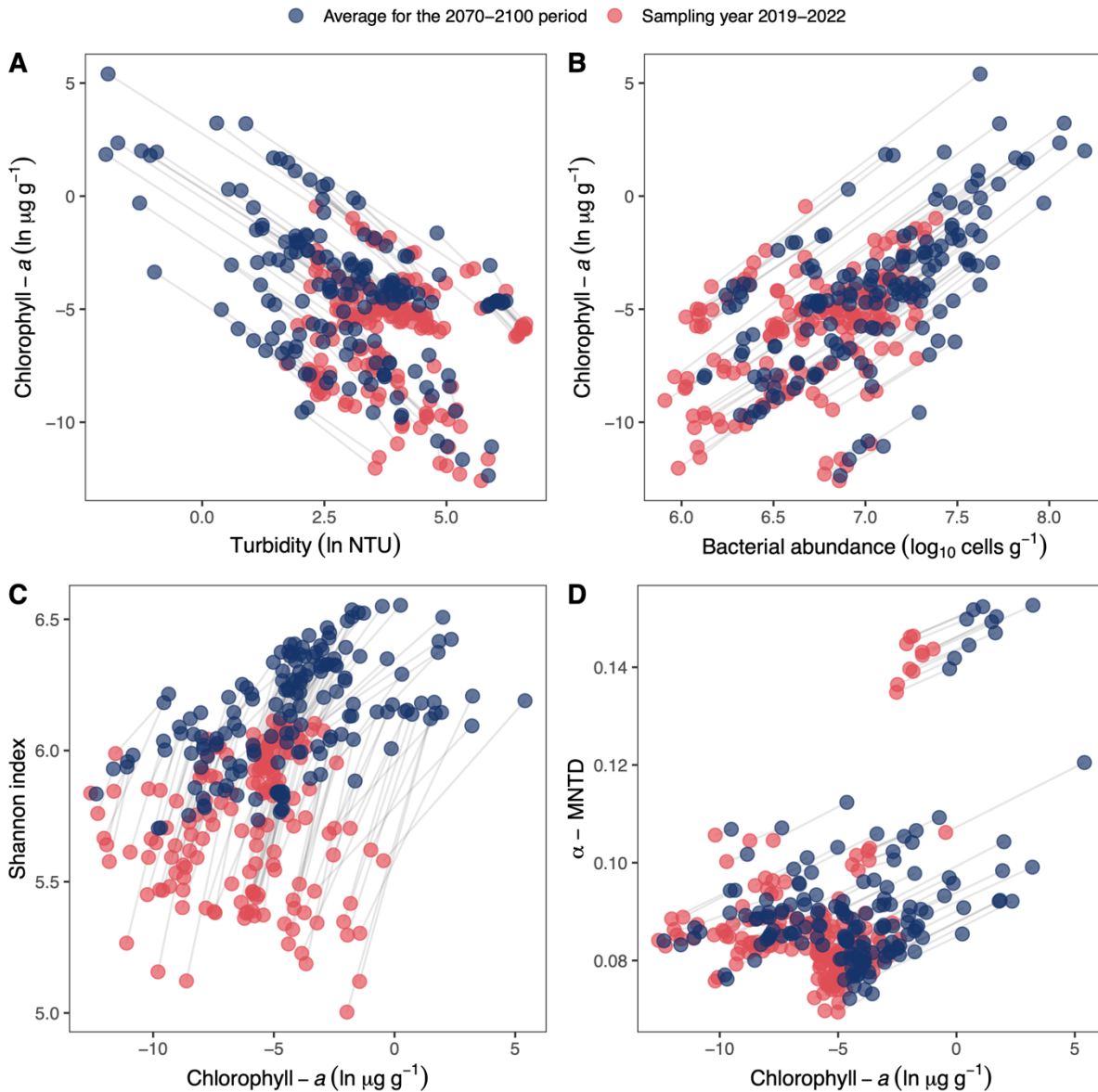
Figure 4. 4 The 'greening' of the world's glacier-fed streams.

(**A**) Scatter plots highlighting the relationship between algal biomass (chlorophyll-*a*) with streamwater turbidity, (**B**) algal biomass with bacterial cell abundance, (**C**) algal biomass with the Shannon index of bacterial communities, and (**D**) algal biomass with the within community mean nearest taxon distance (a-MNTD) of the bacterial communities. Predicted values for the sampling year, and future projections for the 2070-2100 time period are compared, and the current and future conditions for each site are linked with grey lines. The correlations between predicted changes were all highly significant (p < 0.001, Spearman Rho = -0.96, 0.98, 0.69, 0.97 for panels **A**, **B**,**C** and **D**, respectively).
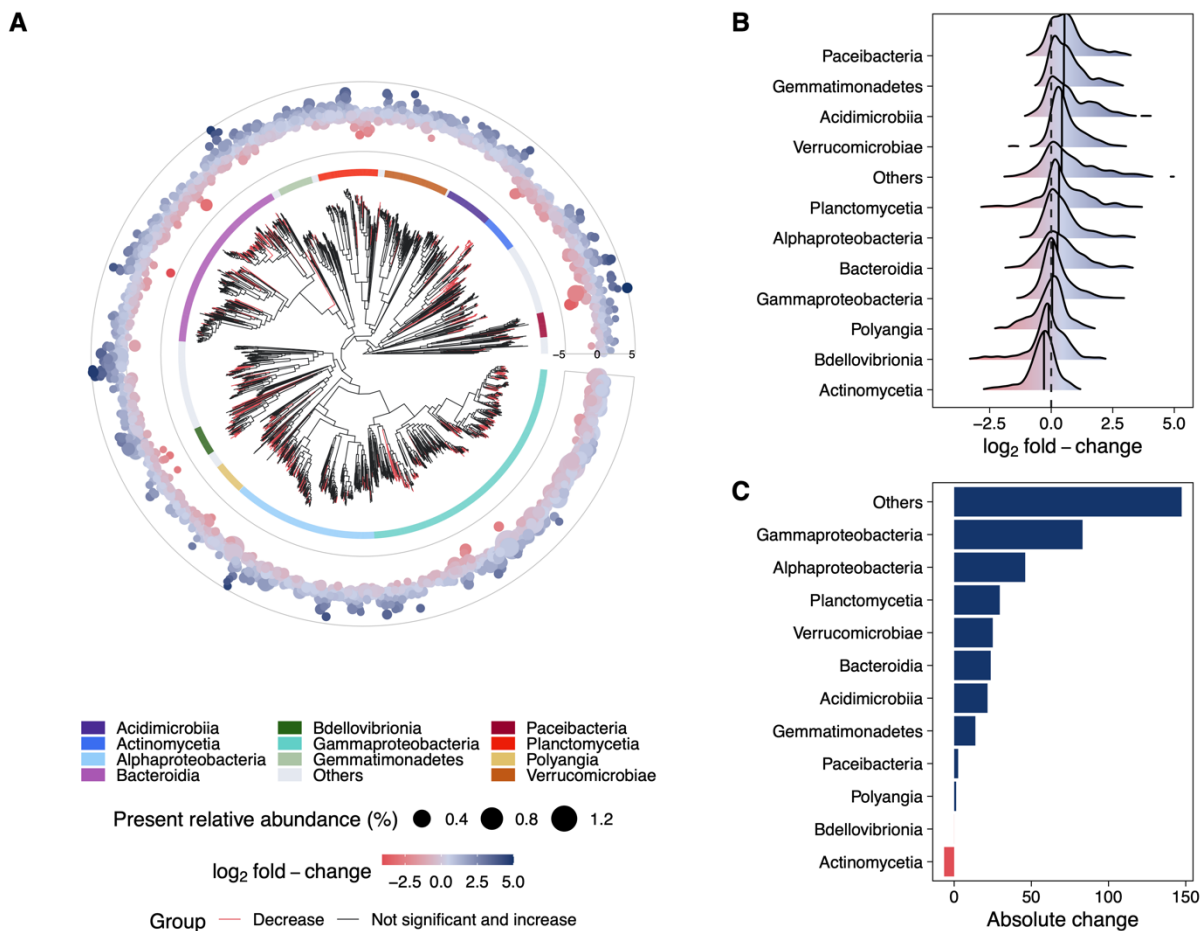
Figure 4. 5 Shift in bacterial communities under climate change

(**A**) Phylogenetic tree showing the log2 fold-change between current and future projections of strains' abundances. The color of the dots matches the log2 fold-change, and the size of the dots represents current relative abundances. In the phylogenetic tree, we highlight tips and clades that are predicted to decrease in abundance significantly by coloring their edges in red, highlighting the high phylogenetic signal in the predicted changes. Additionally, the 11 most abundant bacterial taxonomic classes are highlighted in the inner ring. (**B**) Distributions of the relative, measured as log2 fold-change) and (**C**) summed absolute median changes in abundance for the 11 most abundant classes (highlighted in the tree).
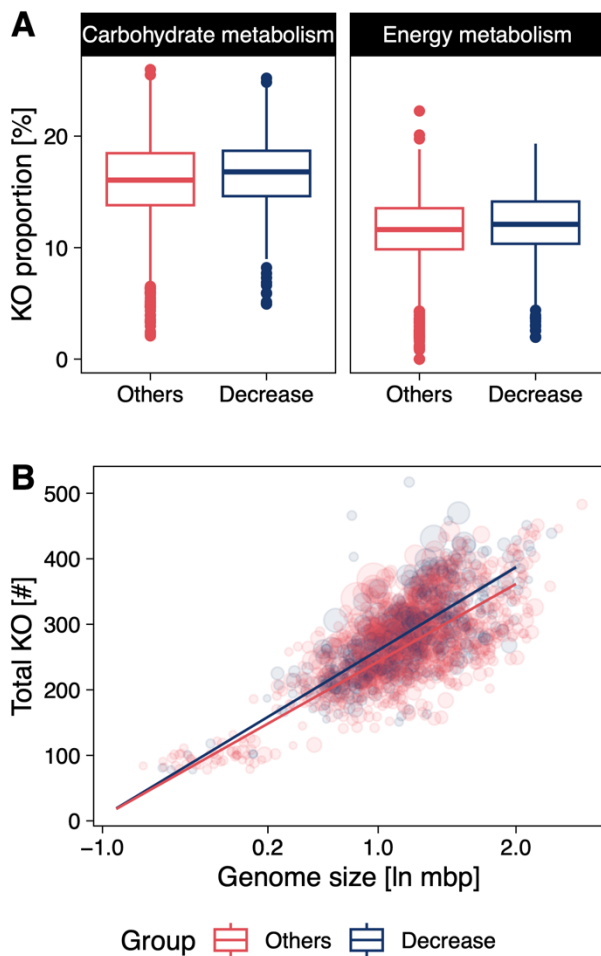
Figure 4. 6 Genomic properties of the strains that are predicted to decrease in abundance

(**A**) Proportion of KEGG orthologous groups (KOs) from these categories (Carbohydrate metabolism and Energy metabolism) in the genome of the strains that are predicted to decrease in abundance (Decrease, blue), and the others (Others, red). The difference in mean was statistically significant for both categories with a fixed effect (ANOVA, $p < 0.001$) in the GAM model approach taking into account completeness, contamination and the N50 of these MAGs. On average (with standard errors), the decreasing strains had 4.02 more KOs in the "Carbohydrate metabolism" category (mean$_{others}$ = 43.5±0.5) and 3.44 more in the "Energy metabolism" category (mean$_{others}$ = 31.2±0.4). (**B**) GAM model with a linear effect of ln genome size on the total number of KOs for both groups, taking into account the completeness, contamination and N50 of the genomes into account with a tensor. The slopes were significantly different for both groups (ANOVA, $p < 0.001$, slope$_{Decrease}$=127.3±1.1, slope$_{Others}$=118.6±0.7).

Chapter 5. Glacier influence shapes the genomic landscape of the downstream aquatic microbiome

Massimo Bourquin[1], Hannes Peter[1], Grégoire Michoud[1], Aileen Geers[1], Susheel Bhanu Busi[2], Tom Ian Battin[1]

[1]River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne, EPFL, Lausanne, Switzerland

[2]UK Centre for Ecology and Hydrology, Wallingford, United Kingdom

❖ *Authors' contribution*

In this chapter, I conceptualised the study and performed all analyses. The dataset used was created for chapter 4. All co-authors gave advice for the analyses, and helped to write the manuscript.

❖ *Publication*

❖ Abstract

The eco-evolutionary interactions that shape microbial genomic landscapes determine the success of microbes and have therefore have fascinated microbial ecologists since long. Here, leveraging 2,855 metagenome-assembled genomes sampled by the *Vanishing Glacier* Project from glacier-fed streams (GFSs), we shed light on the genomic landscape of the benthic microbiome in these harsh ecosystems — now vanishing because of climate change. Owing to the glacier influence, the GFS water is notoriously cold and ultra-oligotrophic, and within highly unstable sediment beds. Along gradients of glacier influence and concomitant variation in benthic algal biomass across 149 GFSs draining Earth's major mountain ranges, we show that GFS bacteria optimise their genomes in terms of size, coding density, redundancy, and translational machinery. We develop a novel, phylogeny-rooted analytical framework that allows pinpointing the phylogenetic depth at which genomic optimization occurs. These analyses reveal deep-branching patterns of genomic optimizations and highlight *Gammaproteobacteria* in shaping community-level genomic landscapes along gradients of glacier influence in GFSs. Using comparative pangenome analyses, we further reveal metabolic gains and losses of the GFS *Gammaproteobacteria*. Our work shows how genomic optimization, beyond genome size reduction, is shaped by selective environmental constraints in an extreme environment. These insights are important as they reveal putatively important eco-evolutionary processes that are now changing at rapid pace due to climate change.

❖ Introduction

How microorganisms optimise genomic features, primarily genome size, has been a mainstay of microbial ecology due to the eco-evolutionary consequences that arise from these adaptive strategies (Giovannoni et al., 2014). Genome size variation and particularly genome streamlining, driven by imbalances between insertions and deletions, have been attributed to environmental constraints such as oligotrophy, to genetic drift in small and isolated populations, or to interactions among symbiotic microbes that allow for adaptive gene loss (Coleman and Chisholm, 2010; Giovannoni et al., 2014). For example, thermophilic microbes thriving in hot springs were found to have small genomes (Sabath et al., 2013), whereas psychrophilic microbes in

cryospheric environments tend to have relatively large genomes (Bourquin et al., 2022), and gene expansion was shown for several cold-adapted species (Dieser et al., 2019; Liu et al., 2020). Increased genome size has been attributed to gene expansion as an adaptive strategy to cope with fluctuating environmental conditions, even under nutrient depletion (Bentkowski et al., 2015; Konstantinidis and Tiedje, 2004; Props et al., 2019). Besides habitat characteristics (e.g. oligotrophy and temperature), microbial lifestyles such as free-living pelagic *versus* attached microbial forms of life, have been associated with genome size variation (Chuckran et al., 2021; Rodríguez-Gijón et al., 2022; Salcher et al., 2019).

Glaciers exert direct influences on the physicochemical characteristics of proglacial streams (Milner et al., 2017). Glacier-fed streams (GFSs) are ultra-oligotrophic, cold and unstable environments, yet they harbour diverse microbial communities (Ezzat et al., 2022; Wilhelm et al., 2013). In GFSs, bacteria forming biofilms attached to sedimentary surfaces dominate microbial life, where they orchestrate important ecosystem functions (Busi et al., 2022; Kohler et al., 2022). These bacterial communities are shaped by selective environmental conditions, which is reflected by deterministic community assembly and microdiversity (Brandani et al., 2023; Fodelianakis et al., 2022). Recent work has shown how climate change may relieve these selective constraints that are associated with glacier influence (Kohler et al., 2022). Glacier meltwaters are generally cold, but GFS water temperature rises with increasing distance to the glacier snout (Brown et al., 2007; Milner et al., 2017). GFS streamwater is often turbid because of high loads of fine suspended sediments eroded and mobilised by glacier activity (Brown et al., 2007; Milner et al., 2017), which reduce light available for primary producers (i.e., benthic algae). Moreover, pronounced oligotrophy of GFSs, linked to the scarcity of vegetation and soils within the catchments they drain, induces energy and nutrient limitation of the microbial communities (Kohler et al., 2024).

Benthic biofilm communities in GFSs represent an ideal model to study eco-evolutionary causes and consequences of genome optimization. The pronounced oligotrophy may favour smaller genomes in GFSs. On the other hand, sessile bacteria in biofilms tend to have large genomes in freshwater (Chiriac et al., 2023). Glacier influence, such as rapid variation in flow and sediment loads and low streamwater temperatures may also promote larger genomes (Sabath et al., 2013). The latter notion is supported by smaller average genome sizes reported from GFSs compared to the genome sizes from tributary streams that are not under glacier influence albeit draining the same proglacial floodplains (Michoud et al., 2023).

Here we analyse metagenomic and environmental data from GFSs sampled by the *Vanishing Glacier* Project from Earth's major mountain ranges to investigate relationships between glacier influence and genomic features of the benthic microbiome. Focusing on the variation of genomic traits along gradients of glacier influence across all GFS, we aim to identify signatures of adaptation to the GFS environment. We consider genomic traits prevalent at stream reaches under high glacial influence to be "optimized" for life in GFS. Moreover, given the tight link between genome size and evolutionary history, it is important to take phylogeny into account when assessing variation in genomic features (Martinez-Gutierrez and Aylward, 2022). To this end, using 2,855 strain-level resolved (i.e., 99% average nucleotide identity, ANI) metagenome-assembled genomes (MAGs), we establish a novel analytical framework to resolve the phylogenetic signatures of genomic features in GFSs. To achieve this, we first identified the main environmental gradients common to the world's GFSs and contributing to selective constraints. Next, we determined the depth at which phylogenetic signal in genomic features arise. These analyses suggest that genomic optimization within *Gammaproteobacteria*, which are widespread, abundant and microdiverse in GFSs (Brandani et al., 2023; Fodelianakis et al., 2022; Michoud et al., 2023), accounts for much of the variation in community averages of genomic features along the main gradients of glacier influence. Finally, we use a pangenome approach to identify specific genomic adaptations (i.e., gene gains and losses) associated with the ecological success of *Gammaproteobacteria* in GFSs. Our work sheds new light on how environmental constraints shape genomic features and contribute to the success of specific clades. These insights are relevant as glacier influence on downstream ecosystems diminishes because of climate change, ultimately reducing selective constraints and potentially threatening microorganisms that are well adapted to the GFS environment. Moreover, our work disentangles the contribution of specific clades to community-level features, thereby providing a phylogeny-rooted framework that provides a lineage-specific understanding of genome variation.

❖ *Material and methods*

Glacier-fed stream sampling and environmental parameters

We sampled benthic biofilms (upper 5 cm of the streambed) from 149 GFSs in the European Alps, Scandinavian Mountains, Himalayas, Pamir and Tian Shan, Ecuadorian Andes, Southwest Greenland, Russian Caucasus, Rwenzori in Africa, and Southern Alps in New Zealand between January 2019 and July 2022. GFSs were sampled in spring or autumn during 'windows of opportunity' when streamflow and streamwater turbidity are relatively low; this sampling strategy facilitates comparability between GFSs. We did not sample GFSs from heavily debris-covered and rock glaciers, and we avoided GFSs downstream of proglacial lakes, with debris flows, or tributaries in the reaches above the sampling sites. At each GFS, we sampled an upstream reach, as close as possible to the glacier snout, and a downstream reach. Within each reach, sandy sediments (250 μm to 3.15 mm size fraction) were collected from three independent patches (approximately 10 m apart). All sampling devices were flame-sterilised in the field. Sediment samples were transferred into sterile cryovials, immediately flash-frozen in liquid nitrogen in the field and subsequently stored at -80°C before and following shipping to Switzerland for DNA extraction and biomass analyses.

For each GFS, the distance to the glacier snout was calculated based on georeferencing (GPSMAPR 66s, GARMIN) of the sampling reach, as well as glacier surface area and glacierized percentage catchment based on satellite imagery (Sentinel-2; Level 2a, March 2019 - July 2022 from *scihub.copernicus.eu*) and a catchment definition derived from the ASTER Global Digital Elevation Model (GDEM) v3. (NASA/Meti/Aist/Japan Spacesystems and US/Japan Aster Science Team, 2019). The glacier index (GI) was calculated as $\frac{\sqrt{Glacier\ area}}{\sqrt{Glacier\ area}\ +\ Distance\ to\ the\ glacier}$ according to Jacobsen & Dangles (2011) (Jacobsen and Dangles, 2012). Benthic chlorophyll-*a*, a proxy for algal biomass, was extracted from the sediment (90% EtOH) in a hot (78°C) water bath for 10 min and further incubated (24 h, 4°C). After vortexing and centrifugation, chlorophyll-*a* concentration in the supernatant was quantified using a plate reader (BioTek Synergy H1; EX/EM: 436/680) and a spinach chlorophyll-a standard (Sigma Aldrich) and normalised to dry mass (DM) of sediment.

Metagenomics

Metagenomes were sequenced for 149 sediment samples. DNA extraction, purification, library preparation, sequencing and metagenome assembly steps were

performed as described in Busi et al. (2022) (Busi et al., 2022). Briefly, 5 g of GFS sediments were treated using a phenol:chloroform-based extraction method subsequently followed by an ethanol precipitation step. This protocol yielded on average 50 ng of DNA per sample which was used for library preparation using the NEBNext Ultra II FS library kit, which also included 6 PCR cycles. Sequencing was performed at the Functional Genomics Centre Zurich using a S4 flowcell on a NovaSeq (Illumina).

The metagenomic sequence data was processed using the Integrated Meta-omic Pipeline (IMP) workflow (version 3.0; commit# 9672c874) (Narayanasamy et al., 2016). Briefly, adapter trimming from reads using *trimmomatic* (Bolger et al., 2014) is followed by an iterative assembly using *MEGAHIT* (Li et al., 2015) and *Flye* (Kolmogorov et al., 2020). To reduce computation time for binning, we removed sequences in the assembly < 1.5 kbp and randomly selected 10% of the pre-processed reads using *seqtk* (v1.3) (Li, 2023). For each individual assembly, we then mapped the selected reads of the 5 spatially closest samples (Euclidean distances of gps coordinates) using *BWA-mem* (v0.7.17). We then used *MetaBAT2* (v2.15) (Kang et al., 2019), *CONCOCT* (v1.1.0) (Alneberg et al., 2013) and *MetaBinner* (v1.4.3) (Wang et al., 2023) using default parameters to obtain bins {see Code availability}. The quality of bins was assessed with *CheckM2* (v1.0.1) (Chklovski et al., 2023), and finally *DASTool* (v1.1.4) (Sieber et al., 2018) was employed to generate a non-redundant set of bins using a score threshold of 0.3.

Bins from all samples (including the ones generated by *IMP3*) with a completeness of more than 50% were then selected for further analyses which accounted for 12,599 bins. We then used *MDMCleaner* (v0.8.3) (Vollmers et al., 2022) to reduce contamination of the bins. Finally, after rerunning *CheckM2* on the bins to get final estimates of completeness and contamination, we used *dRep* (v3.2.2) (Olm et al., 2017) to dereplicate bins using a minimum completeness of 70% and maximum contamination of 10% and an ANI of 99% to obtain 2855 strain-level MAGs. GTDB-Tk (v 2.1) (Chaumeil et al., 2020; Chklovski et al., 2023) was used to assign taxonomy to MAGs. We further used the concatenated alignment of 120 ubiquitous single-copy proteins created by GTDB-Tk to *de novo* generate a phylogenetic tree using FastTree2 (v2.1.11) (Price et al., 2010) under the WAG model of protein evolution with gamma-

distributed rate heterogeneity. Functional annotation of the MAGs was performed with *EggNog-Mapper* (v2.1.9) (Huerta-Cepas et al., 2017) after obtaining coding regions (CDS) with *prodigal* (v2.6.3) (Hyatt et al., 2010). The coverage of MAGs was estimated by mapping reads of samples to the genomic contigs using *CoverM* (v0.6.1, available at https://github.com/wwood/CoverM) using the *trimmed_mean* parameter. We normalised the coverage by similarly mapping reads on the recA gene (K03553). For prevalence, presences were defined as abundance above a 10x recA coverage abundance threshold.

Dimensions of glacier influence and community-weighted mean genomic properties

To identify the main environmental gradients across all GFS samples, Principal Component Analysis (PCA) was performed with the *prcomp* function in R (version 4.3.0), and using a non-redundant set of key physico-chemical as well as glacier-associated measures (glacier area, glacier coverage, glacier index, streamwater temperature, distance to the glacier, benthic chlorophyll-*a*). Community-weighted means (CWM) of genomic features (genome size, gene number, tRNA number, GC content, coding density and redundancy index) were tested with linear effects against the first two principal components using generalised additive models (GAMs) created with the *bam* function of the *mgcv* R package (v1.9.0) (Wood, 2023). For this, genomic features were first normalised using completeness and contamination as follows: $value_{normalised} = value * (1/completeness) * (1 - contamination)$. CWM were then obtained by weighing normalised genomic features by MAG relative abundances and averaging across MAGs present in any given sample. To account for large-scale spatial patterns, we used a smoothed spline (bs = 'sos', k = -1) based on latitude and longitude in the GAMs. Detailed results of these GAMs are available in Supplementary Table 1. Significant linear effects (p<0.01) were visualised using mean and standard errors of predictions across all GFS in the dataset. All figures were created using the *ggplot2* (version 3.4.3) and *ggpubr* (version 0.6.0) *R* packages (R version 4.3.0) (R Core Team, 2023; Wickham, 2016).

Abundance-based phylogenetic permutation

To understand how the phylogenetic structure affects CWM genomic features, we developed a null-model approach that randomly permutes abundances in a phylogenetic-bin based framework. For 40 values of relative phylogenetic height (h) uniformly distributed between zero and one (i.e. scanning the phylogenetic tree from

the root to the tips), we performed phylogenetic agglomeration using the "average" method of the *hclust* R function on the cophenetic distances obtained with the *cophenetic.phylo* function of the *ape* R package (v5.7-1) (Paradis and Schliep, 2019). Subsequently, for each value of h, abundances were randomly permuted within phylogenetic bins (20 iterations). Finally, GAM models accounting for spatial structure (i.e., including a smoothed spline (bs = 'sos', k = -1) on latitude and longitude as covariate) were created, testing for a linear effect of glacier influence on genomic features. Hence, this approach tests for associations between CWM genomic features and environmental parameter compared to null-model expectations across phylogenetic depth. This approach further allows identifying the relative depth at which phylogenetic signal in CWM genomic features appear along the gradients of glacier influence. Significant coefficients were assessed by combining p-values of the linear coefficients over the 20 iterations using Stouffer's method in the *poolr* R package (v1.1-1), the mean and the standard deviation of the coefficients were computed to summarise the null-model permutations (Cinar and Viechtbauer, 2022).

Additionally, this approach allowed us to pinpoint phylogenetic clades contributing to the community-level signal at a specific phylogenetic height. To this end, we used a leave-one-cluster-out approach, computing coefficients with and without a given phylogenetic cluster, and comparing the resulting coefficients' distributions. Wilcoxon tests were used to test for difference in coefficient distributions, a median relative effect was computed comparing the median values with and without the target phylogenetic cluster ((value with – value without) / (value with)).

Taxonomic summary and models

CWM genomic features of *Gammaproteobacteria* were compared to all other MAGs. The taxonomic summary comparing genomic features of strains affiliated to *Gammaproteobacteria* to other taxonomic classes was created using the *dplyr* R package (v1.1.3). Wilcoxon sign rank tests were used to compare the distributions. Relative abundance and prevalence (i.e., the number of occurrences across GFSs) were combined to estimate the 'ecological success' of MAGs. Abundant and wide-spread (i.e., prevalent) MAGs were thus defined as being 'ecologically successful' in GFSs worldwide. GAM models were built using a spline (k=5, bs='ts') for these

'ecological success' covariates, and genome size and coding density were used as response variables. We compared one model with a spline for all MAGs, and one with a different spline for *Gammaproteobacteria* and all other MAGs (using the 'by' argument in the spline). A Bayes factor analysis was used to compare both models, using the *test_performance* function of the *performance* R package (v0.10.5) (Lüdecke (@strengejacke) et al., 2023), and a Bayes factor above 3 was considered significant.

Pangenome analysis

We used *mOTUpan* (Buck et al., 2022) to create pangenomes of high-quality MAGs defined as a completeness higher than 90% (median: 95.06%; IQR: 92.22 - 98.05%), and a contamination threshold of <10% (median: 1.91%; IQR: 0.99 - 3.42%). The first pangenome was created with all MAGs from the phylogenetic cluster (n=127, identified with the phylogenetic permutation) corresponding exactly to all representatives of the *Gammaproteobacteria* class. A second pangenome (n=223) was created with these *Gammaproteobacteria* MAGs and MAGs classified as *Alphaproteobacteria*. Default parameters were used on the .faa files, and the *CheckM2* output was used to initiate completeness values, the --seed parameter was set to 90. KO and CAZyme annotations were used to assign functions to each gene cluster. We then summarised the KOs and CAZymes present in the core genome of both pangenomes. KOs gained by *Gammaproteobacteria* were defined as those that were part of the core genome of the *Gammaproteobacteria* pangenome, but not in the pangenome created including *Alphaproteobacteria* MAGs. Similarly, gene losses were defined as genes that were part of the core genome in the pangenomes of both *Proteobacteria* classes, but not in the pangenome of only *Gammaproteobacteria*. KEGG pathways and categories were then used to summarise gene losses and gains. The description of KOs (available at: https://www.genome.jp/kegg/) was used to identify words that occurred often in the set of KOs gained and lost, the number of descriptions matching queries of interest was computed using regular expression (*regex*) subsetting with the *grepl* function of R (R Core Team, 2023).

❖   *Results and discussion*

Bulk genomic properties of the GFS microbiome under glacier influence

The GFS environment (e.g., discharge, temperature) is directly influenced glaciers, primarily through magnitude and variation of meltwater runoff (Brown et al., 2007; Milner et al., 2017). Runoff determines hydraulic stress, channel stability and sediment loads, whereas streamwater temperature affects metabolic processes (Kohler et al.,

2024; Zhang et al., 2022). These physical processes are largely driven by glacier size, which translates into runoff magnitude and variability (Zhang et al., 2022). Employing PCA on the complete set of measured environmental parameters, the first principal component (PC1; 44.6% explained variance) revealed a gradient of benthic chlorophyll-*a* inversely related to glacier area across all studied GFSs (Figure 5.1A). This is striking given the overall low chlorophyll-*a* content (median: 0.0056 μg g^{-1} dry mass; IQR: 0.0007-0.0272) in the GFSs, which underscores the responsiveness of benthic primary producers to environmental conditions. High runoff and loads of suspended sediments produced by large glaciers abrade benthic algae and attenuate light, thereby inhibiting primary production in GFSs and keeping chlorophyll-*a* concentrations low (Boix Canadell et al., 2021; Uehlinger et al., 2010). PC2 (21.7% explained variance) depicts a gradient of streamwater temperature related to both distance to the glacier snout and glacier area (as encapsulated by the glacier index) across all GFSs. In fact, streamwater warms up with downstream distance from the glacier and depending on the magnitude of runoff from ice melt. Taken together, the PCA reveals two main dimensions quantifying downstream glacial influence at a global scale, and we will explore them as potential underpinning processes of the genomic landscape of the GFS microbiome.

Weighted by relative abundance of MAGs, bacterial genomes across all GFSs were relatively large in terms of size, had a high number of genes, and showed high GC content (Figure 5.1B). These values are bracketed by those reported from other GFSs (Michoud et al., 2023), various cryospheric ecosystems (e.g., permafrost, glacier ice) (Bourquin et al., 2022), and psychrophiles (Sabath et al., 2013). Bacterial genomes contain only little non-coding DNA (on average 87%) (Land et al., 2015), hence variation in gene number and genome size are generally tightly linked (Lynch, 2006), a relationship attributed to the importance of effective population size (Bobay and Ochman, 2017). Additionally, GC content, coding density and genome size have been shown to positively correlate in bacteria (Almpanis et al., 2018; Bentley and Parkhill, 2004; Sabath et al., 2013). However, we find a low number of tRNAs compared to published psychrophilic, mesophilic, and thermophilic cultured isolates (Dutta and Chaudhuri, 2010), which we mainly attribute to the known discrepancy between MAGs and isolates (Meziti et al., 2021). Because translation is energetically expensive, tRNA abundance has been linked to shorter minimal generation time and adaptability to

different environmental conditions (Arella et al., 2021). The redundancy index (i.e., the ratio between the total number of KOs to the number of unique KOs, median RI ~ 1.4) was lower than previously reported in cryoconites biofilms (Zhang et al., 2023), which seems intuitive as the dynamic and unpredictable GFS environment may select for functional plasticity rather than redundancy.

To further explore glacier influence on these genomic properties of the GFS microbiome, we implemented GAMs accounting for large-scale spatial variation and isolating linear effects of environmental parameters on genomic properties. GAMs revealed positive associations between benthic chlorophyll-*a* content (correlated with PC1) with average genome size, gene number, and tRNA number, whereas covariates correlating with PC2 (i.e., glacier distance, water temperature and glacier index) were associated with the redundancy index (Figure 5.1C). This further highlights the importance of functional plasticity in the dynamic environment of streams with high glacier influence. This finding suggests that benthic algae, through the exudation of energy-rich macromolecules, relieve GFS bacteria from energy and carbon limitation, ultimately promoting bacteria with larger genomes as glaciers shrink. Indeed, metabolic interactions between microbial heterotrophs and algae have been repeatedly reported from stream biofilms (Battin et al., 2016; Haack and McFeters, 1982), which may be particularly important in GFSs largely devoid of allochthonous sources of organic carbon (Busi et al., 2022; Kohler et al., 2024). Furthermore, these analyses revealed increasing numbers of tRNAs with diminishing glacier influence, which essentially follows the observed trends of genome size (Figure 5.1C). While tRNAs have been associated with cold adaptation and associated post-translational modifications in bacteria (Dalluge et al., 1997; Lorenz et al., 2017), work on cultured isolates showed that psychrophile genomes have elevated numbers of tRNAs (Dutta and Chaudhuri, 2010). Nevertheless, translational efficiency has been shown to be low in organisms that are able to thrive in multiple habitats, and this could potentially explain the low number of tRNAs that we observed (Arella et al., 2021). Importantly, our analyses have not revealed any major variation in coding density along any of the glaciological variables tested, confirming that in reduced genomes, genome size, number of genes and proportion of non-coding DNA decrease at similar rates.

Gammaproteobacteria contribute to genome optimization in GFS at the microbiome level

Variations in genomic properties of bacterial communities along environmental gradients can either be explained by changes in abundance or the replacement of taxa

with different genomic characteristics. Moreover, shared evolutionary histories among members of the microbiome can bias estimates of relationships between genomic properties and environmental constraints (Martinez-Gutierrez and Aylward, 2022). For example, accounting for phylogenetic dependencies, a previous study on bacterial and archaeal genome size identified deep phylogenetic signatures in genome size variation (Martinez-Gutierrez and Aylward, 2022). Therefore, to assess phylogenetic signatures in genome size, we developed a novel null model-based phylogenetic approach to first identify the phylogenetic depth at which differences in genomic properties arise along environmental gradients. Using a leave-one-out approach of individual clades at this threshold phylogenetic distance, we identify clades that contribute most to this signal. Finally, we investigated whether changes in abundance or replacement of MAGs within clades explain the variation in genomic properties along gradients of glacier influence in GFSs.

We found a significant phylogenetic signature at low depth (i.e., among closely related members, relative phylogenetic tree height between 0 and 0.1) for all tested genomic properties and environmental constraints (Figure 5.2A). This suggests that genomic optimization occurs among closely related taxa (below the median genus-level phylogenetic depth). Relationships between genome redundancy and glacier index, the distance to the glacier and streamwater temperature were exclusively shaped by variation at low phylogenetic distances (i.e., approximately 0.25 relative phylogenetic tree height corresponding to genus-level distance). Relationships between genomic properties and benthic chlorophyll-a arise at greater phylogenetic depth (i.e., approximately 0.6 relative phylogenetic tree height, corresponding to class-level distance) (Figure 5.2B).

Strikingly, the leave-one-cluster-out analysis highlighted the importance of one cluster, comprising all MAGs classified as Gammaproteobacteria and shaping genomic features at the microbiome level across the gradient of chlorophyll-a. We also report the importance of various taxa (i.e., Cyanobacteria, a clade comprising members of Bacteroidota and Fibrobacterota, and one comprising representatives of Acidobacteriota, Desulfobacterota, Myxococcota and Nitrospirota, Supplementary Table 2) in shaping community averages along the benthic chlorophyll-a gradient, and

the relationship between water temperature, glacier index and the distance to the glacier with the redundancy index (Supplementary Table 3).

Our analyses unveil the pivotal role of Gammaproteobacteria in driving variation in microbiome-average genomic features, further emphasising the importance of Gammaproteobacteria for the global GFS microbiome. Given their role, we next compared genomic features of Gammaproteobacteria MAGs to all other MAGs (Figure 5.3). Indeed, we found a strong negative relationship between relative abundance of Gammaproteobacteria and benthic chlorophyll-a (Figure 5.3A). Given that benthic chlorophyll-a diminishes with increasing glacier influence, this finding indicates that Gammaproteobacteria thrive and numerically dominate in the GFSs under high glacier influence. This suggests that members of the Gammaproteobacteria are particularly well adapted to GFS reaches close to glacier snouts, fed by large glaciers and devoid of major autochthonous energy sources. Additionally, we found that Gammaproteobacteria MAGs had significantly increased coding density (median difference: 2%), but fewer tRNAs (median difference: 3.05), and a lower redundancy index (median difference: 0.025) compared to all other MAGs (Fig 3B). Interestingly, however, genome size and gene numbers of Gammaproteobacteria were not significantly different from the other MAGs. This contrasts our findings on community-weighted average genomic features and suggests that abundance differences of Gammaproteobacteria across gradients of glacier influence contribute to the microbiome-weighted averages.

Variation in genomic properties of the GFS microbiome along gradients of glacier influence was also explained by within-clade variation. For instance, at higher levels of benthic chlorophyll-a, genomes were larger in size, had higher gene and tRNA numbers, both among Gammaproteobacteria MAGs and all other MAGs (Figure 5.3C). Interestingly, the number of tRNAs was negatively associated with benthic chlorophyll-a in Gammaproteobacteria, but positively in other clades, putatively because of the role of tRNAs in adaptation to fluctuating environmental conditions, which is characteristic of GFSs (Arella et al., 2021).

Furthermore, to assess the importance of genome optimization of the GFS microbiome, we aligned genomic properties with the ecological success of MAGs, defined as high prevalence and/or mean relative abundance. We found positive relationships between genome size and MAG prevalence and relative abundance (Figure 5.4 C & D, whereas coding density was negatively related to these indicators of ecological success (Figure 5.4 E and F). Using GAMs and a Bayes factor analysis, we

tested whether these relationships differed between Gammaproteobacteria and the other classes. A GAM model with separate splines for Gammaproteobacteria and other MAGs was better supported by the data (Bayes factor > 1000 for all comparisons) than a GAM model with one spline for all MAGs (Figure 5.4). This finding indicates that ecologically successful Gammaproteobacteria combine increased coding density with reduced genome size compared to classes that are similarly successful. While this relationship could be in part attributed to the low prevalence of symbiotic Patescibacteria that have particularly small and thus streamlined genomes (Tian et al., 2020), the difference between Gammaproteobacteria MAGs and the others is clear at higher abundance and prevalence (Figure 5.4).

Collectively, our findings suggest that the selective constraints in GFSs alongside the scarcity of autochthonous organic carbon sources (from primary producers), leads to microbiome-level signatures of genome optimization along gradients of glacier influence. The GFS microbiome shifts genomic features in response to this environmental gradient by changes in the abundance of Gammaproteobacteria compared to other clades. This process is augmented by changes in abundance and replacement of members within Gammaproteobacteria as selective constraints change. We deem these findings critical because the deep phylogenetic rooting of these signatures reflects the long-term and putatively consistent nature of this extreme environment, which is now changing at a rapid pace owing to climate change.

Pangenome analysis reveals specific adaptations of Gammaproteobacteria in GFSs

Given the high coding density of Gammaproteobacteria compared to other GFS community members, and their ecological success under high glacier influence in GFSs, we next aimed at identifying genes that may underpin this difference. To this end, we conducted a pangenome analysis of Gammaproteobacteria in GFS to identify a set of core genes. Subsequently, we compared this core pangenome to the core genes of both Gamma- and Alphaproteobacteria as the latter represents the closest sister group in the GFS microbiome, and used as an outgroup could help us understand the early evolution of Gammaproteobacteria in GFSs. This approach allowed us to identify genes

consistently preserved or lost in the GFS Gammaproteobacteria compared to the common ancestors of both Alpha- and Gammaproteobacteria in GFS.

Our analysis uncovered a core pangenome of 830 gene clusters for Gammaproteobacteria, and 548 gene clusters when Alphaproteobacteria MAGs were also included. Analysing KEGG orthologous groups functions (KOs) of gene gains and losses, we identified 309 KOs gained and 273 KOs lost in the GFS Gammaproteobacteria. Both gains and losses were predominantly concentrated in transporters, enzymes, and the two-component system (Figure 5.5). Additionally, gains were observed in secretion systems, lipopolysaccharide biosynthesis, and bacterial motility, while losses were most common in butanoate metabolism, glyoxylate and dicarboxylate metabolism, amino sugar and nucleotide sugar metabolism, and fructose and mannose metabolism. In the context of GFSs, the losses observed in several metabolism pathways could be associated with the scarcity in organic carbon.

By applying a text-mining approach, we identified higher-level functions associated with these gains and losses. Notably, we observed the gain of 20 ATP-binding transporters related to twitching motility and type IV pili, while one KO associated with flagella was lost. While bacteria in biofilms are typically less motile than their free-living counterparts (Guttenplan and Kearns, 2013), twitching motility through type IV pili is well known for its implications in biofilm formation and movement of prokaryotes on surfaces (Burrows, 2012). Moreover, five KOs related to chemotaxis were gained while one was lost, which further suggests a change in motility during the early evolution of Gammaproteobacteria in GFS. We also noted the loss of nine KOs linked to fermentation, particularly in the butanoate metabolism category, and the gain of seven KOs associated with lipids. While the losses in fermentation and especially butanoate metabolism could be attributed to the high oxygenation of the GFS environment, lipids (e.g., lipopolysaccharides) have been associated with the adaptation of membranes and cell walls to cold conditions (De Maayer et al., 2014). While we find KO gains associated with cold adaptation and a biofilm lifestyle, we observe losses that could be adaptive given the association we observed between genome size and glacier influence. Taken together, comparing the pangenome of Gammaproteobacteria to the closest sister clade in GFSs (i.e., Alphaproteobacteria), we infer that early in the evolution of Gammaproteobacteria in GFSs, adaptations related to motility, the biosynthesis of lipids, and to aerobic metabolism arose.

Conclusion

Eco-evolutionary dynamics shape the genomic landscapes of communities and here we unravel, using a phylogeny-rooted analytical framework, signatures of genome optimization in the world's GFSs. Our results indicate several lines of genome optimization in the GFS microbiome, including genome size reduction, modulation of genomic redundancy and optimization of translation along a gradient of glacier influence. Additionally, we show the importance of *Gammaproteobacteria* at shaping community-level genomic features, especially at high glacier influence. Moreover, we highlight particular genomic features that represent potential adaptations to the environmental conditions and biofilms. As glacier influence diminishes owing to climate change, our findings suggest that the genomic landscape of GFSs may change, thereby altering an ancient and well-adapted microbial biodiversity.

❖ *Acknowledgments*

❖ *Consortium*

The Vanishing Glaciers Field Team

Michael Styllas[1], Martina Schön[1], Matteo Tolosano[1], Vincent de Staercke[1], Hannes Peter[1], Tyler Kohler[2] and Tom J. Battin[1]

[1]*River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland*

[2]Department of Ecology, Faculty of Science, Charles University, Prague, Czechia

❖ *Code availability*

The code and data used in this study will be available on the github repository: https://github.com/Mass23/MAGFS. Additionnally, the code for binning is available on this link: https://github.com/michoug/SnakemakeBinning/.
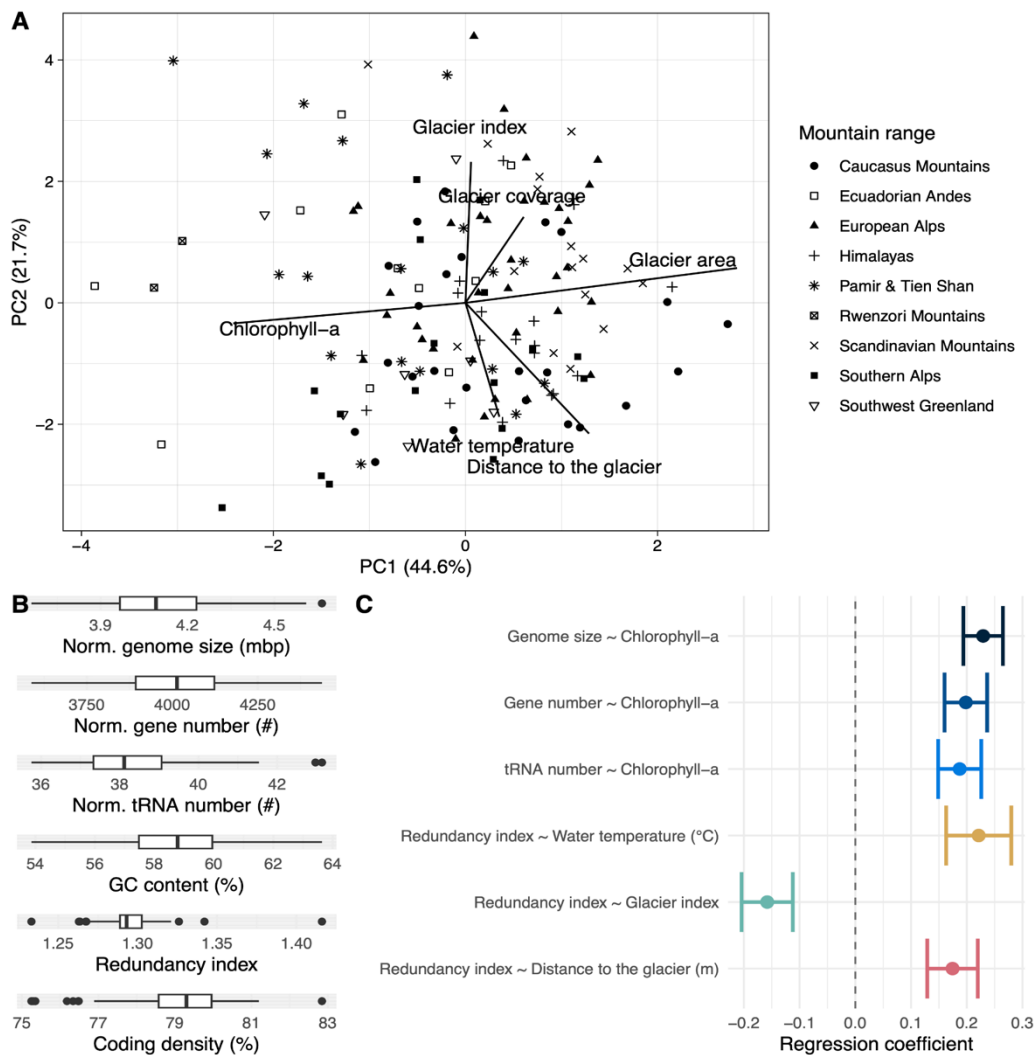
❖ **Figures**

Figure 5. 1 *Dimensions of glacier influence and variation in genomic features.*

(**A**) The first two dimensions of a principal component analysis (PCA) depict associations among key glacier-associated constraints. Symbols represent mountain ranges; arrows depict scores of environmental variables. (**B**) Boxplot showing the distribution of community weighted mean genomic features (i.e., weighted with the relative abundance of MAGs) among GFSs. (**C**) Regression coefficients of the genomic features that correlated significantly with glacial covariates in the generalised additive model (GAM) analysis. GAM models considering spatial variations were fitted adding a linear effect for each pair of genomic features and glaciological variables. Significant relationships after adjusting p-values for multiple testing (Holm's method, $p < 0.05$) are displayed.
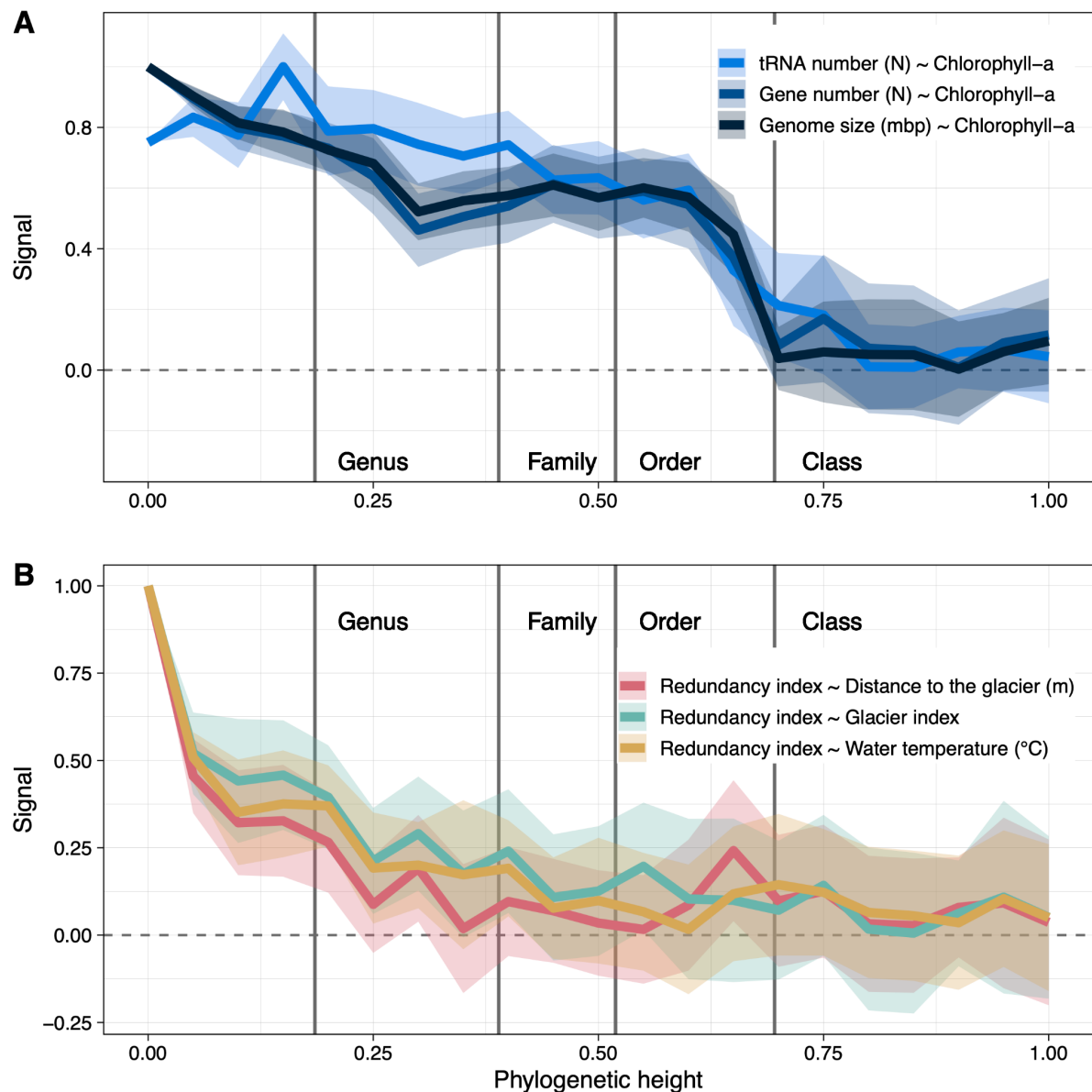
Figure 5. 2 *The signal between community-weighted means of genomic bulk features and glacier influence is structured phylogenetically.*

Line plots displaying the signal in the relationships between redundancy index as response variable and the distance to the glacier, the glacier index and the water temperature as covariates (**A**) and between gene number, genome size and tRNA number and chlorophyll-a as covariate (**B**). The signal was assessed using linear coefficients in the generalised additive models taking spatial variation into account when permuting abundances at various relative phylogenetic heights. Coefficients were normalized by the maximal value for any given glacial covariate-genomic feature pair over the various phylogenetic height values. Shaded areas represent the standard error obtained through 20 null model iterations. Vertical lines indicate median phylogenetic heights for different taxonomic levels and are for visual guidance only.
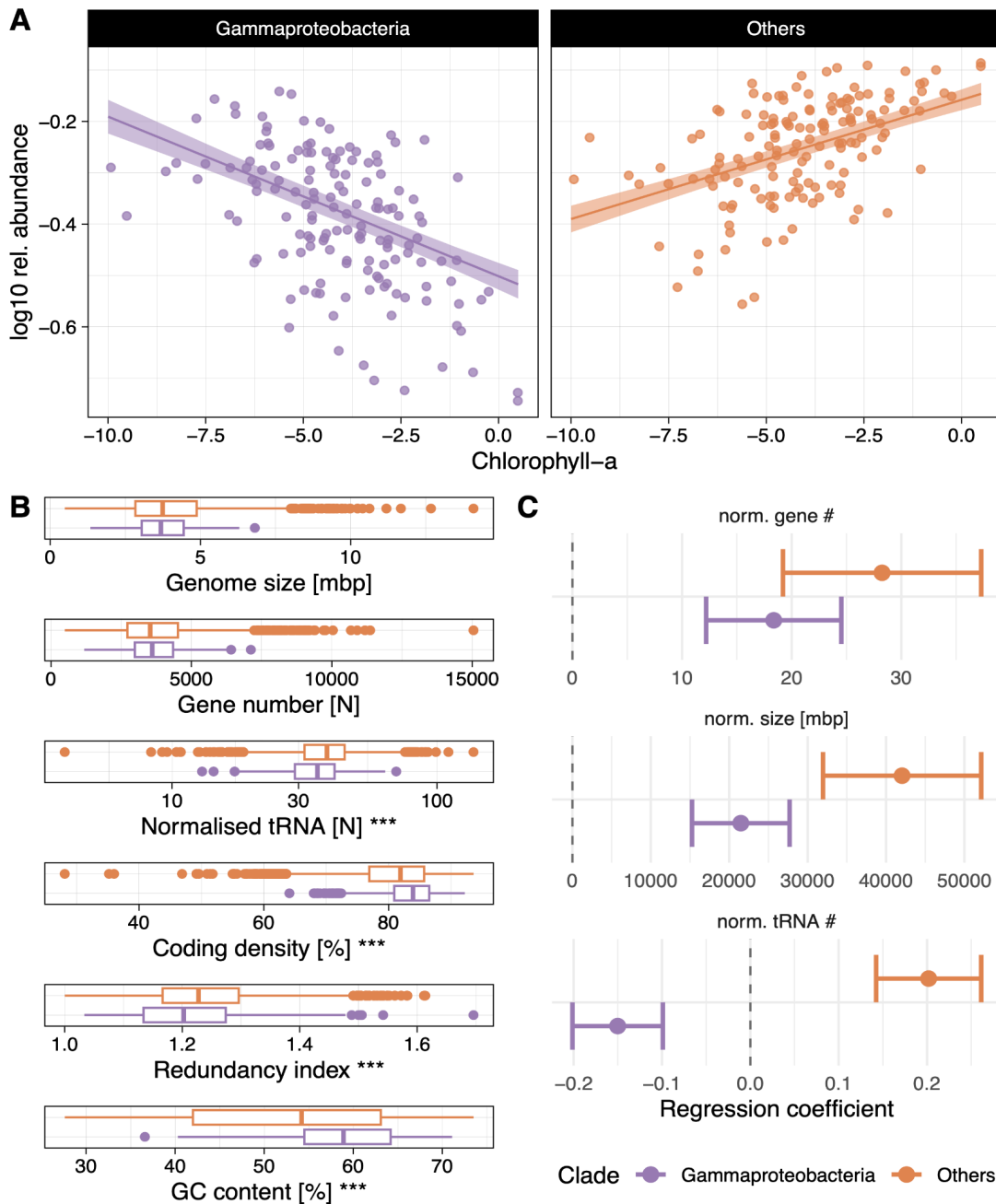
Figure 5. 3 *Gammaproteobacteria drive the variation in genomic features along the gradient of chlorophyll-a*

(**A**) Scatterplot showing the variation in the relative abundance of *Gammaproteobacteria* and all other MAGs along the gradient of benthic chlorophyll-*a* in the world's glacier-fed streams. Lines show linear GAM model fits accounting for large-scale spatial patterns; shaded areas show prediction intervals. (**B**)

Distributions of genomic features for *Gammaproteobacteria* and other MAGs are displayed. Stars denote significance (p<0.01) of Wilcoxon signed rank tests comparing the two groups. (**C**) Linear coefficients representing the variation of genomic feature averages within the *Gammaproteobacteria* (purple) and others (orange), as obtained through the GAM analysis accounting for spatial variation.
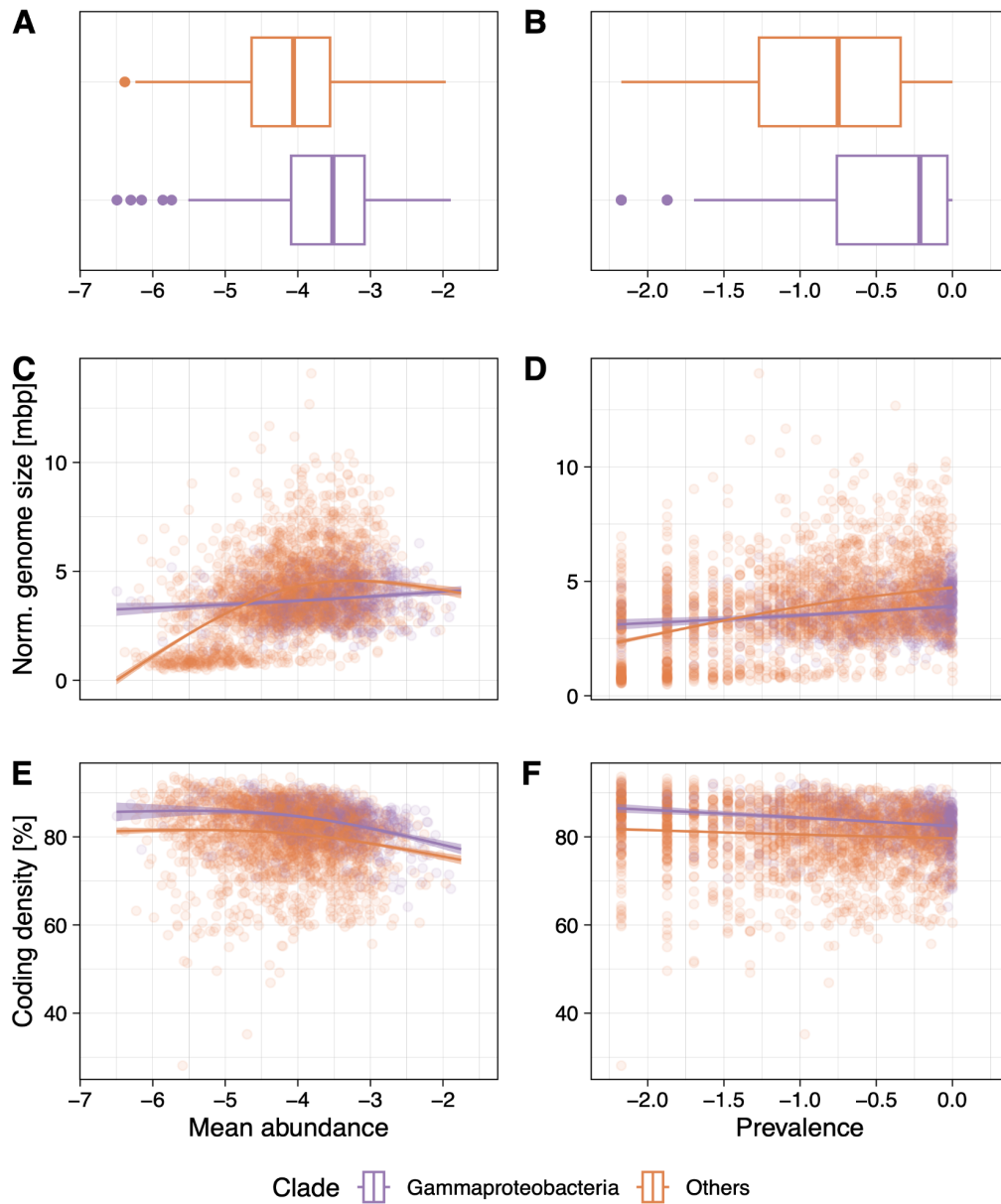
Figure 5. 4 *Gammaproteobacteria MAGs are ecologically successful in GFS and have optimised genomes.*

Comparison of ecological success, measured as relative abundance (**A**) and prevalence (**B**) of MAGs affiliated to *Gammaproteobacteria* (purple) and other classes (orange). Smoothed splines representing the GAM models comparing mean abundance (**C** & **E**) and prevalence (**D** & **F**) with normalised genome

size (**C** & **D**) and coding density (**E** & **F**). Models with separate splines for *Gammaproteobacteria* (purple) and others (orange) MAGs were better supported (Bayes factor > 1,000) than a combined model. While the difference is driven by high abundance and prevalence of large genomes among other classes (or conversely the absence of small genomes at low abundance and prevalence in *Gammaproteobacteria*), for coding density, *Gammaproteobacteria* exhibit increased values across the entire gradient of ecological success.
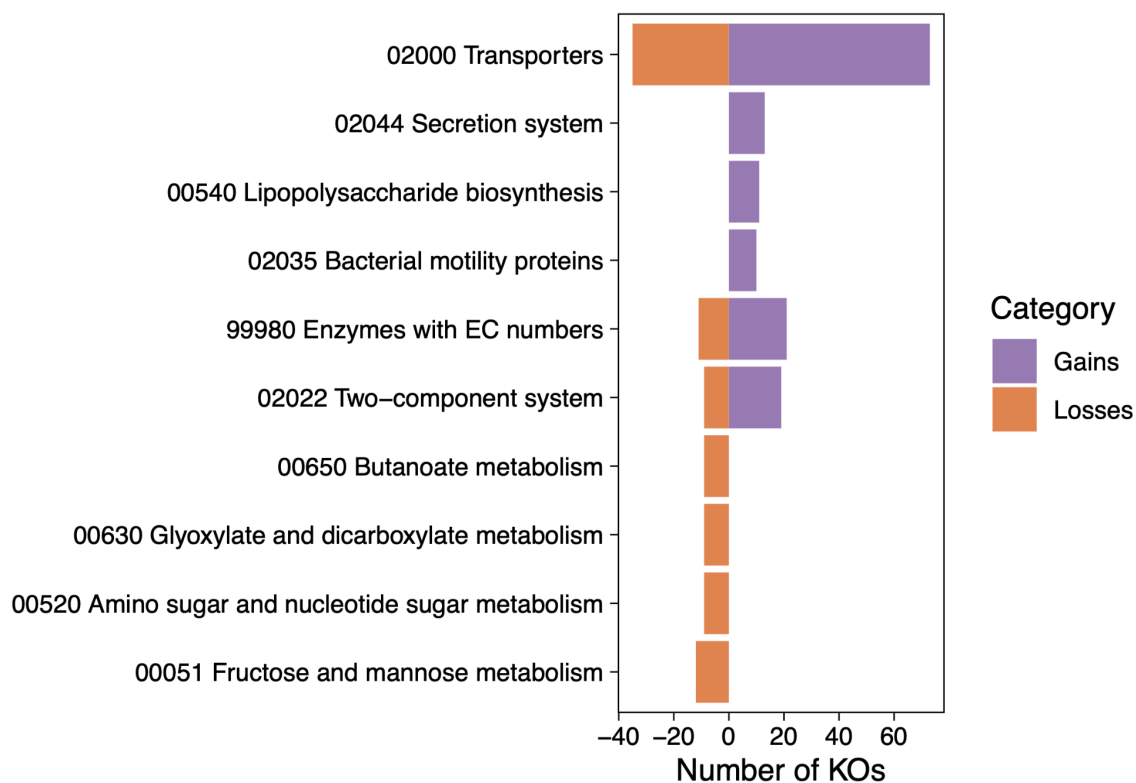
Figure 5. 5 *KO gains and losses of Gammaproteobacteria compared to the closest sister clade (Alphaproteobacteria) in GFS.*

Shown are functional categories that contain the highest number of gene gains and losses in the GFS *Gammaproteobacteria* pangenome analysis. Gains were defined as KOs present in the core genome of *Gammaproteobacteria*, but not in the core genome when *Alphaproteobacteria* MAGs were included. Losses were defined as KOs that were present in the combined core pangenome, but not in the core genome of *Gammaproteobacteria*.

Chapter 6. Discussion

❖   *6.1 Towards a global picture of the microbiome of glacier-fed streams and cryospheric ecosystems*

In chapter 2, we presented a global overview of the microbiome of the cryosphere, encompassing data from ice, snow, terrestrial, freshwater and marine samples. We found that cryospheric ecosystems differ in composition compared to other biomes. Foremost, many bacterial genera affiliated to *Gammaproteobacteria, Alphaproteobacteria, Bacteroidia,* and *Cyanobacteria* were more abundant in the cryosphere compared to other biomes. Some of the most differentially abundant genera were *Sphingomonas, Hymenobacter, Polaromonas,* and *Ferruginibacter*. These taxa are also highly abundant in glacier-fed streams, according to previous work (Ren et al., 2017a; Wilhelm et al., 2014, 2013), and the global sampling of the Vanishing Glaciers Project as seen in chapters 3, 4 and 5. These shared properties of cryospheric ecosystems associated with taxonomy (and thus phylogeny), indeed point towards an association between adaptations to cryospheric conditions and evolutionary history in the bacterial tree of life.

Owing to the large diversity of taxa observed in both cryospheric ecosystems and glacier-fed streams, and given that these taxa are also present in other biomes, the adaptations allowing to thrive in cryospheric conditions seem to be pervasively distributed across the bacterial tree of life, potentially helped by horizontal gene transfers (Dorrell et al., 2021). Thus, to better understand these adaptations, there is the need to account for the evolutionary history by combining phylogenetic data with comparative genomic approaches. In this context, the use of pangenomes has been suggested and successfully applied in microbial ecology (Buck et al., 2022; Eren et al., 2021; Golicz et al., 2020). In chapter 3 we performed a pangenome of the genus *Polaromonas*, comparing the glacier-fed streams MAGs to reference genomes from the RefSeq database, allowing to identify functions that are specific to the strains found in glacier-fed streams. In chapter 5, we performed a pangenome analysis of the glacier-fed streams' *Gammaproteobacteria*, highlighting gene gains and losses that arose early in the evolution of the bacterial class.

In concordance with these shared main community members, we found that the microbiome of cryospheric ecosystems indeed shared functional potential with glacier-fed streams communities. Foremost, genes involved in cell adhesion, quorum sensing, secretion systems and motility all pointed towards the importance of biofilms in both cryospheric ecosystems and glacier-fed streams epilithic biofilms (Chapter 2,

Chapter 3), which is in line with previous reports. This is further developed in glacier-fed streams with the pangenome of *Gammaproteobacteria*, the most abundant bacterial class in glacier-fed streams, that suggested an early change to the type IV pili twitching motility in its evolutionary history in glacier-fed streams; a type of motility that has been associated with biofilms (Burrows, 2012). The network analysis performed on epilithic biofilms in Chapter III suggested cross-domain interactions between bacteria and algae that further highlights the importance of these biofilms. Concordantly, genes involved in the biosynthesis of cobalamin, a compound that has been linked with algal-bacterial relationships, were recovered in the glacier-fed streams metagenomes in Chapter 3 (Amin et al., 2015; Grossman, 2016; Segev et al., 2016).

We also highlighted the importance of chemolithoautotrophy associated genes (e.g, *sox* gene cluster, hydrogen dehydrogenases, etc.) in both cryospheric and glacier-fed stream bacterial communities in chapter 2 and 3. Additionally, we found mixotrophy to be a potentially important adaptation in glacier-fed stream MAGs affiliated to Cyanobacteria, potentially to cope with varying environmental conditions (Stoecker and Lavrentyev, 2018). In these ecosystems with few allochthonous carbon sources, these may indeed represent important adaptations. Nevertheless, these findings are based on metagenomics and thus only consider genomic potential. Moreover, the dataset of 35 cryospheric metagenomes (101 in total) that was compiled for chapter 2, while being global, is limited geographically and contains only a few different habitats. Thus, these findings are potentially not representative of the cryosphere in general.

We also found phylogenetic patterns associated with cryospheric conditions. Analysing phylogenetic metrics of beta diversity, we found that the bacteria composing the microbiome of cryospheric ecosystems are more closely related when compared with other biomes, suggesting niche similarity under the assumption that closely related species share similar functions (Martiny et al., 2013). This indeed suggests potentially shared selective constraints, which is in line with the importance of shared environmental characteristics e.g., low temperatures and oligotrophy. In recent years, the importance of homogeneous selection as been shown to shape the microdiverse microbiome of glacier-fed streams, and has been associated with the extreme conditions of the ecosystem (Fodelianakis et al., 2022). Given the shared environmental conditions across the cryosphere (i.e., low temperature and oligotrophy), the broad overlap in abundant taxa, and the shared functional potential, it

is thus possible that other cryospheric microbiomes could be characterised by similar patterns.

Nevertheless, we also showed how distinct bacterial communities are found throughout different habitats and ecosystems. In chapter 2, we showed that the four mains cryospheric ecosystems types (i.e., ice and snow, terrestrial, freshwater and marine) indeed differ in community composition. In chapter 3, we found that community composition differed between epilithic and epipsammic biofilm communities in glacier-fed streams. Additionally, in chapters 4 and 5, we associated variations in environmental conditions to the abundance of community members. In the context of global climate change, the diversity and the degree of novelty found throughout the cryosphere, including glacier-fed streams, prompts the need to characterise more ecosystems and habitats, as they may harbour analogously unique and diverse communities.

❖ 6.2 Future directions for the microbial ecology of the cryosphere

The metagenomes sequenced for the Vanishing Glacier Project have indeed improved our understanding of glacier-fed streams microbial ecology. Nevertheless, other cryospheric ecosystems and habitats still lack attention. In chapter II, while we compiled globally distributed data, we emphasised the need to urgently characterise more cryospheric ecosystems and their microbial communities. First, the compiled studies were often concentrated in a few locations (e.g., Greenland, Svalbard, etc.). Additionally, we denoted a geographical bias toward polar regions, urging the need to characterise alpine ecosystems. Moreover, the lack of consistency in the sampling methodology made the comparison of environmental parameters impossible.

However, in the meanwhile, some large-scale studies have investigated several cryospheric ecosystems such as the polar arctic ocean (Royo-Llonch et al., 2021), cryoconite holes (Millar et al., 2021), and glaciers for instance (Y. Liu et al., 2022). This recent data could indeed pave the way for new meta-analyses containing more samples, encompassing a wider range of ecosystems, habitats, and geographical locations. Moreover, several of these studies generated metagenome assembled genomes (MAGs) that could be used to characterise more precisely the functional potential of specific taxa (e.g., using a pangenome approach), as opposed to comparing entire metagenomes as in chapter 2 and 3.

While the vanishing glacier project exemplifies how such a global sampling with a consistent methodology can foster our understanding of an ecosystem and its

microbiome, it indeed has some limitations. First, the sampling was performed over the short windows of opportunities in summer, and thus only give a snapshot of the microbiome when the ecosystem is highly productive. Time series data could help understand how the communities change over time in the highly seasonal glacier-fed streams (Boix Canadell et al., 2021; Scotti et al., 2019). This statement is indeed applicable more generally to cryospheric ecosystems in general due to the seasonality of ice and snow accumulation-melt cycles and their importance (Winkel et al., 2022). Additionally, the sampling focused on non-debris covered glaciers while rock glaciers (i.e., glaciers that are covered by rocks) are numerous and harbour distinct stream communities (Tolotti et al., 2020). Furthermore, while metagenomic data unravels the genomic potential of communities, metatranscriptomes could be sequenced to confirm that the genes that we highlight are indeed expressed under natural conditions (Shakya et al., 2019). However, extracting RNA from low-biomass samples is challenging. Moreover, the approach would require a higher number of replicates to obtain meaningful statistical comparisons compared to metagenomes.

The metagenomic approach has additional shortcomings. Foremost, only highly abundant community members are captured by the binning, and are thus included in the analysis of MAGs. In contrast, methods relying on reads and contigs are less precise taxonomically, as only a very small fraction of the genomes is considered. Moreover, MAGs typically are incomplete and contaminated, even more than what the methods used to quantify their quality estimate (Meziti et al., 2021). However, these shortcomings could be partially addressed by the culture of isolates, in combination with whole genome sequencing and phenotyping, used in synergy with metagenomic data (S. Liu et al., 2022; Saheb Kashaf et al., 2022). Nevertheless, metagenomic data represented an important first step, and wss indeed highly informative as MAGs capture the functional potential of communities better than isolates (Albright and Louca, 2023).

❖ 6.3 Functional adaptations to the extreme cryospheric conditions and psychrophiles

Exploring the functional potential of glacier-fed streams and cryospheric ecosystems, we identified genes that have previously been linked with cold-adaptation. In chapter II, we found that cryospheric metagenomes were enriched in chitinase genes that were previously associated with freezing in plants (Ahmed et al., 2012; Gupta and

Deswal, 2014) and upregulated in the cold-adapted Chlamydomonas algae under freezing treatment (Liu et al., 2016). Moreover, we highlighted the importance of biofilm formation in the cryospheric metagenomes as an adaptive strategy that has previously been linked with low temperature environments (Smith et al., 2016). In chapter III, we also reported the abundance of several psychrophilic adaptation genes such as cold-shock proteins in glacier-fed streams metagenomes, and highlighted some of these functions in the pangenome of *Polaromonas* MAGs.

Nevertheless, we identified many other KOs and gene cluster that were abundant in cryospheric ecosystems, and/or glacier-fed streams and that had unknown function, or functions that had not yet been associated with cryospheric conditions. While, some of these genes could represent cold-adaptations, they may as well be linked to other environmental conditions. Low temperature is not the only shared environmental constraint of cryospheric ecosystems, and in natural communities disentangling the effect of various parameters that in addition correlated (e.g., temperature and oligotrophy) is challenging. Thus, further investigations are required to characterise these genes and functions. Combined with the culture of isolates and single-cell technologies, the characterisation of this genomic potential could lead to the discovery of new compounds and enzymes. For example, such an approach has been applied to the biosynthetic potential of the global ocean microbiome: combining genomic potential data with cultivated isolates, unusual bioactive compounds were unravelled (Paoli et al., 2022). Given the large and still understudied biodiversity of cryospheric ecosystems (as shown in chapter 2), using such an approach could foster biotechnology allowing to identify useful molecules whose properties function at cold temperatures (Feller and Gerday, 2003).

❖   6.4 Genomic bulk features are associated with glacier influence in the cryosphere and glacier-fed streams

Since the discovery of DNA and the beginning of the study of genomes, variations in genome size and other genomic bulk features (e.g., GC content, tRNAs, etc.) have fascinated biologists. In microbial ecology, the optimisation of genomic bulk features as an adaptive trait has thus naturally been investigated, especially in high-temperature environments as these particular conditions influence molecular mechanisms. In this context, psychrophiles' genomes were also investigated as cold temperatures are indeed expected to affect the flexibility of molecules. Moreover, more recently and thanks to the advent of metagenomics, oligotrophy has been associated with reduced genome size in several ecosystems (Giovannoni et al., 2014). Given the combination of oligotrophy and low temperatures in cryospheric ecosystems, we thus investigated

these patterns in chapter 2. Moreover, as glacier-fed streams represent a longitudinal gradient of glacier influence (that is also associated with oligotrophy and water temperature), we investigated for patterns in genomic bulk features in chapter 4 and 5.

In chapter 2, 4 and 5, we identified several associations between genomic bulk features and cryospheric conditions. First, we found that the genera that are more abundant in the cryosphere compared to other biomes showed an increased GC content based on their representatives in the RefSeq database (Pruitt et al., 2007). However, this database was primarily constituted of cultivated isolates at the time. Given the importance of habitats and the environment at shaping genomes, the analysis of MAGs data from these ecosystems should be carried away to confirm these patterns, and the gradient of glacier influence in glacier-fed streams allowed us to test the variation in bulk features in the Vanishing Glaciers Project dataset.

In chapter 4, we found that the strains that are forecasted to decrease in abundance under future scenarios of climate change showed distinct values in terms of genomic bulk features with smaller genomes that are functionally less redundant. These findings would be in line with the hypothesis that at higher glacier influence (i.e., under current environmental conditions), the pronounced oligotrophy selects for smaller genomes, while the fluctuating conditions driven by biofilms being less developed due to the harsh conditions drives metabolic flexibility. This is further supported by the analyses performed in chapter 5, linking the functional redundancy with water temperature, and the optimisation of genome size and other genomic bulk features with the gradient of chlorophyll-*a*. If indeed these patterns represent adaptations to oligotrophic and cold conditions, they could thus be present in other cryospheric ecosystems.

However, taking into account phylogenetic signal when investigating genomic bulk features is crucial (Martinez-Gutierrez and Aylward, 2022). To this end, we created an approach that allows to identify clades that shape the genomic landscape across a gradient of environmental parameter in chapter 5. This approach could indeed be used in other ecosystems to identify the forces driving variations in genomic bulk features. In our analysis, the method revealed the deep-branching importance of the class *Gammaproteobacteria* at shaping community averages of genome size, gene number and tRNA number. This prompted us to analyses the MAGs of the class in a comparative genomic approach, and allowed to identify adaptations that arose early in the

evolutionary history of the clade. *Gammaproteobacteria* is also highly successful in many other cryospheric ecosystems (Keuschnig et al., 2023; Li et al., 2019; Royo-Llonch et al., 2021), thus the adaptations that we find potentially apply to other representatives from other ecosystems. Further work could compare glacier-fed stream MAGs with genomes coming from other cryospheric ecosystems or even biomes, with the aim to better understand these adaptations and their link with the environment.

This is especially relevant in the context of climate change, as the diversity of the class is potentially at threat. *Gammaproteobacteria* is predicted to reduce in relative abundance as shown in Chapter 5 with the positive relationship between the clade relative abundance and glacier influence. Additionally, in Chapter 4, we found that *Gammaproteobacteria* is among the taxa that will increase in absolute abundance the least in glacier-fed streams under future scenarios of climate change. Moreover, the monophyletic clades highlighted to be at particular threat in this chapter include several representatives of the class that were affiliated to bacterial genera such as *Ferruginibacter, Polaromonas* and *Methylotenera*.

❖ 6.5 Predictive approach for climate change microbial ecology

In Chapter 4, we showed how the use of a predictive approach in a strain distribution modelling framework can be applied to bacterial communities, and allows to forecast future ecological shifts. As the quality of the model is strain dependent, some strains abundance patterns are more predictable than other based on sample size, or for biological reasons. Thus, for some strains, the output of our predictions was unreliable and they were not considered in further analyses. However, we found that for most of them, the signal captured by the model is informative, and taken together allows to draw general conclusions (e.g., differences in predicted changes across taxa). Thus, we believe such an approach could be applied to other ecosystems in order to better understand how bacterial communities will react to climate change.

Nevertheless, we have to acknowledge that in this regard, glacier-fed streams represent a good study system to apply such a modelling framework. First, the importance of the harsh environmental conditions at shaping the microbiome maybe lessens the importance of biotic interactions, that are notably hard to take into account in species distribution models (Wisz et al., 2013). To account for this in chapter 4, the chlorophyll-*a* content of the sediments was used as a possible covariate in the strain distribution models. However, to better capture these biotic interactions, and thus improve the accuracy of the models, including the abundance of the various eukaryotic taxa found in the metagenomes, or the 18s dataset could be used. However, projecting the abundance of these taxa onto future scenarios of climate change, which would be

required to use them as covariates for the strain distribution models, could prove challenging.

Another reason explaining the satisfactory ability of our models to predict the environmental parameters and the abundance of strains is that previous knowledge on the ecosystem was available and allowed to tailor the modelling framework to the question. For instance, the set of variables that was selected for chapter 4 encompasses many environmental parameters that were already shown to be important for bacterial community composition (e.g., pH and conductivity). Moreover, a previous understanding of the impact of glacier influence on the ecosystem guided the modelling framework: first modelling the effect of glacier influence on the environmental template, and in a second step investigating how these covariates are associated with the bacterial communities. However, while based on cross-validation metrics we obtained satisfactory accuracies for our models, these projections rely only on this one dataset, and no external validation is available. Additionally, while some testing led to the use of stacked generalised additive models for chapter 5, indeed other methodologies (e.g., random forests, deep learning) could be tested in the future to improve these predictions.

❖ 6.6 The greening of glacier-fed streams

While previous conceptual models based on empirical data forecasted the greening of glacier-fed streams, in Chapter IV we created quantitative predictions over future scenarios of climate change. While confirming the link between benthic chlorophyll-a and turbidity through the importance of glacier surface area (and thus glacial weathering), we additionally linked these changes to ecological shifts in the bacterial communities. The main result was an overall increase in the abundance of most strains which was corroborated by our future forecasts of bacterial abundance (i.e., the number of bacterial cells in one gram of sediment), and a previous meta-analysis (Cauvy-Fraunié and Dangles, 2019). Moreover, relying on the coherence of projected changes in abundance over phylogenetic and taxonomic clades, we were able to conclude trends for the main taxa that have many representatives. These ecological shifts were also supported by the lower relative abundance of *Gammaproteobacteria* found at low glacier influence in chapter 5. Moreover, by modelling α-diversity metrics,

we further linked these future projections with changes in the phylogenetic structure of the microbiome.

While under present conditions these communities are characterised by microdiverse clades that are prevalent, we predicted less closely related community members under future scenarios of climate change. Additionally, we showed that the genome of the strains projected to decrease in abundance have different properties than the others, including reduced genome sizes. This is further developed in chapter 5, where differences in genomic bulk features that are consistent with these findings are shown along the gradient of chlorophyll-*a*. While this pattern is mostly drivem by the abundance of *Gammaproteobacteria*, we also find that a significant amount of signal arises by variation at high phylogenetic resolution (i.e., involving the turnover of closely related strains).

This is indeed interesting as the bacterial communities of glacier-fed streams are characterised by microdiversity. Previous work demonstrated this property using patterns of occurrences and abundance with amplicon data. However, we currently lack an understanding of how these patterns relate to the functional potential, and thus how the various representatives of the microdiverse clades are able to colonise various niches. Typically, microdiversity is the occupance of distinct niches by closely related representatives of a clade, that undergo functional differentiation. In the variable conditions of glacier-fed streams, temporal variation in selection pressure for most traits is to be expected, but it remains unclear how does the representatives of the microdiverse clades diversify. The reduction in genome size associated with glacier influence that we observed at high phylogenetic resolution could hypothetically be associated with microdiversity patterns: the strains that lose different genes owing to the selection pressure to reduce genomes and the variable conditions would then be adapted to distinct niches.

However, we currently have not investigated these variations in genome size at high phylogenetic resolution. Moreover, more MAGs that are closely related would be needed to test hypotheses related to microdiversity. Thus, an increased sequencing effort at small spatial and temporal scales would be required to obtain enough high-quality MAGs. Additionally, the presence of many closely related sequences in the samples would be challenging for binning, and thus the use of long read sequencing technologies might be needed.

❖ *6.7 Bacterial diversity and biodiversity loss*

Owing to fast mutation rates in prokaryotic populations, the unclear distribution of taxa, and the lack of taxonomic resolution considered meaningful, the concept of climate-induced biodiversity loss for bacteria is still debated (Thaler, 2021). While in the past, the idea that "everything is everywhere" was associated with the apparent high dispersal capability of bacteria under the neutral theory of ecological selection (O'Malley, 2007). In the past decades, the advent of sequencing technologies and metagenomics fostered our understanding of bacterial biodiversity, unravelling genomic diversity across biomes (Hug et al., 2016; Nayfach et al., 2020). While we still haven't captured the entirety of the Earth' bacterial diversity, empirical data support the idea of discrete taxonomic units. Comparing the similarity of genomes based on nucleotide identity for instance, metagenomic data supports the existence of distinct species (at a 95% average nucleotide identity) (Caro-Quintero and Konstantinidis, 2012), and even more recently strains (at ~99.5%) (Rodriguez-R et al., 2023). However, due to the plasticity of bacterial genomes, the importance of horizontal genes transfers, and the fast rates of evolutionary processes, it remains arguable whether these are meaningful metrics to consider, as an apparent high nucleotide identity could be observed on only a small fraction of the genome (Retchless and Lawrence, 2010, 2007).

Indeed, when comparing bacterial genomes, and this is the case in the pangenomes generated in chapters 3 and 5, gene content across species was highly variable. Even in genomes closely related based on the phylogeny of marker genes, the variation in gene content can be high (Golicz et al., 2020; Horesh et al., 2021; Schubert et al., 2009). This has been shown within species and has been used as an argument to consider only strain-level taxonomy relevant to assess functional potential in the human gut microbiome for instance (Zhu et al., 2015). The proportion of shared genes among representatives of the same taxonomic unit is indeed highly associated with the organisms' lifestyle and ecological conditions (Golicz et al., 2020). Thus, data based on marker genes (such as 16s rRNA amplicon) without pangenome analyses, while allowing to capture phylogenetic signal, does not allow to characterise all aspects of bacterial biodiversity.

Thus, even though the taxa that are abundant in the cryosphere are also present in other biomes, it is likely that biodiversity defined as unique genomic potential, might be

at threat in the microbiome of cryospheric ecosystems. In chapter 2, comparing cryospheric metagenomes to other biome, we found unknown genomic potential that is unique to the cryosphere. The pangenome of *Polaromonas* created for chapter 3 highlighted the phylogenetic and functional novelty of the glacier-fed streams representatives, compared to their database counterparts. This uniqueness of cold-adapted molecules is exemplified by Bowman (2017), that identified proteins unique to psychrophilic genomes. Even though the fast mutation rates of prokaryotes might replace cryosphere-adapted species as glaciers shrink and the associated ecosystems change, unique biodiversity is at threat. This is especially relevant for biotechnology (De Maayer et al., 2014), as some of these molecules possess adaptations to function under cold conditions, but also the exploration of life in a planetary context (Merino et al., 2019). Thus, it is urgent to characterise the microbiome of the cryosphere, and large-scale meta-analyses comparing the functional potential to other biomes, as performed in chapter 2, could help identify such functions. Subsequently, laboratory experiments would be required to characterise these molecules.

❖ *6.7 Conclusion*

Owing to human-induced climate change, cryospheric ecosystems are rapidly shrinking, and thus there is an urgent need to study the microbiome of these endangered ecosystems. Here we analysed global datasets of cryospheric and glacier-fed stream metagenomes, and unveiled diverse, well-adapted bacterial communities that thrive through the diversification of energy pathways, the formation of biofilms, and other genomic adaptations to these extreme conditions. We identified taxonomic, functional and phylogenetic properties of these microbiomes, highlighting the high degree of novelty that they harbour but also the unknowns that remain. Additionally, we used modelling to better understand how glacier influence shapes the environmental template and the microbiome of glacier-fed streams. These models allowed us to forecast future changes in the environment (i.e., "the greening"), and ecological shifts that are associated with the phylogenetic structure of the communities. Finally, we identified the importance of the ecologically successful bacterial class *Gammaproteobacteria* at shaping the genomic landscape of glacier-fed streams over the gradient of glacier influence. This association further highlighted the need to better characterise the microbiome of the melting cryosphere in order to identify potential biodiversity losses and improve our predictions of changes in the functioning of ecosystems.

Appendices

❖ *Supplementary figures*
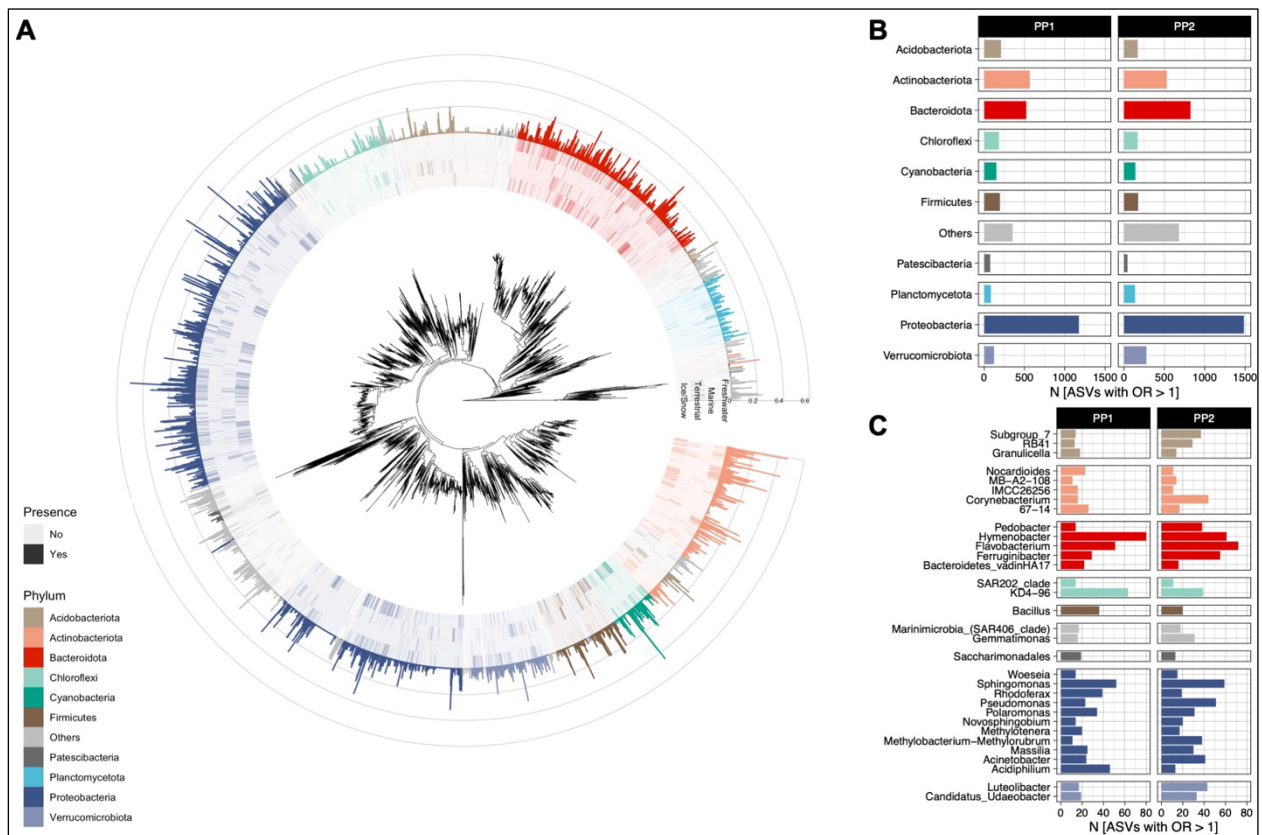
Supplementary figure 2. 1

A) Phylogenetic tree based on the abundant ASVs (relative abundance > 0.5%) found in PP2, demonstrating the cryosphere microbiome diversity across the bacterial tree of life. The highlighted colors represent the ASVs detected in cryospheric ecosystems, and the barplot represents the coefficient for the logistic classification analysis. The number of ASVs at **B)** the phylum-level, and **C)** genus-level taxonomy with an odds-ratio greater than 1 in the logistic classification is shown for PP1 and PP2. Only taxa with the highest numbers are shown.
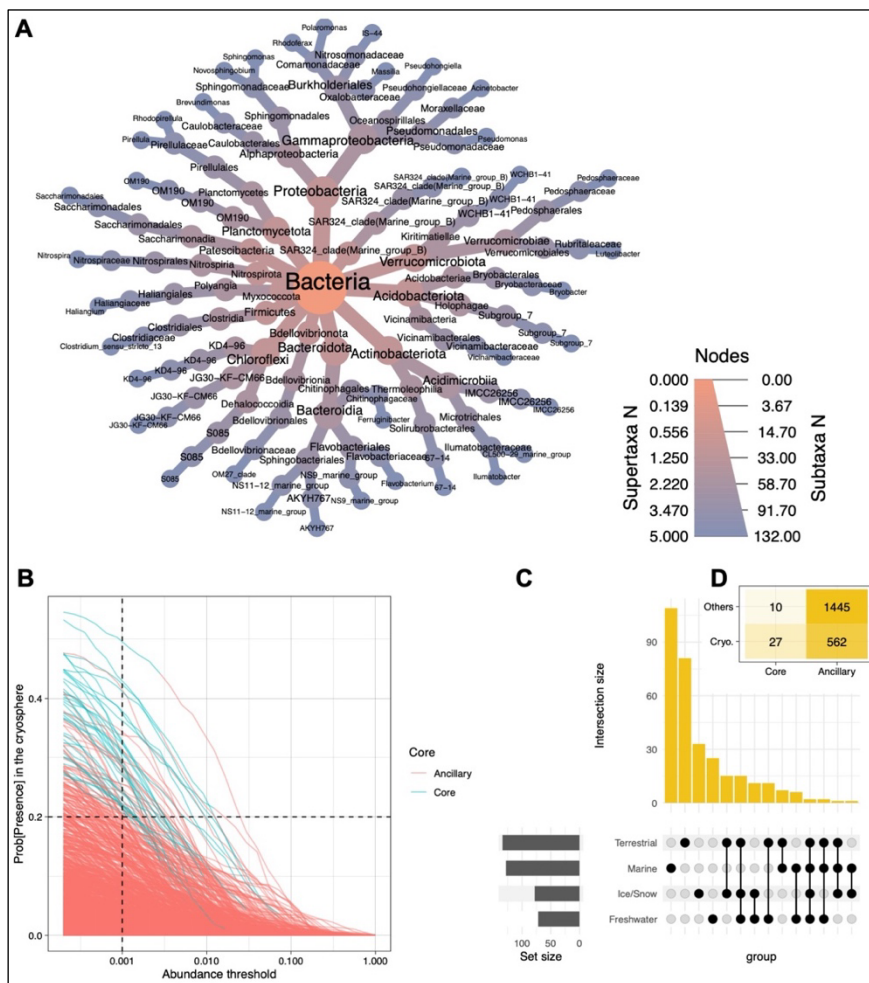
Supplementary figure 2. 2

A) Heat-tree showing the taxonomic classification of the 37 bacterial genera representing the core microbiome of the cryosphere (out of the 2044 total bacterial genera). They represent all bacterial genera with a probability of presence of 20% (calculated in the binomial model analysis, abundance threshold of 0.1%) in the cryosphere, and present in all four ecosystem types. **B)** Line graphs depicting the probability of presence of a given bacterial genus in the cryosphere and their respective abundance, to identify the 'core' (blue) and 'ancillary' (red) genera. The dashed lines represent the chosen thresholds for the core microbiome definition. **C)** Upset plot showing the overlap across the ecosystem types core microbiome, defined at the genus-level, with a prevalence of 20% at an abundance threshold of 0.1% relative abundance. **(D)** Heat plot showing the number of genera per group, highlighting the large overlap between the core microbiome and the cryospheric genera.
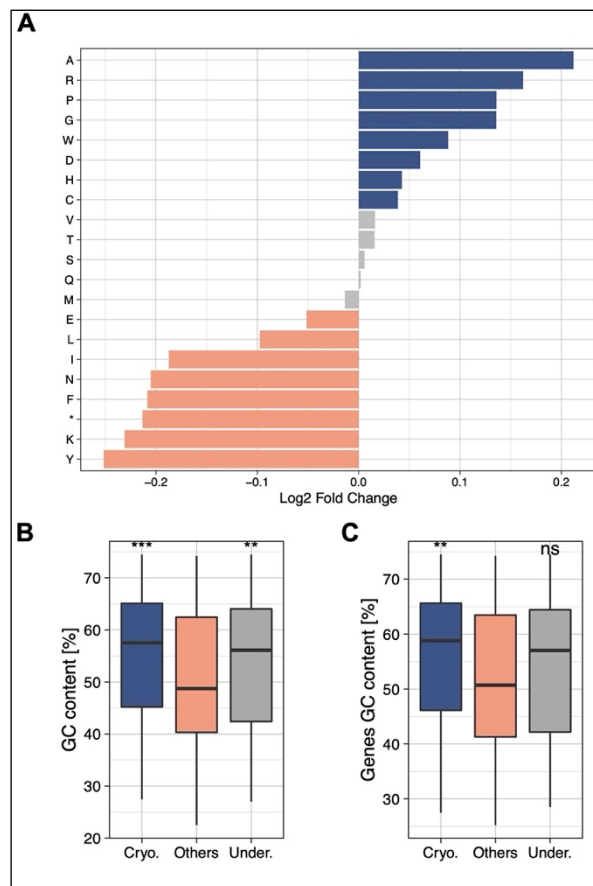
Supplementary figure 2. 3

A) The enriched abundance of aminoacids in the cryosphere (blue; positive fold change) compared to those found in the non-cryospheric ecosystems are depicted. The "*" represents stop codons. **B)** The overall GC% of the genomes belonging to the cryospheric, others and underrepresented genera are depicted (n = 660 total bacterial genera, 197 of which are cryospheric genera, 198 are underrepresented in the cryosphere). **C)** The GC% of the genes predicted in the genomes belonging to the cryospheric, others and underrepresented genera are depicted. The median, 25% and 75% quartiles are represented in the boxplots. Two-sided Wilcoxon tests were performed to assess significance in panels B and C; the Holm method was used to correct for multiple testing (***: 0-0.001, **: 0.001-0.01). Boxplots depict the median and the 25th and 75th quartiles, whiskers extend to values within 1.5 times the interquartile range, and the remaining points are outliers. Exact p-values and medians are listed in Supplementary Table 6.
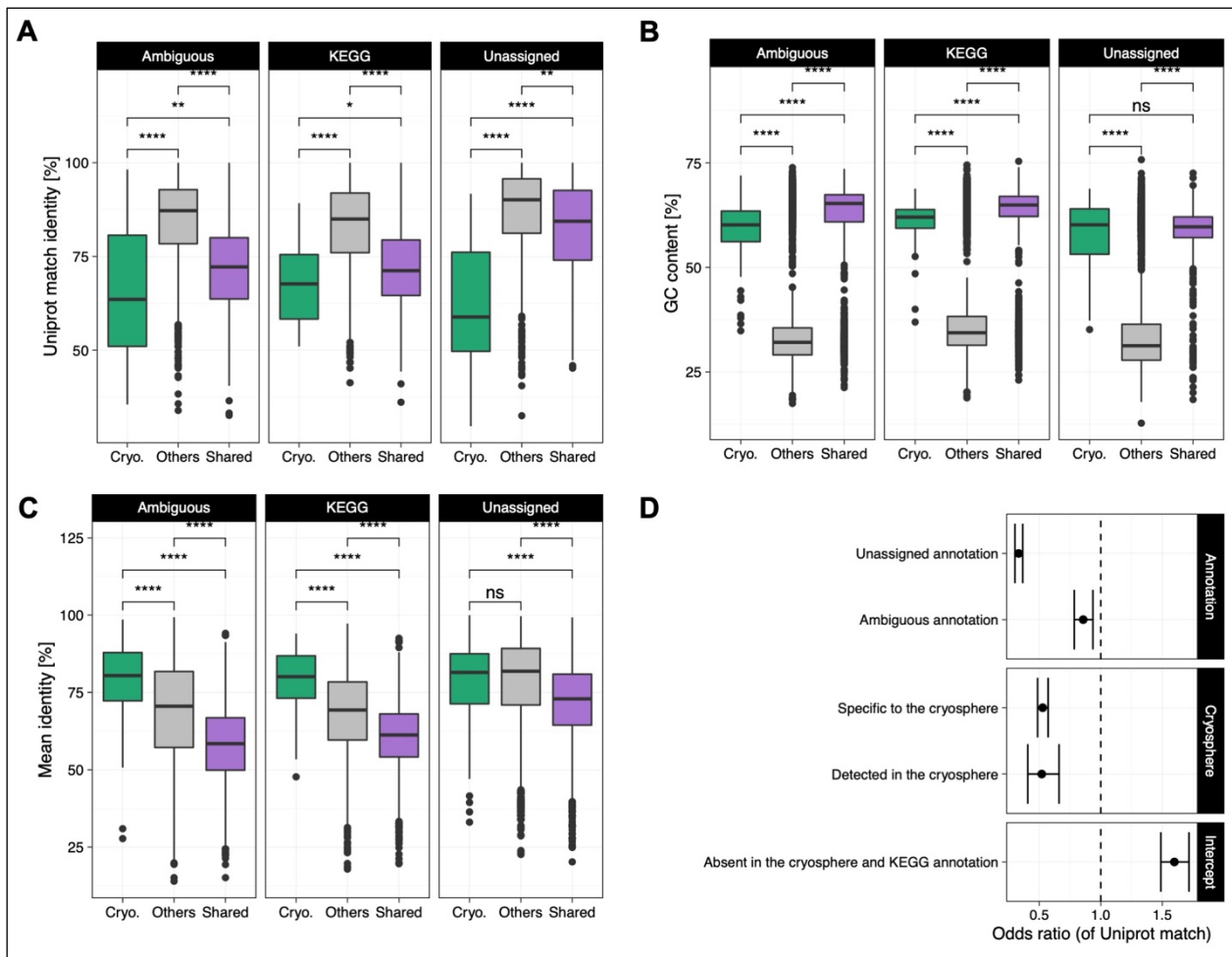
Supplementary figure 2. 4

A) Boxplots indicate the overall identity percentage for representative sequences for each of the gene clusters that matched with UniProt sequences. The figures demonstrate these values for the Cryosphere-only genes, non-cryosphere (Others) and those 'shared' between the two habitats. **B)** The GC content % of all the genes within each cluster is shown. **C)** The pairwise identity of all sequences within each cluster is represented on the boxplots. **D)** Odds-ratio estimations of the UniProt matches with respect to the annotation level, and the presence/or not in the cryospheric metagenomes. Two-sided Wilcoxon tests were performed to assess significance in panels A, B and C; the Holm method was used to correct for multiple testing (****: 0-0.0001, ***: 0.0001-0.001, **: 0.001-0.01, *: 0.01-0.05). Boxplots depict the median and the 25th and 75th quartiles, whiskers extend to values within 1.5 times the interquartile range, and the remaining points are outliers. Sample sizes are listed in Table 1. The exact p-values and test statistics are available in the Supplementary Table 8.
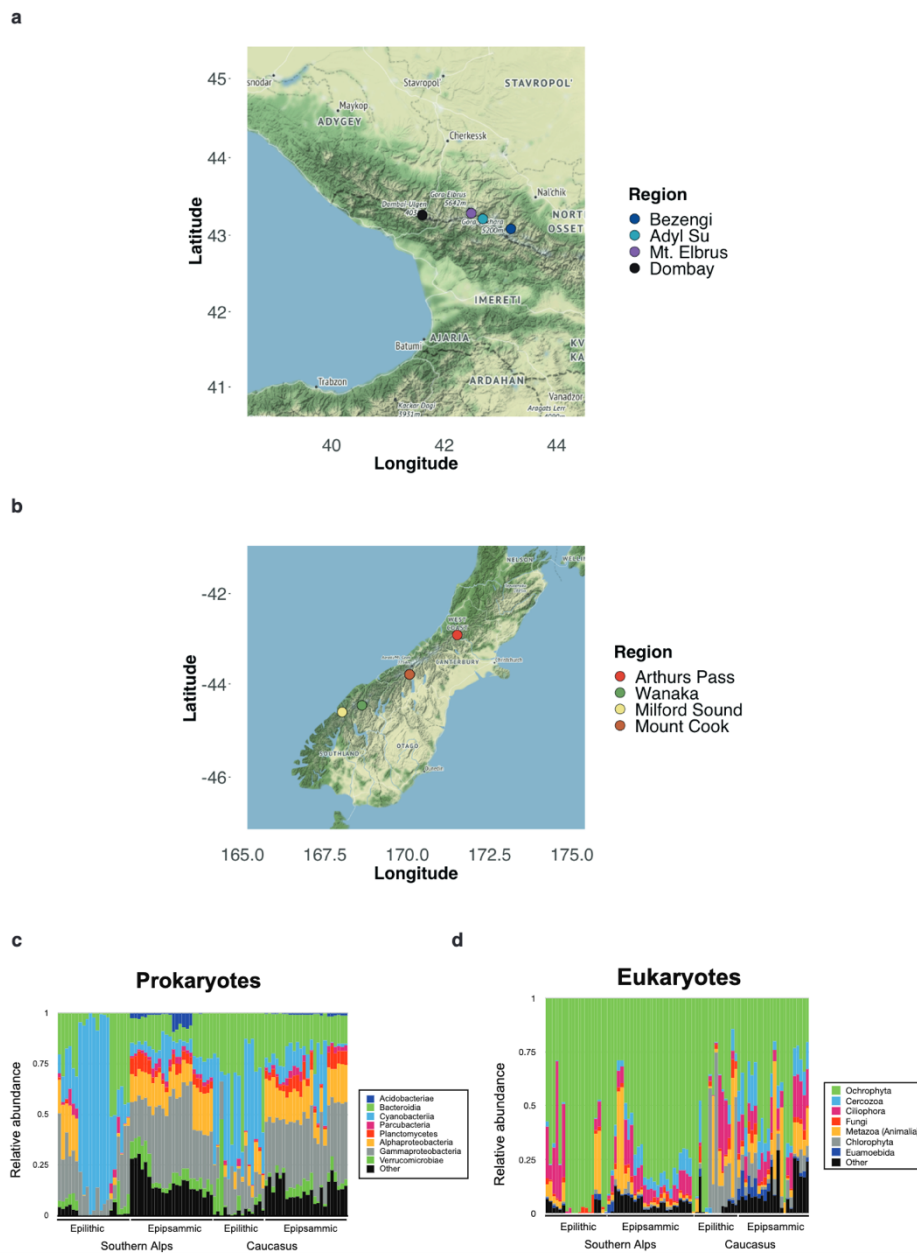
Supplementary figures 3. 1

Glacier-fed streams from where epilithic and epipsammic biofilms were sampled. Regions indicating the collection sites for the epilithic and epipsammic biofilms from (**a**) Caucasus and (**b**) Southern Alps. Relative abundance of prokaryotes (**c**) and eukaryotes (**d**) at the phylum and subdomain levels based on the sequencing of the 16S and 18S rRNA genes, respectively.

a

b
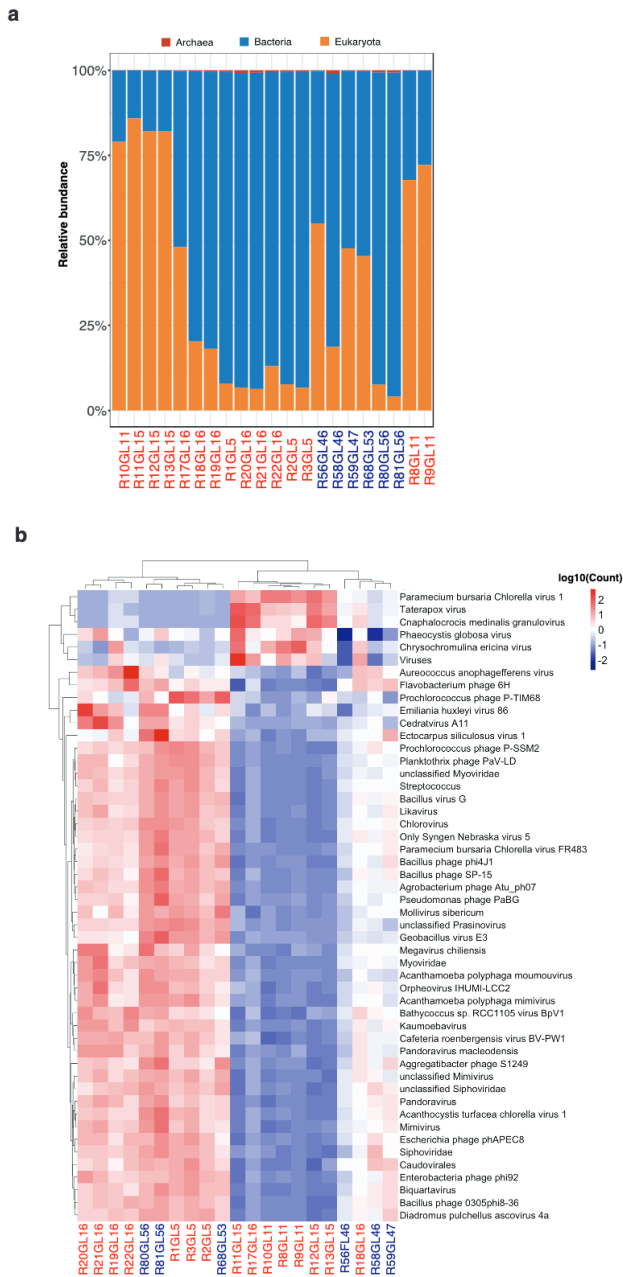
Supplementary figures 3. 2

Epilithic biofilm metagenomic profiles. (**a**) Relative abundance profiles across the three domains of life: archaea, bacteria and eukaryotes in the epilithic biofilms, obtained from the sample metagenomes. Samples from the Southern Alps are indicated in red, while those from Caucasus are shown in blue. (**b**) Virome profile indicating the top 50 viruses. Scaled abundance from low (-2) to high (2) is indicated in the heatmap.

Prokaryotes
- AB1-6_NA
- Acidobacteriota_NA
- Actinobacteriota_UBA10887
- Actinobacteriota_Yonghaparkia
- Bacteroidota_ELB16-189
- Bacteroidota_Emticicia
- Bacteroidota_Haliscomenobacter
- Bacteroidota_Hymenobacter
- Bacteroidota_JJ008
- Bacteroidota_Pedobacter
- Bacteroidota_Spirosoma
- Bacteroidota_UBA1930
- Bdellovibrionota_Bacteriovorax
- Bdellovibrionota_Ga0074139
- Crenarchaeota_NA
- Dependentiae_NA
- Gemmatimonadota_Gemmatimonas
- Myxococcota_NA
- Nanoarchaeota_NA
- Patescibacteria_Aalborg-AAW-1
- Patescibacteria_OLB19
- Patescibacteria_UBA11704
- Patescibacteria_UBA1547
- Planctomycetota_BOG-1363
- Planctomycetota_Fimbriiglobus
- Planctomycetota_Planctomyces_A
- Proteobacteria_AAP99
- Proteobacteria_Aquincola
- Proteobacteria_Brachymonas
- Proteobacteria_Brevundimonas
- Proteobacteria_Flavimaricola
- Proteobacteria_Ga0077545
- Proteobacteria_GCA-2402195
- Proteobacteria_Limnohabitans
- Proteobacteria_Lysobacter_A
- Proteobacteria_Methylotenera_A
- Proteobacteria_Novosphingobium
- Proteobacteria_Phenylobacterium
- Proteobacteria_Polaromonas
- Proteobacteria_Rhodoferax
- Proteobacteria_Sphingomonas_A
- Proteobacteria_UKL13-2
- Proteobacteria_Undibacterium
- Spirochaetota_NA
- Verrucomicrobiota_NA

Eukaryotes
- Chytridiomycetes
- Ochromonas
- Ochromonas
- Ochromonas

Supplementary figures 3. 3

Cross-domain interactions and adaptations of epilithic biofilms. (**a**) Corrplot based on Spearman's correlation between pro- and eukaryotic MAGs aggregated at the phylum level. (**b**) Co-occurrence network of all MAGs across the Southern Alps in New Zealand and Caucasus in Russia. Each node

represents a MAG, while the size represents the degree centrality. The edges represent the positive coefficient of co-occurrence along with the corresponding betweenness centrality between the MAGs. Unconnected nodes represent MAGs with lower betweenness (< 0.5) compared to other MAGs. The color of the nodes represents the individual taxa, while the lines represent the edges connecting the nodes. The thickness of the lines indicates those edges with a betweenness greater than 0.5. Co-occurrence network constructed from pro- and eukaryotic MAGs found in (**c**) the Southern Alps (New Zealand) and (**d**) the Caucasus. The largest connected component of the co-occurrence network from (**e**) the Southern Alps (New Zealand) and (**f**) Caucasus GFSs are depicted.
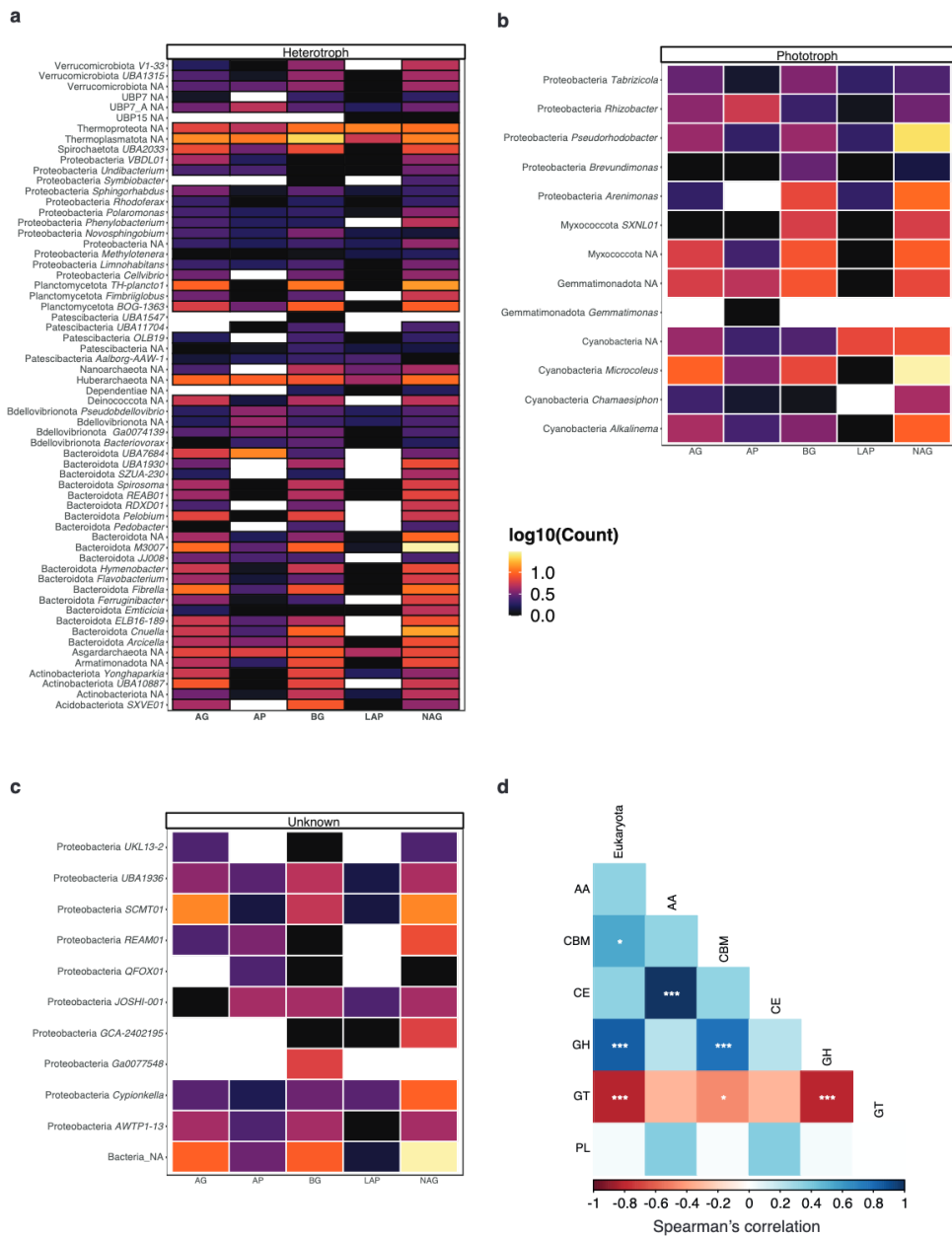
Supplementary figures 3. 4

Extracellular enzyme genes based on lifestyle. The classification at phylum and genus levels of MAGs identified as (**a**) heterotrophs, (**b**) phototrophs, or (**c**) those with 'unknown' trophic metabolisms are depicted, showing the abundance of genes encoding for extracellular enzymes. NA: unclassified genus; AG: α-1,4-glucosidase; BG: β-1,4-glucosidase; LAP: leucine aminopeptidase; NAG: β-1,4-N-acetylglucosaminidase; AP: acid (alkaline) phosphatase. (**d**) (**c**) Spearman's correlation analyses of overall eukaryote relative abundances with the CAZyme abundances. CAZymes include AA: auxilliary activities, CBM: carbohydrate-binding module , CE: carbohydrate esterases , GH: glycoside hydrolases,

GT: glycosyltransferases, PL: polysaccharide lyases. FDR-adjusted p-values were estimated using the 'cor.mtest' function from the corrplot R package and are indicated by *, i.e., * < 0.05, ** < 0.01, *** < 0.001.
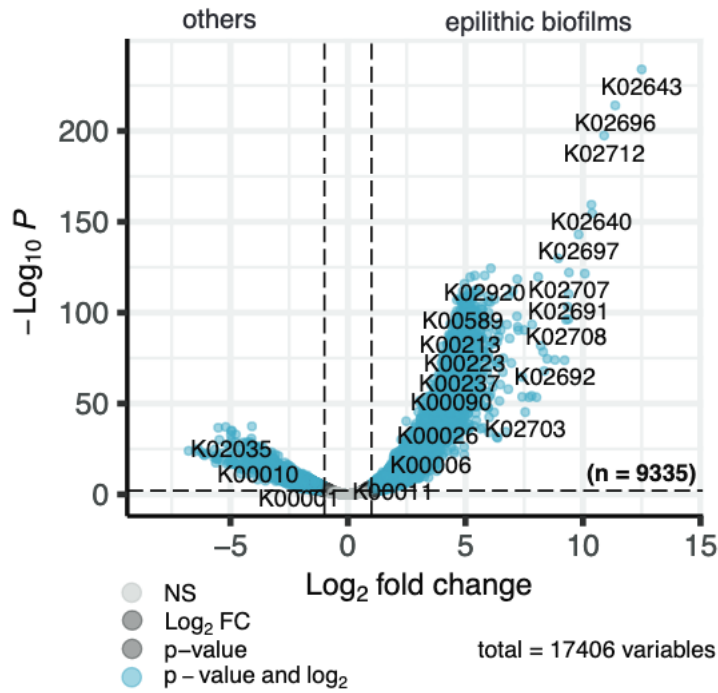
Supplementary figures 3. 5

Comparison to public metagenomes reveals differential gene abundances. Volcano plot indicating the total number of KOs (n = 9,335; total = 17,406) enriched in epilithic biofilms compared to 105 publicly available metagenomes. KO enrichment was assessed using DEseq2, where the adjusted p-value < 0.05 was considered to be significant.
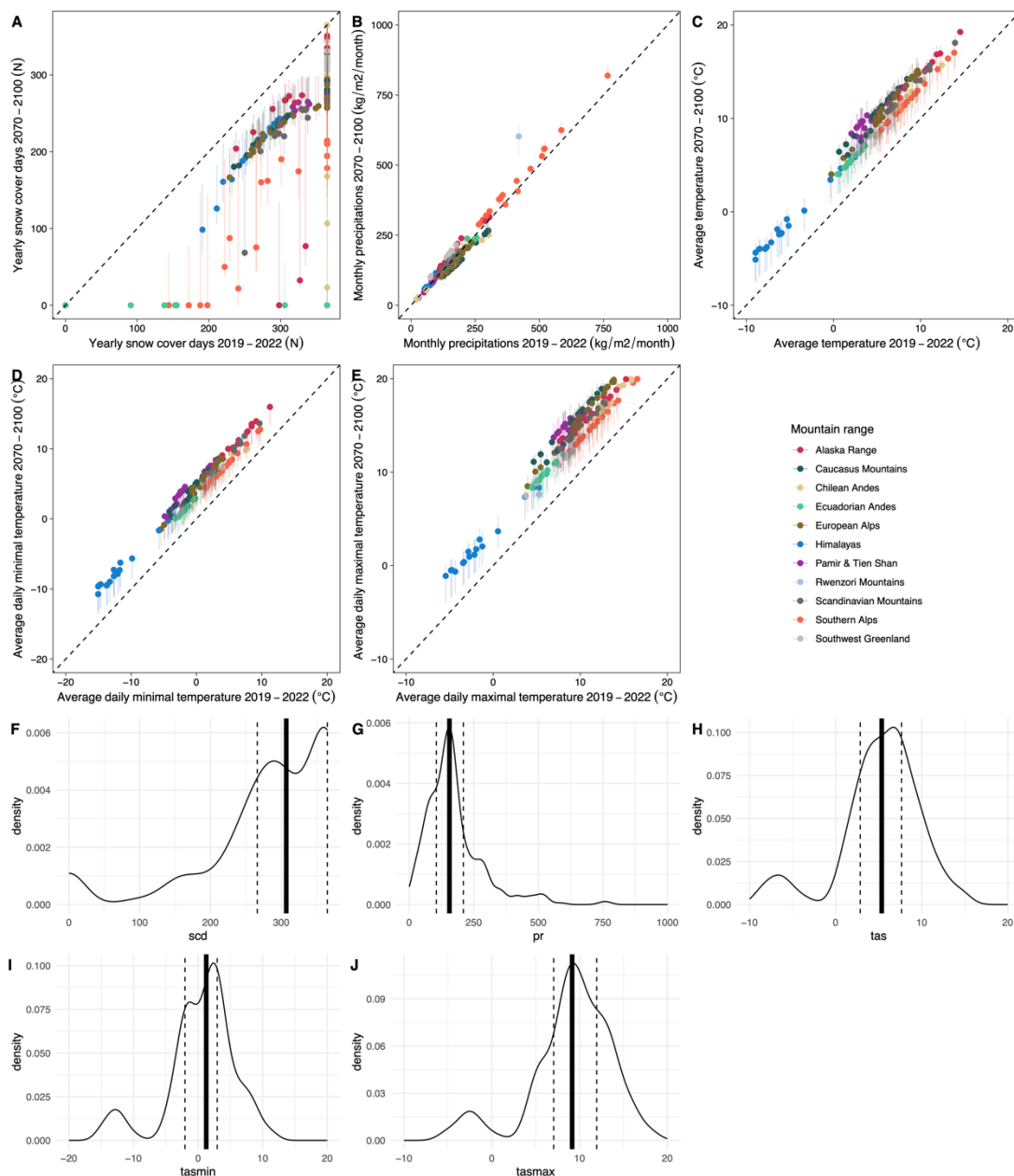
Supplementary figure 4. 1

Climatic changes Comparison of sampling year data (2019-2022) to future projections (2070-100) for the yearly snow cover days (**A**), monthly precipitations (**B**), average temperature (**C**), average minimal daily temperature (**D**), and average maximal daily temperature (**E**). The dots show data for the scenario RCP 4.5, the error bars show the RCP 2.6 and 8.5, and the dashed line the one-to-one line. Plots **F-J** show the median (bold line) and interquartile range (dashed lines) for the 2019-2022 dataset on top of the probability density functions of the historical averages for the 1981-2010 period for the same parameters (scd=yearly snow cover days, pr=monthly precipitations, tas=average temperature, tasmin=average minimal daily temperature, tasmax=average maximal daily temperature). These show

that the dataset created for the 2019-2022 does not exhibit extreme values compared to historical distribution averages.
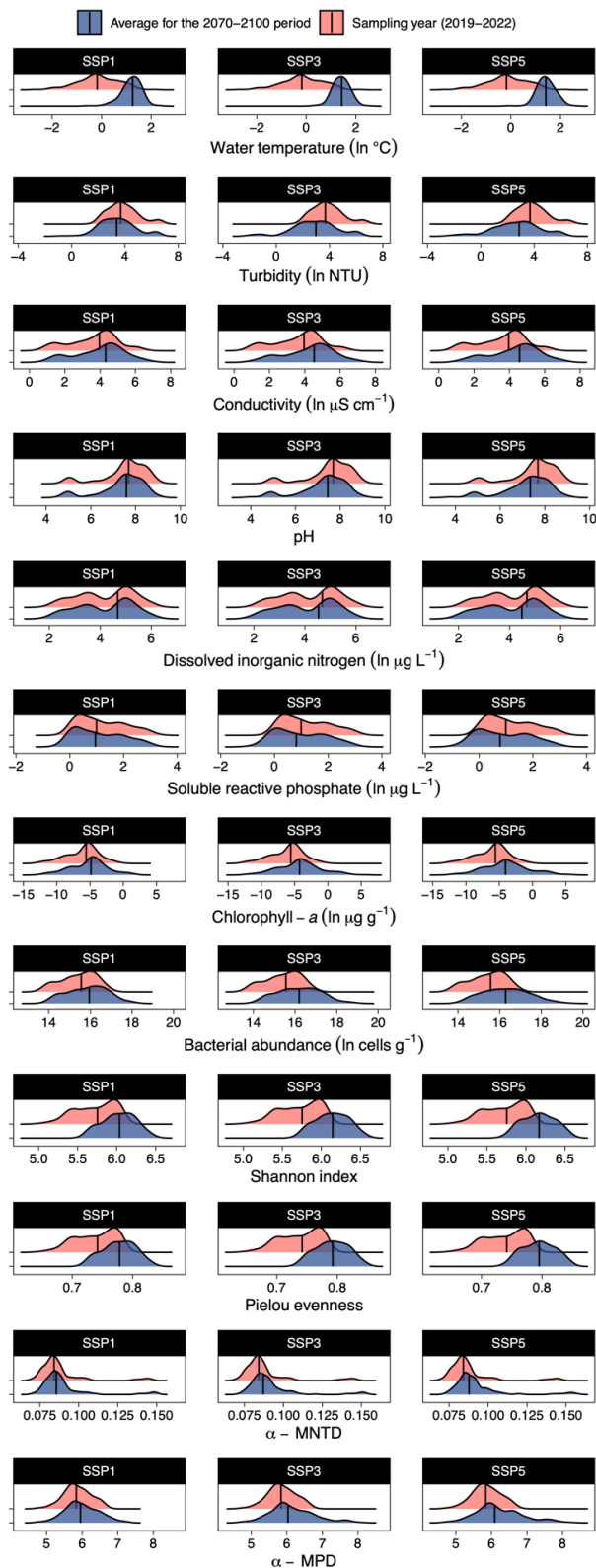
Supplementary figure 4. 2

Proximal drivers changes for all scenarios Distributions showing the predicted values for proximal drivers and biomass models (rows) for the sampling year (red) and future (blue) projections for the scenarios RCP 2.6, 4.5 and 8.5 (facets, from left to right).
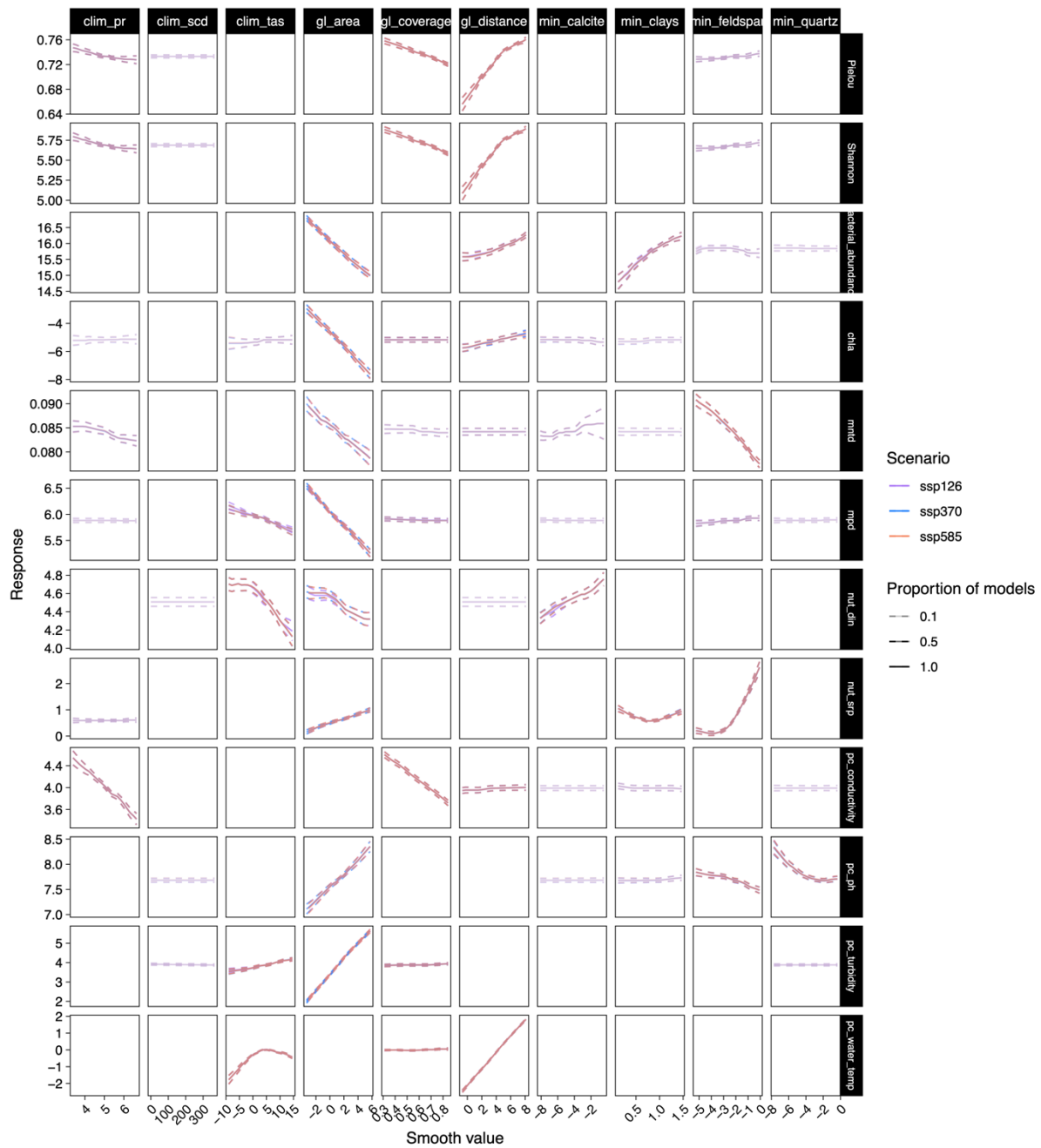
Supplementary figure 4. 3

Proximal drivers and biomass models response curves Response curves showing the effect of the covariates (column) on the proximal drivers and biomass variables (rows) for RCP scenarios 2.6 (purple), 4.5 (blue), and 8.5 (orange). The alpha of the lines represents the proportion of stacked models for which the covariate was selected in the feature selection procedure (more important variables are opaquer). Response curves were computed by predicting values for the response variable for each

covariate separately through the range of covariate values observed, keeping all other at their median. Since spatial splines were used, the median response across sampling sites was computed. The abbreviations for the covariates are the following: gl_dist=distance to the glacier, gl_area=area of the glacier, gl_coverage=coverage of the catchment by the glacier, clim_tas=average monthly temperature, clim_scd=yearly snow cover days, clim_pr=monthly precipitations, min_calcite=calcite content of the sediments, min_clays=clay content of the sediments, min_feldspar=feldspar content of the sediments, min_quartz=quartz content of the sediments. The abbreviations for the variables are the following: bacterial_abundance =abundance of bacterial cells, chla=mass of chlorophyll-a in the sediments, nut_din=streamwater concentration in dissolved inorganic nitrogen, nut_srp=streamwater concentration in soluble reactive phosphate, pc_conductivity=conductivity of the streamwater, pc_ph=pH of the streamwater, pc_turbidity=turbidity of the streamwater, and pc_water_temp=streamwater temperature.
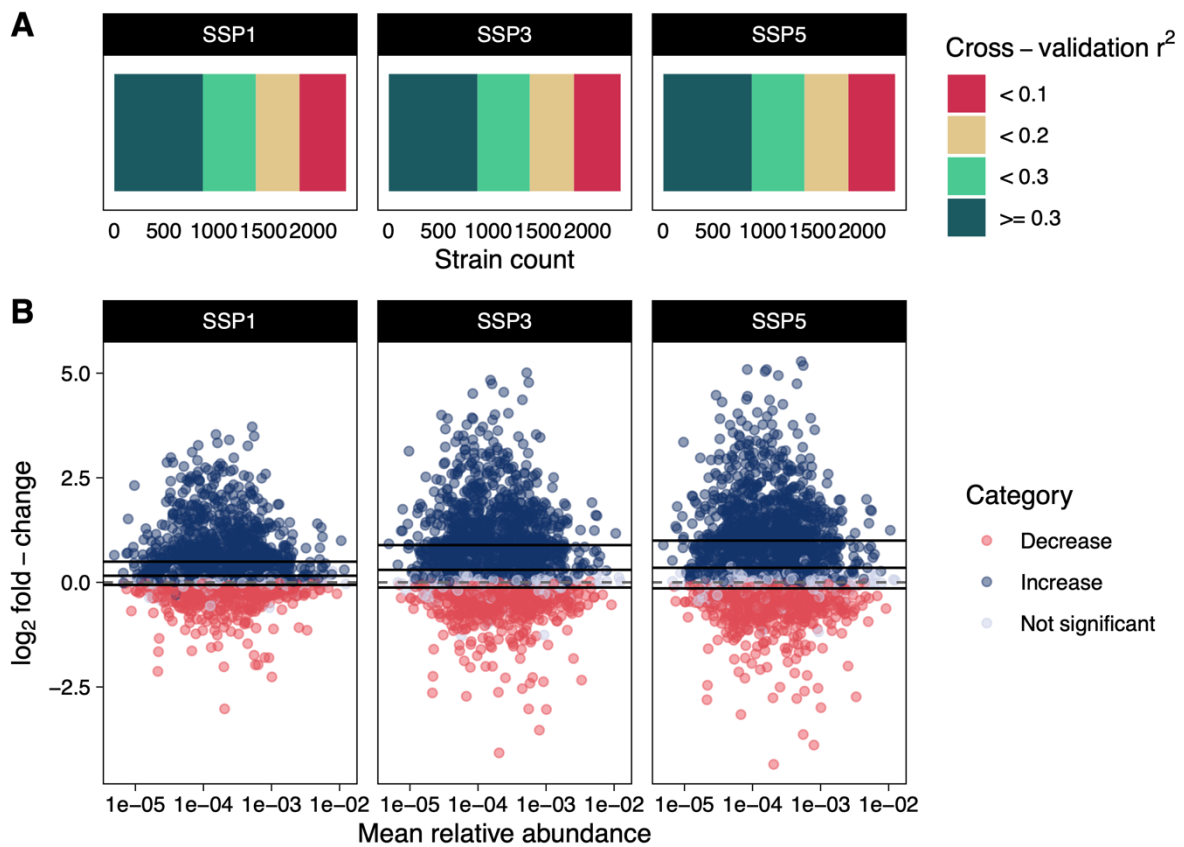
Supplementary figure 4. 4

Strains models performances (**A**) Distribution of the prediction r^2 values for the strain abundance models, shown for RCP 2.6. (126), RCP 4.5 (370), and RCP 8.5 (585). The colours show different categories of prediction r^2 based on thresholds (colours). (**B**) Comparison of the $\log_2$ fold-change to the mean current relative abundance for RCP 2.6 (126), RCP 4.5 (370), and RCP 8.5 (585). The lines show the median and interquartile range of the $\log_2$ fold change distributions. The colours show strains that significantly increase/decrease in abundance in future scenarios of climate change.
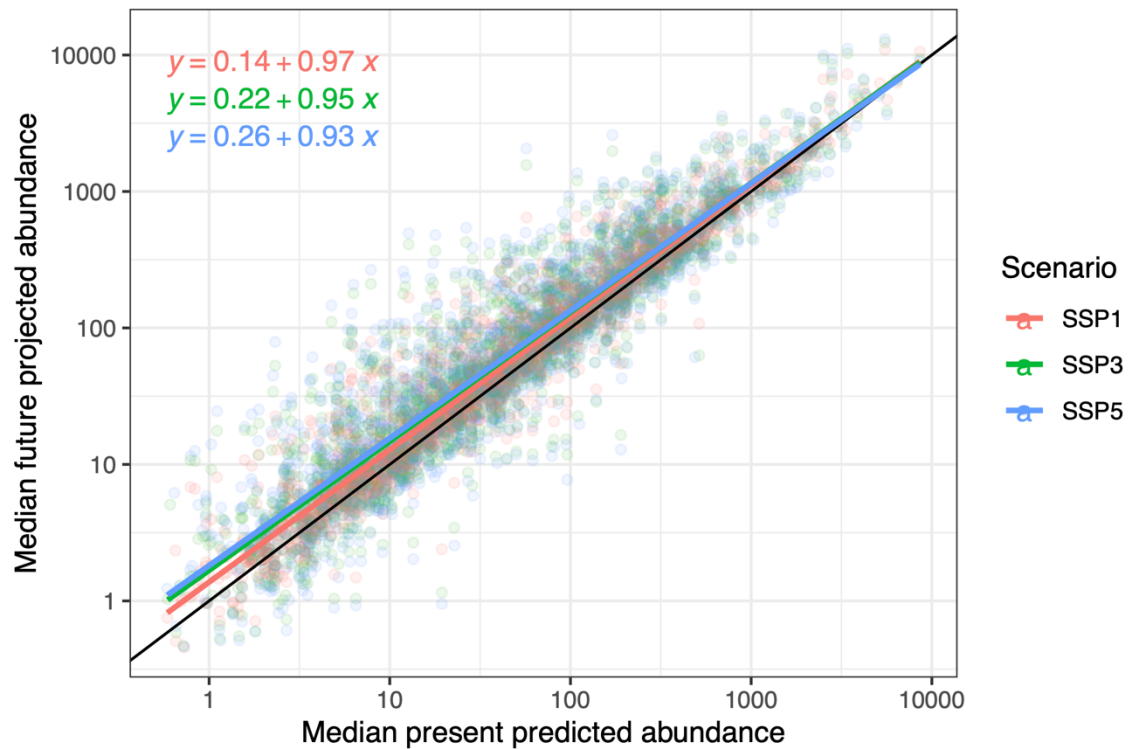
Supplementary figure 4. 5

Evenness analysis Scatter plot with linear regression lines for RCP2.6 (red), 4.5 (green) and 8.5 (blue), of the present median predicted abundance (x-axis) and future projected median (y-axis). Linear model equations are shown on the top left part of the plot, showing that all slopes are smaller than one, meaning that in the future, rare taxa are predicted to be more abundant than in present (i.e., increased evenness). present=sampling year (2019-2022); future=mean for the 2070-2100 time period
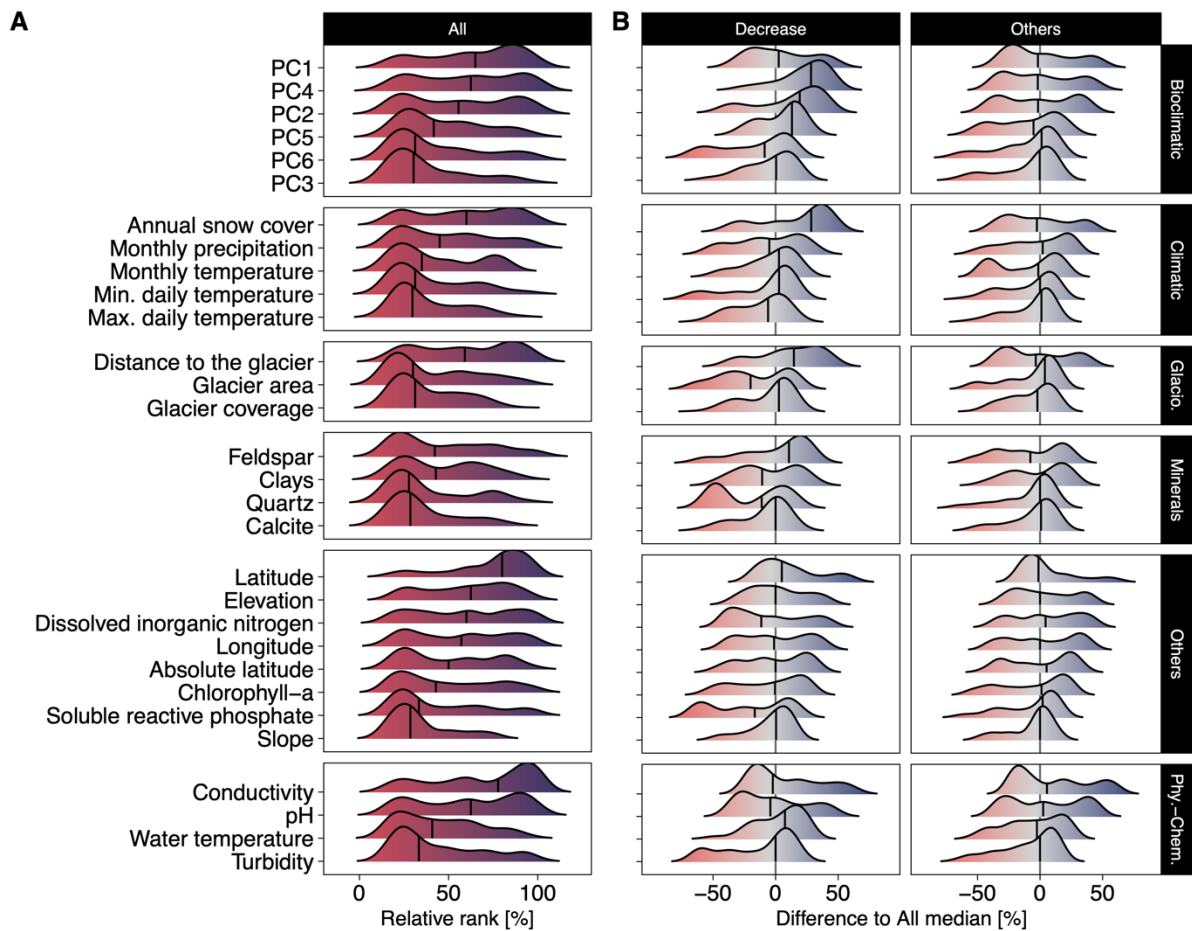
Supplementary figure 4. 6

Importance of features in the strain distribution models. (**A**) Ridgeplots displaying the distribution of the covariates' relative rank in the strain distribution models feature selection procedures, (**B**) and comparing the distributions of the strains that are predicted to decrease in abundance compared to others. The relative ranks were computed by comparing the covariates selection order after the feature selection procedure prior to model creation (*Methods*). The colors represent the relative rank distribution (**A**, darker shade), and the difference between the overall median for all strains, against the ones that are predicted to decrease in abundance (Decrease) and the others (Others) (**B**, lighter shade).
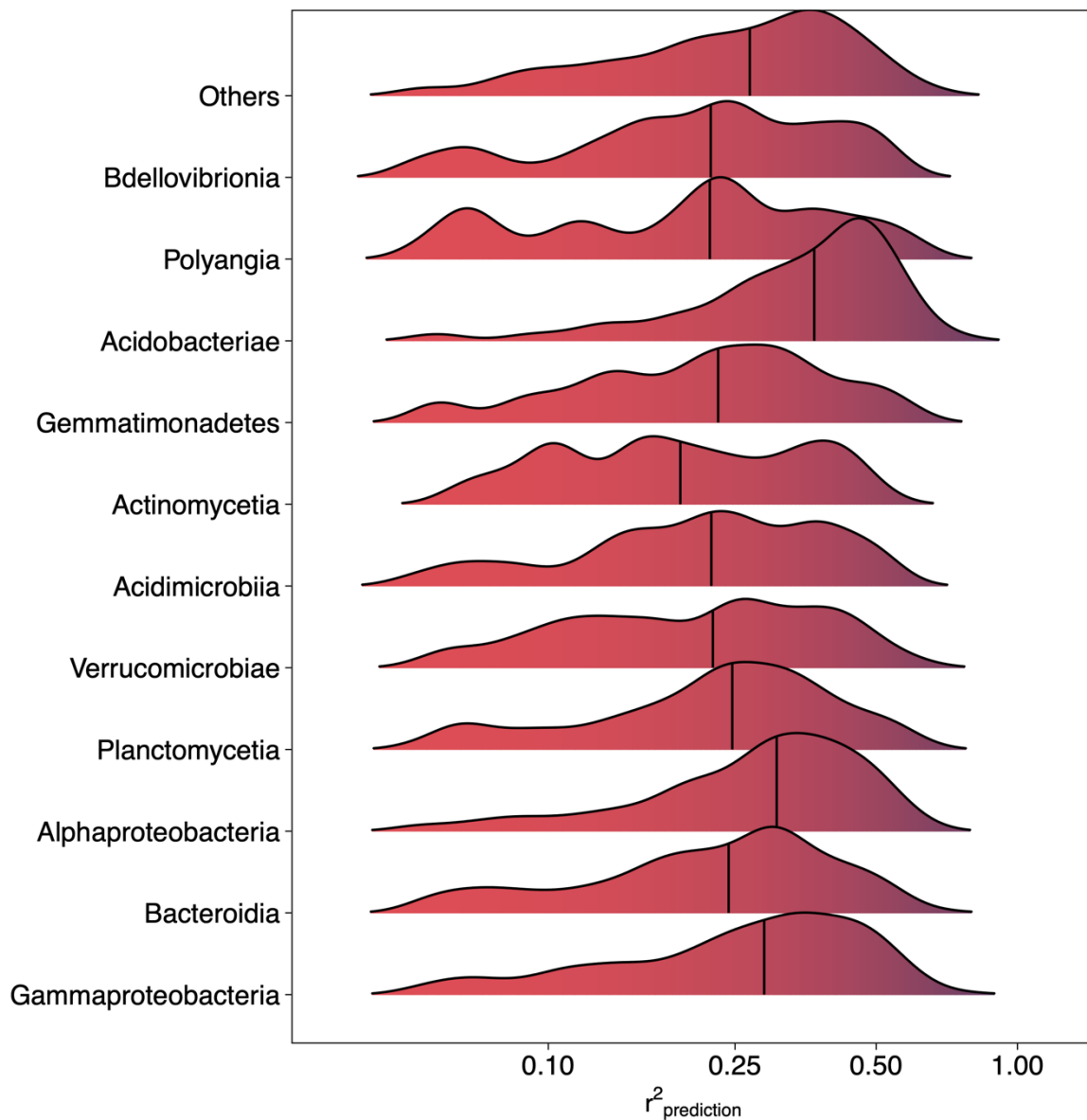
Supplementary figure 4. 7

Predictability of the most abundant bacterial classes Ridge plot depicting the distribution of the predictability (measured as the cross-validated $r^2_{prediction}$) across the 11 most abundant bacterial classes. Predictability was significantly different across these taxa (Kruskal-Wallis test, statistic = 7570.3, p < 0.001).

R = −0.15, *p* < 2.2e−16
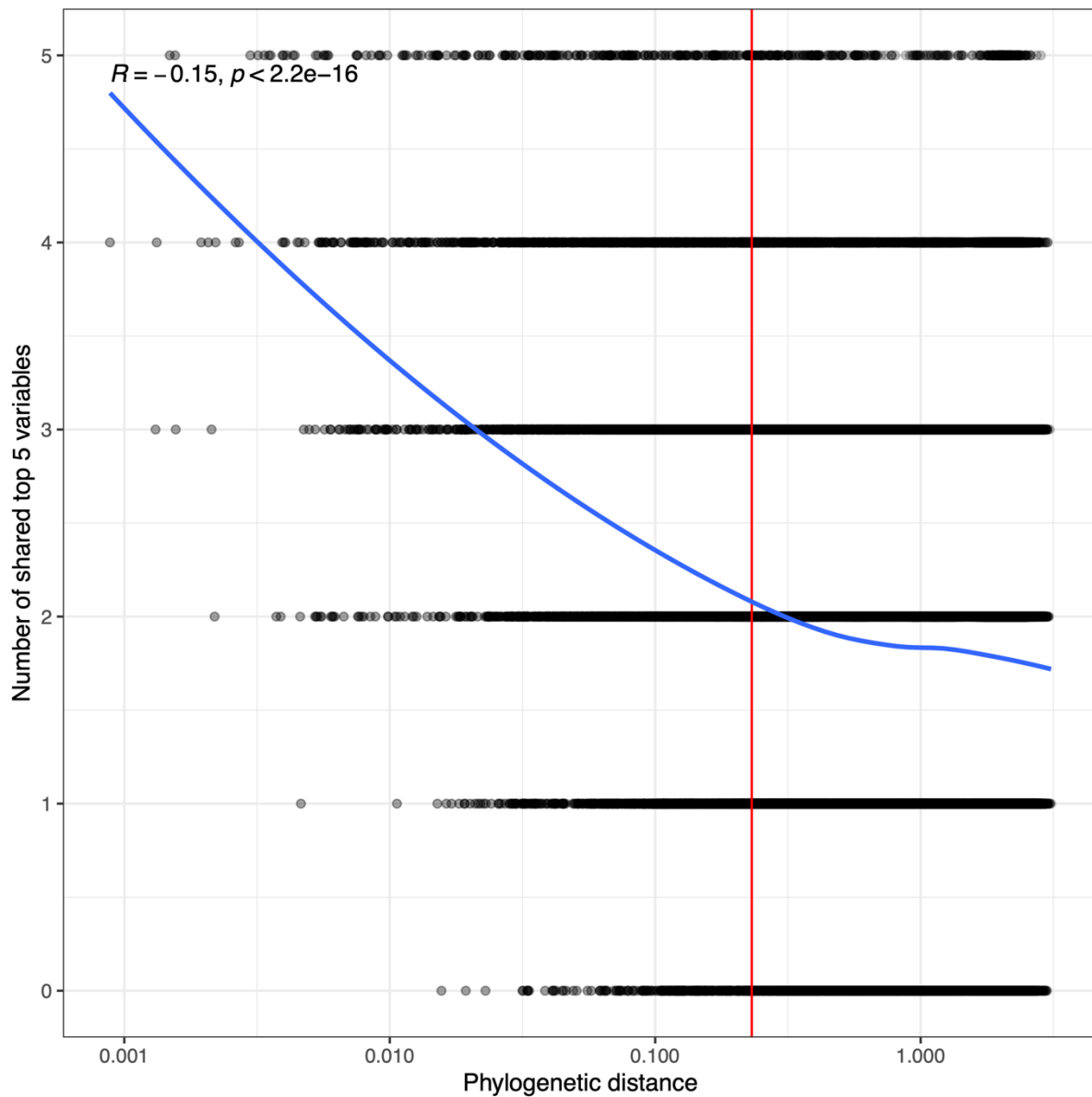
Supplementary figure 4. 8

Variable selection and phylogeny Comparison of the number of shared top 5 variables in the feature selection procedure with the phylogenetic distance, showing that more closely related strains are associated with similar covariates. The blue lines show a loess fit, the red vertical one shows the average within-genus phylogenetic distance. The relationship was significant with a Spearman correlation test (Rho = -0.14, p < 0.001).
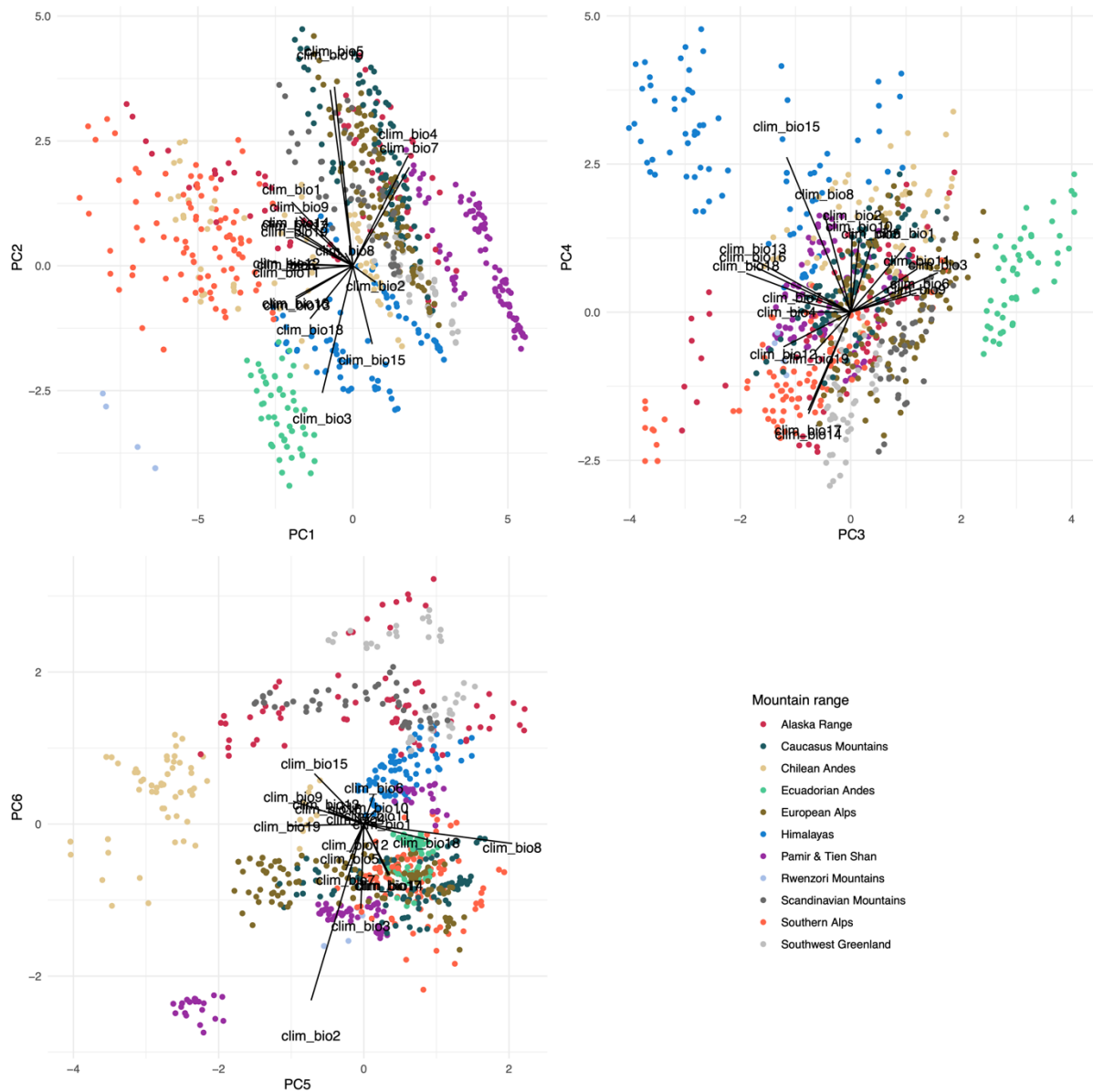
Supplementary figure 4. 9

Bioclimatic PCA Principal components 1 to 6 for the PCA analysis performed on the 19 bioclimatic variables (clim_bio1 to 19), accounting for more than 95% of the variance. The colours correspond to the mountain ranges, arrows representing the scores of the individual variables were added with the variable names as labels.

Supplementary table 2. 1

Dataset	Accuracy [%]	Precision [%]	Recall [%]	AUC [%]
PP1	96.04±1.93	99.99±0.05	92.08±3.86	99.93±0.02
PP2	97.95±1.35	99.93±0.14	95.96±2.71	99.93±0.02

Cryospheric bacterial communities' logistic classification models performance summary for each primer pair dataset.

Supplementary table 2. 2

Dataset	Group 1	Group 2	Sorensen's index			ß-MNTD		
			Corrected p-value	*r*	*Median difference*	*Corrected p-value*	*r*	*Median difference*
PP1	Cryo-Cryo	Cryo-Others	< 2e-16	0.185	0.038	< 2e-16	0.0821	0.012
	Cryo-Cryo	Other-Others	< 2e-16	0.140	0.042	< 2e-16	0.0639	0.01
	Cryo-Others	Other-Others	7.6e-13	0.0174	0.004	< 2e-16	0.144	0.022
PP2	Cryo-Cryo	Cryo-Others	< 2e-16	0.238	0.046	<2e-16	0.183	0.032
	Cryo-Cryo	Other-Others	< 2e-16	0.263	0.057	<2e-16	0.125	0.028
	Cryo-Others	Other-Others	< 2e-16	0.0501	0.011	<2e-16	0.0503	0.004

ß-diversity phylogenetics (Sorensen's Index and ß-MNTD) computed for 50 iterations randomly drawing 50 cryospheric and 50 non-cryospheric samples (sample sizes for each group: $N_{PP1-Sor}$ = 83583, $N_{PP2-Sor}$ = 98142, $N_{PP1-MNTD}$ = 77893, $N_{PP2-MNTD}$=91398). For both datasets and primer pairs, the Kruskal-Wallis tests were highly significant (p-value < 2.2e-16), post-hoc two-sided Wilcoxon tests results are reported in the table, the p-value was corrected using the Holm method. The effect size was computed as r with the *statix* R package.

Supplementary table 2. 3

		α-MPD (model p-value: < 2.2e-16)		
		Estimate	**p-value**	**t-value**
Coefficients	*Intercept*	0.389±0.014	28.667	<2e-16
	Cryosphere	0.077±0.005	13.933	<2e-16
	log(SR)	0.062±0.004	15.973	<2e-16
	DatasetPP2	-0.038±0.004	-9.129	<2e-16
Model	*Adj. R²*	0.108		
	df	4240		
		α-MNTD (model p-value: < 2.2e-16)		
		Estimate	**t-value**	**p-value**
Coefficients	*Intercept*	0.307±0.006	48.933	<2e-16
	Cryosphere	0.015±0.003	5.734	1.05e-08
	log(SR)	-0.053±0.002	-29.399	<2e-16
	DatasetPP2	0.020±0.002	10.302	<2e-16
Model	*Adj. R²*	0.191		
	df	4240		
		α-PD (model p-value: < 2.2e-16)		
		Estimate	**t-value**	**p-value**
Coefficients	*Intercept*	0.450±0.047	9.483	<2e-16
	Cryosphere	0.532±0.049	10.955	<2e-16
	SR	0.110±0.001	89.696	<2e-16
	DatasetPP2	0.398±0.037	10.887	<2e-16
Model	*Adj. R²*	0.664		
	df	4240		

α-diversity phylogenetics (MPD: mean phylogenetic distance, MNTD: mean nearest taxon distance; PD: phylogenetic diversity) linear models testing the influence of the cryosphere on the different metrics, with the species richness (SR, log-transformed) and the dataset as fixed effects (Intercept = non-cryospheric, and PP1).

Supplementary table 2. 4

PERMANOVA	DF	Sum of squares	r²	f-value	p-value
Ecosystem	3	55.774	0.18319	52.702	< 0.001
Dataset	1	5.633	0.01850	15.969	< 0.001
Residual	689	243.055	0.79831		
Total	693	304.462	1.00000		

Snow/Ice – Terr.	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	18.859	0.11779	54.36	< 0.001
Dataset	1	4.906	0.03064	14.14	< 0.001
Residual	393	136.340	0.85157		
Total	395	160.105	1.00000		

Snow/Ice – Marine	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	23.280	0.15758	66.918	< 0.001
Dataset	1	4.087	0.02767	11.749	< 0.001
Residual	346	120.370	0.81476		
Total	348	147.737	1.00000		

Snow/Ice – Fresh.	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	10.991	0.06768	28.6254	< 0.001
Dataset	1	3.195	0.01967	8.3202	< 0.001
Residual	386	148.211	0.91265		
Total	388	162.397	1.00000		

Marine – Terr.	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	25.402	0.20705	85.019	< 0.001
Dataset	1	7.052	0.05748	23.604	< 0.001
Residual	302	90.230	0.73547		
Total	304	122.684	1.00000		

Marine – Fresh.	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	18.464	0.14665	53.295	< 0.001
Dataset	1	5.240	0.04162	15.125	< 0.001
Residual	295	102.202	0.81173		
Total	297	125.906	1.00000		

Terrestrial - Fresh.	DF	Sum of squares	r²	f-value	p-value
Ecosystem	1	16.200	0.11536	47.332	< 0.001
Dataset	1	7.175	0.05109	20.963	< 0.001
Residual	342	117.056	0.83355		
Total	344	140.431	1.00000		

Model summaries for the PERMANOVA and all pairwise.adonis comparisons of cryospheric ecosystem types.

Supplementary table 2. 5

WTS model	Test statistic	df	p-value WTPS	
Ecosystem	112.0236	3	0	
Interaction	103.1681	4	0	
Ecosystem	**Datastet**	**N**	**Mean**	**Variance**
Freshwater	PP1	29	3.72	0.26
Freshwater	PP2	140	2.84	0.48
Ice/Snow	PP1	92	2.76	0.69
Ice/Snow	PP2	128	2.93	0.68
Marine	PP1	88	3.03	0.18
Marine	PP2	41	3.71	0.41
Terrestrial	PP1	92	3.69	0.80
Terrestrial	PP2	84	3.64	0.28

Shannon's index H α-diversity (calculated at the genus taxonomic level) Wald-Type Statistic (WTS) summary. This test was chosen as a non-parametric alternative to ANOVA for non-normally distributed data. "Interaction" represents the interaction between the Ecosystem and Dataset parameters, "Ecosystem" the fixed effect of the ecosystem type. The computed means weighted by sample sizes are 2.987, 2.856, 3.245, and 3.669 for freshwater, snow/ice, marine and terrestrial, respectively.

Supplementary table 2. 6

Metric	Median_{Others (n=265)}	Median_{Cryo (n=197)}	Median_{Under.(n=198)}
GC content [%]	48.7	57.5 (corr. p = 0.0011)	56.1 (corr. p = 0.0086)
Genome size [mbp]	3.97	4.19 (corr. p = 0.17)	4.06 (corr. p = 0.28)
Growth doubling time [d]	4.48	4.12 (corr. p = 0.87)	4.55 (corr. p = 0.51)
Codon usage bias [CUBHE]	0.627	0.627 (corr. p = 1)	0.623 (corr. p = 1)
Consistency [HE]	0.527	0.519 (corr. p = 0.84)	0.529 (corr. p = 0.25)
Codon pair bias [CPB]	-0.375	-0.370 (corr. p = 0.92)	-0.380 (corr. p = 0.74)

RefSeq genomic properties summary. Corrected p-values were computed using two-sided Wilcoxon tests implemented in the *compare_means* function of the *ggpubr* R package, comparing the cryospheric and underrepresented genera against the others.

Supplementary table 2. 7

Dataset	Study ID	Sample n.	Ecosystem type	Description
PP1	Bergk2019	12	Ice/Snow	Snow samples, Svalbard
PP1	NOMIS	10	Freshwater	Glacier-fed stream sediment samples, Russia and New Zealand
PP1	PRJDB9246	11	Freshwater	Microbial mat and water, Antarctica
PP1	PRJEB12640	41	Terrestrial	Soil chronosequence samples, Svalbard
PP1	PRJEB26163	62	Marine	Marine water, Arctic Ocean
PP1	PRJEB29215	32	Ice/Snow	Snow samples, Antarctica
PP1	PRJEB31938	26	Marine	Sea ice, snow, water and sediment, Greenland and the Arctic Ocean
PP1	PRJEB40467	18	Terrestrial	Alpine permafrost, Italy
PP1	PRJNA296475	3	Ice/Snow	Cryoconite hole, Svalbard
PP1	PRJNA320505	1	Ice/Snow	Cryoconite hole, Antarctica
PP1	PRJNA380676	1	Terrestrial	Arctic rock, Svalbard
PP1	PRJNA418054	32	Terrestrial	Permafrost, Alaska
PP1	PRJNA430179	11	Ice/Snow	Glacier Ice/snow, Spain
PP1	PRJNA436954	8	Frehshwater	High-arctic microbial mat
PP1	PRJNA529498	33	Ice/Snow	Cryoconite hole, Antarctica
PP2	PRJEB11496	24	Marine	Marine sediment, Antarctica
PP2	PRJEB23054	59	Terrestrial	Permafrost, Alaska
PP2	PRJNA244335	31	Freshwater	Sediment and water from subglacial lake, Antarctica
PP2	PRJNA255432	70	Freshwater	Arctic lake, Canada
PP2	PRJNA278982	4	Marine	Ice-shelf water cavity, Antarctica
PP2	PRJNA321351	9	Freshwater	Arctic lake, Greenland
PP2	PRJNA324626	9	Terrestrial	Frozen soil, China
PP2	PRJNA430887	26	Ice/Snow	Glacier ice and weather crust, USA
PP2	PRJNA431087	1	Terrestrial	Microbial mat, Antarctica
PP2	PRJNA432743	7	Freshwater	Subglacial aquifer brine, Antarctica
PP2	PRJNA471245	51	Ice/Snow	Water, ice, soil, sediment and microbial mat, Antarctica
PP2	PRJNA480849	1	Ice/Snow	Cryoconite hole, Antarctica
PP2	PRJNA554442	11	Terrestrial	Permafrost, Alaska

PP2	PRJNA593264	13	Marine	Water, sediments and snow, Antarctica
PP2	PRJNA629965	6	Freshwater, Ice/Snow	Snow and glacier melt
PP2	PRJNA744712	72	Ice/Snow	Cryoconite hole, Antarctica

Summary of the cryospheric samples included in the two 16s rRNA amplicon datasets (primer pair 1 = PP1, and primer pair 2 = PP2).

Supplementary table 2. 8

Metric	Group	Comparison	Corr. p-value
Uniprot identity [%]	KEGG	Cryo.–Shared	2.9e-2
		Cryo.–Others	5.7e-10
		Shared-Others	1.2e-101
	Ambiguous	Cryo.–Shared	4.2e-3
		Cryo.–Others	2.6e-12
		Shared-Others	2.7e-87
	Unassigned	Cryo.–Shared	2.6e-7
		Cryo.–Others	2.1e-13
		Shared-Others	2.6e-3
Mean GC [%]	KEGG	Cryo.–Shared	2.7 e-5
		Cryo.–Others	1.1e-17
		Shared-Others	5.50e-292
	Ambiguous	Cryo.–Shared	7.8e-9
		Cryo.–Others	1.1e-50
		Shared-Others	4.20e-297
	Unassigned	Cryo.–Shared	8.6e-1
		Cryo.–Others	1.1e-68
		Shared-Others	0
Mean Cluster Identity [%]	KEGG	Cryo.–Shared	4.3e-19
		Cryo.–Others	2.1e-7
		Shared-Others	3.9e-81
	Ambiguous	Cryo.–Shared	3e-39
		Cryo.–Others	1.9e-11
		Shared-Others	3.9e-78
	Unassigned	Cryo.–Shared	8.8e-12
		Cryo.–Others	5.4e-1
		Shared-Others	5.4e-63

Unassigned functional clusters exact p-values for the pairwise two-sided Wilcoxon tests. Corrected p-values were adjusted using the Holm method. Sample sizes are listed in Table 1.

Supplementary table 4. 1

Sample	Mountain range	date [DD.MM.YYYY]	time [HH:MM]	Glacier name	latitude [DD]	longitude [DD]	elevation [m]	rgi_v6	glims_id
GL1_UP	Southern Alps	22.01.19	10:00	Franz Josef	-43.45	170.1753	470	RGI60-18.02397	G170225E43495S
GL1_DN	Southern Alps	22.01.19	14:26	Franz Josef	-43.4274	170.1734	248	RGI60-18.02397	G170225E43495S
GL2_UP	Southern Alps	25.01.19	08:37	Victoria	-43.4976	170.139	1145	RGI60-18.02270	G170174E43508S
GL2_DN	Southern Alps	25.01.19	11:59	Victoria	-43.4978	170.1291	1093	RGI60-18.02270	G170174E43508S
GL3_UP	Southern Alps	26.01.19	10:30	Fox	-43.5006	170.059	338	RGI60-18.02375	G170162E43537S
GL3_DN	Southern Alps	26.01.19	14:49	Fox	-43.4869	170.0293	226	RGI60-18.02375	G170162E43537S
GL5_UP	Southern Alps	29.01.19	11:47	Lancelot	-42.9249	171.5141	1325	RGI60-18.02895	G171502E42928S
GL5_DN	Southern Alps	29.01.19	16:04	Lancelot	-42.9263	171.5149	1255	RGI60-18.02895	G171502E42928S
GL6_UP	Southern Alps	30.01.19	13:53	Crow	-42.9241	171.5137	1362	RGI60-18.02896	G171511E42916S
GL6_DN	Southern Alps	30.01.19	09:48	Crow	-42.9307	171.5185	1118	RGI60-18.02896	G171511E42916S
GL7_UP	Southern Alps	31.01.19	14:37	White	-42.9984	171.3898	1750	RGI60-18.02839	G171385E43000S
GL7_DN	Southern Alps	31.01.19	17:50	White	-42.9963	171.3913	1687	RGI60-18.02839	G171385E43000S
GL8_UP	Southern Alps	02.02.19	10:54	Marmaduke Dixon	-42.9875	171.3904	1629	RGI60-18.02823	G171383E42988S
GL8_DN	Southern Alps	02.02.19	15:15	Marmaduke Dixon	-42.9884	171.3923	1584	RGI60-18.02823	G171383E42988S
GL9_UP	Southern Alps	03.02.19	10:12	Cahill	-42.9824	171.3978	1510	RGI60-18.02851	G171391E42978S
GL9_DN	Southern Alps	03.02.19	13:44	Cahill	-42.9839	171.4025	1253	RGI60-18.02851	G171391E42978S

GL10_UP	Southern Alps	08.02.19	11:52	Dart	-44.4813	168.6059	1090	RGI60-18.00686	G168609E44455S
GL10_DN	Southern Alps	08.02.19	15:19	Dart	-44.4818	168.5979	1042	RGI60-18.00686	G168609E44455S
GL11_UP	Southern Alps	09.02.19	11:56	Reid	-44.4673	168.6216	1587	RGI60-18.00676	G168623E44461S
GL11_DN	Southern Alps	09.02.19	13:52	Reid	-44.4689	168.6193	1486	RGI60-18.00676	G168623E44461S
GL12_UP	Southern Alps	10.02.19	12:38	Rob Roy	-44.4758	168.7268	749	RGI60-18.01013	G168718E44463S
GL12_DN	Southern Alps	10.02.19	16:18	Rob Roy	-44.4803	168.7265	717	RGI60-18.01013	G168718E44463S
GL13_UP	Southern Alps	13.02.19	13:20	Brewster	-44.0819	169.4317	1699	RGI60-18.01130	G169437E44072S
GL13_DN	Southern Alps	13.02.19	16:20	Brewster	-44.0838	169.4305	1655	RGI60-18.01130	G169437E44072S
GL14_UP	Southern Alps	16.02.19	10:12	Mc Pherson	-44.7565	167.9857	1096	RGI60-18.00367	G167987E44758S
GL14_DN	Southern Alps	16.02.19	13:18	Mc Pherson	-44.7571	167.9865	1071	RGI60-18.00367	G167987E44758S
GL15_UP	Southern Alps	17.02.19	08:52	Age	-44.6121	168.0218	1288	RGI60-18.00179	G168020E44606S
GL15_DN	Southern Alps	17.02.19	05:31	Age	-44.6121	168.0214	1259	RGI60-18.00179	G168020E44606S
GL16_UP	Southern Alps	22.02.19	13:53	Birch Creek	-43.794	170.0643	1320	RGI60-18.01559	G170055E43788S
GL16_DN	Southern Alps	22.02.19	17:09	Birch Creek	-43.7946	170.0674	1206	RGI60-18.01559	G170055E43788S
GL17_UP	Southern Alps	23.02.19	11:50	Tewaewae	-43.6891	170.0822	1236	RGI60-18.01881	G170073E43683S
GL17_DN	Southern Alps	23.02.19	15:38	Tewaewae	-43.6936	170.0849	1006	RGI60-18.01881	G170073E43683S
GL18_UP	Southern Alps	26.02.19	11:07	Charity	-43.8169	169.9252	1208	RGI60-18.01835	G169924E43808S
GL18_DN	Southern Alps	26.02.19	14:50	Charity	-43.8184	169.9279	1113	RGI60-18.01835	G169924E43808S
GL19_UP	Southern Alps	27.02.19	09:37	Richardson	-43.8172	169.9336	1146	RGI60-18.01958	G169947E43804S
GL19_DN	Southern Alps	27.02.19	16:07	Richardson	-43.8239	169.9231	1097	RGI60-18.01958	G169947E43804S

GL20_UP	Southern Alps	01.03.19	12:36	Mawson	-43.418	170.5	1448	RGI60-18.02348	G170508E43421S
GL20_DN	Southern Alps	01.03.19	16:45	Mawson	-43.4074	170.4994	1004	RGI60-18.02348	G170508E43421S
GL21_UP	Southern Alps	02.03.19	10:12	Shackleton	-43.4043	170.5064	1138	RGI60-18.02298	G170521E43401S
GL21_DN	Southern Alps	02.03.19	13:43	Shackleton	-43.406	170.501	982	RGI60-18.02298	G170521E43401S
GL22_UP	European Alps	25.06.19	09:13	Valsorey	45.9166	7.2669	2441	RGI60-11.02927	G007257E45892N
GL22_DN	European Alps	25.06.19	13:36	Valsorey	45.9208	7.257	2387	RGI60-11.02927	G007257E45892N
GL23_UP	European Alps	03.06.20	12:46	Furgg	45.979	7.6868	2735	RGI60-11.02819	G007696E45965N
GL23_DN	European Alps	03.06.20	15:18	Furgg	45.9845	7.6928	2680	RGI60-11.02819	G007696E45965N
GL24_UP	European Alps	04.06.20	09:10	Zmutt	46.0016	7.6511	2317	RGI60-11.02739	G007639E45977N
GL24_DN	European Alps	04.06.20	11:50	Zmutt	46.003	7.6573	2247	RGI60-11.02739	G007639E45977N
GL25_UP	European Alps	05.06.20	11:00	Findelen	46.0108	7.8263	2557	RGI60-11.02773	G007880E45990N
GL25_DN	European Alps	05.06.20	14:30	Findelen	46.0105	7.8199	2508	RGI60-11.02773	G007880E45990N
GL26_UP	European Alps	06.06.20	11:30	Längflue-N	46.031	7.8515	2916	RGI60-11.02742	G007860E46022N
GL26_DN	European Alps	06.06.20	14:31	Längflue-N	46.0318	7.8495	2876	RGI60-11.02742	G007860E46022N
GL27_UP	European Alps	08.06.20	10:12	Arolla (Bas)	45.9901	7.496	2265	RGI60-11.02787	G007490E45965N
GL27_DN	European Alps	08.06.20	14:48	Arolla (Bas)	46.0006	7.4921	2112	RGI60-11.02787	G007490E45965N
GL28_UP	European Alps	09.06.20	08:16	Tsidjiore Nouve	46.016	7.4692	2277	RGI60-11.02755	G007450E45997N
GL28_DN	European Alps	10.06.20	11:05	Tsidjiore Nouve	46.0188	7.4736	2148	RGI60-11.02755	G007450E45997N

GL29_UP	European Alps	11.06.20	10:30	Mont Mine	46.0396	7.5505	2085	RGI60-11.02709	G007553E45994N
GL29_DN	European Alps	11.06.20	13:50	Mont Mine	46.0437	7.5535	1977	RGI60-11.02709	G007553E45994N
GL30_UP	European Alps	30.06.20	11:11	Hohlaub-N	46.1446	7.9939	3081	RGI60-11.02526	G008004E46144N
GL30_DN	European Alps	30.06.20	13:00	Hohlaub-N	46.1449	7.9921	2983	RGI60-11.02526	G008004E46144N
GL31_UP	European Alps	01.07.20	11:15	Schwarzberg	46.0265	7.939	2662	RGI60-11.02746	G007922E46000N
GL31_DN	European Alps	01.07.20	13:30	Schwarzberg	46.0276	7.9397	2659	RGI60-11.02746	G007922E46000N
GL32_UP	Southwest Greenland	15.07.19	20:50	"Nuuk east"	64.1383	-51.1768	613	RGI60-05.07197	G308826E64132N
GL32_DN	Southwest Greenland	16.07.19	20:18	"Nuuk east"	64.1395	-51.1732	550	RGI60-05.07197	G308826E64132N
GL33_UP	Southwest Greenland	16.07.19	12:00	"Nuuk middle"	64.1406	-51.1926	668	RGI60-05.07202	G308801E64139N
GL33_DN	Southwest Greenland	16.07.19	16:00	"Nuuk middle"	64.1418	-51.19	588	RGI60-05.07202	G308801E64139N
GL34_UP	Southwest Greenland	17.07.19	14:37	"Nuuk west"	64.1514	-51.2181	557	RGI60-05.07208	G308763E64146N
GL34_DN	Southwest Greenland	17.07.19	17:30	"Nuuk west"	64.1479	-51.1958	473	RGI60-05.07208	G308763E64146N
GL35_UP	Southwest Greenland	18.07.19	14:00	"Nuuk east"	64.166	-51.0694	757	RGI60-05.07210	G308916E64163N
GL35_DN	Southwest Greenland	18.07.19	13:00	"Nuuk east"	64.1626	-51.059	547	RGI60-05.07210	G308916E64163N
GL36_UP	Southwest Greenland	21.07.19	12:55	Aajuitsup Sermia	64.1211	-51.4805	523	RGI60-05.07199	G308534E64118N
GL36_DN	Southwest Greenland	21.07.19	16:50	Aajuitsup Sermia	64.1262	-51.4901	388	RGI60-05.07199	G308534E64118N
GL37_UP	Southwest Greenland	22.07.19	12:31	Teqqiinngallip Sermia	64.1352	-51.4707	439	RGI60-05.07207	G308542E64131N
GL37_DN	Southwest Greenland	22.07.19	16:12	Teqqiinngallip Sermia	64.1408	-51.4755	249	RGI60-05.07207	G308542E64131N
GL38_UP	Southwest Greenland	29.07.19	11:40	Lyngmarksbrae	69.2904	-53.571	737	RGI60-05.00625	G306419E69295N
GL38_DN	Southwest Greenland	29.07.19	15:25	Lyngmarksbrae	69.2815	-53.5372	444	RGI60-05.00625	G306419E69295N

GL39_UP	Southwest Greenland	30.07.19	12:06	Chamberlin	69.3204	-53.5282	468	RGI60-05.00623	G306428E69317N
GL39_DN	Southwest Greenland	30.07.19	15:19	Chamberlin	69.321	-53.4954	332	RGI60-05.00623	G306428E69317N
GL40_UP	Southwest Greenland	31.07.19	12:25	Petersen	69.3013	-53.5476	576	RGI60-05.00623	G306433E69302N
GL40_DN	Southwest Greenland	31.07.19	15:20	Petersen	69.2999	-53.5273	446	RGI60-05.00623	G306433E69302N
GL41_UP	Southwest Greenland	01.08.19	11:32	Pjetursson	69.2955	-53.3929	629	RGI60-05.00602	G306613E69293N
GL41_DN	Southwest Greenland	01.08.19	16:35	Pjetursson	69.3009	-53.4289	395	RGI60-05.00602	G306613E69293N
GL42_UP	Caucasus Mountains	11.09.19	12:09	Midjirgi	43.0936	43.1692	2637	RGI60-12.01262	G043172E43060N
GL42_DN	Caucasus Mountains	08.09.19	11:35	Midjirgi	43.1122	43.1494	2121	RGI60-12.01262	G043172E43060N
GL43_UP	Caucasus Mountains	09.09.19	10:05	Bezengi	43.1057	43.1318	2190	RGI50-12.00014	G043084E43033N
GL43_DN	Caucasus Mountains	08.09.19	04:48	Bezengi	43.1172	43.1493	2037	RGI50-12.00014	G043084E43033N
GL44_UP	Caucasus Mountains	10.09.19	13:00	"50"/"442"	43.0924	43.199	3481	RGI60-12.00595	G043218E43085N
GL44_DN	Caucasus Mountains	11.09.19	15:33	"50"/"442"	43.0941	43.1733	2660	RGI60-12.00595	G043218E43085N
GL45_UP	Caucasus Mountains	12.09.19	12:02	"50"/"4446"	43.0752	43.1743	3027	RGI60-12.01187	G043188E43071N
GL45_DN	Caucasus Mountains	12.09.19	14:51	"50"/"4446"	43.0751	43.1714	2924	RGI60-12.01187	G043188E43071N
GL46_UP	Caucasus Mountains	13.09.19	11:30	"443"	43.0832	43.1839	3154	RGI60-12.00396	G043201E43079N
GL46_DN	Caucasus Mountains	13.09.19	14:30	"443"	43.0858	43.173	2779	RGI60-12.00396	G043201E43079N
GL47_UP	Caucasus Mountains	17.09.19	10:23	Kashkatash	43.2116	42.6847	2517	RGI60-12.00259	G042693E43189N
GL47_DN	Caucasus Mountains	17.09.19	14:31	Kashkatash	43.2203	42.6839	2177	RGI60-12.00259	G042693E43189N

GL48_UP	Caucasus Mountains	18.09.19	10:59	Djankuat	43.2034	42.7501	2760	RGI50-12.01132	G042766E43192N
GL48_DN	Caucasus Mountains	18.09.19	14:36	Djankuat	43.2088	42.7396	2642	RGI50-12.01132	G042766E43192N
GL49_UP	Caucasus Mountains	19.09.19	11:38	Djantugan	43.198	42.7382	2799	RGI60-12.00314	G042746E43195N
GL49_DN	Caucasus Mountains	19.09.19	14:16	Djantugan	43.1992	42.7342	2741	RGI60-12.00314	G042746E43195N
GL50_UP	Caucasus Mountains	20.09.19	12:03	Shkhelda	43.1952	42.6484	2364	RGI60-12.00426	G042639E43169N
GL50_DN	Caucasus Mountains	20.09.19	15:20	Shkhelda	43.208	42.6511	2212	RGI60-12.00426	G042639E43169N
GL51_UP	Caucasus Mountains	22.09.19	10:13	Bashkara	43.2099	42.7242	2576	RGI60-12.00849	G042727E43193N
GL51_DN	Caucasus Mountains	22.09.19	13:30	Bashkara	43.2133	42.7148	2466	RGI60-12.00849	G042727E43193N
GL52_UP	Caucasus Mountains	23.09.19	11:33	Terskol	43.291	42.5067	2709	RGI60-12.00821	G042482E43317N
GL52_DN	Caucasus Mountains	23.09.19	14:23	Terskol	43.2753	42.5128	2387	RGI60-12.00821	G042482E43317N
GL53_UP	Caucasus Mountains	26.09.19	13:30	Garabashi	43.2846	42.4788	3059	RGI50-12.00161	G042467E43312N
GL53_DN	Caucasus Mountains	24.09.19	15:57	Garabashi	43.2677	42.4881	2400	RGI50-12.00161	G042467E43312N
GL54_UP	Caucasus Mountains	25.09.19	12:19	Bolshoy Azau	43.2721	42.4468	2561	RGI50-12.00080	G042422E43310N
GL54_DN	Caucasus Mountains	25.09.19	15:40	Bolshoy Azau	43.2654	42.4668	2394	RGI50-12.00080	G042422E43310N
GL55_UP	Caucasus Mountains	29.09.19	12:01	Irik	43.3029	42.5498	2661	RGI50-12.00730	G042500E43332N
GL55_DN	Caucasus Mountains	29.09.19	15:37	Irik	43.292	42.5921	2361	RGI50-12.00730	G042500E43332N
GL56_UP	Caucasus Mountains	02.10.19	10:42	Sopruju Sevenriy	43.2617	41.6102	2104	RGI60-12.00474	G041589E43251N
GL56_DN	Caucasus Mountains	02.10.19	14:18	Sopruju Sevenriy	43.2683	41.6149	1848	RGI60-12.00474	G041589E43251N
GL57_UP	Caucasus Mountains	03.10.19	10:30	Amanauz	43.2622	41.6141	1966	RGI60-12.01112	G041628E43233N
GL57_DN	Caucasus Mountains	03.10.19	14:27	Amanauz	43.2685	41.6151	1848	RGI60-12.01112	G041628E43233N

GL58_UP	Caucasus Mountains	04.10.19	10:30	Ptish	43.2356	41.6929	2347	RGI60-12.00063	G041694E43227N
GL58_DN	Caucasus Mountains	04.10.19	13:54	Ptish	43.2512	41.6916	2113	RGI60-12.00063	G041694E43227N
GL59_UP	Caucasus Mountains	05.10.19	12:36	"789"	43.2668	41.6772	1994	RGI60-12.00053	G041670E43254N
GL59_DN	Caucasus Mountains	05.10.19	15:19	"789"	43.2682	41.6782	1927	RGI60-12.00053	G041670E43254N
GL60_UP	Caucasus Mountains	08.10.19	10:40	Dvuyazichniy	43.2983	41.5478	2109	RGI60-12.00842	G041519E43297N
GL60_DN	Caucasus Mountains	08.10.19	14:23	Dvuyazichniy	43.2989	41.5569	1920	RGI60-12.00842	G041519E43297N
GL61_UP	Ecuadorian Andes	08.02.20	11:20	"San Marco's"	0.0402	-77.9969	4698	RGI60-16.01343	G282003E00032N
GL61_DN	Ecuadorian Andes	08.02.20	14:52	"San Marco's"	0.043	-77.9998	4620	RGI60-16.01343	G282003E00032N
GL62_UP	Ecuadorian Andes	09.02.20	11:21	"Laguna Verde West"	0.0167	-78.0065	4861	RGI60-16.01345	G282007E00013N
GL62_DN	Ecuadorian Andes	09.02.20	14:07	"Laguna Verde West"	0.0157	-78.0079	4780	RGI60-16.01345	G282007E00013N
GL63_UP	Ecuadorian Andes	10.02.20	12:13	Hermoso	0.0107	-78.0068	4661	RGI60-16.01345	G282007E00013N
GL63_DN	Ecuadorian Andes	10.02.20	15:17	Hermoso	0.0104	-78.0068	4647	RGI60-16.01345	G282007E00013N
GL64_UP	Ecuadorian Andes	11.02.20	11:28	"Laguna Verde East"	0.0161	-78.0055	4894	RGI60-16.01345	G282007E00013N
GL64_DN	Ecuadorian Andes	11.02.20	14:44	"Laguna Verde East"	0.0148	-78.0072	4801	RGI60-16.01345	G282007E00013N
GL65_UP	Ecuadorian Andes	13.02.20	12:00	Antisana 12/ Los Crespos	-0.4941	-78.1587	4746	RGI60-16.01339	G281855E00490S
GL65_DN	Ecuadorian Andes	13.02.20	14:30	Antisana 12/ Los Crespos	-0.4942	-78.1588	4739	RGI60-16.01339	G281855E00490S
GL66_UP	Ecuadorian Andes	14.02.20	12:05	Antisana 15 beta	-0.4724	-78.1511	4872	RGI60-16.01339	G281855E00490S
GL66_DN	Ecuadorian Andes	14.02.20	15:05	Antisana 15 beta	-0.4716	-78.1549	4783	RGI60-16.01339	G281855E00490S

GL67_UP	Ecuadorian Andes	15.02.20	11:44	Antisana 15 alpha	-0.4739	-78.1543	4832	RGI60-16.01339	G281855E00490S
GL67_DN	Ecuadorian Andes	15.02.20	14:15	Antisana 15 alpha	-0.4718	-78.1555	4782	RGI60-16.01339	G281855E00490S
GL68_UP	Ecuadorian Andes	16.02.20	10:30	Antisana 14	-0.4778	-78.1582	4791	RGI60-16.01339	G281855E00490S
GL68_DN	Ecuadorian Andes	16.02.20	14:09	Antisana 14	-0.4773	-78.1592	4778	RGI60-16.01339	G281855E00490S
GL69_UP	Ecuadorian Andes	20.02.20	12:10	"Cotopaxi 1. east of refuge"	-0.6672	-78.4341	4913	RGI60-16.02944	G281572E00688S
GL69_DN	Ecuadorian Andes	20.02.20	14:47	"Cotopaxi 1. east of refuge"	-0.6656	-78.4332	4827	RGI60-16.02944	G281572E00688S
GL70_UP	Ecuadorian Andes	21.02.20	10:50	Cotopaxi Glaciar Baja	-0.6681	-78.4405	5048	RGI60-16.02943	G281559E00671S
GL70_DN	Ecuadorian Andes	21.02.20	14:03	Cotopaxi Glaciar Baja	-0.6674	-78.4408	4987	RGI60-16.02943	G281559E00671S
GL71_UP	Ecuadorian Andes	22.02.20	10:35	"Cotopaxi 2. east of refuge"	-0.6673	-78.4316	4867	RGI60-16.02944	G281572E00688S
GL71_DN	Ecuadorian Andes	22.02.20	13:31	"Cotopaxi 2. east of refuge"	-0.666	-78.4306	4803	RGI60-16.02944	G281572E00688S
GL72_UP	Ecuadorian Andes	23.02.20	10:31	"Cotopaxi 3. east of refuge"	-0.6663	-78.4263	4757	RGI60-16.02944	G281572E00688S
GL72_DN	Ecuadorian Andes	23.02.20	13:48	"Cotopaxi 3. east of refuge"	-0.6658	-78.4262	4746	RGI60-16.02944	G281572E00688S
GL73_UP	Ecuadorian Andes	27.02.20	11:25	"Chimborazo Rock Glacier"	-1.4612	-78.7769	4541	RGI60-16.01311	G281198E01465S
GL73_DN	Ecuadorian Andes	27.02.20	14:05	"Chimborazo Rock Glacier"	-1.4605	-78.773	4471	RGI60-16.01311	G281198E01465S
GL74_UP	Ecuadorian Andes	29.02.20	11:00	"Carihuarazo"	-1.4051	-78.7559	4764	RGI60-16.01319	G281243E01403S
GL74_DN	Ecuadorian Andes	29.02.20	13:54	"Carihuarazo"	-1.4062	-78.7566	4725	RGI60-16.01319	G281243E01403S
GL75_UP	Ecuadorian Andes	01.03.20	11:34	"Chimborazo North"	-1.4462	-78.804	4936	RGI60-16.01308	G281194E01453S
GL75_DN	Ecuadorian Andes	01.03.20	14:07	"Chimborazo North"	-1.4456	-78.8039	4935	RGI60-16.01308	G281194E01453S
GL76_UP	European Alps	02.07.20	10:19	Trift VS	46.1358	7.9866	2868	RGI60-11.02540	G008002E46138N
GL76_DN	European Alps	02.07.20	13:13	Trift VS	46.1361	7.9823	2788	RGI60-11.02540	G008002E46138N

GL77_UP	European Alps	04.07.20	11:37	Forno	46.3352	9.7018	2254	RGI60-11.02245	G009697E46305N
GL77_DN	European Alps	04.07.20	14:31	Forno	46.3431	9.7006	2223	RGI60-11.02245	G009697E46305N
GL78_UP	European Alps	05.07.20	10:40	Albigna	46.3131	9.6464	2176	RGI60-11.02285	G009641E46297N
GL78_DN	European Alps	05.07.20	13:20	Albigna	46.3151	9.6468	2165	RGI60-11.02285	G009641E46297N
GL79_UP	European Alps	06.07.20	09:49	Morteratsch	46.4197	9.9338	2176	RGI60-11.01946	G009641E46297N
GL79_DN	European Alps	06.07.20	13:15	Morteratsch	46.4227	9.9335	2063	RGI60-11.01946	G009641E46297N
GL80_UP	European Alps	07.07.20	10:53	Roseg	46.385	9.8419	2276	RGI60-11.02119	G009860E46370N
GL80_DN	European Alps	07.07.20	13:48	Roseg	46.3909	9.8443	2161	RGI60-11.02119	G009860E46370N
GL81_UP	European Alps	08.07.20	11:30	Tschierva	46.4035	9.8695	2326	RGI60-11.02051	G009886E46384N
GL81_DN	European Alps	08.07.20	15:20	Tschierva	46.4122	9.8576	2098	RGI60-11.02051	G009886E46384N
GL82_UP	European Alps	10.07.20	09:45	Silvretta	46.8559	10.0569	2474	RGI60-11.00804	G010084E46850N
GL82_DN	European Alps	10.07.20	12:28	Silvretta	46.8545	10.0542	2430	RGI60-11.00804	G010084E46850N
GL83_UP	European Alps	12.07.20	10:50	Hintereis F.	46.8169	10.7994	2489	RGI60-11.00897	G010758E46800N
GL83_DN	European Alps	12.07.20	14:00	Hintereis F.	46.8199	10.8056	2410	RGI60-11.00897	G010758E46800N
GL84_UP	European Alps	13.07.20	11:32	Niederjoch F.	46.7752	10.8593	2930	RGI60-11.00992	G010867E46769N
GL84_DN	European Alps	13.07.20	14:44	Niederjoch F.	46.7792	10.8603	2829	RGI60-11.00992	G010867E46769N
GL85_UP	European Alps	14.07.20	10:10	Tiefenbach F.	46.9147	10.9339	2929	RGI60-11.00674	G010927E46919N
GL85_DN	European Alps	14.07.20	12:24	Tiefenbach F.	46.9149	10.9356	2870	RGI60-11.00674	G010927E46919N

GL86_UP	Scandinavian Mountains	07.08.20	12:07	Westbreen	69.4734	20.0001	393	RGI60-08.00335	G019954E69491N
GL86_DN	Scandinavian Mountains	07.08.20	14:48	Westbreen	69.4723	20.0078	351	RGI60-08.00335	G019954E69491N
GL87_UP	Scandinavian Mountains	08.08.20	11:30	Midbreen	69.4603	19.9521	517	RGI60-08.03209	G019923E69469N
GL87_DN	Scandinavian Mountains	08.08.20	15:06	Midbreen	69.4584	19.9596	400	RGI60-08.03209	G019923E69469N
GL88_UP	Scandinavian Mountains	09.08.20	12:00	"Southeast of Fugldalsvatnet"	69.4854	19.8359	446	RGI60-08.03199	G019849E69478N
GL88_DN	Scandinavian Mountains	09.08.20	14:00	"Southeast of Fugldalsvatnet"	69.4869	19.8298	385	RGI60-08.03199	G019849E69478N
GL89_UP	Scandinavian Mountains	12.08.20	11:30	"Northwest of Blaaisen"	69.4748	19.7763	946	RGI60-08.03193	G019786E69474N
GL89_DN	Scandinavian Mountains	12.08.20	14:30	"Northwest of Blaaisen"	69.4749	19.7736	939	RGI60-08.03193	G019786E69474N
GL90_UP	Scandinavian Mountains	20.08.20	09:55	Tuftebreen	61.6586	7.1539	859	RGI60-08.01125	G007087E61677N
GL90_DN	Scandinavian Mountains	20.08.20	12:40	Tuftebreen	61.6565	7.1551	792	RGI60-08.01125	G007087E61677N
GL91_UP	Scandinavian Mountains	21.08.20	09:46	Fabergstolbreen	61.7124	7.2921	779	RGI60-08.01133	G007202E61739N
GL91_DN	Scandinavian Mountains	21.08.20	12:32	Fabergstolbreen	61.7132	7.2963	716	RGI60-08.01133	G007202E61739N
GL92_UP	Scandinavian Mountains	22.08.20	09:50	Nigardsbreen	61.6774	7.2073	363	RGI60-08.01126	G007099E61715N
GL92_DN	Scandinavian Mountains	22.08.20	12:11	Nigardsbreen	61.676	7.2118	275	RGI60-08.01126	G007099E61715N
GL93_UP	Scandinavian Mountains	23.08.20	10:02	Boverbreen	61.5564	8.0515	1447	RGI60-08.02144	G008095E61549N
GL93_DN	Scandinavian Mountains	23.08.20	12:38	Boverbreen	61.5571	8.0435	1368	RGI60-08.02144	G008095E61549N
GL94_UP	Scandinavian Mountains	24.08.20	09:57	Storbreen	61.5818	8.1625	1441	RGI60-08.00312	G008132E61573N
GL94_DN	Scandinavian Mountains	24.08.20	12:22	Storbreen	61.5841	8.1654	1365	RGI60-08.00312	G008132E61573N
GL95_UP	Scandinavian Mountains	25.08.20	10:46	Storjuvbreen	61.6653	8.2987	1396	RGI60-08.00860	G008289E61636N
GL95_DN	Scandinavian Mountains	25.08.20	12:50	Storjuvbreen	61.6679	8.2997	1357	RGI60-08.00860	G008289E61636N

GL96_UP	European Alps	15.09.20	11:00	Miage	45.7883	6.8885	1812	RGI60-11.03005	G006846E45813N
GL96_DN	European Alps	15.09.20	14:00	Miage	45.7894	6.891	1766	RGI60-11.03005	G006846E45813N
GL97_UP	European Alps	16.09.20	12:36	Très la tête	45.7888	6.7484	2037	RGI60-11.03651	G006784E45784N
GL97_DN	European Alps	16.09.20	15:00	Très la tête	45.7892	6.747	2014	RGI60-11.03651	G006784E45784N
GL98_UP	European Alps	17.09.20	10:44	Mer de glace	45.9357	6.9223	1591	RGI60-11.03643	G006934E45883N
GL98_DN	European Alps	17.09.20	13:29	Mer de glace	45.9375	6.9224	1525	RGI60-11.03643	G006934E45883N
GL99_UP	European Alps	18.09.20	09:40	Pélerins	45.8973	6.8842	2244	RGI60-11.03389	G006889E45892N
GL99_DN	European Alps	18.09.20	12:18	Pélerins	45.8982	6.8827	2205	RGI60-11.03389	G006889E45892N
GL100_UP	European Alps	29.09.20	12:17	Chardon	44.8843	6.2934	2301	RGI60-11.03817	G006302E44881N
GL100_DN	European Alps	29.09.20	15:00	Chardon	44.8863	6.2941	2234	RGI60-11.03817	G006302E44881N
GL101_UP	European Alps	30.09.20	11:59	Des étages	44.8985	6.2624	2440	RGI60-11.03694	G006263E44892N
GL101_DN	European Alps	30.09.20	15:30	Des étages	44.9148	6.2582	2008	RGI60-11.03694	G006263E44892N
GL102_UP	European Alps	01.10.20	11:50	Bonne Pierre	44.9358	6.3183	2436	RGI60-11.03810	G006341E44929N
GL102_DN	European Alps	01.10.20	15:36	Bonne Pierre	44.9387	6.2986	1880	RGI60-11.03810	G006341E44929N
GL103_UP	Himalayas	17.03.21	11:16	Lirung	28.2298	85.56236	4062	RGI60-15.04045	G085556E28239N
GL103_DN	Himalayas	17.03.21	13:50	Lirung	28.22356	85.56226	3989	RGI60-15.04045	G085556E28239N
GL104_UP	Himalayas	18.03.21	10:20	Kyimoshung	28.23401	85.57678	4369	RGI60-15.04075	G085573E28258N
GL104_DN	Himalayas	18.03.21	13:15	Kyimoshung	28.23195	85.57248	4197	RGI60-15.04075	G085573E28258N

GL105_UP	Himalayas	21.03.21	11:07	Langtang	28.23321	85.69581	4488	RGI60-15.04121	G085670E28312N
GL105_DN	Himalayas	21.03.21	13:44	Langtang	28.23187	85.69202	4442	RGI60-15.04121	G085670E28312N
GL106_UP	Himalayas	22.03.21	09:35	Langshisha	28.19602	85.69055	4426	RGI60-15.04176	G085747E28200N
GL106_DN	Himalayas	22.03.21	13:00	Langshisha	28.20647	85.67937	4234	RGI60-15.04176	G085747E28200N
GL107_UP	Himalayas	23.03.21	09:00	Shalbachum	28.21388	85.66135	4255	RGI60-15.04119	G085645E28262N
GL107_DN	Himalayas	23.03.21	11:39	Shalbachum	28.20881	85.66161	4179	RGI60-15.04119	G085645E28262N
GL108_UP	Himalayas	04.04.21	09:45	Gangapurna	28.65324	84.00706	3630	RGI60-15.04768	G083997E28607N
GL108_DN	Himalayas	04.04.21	12:24	Gangapurna	28.6595	84.01666	3508	RGI60-15.04768	G083997E28607N
GL109_UP	Himalayas	05.04.21	09:03	Bhakra	28.64621	84.04721	3543	RGI60-15.04770	G084033E28623N
GL109_DN	Himalayas	05.04.21	12:00	Bhakra	28.64778	84.04729	3496	RGI60-15.04770	G084033E28623N
GL110_UP	Himalayas	07.04.21	11:00	Chulu West	28.74877	83.98823	4829	RGI60-15.04495	G084014E28747N
GL110_DN	Himalayas	07.04.21	13:03	Chulu West	28.74698	83.9882	4762	RGI60-15.04495	G084014E28747N
GL111_UP	Himalayas	09.04.21	13:00	"above TP high camp"	28.77912	83.95431	5093	RGI60-15.04449	G083940E28774N
GL111_DN	Himalayas	09.04.21	10:30	"above TP high camp"	28.78457	83.96045	5003	RGI60-15.04449	G083940E28774N
GL112_UP	Himalayas	10.04.21	10:50	Purpung Himal	28.78145	83.99191	4931	RGI60-15.04748	G084004E28793N
GL112_DN	Himalayas	10.04.21	13:30	Purpung Himal	28.78178	83.98731	4856	RGI60-15.04748	G084004E28793N
GL113_UP	Himalayas	11.04.21	10:17	"Icefall Glacier"	28.78933	83.97688	4893	RGI60-15.04731	G083992E28799N
GL113_DN	Himalayas	11.04.21	13:05	"Icefall Glacier"	28.78656	83.97519	4758	RGI60-15.04731	G083992E28799N
GL114_UP	Himalayas	12.04.21	09:46	Chulu Northwest	28.77362	83.99507	4884	RGI60-15.04760	G084014E28763N
GL114_DN	Himalayas	12.04.21	12:05	Chulu Northwest	28.77479	83.99102	4864	RGI60-15.04760	G084014E28763N

GL115_UP	Himalayas	26.04.21	10:30	Nare	27.8352	86.82957	4528	RGI60-15.03572	G086868E27821N
GL115_DN	Himalayas	26.04.21	14:15	Nare	27.84686	86.82095	4425	RGI60-15.03572	G086868E27821N
GL116_UP	Himalayas	28.04.21	09:45	Lhotse	27.90124	86.87882	4815	RGI60-15.03742	G086915E27927N
GL116_DN	Himalayas	28.04.21	12:14	Lhotse	27.90361	86.8726	4732	RGI60-15.03742	G086915E27927N
GL117_UP	Himalayas	29.04.21	09:30	Nuptse	27.9115	86.86557	4973	RGI60-15.03735	G086868E27946N
GL117_DN	Himalayas	29.04.21	12:41	Nuptse	27.91094	86.86544	4966	RGI60-15.03735	G086868E27946N
GL118_UP	Himalayas	30.04.21	10:00	Lhotse Nup	27.91156	86.88319	4960	RGI60-15.03731	G086891E27940N
GL118_DN	Himalayas	30.04.21	13:08	Lhotse Nup	27.90567	86.87194	4735	RGI60-15.03731	G086891E27940N
GL119_UP	Himalayas	02.05.21	09:50	Cholo	27.91355	86.80015	4391	RGI60-15.03739	G086786E27909N
GL119_DN	Himalayas	02.05.21	11:57	Cholo	27.9122	86.80112	4367	RGI60-15.03739	G086786E27909N
GL120_UP	Pamir & Tien Shan	19.07.21	08:00	Aksai	42.5326	74.5273	3323	RGI60-13.11414	G074544E42510N
GL120_DN	Pamir & Tien Shan	19.07.21	09:00	Aksai	42.5328	74.527	3319	RGI60-13.11414	G074544E42510N
GL121_UP	Pamir & Tien Shan	20.07.21	12:49	Top Karagay	42.5056	74.5048	3334	RGI60-13.11604	G074547E42486N
GL121_DN	Pamir & Tien Shan	20.07.21	14:15	Top Karagay	42.5095	74.5014	3267	RGI60-13.11604	G074547E42486N
GL122_UP	Pamir & Tien Shan	21.07.21	12:20	Golubin	42.4755	74.4829	3371	RGI60-13.11609	G074498E42454N
GL122_DN	Pamir & Tien Shan	21.07.21	13:30	Golubin	42.4778	74.4816	3288	RGI60-13.11609	G074498E42454N
GL123_UP	Pamir & Tien Shan	22.07.21	10:30	Tuyuk	42.473	74.512	3409	RGI60-13.11606	G074530E42461N
GL123_DN	Pamir & Tien Shan	22.07.21	12:17	Tuyuk	42.4806	74.4998	3241	RGI60-13.11606	G074530E42461N

GL124_UP	Pamir & Tien Shan	23.07.21	13:42	Tokragula	42.4383	74.4595	3514	RGI60-13.11419	G074466E42432N
GL124_DN	Pamir & Tien Shan	23.07.21	16:33	Tokragula	42.4439	74.4549	3401	RGI60-13.11419	G074466E42432N
GL125_UP	Pamir & Tien Shan	24.07.21	11:30	Tokragula northeast	42.4399	74.4641	3593	RGI60-13.11643	G074478E42443N
GL125_DN	Pamir & Tien Shan	24.07.21	14:30	Tokragula northeast	42.4444	74.4555	3394	RGI60-13.11643	G074478E42443N
GL126_UP	Pamir & Tien Shan	31.07.21	11:25	Petrovski	39.4713	72.8967	4020	RGI60-13.13094	G072890E39466N
GL126_DN	Pamir & Tien Shan	31.07.21	15:00	Petrovski	39.4818	72.8957	3824	RGI60-13.13094	G072890E39466N
GL127_UP	Pamir & Tien Shan	01.08.21	11:45	North of Lenin	39.4419	72.8983	3970	RGI60-13.13098	G072884E39433N
GL127_DN	Pamir & Tien Shan	01.08.21	14:15	North of Lenin	39.4461	72.9048	3901	RGI60-13.13098	G072884E39433N
GL128_UP	Pamir & Tien Shan	02.08.21	11:30	No. 197a	39.4424	72.9367	3878	RGI60-13.13252	G072967E39424N
GL128_DN	Pamir & Tien Shan	02.08.21	14:00	No. 197a	39.4459	72.9292	3777	RGI60-13.13252	G072967E39424N
GL129_UP	Pamir & Tien Shan	04.08.21	09:30	"No. 199a Lenin"	39.4441	72.9168	3764	RGI60-13.13251	G072928E39397N
GL129_DN	Pamir & Tien Shan	04.08.21	11:50	"No. 199a Lenin"	39.4498	72.9135	3733	RGI60-13.13251	G072928E39397N
GL130_UP	Pamir & Tien Shan	11.08.21	11:05	Bordu	41.8245	78.1545	3838	RGI60-13.08054	G078175E41813N
GL130_DN	Pamir & Tien Shan	11.08.21	14:23	Bordu	41.8313	78.1378	3721	RGI60-13.08054	G078175E41813N
GL131_UP	Pamir & Tien Shan	12.08.21	11:15	No. 354	41.8091	78.1402	3813	RGI60-13.07064	G078164E41793N
GL131_DN	Pamir & Tien Shan	12.08.21	14:15	No. 354	41.811	78.1244	3686	RGI60-13.07064	G078164E41793N
GL132_UP	Pamir & Tien Shan	13.08.21	11:31	"Southwest of 354"	41.8009	78.1335	3869	RGI60-13.07137	G078141E41787N
GL132_DN	Pamir & Tien Shan	13.08.21	14:15	"Southwest of 354"	41.8034	78.1293	3822	RGI60-13.07137	G078141E41787N
GL133_UP	Pamir & Tien Shan	14.08.21	11:00	"Road 1"	41.8879	77.6879	3899	RGI60-13.06829	G077684E41883N
GL133_DN	Pamir & Tien Shan	14.08.21	14:00	"Road 1"	41.8901	77.6894	3796	RGI60-13.06829	G077684E41883N

GL134_UP	Pamir & Tien Shan	16.08.21	10:35	"North of road"	41.9099	77.715	4041	RGI60-13.07981	G077719E41916N
GL134_DN	Pamir & Tien Shan	16.08.21	13:00	"North of road"	41.9076	77.7135	3965	RGI60-13.07981	G077719E41916N
GL135_UP	Pamir & Tien Shan	17.08.21	13:46	West Suek/ Batysh Sook	41.7985	77.7495	3925	RGI60-13.06974	G077749E41787N
GL135_DN	Pamir & Tien Shan	17.08.21	15:45	West Suek/ Batysh Sook	41.7995	77.75	3900	RGI60-13.06974	G077749E41787N
GL136_UP	Pamir & Tien Shan	18.08.21	10:07	"2nd west of West Suek"	41.7929	77.7189	3950	RGI60-13.06972	G077722E41781N
GL136_DN	Pamir & Tien Shan	18.08.21	12:40	"2nd west of West Suek"	41.7943	77.7181	3926	RGI60-13.06972	G077722E41781N
GL137_UP	Pamir & Tien Shan	19.08.21	11:20	"West of West Suek"	41.7925	77.7353	3958	RGI60-13.06973	G077735E41783N
GL137_DN	Pamir & Tien Shan	19.08.21	14:43	"West of West Suek"	41.8019	77.7369	3874	RGI60-13.06973	G077735E41783N
GL138_UP	Pamir & Tien Shan	20.08.21	10:15	East Suek	41.7878	77.7697	4007	RGI60-13.06975	G077771E41782N
GL138_DN	Pamir & Tien Shan	20.08.21	13:05	East Suek	41.7926	77.7672	3909	RGI60-13.06975	G077771E41782N
GL139_UP	Pamir & Tien Shan	21.08.21	11:00	"Ski slope"	41.8182	77.7974	4014	RGI60-13.06980	G077800E41816N
GL139_DN	Pamir & Tien Shan	21.08.21	13:55	"Ski slope"	41.8245	77.7927	3813	RGI60-13.06980	G077800E41816N
GL140_UP	Rwenzori Mountains	03.12.21	11:36	Stanley 003	0.3758	29.8785	4724	RGI50-16.01631	G029875E00377N
GL140_DN	Rwenzori Mountains	03.12.21	13:00	Stanley 003	0.3755	29.8788	4722	RGI50-16.01631	G029875E00377N
GL141_UP	Chilean Andes	25.01.22	12:27	Ventisquiero Yelcho	-43.2891	-72.4689	378	RGI60-17.11003	G287495E43295S
GL141_DN	Chilean Andes	25.01.22	16:06	Ventisquiero Yelcho	-43.2863	-72.4634	312	RGI60-17.11003	G287495E43295S
GL142_UP	Chilean Andes	26.01.22	13:00	Amarillo	-42.8641	-72.4453	498	RGI60-17.11392	G287544E42827S
GL142_DN	Chilean Andes	26.01.22	16:15	Amarillo	-42.8825	-72.4498	430	RGI60-17.11392	G287544E42827S

GL143_UP	Chilean Andes	30.01.22	10:23	"Mocho 1 middle"	-39.9282	-72.0061	1876	RGI60-17.12442	G287988E39929S
GL143_DN	Chilean Andes	30.01.22	13:50	"Mocho 1 middle"	-39.9261	-72	1740	RGI60-17.12442	G287988E39929S
GL144_UP	Chilean Andes	31.01.22	10:32	"Mocho 2 south"	-39.9305	-72.0029	1825	RGI60-17.12442	G287988E39929S
GL144_DN	Chilean Andes	31.01.22	12:55	"Mocho 2 south"	-39.9288	-71.9997	1738	RGI60-17.12442	G287988E39929S
GL145_UP	Chilean Andes	01.02.22	10:30	"Mocho 3 north"	-39.9252	-72.0079	1854	RGI60-17.12442	G287988E39929S
GL145_DN	Chilean Andes	01.02.22	14:14	"Mocho 3 north"	-39.9244	-72.0064	1775	RGI60-17.12442	G287988E39929S
GL146_UP	Chilean Andes	04.02.22	12:22	Pichillancahue	-39.4407	-71.8884	1817	RGI60-17.12517	G288107E39448S
GL146_DN	Chilean Andes	04.02.22	14:45	Pichillancahue	-39.4394	-71.879	1740	RGI60-17.12517	G288107E39448S
GL147_UP	Chilean Andes	05.02.22	12:00	Turbio	-39.4315	-71.8797	1748	RGI60-17.12535	G288088E39427S
GL147_DN	Chilean Andes	05.02.22	15:30	Turbio	-39.4271	-71.8786	1703	RGI60-17.12535	G288088E39427S
GL148_UP	Chilean Andes	08.02.22	11:45	"Nevado Chillan 1 east"	-36.8385	-71.4148	2734	RGI60-17.13045	G288585E36836S
GL148_DN	Chilean Andes	08.02.22	14:15	"Nevado Chillan 1 east"	-36.8405	-71.4166	2610	RGI60-17.13045	G288585E36836S
GL149_UP	Chilean Andes	09.02.22	09:30	"Nevado Chillan 2 west"	-36.8451	-71.4253	2273	RGI60-17.13045	G288585E36836S
GL149_DN	Chilean Andes	09.02.22	12:05	"Nevado Chillan 2 west"	-36.8487	-71.4323	2176	RGI60-17.13045	G288585E36836S
GL150_UP	Chilean Andes	11.02.22	09:22	Universidad	-34.7116	-70.3437	2524	RGI60-17.01218	G289664E34607S
GL150_DN	Chilean Andes	11.02.22	12:30	Universidad	-34.7204	-70.3592	2417	RGI60-17.01218	G289664E34607S
GL151_UP	Chilean Andes	17.02.22	12:45	"El Morado"	-33.736	-70.0439	3227	RGI60-17.13710	G289967E33714S
GL151_DN	Chilean Andes	17.02.22	16:50	"El Morado"	-33.7691	-70.0441	2731	RGI60-17.13710	G289967E33714S
GL152_UP	Chilean Andes	18.02.22	12:50	"Colina"	-33.8605	-69.9229	2987	RGI60-17.13633	G290091E33826S
GL152_DN	Chilean Andes	18.02.22	16:30	"Colina"	-33.8793	-69.9378	2716	RGI60-17.13633	G290091E33826S

179

GL153_UP	Chilean Andes	20.02.22	11:11	"El Morado cliffy"	-33.744	-70.055	3343	RGI60-17.13710	G289967E33714S
GL153_DN	Chilean Andes	20.02.22	14:17	"El Morado cliffy"	-33.744	-70.0527	3293	RGI60-17.13710	G289967E33714S
GL154_UP	Chilean Andes	22.02.22	11:30	"Plomo West"	-33.5739	-69.9214	3495	RGI60-17.13750	G290062E33534S
GL154_DN	Chilean Andes	22.02.22	14:37	"Plomo West"	-33.5825	-69.9175	3359	RGI60-17.13750	G290062E33534S
GL155_UP	Chilean Andes	23.02.22	10:30	"Plomo rock glacier"	-33.6067	-69.9027	3169	RGI60-17.13720	G290109E33559S
GL155_DN	Chilean Andes	23.02.22	13:25	"Plomo rock glacier"	-33.611	-69.9079	3003	RGI60-17.13720	G290109E33559S
GL156_UP	Alaska Range	21.06.22	12:43	Raven	61.066	-149.113	957	RGI60-01.08876	G210925E61062N
GL156_DN	Alaska Range	21.06.22	15:15	Raven	61.07	-149.12	870	RGI60-01.08876	G210925E61062N
GL157_UP	Alaska Range	22.06.22	12:06	Leanard	60.7904	-148.725	233	RGI60-01.09507	G211285E60812N
GL157_DN	Alaska Range	22.06.22	15:56	Leanard	60.784	-148.717	0	RGI60-01.09507	G211285E60812N
GL158_UP	Alaska Range	23.06.22	10:15	Porcupine	59.9992	-149.281	231	RGI60-01.08618	G210738E60001N
GL158_DN	Alaska Range	23.06.22	13:55	Porcupine	59.9999	-149.292	46	RGI60-01.08618	G210738E60001N
GL159_UP	Alaska Range	24.06.22	09:58	Byron	60.7588	-148.852	97	RGI60-01.09245	G211144E60742N
GL159_DN	Alaska Range	24.06.22	13:20	Byron	60.7677	-148.84	57	RGI60-01.09245	G211144E60742N
GL160_UP	Alaska Range	25.06.22	11:50	Milk	61.0454	-149.095	1123	RGI60-01.08875	G210925E61049N
GL160_DN	Alaska Range	25.06.22	14:30	Milk	61.0433	-149.096	1062	RGI60-01.08875	G210925E61049N
GL161_UP	Alaska Range	26.06.22	10:30	"South of Fourth of July Creek"	60.0866	-149.233	538	RGI60-01.08662	G210784E60074N
GL161_DN	Alaska Range	26.06.22	13:15	"South of Fourth of July Creek"	60.088	-149.235	481	RGI60-01.08662	G210784E60074N

GL162_UP	Alaska Range	01.07.22	09:45	"Second west of Powell"	61.6413	-147.268	1520	RGI60-01.23597	G212739E61620N
GL162_DN	Alaska Range	01.07.22	12:20	"Second west of Powell"	61.6433	-147.267	1476	RGI60-01.23597	G212739E61620N
GL163_UP	Alaska Range	02.07.22	09:15	Powell	61.6704	-147.291	949	RGI60-01.10655	G212799E61568N
GL163_DN	Alaska Range	02.07.22	12:15	Powell	61.6736	-147.295	858	RGI60-01.10655	G212799E61568N
GL164_UP	Alaska Range	03.07.22	09:15	"Third west of Powell"	61.6701	-147.309	894	RGI60-01.10621	G212686E61576N
GL164_DN	Alaska Range	03.07.22	12:00	"Third west of Powell"	61.6733	-147.31	848	RGI60-01.10621	G212686E61576N
GL165_UP	Alaska Range	04.07.22	10:00	Nelchina	61.709	-147.074	758	RGI60-01.10683	G213074E61558N
GL165_DN	Alaska Range	04.07.22	12:11	Nelchina	61.7156	-147.072	780	RGI60-01.10683	G213074E61558N
GL166_UP	Alaska Range	05.07.22	09:45	Sylvester	61.6779	-147.067	882	RGI60-01.23046	G212867E61496N
GL166_DN	Alaska Range	05.07.22	12:00	Sylvester	61.683	-147.073	808	RGI60-01.23046	G212867E61496N
GL167_UP	Alaska Range	06.07.22	11:19	Matanuska	61.7752	-147.762	472	RGI60-01.10557	G212412E61677N
GL167_DN	Alaska Range	06.07.22	14:00	Matanuska	61.7877	-147.796	454	RGI60-01.10557	G212412E61677N
GL168_UP	Alaska Range	12.07.22	11:15	Rainbow Cirque	63.3208	-145.611	1496	RGI60-01.24172	G214393E63316N
GL168_DN	Alaska Range	12.07.22	13:43	Rainbow Cirque	63.3247	-145.612	1373	RGI60-01.24172	G214393E63316N
GL169_UP	Alaska Range	13.07.22	09:55	Castner	63.4044	-145.698	816	RGI60-01.00561	G214527E63453N
GL169_DN	Alaska Range	13.07.22	12:30	Castner	63.4024	-145.714	772	RGI60-01.00561	G214527E63453N
GL170_UP	Alaska Range	14.07.22	11:00	Gulkana	63.2551	-145.425	1300	RGI60-01.00570	G214573E63281N
GL170_DN	Alaska Range	14.07.22	14:22	Gulkana	63.2508	-145.445	1174	RGI60-01.00570	G214573E63281N

Data table for sampling sites

Supplementary table 4. 2

Variable	Scenario	Training r^2	Validation $r^2_{prediction}$	N
pc_water_temp	ssp126	0.483694151739488	0.431706502760076	164
pc_water_temp	ssp370	0.483694151739488	0.431706502760076	164
pc_water_temp	ssp585	0.483694151739488	0.431706502760076	164
pc_turbidity	ssp126	0.522411723273106	0.429091161575562	161
pc_turbidity	ssp370	0.522414448424744	0.428957245578879	161
pc_turbidity	ssp585	0.522202567290839	0.428978529364977	161
pc_conductivity	ssp126	0.869520616025708	0.766117596435234	163
pc_conductivity	ssp370	0.869520616025708	0.766117596435234	163
pc_conductivity	ssp585	0.869520616025708	0.766117596435234	163
pc_ph	ssp126	0.700647408099518	0.599720731901965	161
pc_ph	ssp370	0.700513886974243	0.599631224909152	161
pc_ph	ssp585	0.70078533949249	0.600215881508188	161
nut_din	ssp126	0.622775909442637	0.542342181146603	163
nut_din	ssp370	0.623086344263876	0.542454226500516	163
nut_din	ssp585	0.623120952549497	0.542645473717078	163
nut_srp	ssp126	0.588006632675294	0.472814871086123	161
nut_srp	ssp370	0.588021609478706	0.473145106031582	161
nut_srp	ssp585	0.587511702635033	0.47256766896558	161
chla	ssp126	0.583311636837417	0.354547100728673	163

chla	ssp370	0.583155761657876	0.35413988375086	163
chla	ssp585	0.586000614809754	0.353663066988383	163
bacterial_abundance	ssp126	0.443688767675114	0.229704742810531	161
bacterial_abundance	ssp370	0.444758258561565	0.229889068745514	161
bacterial_abundance	ssp585	0.444131144960131	0.229262301469782	161
Shannon	ssp126	0.411848027182637	0.221587052270616	81
Shannon	ssp370	0.411848027182637	0.221587052270616	81
Shannon	ssp585	0.411848027182637	0.221587052270616	81
Pielou	ssp126	0.411572707116054	0.223077856420605	81
Pielou	ssp370	0.411572707116054	0.223077856420605	81
Pielou	ssp585	0.411572707116054	0.223077856420605	81
mntd	ssp126	0.721446539098174	0.517803977019936	81
mntd	ssp370	0.721329393408036	0.519250882692198	81
mntd	ssp585	0.721505243668122	0.518039745055696	81
mpd	ssp126	0.312694729238186	0.0897716609737202	81
mpd	ssp370	0.315673194807295	0.0832937911325424	81
mpd	ssp585	0.315344035545402	0.0816300231165704	81

Statistics for the proximal drivers and biomass models

Supplementary table 4. 3

Scenario	MeanIncrease	MeanDecrease	MeanNotSignificant	Q25	Median	Q75
SSP1	0.611	0.266	0.123	-0.061	0.158	0.495
SSP3	0.606	0.265	0.129	-0.124	0.297	0.890
SSP5	0.603	0.268	0.129	-0.143	0.348	0.999

Proportion of strains that increase/decrease under future scenarios SSP1, 3 and 5. Median and IQR of relative changes for all scenarios across the 2333 strains modelled under future scenarios SSP1, 3 and 5.

Supplementary table 4. 4

Variable	p	logl	logl0	lambda
$\log_2$fold-change	2.18527938781675E-230	-2705.00948543317	-3230.1171965599	0.864908538330015
$r^2_{prediction}$	9.48870948533819E-108	1283.39493194868	1040.28700843249	0.778645212567146

Phylogenetic signal of future changes ($\log_2$ fold-change) and predictability for SSP3.

Supplementary table 4. 5

Phylogenetic depth	N	Taxonomy	Relative phylogenetic depth
1.2570467	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Rhodoferax;s__	0.225403659
1.243231715	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polaromonas;s__	0.233916499
1.317136912	4	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Aquabacterium_A;s__	0.188375872
1.273767246	8	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__LMDS01;s__	0.215100404
0.986627653	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__	0.392036772
1.279421506	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__Methylotenera_A;s__	0.21161623
1.223037269	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Casimicrobiaceae;g__Casimicrobium;s__	0.246360383
1.195615144	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__SZUA-5;g__;s__	0.263257988
1.279635137	5	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__SCMT01;s__	0.21148459
0.908188277	3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Nevskiales;f__Nevskiaceae;g__	0.440371375
1.194060184	3	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Acetobacterales;f__Acetobacteraceae;g__;s__	0.264216159
1.232826515	4	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomicrobium;s__	0.240328218
1.250610621	3	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Phenylobacterium;s__	0.229369593
1.055489776	7	d__Bacteria;p__Acidobacteriota;c__Blastocatellia;o__Pyrinomonadales;f__Pyrinomonadaceae;g__	0.349603704

1.198249271	3	d__Bacteria;p__Acidobacteriota;c__Acidobacteriae;o__Bryobacter ales;f__Bryobacteraceae;g__;s__	0.26163483
1.121746336	3	d__Bacteria;p__Bdellovibrionota;c__Bdellovibrionia;o__Bdellovibri onales;f__Bdellovibrionaceae;g__;s__	0.308776192
1.367694437	5	d__Bacteria;p__Bdellovibrionota;c__Bdellovibrionia;o__Bdellovibri onales;f__Bdellovibrionaceae;g__Pseudobdellovibrio;s__	0.157222157
1.010345614	3	d__Bacteria;p__Myxococcota;c__Polyangia;o__Polyangiales;f__Po lyangiaceae;g__	0.377421686
1.218611259	4	d__Bacteria;p__Myxococcota;c__Polyangia;o__Polyangiales;f__Po lyangiaceae;g__	0.249087705
1.194489325	3	d__Bacteria;p__Myxococcota;c__Polyangia;o__Polyangiales;f__Po lyangiaceae;g__	0.263951721
1.112980418	3	d__Bacteria;p__Myxococcota;c__Myxococcia;o__Myxococcales;f_ _Myxococcaceae;g__JAEUJT01;s__	0.31417778
1.250220875	4	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Sphingobacteriale s;f__	0.229609755
1.350284278	4	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f __Chitinophagaceae;g__	0.167950355
1.369200672	3	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f __Chitinophagaceae;g__Ferruginibacter;s__	0.15629401
1.398513364	5	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f __Chitinophagaceae;g__Ferruginibacter;s__	0.138231432
1.408452528	3	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f __Chitinophagaceae;g__Ferruginibacter;s__	0.132106887
1.401133269	3	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Chitinophagales;f __Saprospiraceae;g__M3007;s__	0.136617038
1.122756325	3	d__Bacteria;p__Planctomycetota;c__Planctomycetia;o__Isosphaer ales;f__Isosphaeraceae;g__;s__	0.308153833
1.216809843	3	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verruc omicrobiales;f__Verrucomicrobiaceae;g__VFKE01;s__	0.250197744
0.459872106	4	d__Bacteria;p__Chloroflexota;c__Chloroflexia;o__	0.716625285
1.293865648	3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrob iales;f__RAAP-2;g__RAAP-2;s__	0.202715702
0.952046238	15	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Mycobacter iales;f__SCTD01;g__	0.413345954

1.105178794	8	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomyce tales;f__Dermatophilaceae;g__UBA4719;s__	0.318985166
1.358744183	13	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomyce tales;f__Microbacteriaceae;g__Lacisediminihabitans;s__	0.162737333
1.024311732	3	d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__Cyanobacter iales;f__Microcoleaceae;g__Microcoleus;s__	0.368815718
0.911200145	7	d__Bacteria;p__Armatimonadota;c__Chthonomonadetes;o__Chtho nomonadales;f__Chthonomonadaceae;g__;s__	0.438515452
0.913730191	4	d__Bacteria;p__Deinococcota;c__Deinococci;o__Deinococcales;f_ _Deinococcaceae;g__Deinococcus;s__	0.436956429

Monophyletic clades with all representatives decreasing in the future under RCP4.5

Supplementary table 4. 6

KEGG_ko	n	Description
K03117	10	sec-independent protein translocase protein TatB - Importance for cold-shock in Shewanella oneidensis (https://journals.asm.org/doi/full/10.1128/jb.01908-05)
K03168	10	DNA topoisomerase I topA - cold sensitivity of mutants lacking this gene in E. coli (https://www.jbc.org/article/S0021-9258(20)76598-1/fulltext)
K02860	9	16S rRNA processing protein RimM - expression linked with the cold-adaptation RbfA gene (https://www.annualreviews.org/doi/pdf/10.1146/annurev-biochem-062608-160432)
K03814	9	monofunctional glycosyltransferase mtgA - peptidogylacan maturation (https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/0014-5793%2896%2900809-5)
K13288	9	oligoribonuclease orn, REX2, REXO2 - associated with biofilm formation (https://www.pnas.org/doi/full/10.1073/pnas.1507245112)
K00140	8	malonate-semialdehyde dehydrogenase (acetylating) / methylmalonate-semialdehyde dehydrogenase mmsA, iolA, ALDH6A1 - differentially expressed at cold temperatures in Psychrobacter (https://ami-journals.onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.13578)
K00847	8	fructokinase scrK
K01800	8	maleylacetoacetate isomerase maiA, GSTZ1
K02339	8	DNA polymerase III subunit chi holC
K03179	8	4-hydroxybenzoate polyprenyltransferase ubiA
K03216	8	tRNA (cytidine/uridine-2'-O-)-methyltransferase trmL, cspR - tRNA and cold adaptation (https://www.mdpi.com/2218-273X/7/2/35)
K03558	8	membrane protein required for colicin V production cvpA - bacteriocin (https://journals.asm.org/doi/full/10.1128/mmbr.00036-06)
K03821	8	poly[(R)-3-hydroxyalkanoate] polymerase subunit phaC, phbC - PHAs in the baltic sea, adaptation to cold (https://link.springer.com/article/10.1007/s00792-014-0699-9, https://link.springer.com/article/10.1007/s00792-014-0699-9#ref-CR7)
K04760	8	transcription elongation factor greB
K06195	8	ApaG protein apaG
K07396	8	putative protein-disulfide isomerase, participates in the insertion of lipopolysaccharides in the outer membrane (https://www.sciencedirect.com/science/article/pii/S0021925820526628?via%3Dihub)
K07518	8	hydroxybutyrate-dimer hydrolase
K09748	8	ribosome maturation factor rimP - associated with the cold-adaptation RbfA gene, important at higher temperatures (https://www.sciencedirect.com/science/article/pii/S0022283608016148, https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.052209-0)
K11443	8	two-component system, cell cycle response regulator DivK
K14472	8	succinyl-CoA:(S)-malate CoA-transferase subunit B, smtB - carbon metabolism
K14540	8	ribosome biogenesis GTPase A rbgA

Top KEGG orthologous groups highlighted in the phylogenetic random forest analysis.

Supplementary table 4. 7

Pathway	p-value	Odds ratio	Adjusted p-value
09101 Carbohydrate metabolism	5.67364858822658E-12	2.19814468727323	1.02125674588078E-10
09102 Energy metabolism	0.000157647561013464	1.79852401019023	0.00283765609824235
09103 Lipid metabolism	1	0.983936076105019	1
09104 Nucleotide metabolism	0.0263429482376568	1.65424202906632	0.474173068277823
09105 Amino acid metabolism	0.0684907886634533	1.2970060651012	1
09106 Metabolism of other amino acids	0.28995021507565	1.32556634065616	1
09107 Glycan biosynthesis and metabolism	0.147697932724234	0.625876910303831	1
09108 Metabolism of cofactors and vitamins	0.359915570092523	1.18853793964322	1
09109 Metabolism of terpenoids and polyketides	0.278403713568717	1.31358379332405	1
09110 Biosynthesis of other secondary metabolites	0.628511243750385	0.806217749825693	1
09111 Xenobiotics biodegradation and metabolism	0.442825854647123	1.17259731591119	1
09121 Transcription	0.227407317237828	2.61145370535302	1
09182 Protein families: genetic information processing	0.000374756903304388	0.660246685402945	0.00674562425947898
09183 Protein families: signaling and cellular processes	8.53996481003051E-11	0.419593451114279	1.53719366580549E-09
09191 Unclassified: metabolism	0.517720020642494	0.871955704845446	1
09192 Unclassified: genetic information processing	0.310081669412409	0.54589648188649	1
09193 Unclassified: signaling and cellular processes	0.104096375090835	0.413315645603366	1
09194 Poorly characterized	0.00169446888027172	0.314728576292485	0.030500439844891

Enrichment analysis of the KEGG orthologous groups associated with the strains forecasted to decrease.

Supplementary table 4. 8

Parameter	values
Regularization.factor	0.2,0.4,0.6,0.8,1
Num.trees	250,500,750,1000,1500
Max.depth	5,15,30,50
Regularization.usedepth	0.0263429482376568
mtry	20,30,40,50
splitrule	'hellinger','extratrees','gini'
Sample.fraction	0.5,0.6,0.7,0.8

Ranger random forest analysis hyperparameters random grid.

Supplementary table 5. 1

covariate	resp_var	t	coef	coef_se	p	spatial_f	spatial_p	padj
water_temp	mean_norm_size	1.317	0.074	0.056	0.380	1.809	0.000	1.000
water_temp	mean_GC	-0.763	-0.041	0.054	0.894	2.299	0.000	1.000
water_temp	mean_norm_gene_number	1.404	0.083	0.059	0.325	1.354	0.000	1.000
water_temp	mean_norm_tRNAs	1.473	0.082	0.056	0.286	2.019	0.000	1.000
water_temp	mean_redundancy_index	3.788	0.221	0.058	0.000	0.822	0.000	0.014
water_temp	mean_coding_density	0.286	0.014	0.049	1.551	3.382	0.000	1.000
chla	mean_norm_size	6.457	0.229	0.036	0.000	1.144	0.000	0.000
chla	mean_GC	-1.949	-0.077	0.039	0.107	1.980	0.000	1.000
chla	mean_norm_gene_number	5.190	0.198	0.038	0.000	0.908	0.000	0.000
chla	mean_norm_tRNAs	4.842	0.187	0.039	0.000	1.782	0.000	0.000
chla	mean_redundancy_index	0.889	0.038	0.042	0.751	1.044	0.000	1.000
chla	mean_coding_density	-2.907	-0.103	0.035	0.009	2.979	0.000	0.256
gl_area	mean_norm_size	-2.250	-0.119	0.053	0.052	1.749	0.000	1.000
gl_area	mean_GC	-0.468	-0.025	0.053	1.281	1.966	0.000	1.000
gl_area	mean_norm_gene_number	-2.374	-0.130	0.055	0.038	1.301	0.000	0.950
gl_area	mean_norm_tRNAs	-1.368	-0.075	0.055	0.347	2.001	0.000	1.000
gl_area	mean_redundancy_index	0.357	0.020	0.057	1.443	0.988	0.000	1.000
gl_area	mean_coding_density	0.403	0.019	0.048	1.375	3.399	0.000	1.000
gl_dist	mean_norm_size	0.612	0.027	0.044	1.083	1.744	0.000	1.000
gl_dist	mean_GC	-0.338	-0.015	0.043	1.472	2.100	0.000	1.000
gl_dist	mean_norm_gene_number	0.912	0.042	0.046	0.727	1.329	0.000	1.000
gl_dist	mean_norm_tRNAs	2.438	0.106	0.044	0.032	2.091	0.000	0.837
gl_dist	mean_redundancy_index	3.849	0.174	0.045	0.000	1.012	0.000	0.012
gl_dist	mean_coding_density	1.374	0.053	0.039	0.344	3.526	0.000	1.000
gl_cov	mean_norm_size	-1.519	-0.558	0.368	0.262	1.850	0.000	1.000

resp_var		t	coef	coef_se	p	spatial_f	spatial_p	padj
gl_cov	mean_GC	1.211	0.441	0.364	0.456	2.193	0.000	1.000
gl_cov	mean_norm_gene_number	-1.388	-0.526	0.379	0.335	1.345	0.000	1.000
gl_cov	mean_norm_tRNAs	-2.611	-0.969	0.371	0.020	2.005	0.000	0.564
gl_cov	mean_redundancy_index	-2.510	-0.946	0.377	0.026	0.849	0.000	0.714
gl_cov	mean_coding_density	0.237	0.078	0.332	1.626	3.276	0.000	1.000
gl_index	mean_norm_size	-1.478	-0.065	0.044	0.283	1.865	0.000	1.000
gl_index	mean_GC	0.170	0.007	0.043	1.731	2.290	0.000	1.000
gl_index	mean_norm_gene_number	-1.836	-0.084	0.046	0.137	1.440	0.000	1.000
gl_index	mean_norm_tRNAs	-2.827	-0.123	0.043	0.011	2.161	0.000	0.315
gl_index	mean_redundancy_index	-3.442	-0.158	0.046	0.002	0.909	0.000	0.047
gl_index	mean_coding_density	-1.162	-0.045	0.039	0.495	3.489	0.000	1.000

GAM models output for the community averages models. resp_var = response variable, t = t-value of the linear effect, coef = coefficient of the linear effect, coef_se = standard error of the estimated coefficient, p = p-value for the linear effect, spatial_f = f-value of the spatial spline, spatial_p = p-value for the spatial spline, padj = adjusted p-value for the linear effect.

Supplementary table 5. 2

cluster	p	median_effect	padj	test
1	0.000	0.493	0.000	norm_size ~ chla
2	0.192	-0.081	1.000	norm_size ~ chla
3	0.640	0.028	1.000	norm_size ~ chla
4	0.925	0.005	1.000	norm_size ~ chla
5	0.134	-0.104	1.000	norm_size ~ chla
6	0.620	0.043	1.000	norm_size ~ chla
7	0.314	0.085	1.000	norm_size ~ chla
8	0.602	0.045	1.000	norm_size ~ chla
9	0.183	0.119	1.000	norm_size ~ chla
10	0.738	-0.019	1.000	norm_size ~ chla
11	0.862	-0.016	1.000	norm_size ~ chla
12	0.989	-0.035	1.000	norm_size ~ chla
13	0.947	0.035	1.000	norm_size ~ chla
14	0.620	0.009	1.000	norm_size ~ chla
15	0.718	-0.013	1.000	norm_size ~ chla
16	0.718	-0.014	1.000	norm_size ~ chla
17	0.718	-0.076	1.000	norm_size ~ chla
18	0.779	-0.069	1.000	norm_size ~ chla
19	0.820	-0.032	1.000	norm_size ~ chla
20	0.862	-0.059	1.000	norm_size ~ chla
21	0.989	-0.029	1.000	norm_size ~ chla
22	0.201	-0.084	1.000	norm_size ~ chla
23	0.547	0.118	1.000	norm_size ~ chla
24	0.925	0.018	1.000	norm_size ~ chla
25	0.862	-0.019	1.000	norm_size ~ chla
26	0.341	-0.084	1.000	norm_size ~ chla
27	0.883	0.007	1.000	norm_size ~ chla
28	0.043	0.150	1.000	norm_size ~ chla
29	0.758	-0.012	1.000	norm_size ~ chla
30	0.000	0.288	0.003	norm_size ~ chla
31	0.841	0.031	1.000	norm_size ~ chla
32	0.968	-0.012	1.000	norm_size ~ chla
33	0.841	0.004	1.000	norm_size ~ chla
1	0.000	0.705	0.000	norm_gene ~ chla
2	0.369	-0.050	1.000	norm_gene ~ chla
3	0.659	0.125	1.000	norm_gene ~ chla

4	0.758	0.134	1.000	norm_gene ~ chla
5	0.005	-0.173	0.169	norm_gene ~ chla
6	0.529	0.098	1.000	norm_gene ~ chla
7	0.529	0.091	1.000	norm_gene ~ chla
8	0.495	0.092	1.000	norm_gene ~ chla
9	0.355	0.147	1.000	norm_gene ~ chla
10	0.799	0.136	1.000	norm_gene ~ chla
11	0.862	0.034	1.000	norm_gene ~ chla
12	0.738	0.092	1.000	norm_gene ~ chla
13	0.718	0.139	1.000	norm_gene ~ chla
14	0.583	0.054	1.000	norm_gene ~ chla
15	0.820	0.139	1.000	norm_gene ~ chla
16	0.820	0.138	1.000	norm_gene ~ chla
17	0.904	-0.018	1.000	norm_gene ~ chla
18	0.841	0.038	1.000	norm_gene ~ chla
19	0.904	-0.018	1.000	norm_gene ~ chla
20	0.862	0.026	1.000	norm_gene ~ chla
21	0.968	0.051	1.000	norm_gene ~ chla
22	0.265	-0.003	1.000	norm_gene ~ chla
23	0.602	0.203	1.000	norm_gene ~ chla
24	0.820	0.092	1.000	norm_gene ~ chla
25	1.000	0.074	1.000	norm_gene ~ chla
26	0.640	-0.011	1.000	norm_gene ~ chla
27	0.738	0.077	1.000	norm_gene ~ chla
28	0.056	0.217	1.000	norm_gene ~ chla
29	0.583	0.085	1.000	norm_gene ~ chla
30	0.000	0.469	0.000	norm_gene ~ chla
31	0.862	0.107	1.000	norm_gene ~ chla
32	0.862	0.030	1.000	norm_gene ~ chla
33	0.758	0.088	1.000	norm_gene ~ chla
1	0.000	0.707	0.000	norm_trna ~ chla
2	0.017	-0.202	0.552	norm_trna ~ chla
3	0.947	-0.061	1.000	norm_trna ~ chla
4	0.968	-0.083	1.000	norm_trna ~ chla
5	0.000	-0.424	0.009	norm_trna ~ chla
6	0.883	-0.076	1.000	norm_trna ~ chla
7	0.968	-0.107	1.000	norm_trna ~ chla
8	1.000	-0.139	1.000	norm_trna ~ chla
9	0.495	-0.043	1.000	norm_trna ~ chla
10	0.355	-0.033	1.000	norm_trna ~ chla
11	0.429	0.018	1.000	norm_trna ~ chla
12	0.369	-0.126	1.000	norm_trna ~ chla
13	0.678	-0.078	1.000	norm_trna ~ chla
14	0.149	0.036	1.000	norm_trna ~ chla

15	0.355	-0.006	1.000	norm_trna ~ chla
16	0.355	-0.006	1.000	norm_trna ~ chla
17	0.738	0.016	1.000	norm_trna ~ chla
18	0.547	-0.099	1.000	norm_trna ~ chla
19	0.301	-0.141	1.000	norm_trna ~ chla
20	0.221	-0.108	1.000	norm_trna ~ chla
21	0.989	-0.056	1.000	norm_trna ~ chla
22	0.779	-0.109	1.000	norm_trna ~ chla
23	0.779	-0.075	1.000	norm_trna ~ chla
24	0.758	-0.098	1.000	norm_trna ~ chla
25	0.583	-0.006	1.000	norm_trna ~ chla
26	0.165	0.048	1.000	norm_trna ~ chla
27	0.461	-0.112	1.000	norm_trna ~ chla
28	0.000	0.240	0.001	norm_trna ~ chla
29	0.301	-0.094	1.000	norm_trna ~ chla
30	0.495	-0.075	1.000	norm_trna ~ chla
31	0.989	-0.032	1.000	norm_trna ~ chla
32	0.698	-0.079	1.000	norm_trna ~ chla
33	0.989	-0.088	1.000	norm_trna ~ chla

Leave-one-cluster-out tests output for the models with chlorophyll-a as covariate. cluster = identifier of the phylogenetic cluster, p = p-value of the Wilcoxon signed rank test, median_effect = relative median difference in coefficient, padj = adjusted p-value, test = relationship that was tested.

Supplementary table 5. 3

cluster	p	median_effect	padj	test
1	0.017	0.285	1.000	KO redundancy ~ gl_dist
2	0.000	0.656	0.047	KO redundancy ~ gl_dist
3	0.002	0.435	0.853	KO redundancy ~ gl_dist
4	0.000	0.869	0.031	KO redundancy ~ gl_dist
5	0.001	0.555	0.410	KO redundancy ~ gl_dist
6	0.000	0.566	0.186	KO redundancy ~ gl_dist
7	0.001	0.513	0.263	KO redundancy ~ gl_dist
8	0.000	0.604	0.147	KO redundancy ~ gl_dist
9	0.002	0.395	0.943	KO redundancy ~ gl_dist
10	0.017	0.287	1.000	KO redundancy ~ gl_dist
11	0.102	0.289	1.000	KO redundancy ~ gl_dist
12	0.012	0.518	1.000	KO redundancy ~ gl_dist
13	0.327	0.115	1.000	KO redundancy ~ gl_dist
14	0.004	0.461	1.000	KO redundancy ~ gl_dist
15	0.000	0.934	0.047	KO redundancy ~ gl_dist
16	0.004	0.695	1.000	KO redundancy ~ gl_dist
17	0.010	0.476	1.000	KO redundancy ~ gl_dist
18	0.000	0.869	0.004	KO redundancy ~ gl_dist
19	0.018	0.486	1.000	KO redundancy ~ gl_dist
20	0.000	0.759	0.036	KO redundancy ~ gl_dist
21	0.033	0.322	1.000	KO redundancy ~ gl_dist
22	0.000	0.868	0.031	KO redundancy ~ gl_dist
23	0.004	0.373	1.000	KO redundancy ~ gl_dist
24	0.007	0.379	1.000	KO redundancy ~ gl_dist
25	0.005	0.552	1.000	KO redundancy ~ gl_dist
26	0.000	0.761	0.054	KO redundancy ~ gl_dist
27	0.002	0.428	0.696	KO redundancy ~ gl_dist
28	0.007	0.524	1.000	KO redundancy ~ gl_dist
29	0.000	0.808	0.079	KO redundancy ~ gl_dist
30	0.072	0.186	1.000	KO redundancy ~ gl_dist
31	0.012	0.454	1.000	KO redundancy ~ gl_dist
32	0.003	0.474	1.000	KO redundancy ~ gl_dist
33	0.005	0.513	1.000	KO redundancy ~ gl_dist
34	0.000	0.505	0.186	KO redundancy ~ gl_dist
35	0.001	0.499	0.457	KO redundancy ~ gl_dist
36	0.018	0.561	1.000	KO redundancy ~ gl_dist
37	0.012	0.343	1.000	KO redundancy ~ gl_dist
38	0.009	0.743	1.000	KO redundancy ~ gl_dist
39	0.000	0.711	0.090	KO redundancy ~ gl_dist
40	0.000	0.590	0.090	KO redundancy ~ gl_dist
41	0.001	0.591	0.209	KO redundancy ~ gl_dist

42	0.002	0.627	0.771	KO redundancy ~ gl_dist
43	0.001	0.557	0.263	KO redundancy ~ gl_dist
44	0.001	0.557	0.263	KO redundancy ~ gl_dist
45	0.000	0.768	0.011	KO redundancy ~ gl_dist
46	0.001	0.563	0.263	KO redundancy ~ gl_dist
47	0.001	0.678	0.235	KO redundancy ~ gl_dist
48	0.000	0.565	0.165	KO redundancy ~ gl_dist
49	0.000	0.668	0.047	KO redundancy ~ gl_dist
50	0.001	0.587	0.209	KO redundancy ~ gl_dist
51	0.001	0.566	0.209	KO redundancy ~ gl_dist
52	0.001	0.562	0.209	KO redundancy ~ gl_dist
53	0.001	0.567	0.209	KO redundancy ~ gl_dist
54	0.001	0.567	0.209	KO redundancy ~ gl_dist
55	0.001	0.552	0.457	KO redundancy ~ gl_dist
56	0.001	0.567	0.209	KO redundancy ~ gl_dist
57	0.001	0.567	0.209	KO redundancy ~ gl_dist
58	0.001	0.567	0.209	KO redundancy ~ gl_dist
59	0.000	0.642	0.102	KO redundancy ~ gl_dist
60	0.000	0.771	0.007	KO redundancy ~ gl_dist
61	0.001	0.561	0.294	KO redundancy ~ gl_dist
62	0.000	0.666	0.027	KO redundancy ~ gl_dist
63	0.000	0.465	0.165	KO redundancy ~ gl_dist
64	0.004	0.420	1.000	KO redundancy ~ gl_dist
65	0.000	0.492	0.102	KO redundancy ~ gl_dist
66	0.001	0.662	0.368	KO redundancy ~ gl_dist
67	0.001	0.468	0.457	KO redundancy ~ gl_dist
68	0.001	0.560	0.565	KO redundancy ~ gl_dist
69	0.035	0.212	1.000	KO redundancy ~ gl_dist
70	0.000	0.521	0.147	KO redundancy ~ gl_dist
71	0.000	0.527	0.165	KO redundancy ~ gl_dist
72	0.000	0.469	0.147	KO redundancy ~ gl_dist
73	0.001	0.436	0.457	KO redundancy ~ gl_dist
74	0.000	0.437	0.186	KO redundancy ~ gl_dist
75	0.001	0.544	0.368	KO redundancy ~ gl_dist
76	0.001	0.519	0.294	KO redundancy ~ gl_dist
77	0.000	0.599	0.186	KO redundancy ~ gl_dist
78	0.000	0.573	0.165	KO redundancy ~ gl_dist
79	0.001	0.575	0.209	KO redundancy ~ gl_dist
80	0.001	0.562	0.329	KO redundancy ~ gl_dist

81	0.001	0.563	0.329	KO redundancy ~ gl_dist
82	0.002	0.378	0.696	KO redundancy ~ gl_dist
83	0.000	0.499	0.186	KO redundancy ~ gl_dist
84	0.000	0.497	0.115	KO redundancy ~ gl_dist
85	0.001	0.492	0.294	KO redundancy ~ gl_dist
86	0.001	0.504	0.263	KO redundancy ~ gl_dist
87	0.002	0.457	0.696	KO redundancy ~ gl_dist
88	0.000	0.508	0.186	KO redundancy ~ gl_dist
89	0.001	0.496	0.368	KO redundancy ~ gl_dist
90	0.002	0.462	0.628	KO redundancy ~ gl_dist
91	0.003	0.394	1.000	KO redundancy ~ gl_dist
92	0.002	0.464	0.628	KO redundancy ~ gl_dist
93	0.002	0.458	0.696	KO redundancy ~ gl_dist
94	0.002	0.461	0.696	KO redundancy ~ gl_dist
95	0.002	0.461	0.696	KO redundancy ~ gl_dist
96	0.001	0.496	0.410	KO redundancy ~ gl_dist
97	0.002	0.468	0.943	KO redundancy ~ gl_dist
98	0.002	0.441	0.943	KO redundancy ~ gl_dist
99	0.003	0.495	1.000	KO redundancy ~ gl_dist
100	0.003	0.497	1.000	KO redundancy ~ gl_dist
101	0.017	0.285	1.000	KO redundancy ~ gl_dist
102	0.017	0.280	1.000	KO redundancy ~ gl_dist
103	0.003	0.501	1.000	KO redundancy ~ gl_dist
104	0.017	0.284	1.000	KO redundancy ~ gl_dist
105	0.002	0.481	0.628	KO redundancy ~ gl_dist
106	0.001	0.756	0.294	KO redundancy ~ gl_dist
107	0.014	0.439	1.000	KO redundancy ~ gl_dist
108	0.000	0.701	0.079	KO redundancy ~ gl_dist
109	0.015	0.283	1.000	KO redundancy ~ gl_dist
110	0.012	0.337	1.000	KO redundancy ~ gl_dist
111	0.001	0.629	0.294	KO redundancy ~ gl_dist
112	0.002	0.747	0.853	KO redundancy ~ gl_dist
113	0.000	0.720	0.090	KO redundancy ~ gl_dist
114	0.000	0.703	0.115	KO redundancy ~ gl_dist
115	0.001	0.659	0.294	KO redundancy ~ gl_dist
116	0.142	0.129	1.000	KO redundancy ~ gl_dist
117	0.001	0.574	0.368	KO redundancy ~ gl_dist
118	0.000	0.816	0.079	KO redundancy ~ gl_dist
119	0.000	0.717	0.054	KO redundancy ~ gl_dist
120	0.026	0.343	1.000	KO redundancy ~ gl_dist
121	0.026	0.344	1.000	KO redundancy ~ gl_dist
122	0.002	0.513	0.771	KO redundancy ~ gl_dist
123	0.000	0.634	0.061	KO redundancy ~ gl_dist
124	0.020	0.324	1.000	KO redundancy ~ gl_dist

125	0.020	0.323	1.000	KO redundancy ~ gl_dist
126	0.000	0.865	0.006	KO redundancy ~ gl_dist
127	0.000	0.887	0.130	KO redundancy ~ gl_dist
128	0.035	0.259	1.000	KO redundancy ~ gl_dist
129	0.035	0.259	1.000	KO redundancy ~ gl_dist
130	0.091	0.216	1.000	KO redundancy ~ gl_dist
131	0.002	0.503	0.943	KO redundancy ~ gl_dist
132	0.014	0.504	1.000	KO redundancy ~ gl_dist
133	0.000	0.965	0.001	KO redundancy ~ gl_dist
134	0.038	0.465	1.000	KO redundancy ~ gl_dist
135	0.052	0.276	1.000	KO redundancy ~ gl_dist
136	0.072	0.282	1.000	KO redundancy ~ gl_dist
137	0.002	0.630	0.696	KO redundancy ~ gl_dist
138	0.030	0.272	1.000	KO redundancy ~ gl_dist
139	0.030	0.272	1.000	KO redundancy ~ gl_dist
140	0.026	0.175	1.000	KO redundancy ~ gl_dist
141	0.026	0.176	1.000	KO redundancy ~ gl_dist
142	0.026	0.174	1.000	KO redundancy ~ gl_dist
143	0.026	0.178	1.000	KO redundancy ~ gl_dist
144	0.020	0.198	1.000	KO redundancy ~ gl_dist
145	0.000	0.994	0.024	KO redundancy ~ gl_dist
146	0.023	0.182	1.000	KO redundancy ~ gl_dist
147	0.000	1.022	0.021	KO redundancy ~ gl_dist
148	0.000	0.936	0.070	KO redundancy ~ gl_dist
149	0.023	0.181	1.000	KO redundancy ~ gl_dist
150	0.028	0.196	1.000	KO redundancy ~ gl_dist
151	0.028	0.196	1.000	KO redundancy ~ gl_dist
152	0.028	0.197	1.000	KO redundancy ~ gl_dist
153	0.000	0.944	0.041	KO redundancy ~ gl_dist
154	0.030	0.205	1.000	KO redundancy ~ gl_dist
155	0.002	0.695	0.628	KO redundancy ~ gl_dist
156	0.000	0.717	0.006	KO redundancy ~ gl_dist
157	0.040	0.216	1.000	KO redundancy ~ gl_dist
158	0.038	0.218	1.000	KO redundancy ~ gl_dist
159	0.040	0.217	1.000	KO redundancy ~ gl_dist
160	0.000	0.591	0.186	KO redundancy ~ gl_dist
161	0.028	0.209	1.000	KO redundancy ~ gl_dist
162	0.000	0.900	0.130	KO redundancy ~ gl_dist
163	0.052	0.156	1.000	KO redundancy ~ gl_dist

164	0.052	0.155	1.000	KO redundancy ~ gl_dist
165	0.052	0.156	1.000	KO redundancy ~ gl_dist
166	0.052	0.155	1.000	KO redundancy ~ gl_dist
167	0.056	0.147	1.000	KO redundancy ~ gl_dist
168	0.000	0.852	0.102	KO redundancy ~ gl_dist
169	0.001	0.546	0.329	KO redundancy ~ gl_dist
170	0.000	0.808	0.090	KO redundancy ~ gl_dist
171	0.018	0.283	1.000	KO redundancy ~ gl_dist
172	0.000	0.545	0.079	KO redundancy ~ gl_dist
173	0.018	0.283	1.000	KO redundancy ~ gl_dist
174	0.018	0.283	1.000	KO redundancy ~ gl_dist
175	0.018	0.283	1.000	KO redundancy ~ gl_dist
176	0.000	0.544	0.079	KO redundancy ~ gl_dist
177	0.018	0.283	1.000	KO redundancy ~ gl_dist
178	0.000	0.544	0.079	KO redundancy ~ gl_dist
179	0.000	0.793	0.061	KO redundancy ~ gl_dist
180	0.015	0.294	1.000	KO redundancy ~ gl_dist
181	0.015	0.294	1.000	KO redundancy ~ gl_dist
182	0.015	0.292	1.000	KO redundancy ~ gl_dist
183	0.015	0.294	1.000	KO redundancy ~ gl_dist
184	0.201	0.090	1.000	KO redundancy ~ gl_dist
185	0.000	0.589	0.090	KO redundancy ~ gl_dist
186	0.004	0.528	1.000	KO redundancy ~ gl_dist
187	0.001	0.595	0.263	KO redundancy ~ gl_dist
188	0.000	0.587	0.090	KO redundancy ~ gl_dist
189	0.000	0.769	0.165	KO redundancy ~ gl_dist
190	0.000	0.808	0.031	KO redundancy ~ gl_dist
191	0.001	0.530	0.565	KO redundancy ~ gl_dist
192	0.001	0.536	0.565	KO redundancy ~ gl_dist
193	0.000	0.424	0.102	KO redundancy ~ gl_dist
194	0.010	0.362	1.000	KO redundancy ~ gl_dist
195	0.000	0.661	0.165	KO redundancy ~ gl_dist
196	0.008	0.422	1.000	KO redundancy ~ gl_dist
197	0.000	0.550	0.079	KO redundancy ~ gl_dist
198	0.001	0.563	0.509	KO redundancy ~ gl_dist
199	0.000	0.589	0.070	KO redundancy ~ gl_dist
200	0.086	0.324	1.000	KO redundancy ~ gl_dist
201	0.007	0.718	1.000	KO redundancy ~ gl_dist
202	0.001	0.508	0.294	KO redundancy ~ gl_dist
203	0.009	0.613	1.000	KO redundancy ~ gl_dist
204	0.013	0.498	1.000	KO redundancy ~ gl_dist
205	0.035	0.194	1.000	KO redundancy ~ gl_dist
206	0.035	0.164	1.000	KO redundancy ~ gl_dist
207	0.002	0.712	0.696	KO redundancy ~ gl_dist

208	0.134	0.256	1.000	KO redundancy ~ gl_dist
209	0.012	0.582	1.000	KO redundancy ~ gl_dist
210	0.021	0.545	1.000	KO redundancy ~ gl_dist
211	0.002	0.498	0.696	KO redundancy ~ gl_dist
212	0.002	0.498	0.696	KO redundancy ~ gl_dist
213	0.021	0.546	1.000	KO redundancy ~ gl_dist
214	0.021	0.546	1.000	KO redundancy ~ gl_dist
215	0.000	0.654	0.090	KO redundancy ~ gl_dist
216	0.023	0.361	1.000	KO redundancy ~ gl_dist
217	0.000	0.820	0.079	KO redundancy ~ gl_dist
218	0.020	0.269	1.000	KO redundancy ~ gl_dist
219	0.002	0.527	0.943	KO redundancy ~ gl_dist
220	0.002	0.666	0.943	KO redundancy ~ gl_dist
221	0.001	0.685	0.509	KO redundancy ~ gl_dist
222	0.002	0.671	0.853	KO redundancy ~ gl_dist
223	0.002	0.669	0.853	KO redundancy ~ gl_dist
224	0.002	0.668	0.943	KO redundancy ~ gl_dist
225	0.108	0.304	1.000	KO redundancy ~ gl_dist
226	0.004	0.388	1.000	KO redundancy ~ gl_dist
227	0.000	0.748	0.102	KO redundancy ~ gl_dist
228	0.001	0.507	0.368	KO redundancy ~ gl_dist
229	0.000	0.911	0.079	KO redundancy ~ gl_dist
230	0.004	0.373	1.000	KO redundancy ~ gl_dist
231	0.004	0.374	1.000	KO redundancy ~ gl_dist
232	0.000	0.741	0.024	KO redundancy ~ gl_dist
233	0.030	0.263	1.000	KO redundancy ~ gl_dist
234	0.000	0.679	0.115	KO redundancy ~ gl_dist
235	0.021	0.232	1.000	KO redundancy ~ gl_dist
236	0.021	0.235	1.000	KO redundancy ~ gl_dist
237	0.001	0.680	0.410	KO redundancy ~ gl_dist
238	0.001	0.657	0.209	KO redundancy ~ gl_dist
239	0.040	0.231	1.000	KO redundancy ~ gl_dist
240	0.038	0.240	1.000	KO redundancy ~ gl_dist
241	0.038	0.238	1.000	KO redundancy ~ gl_dist
242	0.038	0.239	1.000	KO redundancy ~ gl_dist
243	0.038	0.239	1.000	KO redundancy ~ gl_dist
244	0.001	0.513	0.457	KO redundancy ~ gl_dist
245	0.000	0.870	0.018	KO redundancy ~ gl_dist
246	0.005	0.555	1.000	KO redundancy ~ gl_dist

247	0.005	0.555	1.000	KO redundancy ~ gl_dist
248	0.004	0.565	1.000	KO redundancy ~ gl_dist
249	0.012	0.271	1.000	KO redundancy ~ gl_dist
250	0.006	0.541	1.000	KO redundancy ~ gl_dist
251	0.033	0.236	1.000	KO redundancy ~ gl_dist
252	0.005	0.509	1.000	KO redundancy ~ gl_dist
253	0.005	0.509	1.000	KO redundancy ~ gl_dist
254	0.002	0.498	0.853	KO redundancy ~ gl_dist
255	0.002	0.472	0.628	KO redundancy ~ gl_dist
256	0.015	0.296	1.000	KO redundancy ~ gl_dist
257	0.002	0.560	0.771	KO redundancy ~ gl_dist
258	0.000	0.815	0.015	KO redundancy ~ gl_dist
259	0.014	0.300	1.000	KO redundancy ~ gl_dist
260	0.005	0.651	1.000	KO redundancy ~ gl_dist
261	0.002	0.544	0.628	KO redundancy ~ gl_dist
262	0.002	0.527	0.853	KO redundancy ~ gl_dist
263	0.002	0.671	0.943	KO redundancy ~ gl_dist
264	0.000	0.767	0.013	KO redundancy ~ gl_dist
265	0.001	0.497	0.209	KO redundancy ~ gl_dist
266	0.024	0.170	1.000	KO redundancy ~ gl_dist
267	0.010	0.375	1.000	KO redundancy ~ gl_dist
268	0.108	0.114	1.000	KO redundancy ~ gl_dist
269	0.023	0.431	1.000	KO redundancy ~ gl_dist
270	0.001	0.467	0.329	KO redundancy ~ gl_dist
271	0.000	0.529	0.061	KO redundancy ~ gl_dist
272	0.021	0.405	1.000	KO redundancy ~ gl_dist
273	0.000	0.611	0.102	KO redundancy ~ gl_dist
274	0.000	0.822	0.090	KO redundancy ~ gl_dist
275	0.001	0.613	0.294	KO redundancy ~ gl_dist
276	0.011	0.417	1.000	KO redundancy ~ gl_dist
277	0.001	0.664	0.209	KO redundancy ~ gl_dist
278	0.001	0.668	0.209	KO redundancy ~ gl_dist
279	0.001	0.486	0.294	KO redundancy ~ gl_dist
280	0.046	0.291	1.000	KO redundancy ~ gl_dist
281	0.012	0.491	1.000	KO redundancy ~ gl_dist
282	0.000	0.746	0.001	KO redundancy ~ gl_dist
283	0.002	0.576	0.771	KO redundancy ~ gl_dist
284	0.035	0.185	1.000	KO redundancy ~ gl_dist
285	0.004	0.638	1.000	KO redundancy ~ gl_dist
286	0.010	0.243	1.000	KO redundancy ~ gl_dist
287	0.009	0.251	1.000	KO redundancy ~ gl_dist
288	0.003	0.594	1.000	KO redundancy ~ gl_dist
289	0.000	0.587	0.027	KO redundancy ~ gl_dist
290	0.006	0.669	1.000	KO redundancy ~ gl_dist

291	0.026	0.339	1.000	KO redundancy ~ gl_dist
292	0.026	0.340	1.000	KO redundancy ~ gl_dist
293	0.026	0.339	1.000	KO redundancy ~ gl_dist
294	0.026	0.338	1.000	KO redundancy ~ gl_dist
295	0.000	0.545	0.015	KO redundancy ~ gl_dist
296	0.003	0.585	1.000	KO redundancy ~ gl_dist
297	0.000	0.731	0.006	KO redundancy ~ gl_dist
298	0.028	0.263	1.000	KO redundancy ~ gl_dist
299	0.000	0.626	0.090	KO redundancy ~ gl_dist
300	0.014	0.499	1.000	KO redundancy ~ gl_dist
301	0.015	0.509	1.000	KO redundancy ~ gl_dist
302	0.018	0.505	1.000	KO redundancy ~ gl_dist
303	0.072	0.231	1.000	KO redundancy ~ gl_dist
304	0.006	0.463	1.000	KO redundancy ~ gl_dist
305	0.091	0.311	1.000	KO redundancy ~ gl_dist
306	0.000	0.895	0.102	KO redundancy ~ gl_dist
307	0.142	0.136	1.000	KO redundancy ~ gl_dist
308	0.002	0.629	0.771	KO redundancy ~ gl_dist
309	0.007	0.637	1.000	KO redundancy ~ gl_dist
310	0.004	0.514	1.000	KO redundancy ~ gl_dist
311	0.017	0.415	1.000	KO redundancy ~ gl_dist
312	0.017	0.420	1.000	KO redundancy ~ gl_dist
313	0.005	0.593	1.000	KO redundancy ~ gl_dist
314	0.000	1.666	0.000	KO redundancy ~ gl_dist
315	0.004	0.481	1.000	KO redundancy ~ gl_dist
316	0.004	0.481	1.000	KO redundancy ~ gl_dist
317	0.004	0.543	1.000	KO redundancy ~ gl_dist
318	0.000	0.909	0.004	KO redundancy ~ gl_dist
319	0.003	0.471	1.000	KO redundancy ~ gl_dist
320	0.000	0.915	0.102	KO redundancy ~ gl_dist
321	0.004	0.429	1.000	KO redundancy ~ gl_dist
322	0.005	0.416	1.000	KO redundancy ~ gl_dist
323	0.003	0.476	1.000	KO redundancy ~ gl_dist
324	0.005	0.416	1.000	KO redundancy ~ gl_dist
325	0.005	0.415	1.000	KO redundancy ~ gl_dist
326	0.000	0.893	0.002	KO redundancy ~ gl_dist
327	0.002	0.482	0.943	KO redundancy ~ gl_dist
328	0.002	0.481	0.943	KO redundancy ~ gl_dist
329	0.000	0.639	0.024	KO redundancy ~ gl_dist

330	0.005	0.513	1.000	KO redundancy ~ gl_dist
331	0.000	0.889	0.008	KO redundancy ~ gl_dist
332	0.000	0.800	0.047	KO redundancy ~ gl_dist
333	0.020	0.312	1.000	KO redundancy ~ gl_dist
334	0.017	0.305	1.000	KO redundancy ~ gl_dist
335	0.001	0.635	0.209	KO redundancy ~ gl_dist
336	0.004	0.516	1.000	KO redundancy ~ gl_dist
337	0.010	0.758	1.000	KO redundancy ~ gl_dist
338	0.003	0.542	1.000	KO redundancy ~ gl_dist
339	0.000	0.506	0.186	KO redundancy ~ gl_dist
340	0.000	0.506	0.186	KO redundancy ~ gl_dist
341	0.000	0.888	0.061	KO redundancy ~ gl_dist
342	0.001	0.454	0.509	KO redundancy ~ gl_dist
343	0.001	0.470	0.565	KO redundancy ~ gl_dist
344	0.000	0.587	0.186	KO redundancy ~ gl_dist
345	0.002	0.370	0.943	KO redundancy ~ gl_dist
346	0.002	0.370	0.943	KO redundancy ~ gl_dist
347	0.002	0.371	0.943	KO redundancy ~ gl_dist
348	0.000	0.672	0.018	KO redundancy ~ gl_dist
349	0.002	0.507	0.771	KO redundancy ~ gl_dist
350	0.001	0.526	0.329	KO redundancy ~ gl_dist
351	0.011	0.501	1.000	KO redundancy ~ gl_dist
352	0.001	0.525	0.329	KO redundancy ~ gl_dist
353	0.010	0.511	1.000	KO redundancy ~ gl_dist
354	0.000	0.715	0.036	KO redundancy ~ gl_dist
355	0.289	0.034	1.000	KO redundancy ~ gl_dist
356	0.002	0.535	0.853	KO redundancy ~ gl_dist
357	0.000	0.895	0.041	KO redundancy ~ gl_dist
358	0.017	0.453	1.000	KO redundancy ~ gl_dist
359	0.000	0.696	0.070	KO redundancy ~ gl_dist
360	0.000	0.758	0.061	KO redundancy ~ gl_dist
361	0.018	0.338	1.000	KO redundancy ~ gl_dist
362	0.001	0.693	0.209	KO redundancy ~ gl_dist
363	0.004	0.498	1.000	KO redundancy ~ gl_dist
364	0.000	0.860	0.079	KO redundancy ~ gl_dist
365	0.011	0.550	1.000	KO redundancy ~ gl_dist
366	0.001	0.533	0.565	KO redundancy ~ gl_dist
367	0.012	0.343	1.000	KO redundancy ~ gl_dist
368	0.001	0.633	0.565	KO redundancy ~ gl_dist
369	0.000	0.717	0.186	KO redundancy ~ gl_dist
370	0.000	0.850	0.018	KO redundancy ~ gl_dist
371	0.012	0.317	1.000	KO redundancy ~ gl_dist
372	0.000	0.756	0.054	KO redundancy ~ gl_dist
373	0.001	0.550	0.263	KO redundancy ~ gl_dist

374	0.002	0.734	0.853	KO redundancy ~ gl_dist
375	0.000	0.562	0.186	KO redundancy ~ gl_dist
376	0.002	0.375	0.853	KO redundancy ~ gl_dist
377	0.002	0.375	0.853	KO redundancy ~ gl_dist
378	0.002	0.377	0.853	KO redundancy ~ gl_dist
379	0.030	0.394	1.000	KO redundancy ~ gl_dist
380	0.003	0.402	1.000	KO redundancy ~ gl_dist
381	0.002	0.739	0.628	KO redundancy ~ gl_dist
382	0.002	0.483	0.696	KO redundancy ~ gl_dist
383	0.003	0.453	1.000	KO redundancy ~ gl_dist
384	0.005	0.585	1.000	KO redundancy ~ gl_dist
385	0.020	0.415	1.000	KO redundancy ~ gl_dist
386	0.659	-0.003	1.000	KO redundancy ~ gl_dist
387	0.008	0.589	1.000	KO redundancy ~ gl_dist
388	0.000	0.488	0.115	KO redundancy ~ gl_dist
389	0.009	0.672	1.000	KO redundancy ~ gl_dist
390	0.009	0.392	1.000	KO redundancy ~ gl_dist
391	0.002	0.663	0.853	KO redundancy ~ gl_dist
392	0.002	0.713	0.771	KO redundancy ~ gl_dist
393	0.000	0.760	0.010	KO redundancy ~ gl_dist
394	0.049	0.324	1.000	KO redundancy ~ gl_dist
1	0.277	0.128	1.000	KO redundancy ~ watemp
2	0.033	0.188	1.000	KO redundancy ~ watemp
3	0.142	0.195	1.000	KO redundancy ~ watemp
4	0.010	0.276	1.000	KO redundancy ~ watemp
5	0.052	0.194	1.000	KO redundancy ~ watemp
6	0.060	0.163	1.000	KO redundancy ~ watemp
7	0.063	0.136	1.000	KO redundancy ~ watemp
8	0.004	0.289	1.000	KO redundancy ~ watemp
9	0.035	0.229	1.000	KO redundancy ~ watemp
10	0.277	0.129	1.000	KO redundancy ~ watemp
11	0.142	0.112	1.000	KO redundancy ~ watemp
12	0.001	0.482	0.368	KO redundancy ~ watemp
13	0.925	-0.072	1.000	KO redundancy ~ watemp
14	0.142	0.176	1.000	KO redundancy ~ watemp
15	0.211	0.178	1.000	KO redundancy ~ watemp
16	0.091	0.130	1.000	KO redundancy ~ watemp
17	0.096	0.218	1.000	KO redundancy ~ watemp
18	0.026	0.172	1.000	KO redundancy ~ watemp

19	0.086	0.165	1.000	KO redundancy ~ watemp
20	0.242	0.165	1.000	KO redundancy ~ watemp
21	0.355	0.048	1.000	KO redundancy ~ watemp
22	0.086	0.160	1.000	KO redundancy ~ watemp
23	0.201	0.168	1.000	KO redundancy ~ watemp
24	0.121	0.119	1.000	KO redundancy ~ watemp
25	0.086	0.173	1.000	KO redundancy ~ watemp
26	0.127	0.103	1.000	KO redundancy ~ watemp
27	0.461	0.030	1.000	KO redundancy ~ watemp
28	0.355	0.025	1.000	KO redundancy ~ watemp
29	0.114	0.265	1.000	KO redundancy ~ watemp
30	0.314	0.153	1.000	KO redundancy ~ watemp
31	0.327	0.133	1.000	KO redundancy ~ watemp
32	0.096	0.241	1.000	KO redundancy ~ watemp
33	0.183	0.178	1.000	KO redundancy ~ watemp
34	0.201	0.146	1.000	KO redundancy ~ watemp
35	0.096	0.201	1.000	KO redundancy ~ watemp
36	0.068	0.205	1.000	KO redundancy ~ watemp
37	0.063	0.246	1.000	KO redundancy ~ watemp
38	0.017	0.259	1.000	KO redundancy ~ watemp
39	0.063	0.224	1.000	KO redundancy ~ watemp
40	0.060	0.226	1.000	KO redundancy ~ watemp
41	0.030	0.174	1.000	KO redundancy ~ watemp
42	0.026	0.272	1.000	KO redundancy ~ watemp
43	0.063	0.227	1.000	KO redundancy ~ watemp
44	0.063	0.226	1.000	KO redundancy ~ watemp
45	0.007	0.324	1.000	KO redundancy ~ watemp
46	0.046	0.156	1.000	KO redundancy ~ watemp
47	0.024	0.264	1.000	KO redundancy ~ watemp
48	0.038	0.165	1.000	KO redundancy ~ watemp
49	0.006	0.289	1.000	KO redundancy ~ watemp
50	0.056	0.161	1.000	KO redundancy ~ watemp
51	0.060	0.155	1.000	KO redundancy ~ watemp
52	0.060	0.152	1.000	KO redundancy ~ watemp
53	0.060	0.154	1.000	KO redundancy ~ watemp
54	0.060	0.154	1.000	KO redundancy ~ watemp
55	0.052	0.193	1.000	KO redundancy ~ watemp
56	0.060	0.154	1.000	KO redundancy ~ watemp
57	0.060	0.154	1.000	KO redundancy ~ watemp
58	0.060	0.154	1.000	KO redundancy ~ watemp
59	0.023	0.299	1.000	KO redundancy ~ watemp
60	0.007	0.326	1.000	KO redundancy ~ watemp
61	0.052	0.193	1.000	KO redundancy ~ watemp
62	0.018	0.199	1.000	KO redundancy ~ watemp

63	0.018	0.283	1.000	KO redundancy ~ watemp
64	0.035	0.211	1.000	KO redundancy ~ watemp
65	0.014	0.203	1.000	KO redundancy ~ watemp
66	0.020	0.238	1.000	KO redundancy ~ watemp
67	0.052	0.181	1.000	KO redundancy ~ watemp
68	0.028	0.289	1.000	KO redundancy ~ watemp
69	0.429	-0.022	1.000	KO redundancy ~ watemp
70	0.013	0.260	1.000	KO redundancy ~ watemp
71	0.020	0.223	1.000	KO redundancy ~ watemp
72	0.014	0.279	1.000	KO redundancy ~ watemp
73	0.033	0.211	1.000	KO redundancy ~ watemp
74	0.024	0.218	1.000	KO redundancy ~ watemp
75	0.056	0.140	1.000	KO redundancy ~ watemp
76	0.023	0.254	1.000	KO redundancy ~ watemp
77	0.013	0.216	1.000	KO redundancy ~ watemp
78	0.006	0.278	1.000	KO redundancy ~ watemp
79	0.035	0.179	1.000	KO redundancy ~ watemp
80	0.006	0.271	1.000	KO redundancy ~ watemp
81	0.006	0.272	1.000	KO redundancy ~ watemp
82	0.014	0.235	1.000	KO redundancy ~ watemp
83	0.009	0.286	1.000	KO redundancy ~ watemp
84	0.017	0.221	1.000	KO redundancy ~ watemp
85	0.021	0.223	1.000	KO redundancy ~ watemp
86	0.009	0.289	1.000	KO redundancy ~ watemp
87	0.017	0.291	1.000	KO redundancy ~ watemp
88	0.010	0.289	1.000	KO redundancy ~ watemp
89	0.035	0.286	1.000	KO redundancy ~ watemp
90	0.046	0.289	1.000	KO redundancy ~ watemp
91	0.043	0.240	1.000	KO redundancy ~ watemp
92	0.043	0.291	1.000	KO redundancy ~ watemp
93	0.046	0.288	1.000	KO redundancy ~ watemp
94	0.043	0.290	1.000	KO redundancy ~ watemp
95	0.046	0.290	1.000	KO redundancy ~ watemp
96	0.026	0.324	1.000	KO redundancy ~ watemp
97	0.035	0.273	1.000	KO redundancy ~ watemp
98	0.030	0.288	1.000	KO redundancy ~ watemp
99	0.038	0.289	1.000	KO redundancy ~ watemp
100	0.398	-0.021	1.000	KO redundancy ~ watemp
101	0.277	0.128	1.000	KO redundancy ~ watemp

102	0.265	0.130	1.000	KO redundancy ~ watemp
103	0.383	-0.015	1.000	KO redundancy ~ watemp
104	0.277	0.128	1.000	KO redundancy ~ watemp
105	0.102	0.129	1.000	KO redundancy ~ watemp
106	0.221	0.060	1.000	KO redundancy ~ watemp
107	0.060	0.135	1.000	KO redundancy ~ watemp
108	0.040	0.145	1.000	KO redundancy ~ watemp
109	0.565	0.029	1.000	KO redundancy ~ watemp
110	0.495	0.042	1.000	KO redundancy ~ watemp
111	0.114	0.110	1.000	KO redundancy ~ watemp
112	0.002	0.385	0.771	KO redundancy ~ watemp
113	0.134	0.100	1.000	KO redundancy ~ watemp
114	0.142	0.095	1.000	KO redundancy ~ watemp
115	0.398	0.012	1.000	KO redundancy ~ watemp
116	0.211	0.146	1.000	KO redundancy ~ watemp
117	0.007	0.366	1.000	KO redundancy ~ watemp
118	0.006	0.313	1.000	KO redundancy ~ watemp
119	0.001	0.327	0.509	KO redundancy ~ watemp
120	0.583	0.100	1.000	KO redundancy ~ watemp
121	0.583	0.100	1.000	KO redundancy ~ watemp
122	0.265	0.092	1.000	KO redundancy ~ watemp
123	0.114	0.266	1.000	KO redundancy ~ watemp
124	0.529	0.099	1.000	KO redundancy ~ watemp
125	0.547	0.099	1.000	KO redundancy ~ watemp
126	0.086	0.258	1.000	KO redundancy ~ watemp
127	0.004	0.372	1.000	KO redundancy ~ watemp
128	0.947	0.088	1.000	KO redundancy ~ watemp
129	0.947	0.088	1.000	KO redundancy ~ watemp
130	0.841	0.031	1.000	KO redundancy ~ watemp
131	0.201	0.098	1.000	KO redundancy ~ watemp
132	0.060	0.271	1.000	KO redundancy ~ watemp
133	0.086	0.085	1.000	KO redundancy ~ watemp
134	0.149	0.096	1.000	KO redundancy ~ watemp
135	0.076	0.151	1.000	KO redundancy ~ watemp
136	0.461	0.132	1.000	KO redundancy ~ watemp
137	0.033	0.202	1.000	KO redundancy ~ watemp
138	0.779	0.025	1.000	KO redundancy ~ watemp
139	0.779	0.025	1.000	KO redundancy ~ watemp
140	0.640	-0.022	1.000	KO redundancy ~ watemp
141	0.640	-0.022	1.000	KO redundancy ~ watemp
142	0.640	-0.022	1.000	KO redundancy ~ watemp
143	0.640	-0.022	1.000	KO redundancy ~ watemp
144	0.620	-0.022	1.000	KO redundancy ~ watemp
145	0.040	0.177	1.000	KO redundancy ~ watemp

146	0.620	-0.012	1.000	KO redundancy ~ watemp
147	0.013	0.199	1.000	KO redundancy ~ watemp
148	0.242	0.168	1.000	KO redundancy ~ watemp
149	0.640	-0.012	1.000	KO redundancy ~ watemp
150	0.841	-0.049	1.000	KO redundancy ~ watemp
151	0.841	-0.049	1.000	KO redundancy ~ watemp
152	0.841	-0.049	1.000	KO redundancy ~ watemp
153	0.253	0.169	1.000	KO redundancy ~ watemp
154	0.883	-0.050	1.000	KO redundancy ~ watemp
155	0.026	0.240	1.000	KO redundancy ~ watemp
156	0.231	0.064	1.000	KO redundancy ~ watemp
157	0.718	-0.065	1.000	KO redundancy ~ watemp
158	0.718	-0.064	1.000	KO redundancy ~ watemp
159	0.718	-0.064	1.000	KO redundancy ~ watemp
160	0.081	0.194	1.000	KO redundancy ~ watemp
161	0.779	-0.030	1.000	KO redundancy ~ watemp
162	0.221	0.153	1.000	KO redundancy ~ watemp
163	0.862	-0.076	1.000	KO redundancy ~ watemp
164	0.862	-0.076	1.000	KO redundancy ~ watemp
165	0.862	-0.075	1.000	KO redundancy ~ watemp
166	0.862	-0.075	1.000	KO redundancy ~ watemp
167	0.904	-0.078	1.000	KO redundancy ~ watemp
168	0.242	0.162	1.000	KO redundancy ~ watemp
169	0.102	0.169	1.000	KO redundancy ~ watemp
170	0.063	0.260	1.000	KO redundancy ~ watemp
171	0.414	0.114	1.000	KO redundancy ~ watemp
172	0.149	0.116	1.000	KO redundancy ~ watemp
173	0.414	0.114	1.000	KO redundancy ~ watemp
174	0.414	0.115	1.000	KO redundancy ~ watemp
175	0.414	0.114	1.000	KO redundancy ~ watemp
176	0.149	0.115	1.000	KO redundancy ~ watemp
177	0.414	0.114	1.000	KO redundancy ~ watemp
178	0.149	0.114	1.000	KO redundancy ~ watemp
179	0.461	0.010	1.000	KO redundancy ~ watemp
180	0.383	0.127	1.000	KO redundancy ~ watemp
181	0.383	0.127	1.000	KO redundancy ~ watemp
182	0.383	0.127	1.000	KO redundancy ~ watemp
183	0.383	0.127	1.000	KO redundancy ~ watemp
184	0.429	-0.126	1.000	KO redundancy ~ watemp

185	0.142	0.134	1.000	KO redundancy ~ watemp
186	0.314	0.093	1.000	KO redundancy ~ watemp
187	0.369	0.168	1.000	KO redundancy ~ watemp
188	0.149	0.119	1.000	KO redundancy ~ watemp
189	0.429	0.165	1.000	KO redundancy ~ watemp
190	0.242	0.266	1.000	KO redundancy ~ watemp
191	0.174	0.122	1.000	KO redundancy ~ watemp
192	0.265	0.138	1.000	KO redundancy ~ watemp
193	0.114	0.086	1.000	KO redundancy ~ watemp
194	0.277	0.127	1.000	KO redundancy ~ watemp
195	0.052	0.171	1.000	KO redundancy ~ watemp
196	0.314	0.138	1.000	KO redundancy ~ watemp
197	0.165	0.066	1.000	KO redundancy ~ watemp
198	0.010	0.184	1.000	KO redundancy ~ watemp
199	0.081	0.117	1.000	KO redundancy ~ watemp
200	0.221	0.075	1.000	KO redundancy ~ watemp
201	0.149	0.109	1.000	KO redundancy ~ watemp
202	0.157	0.127	1.000	KO redundancy ~ watemp
203	0.221	0.176	1.000	KO redundancy ~ watemp
204	0.221	0.099	1.000	KO redundancy ~ watemp
205	0.414	0.044	1.000	KO redundancy ~ watemp
206	0.495	0.034	1.000	KO redundancy ~ watemp
207	0.012	0.187	1.000	KO redundancy ~ watemp
208	0.565	-0.008	1.000	KO redundancy ~ watemp
209	0.127	0.097	1.000	KO redundancy ~ watemp
210	0.174	0.125	1.000	KO redundancy ~ watemp
211	0.211	0.069	1.000	KO redundancy ~ watemp
212	0.211	0.069	1.000	KO redundancy ~ watemp
213	0.174	0.125	1.000	KO redundancy ~ watemp
214	0.174	0.125	1.000	KO redundancy ~ watemp
215	0.253	0.005	1.000	KO redundancy ~ watemp
216	0.327	0.067	1.000	KO redundancy ~ watemp
217	0.121	0.156	1.000	KO redundancy ~ watemp
218	0.429	0.014	1.000	KO redundancy ~ watemp
219	0.192	0.154	1.000	KO redundancy ~ watemp
220	0.231	0.126	1.000	KO redundancy ~ watemp
221	0.221	0.131	1.000	KO redundancy ~ watemp
222	0.221	0.130	1.000	KO redundancy ~ watemp
223	0.221	0.128	1.000	KO redundancy ~ watemp
224	0.221	0.127	1.000	KO redundancy ~ watemp
225	0.678	-0.025	1.000	KO redundancy ~ watemp
226	0.174	0.100	1.000	KO redundancy ~ watemp
227	0.004	0.312	1.000	KO redundancy ~ watemp
228	0.127	0.194	1.000	KO redundancy ~ watemp

229	0.108	0.199	1.000	KO redundancy ~ watemp
230	0.183	0.170	1.000	KO redundancy ~ watemp
231	0.183	0.169	1.000	KO redundancy ~ watemp
232	0.068	0.170	1.000	KO redundancy ~ watemp
233	0.398	0.040	1.000	KO redundancy ~ watemp
234	0.052	0.178	1.000	KO redundancy ~ watemp
235	0.383	0.028	1.000	KO redundancy ~ watemp
236	0.369	0.029	1.000	KO redundancy ~ watemp
237	0.142	0.168	1.000	KO redundancy ~ watemp
238	0.369	-0.006	1.000	KO redundancy ~ watemp
239	0.242	0.105	1.000	KO redundancy ~ watemp
240	0.461	0.040	1.000	KO redundancy ~ watemp
241	0.478	0.038	1.000	KO redundancy ~ watemp
242	0.478	0.039	1.000	KO redundancy ~ watemp
243	0.478	0.039	1.000	KO redundancy ~ watemp
244	0.165	0.037	1.000	KO redundancy ~ watemp
245	0.063	0.214	1.000	KO redundancy ~ watemp
246	0.086	0.175	1.000	KO redundancy ~ watemp
247	0.086	0.176	1.000	KO redundancy ~ watemp
248	0.081	0.179	1.000	KO redundancy ~ watemp
249	0.779	-0.043	1.000	KO redundancy ~ watemp
250	0.091	0.173	1.000	KO redundancy ~ watemp
251	0.758	-0.069	1.000	KO redundancy ~ watemp
252	0.461	0.024	1.000	KO redundancy ~ watemp
253	0.461	0.024	1.000	KO redundancy ~ watemp
254	0.174	0.095	1.000	KO redundancy ~ watemp
255	0.157	0.042	1.000	KO redundancy ~ watemp
256	0.640	0.000	1.000	KO redundancy ~ watemp
257	0.102	0.184	1.000	KO redundancy ~ watemp
258	0.052	0.226	1.000	KO redundancy ~ watemp
259	0.678	-0.027	1.000	KO redundancy ~ watemp
260	0.314	0.063	1.000	KO redundancy ~ watemp
261	0.565	0.072	1.000	KO redundancy ~ watemp
262	0.231	0.152	1.000	KO redundancy ~ watemp
263	0.277	0.036	1.000	KO redundancy ~ watemp
264	0.021	0.233	1.000	KO redundancy ~ watemp
265	0.060	0.219	1.000	KO redundancy ~ watemp
266	0.327	0.114	1.000	KO redundancy ~ watemp
267	0.445	0.008	1.000	KO redundancy ~ watemp

268	0.529	0.006	1.000	KO redundancy ~ watemp
269	0.102	0.180	1.000	KO redundancy ~ watemp
270	0.341	0.057	1.000	KO redundancy ~ watemp
271	0.114	0.185	1.000	KO redundancy ~ watemp
272	0.383	0.126	1.000	KO redundancy ~ watemp
273	0.026	0.242	1.000	KO redundancy ~ watemp
274	0.076	0.224	1.000	KO redundancy ~ watemp
275	0.068	0.154	1.000	KO redundancy ~ watemp
276	0.461	0.019	1.000	KO redundancy ~ watemp
277	0.108	0.210	1.000	KO redundancy ~ watemp
278	0.108	0.212	1.000	KO redundancy ~ watemp
279	0.231	0.110	1.000	KO redundancy ~ watemp
280	0.659	-0.022	1.000	KO redundancy ~ watemp
281	0.063	0.119	1.000	KO redundancy ~ watemp
282	0.001	0.318	0.209	KO redundancy ~ watemp
283	0.096	0.107	1.000	KO redundancy ~ watemp
284	0.547	0.012	1.000	KO redundancy ~ watemp
285	0.414	0.050	1.000	KO redundancy ~ watemp
286	0.383	0.061	1.000	KO redundancy ~ watemp
287	0.383	0.064	1.000	KO redundancy ~ watemp
288	0.565	0.015	1.000	KO redundancy ~ watemp
289	0.015	0.239	1.000	KO redundancy ~ watemp
290	0.221	0.099	1.000	KO redundancy ~ watemp
291	0.211	0.162	1.000	KO redundancy ~ watemp
292	0.211	0.162	1.000	KO redundancy ~ watemp
293	0.211	0.162	1.000	KO redundancy ~ watemp
294	0.211	0.161	1.000	KO redundancy ~ watemp
295	0.006	0.215	1.000	KO redundancy ~ watemp
296	0.174	0.167	1.000	KO redundancy ~ watemp
297	0.043	0.236	1.000	KO redundancy ~ watemp
298	0.242	0.135	1.000	KO redundancy ~ watemp
299	0.015	0.197	1.000	KO redundancy ~ watemp
300	0.414	0.141	1.000	KO redundancy ~ watemp
301	0.461	0.109	1.000	KO redundancy ~ watemp
302	0.355	0.132	1.000	KO redundancy ~ watemp
303	0.265	0.127	1.000	KO redundancy ~ watemp
304	0.174	0.255	1.000	KO redundancy ~ watemp
305	0.369	0.021	1.000	KO redundancy ~ watemp
306	0.004	0.198	1.000	KO redundancy ~ watemp
307	0.327	0.246	1.000	KO redundancy ~ watemp
308	0.011	0.243	1.000	KO redundancy ~ watemp
309	0.314	0.049	1.000	KO redundancy ~ watemp
310	0.035	0.249	1.000	KO redundancy ~ watemp
311	0.253	0.164	1.000	KO redundancy ~ watemp

312	0.253	0.164	1.000	KO redundancy ~ watemp
313	0.174	0.184	1.000	KO redundancy ~ watemp
314	0.000	0.550	0.001	KO redundancy ~ watemp
315	0.046	0.257	1.000	KO redundancy ~ watemp
316	0.046	0.257	1.000	KO redundancy ~ watemp
317	0.033	0.299	1.000	KO redundancy ~ watemp
318	0.003	0.354	1.000	KO redundancy ~ watemp
319	0.096	0.242	1.000	KO redundancy ~ watemp
320	0.004	0.258	1.000	KO redundancy ~ watemp
321	0.277	0.171	1.000	KO redundancy ~ watemp
322	0.277	0.162	1.000	KO redundancy ~ watemp
323	0.091	0.242	1.000	KO redundancy ~ watemp
324	0.277	0.163	1.000	KO redundancy ~ watemp
325	0.277	0.163	1.000	KO redundancy ~ watemp
326	0.004	0.366	1.000	KO redundancy ~ watemp
327	0.121	0.202	1.000	KO redundancy ~ watemp
328	0.121	0.202	1.000	KO redundancy ~ watemp
329	0.020	0.248	1.000	KO redundancy ~ watemp
330	0.183	0.178	1.000	KO redundancy ~ watemp
331	0.006	0.367	1.000	KO redundancy ~ watemp
332	0.001	0.275	0.329	KO redundancy ~ watemp
333	0.211	0.091	1.000	KO redundancy ~ watemp
334	0.231	0.076	1.000	KO redundancy ~ watemp
335	0.049	0.266	1.000	KO redundancy ~ watemp
336	0.277	0.161	1.000	KO redundancy ~ watemp
337	0.001	0.324	0.509	KO redundancy ~ watemp
338	0.038	0.235	1.000	KO redundancy ~ watemp
339	0.201	0.146	1.000	KO redundancy ~ watemp
340	0.201	0.146	1.000	KO redundancy ~ watemp
341	0.017	0.225	1.000	KO redundancy ~ watemp
342	0.174	0.222	1.000	KO redundancy ~ watemp
343	0.142	0.132	1.000	KO redundancy ~ watemp
344	0.023	0.182	1.000	KO redundancy ~ watemp
345	0.121	0.133	1.000	KO redundancy ~ watemp
346	0.121	0.136	1.000	KO redundancy ~ watemp
347	0.121	0.133	1.000	KO redundancy ~ watemp
348	0.002	0.278	0.628	KO redundancy ~ watemp
349	0.242	0.132	1.000	KO redundancy ~ watemp
350	0.068	0.207	1.000	KO redundancy ~ watemp

351	0.369	0.222	1.000	KO redundancy ~ watemp
352	0.068	0.207	1.000	KO redundancy ~ watemp
353	0.086	0.277	1.000	KO redundancy ~ watemp
354	0.002	0.339	0.943	KO redundancy ~ watemp
355	0.301	0.120	1.000	KO redundancy ~ watemp
356	0.086	0.259	1.000	KO redundancy ~ watemp
357	0.015	0.270	1.000	KO redundancy ~ watemp
358	0.127	0.228	1.000	KO redundancy ~ watemp
359	0.001	0.342	0.565	KO redundancy ~ watemp
360	0.001	0.346	0.410	KO redundancy ~ watemp
361	0.134	0.191	1.000	KO redundancy ~ watemp
362	0.007	0.288	1.000	KO redundancy ~ watemp
363	0.020	0.229	1.000	KO redundancy ~ watemp
364	0.003	0.293	1.000	KO redundancy ~ watemp
365	0.231	0.141	1.000	KO redundancy ~ watemp
366	0.052	0.212	1.000	KO redundancy ~ watemp
367	0.052	0.255	1.000	KO redundancy ~ watemp
368	0.301	0.140	1.000	KO redundancy ~ watemp
369	0.108	0.254	1.000	KO redundancy ~ watemp
370	0.003	0.302	1.000	KO redundancy ~ watemp
371	0.021	0.212	1.000	KO redundancy ~ watemp
372	0.003	0.297	1.000	KO redundancy ~ watemp
373	0.086	0.143	1.000	KO redundancy ~ watemp
374	0.072	0.326	1.000	KO redundancy ~ watemp
375	0.076	0.127	1.000	KO redundancy ~ watemp
376	0.072	0.305	1.000	KO redundancy ~ watemp
377	0.072	0.306	1.000	KO redundancy ~ watemp
378	0.072	0.308	1.000	KO redundancy ~ watemp
379	0.253	0.063	1.000	KO redundancy ~ watemp
380	0.063	0.327	1.000	KO redundancy ~ watemp
381	0.017	0.267	1.000	KO redundancy ~ watemp
382	0.253	0.100	1.000	KO redundancy ~ watemp
383	0.192	0.122	1.000	KO redundancy ~ watemp
384	0.010	0.333	1.000	KO redundancy ~ watemp
385	0.035	0.139	1.000	KO redundancy ~ watemp
386	0.355	0.084	1.000	KO redundancy ~ watemp
387	0.006	0.280	1.000	KO redundancy ~ watemp
388	0.060	0.159	1.000	KO redundancy ~ watemp
389	0.081	0.193	1.000	KO redundancy ~ watemp
390	0.081	0.179	1.000	KO redundancy ~ watemp
391	0.081	0.296	1.000	KO redundancy ~ watemp
392	0.063	0.167	1.000	KO redundancy ~ watemp
393	0.005	0.224	1.000	KO redundancy ~ watemp
394	0.289	-0.220	1.000	KO redundancy ~ watemp

Leave-one-cluster-out tests output for the models of genome redundancy index. cluster = identifier of the phylogenetic cluster, p = p-value of the Wilcoxon signed rank test, median_effect = relative median difference in coefficient, padj = adjusted p-value, test = relationship that was tested.

❖ *Supplementary methods*

<u>A) DNA extraction protocol from alpine stream biofilms (rDNA)</u>

Remark: Every time you open the tubes make sure that there is no liquid on the lids by applying a short spin

   1. In a 1.5-ml tube add 10-20% (~300 ul) 0.1 mm Zirconium beads (Cole-Parmer 36270-62)

   per volume and 750 ml of Lysis buffer mixed with 0.5ul of RNase (100 mg/ml, Qiagen

   19101)

   2. Add 0.05 to 0.1 g and bead-beat at 6000 r/min, 2x 15sec-break 15sec (Precellys 24

   homogenizer)

   3. Incubate at 37 °C for 1 h with gentle agitation

   4. Spin samples, add 5 ul Proteinase K (20 mg/ml, Fisher Scientific Cat.No. 25530049) and

   mix a few times

   5. Incubate statically at 70 °C for 10 min

   6. Centrifuge at 12.000 x g for 1 min and transfer all supernatant to a new 1.5 ml microtube

   7. Spin samples and add 1 vol of Phenol:CHCl3:IAA (Fisher Scientific, 15593049)

   8. Mix thoroughly and centrifuge at 13.000 x g for 10 min

   9. Transfer aqueous phase into a new 1.5-ml tube and add 1 vol ml Chloroform – isoamyl

   alcohol mixture (Sigma, 25666)

   10. Mix thoroughly and centrifuge at 13.000 x g for 5 min

   11. Transfer supernatant to a new 2ml tube and then add 1/10th volume of 3M sodium acetate

   (pH 5.2) (Sigma S7899)

   12. Add 0.7 volumes of ice-cold Isopropanol (Sigma I9516) and mix thoroughly

   13. Precipitate DNA at -20 °C overnight

   14. Centrifuge at 12.000 x g at 4 °C for 15 min

   15. Remove supernatant and discard without disturbing the pellet

   16. Wash 2 times with 0.4 ml of 70% EtOH and centrifuge at 13.000 g at 4 °C for 10 min

17. Air-dry the pellet, and elute with 100 ul RNase-free, DNase-free water (Qiagen 129112)

18. Let DNA pellet to dissolve o/n at 4 °C

19. Use 2 ul sample to quantify DNA using Qubit HS dsDNA (Invitrogen Q32854)

B) NCBI Accessions for Polaromonas genomes

Name	AccessionID
OUT1	GCF_001955735.1_ASM195573v1
DB1	GCF_000013865.1_ASM1386v1
DB2	GCF_000015505.1_ASM1550v1
DB3	GCF_000282655.1_Polaromonas.strCF318_v1.0
DB4	GCF_000688115.1_ASM68811v1
DB5	GCF_000709345.1_Polaromonas_sp.
DB6	GCF_001598235.1_ASM159823v1
DB7	GCF_002001015.1_ASM200101v1
DB8	GCF_002002705.1_ASM200270v1
DB9	GCF_002379085.1_ASM237908v1
DB10	GCF_002379095.1_ASM237909v1
DB11	GCF_003711205.1_ASM371120v1
DB12	GCF_009664225.1_ASM966422v1
DB13	GCF_012584515.1_ASM1258451v1
DB14	GCF_014641715.1_ASM1464171v1
DB15	GCF_015751795.1_ASM1575179v1
DB16	GCF_015752205.1_ASM1575220v1
DB17	GCF_015752225.1_ASM1575222v1
DB18	GCF_900103405.1_IMG-taxon_2636416056_annotated_assembly

DB19	GCF_900112285.1_IMG-taxon_2609459740_annotated_assembly
DB20	GCF_900116715.1_IMG-taxon_2615840640_annotated_ass

❖ *Supplementary note*

Sloan model summary: The dispersal rate coefficient (m) and the goodness of fit of the beta distribution model (R2) based on the Sloan neutral model analyses are indicated for New Zealand and Caucasus with respect to the metabarcoding information per amplicon sequence variant (ASV).

	16S rRNA gene amplicons		18S rRNA gene amplicons	
	R^2	m	R^2	m
New Zealand	-0.726	3.7E-04±4E-04	-0.204	4.3E-04±3E-04
Caucasus	-1.11	6.4E-04±5E-04	-0.527	0.57±0.31

❖ *Supplementary data legends*

Supplementary data tables are available at the following link:

https://www.nature.com/articles/s41467-022-29914-0

Supplementary Data 1. CAZyme abundances

Normalised abundances of the carbohydrate-active enzymes (CAZymes) across all samples. AA: auxilliary activities, CBM: non-catalytic carbohydrate-binding modules, CE: carbohydrate esterases, GH: glycoside hydrolases, GT: glycosyltransferases, PL: polysaccharide lyases, and SLH: S-layer homology domain enzymes.

Supplementary Data 2. Public metagenomes

Metadata including ecosystems and location of the publicly-available metagenomes used for comparing Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs.

Supplementary Data 3. Enriched KEGG orthologs in epilithic biofilms.

KEGG orthology (KO) genes enriched in epilithic biofilms compared to other metagenomic datasets. Gene enrichment was assessed using DEseq2, where the adjusted p-value < 0.05 was considered to be significant.

Supplementary Data 4. COG functions enriched in GFS Polaromonas spp..

Clustered-orthologous genes (COG20) functions enriched in Polaromonas spp. compared to genomes available via RefSeq. Gene enrichment was assessed using DEseq2, where the adjusted p-value < 0.05 was considered to be significant.

Supplementary Data 5. Sample metadata.

Sample metadata including physico-chemical parameters such as pH, turbidity, conductivity, dissolved organic carbon, temperature, and CO2 saturation.

Supplementary Data 6. Accession information.

NCBI sequence read archive (SRA) accession IDs for all samples used in the study including hyperlinks for each sample.

Supplementary Data 7. Osmotic stress genes

Gene counts for osmotic stress found in respective Phyla.

Bibliography

Ahmed NU, Park J-I, Jung H-J, Kang K-K, Hur Y, Lim Y-P, Nou I-S. 2012. Molecular characterization of stress resistance-related chitinase genes of Brassica rapa. *Plant Physiol Biochem* **58**:106–115. doi:10.1016/j.plaphy.2012.06.015

Albright S, Louca S. 2023. Trait biases in microbial reference genomes. *Sci Data* **10**:84. doi:10.1038/s41597-023-01994-7

Almpanis A, Swain M, Gatherer D, McEwan N 2018. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microb Genomics* **4**:e000168. doi:10.1099/mgen.0.000168

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**:1144–1146. doi:10.1038/nmeth.3103

Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Loman NJ, Andersson AF, Quince C. 2013. CONCOCT: Clustering cONtigs on COverage and ComposiTion. doi:10.48550/arXiv.1312.4038

Alonso-Sáez L, Zeder M, Harding T, Pernthaler J, Lovejoy C, Bertilsson S, Pedrós-Alió C. 2014. Winter bloom of a rare betaproteobacterium in the Arctic Ocean. *Front Microbiol* **5**.

Altmann A, Toloşi L, Sander O, Lengauer T. 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* **26**:1340–1347. doi:10.1093/bioinformatics/btq134

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2

Alves RJE, Kerou M, Zappe A, Bittner R, Abby SS, Schmidt HA, Pfeifer K, Schleper C. 2019. Ammonia Oxidation by the Arctic Terrestrial Thaumarchaeote Candidatus Nitrosocosmicus arcticus Is Stimulated by Increasing Temperatures. *Front Microbiol* **10**.

Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, Morales RL, Berthiaume CT, Parker MS, Djunaedi B, Ingalls AE, Parsek MR, Moran MA, Armbrust EV. 2015. Interaction and signalling between a cosmopolitan

phytoplankton and associated bacteria. *Nature* **522**:98–101. doi:10.1038/nature14488

Anesio AM, Hodson AJ, Fritz A, Psenner R, Sattler B. 2009. High microbial activity on glaciers: importance to the global carbon cycle. *Glob Change Biol* **15**:955–960. doi:https://doi.org/10.1111/j.1365-2486.2008.01758.x

Anesio AM, Laybourn-Parry J. 2012. Glaciers and ice sheets as a biome. *Trends Ecol Evol* **27**:219–225. doi:10.1016/j.tree.2011.09.012

Anesio AM, Lutz S, Chrismas NAM, Benning LG. 2017. The microbiome of glaciers and ice sheets. *Npj Biofilms Microbiomes* **3**:1–11. doi:10.1038/s41522-017-0019-0

Anesio AM, Mindl B, Laybourn-Parry J, Hodson AJ, Sattler B. 2007. Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J Geophys Res Biogeosciences* **112**. doi:10.1029/2006JG000350

Araújo MB, Anderson RP, Márcia Barbosa A, Beale CM, Dormann CF, Early R, Garcia RA, Guisan A, Maiorano L, Naimi B, O'Hara RB, Zimmermann NE, Rahbek C. 2019. Standards for distribution models in biodiversity assessments. *Sci Adv* **5**:eaat4858. doi:10.1126/sciadv.aat4858

Arella D, Dilucca M, Giansanti A. 2021. Codon usage bias and environmental adaptation in microbial organisms. *Mol Genet Genomics* **296**:751–762. doi:10.1007/s00438-021-01771-4

Avcı B, Krüger K, Fuchs BM, Teeling H, Amann RI. 2020. Polysaccharide niche partitioning of distinct Polaribacter clades during North Sea spring algal blooms. *ISME J* **14**:1369–1383. doi:10.1038/s41396-020-0601-y

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. doi:10.1128/AEM.02101-09

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Res* **33**:D154–D159. doi:10.1093/nar/gki070

Barnett TP, Adam JC, Lettenmaier DP. 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **438**:303–309. doi:10.1038/nature04141

Barry R, Gan TY. 2011. The Global Cryosphere: Past, Present and Future. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511977947

Battin TJ, Besemer K, Bengtsson MM, Romani AM, Packmann AI. 2016. The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* **14**:251–263. doi:10.1038/nrmicro.2016.15

Battin TJ, Kaplan LA, Denis Newbold J, Hansen CME. 2003. Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* **426**:439–442. doi:10.1038/nature02152

Battin TJ, Sloan WT, Kjelleberg S, Daims H, Head IM, Curtis TP, Eberl L. 2007. Microbial landscapes: new paths to biofilm research. *Nat Rev Microbiol* **5**:76–81. doi:10.1038/nrmicro1556

Battin TJ, Wille A, Psenner R, Richter A. 2004. Large-scale environmental controls on microbial biofilms in high-alpine streams. *Biogeosciences* **1**:159–171. doi:10.5194/bg-1-159-2004

BBMap: A Fast, Accurate, Splice-Aware Aligner (Conference) | OSTI.GOV. https://www.osti.gov/biblio/1241166

Becquet J, Lamouroux N, Condom T, Gouttevin I, Forcellini M, Launay B, Rabatel A, Cauvy-Fraunié S. 2022. Macroinvertebrate distribution associated with environmental variables in alpine streams. *Freshw Biol* **67**:1815–1831. doi:10.1111/fwb.13977

Bellas CM, Schroeder DC, Edwards A, Barker G, Anesio AM. 2020. Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. *Nat Commun* **11**:4403. doi:10.1038/s41467-020-18236-8

Bentkowski P, Van Oosterhout C, Mock T. 2015. A Model of Genome Size Evolution for Prokaryotes in Stable and Fluctuating Environments. *Genome Biol Evol* **7**:2344–2351. doi:10.1093/gbe/evv148

Bentley SD, Parkhill J. 2004. Comparative Genomic Structure of Prokaryotes. *Annu Rev Genet* **38**:771–791. doi:10.1146/annurev.genet.38.072902.094318

Bernhardt ES, Blaszczak JR, Ficken CD, Fork ML, Kaiser KE, Seybold EC. 2017. Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept. *Ecosystems* **20**:665–682. doi:10.1007/s10021-016-0103-y

Besemer K, Singer G, Hödl I, Battin TJ. 2009. Bacterial Community Composition of Stream Biofilms in Spatially Variable-Flow Environments. *Appl Environ Microbiol* **75**:7189–7195. doi:10.1128/AEM.01284-09

Biolabs NE. 2020. E7805 NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® Protocol for use with Inputs ≤ 100 ng.

Blagojevic DP, Grubor-Lajsic GN, Spasic MB. 2011. Cold defence responses: the role of oxidative stress. *Front Biosci-Sch* **3**:416–427. doi:10.2741/S161

Blighe K. 2023. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling.

Bobay L-M, Ochman H. 2017. The Evolution of Bacterial Genome Architecture. *Front Genet* **8**.

Boetius A, Anesio AM, Deming JW, Mikucki JA, Rapp JZ. 2015. Microbial ecology of the cryosphere: sea ice and glacial habitats. *Nat Rev Microbiol* **13**:677–690. doi:10.1038/nrmicro3522

Boix Canadell M, Escoffier N, Ulseth AJ, Lane SN, Battin TJ. 2019. Alpine Glacier Shrinkage Drives Shift in Dissolved Organic Carbon Export From Quasi-Chemostasis to Transport Limitation. *Geophys Res Lett* **46**:8872–8881. doi:10.1029/2019GL083424

Boix Canadell M, Gómez-Gener L, Ulseth AJ, Clémençon M, Lane SN, Battin TJ. 2021. Regimes of primary production and their drivers in Alpine streams. *Freshw Biol* **66**:1449–1463. doi:10.1111/fwb.13730

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**:90. doi:10.1186/s40168-018-0470-z

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. doi:10.1093/bioinformatics/btu170

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X,

Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**:852–857. doi:10.1038/s41587-019-0209-9

Bourquin M, Busi SB, Fodelianakis S, Peter H, Washburne A, Kohler TJ, Ezzat L, Michoud G, Wilmes P, Battin TJ. 2022. The microbiome of cryospheric ecosystems. *Nat Commun* **13**:3087. doi:10.1038/s41467-022-30816-4

Bowman JP. 2017. Genomics of Psychrophilic Bacteria and Archaea In: Margesin R, editor. Psychrophiles: From Biodiversity to Biotechnology. Cham: Springer International Publishing. pp. 345–387. doi:10.1007/978-3-319-57057-0_15

Boyd ES, Lange RK, Mitchell AC, Havig JR, Hamilton TL, Lafrenière MJ, Shock EL, Peters JW, Skidmore M. 2011. Diversity, Abundance, and Potential Activity of Nitrifying and Nitrate-Reducing Microbial Assemblages in a Subglacial Ecosystem. *Appl Environ Microbiol* **77**:4778–4787. doi:10.1128/AEM.00376-11

Boyd ES, Skidmore M, Mitchell AC, Bakermans C, Peters JW. 2010. Methanogenesis in subglacial sediments. *Environ Microbiol Rep* **2**:685–692. doi:https://doi.org/10.1111/j.1758-2229.2010.00162.x

Brandani J, Peter H, Busi SB, Kohler TJ, Fodelianakis S, Ezzat L, Michoud G, Bourquin M, Pramateftaki P, Roncoroni M, Lane SN, Battin TJ. 2022. Spatial patterns of benthic biofilm diversity among streams draining proglacial floodplains. *Front Microbiol* **13**.

Brandani J, Peter H, Fodelianakis S, Kohler TJ, Bourquin M, Michoud G, Busi SB, Ezzat L, Lane S, Battin TJ. 2023. Homogeneous Environmental Selection Structures the Bacterial Communities of Benthic Biofilms in Proglacial Floodplain Streams. *Appl Environ Microbiol* **89**:e02010-22. doi:10.1128/aem.02010-22

Brown LE, Hannah DM, Milner AM. 2007. Vulnerability of alpine stream biodiversity to shrinking glaciers and snowpacks. *Glob Change Biol* **13**:958–966. doi:10.1111/j.1365-2486.2007.01341.x

Brown LE, Khamis K, Wilkes M, Blaen P, Brittain JE, Carrivick JL, Fell S, Friberg N, Füreder L, Gislason GM, Hainie S, Hannah DM, James WHM, Lencioni V, Olafsson JS, Robinson CT, Saltveit SJ, Thompson C, Milner AM. 2018. Functional diversity and community assembly of river invertebrates show globally consistent responses to decreasing glacier cover. *Nat Ecol Evol* **2**:325–333. doi:10.1038/s41559-017-0426-x

Brown SP, Jumpponen A. 2019. Microbial Ecology of Snow Reveals Taxa-Specific Biogeographical Structure. *Microb Ecol* **77**:946–958. doi:10.1007/s00248-019-01357-z

Buck M, Mehrshad M, Bertilsson S. 2022. mOTUpan: a robust Bayesian approach to leverage metagenome-assembled genomes for core-genome estimation. *NAR Genomics Bioinforma* **4**:lqac060. doi:10.1093/nargab/lqac060

Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJ. 2016. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* **10**:655–664. doi:10.1038/ismej.2015.142

Burrows LL. 2012. Pseudomonas aeruginosa Twitching Motility: Type IV Pili in Action. *Annu Rev Microbiol* **66**:493–520. doi:10.1146/annurev-micro-092611-150055

Busby JR. 1991. BIOCLIM-a bioclimate analysis and prediction system. *Plant Prot Q Aust*.

Busi SB, Bourquin M, Fodelianakis S, Michoud G, Kohler TJ, Peter H, Pramateftaki P, Styllas M, Tolosano M, De Staercke V. 2022. Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams. *Nat Commun* **13**:2168.

Busi SB, Pramateftaki P, Brandani J, Fodelianakis S, Peter H, Halder R, Wilmes P, Battin TJ. 2020. Optimised biomolecular extraction for metagenomic analysis of microbial biofilms from high-mountain streams. *PeerJ* **8**:e9973. doi:10.7717/peerj.9973

Cai M, Richter-Heitmann T, Yin X, Huang W-C, Yang Y, Zhang C, Duan C, Pan J, Liu Yang, Liu Yue, Friedrich MW, Li M. 2021. Ecological features and global distribution of Asgard archaea. *Sci Total Environ* **758**:143581. doi:10.1016/j.scitotenv.2020.143581

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. doi:10.1093/bioinformatics/btp348

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**:1621–1624. doi:10.1038/ismej.2012.8

Carey CJ, Hart SC, Aciego SM, Riebe CS, Blakowski MA, Aronson EL. 2016. Microbial Community Structure of Subalpine Snow in the Sierra Nevada, California. *Arct Antarct Alp Res* **48**:685–701. doi:10.1657/AAAR0015-062

Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist, metagenomics reveal. *Environ Microbiol* **14**:347–355. doi:10.1111/j.1462-2920.2011.02668.x

Casillo A, Parrilli E, Sannino F, Mitchell DE, Gibson MI, Marino G, Lanzetta R, Parrilli M, Cosconati S, Novellino E, Randazzo A, Tutino ML, Corsaro MM. 2017. Structure-activity relationship of the exopolysaccharide from a psychrophilic bacterium: A strategy for cryoprotection. *Carbohydr Polym* **156**:364–371. doi:10.1016/j.carbpol.2016.09.037

Cauvy-Fraunié S, Dangles O. 2019. A global synthesis of biodiversity responses to glacier retreat. *Nat Ecol Evol* **3**:1675–1685. doi:10.1038/s41559-019-1042-8

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW, Danovaro R, Foreman CM, Huisman J, Hutchins DA, Jansson JK, Karl DM, Koskella B, Mark Welch DB, Martiny JBH, Moran MA, Orphan VJ, Reay DS, Remais JV, Rich VI, Singh BK, Stein LY, Stewart FJ, Sullivan MB, van Oppen MJH, Weaver SC, Webb EA, Webster NS. 2019. Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol* **17**:569–586. doi:10.1038/s41579-019-0222-5

Center S. 1989. Diel fluctuations in bacterial activity on streambed substrata during vernal algal blooms: effects of temperature, water chemistry, and habitat. *Stroud Water Res Cent*. https://stroudcenter.org/publications/diel-fluctuations-bacterial-activity-streambed-substrata-vernal-algal-blooms-effects-temperature-water-chemistry-habitat/

Chaudhari NM, Overholt WA, Figueroa-Gonzalez PA, Taubert M, Bornemann TLV, Probst AJ, Hölzer M, Marz M, Küsel K. 2021. The economical lifestyle of CPR bacteria in

groundwater allows little preference for environmental drivers. *Environ Microbiome* **16**:24. doi:10.1186/s40793-021-00395-w

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**:1925–1927. doi:10.1093/bioinformatics/btz848

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. doi:10.1093/bioinformatics/bty560

Chiriac M-C, Haber M, Salcher MM. 2023. Adaptive genetic traits in pelagic freshwater microbes. *Environ Microbiol* **25**:606–641. doi:10.1111/1462-2920.16313

Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* **20**:1203–1212. doi:10.1038/s41592-023-01940-w

Christie-Oleza JA, Sousoni D, Lloyd M, Armengaud J, Scanlan DJ. 2017. Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nat Microbiol* **2**:1–10. doi:10.1038/nmicrobiol.2017.100

Christner BC, Kvitko BH, Reeve JN. 2003. Molecular identification of Bacteria and Eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles* **7**:177–183. doi:10.1007/s00792-002-0309-0

Chróst RJ, editor. 1991. Microbial Enzymes in Aquatic Environments, Brock/Springer Series in Contemporary Bioscience. New York, NY: Springer. doi:10.1007/978-1-4612-3090-8

Chuckran PF, Hungate BA, Schwartz E, Dijkstra P. 2021. Variation in genomic traits of microbial communities among ecosystems. *FEMS Microbes* **2**:xtab020. doi:10.1093/femsmc/xtab020

Cinar O, Viechtbauer W. 2022. The poolr Package for Combining Independent and Dependent p Values. *J Stat Softw* **101**:1–42. doi:10.18637/jss.v101.i01

Clason C, Rangecroft S, Owens PN, Łokas E, Baccolo G, Selmes N, Beard D, Kitch J, Dextre RM, Morera S, Blake W. 2023. Contribution of glaciers to water, energy and food security in mountain regions: current perspectives and future priorities. *Ann Glaciol* 1–6. doi:10.1017/aog.2023.14

Cockell CS, Rettberg P, Horneck G, Wynn-Williams DD, Scherer K, Gugg-Helminger A. 2002. Influence of ice and snow covers on the UV exposure of terrestrial microbial communities: dosimetric studies. *J Photochem Photobiol B* **68**:23–32. doi:10.1016/S1011-1344(02)00327-5

Cole JJ. 1982. Interactions Between Bacteria and Algae in Aquatic Ecosystems. *Annu Rev Ecol Syst* **13**:291–314.

Coleman D, Bevitt R, Reinfelds I. 2021. Predicting the Thermal Regime Change of a Regulated Snowmelt River Using a Generalised Additive Model and Analogue Reference Streams. *Environ Process* **8**:511–531. doi:10.1007/s40710-021-00501-7

Coleman ML, Chisholm SW. 2010. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci* **107**:18634–18639. doi:10.1073/pnas.1009480107

Collart F, Guisan A. 2023. Small to train, small to test: Dealing with low sample size in model evaluation. *Ecol Inform* **75**:102106. doi:10.1016/j.ecoinf.2023.102106

Collins RE, Rocap G, Deming JW. 2010. Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ Microbiol* **12**:1828–1841. doi:10.1111/j.1462-2920.2010.02179.x

Colón-González FJ, Fezzi C, Lake IR, Hunter PR. 2013. The Effects of Weather and Climate Change on Dengue. *PLoS Negl Trop Dis* **7**:e2503. doi:10.1371/journal.pntd.0002503

Comte J, Culley AI, Lovejoy C, Vincent WF. 2018. Microbial connectivity and sorting in a High Arctic watershed. *ISME J* **12**:2988–3000. doi:10.1038/s41396-018-0236-4

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**:2938–2940. doi:10.1093/bioinformatics/btx364

Correa-Garcia S, Constant P, Yergeau E. 2023. The forecasting power of the microbiome. *Trends Microbiol* **31**:444–452. doi:10.1016/j.tim.2022.11.013

Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**:90–93. doi:10.1038/nature04056

Crook JA, Forster PM. 2014. Comparison of surface albedo feedback in climate models and observations. *Geophys Res Lett* **41**:1717–1723. doi:10.1002/2014GL059280

Csardi G, Nepusz T. The igraph software package for complex network research.

Dalluge JJ, Hamamoto T, Horikoshi K, Morita RY, Stetter KO, McCloskey JA. 1997. Posttranscriptional modification of tRNA in psychrophilic bacteria. *J Bacteriol* **179**:1918. doi:10.1128/jb.179.6.1918-1923.1997

D'Amico S, Collins T, Marx J, Feller G, Gerday C, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**:385–389. doi:10.1038/sj.embor.7400662

De Maayer P, Anderson D, Cary C, Cowan DA. 2014. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* **15**:508–517. doi:10.1002/embr.201338170

Dieser M, Smith HJ, Ramaraj T, Foreman CM. 2019. Janthinobacterium CG23_2: Comparative Genome Analysis Reveals Enhanced Environmental Sensing and Transcriptional Regulation for Adaptation to Life in an Antarctic Supraglacial Stream. *Microorganisms* **7**:454. doi:10.3390/microorganisms7100454

Dini-Andreote F, Stegen JC, Elsas JD van, Salles JF. 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci* **112**:E1326–E1332. doi:10.1073/pnas.1414261112

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**:927–930. doi:10.1111/j.1654-1103.2003.tb02228.x

Dobinski W. 2011. Permafrost. *Earth-Sci Rev* **108**:158–169. doi:10.1016/j.earscirev.2011.06.007

Dormann CF, Fründ J, Blüthgen N, Gruber B. 2009. Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks.

Dorrell RG, Kuo A, Füssy Z, Richardson E, Salamov A, Zarevski N, Freyria NJ, Ibarbalz FM, Jenkins J, Karlusich JJP, Steindorff AS, Edgar RE, Handley L, Lail K, Lipzen A, Lombard V, McFarlane J, Nef C, Vanclová AMGN, Peng Y, Plott C, Potvin M, Vieira FRJ, Barry K, Dacks JB, Vargas C de, Henrissat B, Pelletier E, Schmutz J, Wincker P, Bowler C, Grigoriev IV, Lovejoy C. 2021. Within-Arctic horizontal gene transfer as a driver of convergent evolution in distantly related microalgae. doi:10.1101/2021.07.31.454568

Dunham EC, Dore JE, Skidmore ML, Roden EE, Boyd ES. 2021. Lithogenic hydrogen supports microbial primary production in subglacial and proglacial environments. *Proc Natl Acad Sci* **118**:e2007051117. doi:10.1073/pnas.2007051117

Dutta A, Chaudhuri K. 2010. Analysis of tRNA composition and folding in psychrophilic, mesophilic and thermophilic genomes: indications for thermal adaptation. *FEMS Microbiol Lett* **305**:100–108. doi:10.1111/j.1574-6968.2010.01922.x

Dzubakova K, Peter H, Bertuzzo E, Juez C, Franca MJ, Rinaldo A, Battin TJ. 2018. Environmental heterogeneity promotes spatial resilience of phototrophic biofilms in streambeds. *Biol Lett* **14**:20180432. doi:10.1098/rsbl.2018.0432

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLOS Comput Biol* **7**:e1002195. doi:10.1371/journal.pcbi.1002195

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797. doi:10.1093/nar/gkh340

Elek A, Kuzman M, Vlahoviček K. 2019. coRdon: codon usage analysis and prediction of gene expressivity https://github.com/BioinfoHR/coRdon. *R Package Version* **1**.

Elser JJ, Wu C, González AL, Shain DH, Smith HJ, Sommaruga R, Williamson CE, Brahney J, Hotaling S, Vanderwall J, Yu J, Aizen V, Aizen E, Battin TJ, Camassa R, Feng X, Jiang H, Lu L, Qu JJ, Ren Z, Wen J, Wen L, Woods HA, Xiong X, Xu J, Yu G, Harper JT, Saros JE. 2020. Key rules of life and the fading cryosphere: Impacts in alpine lakes and streams. *Glob Change Biol* **26**:6644–6656. doi:10.1111/gcb.15362

Envidat. https://envidat.ch/#/metadata/bioclim_plus

Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**:e1319. doi:10.7717/peerj.1319

Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**:3–6. doi:10.1038/s41564-020-00834-3

Ezzat L, Fodelianakis S, Kohler TJ, Bourquin M, Brandani J, Busi SB, Daffonchio D, Staercke VD, Marasco R, Michoud G, Oppliger E, Peter H, Pramateftaki P, Schön M, Styllas M, Tadei V, Tolosano M, Battin TJ. 2022. Benthic Biofilms in Glacier-Fed

Streams from Scandinavia to the Himalayas Host Distinct Bacterial Communities Compared with the Streamwater. *Appl Environ Microbiol* **88**:e00421-22. doi:10.1128/aem.00421-22

Federhen S. 2012. The NCBI Taxonomy database. *Nucleic Acids Res* **40**:D136–D143. doi:10.1093/nar/gkr1178

Fell SC, Carrivick JL, Brown LE. 2017. The Multitrophic Effects of Climate Change and Glacier Retreat in Mountain Rivers. *BioScience* **67**:897–911. doi:10.1093/biosci/bix107

Feller G, Gerday C. 2003. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1**:200–208. doi:10.1038/nrmicro773

Feller G, Narinx E, Arpigny JL, Aittaleb M, Baise E, Genicot S, Gerday C. 1996. Enzymes from psychrophilic organisms. *FEMS Microbiol Rev* **18**:189–202. doi:10.1111/j.1574-6976.1996.tb00236.x

Fellman JB, Hood E, Raymond PA, Hudson J, Bozeman M, Arimitsu M. 2015. Evidence for the assimilation of ancient glacier organic carbon in a proglacial stream food web. *Limnol Oceanogr* **60**:1118–1128. doi:10.1002/lno.10088

Fodelianakis S, Washburne AD, Bourquin M, Pramateftaki P, Kohler TJ, Styllas M, Tolosano M, De Staercke V, Schön M, Busi SB, Brandani J, Wilmes P, Peter H, Battin TJ. 2022. Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. *ISME J* **16**:666–675. doi:10.1038/s41396-021-01106-6

Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, Saul DJ, Lawson W. 2004. Culturable Bacteria in Subglacial Sediments and Ice from Two Southern Hemisphere Glaciers. *Microb Ecol* **47**:329–340. doi:10.1007/s00248-003-1036-5

Foster ZSL, Sharpton TJ, Grünwald NJ. 2017. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput Biol* **13**:e1005404. doi:10.1371/journal.pcbi.1005404

Fountain AG, Campbell JL, Schuur EAG, Stammerjohn SE, Williams MW, Ducklow HW. 2012. The Disappearing Cryosphere: Impacts and Ecosystem Responses to Rapid Cryosphere Loss. *BioScience* **62**:405–415. doi:10.1525/bio.2012.62.4.11

Frémont P, Gehlen M, Vrac M, Leconte J, Delmont TO, Wincker P, Iudicone D, Jaillon O. 2022. Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat Clim Change* **12**:393–401. doi:10.1038/s41558-022-01314-8

Frey B, Rime T, Phillips M, Stierli B, Hajdas I, Widmer F, Hartmann M. 2016. Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiol Ecol* **92**. doi:10.1093/femsec/fiw018

Friedman J, Alm EJ. 2012. Inferring Correlation Networks from Genomic Survey Data. *PLOS Comput Biol* **8**:e1002687. doi:10.1371/journal.pcbi.1002687

GDAL Development Team. GDAL - Geospatial Data Abstraction Library, Version x.x.x. Open Source Geospatial Foundation.

Giersch JJ, Hotaling S, Kovach RP, Jones LA, Muhlfeld CC. 2017. Climate-induced glacier and snow loss imperils alpine stream insects. *Glob Change Biol* **23**:2577–2589. doi:10.1111/gcb.13565

Gillies S, others. 2013. Rasterio: geospatial raster I/O for Python programmers.

Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J* **8**:1553–1565. doi:10.1038/ismej.2014.60

Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. 2020. Pangenomics Comes of Age: From Bacteria to Plant and Animal Applications. *Trends Genet* **36**:132–145. doi:10.1016/j.tig.2019.11.006

Goordial J. 2021. Cryomicrobial Ecology: Still Much To Learn about Life Left Out in the Cold. *mSystems* **6**:e00852-21. doi:10.1128/mSystems.00852-21

Goordial J, Raymond-Bouchard I, Ronholm J, Shapiro N, Woyke T, Whyte L, Bakermans C. 2015. Improved-high-quality draft genome sequence of Rhodococcus sp. JG-3, a eurypsychrophilic Actinobacteria from Antarctic Dry Valley permafrost. *Stand Genomic Sci* **10**:61. doi:10.1186/s40793-015-0043-8

Gooseff MN, McKnight DM, Runkel RL, Duff JH. 2004. Denitrification and hydrologic transient storage in a glacial meltwater stream, McMurdo Dry Valleys, Antarctica. *Limnol Oceanogr* **49**:1884–1895. doi:10.4319/lo.2004.49.5.1884

Gottfried M, Pauli H, Futschik A, Akhalkatsi M, Barančok P, Benito Alonso JL, Coldea G, Dick J, Erschbamer B, Fernández Calzado MR, Kazakis G, Krajči J, Larsson P, Mallaun M, Michelsen O, Moiseev D, Moiseev P, Molau U, Merzouki A, Nagy L, Nakhutsrishvili G, Pedersen B, Pelino G, Puscas M, Rossi G, Stanisci A, Theurillat J-P, Tomaselli M, Villar L, Vittoz P, Vogiatzakis I, Grabherr G. 2012. Continent-

wide response of mountain vegetation to climate change. *Nat Clim Change* **2**:111–115. doi:10.1038/nclimate1329

Graham ED, Heidelberg JF, Tully BJ. 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* **12**:1861–1866. doi:10.1038/s41396-018-0091-3

Grossman A. 2016. Nutrient Acquisition: The Generation of Bioactive Vitamin B12 by Microalgae. *Curr Biol* **26**:R319–R321. doi:10.1016/j.cub.2016.02.047

Grüss A, Yemane D, Fairweather T. 2016. Exploring the spatial distribution patterns of South African Cape hakes using generalised additive models. *Afr J Mar Sci* **38**:395–409. doi:10.2989/1814232X.2016.1218367

Gupta R, Deswal R. 2014. Refolding of β-Stranded Class I Chitinases of Hippophae rhamnoides Enhances the Antifreeze Activity during Cold Acclimation. *PLOS ONE* **9**:e91723. doi:10.1371/journal.pone.0091723

Guttenplan SB, Kearns DB. 2013. Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev* **37**:849–871. doi:10.1111/1574-6976.12018

Haack TK, McFeters GA. 1982. Nutritional relationships among microorganisms in an epilithic biofilm community. *Microb Ecol* **8**:115–126. doi:10.1007/BF02010445

Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, Wampach L, Schneider JG, Hogan A, de Beaufort C, Wilmes P. 2016. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol* **2**:1–13. doi:10.1038/nmicrobiol.2016.180

Hernández M, Vera-Gargallo B, Calabi-Floody M, King GM, Conrad R, Tebbe CC. 2020. Reconstructing Genomes of Carbon Monoxide Oxidisers in Volcanic Deposits Including Members of the Class Ktedonobacteria. *Microorganisms* **8**:1880. doi:10.3390/microorganisms8121880

Hickl O, Queirós P, Wilmes P, May P, Heintz-Buschart A. 2022. binny: an automated binning algorithm to recover high-quality genomes from complex metagenomic datasets. *Brief Bioinform* **23**:bbac431. doi:10.1093/bib/bbac431

Hood E, Battin TJ, Fellman J, O'Neel S, Spencer RGM. 2015. Storage and release of organic carbon from glaciers and ice sheets. *Nat Geosci* **8**:91–96. doi:10.1038/ngeo2331

Hood E, Fellman JB, Spencer RGM. 2020. Glacier Loss Impacts Riverine Organic Carbon Transport to the Ocean. *Geophys Res Lett* **47**:e2020GL089804. doi:10.1029/2020GL089804

Horesh G, Taylor-Brown A, McGimpsey S, Lassalle F, Corander J, Heinz E, Thomson NR. 2021. Different evolutionary trends form the twilight zone of the bacterial pan-genome. *Microb Genomics* **7**:000670. doi:10.1099/mgen.0.000670

Horgby Å, Segatto PL, Bertuzzo E, Lauerwald R, Lehner B, Ulseth AJ, Vennemann TW, Battin TJ. 2019. Unexpected large evasion fluxes of carbon dioxide from turbulent streams draining the world's mountains. *Nat Commun* **10**:4888. doi:10.1038/s41467-019-12905-z

Hornung C, Poehlein A, Haack FS, Schmidt M, Dierking K, Pohlen A, Schulenburg H, Blokesch M, Plener L, Jung K, Bonge A, Krohn-Molt I, Utpatel C, Timmermann G, Spieck E, Pommerening-Röser A, Bode E, Bode HB, Daniel R, Schmeisser C, Streit WR. 2013. The Janthinobacterium sp. HH01 Genome Encodes a Homologue of the V. cholerae CqsA and L. pneumophila LqsA Autoinducer Synthases. *PLOS ONE* **8**:e55045. doi:10.1371/journal.pone.0055045

Hotaling S, Hood E, Hamilton TL. 2017. Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. *Environ Microbiol* **19**:2935–2948. doi:10.1111/1462-2920.13766

Hoyle JT, Kilroy C, Hicks DM, Brown L. 2017. The influence of sediment mobility and channel geomorphology on periphyton abundance. *Freshw Biol* **62**:258–273. doi:10.1111/fwb.12865

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. doi:10.1093/molbev/msx148

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**:D309–D314. doi:10.1093/nar/gky1085

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* **1**:1–6. doi:10.1038/nmicrobiol.2016.48

Hugonnet R, McNabb R, Berthier E, Menounos B, Nuth C, Girod L, Farinotti D, Huss M, Dussaillant I, Brun F, Kääb A. 2021. Accelerated global glacier mass loss in the

early twenty-first century. *Nature* **592**:726–731. doi:10.1038/s41586-021-03436-z

Huss M, Bookhagen B, Huggel C, Jacobsen D, Bradley R s., Clague J j., Vuille M, Buytaert W, Cayan D r., Greenwood G, Mark B g., Milner A m., Weingartner R, Winder M. 2017. Toward mountains without permanent snow and ice. *Earths Future* **5**:418–435. doi:10.1002/2016EF000514

Huss M, Hock R. 2018. Global-scale hydrological response to future glacier mass loss. *Nat Clim Change* **8**:135–140. doi:10.1038/s41558-017-0049-x

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119. doi:10.1186/1471-2105-11-119

Immerzeel WW, Lutz AF, Andrade M, Bahl A, Biemans H, Bolch T, Hyde S, Brumby S, Davies BJ, Elmore AC, Emmer A, Feng M, Fernández A, Haritashya U, Kargel JS, Koppes M, Kraaijenbrink PDA, Kulkarni AV, Mayewski PA, Nepal S, Pacheco P, Painter TH, Pellicciotti F, Rajaram H, Rupper S, Sinisalo A, Shrestha AB, Viviroli D, Wada Y, Xiao C, Yao T, Baillie JEM. 2020. Importance and vulnerability of the world's water towers. *Nature* **577**:364–369. doi:10.1038/s41586-019-1822-y

Mazel F, Davies TJ, Gallien L, Renaud J, Groussin M, Münkemüller T, Thuiller W. 2016. Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography* **39**:913-920. doi:10.1111/ecog.01694

Inkscape. *Guide Books*. doi:10.5555/2028921

Jacobsen D, Dangles O. 2012. Environmental harshness and global richness patterns in glacier-fed streams. *Glob Ecol Biogeogr* **21**:647–656. doi:10.1111/j.1466-8238.2011.00699.x

Jacobsen D, Milner AM, Brown LE, Dangles O. 2012. Biodiversity under threat in glacier-fed river systems. *Nat Clim Change* **2**:361–364. doi:10.1038/nclimate1435

Jansson JK, Taş N. 2014. The microbial ecology of permafrost. *Nat Rev Microbiol* **12**:414–425. doi:10.1038/nrmicro3262

Jowett IG, Parkyn SM, Richardson J. 2008. Habitat characteristics of crayfish (Paranephrops planifrons) in New Zealand streams using generalised additive models (GAMs). *Hydrobiologia* **596**:353–365. doi:10.1007/s10750-007-9108-z

Kahle D, Wickham H. 2013. ggmap: Spatial Visualization with ggplot2. *R J* **5**:144–161.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**:27–30. doi:10.1093/nar/28.1.27

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**:e7359. doi:10.7717/peerj.7359

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Sci Data* **4**:170122. doi:10.1038/sdata.2017.122

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**:772–780. doi:10.1093/molbev/mst010

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**:1463–1464. doi:10.1093/bioinformatics/btq166

Keuschnig C, Vogel TM, Barbaro E, Spolaor A, Koziol K, Björkman MP, Zdanowicz C, Gallet J-C, Luks B, Layton R, Larose C. 2023. Selection processes of Arctic seasonal glacier snowpack bacterial communities. *Microbiome* **11**:35. doi:10.1186/s40168-023-01473-6

Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**:90. doi:10.1186/s40168-020-00867-0

Kim B-C, Oh HW, Kim H, Park D-S, Hong SG, Lee HK, Bae KS. 2013. Polaribacter sejongensis sp. nov., isolated from Antarctic soil, and emended descriptions of the genus Polaribacter, Polaribacter butkevichii and Polaribacter irgensii. *Int J Syst Evol Microbiol* **63**:4000–4005. doi:10.1099/ijs.0.047100-0

Klawonn I, Van den Wyngaert S, Parada AE, Arandia-Gorostidi N, Whitehouse MJ, Grossart H-P, Dekas AE. 2021. Characterizing the "fungal shunt": Parasitic fungi on diatoms affect carbon flow and bacterial communities in aquatic microbial food webs. *Proc Natl Acad Sci* **118**:e2102225118. doi:10.1073/pnas.2102225118

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**:e1–e1. doi:10.1093/nar/gks808

Kohler J. 2024. In press.

Kohler TJ, Fodelianakis S, Michoud G, Ezzat L, Bourquin M, Peter H, Busi SB, Pramateftaki P, Deluigi N, Styllas M, Tolosano M, de Staercke V, Schön M, Brandani J, Marasco R, Daffonchio D, Wilmes P, Battin TJ. 2022. Glacier shrinkage will accelerate downstream decomposition of organic matter and alters microbiome structure and function. *Glob Change Biol* **28**:3846–3859. doi:10.1111/gcb.16169

Kohler TJ, Peter H, Fodelianakis S, Pramateftaki P, Styllas M, Tolosano M, de Staercke V, Schön M, Busi SB, Wilmes P, Washburne A, Battin TJ. 2020a. Patterns and Drivers of Extracellular Enzyme Activity in New Zealand Glacier-Fed Streams. *Front Microbiol* **11**.

Kohler TJ, Vinšová P, Falteisek L, Žárský JD, Yde JC, Hatton JE, Hawkings JR, Lamarche-Gagnon G, Hood E, Cameron KA, Stibal M. 2020b. Patterns in Microbial Assemblages Exported From the Meltwater of Arctic and Sub-Arctic Glaciers. *Front Microbiol* **11**:669. doi:10.3389/fmicb.2020.00669

Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* **17**:1103–1110. doi:10.1038/s41592-020-00971-x

Konings WN, Albers S-V, Koning S, Driessen AJM. 2002. The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments. *Antonie Van Leeuwenhoek* **81**:61–72. https://doi.org/10.1023/A:1020573408652

Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, Schada von Borzyskowski L, Erb TJ, Stahl DA, Berg IA. 2014. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc Natl Acad Sci* **111**:8239–8244. doi:10.1073/pnas.1402028111

Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci* **101**:3160–3165. doi:10.1073/pnas.0308653100

Kosanic A, Lambers K, Galata S, Kothieringer K, Abderhalden A. 2023. Importance of Cultural Ecosystem Services for Cultural Identity and Wellbeing in the Lower Engadine, Switzerland. *Land* **12**:2156. doi:10.3390/land12122156

Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**:2520–2522. doi:10.1093/bioinformatics/bts480

Krinos AI, Hu SK, Cohen NR, Alexander H. 2020. EUKulele: Taxonomic annotation of the unsung eukaryotic microbes. doi:10.48550/arXiv.2011.00089

Kuhn M. 2001. The nutrient cycle through snow and ice, a review. *Aquat Sci* **63**:150–167. doi:10.1007/PL00001348

Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Comput Biol* **11**:e1004226. doi:10.1371/journal.pcbi.1004226

Kuypers MMM, Marchant HK, Kartal B. 2018. The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**:263–276. doi:10.1038/nrmicro.2018.9

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. doi:10.1007/s10142-015-0433-4

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947–2948. doi:10.1093/bioinformatics/btm404

Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**:4162–4164. doi:10.1093/bioinformatics/btz188

Levy Karin E, Mirdita M, Söding J. 2020. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* **8**:48. doi:10.1186/s40168-020-00808-x

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**:1674–1676. doi:10.1093/bioinformatics/btv033

Li H. 2023. lh3/seqtk.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi:10.48550/arXiv.1303.3997

Li Y, Cha Q-Q, Dang Y-R, Chen X-L, Wang M, McMinn A, Espina G, Zhang Y-Z, Blamey JM, Qin Q-L. 2019. Reconstruction of the Functional Ecosystem in the High Light, Low Temperature Union Glacier Region, Antarctica. *Front Microbiol* **10**.

Liang Y, Jiang P, Yao B, Jiao Y, Li J. 2021. Lacisediminihabitans changchengi sp. nov., an actinobacterium isolated from Antarctic swamplands mud. *Arch Microbiol* **203**:5519–5524. doi:10.1007/s00203-021-02531-z

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923–930. doi:10.1093/bioinformatics/btt656

Lin FH. 2019. Third release of ANCOM. doi:10.5281/zenodo.3577802

Liu C, Wang Xiuliang, Wang Xingna, Sun C. 2016. Acclimation of Antarctic Chlamydomonas to the sea-ice environment: a transcriptomic analysis. *Extremophiles* **20**:437–450. doi:10.1007/s00792-016-0834-x

Liu Q, Li W, Liu D, Li L, Li J, Lv N, Liu F, Zhu B, Zhou Y, Xin Y, Dong X. 2021. Light stimulates anoxic and oligotrophic growth of glacial Flavobacterium strains that produce zeaxanthin. *ISME J* **15**:1844–1857. doi:10.1038/s41396-020-00891-w

Liu S, Moon CD, Zheng N, Huws S, Zhao S, Wang J. 2022. Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* **10**:76. doi:10.1186/s40168-022-01272-5

Liu Y, Ji M, Yu T, Zaugg J, Anesio AM, Zhang Z, Hu S, Hugenholtz P, Liu K, Liu P, Chen Y, Luo Y, Yao T. 2022. A genome and gene catalog of glacier microbiomes. *Nat Biotechnol* **40**:1341–1348. doi:10.1038/s41587-022-01367-2

Liu Y, Makarova KS, Huang W-C, Wolf YI, Nikolskaya AN, Zhang X, Cai M, Zhang C-J, Xu W, Luo Z, Cheng L, Koonin EV, Li M. 2021. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**:553–557. doi:10.1038/s41586-021-03494-3

Liu Y, Shen L, Zeng Y, Xing T, Xu B, Wang N. 2020. Genomic Insights of Cryobacterium Isolated From Ice Core Reveal Genome Dynamics for Adaptation in Glacier. *Front Microbiol* **11**. doi:10.3389/fmicb.2020.01530

Lorenz C, Lünse CE, Mörl M. 2017. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **7**. doi:10.3390/biom7020035

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

Lüdecke (@strengejacke) D, Makowski (@Dom_Makowski) D, Ben-Shachar (@mattansb) MS, Patil (@patilindrajeets) I, Waggoner P, Wiernik (@bmwiernik) BM, Arel-Bundock V, Thériault (@rempsyc) R, Jullum M, gjo11, Bacher E. 2023. performance: Assessment of Regression Models Performance.

Lynch M. 2006. Streamlining and Simplification of Microbial Genome Architecture. *Annu Rev Microbiol* **60**:327–349. doi:10.1146/annurev.micro.60.080805.142300

Maillot NJ, Honoré FA, Byrne D, Méjean V, Genest O. 2019. Cold adaptation in the environmental bacterium Shewanella oneidensis is controlled by a J-domain co-chaperone protein network. *Commun Biol* **2**:1–10. doi:10.1038/s42003-019-0567-3

Margesin R, Collins T. 2019. Microbial ecology of the cryosphere (glacial and permafrost habitats): current knowledge. *Appl Microbiol Biotechnol* **103**:2537–2549. doi:10.1007/s00253-019-09631-3

Martinez Arbizu, P. 2020. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4 https://github.com/pmartinezarbizu/pairwiseAdonis.

Martinez-Gutierrez CA, Aylward FO. 2022. Genome size distributions in bacteria and archaea are strongly linked to evolutionary history at broad phylogenetic scales. *PLOS Genet* **18**:e1010220. doi:10.1371/journal.pgen.1010220

Martiny AC, Treseder K, Pusch G. 2013. Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**:830–838. doi:10.1038/ismej.2012.160

Marx JG, Carpenter SD, Deming JW. 2009. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium Colwellia psychrerythraea strain 34H under extreme conditionsThis article is one of a selection of papers in the Special Issue on Polar and Alpine Microbiology. *Can J Microbiol* **55**:63–72. doi:10.1139/W08-130

McDaniel EA, Anantharaman K, McMahon KD. 2019. metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. doi:10.1101/2019.12.20.884627

Merino N, Aronson HS, Bojanova DP, Feyhl-Buska J, Wong ML, Zhang S, Giovannelli D. 2019. Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context. *Front Microbiol* **10**:780. doi:10.3389/fmicb.2019.00780

Methé BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X, Moult J, Madupu R, Nelson WC, Dodson RJ, Brinkac LM, Daugherty SC, Durkin AS, DeBoy RT, Kolonay JF, Sullivan SA, Zhou L, Davidsen TM, Wu M, Huston AL, Lewis M, Weaver B, Weidman JF, Khouri H, Utterback TR, Feldblyum TV, Fraser CM. 2005. The psychrophilic lifestyle as revealed by the genome sequence of Colwellia

psychrerythraea 34H through genomic and proteomic analyses. *Proc Natl Acad Sci* **102**:10913–10918. doi:10.1073/pnas.0504766102

Meziti A, Rodriguez-R LM, Hatt JK, Peña-Gonzalez A, Levy K, Konstantinidis KT. 2021. The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Appl Environ Microbiol* **87**:e02593-20. doi:10.1128/AEM.02593-20

Michoud G, Kohler TJ, Peter H, Brandani J, Busi SB, Battin TJ. 2023. Unexpected functional diversity of stream biofilms within and across proglacial floodplains despite close spatial proximity. *Limnol Oceanogr* **68**:2183–2194. doi:10.1002/lno.12415

Millar JL, Bagshaw EA, Edwards A, Poniecka EA, Jungblut AD. 2021. Polar Cryoconite Associated Microbiota Is Dominated by Hemispheric Specialist Genera. *Front Microbiol* **12**.

Milner AM, Khamis K, Battin TJ, Brittain JE, Barrand NE, Füreder L, Cauvy-Fraunié S, Gíslason GM, Jacobsen D, Hannah DM, Hodson AJ, Hood E, Lencioni V, Ólafsson JS, Robinson CT, Tranter M, Brown LE. 2017. Glacier shrinkage driving global changes in downstream systems. *Proc Natl Acad Sci* **114**:9770–9778. doi:10.1073/pnas.1619807114

Milner AM, Petts GE. 1994. Glacial rivers: physical habitat and ecology. *Freshw Biol* **32**:295–307. doi:10.1111/j.1365-2427.1994.tb01127.x

Mod HK, Buri A, Yashiro E, Guex N, Malard L, Pinto-Figueroa E, Pagni M, Niculita-Hirzel H, van der Meer JR, Guisan A. 2021. Predicting spatial patterns of soil bacteria under current and future environmental conditions. *ISME J* **15**:2547–2560. doi:10.1038/s41396-021-00947-5

Mujakić I, Piwosz K, Koblížek M. 2022. Phylum Gemmatimonadota and Its Role in the Environment. *Microorganisms* **10**:151. doi:10.3390/microorganisms10010151

Mykytczuk NCS, Foote SJ, Omelon CR, Southam G, Greer CW, Whyte LG. 2013. Bacterial growth at −15 °C; molecular insights from the permafrost bacterium Planococcus halocryophilus Or1. *ISME J* **7**:1211–1226. doi:10.1038/ismej.2013.8

Narayanasamy S, Jarosz Y, Muller EEL, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Pinel N, May P, Wilmes P. 2016. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol* **17**:260. doi:10.1186/s13059-016-1116-8

Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**:578–585. doi:10.1038/s41587-020-00774-7

Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirton E, Faria JP, Edirisinghe JN, Henry CS, Jungbluth SP, Chivian D, Dehal P, Wood-Charlson EM, Arkin AP, Tringe SG, Visel A, Woyke T, Mouncey NJ, Ivanova NN, Kyrpides NC, Eloe-Fadrosh EA. 2020. A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 1–11. doi:10.1038/s41587-020-0718-6

Nelson W, Stegen J. 2015. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Front Microbiol* **6**.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**:268–274. doi:10.1093/molbev/msu300

Niedrist GH, Füreder L. 2018. When the going gets tough, the tough get going: The enigma of survival strategies in harsh glacial stream environments. *Freshw Biol* **63**:1260–1272. doi:10.1111/fwb.13131

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H, others. 2013. Package 'vegan.' *Community Ecol Package Version* **2**:1–295.

Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**:2864–2868. doi:10.1038/ismej.2017.126

O'Malley MA. 2007. The nineteenth century roots of "everything is everywhere." *Nat Rev Microbiol* **5**:647–651. doi:10.1038/nrmicro1711

Paillex A, Siebers AR, Ebi C, Mesman J, Robinson CT. 2020. High stream intermittency in an alpine fluvial network: Val Roseg, Switzerland. *Limnol Oceanogr* **65**:557–568. doi:10.1002/lno.11324

Palmer MA, Swan CM, Nelson K, Silver P, Alvestad R. 2000. Streambed landscapes: evidence that stream invertebrates respond to the type and spatial arrangement of patches. *Landsc Ecol* **15**:563–576. doi:10.1023/A:1008194130695

Paoli L, Ruscheweyh H-J, Forneris CC, Hubrich F, Kautsar S, Bhushan A, Lotti A, Clayssen Q, Salazar G, Milanese A, Carlström CI, Papadopoulou C, Gehrig D, Karasikov M, Mustafa H, Larralde M, Carroll LM, Sánchez P, Zayed AA, Cronin DR, Acinas SG, Bork P, Bowler C, Delmont TO, Gasol JM, Gossert AD, Kahles A, Sullivan MB, Wincker P, Zeller G, Robinson SL, Piel J, Sunagawa S. 2022. Biosynthetic potential of the global ocean microbiome. *Nature* **607**:111–118. doi:10.1038/s41586-022-04862-3

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528. doi:10.1093/bioinformatics/bty633

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. doi:10.1101/gr.186072.114

Payne AT, Davidson AJ, Kan J, Peipoch M, Bier R, Williamson K. 2020. Widespread cryptic viral infections in lotic biofilms. *Biofilm* **2**:100016. doi:10.1016/j.bioflm.2019.100016

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, others. 2011. Scikit-learn: Machine learning in Python. *J Mach Learn Res* **12**:2825–2830.

Pörtner H-O, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M, Poloczanska E, Mintenbeck K, Nicolai M, Okem A, Petzold J. 2019. IPCC special report on the ocean and cryosphere in a changing climate. *IPCC Intergov Panel Clim Change IPCC*.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**:e9490. doi:10.1371/journal.pone.0009490

Props R, Monsieurs P, Vandamme P, Leys N, Denef VJ, Boon N. 2019. Gene Expansion and Positive Selection as Bacterial Adaptations to Oligotrophic Conditions. *mSphere* **4**:10.1128/mspheredirect.00011-19. doi:10.1128/mspheredirect.00011-19

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**:D61–D65. doi:10.1093/nar/gkl842

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**:D590–D596. doi:10.1093/nar/gks1219

Queirós P, Delogu F, Hickl O, May P, Wilmes P. 2021. Mantis: flexible and consensus-driven genome annotation. *GigaScience* **10**:giab042. doi:10.1093/gigascience/giab042

Quick AM, Reeder WJ, Farrell TB, Tonina D, Feris KP, Benner SG. 2019. Nitrous oxide from streams and rivers: A review of primary biogeochemical pathways and environmental variables. *Earth-Sci Rev* **191**:224–262. doi:10.1016/j.earscirev.2019.02.021

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Ravindra K, Rattan P, Mor S, Aggarwal AN. 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. *Environ Int* **132**:104987. doi:10.1016/j.envint.2019.104987

Ren Z, Gao H, Elser JJ. 2017a. Longitudinal variation of microbial communities in benthic biofilms and association with hydrological and physicochemical conditions in glacier-fed streams. *Freshw Sci* **36**:479–490. doi:10.1086/693133

Ren Z, Gao H, Elser JJ, Zhao Q. 2017b. Microbial functional genes elucidate environmental drivers of biofilm metabolism in glacier-fed streams. *Sci Rep* **7**:12668. doi:10.1038/s41598-017-13086-9

Ren Z, Martyniuk N, Oleksy IA, Swain A, Hotaling S. 2019. Ecological Stoichiometry of the Mountain Cryosphere. *Front Ecol Evol* **7**.

Retchless AC, Lawrence JG. 2010. Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. *Proc Natl Acad Sci* **107**:11453–11458. doi:10.1073/pnas.1001291107

Retchless AC, Lawrence JG. 2007. Temporal Fragmentation of Speciation in Bacteria. *Science* **317**:1093–1096. doi:10.1126/science.1144876

Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* **3**:217–223. doi:10.1111/j.2041-210X.2011.00169.x

Risse-Buhl U, Anlanger C, Chatzinotas A, Noss C, Lorke A, Weitere M. 2020. Near streambed flow shapes microbial guilds within and across trophic levels in fluvial biofilms. *Limnol Oceanogr* **65**:2261–2277. doi:10.1002/lno.11451

Robison AL, Deluigi N, Rolland C, Manetti N, Battin T. 2023. Glacier loss and vegetation expansion alter organic and inorganic carbon dynamics in high-mountain streams. *Biogeosciences* **20**:2301–2316. doi:10.5194/bg-20-2301-2023

Rodríguez-Gijón A, Nuy JK, Mehrshad M, Buck M, Schulz F, Woyke T, Garcia SL. 2022. A Genomic Perspective Across Earth's Microbiomes Reveals That Genome Size in Archaea and Bacteria Is Linked to Ecosystem Type and Trophic Strategy. *Front Microbiol* **12**.

Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, Venter SN, Orellana LH, Amann R, Rossello-Mora R, Konstantinidis KT. 2023. An ANI gap within bacterial species that advances the definitions of intra-species units. *mBio* **0**:e02696-23. doi:10.1128/mbio.02696-23

Roncoroni M, Brandani J, Battin TI, Lane SN. 2019. Ecosystem engineers: Biofilms and the ontogeny of glacier floodplain ecosystems. *WIREs Water* **6**:e1390. doi:10.1002/wat2.1390

Rounce DR, Hock R, Maussion F, Hugonnet R, Kochtitzky W, Huss M, Berthier E, Brinkerhoff D, Compagno L, Copland L, Farinotti D, Menounos B, McNabb RW. 2023. Global glacier change in the 21st century: Every increase in temperature matters. *Science* **379**:78–83. doi:10.1126/science.abo1324

Royo-Llonch M, Sánchez P, Ruiz-González C, Salazar G, Pedrós-Alió C, Sebastián M, Labadie K, Paoli L, M. Ibarbalz F, Zinger L, Churcheward B, Chaffron S, Eveillard D, Karsenti E, Sunagawa S, Wincker P, Karp-Boss L, Bowler C, Acinas SG. 2021. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol* **6**:1561–1574. doi:10.1038/s41564-021-00979-9

Rumpf SB, Gravey M, Brönnimann O, Luoto M, Cianfrani C, Mariethoz G, Guisan A. 2022. From white to green: Snow cover loss and increased vegetation productivity in the European Alps. *Science* **376**:1119–1122. doi:10.1126/science.abn6697

Saary P, Mitchell AL, Finn RD. 2020. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. *Genome Biol* **21**:244. doi:10.1186/s13059-020-02155-4

Sabath N, Ferrada E, Barve A, Wagner A. 2013. Growth Temperature and Genome Size in Bacteria Are Negatively Correlated, Suggesting Genomic Streamlining During Thermal Adaptation. *Genome Biol Evol* **5**:966–977. doi:10.1093/gbe/evt050

Saheb Kashaf S, Proctor DM, Deming C, Saary P, Hölzer M, Taylor ME, Kong HH, Segre JA, Almeida A, Finn RD. 2022. Integrating cultivation and metagenomics for a

multi-kingdom view of skin microbiome diversity and functions. *Nat Microbiol* **7**:169–179. doi:10.1038/s41564-021-01011-w

Sajjad W, Ali B, Bahadur A, Ghimire PS, Kang S. 2021. Bacterial Diversity and Communities Structural Dynamics in Soil and Meltwater Runoff at the Frontier of Baishui Glacier No.1, China. *Microb Ecol* **81**:370–384. doi:10.1007/s00248-020-01600-y

Salcher MM, Schaefle D, Kaspar M, Neuenschwander SM, Ghai R. 2019. Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in Methylophilaceae. *ISME J* **13**:2764–2777. doi:10.1038/s41396-019-0471-3

Sánchez Barranco V, Van der Meer MTJ, Kagami M, Van den Wyngaert S, Van de Waal DB, Van Donk E, Gsell AS. 2020. Trophic position, elemental ratios and nitrogen transfer in a planktonic host–parasite–consumer food chain including a fungal parasite. *Oecologia* **194**:541–554. doi:10.1007/s00442-020-04721-w

Satapathy SS, Dutta M, Ray SK. 2010. Higher tRNA diversity in thermophilic bacteria: A possible adaptation to growth at high temperature. *Microbiol Res* **165**:609–616. doi:10.1016/j.micres.2009.12.003

Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, Hoffmann C, Weinert K, Tenaillon O, Matic I, Denamur E. 2009. Role of Intraspecies Recombination in the Spread of Pathogenicity Islands within the Escherichia coli Species. *PLOS Pathog* **5**:e1000257. doi:10.1371/journal.ppat.1000257

Schütz C, Wallinger M, Burger R, Füreder L. 2001. Effects of snow cover on the benthic fauna in a glacier-fed stream. *Freshw Biol* **46**:1691–1704. doi:10.1046/j.1365-2427.2001.00852.x

Schuur E a. G, McGuire AD, Schädel C, Grosse G, Harden JW, Hayes DJ, Hugelius G, Koven CD, Kuhry P, Lawrence DM, Natali SM, Olefeldt D, Romanovsky VE, Schaefer K, Turetsky MR, Treat CC, Vonk JE. 2015. Climate change and the permafrost carbon feedback. *Nature* **520**:171–179. doi:10.1038/nature14338

Scotti A, Jacobsen D, Tappeiner U, Bottarin R. 2019. Spatial and temporal variation of benthic macroinvertebrate assemblages during the glacial melt season in an Italian glacier-fed stream. *Hydrobiologia* **827**:123–139. doi:10.1007/s10750-018-3731-8

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068–2069. doi:10.1093/bioinformatics/btu153

Segev E, Wyche TP, Kim KH, Petersen J, Ellebrandt C, Vlamakis H, Barteneva N, Paulson JN, Chai L, Clardy J, Kolter R. 2016. Dynamic metabolic exchange governs a marine algal-bacterial interaction. *eLife* **5**:e17473. doi:10.7554/eLife.17473

Selle B, Knorr K-H, Lischeid G. 2019. Mobilisation and transport of dissolved organic carbon and iron in peat catchments—Insights from the Lehstenbach stream in Germany using generalised additive models. *Hydrol Process* **33**:3213–3225. doi:10.1002/hyp.13552

Seymour JR, Amin SA, Raina J-B, Stocker R. 2017. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* **2**:1–12. doi:10.1038/nmicrobiol.2017.65

Shakya M, Lo C-C, Chain PSG. 2019. Advances and Challenges in Metatranscriptomic Analysis. *Front Genet* **10**.

Shannon CE, Weaver W. 1949. The mathematical theory of information. *Urbana Univ Ill Press* **97**.

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**:e0163962. doi:10.1371/journal.pone.0163962

Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, Fernandez-Guerra A, Liebeke M, Schweder T, Polz MF, Hehemann J-H. 2020. Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nat Microbiol* **5**:1026–1039. doi:10.1038/s41564-020-0720-2

Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**:836–843. doi:10.1038/s41564-018-0171-1

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. doi:10.1093/bioinformatics/btv351

Singer GA, Fasching C, Wilhelm L, Niggemann J, Steier P, Dittmar T, Battin TJ. 2012. Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate. *Nat Geosci* **5**:710–714. doi:10.1038/ngeo1581

Sinsabaugh RL, Hill BH, Follstad Shah JJ. 2009. Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. *Nature* **462**:795–798. doi:10.1038/nature08632

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**:732–740. doi:10.1111/j.1462-2920.2005.00956.x

Smith HJ, Schmit A, Foster R, Littman S, Kuypers MM, Foreman CM. 2016. Biofilms on glacial surfaces: hotspots for biological activity. *Npj Biofilms Microbiomes* **2**:1–4. doi:10.1038/npjbiofilms.2016.8

Sommaruga R. 2001. The role of solar UV radiation in the ecology of alpine lakes. *J Photochem Photobiol B*, Impacts of Ultraviolet Radiation on Aquatic and Terrestrial Ecosystems **62**:35–42. doi:10.1016/S1011-1344(01)00154-3

Southwood TRE. 1977. Habitat, the Templet for Ecological Strategies? *J Anim Ecol* **46**:337–365. doi:10.2307/3817

St. Pierre KA, St. Louis VL, Schiff SL, Lehnherr I, Dainard PG, Gardner AS, Aukes PJK, Sharp MJ. 2019. Proglacial freshwaters are significant and previously unrecognized sinks of atmospheric CO2. *Proc Natl Acad Sci* **116**:17690–17695. doi:10.1073/pnas.1904241116

Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**:1026–1028. doi:10.1038/nbt.3988

Stibal M, Šabacká M, Žárský J. 2012. Biological processes on glacier and ice sheet surfaces. *Nat Geosci* **5**:771–774. doi:10.1038/ngeo1611

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19**:21–31. doi:10.1111/j.1365-294X.2009.04480.x

Stoecker DK, Lavrentyev PJ. 2018. Mixotrophic Plankton in the Polar Seas: A Pan-Arctic Review. *Front Mar Sci* **5**.

Su B, Xiao C, Chen D, Qin D, Ding Y. 2019. Cryosphere Services and Human Well-Being. *Sustainability* **11**:4365. doi:10.3390/su11164365

Tang J, Du L-M, Liang Y-M, Daroch M. 2019. Complete Genome Sequence and Comparative Analysis of Synechococcus sp. CS-601 (SynAce01), a Cold-

Adapted Cyanobacterium from an Oligotrophic Antarctic Habitat. *Int J Mol Sci* **20**:152. doi:10.3390/ijms20010152

Tange O. 2018. GNU Parallel 2018. Lulu.com.

Tay JK, Narasimhan B, Hastie T. 2023. Elastic Net Regularization Paths for All Generalized Linear Models. *J Stat Softw* **106**:1–31. doi:10.18637/jss.v106.i01

Thaler DS. 2021. Is Global Microbial Biodiversity Increasing, Decreasing, or Staying the Same? *Front Ecol Evol* **9**.

Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**:457–463. doi:10.1038/nature24621

Tian R, Ning D, He Z, Zhang P, Spencer SJ, Gao S, Shi W, Wu L, Zhang Y, Yang Y, Adams BG, Rocha AM, Detienne BL, Lowe KA, Joyner DC, Klingeman DM, Arkin AP, Fields MW, Hazen TC, Stahl DA, Alm EJ, Zhou J. 2020. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* **8**:51. doi:10.1186/s40168-020-00825-w

Ting L, Williams TJ, Cowley MJ, Lauro FM, Guilhaus M, Raftery MJ, Cavicchioli R. 2010. Cold adaptation in the marine bacterium, Sphingopyxis alaskensis, assessed using quantitative proteomics. *Environ Microbiol* **12**:2658–2676. doi:10.1111/j.1462-2920.2010.02235.x

Tockner K, Malard F, Uehlinger U, Ward JV. 2002. Nutrients and organic matter in a glacial river—floodplain system (Val Roseg, Switzerland). *Limnol Oceanogr* **47**:266–277. doi:10.4319/lo.2002.47.1.0266

Tolotti M, Cerasino L, Donati C, Pindo M, Rogora M, Seppi R, Albanese D. 2020. Alpine headwaters emerging from glaciers and rock glaciers host different bacterial communities: Ecological implications for the future. *Sci Total Environ* **717**:137101. doi:10.1016/j.scitotenv.2020.137101

Tranter M, Brown G, Raiswell R, Sharp M, Gurnell A. 1993. A conceptual model of solute acquisition by Alpine glacial meltwaters. *J Glaciol* **39**:573–581. doi:10.3189/S0022143000016464

Tranter M, Mills R, Raiswell R. 1989. Chemical weathering reactions in Alpine glacial meltwaters. Presented at the International symposium on water-rock interaction. pp. 687–690.

Tribelli PM, López NI. 2018. Reporting Key Features in Cold-Adapted Bacteria. *Life* **8**:8. doi:10.3390/life8010008

Tribelli PM, Venero ECS, Ricardi MM, Gómez-Lozano M, Iustman LJR, Molin S, López NI. 2015. Novel Essential Role of Ethanol Oxidation Genes at Low Temperature Revealed by Transcriptome Analysis in the Antarctic Bacterium Pseudomonas extremaustralis. *PLOS ONE* **10**:e0145353. doi:10.1371/journal.pone.0145353

Tripathi BM, Kim M, Kim Y, Byun E, Yang J-W, Ahn J, Lee YK. 2018. Variations in bacterial and archaeal communities along depth profiles of Alaskan soil cores. *Sci Rep* **8**:504. doi:10.1038/s41598-017-18777-x

Uehlinger U, Robinson CT, Hieber M, Zah R. 2010. The physico-chemical habitat template for periphyton in alpine glacial streams under a changing climate In: Stevenson RJ, Sabater S, editors. Global Change and River Ecosystems—Implications for Structure, Function and EcosystemServices, Developments in Hydrobiology 215. Dordrecht: Springer Netherlands. pp. 107–121. doi:10.1007/978-94-007-0608-8_8

Varin T, Lovejoy C, Jungblut AD, Vincent WF, Corbeil J. 2012. Metagenomic Analysis of Stress Genes in Microbial Mat Communities from Antarctica and the High Arctic. *Appl Environ Microbiol* **78**:549–559. doi:10.1128/AEM.06354-11

Varin T, Lovejoy C, Jungblut AD, Vincent WF, Corbeil J. 2010. Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnol Oceanogr* **55**:1901–1911. doi:https://doi.org/10.4319/lo.2010.55.5.1901

Varrette S, Bouvry P, Cartiaux H, Georgatos F. 2014. Management of an academic HPC cluster: The UL experience2014 International Conference on High Performance Computing Simulation (HPCS). Presented at the 2014 International Conference on High Performance Computing Simulation (HPCS). pp. 959–967. doi:10.1109/HPCSim.2014.6903792

Vigneron A, Cruaud P, Langlois V, Lovejoy C, Culley AI, Vincent WF. 2020. Ultra-small and abundant: Candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. *Limnol Oceanogr Lett* **5**:212–220. doi:10.1002/lol2.10132

Vincent WF, Downes MT, Castenholz RW, Howard-Williams C. 1993. Community structure and pigment organisation of cyanobacteria-dominated microbial mats in Antarctica. *Eur J Phycol* **28**:213–221. doi:10.1080/09670269300650321

Vollmers J, Wiegand S, Lenk F, Kaster A-K. 2022. How clear is our current view on microbial dark matter? (Re-)assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Res* **50**:e76. doi:10.1093/nar/gkac294

Wagner K, Bengtsson MM, Findlay RH, Battin TJ, Ulseth AJ. 2017. High light intensity mediates a shift from allochthonous to autochthonous carbon use in phototrophic stream biofilms. *J Geophys Res Biogeosciences* **122**:1806–1820. doi:10.1002/2016JG003727

Waibel A, Peter H, Sommaruga R. 2019. Importance of mixotrophic flagellates during the ice-free season in lakes located along an elevational gradient. *Aquat Sci* **81**:45. doi:10.1007/s00027-019-0643-2

Wang Z, Huang P, You R, Sun F, Zhu S. 2023. MetaBinner: a high-performance and stand-alone ensemble binning method to recover individual genomes from complex microbial communities. *Genome Biol* **24**:1. doi:10.1186/s13059-022-02832-6

Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc Natl Acad Sci* **118**. doi:10.1073/pnas.2016810118

West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. 2018. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res* **28**:569–580. doi:10.1101/gr.228429.117

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. *J Open Source Softw* **4**:1686. doi:10.21105/joss.01686

Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA, Quince C, Battin TJ. 2014. Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environ Microbiol* **16**:2514–2524. doi:https://doi.org/10.1111/1462-2920.12392

Wilhelm L, Besemer K, Fragner L, Peter H, Weckwerth W, Battin TJ. 2015. Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms. *ISME J* **9**:2454–2464. doi:10.1038/ismej.2015.56

Wilhelm L, Singer GA, Fasching C, Battin TJ, Besemer K. 2013. Microbial biodiversity in glacier-fed streams. *ISME J* **7**:1651–1660. doi:10.1038/ismej.2013.44

wilkelab/ggridges: Ridgeline plots in ggplot2. https://github.com/wilkelab/ggridges

Wilkes MA, Carrivick JL, Castella E, Ilg C, Cauvy-Fraunié S, Fell SC, Füreder L, Huss M, James W, Lencioni V, Robinson C, Brown LE. 2023. Glacier retreat reorganizes river habitats leaving refugia for Alpine invertebrate biodiversity poorly protected. *Nat Ecol Evol* **7**:841–851. doi:10.1038/s41559-023-02061-5

Winkel M, Trivedi CB, Mourot R, Bradley JA, Vieth-Hillebrand A, Benning LG. 2022. Seasonality of Glacial Snow and Ice Microbial Communities. *Front Microbiol* **13**.

Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API.

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes J-A, Guisan A, Heikkinen RK, Høye TT, Kühn I, Luoto M, Maiorano L, Nilsson M-C, Normand S, Öckinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning J-C. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol Rev* **88**:15–30. doi:10.1111/j.1469-185X.2012.00235.x

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**:257. doi:10.1186/s13059-019-1891-0

Wood S. 2023. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.

Wu H, Zhang Z, Hu S, Yu J. 2012. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol Direct* **7**:2. doi:10.1186/1745-6150-7-2

Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**:605–607. doi:10.1093/bioinformatics/btv638

Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, Tang W, Feng T, Chen M, Zhan L, Wu T, Hu E, Jiang Y, Bo X, Yu G. 2021. ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Mol Biol Evol* **38**:4039–4042. doi:10.1093/molbev/msab166

Yu G, Lam TT-Y, Zhu H, Guan Y. 2018. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**:3041–3043. doi:10.1093/molbev/msy194

Zablocki O, Jang HB, Bolduc B, Sullivan MB. 2019. vConTACT 2: A Tool to Automate Genome-Based Prokaryotic Viral Taxonomy. Presented at the Plant and Animal Genome XXVII Conference
(January 12- 16, 2019). PAG.

Zhang T, Li D, East AE, Walling DE, Lane S, Overeem I, Beylich AA, Koppes M, Lu X. 2022. Warming-driven erosion and sediment transport in cold regions. *Nat Rev Earth Environ* **3**:832–851. doi:10.1038/s43017-022-00362-0

Zhang Z, Liu Y, Zhao W, Ji M. 2023. Radiation impacts gene redundancy and biofilm regulation of cryoconite microbiomes in Northern Hemisphere glaciers. *Microbiome* **11**:228. doi:10.1186/s40168-023-01621-y

Zhou J, Lyu Y, Richlen ML, Anderson DM, Cai Z. 2016. Quorum Sensing Is a Language of Chemical Signals and Plays an Ecological Role in Algal-Bacterial Interactions. *Crit Rev Plant Sci* **35**:81–105. doi:10.1080/07352689.2016.1172461

Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, Karaoz U, Anantharaman K. 2022. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**:33. doi:10.1186/s40168-021-01213-8

Zhu A, Sunagawa S, Mende DR, Bork P. 2015. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol* **16**:82. doi:10.1186/s13059-015-0646-9

Massimo Bourquin (bioinformatician)

Chemin des Paleyres 14, 1006 Lausanne (Switzerland), tel. +41 79 730 60 56
massimo.bourquin@gmail.com, twitter: @MassB23, researchgate: Massimo Bourquin

Education

2020 - PhD student in Civil and Environmental engineering at the RIVER laboratory, EPFL

- École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
- Vanishing Glaciers Project (glacierstreams.ch)
- Microbial ecology and evolution of cryospheric bacterial communities
- Metagenomics of glacier-fed streams and predicting the response of bacterial communities to climate change.
- Teaching: Multivariate statistics in R, Fundamentals in ecology

2017 – Master of science in computational ecology and evolution

- Université de Lausanne (UNIL), Switzerland
- Comparative genomics, population genomics and evolutionary biology

2014 – Bachelor of science in Biology

- Université de Neuchâtel (UNINE), Switzerland
- Projects in ecology and evolution

Work experience

2019 – 2020 Swiss civilian service assignment (bioinformatician)

- École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

2017 – 2018 Assistant student work

- Université de Lausanne (UNIL), Switzerland

2017 Research internship and rearing work

- Université de Neuchâtel (UNINE), Switzerland

Publications

2023

- Brandani J, Peter H, Fodelianakis S, Kohler TJ, **Bourquin M**, et al. *Homogeneous Environmental Selection Structures the Bacterial Communities of Benthic Biofilms in Proglacial Floodplain Streams.* Applied and Environmental Microbiology 89:3:e02010-22.

- Michoud, G. et al. *The dark side of the moon: first insights into the microbiome structure and function of one of the last glacier-fed streams in Africa.* R. Soc. Open Sci. 10, 230329 (2023).

Research skills

- Data science analyses in R including data processing, modelling and visualisations
- Processing of sequencing data (whole genome, amplicon and metagenomes)
- UNIX, python and R for bioinformatics
- Machine learning (Python and R)
- Phylogenetics and microbial ecology analyses
- Functional annotation, genome content analyses and comparative genomics

Github:
https://github.com/Mass23

Public outreach

- Press releases
- Public presentations
- Interviews (radio, written) on the Vanishing glaciers project (glacierstreams.ch/outreach)

References

- Tom Ian Battin, Thesis director. EPFL, RIVER laboratory

  tom.battin@epfl.ch

  +41216930837

Languages

- French: mother tongue
- English: excellent knowledge, study and work language
- Italian/German: basic knowledge

Arts and sport

- Music, scenic and visual arts
- Hiking, climbing, running

- Busi SB, De Nies L, Busi SB, Pramateftaki P, **Bourquin M**, et al. *Glacier-Fed Stream Biofilms Harbor Diverse Resistomes and Biosynthetic Gene Clusters.* **Microbiology spectrum** 11:1:e04069-22.

2022
- Brandani J, Peter H, Busi SB, Kohler TJ, Fodelianakis S, Ezzat L, Michoud G, **Bourquin M**, et al. *Spatial patterns of benthic biofilm diversity among streams draining proglacial floodplains.* **Frontiers in Microbiology** 13:948165.

- Ezzat L, Fodelianakis S, Kohler TJ, **Bourquin M**, et al. *Benthic Biofilms in Glacier-Fed Streams from Scandinavia to the Himalayas Host Distinct Bacterial Communities Compared with the Streamwater.* **Applied and Environmental Microbiology** 88:12:e00421-22.

- **Bourquin M**, Busi SB, Fodelianakis S, et al. *The microbiome of cryospheric ecosystems.* **Nature Communications** 13, 3087.

- Busi SB, **Bourquin M**, Fodelianakis S, et al. *Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams.* **Nature Communications** 13, 2168.

- Kohler TJ, Fodelianakis S, Michoud G, Ezzat L, **Bourquin M**, et al. *Glacier shrinkage will accelerate downstream decomposition of organic matter and alters microbiome structure and function.* **Global Change Biology**, 28, 3846–3859.

- Fodelianakis S, Washburne AD, **Bourquin M**, et al. *Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome.* **ISME J** 16, 666–675.

2019
- Bekliz M, Brandani J, **Bourquin M**, Battin TJ, Peter H. *Benchmarking protocols for the metagenomic analysis of stream biofilm viromes.* **PeerJ** 7:e8187.

Conferences

2023

- EDCE Research Day 2023, EPFL – Poster presentation: Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change, *Best poster jury award*
- ENAC Research Day 2023, EPFL – Poster presentation: Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change
- EGU23, European geosciences union general assembly – Poster presentation: Predicting the global response of the glacier-fed streams and their bacterial microbiome to climate change

2022

- ISME18, International symposium on microbial ecology – Poster presentation: Predicting the impact of climate change on the microbiome of glacier-fed streams
- INTECOL 2022, International association for ecology conference – Poster presentation: Modelling the impact of climate change on glacier-fed streams biofilms