



HAL
open science

Exploration des interactions virus-hôte et leur importance pour l'adaptation microbienne à travers du CRISPRs

Laura Sanguino Casado

► **To cite this version:**

Laura Sanguino Casado. Exploration des interactions virus-hôte et leur importance pour l'adaptation microbienne à travers du CRISPRs. Autre. Ecole Centrale de Lyon, 2015. Français. NNT : 2015ECDL0033 . tel-01546401

HAL Id: tel-01546401

<https://theses.hal.science/tel-01546401>

Submitted on 23 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Présentée devant :

ÉCOLE CENTRALE DE LYON

Pour obtenir le grade de :

DOCTEUR

De l'école doctorale :

Électronique, électrotechnique, automatique

UNIVERSITÉ DE LYON

Spécialité : Microbiologie Environnementale

Par

Laura Sanguino Casado

**Exploration des interactions virus-hôte et leur importance pour l'adaptation
microbienne à travers du CRISPRs.**

Jury :

Timothy M. Vogel

Professeur, Université Claude Bernard Lyon 1, Ecole Centrale de Lyon, Ecully, France

Directeur de Thèse

Catherine Larose

CR2 CNRS, Ecole Centrale de Lyon, Ecully, France

Co-Directrice de Thèse

Francisco Rodríguez Valera

Catedrático, Universidad Miguel Hernández, Alicante, Spain

Rapporteur

Alexandre Anesio

Professor, University of Bristol, Bristol, United Kingdom

Rapporteur

Matthew Sullivan

Associate Professor, Ohio State University, Ohio, United States of America

Examineur

Michael Schloter

Professor, Helmholtz Zentrum München, Munich, Germany

Examineur

Pascal Simonet

DR CNRS, Ecole Centrale de Lyon, Ecully, France

Examineur

**EXPLORING ENVIRONMENTAL VIRUS-HOST
INTERACTIONS AND THEIR RELEVANCE TO
MICROBIAL ADAPTATION USING CRISPRs**

Laura Sanguino Casado

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

AUTUMN 2015

*“You cannot hope to build a better world without
improving the individuals.”*

Maria Skłodowska-Curie



This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no. [289949].

Acknowledgments

The people behind this thesis and its work are numerous, starting with my supervisors Tim Vogel and Catherine Larose. Thank you Tim for your insight, for making us rack our brains, for never answering our questions but with other questions. Thank you for your help in and out of the lab. Thank you Cath, for your advice and help at all times and for a fantastic trip to the Arctic. I am also grateful to my thesis committee members; Alexandre Anesio, Francisco Rodriguez-Valera, Matthew Sullivan and Michael Schloter for their guidance and for agreeing to form part of this thesis.

I would like to acknowledge all the members of the Laboratoire Ampère, which has hosted me during these three years. I am grateful to Pascal Simonet for his scientific questions during my practice talks, for his help and for all the literature he sent. Special thanks to Edith and Marie Christine for all their efficient work and their cheerfulness, because they made bureaucracy seem easy. I have to credit Laure and Sandrine for their patience, their management and their help in the lab. Alice provided her incredible gaiety every day and I thank her for it. Without Richard and his skills I would probably still be filtering ice, so thank you very much. Thanks to all the students with whom I shared lunch, discussions and beers: Joseph, Agathe, David, Alban, Natalia, Samuel, Jeremy, Thibault, and many other I'm sure I forget. Especially to those who took part in the Friday morning meetings, where we got out with more questions than we had entered with. Thank you for helping me shape my research and my ideas. To Graeme Nicol and Cecile Rangin I owe my masters and finding this lab, for that I will be always thankful.

My office colleagues; Jean-Sébastien Beaulne, Alexiane Godain, Lorrie Maccario, Christoph Keuschnig, and the intruder Sébastien Cecillon, I hold responsible for throwing overboard the little mental sanity I brought with me into the lab. I cannot thank you enough for all the laughter and aid throughout these three years. I never knew a PhD could be so much fun. I am especially indebted to Seb, who I must credit for pushing me ever further into the scientific reasoning and bioinformatics, at times even against my own will. His interest and critics have been essential in the development of the work that I here present. To Lorrie I will be always grateful, she

became a friend in a foreign country, my French teacher, my gym motivation and my escape valve.

Some people remain oblivious to the fact that we now live thousands of kilometres apart and it always seems you saw them yesterday. Alicia Acevedo, Rebeca Arias and Sandra Garcia have become masters of this aptitude. Thank you girls, it's comforting to know that Europe is colonized by friends. With my brother Diego I shared a fantastic childhood; thanks for bearing with the loving tortures of your sister, and I hope we can share a life-lasting friendship. To Patxi I have to thank his constant love and support to help me become the best version of myself. I look forward to this new phase we are starting together. Thank you for all there is to come.

My parents, Luis and Diana, are, and will always be, ultimately responsible for my achievements. Their aspirations for me and my brother and their endeavour have equipped us with all we need to be whoever we wish to be. For that, and for so much more: thank you.

Summary

Interactions between the members of a microbial community can be a means of adaptation in the environment. Among the many interactions that take place in an ecosystem and have been seen to play a major role on microbial diversity and population dynamics is that of prokaryotic viruses and their hosts. Viruses can also mediate the transfer of genetic material between prokaryotes (transduction), which could be a mechanism for rapid adaptation. In order to determine the potential impact of viruses and transduction, we need a better understanding of the dynamics of interactions between viruses and their hosts in the environment. Data on environmental viruses are scarce, and methods for tracking their interactions with prokaryotes are needed. Clustered regularly interspaced short palindromic repeats (CRISPRs), which contain viral sequences in bacterial genomes, might help document the history of virus-host interactions in the environment. Thus, this thesis aimed to explore virus-host interactions in a given environment through CRISPRs.

Viruses in the cryosphere have been seen to be abundant, highly active and with broad host ranges. These characteristics could make viral transduction a key driver of adaptation in these environments. Public metagenomes created from environments over a range of temperatures were examined through sequence and CRISPR analysis. In this fashion, certain virus-host interaction dynamics were found to have a correlation with temperature. A workflow was then developed to create a network linking viruses and their hosts using CRISPR sequences obtained from metagenomic data from Arctic glacial ice and soil. The creation of CRISPR-based infection networks provided a new perspective on virus-host interactions in glacial ice. Moreover, we searched for transduction events in metagenomic data by looking for viral sequences containing microbial DNA. Further analysis of the viral sequences in the CRISPRs indicated that *Ralstonia* phages might be agents of transduction in Arctic glacial ice.

Index

.....	i
Acknowledgments	i
Summary.....	iii
Index of Figures	1
Index of Tables.....	2
Abbreviations.....	3
1. Introduction.....	5
1.1 Viruses in Microbial Ecology.....	5
1.1.1 The Discovery of Prokaryotic Viruses	5
1.1.2 Viral Abundance and Diversity	6
1.1.3 Viral Regulation of Microbial Populations and Diversity.....	7
1.1.4 Viral Role on Microbial Evolution	8
1.2 CRISPRs as a tool.....	11
1.2.1 What are Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)?.....	11
1.2.2 The Discovery of CRISPRs	14
1.2.3 CRISPR/Cas9 Genome-Editing	14
1.2.4 Use of CRISPRs in Linking Viruses with their Hosts	16
1.3 Microbial Communities in Cold Environments.....	17
1.3.1 The Relevance of Cold Environments in Viral Studies.....	17
1.3.2 Habitats within Glacial Ice	19
1.3.3 Microbial Diversity and Abundances in Glacial Ice.....	25
2. Objectives	29
3. Material and Methods.....	30
3.1 Model Ecosystem (Kongsvegen Glacier) and Sampling.....	30

3.2 DNA Extractions	31
3.3 Sequencing	32
3.4 Bioinformatics.....	33
3.4.1 Recovery of Public Metagenomic Data	33
3.4.2 CRISPR Recovery from Metagenomic Data	34
3.4.3 Alpha Diversity Indexes	35
3.4.4 Sequence Clustering	36
3.4.5 MG-RAST.....	36
3.4.6 Metagenomic Assembly	37
3.4.7 Metavir	37
3.4.8 16S rRNA Gene Taxa Assignment	38
3.4.9 CRISPR Network Workflow	38
3.4.10 Analysis of Viral Sequences for Transduction	41
3.4.11 Analysis of <i>Ralstonia</i> Phage Presence in Microbial Ice Sequences	42
3.5 Statistics.....	42
3.5.1 Correlations	42
3.5.2 Principal Component Analysis	42
3.5.3 Rarefaction Curves	42
4. Results & Discussion	43
4.1 Distribution of CRISPRs in natural environments.....	43
4.1.1 Viral Diversity.....	47
4.1.2 Viral Host Range	51
4.2 Linking viruses and their hosts through CRISPRs	54
4.2.1 Overall Description of Arctic Ice and Soil Metagenomic Data	54
4.2.2 Creating the Network	63
4.3 Transduction Events	75

5. General Discussion	81
5.1 Viral Populations in Environments with a Range of Temperatures	81
5.2 CRISPR-based Infection Networks	84
5.3 Transduction in Glacial Ice	86
6. Conclusions	89
7. Perspectives	90
8. References	92
9. Appendix	115
9.1 Appendix 1: Diversity and CRISPR Abundance in Public Metagenomes	115
9.2 Appendix 2: Rarefaction Curves	118
9.3 Appendix 3: CRISPR Sequences in Arctic Ice and Soil Metagenomes	123
9.3 Appendix 4: Publications	134
.....	135

Index of Figures

- Figure 1.** One of the first electron microscopy images of bacteriophages.
- Figure 2.** Schematic summary of the three major horizontal transfer mechanisms.
- Figure 3.** Schematic diagram of the CRISPR/Cas mechanism.
- Figure 4.** Cell abundances and viral abundances in polar environments.
- Figure 5.** Location of the Kongsvegen glacier and a diagram of its thermal budget.
- Figure 6.** A glacial ecosystem with some of its niches.
- Figure 7.** Cell abundances in glacial environments.
- Figure 8.** Workflow constructed to identify viruses and their hosts through CRISPRs.
- Figure 9.** Curated workflow to identify viruses and their hosts through CRISPRs.
- Figure 10.** Viral diversity of the environments plotted against sample temperature.
- Figure 11.** Weighted average percentage of CRISPRs found in the metagenomes.
- Figure 12.** Number of spacer clusters plotted against temperature of the samples.
- Figure 13.** Duplicate spacers in an environment plotted against temperature.
- Figure 14.** Proportion of CRISPRs plotted against sample temperature.
- Figure 15.** Classification of metagenomic reads by the MG-RAST pipeline.
- Figure 16.** MG-RAST taxonomic classification for the bacterial domain.
- Figure 17.** PCA using MG-RAST based on organism abundance.
- Figure 18.** PCA using MG-RAST based on functional abundance.
- Figure 19.** Virus-host networks inferred from CRISPRs recovered from ice and soil.
- Figure 20.** Virus-host networks with shaded shapes based on workflow identification.
- Figure 21.** Networks with the DR and spacers identified by the first workflow.
- Figure 22.** Networks with the DR and spacers identified by the second workflow.
- Figure 23.** Networks with potential transduction agents and their hosts.

Figure 24. Rarefaction curves for all the microbial metagenomes used in this study.

Figure 25. Rarefaction curves for the arctic microbial metagenomes generated in this study.

Figure 26. Rarefaction curves for all the viral metagenomes used in this study.

Figure 27. Rarefaction curves for the arctic viral metagenomes generated in this study.

Index of Tables

Table 1. Metagenomes from public archives with their identification numbers.

Table 2. Characteristics of the public metagenomes used.

Table 3. Spearman's rank correlation coefficient for the different variables studied.

Table 4. Metagenomes created for this study.

Table 5. Taxonomic assignment of the viral reads and viral contigs by Metavir.

Table 6. The identified prokaryotes and the identified viruses from Figure 21.

Table 7. Hybrid reads taken as hallmarks of specialized transduction.

Table 8. Hybrid contigs taken as hallmarks of specialized transduction.

Table 9. Details of the 30 metagenomes used in the analyses in section 4.1.

Table 10. CRISPR sequences found by metaCRT in the ice and soil metagenomes.

Abbreviations

BLAST: basic local alignment search tool

Bp: base pairs

BREX: bacteriophage exclusion

ca.: *circa* (approximately)

CFB: Cytophaga-Flavobacteria-Bacteroides group

CFU: colony forming unit

CRISPR: clustered regularly interspaced short palindromic repeats

DB: database

DR: direct repeat

dsDNA: double stranded DNA

etc.: *et cetera* (and the rest)

e.g.: *exempli gratia* (for instance)

EPS: extracellular polymeric substances

et al.: *et alii* (and co-workers)

GRACE: gravity recovery and climate experiment

GTA: gene transfer agents

HGT: horizontal gene transfer

HIV: human immunodeficiency virus

HR: homologous recombination

ICE: integrated conjugative elements

i.e.: *id est* (that is)

Kb: kilo base pairs

MDA: multiple displacement amplification

NCBI: national center for biotechnology information

NHEJ: non-homologous end joining

nm: nanometer

OTU: operational taxonomic unit

ρ : Spearman's rank correlation coefficient

PAM: protospacer adjacent motifs

PAR: photosynthetically active radiation

PBS: phosphate buffered saline

PCA: principal component analysis

PCR: polymerase chain reaction

QC: quality control

ssDNA: single strand DNA

ssRNA: single strand RNA

sp: spacer

T: temperature

UV: ultraviolet

VLP: viral like particle

1. Introduction

1.1 Viruses in Microbial Ecology

1.1.1 The Discovery of Prokaryotic Viruses

Prokaryotic viruses were first described in 1915 by William Twort (Twort, 1915), although they were not identified as such until 1917 when Felix d’Herelle introduced the term bacteriophage (d’Herelle, 1917). Fleming would not discover antibiotics until 1928 (Fleming, 1929) and thus, in a pre-antibiotics era, d’Herelle saw the potential of bacteriophages in the treatment of human bacterial diseases (d’Herelle, 1929). However, members of the scientific community at the time were not convinced by d’Herelle’s affirmations about the existence of bacterial viruses and rather thought that his observations were due to bacterial lysozymes (Ninetieth Annual Meeting of the British Medical Association, 1922). This scientific dispute led to the in-depth study of the nature of bacteriophages (d’Hérelle and Smith, 1926) and was only ended with the direct visualization of phages by electron microscopy (Figure 1) (Ruska, 1940; Luria and Anderson, 1942).

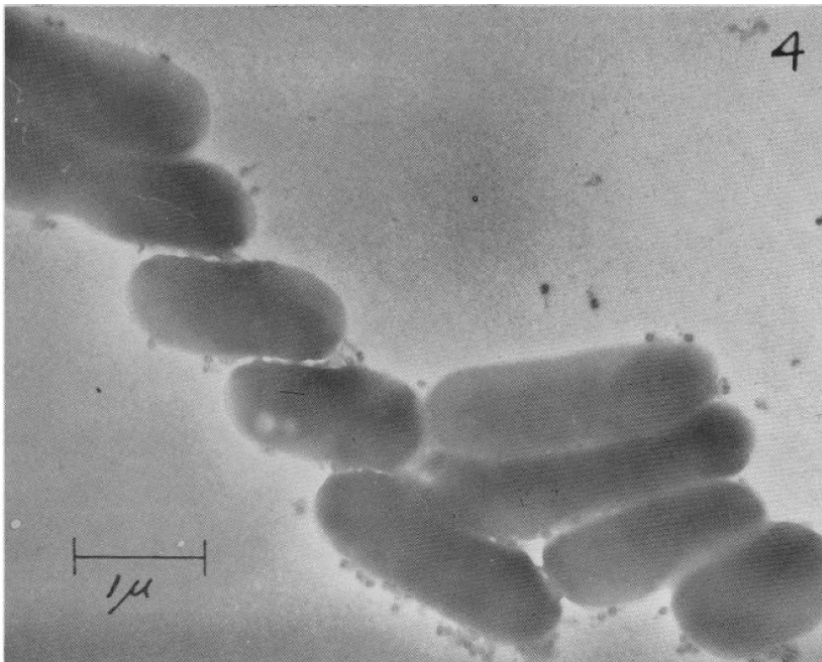


Figure 1. One of the first electron microscopy images of bacteriophages. *E. coli* cells in a suspension of “*bacteriophage anti-coli PC*”. In Luria and Anderson, 1942.

In the 1970's, several bacteriophages were reported for microbial species that would later be classified as Archaea (Torsvik and Dundas, 1974; Wais et al., 1975). Eventually, archaeal viruses were shown to have exceptional morphologies, few of which are shared with bacteriophages (Prangishvili, 2013). Many of their genomes are also unrelated to bacterial viruses, suggesting different evolutionary origins (Prangishvili, Forterre, et al., 2006; Prangishvili, Garrett, et al., 2006).

The term bacteriophage relates only to bacterial viruses, phage being an accepted synonym. In the case of Archaea, the term virus is used rather than phage. Phage derives from the greek *phagein* "to devour" while virus derives from the latin *virus* "toxin" or "venom". There seems to be a lack of consensus as to the use of phage or virus (Abedon and Murray, 2013). Thus in the writing of this thesis, the terms phage and virus are used as synonyms, always referring to archaeal and bacterial viruses in general.

1.1.2 Viral Abundance and Diversity

Since their discovery, prokaryotic viruses have been shown to outnumber prokaryotes and are the most abundant biological entities in a range of ecosystems (Fuhrman, 1999; Williamson et al., 2013) and possibly in the entire biosphere. The total number of viral particles on Earth has been estimated to be in the order of 10^{31} , one order of magnitude higher than that of bacteria (10^{30}) (Whitman et al., 1998; Suttle, 2005; Wommack et al., 2015). They are not only incredibly abundant but also extremely diverse. Their genomes can be linear or circular, they can be made out of double stranded DNA (with or without an RNA stage), single stranded DNA, positive single stranded RNA (with or without a DNA intermediate), negative single stranded RNA and double stranded RNA. Their structure is also diverse, with very different morphologies and capsid structures with or without an envelope and with contractile or non-contractile tails of different sizes. These features are used to classify them into the different viral taxons (López-García and Moreira, 2012).

Accordingly, up to a million different viral genotypes were found in 1kg of sediment (Breitbart et al., 2004) and the first viral metagenomic studies found an immense viral diversity hardly represented in the databases (Edwards and Rohwer, 2005). If the local

diversity is so high, it follows that the global diversity would also be large and richness estimates are in the order of hundreds of thousands of species for the marine environment alone (Angly et al., 2006). However, marine data supports the idea of a high local diversity sustained by a global pool of restricted diversity in what has been named a seed-bank viral community structure (Breitbart and Rohwer, 2005; Brum, Ignacio-Espinoza, et al., 2015). In this case, viruses might actually be able to change niches and even biomes (Sano et al., 2004). The pattern of distribution of viruses in the environment is still a subject of study with every new study describing previously unknown viral genomes, groups and even genera (Holmfeldt et al., 2013; Labonté and Suttle, 2013; Mizuno et al., 2013; Bellas et al., 2015; Roux et al., 2015).

1.1.3 Viral Regulation of Microbial Populations and Diversity

In every environment where microorganisms dwell, viruses are bound to exist because their life cycle keeps them tightly linked to their hosts. Lytic viruses replicate their DNA within the bacterial cell and after assembly, they lyse and exit the cell (Clokier et al., 2011). Alternatively, lysogenic viruses integrate their genome within the bacterial genome or stay as a plasmid and thus become a prophage. Prophages will be replicated and passed to next generations until the phage enters the lytic cycle (Weinbauer, 2004). Phages capable of leading lysogenic and lytic cycles are known as temperate viruses. One other life cycle is chronic infection, where viruses are slowly synthesized and shed by the cell without obvious cell death (Russel, 1995; Ackermann and Prangishvili, 2012).

These various life cycles can shape their hosts and the microbial communities in different ways. Cells with prophages can eventually undergo phenotypic changes due to lysogenic conversion, a well-known example is the bacterial expression of cholera toxins encoded by a phage (Waldor and Mekalanos, 1996; Bondy-Denomy and Davidson, 2014). While viruses in the lytic cycle can regulate population dynamics, microblooms of a single bacterial species will lead to an increase in the production of phages specific for this host, which in turn will reduce the number of cells in a process named “kill the winner” (Thingstad and Lignell, 1997; Wommack et al., 1999). This process not only limits microbial numbers, but can also help maintain microbial

diversity by killing cell blooms of ecologically fitter individuals, thereby protecting bacteria with less ecological success (Rodriguez-Valera et al., 2009, 2014). A collateral effect of this microbial diversity maintenance is the regulation of resource exploitation. The “kill the winner” process suppresses the growth of single bacterial species feeding on one resource and instead preserves a diverse community of species feeding on different resources (Rodriguez-Valera et al., 2009). Thus, prokaryotic phages do not only impact the microbial community, but the whole ecosystem. Moreover, the lysis of microbial cells might even have an impact on carbon cycling and hence on global biogeochemical processes (Suttle, 2005, 2007).

1.1.4 Viral Role on Microbial Evolution

Prokaryotic viruses play yet another role in microbial ecology as drivers of evolution. The “Red Queen Hypothesis” states that to keep the same *status quo* you need to continually evolve (Van Valen, 1974). The name of the hypothesis is based on a quote from the book “Through the Looking-Glass” (Carroll, 1871), where the Queen says to Alice: “Now here, you see, it takes all the running you can do, to keep in the same place”. Under these terms, bacteria develop new mechanisms to prevent viral infection and viruses develop new strategies in an endless co-evolution (Stern and Sorek, 2011). This co-evolution can shape the microbial community by changing their reservoir of genetic diversity, their phenotype and their interactions with other community members (Buckling and Rainey, 2002; Giovannoni et al., 2013; Zhao et al., 2013; Koskella and Brockhurst, 2014).

The effect of phages on microbial evolution does not only rely on the development of mechanisms to escape infection, a much more direct path of microbial genetic alteration exists and is called transduction. Transduction along with transformation and conjugation are the three best known mechanisms of horizontal gene transfer (HGT) (Figure 2). Transformation is the integration of exogenous DNA in the genome, however cells must have the ability to be transformed, and so far, only 82 species have been shown to be transformable, mostly in laboratory settings (Johnston et al., 2014). The direct transfer of DNA from one bacteria to another is known as conjugation. This mechanism also requires the donor cells to have conjugative plasmids or

chromosomally integrated conjugative elements (ICEs) (Llosa et al., 2002; Frost et al., 2005). Transduction involves the packaging of microbial genetic material into the viral capsid during the lytic cycle and its transfer into another host (Canchaya et al., 2003). There are two types of transduction, specialized and generalized transduction (Figure 2). Generalized transduction is when a virus, generally a lytic one, encapsizes random pieces of microbial DNA instead of its own genome. On the other hand, when a lysogenic virus integrates into the host's genome and takes an adjacent piece of microbial DNA along with its viral genome, it is called specialized transduction (Weinbauer, 2004). Other mechanisms of HGT exist, such as gene transfer agents (GTA) and nanotubes (Popa and Dagan, 2011). GTAs are phage-like particles which carry DNA from the producing cell and are not able to replicate (Lang et al., 2012). As for nanotubes, DNA can be interchanged through these structures that can link neighbouring cells (Dubey and Ben-Yehuda, 2011).

The different mechanisms of gene transfer lead to different kinds of exchanges. For example, conjugation enables the transfer of bigger DNA sequences between cells (several hundred kb) (Norman et al., 2009) compared to transformation (~10kb) and transduction (~100kb) (Majewski, 2001). Advantages of transformation and transduction are that no direct contact between cells is needed (Barkay and Smets, 2005; Frost et al., 2005). While transduction is believed to be the least inhibited by physical distance, this process is limited by the virus host range (Majewski, 2001).

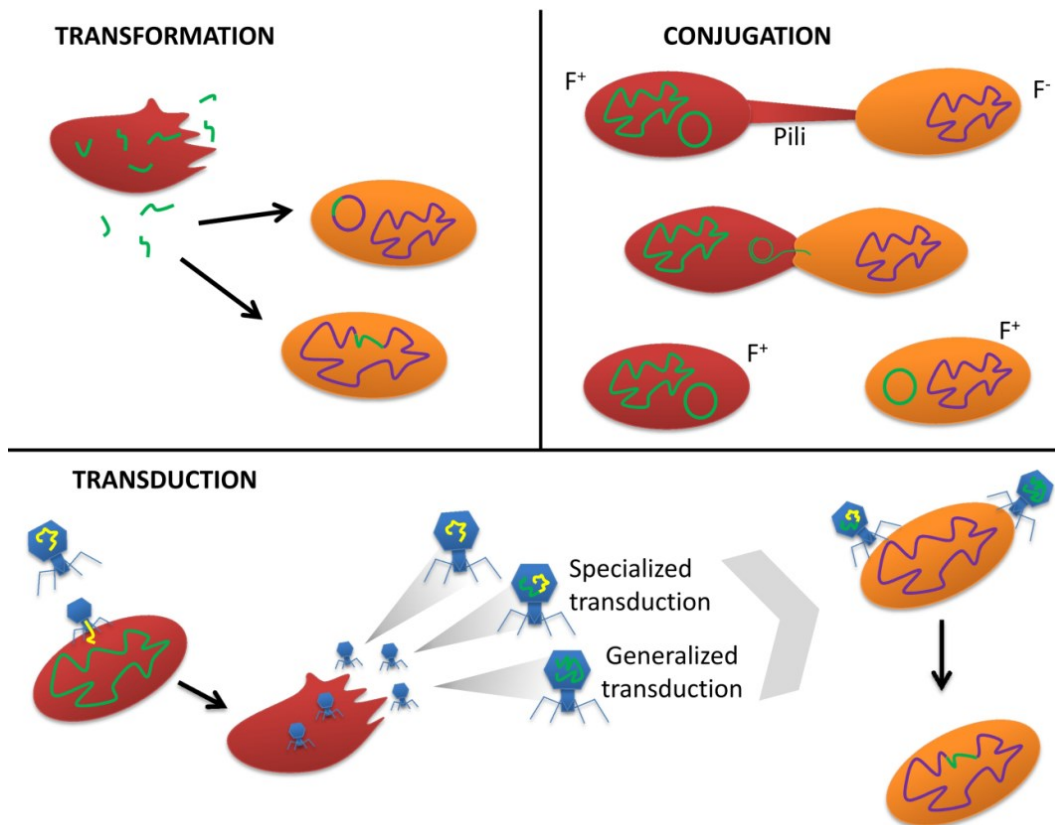


Figure 2. Schematic summary of the three major horizontal transfer mechanisms.

Even though transduction has been found to take place in a series of environments (Jiang and Paul, 1998; Del Casale et al., 2011; Solheim et al., 2013), we do not know its real implications in bacterial adaptation and evolution. Examples of specific viral transduction events with potential profit for the bacteria host have been reported. For example, marine viruses can carry a range of carbon metabolism genes (Lindell et al., 2004, 2005; Hurwitz et al., 2013). Taking into account that the rate of gene transfer by transduction in oceans has been calculated to be as high as 10^{24} genes per year (Jiang and Paul, 1998; Rohwer and Thurber, 2009), this process could be shaping microbial metabolism in the oceans. However, there are some factors that could limit the relevance of transduction as a mechanism for bacterial adaptation. One of these factors is host survival after viral infection. Available data shows that up to 20% of cells having acquired genes through transduction are able to survive the infection of the transduction agent (Kenzaka et al., 2010). Another limitation of genetic exchange through transduction is related to viral host ranges. Traditionally, phages were

believed to be highly host specific, which would limit gene exchange to closely related cells. However, with the increase in environmental virus studies, it has become apparent that phage specificity in natural environments is largely unknown (Koskella and Meaden, 2013). Several studies have identified environmental viruses with broad host ranges, capable of infecting bacteria from different groups or even different genera (Sullivan et al., 2003; Filée et al., 2005; Kawasaki et al., 2009; Matilla and Salmond, 2014). However, these could be exceptions rather than the general rule.

Consequently, transduction rates and their impact on microbial adaptation in a given ecosystem will be determined by the nature of virus-host interactions: which viruses infect which bacteria, if there are more lytic or lysogenic phages, viral host ranges, what proportion of viruses can act as transduction agents, etc. Due to the lack of knowledge on environmental viruses and their interactions with their hosts, assessing the impact of transduction on microbial evolution and adaptation is, as of today, unfeasible. The development of new techniques and approaches to try and elucidate the different variables involved in virus-hosts interactions are required.

1.2 CRISPRs as a tool

1.2.1 What are Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)?

Given their nature as libraries of invading DNA sequences, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs), could be used as tools to recover information on virus-host interactions in natural environments and in this fashion, help us assess the impact of transduction on microbial adaptation and evolution. CRISPRs are heritable and widespread systems (Fineran and Charpentier, 2012), nearly all Archaea and about half of Bacteria have been found to have them (Jansen et al., 2002; Terns and Terns, 2011). CRISPRs are made of a series of direct repeats (DR) of microbial origin interspaced by short sequences called spacers of viral origin (among others) (Figure 3) (Mojica et al., 2005). The analysis of these sequences could be used to link viruses with their hosts (Anderson et al., 2011; Garcia-Heredia et al., 2012; Stern et al., 2012).

CRISPR arrays have associated genes named *cas* and are known as CRISPR/Cas systems. Both the length of these systems and the quantity contained within a single genome are highly variable, one prokaryote appears to host up to 18 CRISPR arrays (Bult et al., 1996). Within the same CRISPR array, direct repeats (DRs) are highly conserved and they are similar or identical among close species (Jansen et al., 2002). Size range of DRs is usually between 20-50 bp (Sorek et al., 2013). They also seem to share a conserved motif at the 3' end which could act as a binding site for Cas proteins (Godde and Bickerton, 2006; Kunin et al., 2007). Some DRs have palindromic sequences that could form functional RNA secondary structures (Kunin et al., 2007). As for spacers, they specifically match viral, plasmidic and other extrachromosomal sequences (Bolotin et al., 2005; Mojica et al., 2005; Barrangou et al., 2007). Although, in certain cases, spacers have been reported to match sequences from the host genomes (Stern et al., 2010). In addition, proto-spacer adjacent motifs (PAM), which are short sequences in phage genomes, are involved in recognition and spacer uptake (Deveau et al., 2008). An AT rich leader sequence usually flanks the CRISPR (Jansen et al., 2002) and it is at this side of the loci where new spacers are inserted (Pourcel et al., 2005). The *cas* genes are generally adjacent to the CRISPR array. Up to 45 different *cas* genes have been described to date (Haft et al., 2005) yet the function of many of these proteins is still unknown (Haft et al., 2005; Makarova et al., 2015). Both DRs and *cas* genes have been used to create a classification of the different CRISPR/Cas systems discovered to date (Kunin et al., 2007; Makarova et al., 2012, 2015).

The CRISPR defense mechanism is divided into three stages: CRISPR adaptation, crRNA biogenesis and crRNA-guided interference (Sorek et al., 2013; van der Oost et al., 2014) (1, 2 and 3 in Figure 3). The first stage, CRISPR adaptation, is the uptake of a new spacer sequence. After a viral infection (or entrance into the cell of extrachromosomal DNA), short sequences from the invading DNA (proto-spacers) are inserted into the CRISPR, adjacent to the leader sequence (Erdmann and Garrett, 2012). Information on the Cas machinery involved or other details of the adaptation stage are still mostly unknown. These short specific sequences (spacers) are used by the host to recognize the virus (or other) and protect itself against future infections (Barrangou et al., 2007; Mojica et al., 2009). In order to achieve this, the next stage must take place: crRNA

biogenesis. During this stage, a transcript of the CRISPR array is synthesized, this transcript is then matured into crRNA species (Brouns et al., 2008). These crRNA will form complexes with Cas proteins (crRNA-guided surveillance complexes) (Jackson et al., 2014). The different CRISPR/Cas types have different assembly mechanisms with different Cas proteins. The crRNA-guided interference stage is the last stage, it involves the recognition and silencing of the foreign DNA by the crRNA-guided surveillance complexes (Garneau et al., 2010).

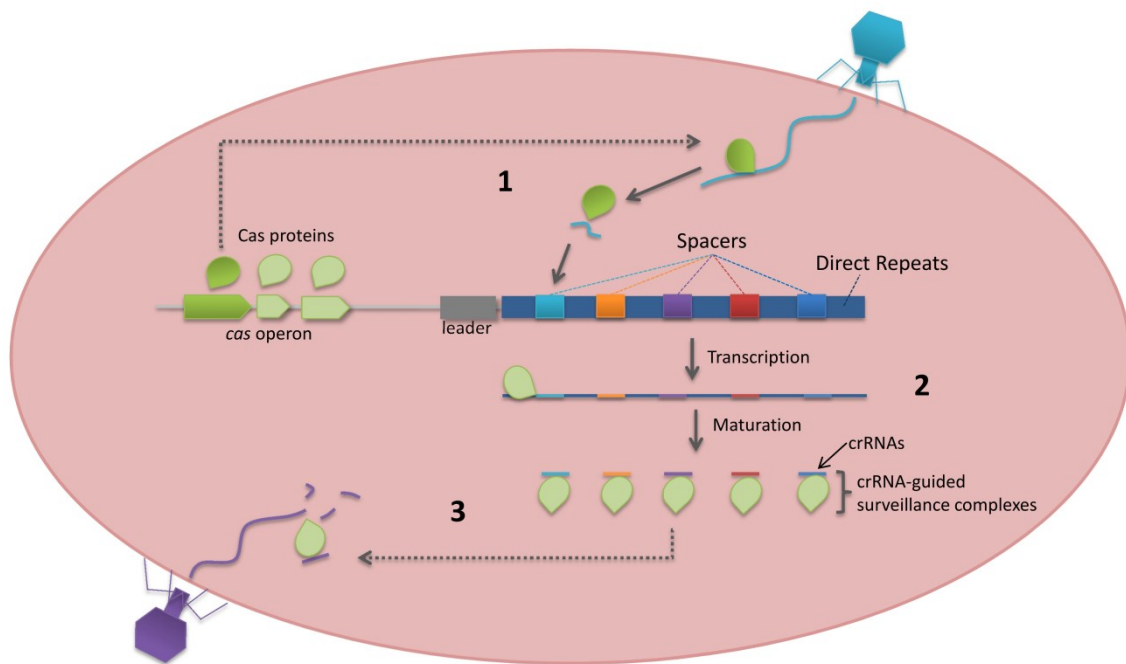


Figure 3. Schematic diagram of the CRISPR/Cas mechanism. The numbers in black indicate the different stages of the CRISPR defense pathway: **1** CRISPR adaptation, **2** crRNA biogenesis and **3** crRNA-guided interference.

Recently, phages have been seen to host CRISPR/Cas systems (Garcia-Heredia et al., 2012; Seed et al., 2013; Bellas et al., 2015). In one case, the CRISPR was used by the phage to successfully infect a bacterial host by counteracting a phage inhibitory chromosomal island (Seed et al., 2013). In another case, a lysogenic phage was found to have a CRISPR/Cas system with spacers against other phages, it was thus suggested that the lysogenic phage might provide immunity against other viruses to its host (Bellas et al., 2015). These are good examples of the Red Queen hypothesis and the constant co-evolution of prokaryotes and their viruses.

1.2.2 The Discovery of CRISPRs

The first description of a CRISPR sequence was published in 1987 when several interspaced repeat sequences were found in the genome of *E. coli* (Ishino et al., 1987). Similar sequences were found in the genomes of other prokaryotes in the following years (Hermans et al., 1991; Mojica et al., 1995; Bult et al., 1996). In 2000, they were seen to be present in a large group of prokaryotes through computational analysis, were deemed to play an essential role and were given the name of Short Regularly Spaced Repeats (SRSRs) (Mojica et al., 2000). In 2002 Mojica *et al.* and Jansen *et al.* named these sequences Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) (Jansen et al., 2002). In this same publication, *cas* genes were described and often found to be adjacent to the CRISPR array. But the role of CRISPR/Cas systems was not discovered until the finding of spacer sequences matching phage and plasmid DNA by three different research groups in the same year, it was then proposed that they could mediate against extrachromosomal DNA invasions (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Scientific evidence of their role was provided by a study showing that spacers acquired after a phage infection conferred immunity to the bacteria against that phage (Barrangou et al., 2007).

1.2.3 CRISPR/Cas9 Genome-Editing

The very recent discoveries and applications surrounding Cas9 proteins are worth mentioning. CRISPR/Cas9 genome-editing enables targeted and efficient gene restoration strategies (Dow et al., 2015) as well as elucidation of gene function at a genome-wide scale (Shalem et al., 2014). It is a promising technology that will likely lead to many interesting discoveries and hopefully to effective genetic therapy.

Type II bacterial CRISPR/Cas systems rely on the Cas9 protein for all three stages; adaptation, biogenesis and interference (Charpentier et al., 2015). In the adaptation stage, the protein selects the spacers by recognizing the PAM sequence (Heler et al., 2015). Cas9 is involved in the biogenesis by stabilizing the crRNA-tracrRNA complex during the maturation process (Deltcheva et al., 2011). tracrRNA or trans-activating crRNA is only present in type II CRISPR/Cas systems, it is complementary to the crRNA and it is involved in its maturation (Chylinski et al., 2013). The crRNA-tracrRNA complex

is also called the RNA duplex and forms a complex with the Cas9 protein. This complex will recognize and introduce double-stranded breaks in the target DNA (Jinek et al., 2012).

The CRISPR/Cas9 genome-editing tools that are being developed rely on the Cas9 endonuclease activity which produces breaks in targeted DNA and triggers the subsequent cellular repair process (Zhang et al., 2014). The cellular repair process might be either non-homologous end joining (NHEJ), which is prone to error, but quick, or the homologous recombination (HR), which is error-free (Ma et al., 2014). NHEJ will introduce small deletions or insertions at known sites, these can inactivate a targeted gene. The use of CRISPR/Cas9 followed by NHEJ mechanism has been deployed for gene editing with apparent success (Wang et al., 2013; Ousterout et al., 2015). In the presence of a homologous template DNA, the HR repair process takes place. Although the requirement of donor DNA makes the process more complicated from the genome-editing point of view, it allows for precise modifications (Ma et al., 2014). Nevertheless, promising studies have managed to integrate foreign DNA in targeted sites through CRISPR/Cas9 followed by homologous recombination repair processes (Mali et al., 2013).

So far, the CRISPR/Cas 9 tools have been used to disrupt the latently integrated provirus genome of HIV in infected human *in vitro* cells, blocking its expression (Ebina et al., 2013). It could even provide a defense strategy against new viral infections of HIV or other viruses (Liao et al., 2015). A recent publication claimed to be able to correct dystrophin mutations that cause Duchenne muscular dystrophy *in vitro* through this technique with promising results in mice (Ousterout et al., 2015). Another set of studies use the system to induce and model the emergence of human carcinogenic tumors (Torres et al., 2014; Matano et al., 2015; Zuckermann et al., 2015). All of these studies, and many others not mentioned here, highlight the relevance of CRISPR/Cas9 tools and the possible impact they could have in prevention and treatment of human diseases.

1.2.4 Use of CRISPRs in Linking Viruses with their Hosts

Given that they stock sequences from past infections (spacers), CRISPRs have been suggested as a potential tool to identify virus-hosts interactions in complex environmental communities (Anderson et al., 2011; Stern et al., 2012). The use of CRISPRs has proved to be useful in the recovery of information on viral populations and dynamics (Snyder et al., 2010; Brodt et al., 2011). In addition, the spacers have been proposed to insert in the leader end of the CRISPR could also allow for chronological inferences of the infections (Terns and Terns, 2011). CRISPR could therefore become a tool for the exploration of microbial communities in natural environments. They provide a new way of outsourcing information from sequencing data other than by sequence similarity.

Several software programs have been developed to browse for CRISPRs in genomic and metagenomic data (Bland et al., 2007; Edgar, 2007; Grissa et al., 2007a; Rho et al., 2012; Skennerton et al., 2013). This software helped recover CRISPRs from high throughput sequencing data from environmental samples. Direct repeats within CRISPRs are specific for taxonomic groups, often at the strain or species level (Mojica et al., 1995; Jansen et al., 2002; Kunin et al., 2007). Thus, we suggest that they could help identify which bacteria have which CRISPRs. At the same time, the spacers could identify the viruses that have infected the cell using sequence comparison programs. This kind of analysis could lead to the construction of infection networks (Weitz et al., 2013). CRISPR-based infection networks could provide information on host ranges in a given environment by identifying which virus infects which bacteria and also. Even if we cannot determine which viral species has which spacer or which bacteria has which DR, they could still provide an ecological picture of the microbial community and their related phages without being limited by the depth of databases. This would be done by exploring how many different viral derived spacers are linked to a specific DR via CRISPRs. Determining whether the number of spacers is highly variable or conserved, or comparing CRISPR-based infection networks from different environments might provide further information on the nature of viral-microbial communities.

In order to test the feasibility of creating infection networks with CRISPRs, we needed to choose a model ecosystem. Given that the interest behind the creation of this

network was to gather data on viral-host interactions to help assess the role of transduction in microbial adaptation, we decided to search for an environment where transduction could have an important role.

1.3 Microbial Communities in Cold Environments

1.3.1 The Relevance of Cold Environments in Viral Studies

Cold environments have been described as possible hot spots for microbial evolution driven by viruses (Anesio and Bellas, 2011) and could provide good model ecosystems for our study. Reasons behind this proposition include abundant concentrations of viruses in several polar ecosystems as measured through epifluorescence microscopy (Figure 4) (S awstr om, Anesio, et al., 2007; Foreman et al., 2011, 2013; Bellas et al., 2013), although viral concentrations are highly variable and seem to be niche-dependent. The highest concentrations reported are in cryoconite sediments from arctic glaciers (Bellas et al., 2013), reaching values in the order of 10^9 viruses per gram, which are comparable to viral abundances found in temperate soils (Srinivasiah et al., 2008). Cryoconite holes, antarctic lakes and sea-ice have also been found to host high viral diversity (L opez-Bueno et al., 2009; Bellas and Anesio, 2013; Sen ילו et al., 2014), with similar numbers of viral groups or viral capsid genes as those measured in marine systems. The lowest abundances observed to date in cold environments are from a supraglacial stream in Cotton Glacier, Antarctica (Foreman et al., 2013), with values in the order of 10^4 viruses per ml, which is two orders of magnitude less than average marine systems, but comparable to values from hot springs (Srinivasiah et al., 2008). Maybe the most notable inference from these datasets is that cold environments are comparable to other earth ecosystems with niches that can be as prolific and diverse as the most fertile natural environments.

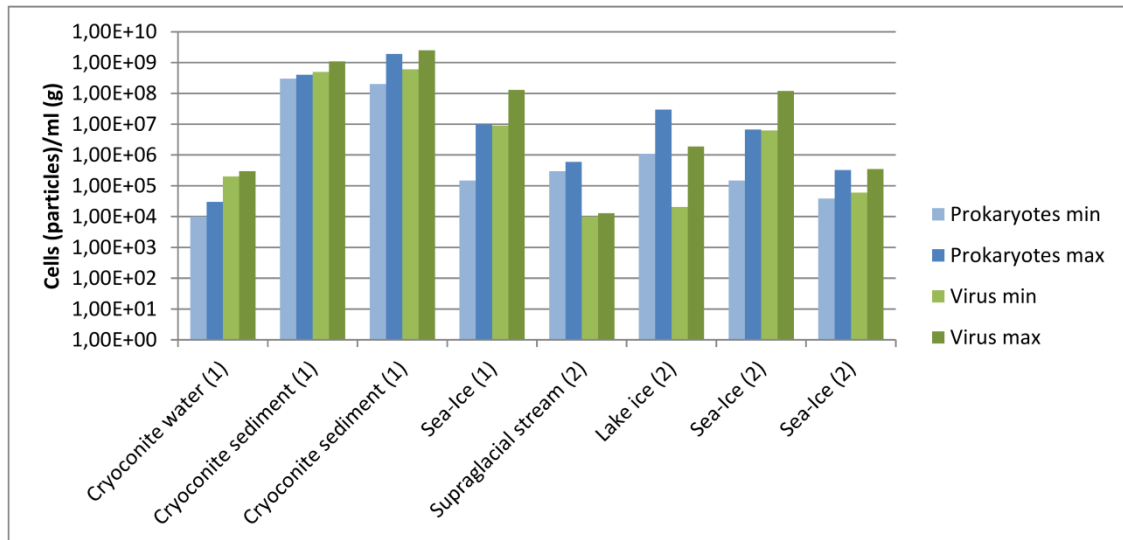


Figure 4. Cell abundances and viral abundances in polar environments (maximum and minimum values). Values are in cells/ml or virus like particles/ml for water systems and cells/g or virus like particles/g in sediments. Numbers indicate location: (1) is for arctic studies and (2) for antarctic studies. (Maranger et al., 1994; Gowing et al., 2004; Säwström, Granéli, et al., 2007; Foreman et al., 2011, 2013; Paterson and Laybourn-Parry, 2012; Bellas et al., 2013).

Another very relevant variable to take into account is host range. Studies on cryoconite holes and sea-ice have reported viruses with broad host ranges (Anesio et al., 2007; Colangelo-Lillis and Deming, 2013) that would enable them to infect different bacterial species within the same environment and possibly bacteria from different environments (Sano et al., 2004; Anesio et al., 2007; Bellas and Anesio, 2013). However, as discussed above, viral specificity in natural environments is as yet undetermined and extrapolation of specific studies to all ecosystems could be misleading. High infectivity is yet another argument for cold environments as hot spots of viral activity. Frequency of visibly infected cells were reported to be high in arctic cryoconite holes, antarctic lakes and sea-ice (Säwström, Granéli, et al., 2007; Paterson and Laybourn-Parry, 2012; Bellas et al., 2013), although with small burst sizes (*i.e.* low numbers of released viruses). Several of these values suggest that half of the bacterial population is visibly infected by viruses at any given moment. High infection rates and low numbers of released viruses could be a psychrophilic life style of viruses, although data is lacking to conclude on the nature of cold-active viruses (Wells, 2008). Nevertheless, infection rates could be especially relevant when assessing the effect of viruses in microbial adaptation in cold environments. This increased infection rate

could be explained by higher contact rates between viruses and bacteria in icy ecosystems. In this kind of environment, as water freezes channels or veins of liquid water are created, particles are excluded into these veins which can be niches for microorganisms (Price, 2000). Models of particle exclusion in glacial ice veins predict that the concentration of cells could be a thousand times higher in the veins than in the bulk melt water (Mader et al., 2006). With cells and viruses concentrated into smaller niches, contact rates between them might be higher than those estimated for bulk ice (Wells and Deming, 2006). Concentrations of extracellular DNA and viruses were calculated to be high in sea-ice brines (Collins and Deming, 2011).

The relevance of microbial processes in cold environments is emphasized by the extent of the Earth they cover. The term cryosphere, first introduced by Dobrowolski in 1923 (Barry et al., 2011), is derived from the greek *cryos*, which means "cold" or "ice" and refers to areas where water is in a frozen state (Miteva, 2008). Snow, glacial ice and sea ice, main components of the cryosphere, cover up to 1.06×10^8 km² of the Earth's surface. Snow can cover up to 12% of the Earth's surface in the winter, which is about 61 million km² (Marshall, 2011). Approximately 10% of the planet's land, about 15 million km², is covered by glacial ice in the form of ice caps, ice sheets or glaciers, storing 75% of the world's fresh water (GLIMS, and National Snow and Ice Data Center, 2012). Given their massive coverage at a global scale, ecosystems within the cryosphere could have a major and underestimated role in global biogeochemical cycling.

Overall, it seems that niches within cold environments host abundant and diverse viral populations that could have important roles in the development, management and evolution of the microbial communities. These kinds of environments could provide useful model systems for our study since transduction might have an important impact on the microbial communities inhabiting them with implications at a global scale.

1.3.2 Habitats within Glacial Ice

Interactions between members of a community are not only driven by its members but also by the structure of the ecosystem and its physico-chemical parameters. Understanding the ecology of glacial ice habitats before trying to establish infection

networks is therefore important. Low temperature, high UV light, low microbial productivity and other variables have been correlated to the increase of lysogenic versus lytic viruses (Payet and Suttle, 2013; Mojica and Brussaard, 2014). This is a good example of how physico-chemical parameters can shape virus-host interactions. This section will provide an overview of the physico chemical characteristics that define niches in glaciers.

Ice is found in the coldest places on the planet and falls within group E of the Köppen climate classification or in polar and alpine climates, which are characterized by monthly mean temperatures below 10°C (Peel et al., 2007). Mean temperatures observed in glacial ice environments can be highly variable at different depths, sites or seasons. They can reach values as low as – 30°C in the Arctic (Laybourn-Parry et al., 2012) and -50°C in the Antarctic (Price et al., 2002) and can be as warm as 0°C during the melt period in summer. Low temperatures can be deleterious for cells due to increases in membrane rigidity that will limit substrate exchange, reduce enzyme activity and damage cells with cytoplasmic water crystal formation. Even though in laboratory experiments with permafrost the coldest temperature where bacterial growth has been observed is -20°C (Tuorto et al., 2014) and activity was detected at temperatures as low as -40°C (Panikov et al., 2006), cell growth could be slow or even arrested in glacial ice (Price and Sowers, 2004). This would imply that adaptation in glacial ice needs to rely on mechanisms not involving DNA replication such as HGT, which could therefore play a major role in these ecosystems as a mechanism of evolution (Jain et al., 2003). Thermal budgets in glaciers originate from air temperature and solar radiation (Hodson et al., 2008) and can vary between different glaciers. The thermal structure of a glacier regulates the availability and distribution of liquid water, defining the different habitats that can be found in the glacier. Since viruses do not have the capability of moving and can only diffuse, the content of water and its movement within a glacier can play an essential role in virus contact rates and in virus-host interactions.

Glaciers can sustain different ecological systems on their surface (supraglacial), within the ice (englacial) and underneath it (subglacial). All of the niches within these systems are exposed to different stresses that can shape the microbial communities that

inhabit them and their interactions. Supraglacial lakes, ponds and streams are created by the melting of snow and ice on the surface of glaciers. They tend to be ephemeral systems due to vertical draining through crevasses, moulins and other channels which deliver the water to englacial and basal areas of the glacier (McMillan et al., 2007; Chu, 2013; Leeson et al., 2015; Smith et al., 2015). The most studied supraglacial systems are probably cryoconite holes (Figure 6). From the greek kruos (ice) and konis (dust), cryoconites are small aggregate granules of mineral particles, organic matter and microorganisms (MacDonell and Fitzsimons, 2008). These granules decrease the glacier's surface albedo leading to the melting of snow and ice and the creation of cryoconite holes which are filled with liquid water and a bottom layer of sediment. These holes serve as nutrient-enriched areas providing semistagnant aquatic habitats in glaciers which have been considered to be eutrophic (Takeuchi, 2001; Margesin et al., 2002). As has been mentioned before, diversity and abundance of viruses in these systems is high and their role in bacterial adaptation might be specially relevant and even unusual (Bellas et al., 2013, 2015). Bare ice can also provide a niche for some organisms as ice-mats can develop on surface of glaciers and its intense pigmentation gives the ice a red or brown colour (Vincent et al., 2004; Hodson et al., 2008).

Glacial ice is formed by layers of snow that accumulate over winter with each snowfall. As snow is transformed and frozen into ice crystals, impurities are excluded from the ice creating hyper-saline and hyper-acidic vein networks (Figure 6) (Barletta et al., 2012; Dani et al., 2012). High salinity could induce slower adsorption of the phages to the host and lower decay rates of the viral particles (Mojica and Brussaard, 2014), thus influencing viral-host relationships in these systems. Concentrations of nitrate and sulfate in ice veins from Greenland and Antarctica were 10^3 to 10^5 times more concentrated than in bulk ice (Barletta et al., 2012). These levels of nutrients are similar to what can be found in rich growth media (Mader et al., 2006) and are likely sufficient to support microbial activity in deep glacial ice (Price, 2000; Rohde and Price, 2007). Microorganisms within glacial veins are confronted with desiccation and relatively high osmotic pressure that could result in cytoplasmic water loss and intracellular ionic imbalance. Potential microbial adaptation to low water availability has also been detected. For example, genes encoding the synthesis of some

osmoprotectants, such as glycine, betaine, choline, sarcosine, and glutamate, have been detected in alpine glacier ice core metagenomes (Simon et al., 2009). Using selective hyperosmotic (sugar or salt) media, up to $9 \cdot 10^3$ CFU of xerophilic fungi identified as black yeast, yeast or a large variety of *Penicillium* were isolated from an Arctic glacial ice core (Gunde-Cimerman et al., 2003).

Within glacial ice, ice veins might not be the only habitat able to support microbial life. Microbes might be also present on the surface of trapped minerals (Figure 6) (Price, 2007). Laboratory ice formation simulations suggested that while larger particles (5 μm and 10 μm) get trapped in the bulk ice, microbial cells and smaller particles are excluded from the ice and into the veins (Mader et al., 2006). However, Rohde and Price proposed that microorganisms could also survive trapped directly in ice crystals by a coat (1 nm) of water that would be sufficient for nutrient exchange (Rohde and Price, 2007). Microorganisms may extract energy from redox reactions with putative methane production based on observations from ice cores (Tung et al., 2005).

An intrinsic property of glaciers is movement due to the pressure of their weight. The thickness of the ice, the pressure issued by it, the friction created by the movement of the ice mass and geothermal heat can melt the ice creating a layer of liquid water between the ice and the bedrock (Dowdeswell and Siegert, 2003; Llubes et al., 2006; Kyrke-Smith et al., 2014). Additionally, in some glaciers, superficial water can be delivered through crevasses and moulins to the bed of the glacier (Pimentel and Flowers, 2011; Willis et al., 2015). Along with water, organic carbon and minerals sustain the development of subglacial microbial communities in what seems a largely anoxic environment (Skidmore et al., 2005; Stibal et al., 2012; Boyd et al., 2014). The melted water under glaciers can also flow and accumulate at topographic depressions of the basin giving rise to subglacial lakes (Siegert, 2005). Up to 379 subglacial lakes have been identified in Antarctica (Wright and Siegert, 2012) and, to a lesser extent, some subglacial lakes have also been identified in the Arctic (Palmer et al., 2013; Mansutti et al., 2015). These lakes have two types of ice forming their cover, the top layer is glacial ice originated through the deposition of snow whilst the lower layer is accretion ice (refrozen lake water) (Souchez et al., 2000). In Lake Vostok, both types of ice have been seen to contain microorganisms (Christner et al., 2006). Non-subglacial

lakes can also develop ice covers that are seasonal or perennial depending on the climate, lake morphology and elevation (Williams et al., 2004). These ice covers can also harbor microorganisms (Mosier et al., 2007; Bramucci et al., 2013).

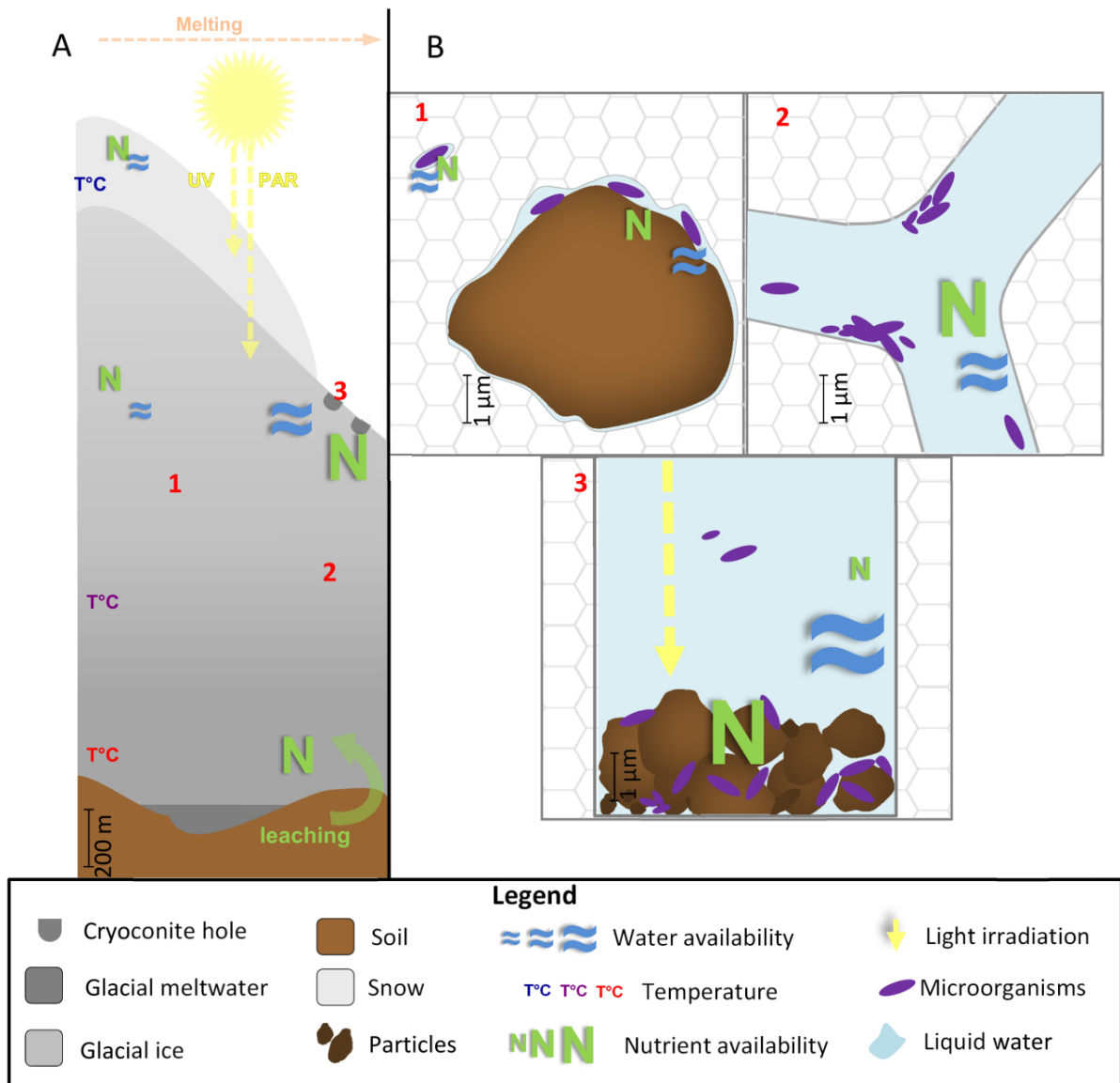


Figure 6. A glacial ecosystem (A) with some of its niches (B). Niche 1 represents a particle immersed in glacial ice, 2 an ice vein and 3 a cryoconite hole. (PAR: Photosynthetically active radiation).

All of these different niches within a glacier have different physico-chemical characteristics as can be seen in Figure 6 and could possibly host different microbial populations and virus-host interactions. They are all subject to low temperatures (albeit with some differences), while some of them have particular stresses, such as high salinity in ice veins or exposition to solar radiation in supraglacial habitats.

Another stress that might be present in these ecosystems and to which microbial cells would have to adapt is deposition of pollutants from the atmosphere. Snow efficiently scavenges organic contaminants from the atmosphere, depositing them on the surface of glaciers (Wania, 1997). The conversion of snow into ice could further transport these contaminants to englacial niches. An example of one of these contaminants is mercury, which has been measured in the arctic snowpack (Dommergue et al., 2010) and was seen to influence the snow microbial community structure (Larose, Prestat, et al., 2013). In accordance, some studies have reported possible microbial adaptation to these environmental stresses through HGT, such as mercury resistance plasmids in arctic snow (Møller et al., 2014) and plasmids encoding ultraviolet radiation tolerance genes in antarctic glacier ice (Dziewit et al., 2013). In addition, seasonal changes could also induce changes in virus-host interactions. Indeed, temperate viruses have been shown to be more abundant in polar environments compared to other regions, carrying lytic cycles in summer and lysogenic cycles in spring (Brum, Hurwitz, et al., 2015).

If seasonal variations impose changes in virus-host interactions in polar environments, we could imagine that climate change would lead to an alteration of this seasonal variation in viral cycles. Furthermore, niches within the glacier might suffer transformations such as ice melt which could impact microbial communities (Vincent, 2010), (*i.e.* higher water content in glacial ecosystems could impact microbial interactions with viruses). Glaciers grow or shrink depending on the climate, and glacier retreat has been an issue of concern for decades now. However, the different measures delivered by different techniques have created controversy on what to expect for the future (Kerr, 2013). Only 120 glaciers and ice caps have had their annual mass balance directly measured, *i.e.* 0.075% of the world's total, and only 37 of these have been monitored for more than 30 years (Bamber, 2012). Recent analysis of the data produced by GRACE (Gravity Recovery and Climate Experiment) satellites since 2002 and its comparison to previous measurements, point to a slow mass loss in Antarctica and a moderate loss in Greenland (Jacob et al., 2012; Hanna et al., 2013). Further research on GRACE's data is underway and might provide a more complete picture of the effects of climate change on ice masses.

1.3.3 Microbial Diversity and Abundances in Glacial Ice

The composition of microbial communities within glacial ice, their diversity and abundance of cells are essential factors to take into account when determining the nature of interactions with viruses. Over the past twenty years, microorganisms inhabiting the cryosphere have been increasingly studied for the potential discovery of enzymes with biotechnological interest and for fundamental research on the ecology of “extreme” environments (Miteva, 2008; Deming, 2009; Margesin and Miteva, 2011; Larose, Dommergue, et al., 2013; Arrigo, 2014). Among these cold environments, glacial systems have been seen to host unexpectedly high microbial abundances and diversity. The ecosystems within glacial ice with the greatest abundance of microbial cells registered are cryoconite hole sediments (10^9 cells/g (Bellas et al., 2013)) and debris-rich glacier ice (10^7 cells per ml of melted bulk ice (Miteva et al., 2004, 2009; Ball et al., 2014)) (Figure 7). These cell concentrations are equivalent to those of marine sediments or different soils (Whitman et al., 1998). Again, while microbial abundance is expressed by total volume of melted samples and averaged over the sample size, cells are probably concentrated into the liquid phase of the ice (Mader et al., 2006). Additionally, due to the potential small cell sizes of cryosphere microorganisms, studies have suggested that the microbial abundance of frozen water ecosystems like snow and ice might be underestimated, given that the methods used for cell count measurements generally involve filtration and the entire community may not be retained by the filters (Karl et al., 1999; Miteva and Brenchley, 2005). Altogether, this would mean that contact rates between viruses and host in ice systems might be even higher than previously estimated (Wells and Deming, 2006).

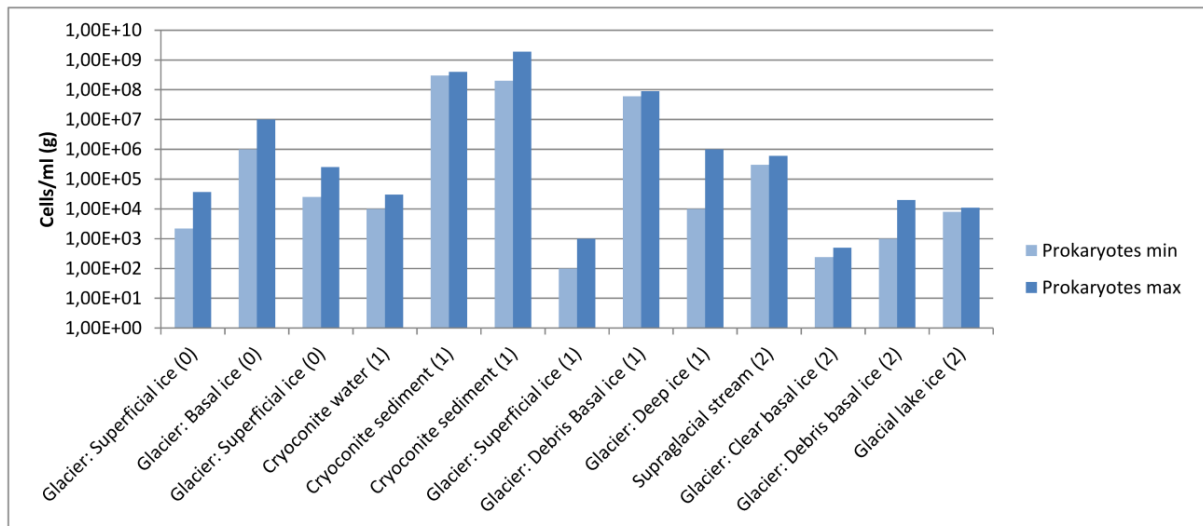


Figure 7. Cell abundances in glacial environments (maximum and minimum values). Values are in cells/ml water systems and cells/g in sediments. Numbers indicate location: (0) is for alpine studies, (1) is for arctic studies and (2) for antarctic studies. (Karl et al., 1999; Miteva et al., 2004, 2014; Mader et al., 2006; Sävström, Granéli, et al., 2007; Zhang et al., 2008; Bellas et al., 2013; Doyle et al., 2013; Foreman et al., 2013; Ball et al., 2014; Cappa et al., 2014).

Bacteria seem to be ubiquitous in ice, as they have been observed systematically in all ice environments studied to date. The bacteria identified through sequencing or culturing in these studies belong to numerous taxa, although mostly from Proteobacteria (Alpha-, Beta- and Gamma-), the Cytophaga-Flexibacter-Bacteriodes group, Actinobacteria, and Firmicutes (Miteva et al., 2004, 2009, 2014; Zhang et al., 2008; Simon et al., 2009; Doyle et al., 2013; Edwards et al., 2013; Foreman et al., 2013; Y.-X. Zeng et al., 2013; Ball et al., 2014). Unlike Bacteria, Archaea are only rarely observed. Sequences associated with the Archaeal domain were detected with a relative abundance below 1% in cryoconite sediments and arctic and alpine glacial ice (Miteva et al., 2009; Simon et al., 2009; Edwards et al., 2013). Ice also provides habitats for a range of eukaryotic microorganisms such as fungi, microalgae or heterotrophic protists (Ma et al., 2000; Jungblut et al., 2012). Algae can be found within cryoconite holes or on the surface of glaciers (Uetake et al., 2010), creating what has been named “grey ice” (Lutz et al., 2014).

Interactions, not only with viruses, but among prokaryotes can shape and partly define an ecosystem (Konopka, 2009). In the cryosphere, these relationships could be an important asset in the colonization of the different habitats. For instance, production

and excretion of certain proteins by bacteria can help in the development of microbial communities. An example of this phenomenon in ice is the production of extracellular polymeric substances (EPS) or ice binding proteins (Raymond et al., 2008; Varin et al., 2012; Singh et al., 2013). In glacier and lake ice, the production of ice binding proteins by bacteria can modify ice vein structure by inhibiting recrystallization, which seems to reduce diffusion in veins (Achberger et al., 2011; Brown et al., 2014), which could in turn decrease viral movement and thus infection rates. Cryoconite aggregates on the surface of glaciers, which can eventually lead to the formation of cryoconite holes, are an example of algal interactions with other organisms in the cryosphere (Edwards et al., 2013). Filamentous cyanobacteria along with algal and bacterial EPS production determine the size and stability of these aggregates (Langford et al., 2014). Mineral particles trapped inside as well as photosynthetic primary production are used as nutrient sources to sustain heterotrophic bacteria (Takeuchi, 2001) along with fungi and archaea (Edwards et al., 2013; Zarsky et al., 2013).

A large diversity of non-microbial organisms has also been observed in ice. Meiofauna identified in cryoconite holes includes nematodes, rotifers and tardigrades among others (De Smet and Van Rompu, 1994; Zawierucha et al., 2014). Even macrofauna, such as species of the genus *Mesenchytraeus*, a genus of annelids (ice worms) has recurrently been found inhabiting glacial ice and actually live their entire lifespans in this system (Hartzell et al., 2005). A recent study noted that ice worms from the species *Mesenchytraeus solifugus* have bacteria associated with their gut walls in what seems to be a symbiotic relationship (Murakami et al., 2015). Thus, the macrofauna provide yet another niche for microorganisms in ice. Additionally, these annelids could graze on microbial algae (Goodman, 1971) and maybe also bacteria, and thus, control their populations.

Moreover, the possibility that this icy microbial life style could have a remote or ancient origin cannot be discarded. Although the origin of life and the emergence of microorganisms on Earth remain unknown, some proposals for a cold origin of life have been made (Price, 2007, 2009). This hypothesis is supported by various studies including assays on prebiotics polymerisation in icy matrix (Miyakawa et al., 2002; Trinks et al., 2005; Menor-Salván and Marín-Yaseli, 2012; Lehman, 2013) and bacterial

phylogeny (Galtier et al., 1999). This adds significance to the quest of understanding the adaptation and evolution mechanisms of microbial communities in cold environments.

Superficial arctic glacial ice, an environment with high viral abundance and diversity concentrated into the veins of liquid water, was selected as the model ecosystem for this study based on the evidence provided above. In this niche, viruses could be a major motor of microbial adaptation and their interactions with their hosts could potentially be traced through CRISPRs. It is an environment which covers a significant portion of the Earth's surface and is subject to changes imposed by climate change, and yet its microbial communities are fairly unknown. The application of a CRISPR-based analysis to this system would serve as a "proof of concept". It might help us understand general traits of virus-host interactions and transduction and could also address more specific questions on the nature of viruses in cold environments.

2. Objectives

Viruses might be drivers of microbial adaptation to environmental conditions by enabling the acquisition of whole segments of genomic information that can recombine or integrate into the host genome. Based on current literature, this role could be especially prominent in cold environments. Thus, the first objective of this thesis was to analyze whether there was any evidence suggesting that environments with different temperatures hosted different types of viral communities. In addition, we aimed to reconstruct virus-host interactions in a model environment. By constructing infection networks through the use of CRISPRs, we gathered information on host ranges and community ecology. This knowledge might help us understand the potential relevance of transduction mechanisms in glacial ice. In addition, we aimed to identify transduction agents in the model ecosystem and the genes they were potentially transferring among organisms.

Hence this thesis had the following specific objectives:

- 1) Analyze viral diversity and CRISPR composition in environments over a wide range of temperatures.
- 2) Use CRISPRs as a tool to create infection networks and learn about specific virus-host interactions in the selected model environment.
- 3) Find evidence of the presence of transduction agents in the model ecosystem and their possible role in microbial adaptation.

3. Material and Methods

3.1 Model Ecosystem (Kongsvegen Glacier) and Sampling

We chose glacial ice as our model ecosystem for this study based on the literature discussed in the introduction. Kongsvegen glacier (79N, 13E) is located on the Spitsbergen Island in the Svalbard archipelago (Figure 5). Its mass balance has been recorded since 1987 (Hagen, 1988) and has been extensively studied, both from the geological and biological perspective (Obleitner and Lehning, 2004; Amato et al., 2007; Dommergue et al., 2010; Jacques et al., 2015).

Kongsvegen glacier is 27 km long and 189 km², and has periods of rapid movement and long periods of inactivity in what is known as a surge-type glacier (Woodward et al., 2002). It is a polythermal glacier, which means that it has both temperate and cold ice all year round. Temperate ice is at the *pressure melting point* (temperature at which the ice melts at a given pressure) and has coexisting ice and liquid water while cold ice is below the *pressure melting point* and has very little liquid water content (Irvine-Fynn et al., 2011). In Kongsvegen, the cold ice is at the ablation zone while the temperate ice is in the accumulation and basal area (Figure 5) (Björnsson et al.). Samples used in this study were taken from the accumulation area where the content of water in the ice was higher.

Ice samples were recovered from Kongsvegen glacier in Svalbard (N78°45'19.8", E13°20'11"), Norway, in the spring of 2007. A 15 meter ice core was extracted from the accumulation area of the glacier, at an altitude of 670m over sea level, using a manual drill (PICO corer) without any drilling fluids. The ice was stored in clean, sterile bags and frozen (-20°C) until further use. The core was composed of ice and firn layers, only ice layers were used in this study. On the other hand, soil was collected at the outskirts of Ny-Alesund (N78°45'20.16", E13°20'11.4") and was also stored frozen (-20°C). The soil was covered by snow and had a lichen crust over it. Samples were recovered from the first 5-10cm and were found to be very rich in clay and organic matter.

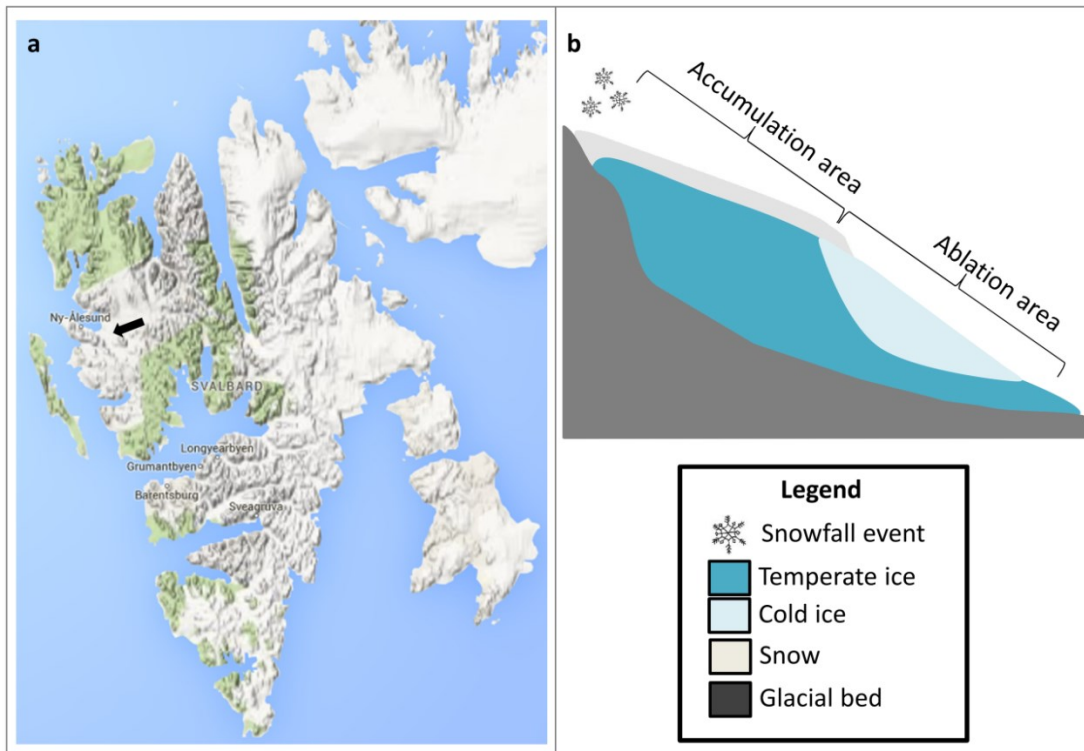


Figure 5. Location of the Kongsvegen glacier and a diagram of its thermal budget. **a.** A map of the Svalbard archipelago. The black arrow points at the location of the Kongsvegen glacier. **b.** A diagram of the thermal budget of the Kongsvegen glacier, showing the different ice types and their locations within the glacier.

3.2 DNA Extractions

Nine liters of ice were melted at room temperature and filtered simultaneously through 10 μ m 47mm (Millipore) filters and through 0.2 μ m 47mm (Millipore) to recover the bacterial fraction. Filtrate was then filtered through 0.02 μ m 25mm (Anodisc, Anopore Inorganic Membranes, Whatman) to recover viral fraction. The 0.02 μ m filters were left overnight in PBS (without Ca or Mg, autoclaved and filtered through 0.02 μ m) then put in the shaker for 2 hours before discarding the filter and proceeding with the extraction protocol. To assess the presence of viral particles and/or microbial cells, aliquots of the samples were filtered through 0.02 μ m filters, stained with SYBR Gold (Life Technologies) and examined through epifluorescence microscopy as described by Thurber *et al.*, 2009.

Nine hundred grams of soil were used for DNA extraction, following a modified protocol from Thurber *et al.*, 2009. Briefly, 20 ml of PBS (without Ca or Mg in order to

decrease the adherence of cells to soil particles, autoclaved and filtered through 0.02µm filters) were added to every 50g of soil and then blended with a Waring blender. The mixture was then centrifuged for 10 minutes at 2500g and the supernatant was retrieved with sterile syringes. The supernatant was filtered through 0.45µm 47mm filters (Millipore) and then through 0.2µm 47mm filters. Examination of the viral concentrate through epifluorescence microscopy was done as described above. Microbial DNA was recovered from the filters while viral DNA was recovered from the filtrate.

Microbial extraction of soil was done along with microbial extraction of ice as described by Larose et al., 2010. The extraction yielded 28 ng/µl of soil microbial DNA and in the order of 2ng/µl for ice microbial DNA, samples were stored at -20°C until further use.

Viral DNA was extracted following the protocol from Thurber et al., 2009, yielding 0.502ng/µl of DNA from ice and 3.83ng/µl of DNA from soil, based on Qubit DNA HS fluorometric quantitation (Life Technologies). Multiple displacement amplification was performed for viral and microbial DNA (illustra GenomiPhi HY DNA Amplification Kit, GE Healthcare Life Sciences, NJ, USA) following manufacturer instructions in order to have enough DNA for sequencing. DNA was then cleaned with GeneClean Turbo Kit (MP Biomedicals, France).

3.3 Sequencing

DNA from viral and microbial extractions from ice and soil was sent for metagenomic sequencing using GS FLX System (Roche-454 Life Science, Beckman Coulter Inc., MA, USA).

16S rRNA gene sequencing from ice and soil microbial DNA extracts was done through 454 Pyrotag sequencing for regions V4 to V6. Barcoded amplicons were prepared using MID-primers, each 50µl PCR reaction contained: 37µl water, 5µl Titanium *Taq* PCR Buffer, 1µl of each MID-primer pair, 1µl of dNTP Mix, 1µl Titanium *Taq* DNA Polymerase (Clontech) and 4µl (10ng) of template DNA. The PCR cycling conditions were: 1 min at 95°C for denaturation followed by 30 cycles of 30 seconds denaturation

at 95°C, 3 min annealing and elongation at 68°C and one cycle of 3 min at 68°C. The amplified DNA was size selected by excising the appropriate gel bands and purified using Illustra GFX PCR DNA and gel band purification kit. The DNA was quantified using Qubit DNA HS assay (Life Technologies). Amplicons were pooled in equimolar concentrations. The amplicon libraries were sequenced by Beckman Coulter Inc. (MA, USA) using 454 platforms.

3.4 Bioinformatics

3.4.1 Recovery of Public Metagenomic Data

In order to test a relationship between viral diversity, host range and temperature 30 metagenomes were downloaded from different web based archives as presented in Table 1.

Table 1. Metagenomes downloaded from public archives with their identification numbers. Archives used: MG-RAST (<http://metagenomics.anl.gov/>), European Bioinformatics Institute (EBI) (<https://www.ebi.ac.uk/metagenomics>) and National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>).

Metagenome	Download Archive	Archive ID
Albufera 1	MG-RAST	4516288.3
Albufera 2	MG-RAST	4516289.3
Albufera 3	MG-RAST	4516290.3
Amazonian Soil 57	MG-RAST	4477757.3
Amazonian Soil 90	MG-RAST	4477790.3
Amazonian Soil 94	MG-RAST	4478294.3
Antarctica 104	EBI	SRS010104
Antarctica 106	EBI	SRS010106
Antarctica 115	EBI	SRS010115
Arctic Soil 125	MG-RAST	4450125.3
Arctic Soil 126	MG-RAST	4450126.3
Arctic Soil 127	MG-RAST	4450127.3
Baltic Sea 1	EBI	ERS240712
Baltic Sea 2	EBI	ERS240713
Baltic Sea 3	EBI	ERS240714

Coral Reef California	MG-RAST	4447483.3
Coral Reef PAB	MG-RAST	4447551.3
Coral Reef Timbebas	MG-RAST	4447862.3
Glacier1	NCBI	SRR002326
Glacier2	NCBI	SRR002327
Glacier3	NCBI	SRR002328
Human Gut 31	MG-RAST	4478631.3
Human Gut 32	MG-RAST	4478632.3
Human Gut 86	MG-RAST	4473686.3
Polar Mats McMurdo	MG-RAST	4445845.3
Polar Mats MIS	MG-RAST	4445126.3
Polar Mats WHI	MG-RAST	4445129.3
Puget Sound P1	MG-RAST	4460178.3
Puget Sound P28	MG-RAST	4460180.3
Puget Sound P5	MG-RAST	4460188.3

3.4.2 CRISPR Recovery from Metagenomic Data

Several software programs have been developed in order to search for CRISPR sequences in metagenomic data, using different algorithms. In this thesis, three different software were used: metaCRT, Crass and Piler-CR. All programs were installed and run on a Unix system.

MetaCRT is a tool for *de novo* identification of CRISPR sequences, designed to be used with assembled metagenomic data. It can be used with unassembled reads but works best with long sequences. It was developed by Rho *et al.*, 2012 based on the software CRT (Bland et al., 2007), to take into account incomplete repeats that could appear in metagenomics contigs. MetaCRT looks for short repeated kmer's which are spaced by a similar distance, then it elongates them to the actual length of the direct repeat. Once the direct repeats are found the algorithm checks if the spacers are non-repeated and similarly sized. The flanks of the candidate CRISPRs are then checked for

similar repeats in order to see if the CRISPR sequence can be elongated taking into account that repeats might not be identical.

Crass (Skenner et al., 2013) identifies and reconstructs CRISPRs from unassembled metagenomic data. Crass carries out a different approach depending on the length of the read, thus it first separates long reads (≥ 177 bp) from short reads. For short reads, the algorithm searches two direct repeats in the same read (default length 23bp), spaced by 26bps and then tries to elongate the CRISPR. For long reads, the algorithm searches for two identical short repeated kmer's (8bp) spaced by a known interval. It then uses the structure as a seed to find the same structure at the same interval across the read at least three times. Once candidate CRISPR sequences are found, they are filtered to discard sequences where DR or spacers have internal repeats motifs and with lengths that do not fall within default ranges. CRISPRs where spacers share homology with each other or where the DR's share homology with the spacers are also discarded. Crass then clusters the DR's found and creates a set of DR's types, it then searches for reads that might have only one copy of one of the DR types. Crass defines the correct DR sequences through alignment of the DR clustered into one type, this allows Crass to reconstruct CRISPR sequences from different reads.

Piler-CR (Edgar, 2007) was developed to be used with genomic data but can also be used for metagenomic data. Piler-CR uses local alignments to find hits of the genomic data to itself. It then searches for alignments of two regions that are close to each other. The algorithm calculates the coverage of the repeat in the sequencing data, identical repeats are connected together. The order of the repeats of a connected component is determined and the candidate CRISPRs are built. Adjacent CRISPR sequences are merged and consensus DR sequences are computed.

3.4.3 Alpha Diversity Indexes

Viral alpha diversity indexes were computed by the construction of OTU tables through BlastX (Altschul et al., 1990) alignments and using the Shannon diversity method. In order to do so, a protein database for prokaryotic viruses was downloaded from the Phantome, PHage ANnotation TOols and MMethods (<http://www.phantome.org/>), web page. The database had 116,112 protein sequences and had been updated in April

2015. A BlastX alignment was done using DIAMOND (Buchfink et al., 2015) with the different metagenomes against the viral proteins database. Best hits with a maximum e-value of 10^{-5} and an alignment length of at least 60% were used to create OTU tables. The Shannon Index was calculated from the OTU tables using the R package (R Core Team, 2014) Vegan 2.3-0 (Oksanen et al., 2015).

Bacterial diversity indexes were computed by the construction of OTU tables through Blast alignments and using the Shannon diversity method. The Blast alignment of the metagenomes was done against a 16S rRNA gene database of unaligned sequences from the Ribosomal Database Project (Cole et al., 2014) from October 2011. Best hits with at least 97% identity were used to create the OTU tables. The Shannon Index was calculated as described above.

3.4.4 Sequence Clustering

Spacer sequences were subjected to clustering. The spacers issued from the CRISPRs found by the three CRISPR finding software and from the three metagenomes from each environment were pooled together and then clustered at 95% identity with the use of UCLUST (Edgar, 2010).

3.4.5 MG-RAST

Metagenomes from arctic glacial ice and soil were uploaded to the MG-RAST analysis web server (<http://metagenomics.anl.gov/>)(Meyer et al., 2008) in order to obtain a global picture of the viral and microbial communities inhabiting these environments. Raw metagenomic sequences in standard flowgram format (sff) format were uploaded to the server in March 2013 along with the metadata files. The files were converted to fastq in the MG-RAST server.

Detailed description of the MG-RAST pipeline can be found in their manual (Wilke et al., 2013). Briefly, after sequences are uploaded they are preprocessed to trim low-quality regions, the yielded sequences are then dereplicated. The next step screens for reads that are near exact matches to a set of defined organisms (mouse, human, cow, etc.) and removes them from the pool. The remaining reads go through the gene calling process to profile coding sequences. After clustering and through sequence

similarity of the longest sequence from each cluster, the species composition of the samples is reconstructed. A protein database (M5NR) (Wilke et al., 2012) is used to annotate the reads through sequence similarity searches to create functional profiles of the samples. In parallel, rRNA sequences are identified through sequence similarity searches, yielded reads are then clustered. Similarity searches of representatives from each cluster against the M5rna database are performed. Data completes the taxonomic annotation of the samples.

3.4.6 Metagenomic Assembly

Viral and microbial 454 reads were assembled into contigs using the software provided by Roche Diagnostics Corporation, GS De Novo Assembler. Sequence redundancy had been previously reduced using CD-HIT-454 (Fu et al., 2012). Assembly of the viral sequences from the ice yielded 689 contigs, with a mean length of 1423bp. In the case of viral soil sequences, assembly yielded 2076 contigs with a mean length of 711bp. The assembly of the microbial reads from the ice yielded 380 contigs with a mean length of 954bp. As for the microbial reads from the soil, the assembly yielded 19 contigs with a mean length of 333bp. Assembly was also carried out for the reads from the soil DNA subjected to MDA, it yielded 76 contigs with a mean length of 820bp.

3.4.7 Metavir

Unassembled and assembled viral metagenomes were uploaded to the Metavir analysis web server (<http://metavir-meb.univ-bpclermont.fr/>) (Roux et al., 2011, 2014) to gather a general picture of the viral community in arctic ice and soil. Metagenomes were uploaded in April 2013. Data was uploaded in fasta format following the servers' instructions.

Metavir computes taxonomical assignments through sequence similarity using complete viral genomes proteins from the RefSeq database from NCBI (Tatusova et al., 2014). In the case of assembled reads, open reading frames (ORFs) are first predicted and are then used in the sequence similarity search against the database. While other threshold options are available, best hits with threshold 50 on the bit score were

selected for the analyses carried out in this study given that it is a normalized value which is comparable among different searches (Fassler and Cooper, 2011).

3.4.8 16S rRNA Gene Taxa Assignment

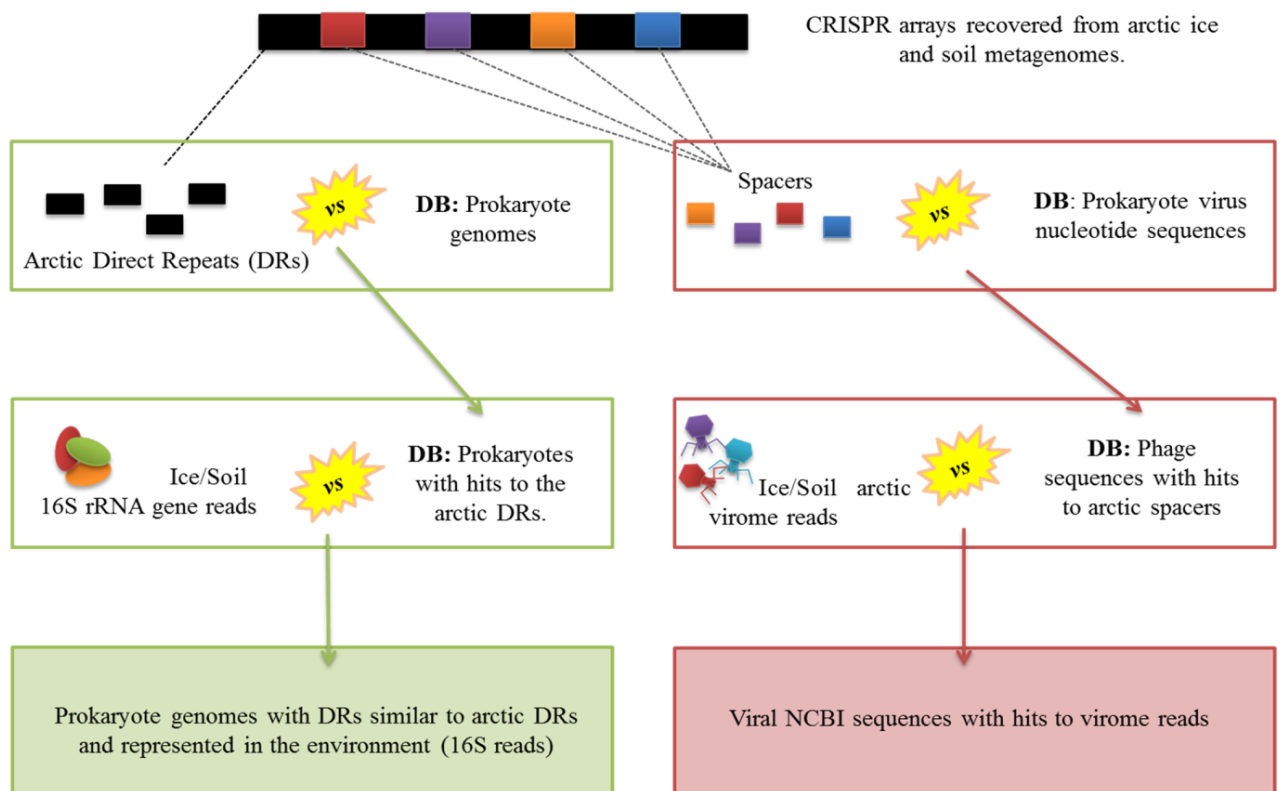
454 Pyrotag sequencing for regions V4 to V6 data was analyzed following QIIME pipeline (Caporaso et al., 2011) for *de novo* OTU picking and diversity analysis using 454 data. This pipeline involves a first step of demultiplexing the reads using mapping files defined by the user. The renamed reads from the demultiplexing are clustered into OTUs based on sequence similarity (97%). A representative sequence from each cluster is selected for downstream analysis. QIIME uses the UCLUST consensus taxonomy classifier to assign taxonomy to each of the selected representative sequences.

3.4.9 CRISPR Network Workflow

CRISPRs were searched for in the ice and soil metagenomes using metaCRT (Rho et al., 2012). The different CRISPRs found in the three metagenomes (ice, soil and soil subjected to MDA techniques) were represented in a set of infection networks (see Results and Discussion section) that were built using the software Cytoscape (Saito et al., 2012).

In order to try and identify the bacteria and viruses present in the infection networks, a workflow was constructed. The workflow followed is summarized in Figure 8. CRISPR sequences outsourced from metagenomes using metaCRT were decomposed into a DR file and a spacer file. DRs were blasted against an archaeal and bacterial genome database. The database used was a GenBank database (5193 full genomes), downloaded from their ftp site in May 2014. The organisms retrieved through the blast of the DRs against the database were considered candidate sources of the CRISPRs (default stringency parameters: maximum *e*-value 10). We then checked to see if these prokaryotic genomes were represented in our 16S rRNA gene libraries by doing a blast of the 16S rRNA gene sequences against the prokaryotic genomes we had left at this point as putative originators of the metagenomics CRISPRs (maximum *e*-value 9×10^{-10}). The result of this blast gave us a list of prokaryotes with hits to the DRs and which

were represented in our 16S rRNA database. Spacers outsourced from the CRISPRs from the ice were blasted against a prokaryotic viruses GenBank database sourced from NCBI in May 2014 (12.954 nucleotide sequences). The viruses recovered (default stringency parameters: maximum *e*-value 10) were considered to be putative sources of the spacers. In a similar way as for the DR, we checked if these viruses were represented in our metagenomes. Since there is no analogous sequence to the 16S rRNA gene for viruses, we decided to do a blast of the viral metagenomes against the putative sources (maximum *e*-value 9×10^{-10} to account for the small size of the database). A set of infection networks were constructed using the results outsourced from the application of this workflow (see Results and Discussion section) through the use of the software Cytoscape.



This symbol stands for a BLAST operation.

Figure 8. Workflow constructed to identify viruses and their hosts through CRISPRs.

To further curate the results obtained from the above mentioned workflow, a few steps were added and a new workflow was constructed as shown in Figure 9. The identification procedure of the viruses from which the spacer came (right part of the workflow Figures 8 and 9) was kept unchanged. However, the identification of the bacteria the DR came from was curated adding two more steps. First, we decided to check if the bacteria selected as candidate sources of the CRISPRs actually had CRISPR arrays in their genomes. In order to do so, the set of bacterial genomes selected through the initial blast of the DRs against the archaeal and bacterial genome database were analyzed with metaCRT. The next added step was to compare the two sets of DRs against each other. Thus, we blasted the DR issued from the ice and soil metagenomic CRISPRs against the database of DRs found in the candidate bacterial sources of the CRISPRs (default stringency parameters: maximum *e*-value 10). This left us with a list of bacterial genomes with CRISPRs whose DRs were similar to the DR we had outsourced from our metagenomic data. We then checked to see if these bacterial genomes were represented in our 16S rRNA gene libraries, as in the precedent workflow, by doing a blast of the 16S rRNA gene sequences against the prokaryotic genomes we had left at this point as putative originators of the metagenomics CRISPRs (97% identity over a minimum of 200bp). Finally a new set of infection networks were constructed through the use of the software Cytoscape (see Results and Discussion section).

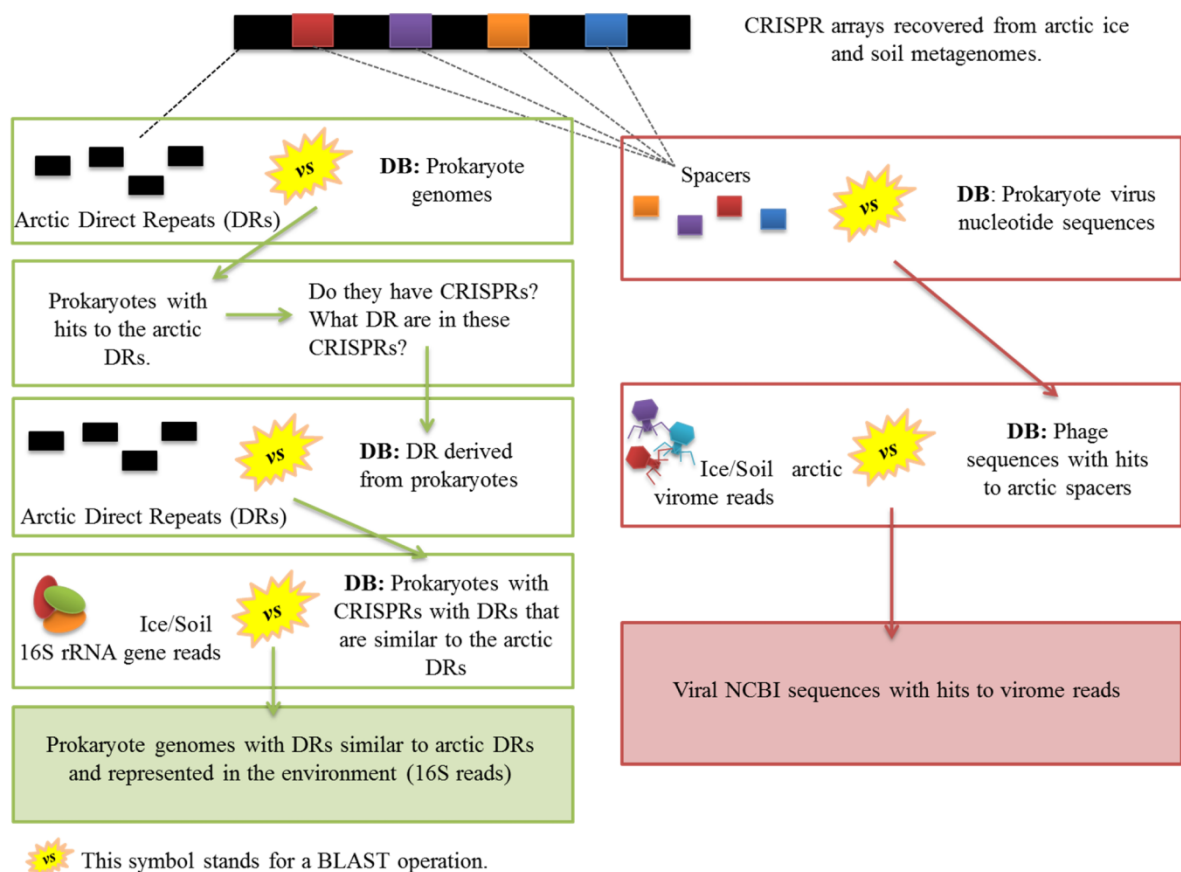


Figure 9. Curated workflow with added steps to identify viruses and their hosts through CRISPRs at a higher stringency level.

3.4.10 Analysis of Viral Sequences for Transduction

GenBank protein sequences for mercury resistance, cold shock and efflux pump genes were downloaded from NCBI in May 2014. Each of these set of sequences were blasted against the ice and the soil viral reads and contigs. Reads or contigs with hits to the protein sequences (maximum blast e -value 9×10^{-10}) were taken as sequences with microbial DNA within them. In order to test if these sequences also had viral DNA within them, we blasted them against the prokaryotic virus database. The sequences with hits to the prokaryotic virus database (maximum blast e -value 9×10^{-10}) were taken as possible tracers of transduction since they shared both microbial and viral information. The viral part of the sequence helped us to identify possible transduction agents.

3.4.11 Analysis of *Ralstonia* Phage Presence in Microbial Ice Sequences

Ralstonia phage was identified as a potential transduction agent for all three genes tested (mercury resistance, cold shock and efflux pumps genes). It was, therefore, decided to search for *Ralstonia* phage proteins in the assembled metagenomic microbial sequences of the ice. *Ralstonia* phage protein sequences were downloaded from the NCBI protein database in July 2014. A blast database was created with the bacterial contigs from the ice and the *Ralstonia* phage proteins were blasted against it. Results were filter by *e*-value taking only hits with an *e*-value of 9×10^{-10} or less. In order to assign the contigs of interest to a bacterial species, they were blasted against the GenBank database of full prokaryotic genomes. Results were again filtered by *e*-value (min 9×10^{-5}) and then by best hit.

3.5 Statistics

3.5.1 Correlations

Most of the data analyzed for correlations in this work did not have a normal distribution. Therefore, the non-parametric Spearman's rank correlation coefficient was used to test correlations. The coefficient was calculated using Minitab 17 Statistical Software (State College, PA, USA).

3.5.2 Principal Component Analysis

Principal component analyses (PCAs) were carried out with the use of MG-RAST to explore the variation among samples. This is a non-parametric test and therefore does not require a normal distribution of the data. PCA is a multivariate analysis that transforms the data on a coordinate system where the greatest variance is present in the first axis (first principal component) and the second greatest in the second axis (second principal component) and so on.

3.5.3 Rarefaction Curves

Rarefaction curves were generated through the use of the R package Vegan (Oksanen et al., 2015).

4. Results & Discussion

The results discussed in this section follow the three objectives of this thesis. As described before, our first aim was to assess if there is evidence supporting the hypothesis that cold environments have viral communities with characteristics favoring microbial adaptation through viruses. Some of the analyses carried out were done through CRISPR sequences, based on the idea that CRISPRs can provide feedback about how viruses drive adaptation and evolution. Results yielded from these analyses helped us choose a model ecosystem for our second objective: to reconstruct virus-host interactions in a model environment by constructing infection networks through the use of CRISPRs. Lastly, transduction events and their causing agents were searched for in this model environment.

4.1 Distribution of CRISPRs in natural environments

As discussed in the introduction, cold environments have been suggested to support a diverse viral community which could have broad host ranges (Anesio et al., 2007; López-Bueno et al., 2009; Colangelo-Lillis and Deming, 2013). Under these conditions viruses in cold environments could have an increased relevance for microbial communities. Not only from the point of view of microbial population control, but also from an adaptive one. In terms of transduction, higher viral diversity could mean a wider number and variety of transduction agents, and broad host ranges suggest an efficient exchange of the transducing material. It was therefore decided to carry out a preliminary analysis with publically available data. The objective was to see if the analysis of this data offered evidence consistent with the idea of environmental temperature and viral dynamics being related.

A set of 30 publicly available metagenomes were selected for these analyzes. These metagenomes belonged to 10 different environments with different temperatures. Details for each of the metagenomes used, such as number of sequences, mean sequence length and temperature of sample can be found in Table 2 along with the references of each study that produced the metagenomes. The metagenomes came from the human gut, amazonian agricultural soil, coral reef, freshwater lake in Spain, Puget Sound water, arctic soil, Baltic Sea water, polar microbial mats, antarctic lake

water and glacial ice. Although there were no true replicates sequenced, all of these environments had at least three metagenomes of close-by samples sequenced at the same time and with 454 technology. The number of reads of all the metagenomes was in the order of 10^5 reads with the exception of one of the metagenomes from the polar microbial mats that had 83,271 reads. As for the mean read length, the longest was 534bp for one of the metagenomes from Baltic Sea water, while the shortest was 183bp for one of the metagenomes from polar microbial mats. The metadata for some of the metagenomes had precise temperature data for the samples, but for others, a mean temperature of the environment was provided. Only the glacial ice metagenomes had no specified temperature for the samples or the environment. In this case, an estimation for alpine superficial ice temperature was made based on glacial thermal regime studies (Hodson et al., 2008; Irvine-Fynn et al., 2011). Altogether, samples from which the metagenomes were issued ranged from 37°C to -1°C approximately.

An advantage of sequencing technologies is the possibility of sharing and exploiting metagenomic data among scientists all around the world. However, since publicly available metagenomes are created by a variety of people using different techniques and different samples, these analyzes are always confronted by the same sort of challenges: scarce metadata, different sequencing technologies and different data-set sizes among others. Thus, when searching for public metagenomes to assess the differences among viral populations from a range of temperatures, we decided to use metagenomes which shared a set of parameters. We decided only to select metagenomes that had been sequenced with 454-technology, given that late 454-technology offered relatively long reads of about 300-600 bp, which are an asset when looking for CRISPRs through bioinformatics, since algorithms in general look for the presence of repeats interspaced by a short sequence. The longer the read the more repeats and spacers it would be able to host, and the CRISPRs sourced would be more reliable (Skennerton et al., 2013). It also means you can have more complete CRISPRs with a higher number of spacers and thus more information. Read length would not represent an obstacle if we could assemble the reads and then look for CRISPRs in the contigs, in this case other sequencing technologies such as Illumina would maybe be a

better choice since this technology is less error prone. However, several tests were done searching for CRISPRs with the available software in assembled metagenomes and up to a 90% -100% of the CRISPRs found in the reads would disappear in the assembled data. Assembly mechanisms seemed to filter out reads with CRISPRs, probably because of the repeats within them. It was thus decided to carry out our analysis with raw 454-sequenced reads.

Because of the evolution of 454-technology, metagenomes sequenced with it could have very different attributes. It was therefore decided to only take metagenomes with a number of reads in the 10^5 order, and with mean read lengths between 300-600 bp. However, finding studies that had created at least three metagenomes from similar samples simultaneously, with 454-technology and that satisfied the conditions mentioned above, was a difficult task. It was notably so with metagenomes from the cryosphere, which usually had shorter reads and in lower numbers. Because the interest that drove this analysis was to test if cold environments had a particular viral community, leaving these metagenomes out defeated the purpose of the study. We thus decided to carry on with the analyses even though, as can be seen in Table 2, metagenomes from polar microbial mats and glacier ice have mean sequence lengths below 300bp and one of the metagenomes from polar microbial mats has a total number of reads in the order of 10^4 .

Table 2. Summary table of the metagenomes used for this analysis, the environment they came from, the number of reads and their mean sequence length. Temperature values were recovered from the publications when available. (* temperature not specified in the publication, approximate value).

Environment	Metagenome	T°	Number of reads	Mean read length	References
Human Gut	Human Gut 86	37	672251	449	(Wylie et al., 2012)
	Human Gut 31	37	620142	402	
	Human Gut 32	37	644686	432	
Amazonian Soil	Amazonia Soil 57	28	144908	520	(Mendes et al., 2014)
	Amazonia Soil 90	28	144587	522	
	Amazonia Soil 94	28	138556	525	
Coral Reef	Timbebas	24	149734	393	(Bruce et al., 2012)
	Pab	24	126741	425	
	California	24	167513	447	
Freshwater lake (Albufera)	Albufera1	19.5	253260	347	(Ghai et al., 2012)
	Albufera2	19.5	262320	377	
	Albufera3	19.5	166180	375	
Puget Sound	Puget Sound P1	11.5	198750	352	(Port et al., 2012)
	Puget Sound P28	11.5	173539	354	
	Puget Sound P5	11.5	433841	356	
Arctic Soil	Arctic Soil 125	6	103366	405	(Tveit et al., 2013)
	Arctic Soil 126	4	101781	407	
	Arctic Soil 127	2	136679	411	
Baltic Sea Water	Balctic Sea 1	4.1	185809	504	(Thureborn et al., 2013)
	Balctic Sea 2	4.7	354912	515	
	Balctic Sea 3	5.7	434236	534	
Polar Microbial Mats	McMurdo	7.4	83271	371	(Varin et al., 2010)
	WHI	0.9	335705	183	
	MIS	1.8	256849	207	
Antarctica Lake	Antarctica 104	2.71	543641	376	(Lauro et al., 2011)
	Antarctica 106	0.42	593365	394	
	Antarctica 115	1.2	608567	409	
Glacial Ice	Glacier1	-1*	538886	257	(Simon et al., 2009)
	Glacier2	-1*	255505	256	
	Glacier3	-1*	282148	256	

4.1.1 Viral Diversity

In order to test if there was any apparent relationship between viral diversity and environmental temperature diversity estimates for the different metagenomes were calculated. These were done both through common diversity index calculation and CRISPR sequence analysis. The estimates were then compared among each other and related to temperature to detect possible relationships between these variables.

Viral assignments were carried out for the different metagenomes through Blast against a viral database and viral alpha diversity indexes were calculated (please refer to material and methods for details). The percentage of reads per metagenome that were assigned to viruses can be seen in Appendix1 (Table 9). A scatterplot showing the distribution of viral diversity values and temperatures is shown in Figure 10. Spearman's rank correlation coefficient (ρ) between viral diversity and temperature was -0.382 with a P-Value of 0.037. For a significance level $\alpha=0.05$, an association between the two variables can be inferred. This data supports the idea of a higher viral diversity in environments with lower temperatures.

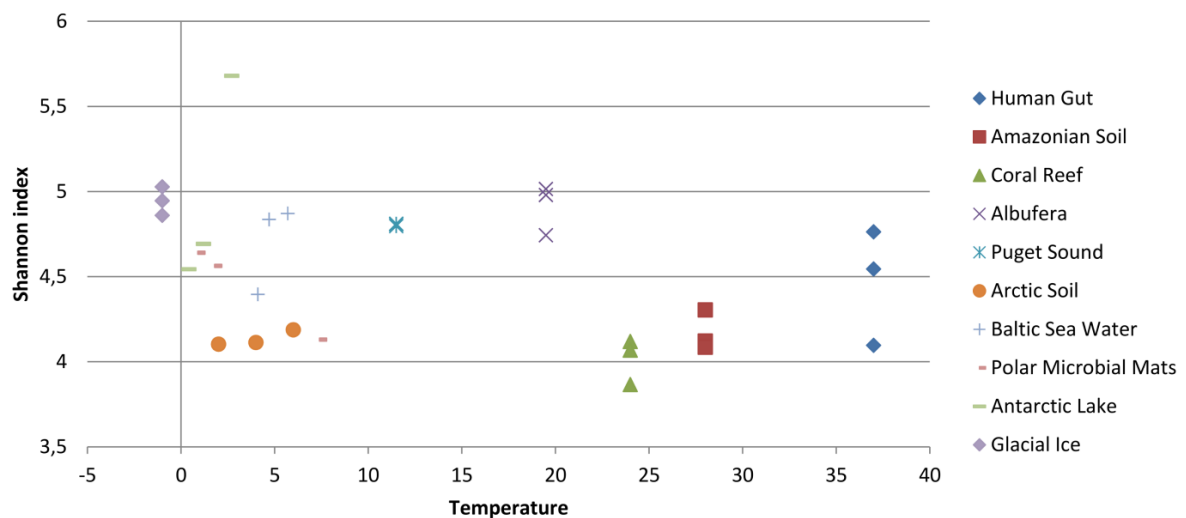


Figure 10. Shannon indexes for viral diversity of each of the ten environments plotted against samples temperature.

Given that in this work, we propose CRISPRs as a tool to learn more about viral dynamics in natural environments, we decided to see if an analysis of the environmental CRISPRs would be in accordance with the previous results on viral diversity and temperature. If a higher viral diversity in cold environments is assumed

we could also expect to see the diversity of CRISPR spacers increase with lower temperatures. Therefore, CRISPR sequences were searched for in the metagenomes using three different software: metaCRT, Crass and Piler-CR. For further information on the different algorithms used by these three software, please refer to the Material and Methods section. Each software issued a different number of CRISPRs per environment (Figure 11), with metaCRT always yielding the highest number of CRISPR sequences as compared to the other two programs.

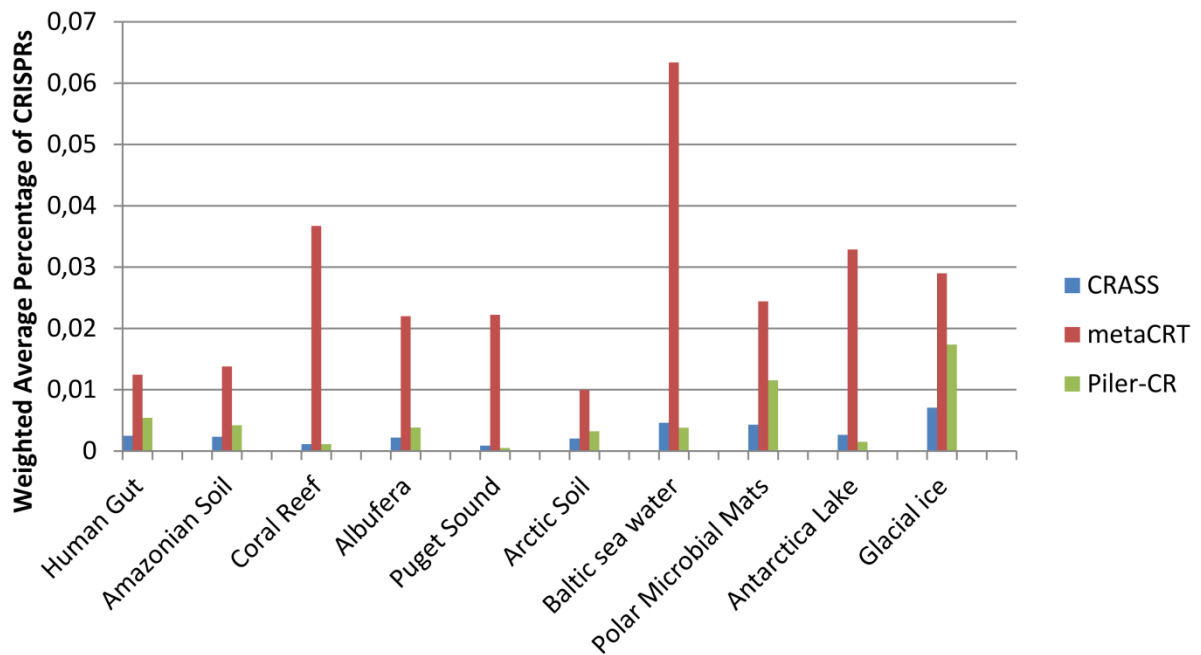


Figure 11. Weighted average percentage of the numbers of CRISPRs found in each set of metagenomes. Each value is an average of the number of CRISPRs found in the three metagenomes from each environment. Weighted average percentages were calculated to take into account the different metagenome sizes.

Spacers were outsourced from the CRISPRs found by the different software. All spacers from a given environment (3 metagenomes per environment) found by any of the 3 programs were pooled together. They were then clustered at 95% identity with the use of UCLUST (Edgar, 2010). The number of clusters per environment was plotted against temperature (the scatterplot is shown in Figure 12). The results yielded by the cluster analysis of the CRISPR spacers are in agreement with the results of the viral diversity. Environments subjected to lower temperatures present a higher number of clusters than warmer environments.

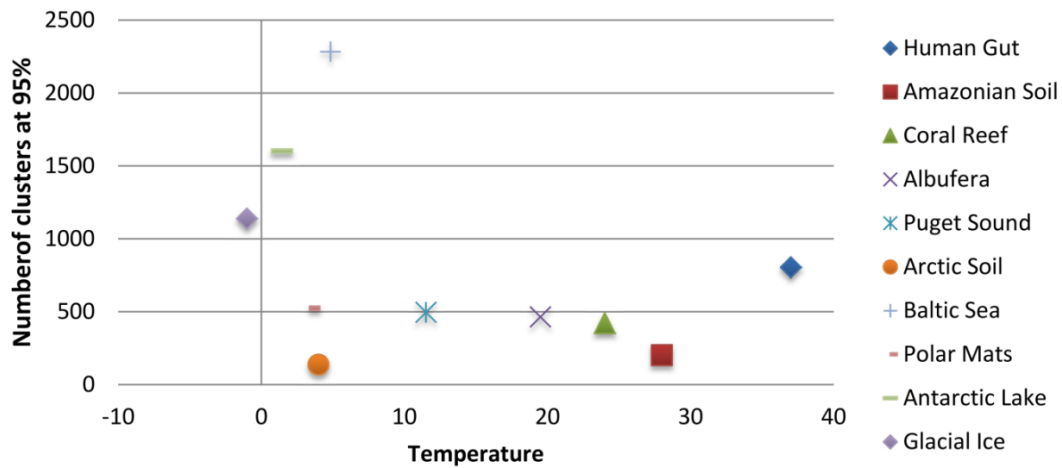


Figure 12. Scatterplot with the number of spacer clusters (at 95 % identity) per environment, plotted against mean temperature of the samples.

Estimates of bacterial diversity were also calculated for all the studied metagenomes as described in the Material and Methods section. These indexes of bacterial diversity did not have a significant correlation with temperature (ρ 0.041, P-value 0.829), unlike viral diversity. More surprisingly, viral diversity and bacterial diversity also seemed to be uncoupled based on the results yielded in this study (ρ 0.123, P-value 0.516). No significant correlation was found between viral or bacterial diversity and CRISPR proportions. Correlation coefficients for temperature, viral and bacterial diversity and proportion of CRISPRs can be found in Table 3. An exhaustive table with all the data used for these correlations, such as proportion of CRISPRs and diversity indexes per metagenome, can be found in Appendix 1. Additionally, rarefaction curves for all the viral and microbial metagenomes studied throughout this work can be found in Appendix 2.

Table 3. Spearman’s rank correlation coefficient (left columns) and their relative P-Values (right columns) for the different variables studied. Viral and bacterial diversity were estimated through Shannon index computation. CRISPR correlations were done using proportion of CRISPRs found in the different metagenomes by the different software. Cells in green are those were correlations were found to be significant.

	T°		Viral Diversity		Bacterial Diversity		CRISPRs (Crass)		CRISPRs (metaCRT)		CRISPRs (Piler-CR)	
T°	1	*	-0.382	0.037	0.113	0.553	-0.437	0.016	-0.409	0.025	-0.271	0.148
Viral Diversity			1	*	0.123	0.516	0.201	0.286	0.248	0.187	0.173	0.361
Bacterial Diversity					1	*	-0.257	0.171	-0.321	0.084	-0.107	0.575
CRISPRs (Crass)							1	*	0.485	0.007	0.712	0
CRISPRs (metaCRT)									1	*	0.158	0.404
CRISPRs (Piler-CR)											1	*

Overall, the results obtained from the analysis of public metagenomes and their CRISPRs suggest that there might be a relation between viral diversity and temperature. However, the reason why this might be so is not evident. Lower microbial abundances or sparse distribution of cells in cold environments might lead to a higher diversity of viruses. In what might be an evolutionary effort to subsist, the viral community might have evolved to have an increased richness in order to have higher probabilities of infecting cells. For this hypothesis to be valid, contact rates between viral particles and cells in ice should be lower than in other environments. Nevertheless, as was commented in the introduction, modeled contact rates in arctic sea-ice were expected to be higher in sea-ice brine channels than in underlying seawater (Wells and Deming, 2006). In addition, frequencies of infected cells in Arctic and Antarctic freshwater suggest high infection rates (Sävström, Granéli, et al., 2007). Although, glacial ice differs from sea-ice and from freshwater in structure and composition (Barletta et al., 2012), these results do not seem to point to low probabilities of cell encounter by viruses in ice or cold ecosystems. Whatever the reason might be, studies of viruses in polar environments support the idea of a high viral diversity (López-Bueno et al., 2009; Bellas and Anesio, 2013) and the analysis of the CRISPR spacers carried out in this section supported these findings.

4.1.2 Viral Host Range

Lastly, we tested the presence of broad host ranges in cold environments. Spacers are not arbitrary pieces of DNA from the infecting phage. A short sequence known as PAM (Protospacer Adjacent Motif) is required for the spacer uptake (Deveau et al., 2008). Even though there might be several PAM sequences within a viral genome, spacers from the same virus could be conserved throughout different CRISPRs from different bacteria (Paez-Espino et al., 2013). This being the case, we could assume that a repeated spacer through different CRISPRs could point to the same virus infecting several bacteria. In addition, the same mechanisms that limit the growth of CRISPRs (spacer deletions and recombination) could help reduce the appearance of spacer duplicates within the same CRISPR (Anderson et al., 2011). Thus, spacers from the same environment were pooled together and the numbers of duplicate sequences were plotted against temperature to see if there was a trend (Figure 13).

In all three cases, glacial ice and human gut metagenomes had higher numbers of duplicated spacers. In the case of the CRISPRs found with metaCRT Antarctic lake metagenomes had the highest number. These environments could have viruses with broad host ranges based on the principles explained above. Glacial environments have previously been suggested to have viruses with broad host ranges (Anesio et al., 2007). The analysis of the CRISPR spacers carried out in this section was in accordance with this idea.

The abundance of CRISPRs in natural environments could be an indicator for viral dynamics. Higher viral diversity, higher infection rates and broad host ranges in an environment could lead to an increase in the microbial mechanisms against viral infection, such as CRISPRs. The number of CRISPRs found with each of the three programs was plotted against temperature as shown in Figure 14. Spearman's rank correlation coefficient between the number of CRISPRs found with metaCRT and temperature was -0.409 with a P-Value of 0.025. As for the correlation between the number of CRISPRs found with Crass and temperature the Spearman's rank correlation coefficient was -0,437 with a P-Value of 0.016. Lastly, the Spearman's rank correlation coefficient between the number of CRISPRs found with Piler-CR and temperature was -0,271 with a P-Value of 0.148. The correlation results for the number of CRISPRs found

with metaCRT and Crass were in agreement with the idea that there are higher numbers of CRISPRs in cold environments. However, the correlation between the number of CRISPRs found with Piler-CR and temperature was not significant.

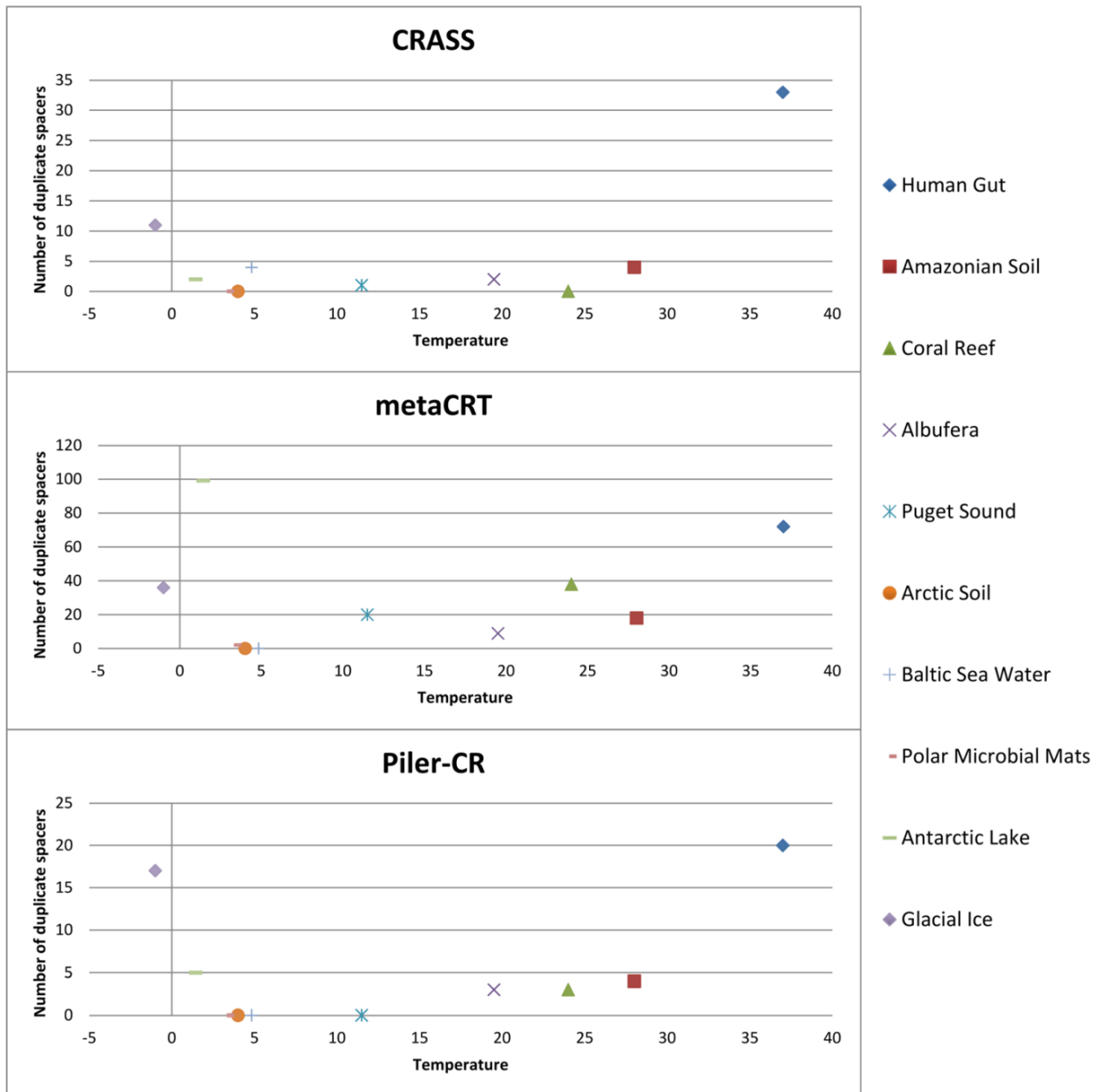


Figure 13. The number of duplicate spacers in an environment plotted against temperature for each of the different CRISPR searching softwares.

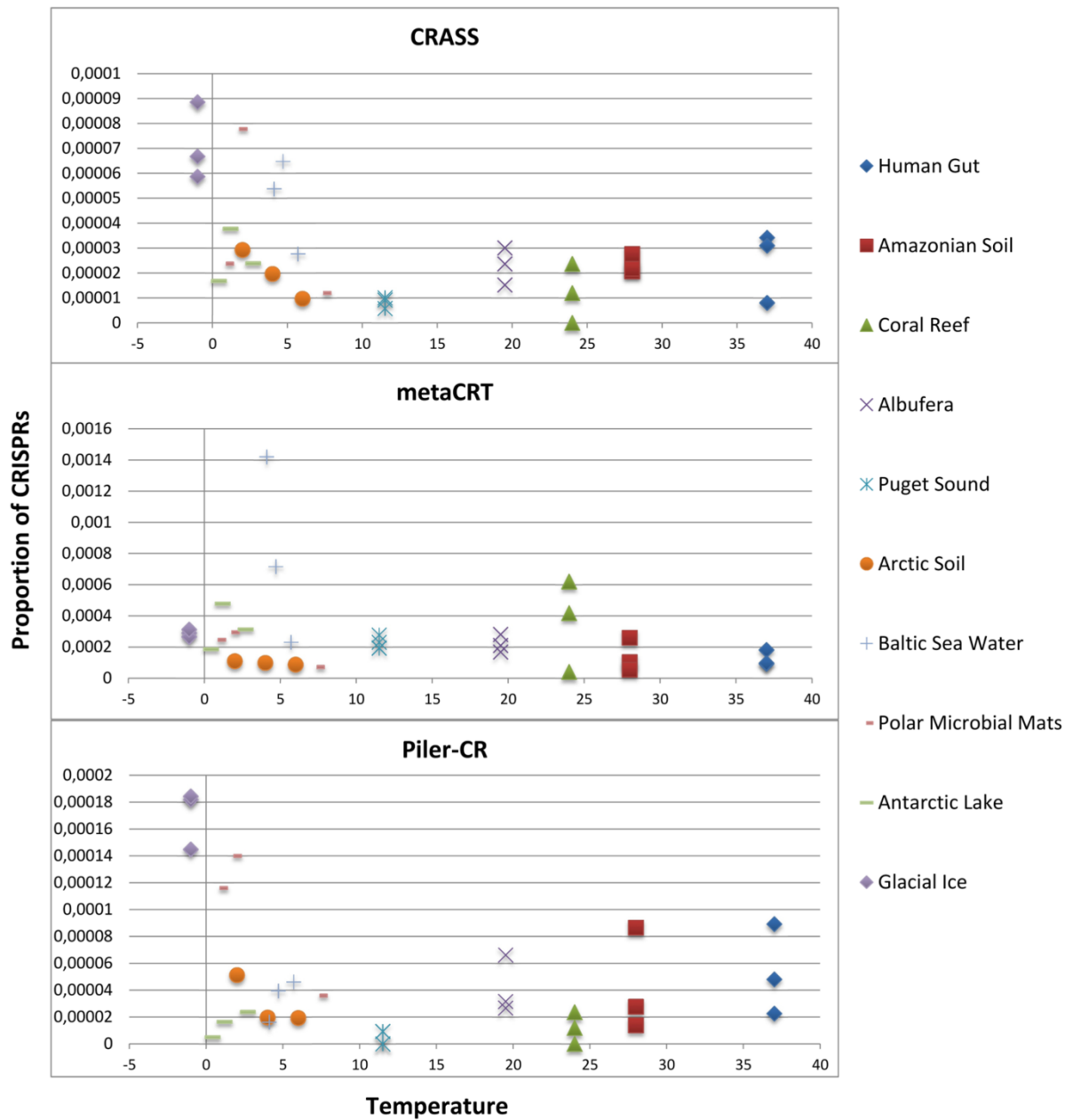


Figure 14. Scatterplots for the three different software used. The proportion of CRISPRs per number of reads of the metagenome was plotted against sample temperature.

On the whole, the work carried out with CRISPR sequences from publicly available metagenomes upholds cold environments, and particularly glaciers, as ecosystems with viral dynamics that could enhance the adaptive potential of microorganisms through interactions with viruses. In order to further explore the nature of these interactions, we focused on CRISPRs in glacial ice and soil samples from Svalbard.

4.2 Linking viruses and their hosts through CRISPRs

Samples from arctic glacial ice and soil were recovered and sequenced as described in the Material and Methods section. A first exploration of the created metagenomes was done through web based tools to have a better picture of the communities inhabiting these environments. Subsequently, CRISPR arrays were search for in the metagenomic datasets and a workflow was designed to link viruses with their hosts.

4.2.1 Overall Description of Arctic Ice and Soil Metagenomic Data

With the aim of having a global picture of the viral and microbial communities inhabiting ice and soil, metagenomic data was explored through different web based tools. Metagenomes were created from both bacterial and viral DNA from arctic ice and soil. The experimental procedures carried out to create these metagenomes are described in the Material and Methods section. Due to the low amount of extracted DNA, Multiple Displacement Amplification (MDA) techniques were used in order to have enough DNA for sequencing. Only the microbial fraction from arctic soil yielded enough DNA to enable sequencing without amplification. However, it was decided to also sequence the DNA from the same sample subjected to MDA so that the bias induced by this technique could be partially evaluated. Thus, a total of five metagenomes were created, and are presented in Table 4 with the number of reads and mean read length.

Table 4. Metagenomes created for this study with details of their contents.

Environment	Sequence Type	MDA	Nº Reads	Mean Read Length
Arctic Ice	Microbial	Yes	47,626	374
Arctic Ice	Viral	Yes	65,599	486
Arctic Soil	Microbial	No	42,078	341
Arctic Soil	Microbial	Yes	33,241	393
Arctic Soil	Viral	Yes	211,571	495

Metagenomes were uploaded to the MG-RAST metagenomics analysis server. The MG-RAST pipeline annotated over 36% of the microbial sequences from the soil metagenomes to known proteins and 20% of the microbial sequences from the ice as

shown in Figure 15. For all three microbial metagenomes, around 40% of the sequences were classified as unknown proteins and approximately 20% did not pass the quality filters. The two viral metagenomes had only 6% of their sequences that did not pass the quality filters, yet 90% of the sequences were classified as unknown. Hence, the MG-RAST analysis of the viral metagenomes should be considered with caution given the scant information the system has used to yield its results. Viral metagenomes were uploaded to another analysis server specialized in viruses, as will be discussed below, in order to address this issue.

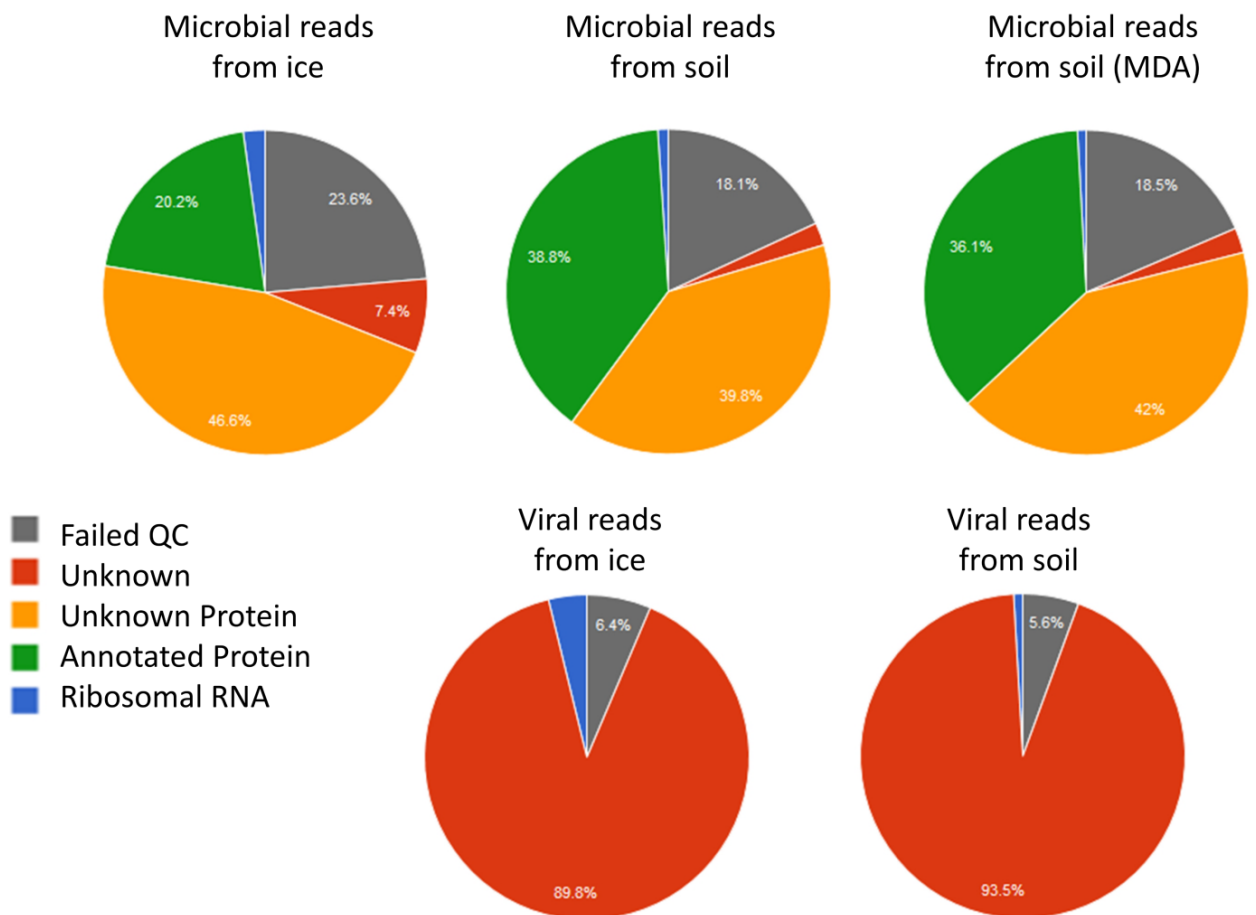


Figure 15. Classification of metagenomic reads by the MG-RAST pipeline.

The annotated sequences of the microbial metagenomes were predominantly assigned to the bacterial domain. Nonetheless, the microbial metagenome from the ice had 22.3% of the annotated sequences assigned to the domain Eukarya. Twenty-four percent of the eukaryotic sequences (5% of the total) were assigned to the *Chordata*

phylum while 23% were assigned to the phylum *Nematoda*. The relative abundances of the different phyla within the domain *Bacteria* are represented in Figure 16. The relative bacterial abundances of the soil metagenomes with and without MDA are quite similar with a slight decrease in representatives from the *Proteobacteria* phylum in the soil subjected to MDA. Regarding diversity, the rarefaction curves showed a higher diversity of the soil microbial metagenomes as compared to the ice microbial metagenome. Bacterial diversity estimates (Shannon index), computed as described in the Material and Methods section, were in agreement with these results. The Shannon index for ice was of 6.58 and for soil of 8.6 and 8.3 (with and without MDA, respectively).

16S rRNA gene sequencing from the ice and soil microbial fractions were also carried out as described in the Material and Methods section. The main purpose of this sequencing was related to the CRISPR-based infection network development and will be discussed in the following section. However, 16S rRNA gene assignments were done through the QIIME pipeline in order to compare the assignments with those yielded by the MG-RAST pipeline. It was noted that the general traits of bacterial relative abundances were highly similar between pipelines. Rarefaction curves for the microbial and viral metagenomes can be found in Appendix 2.

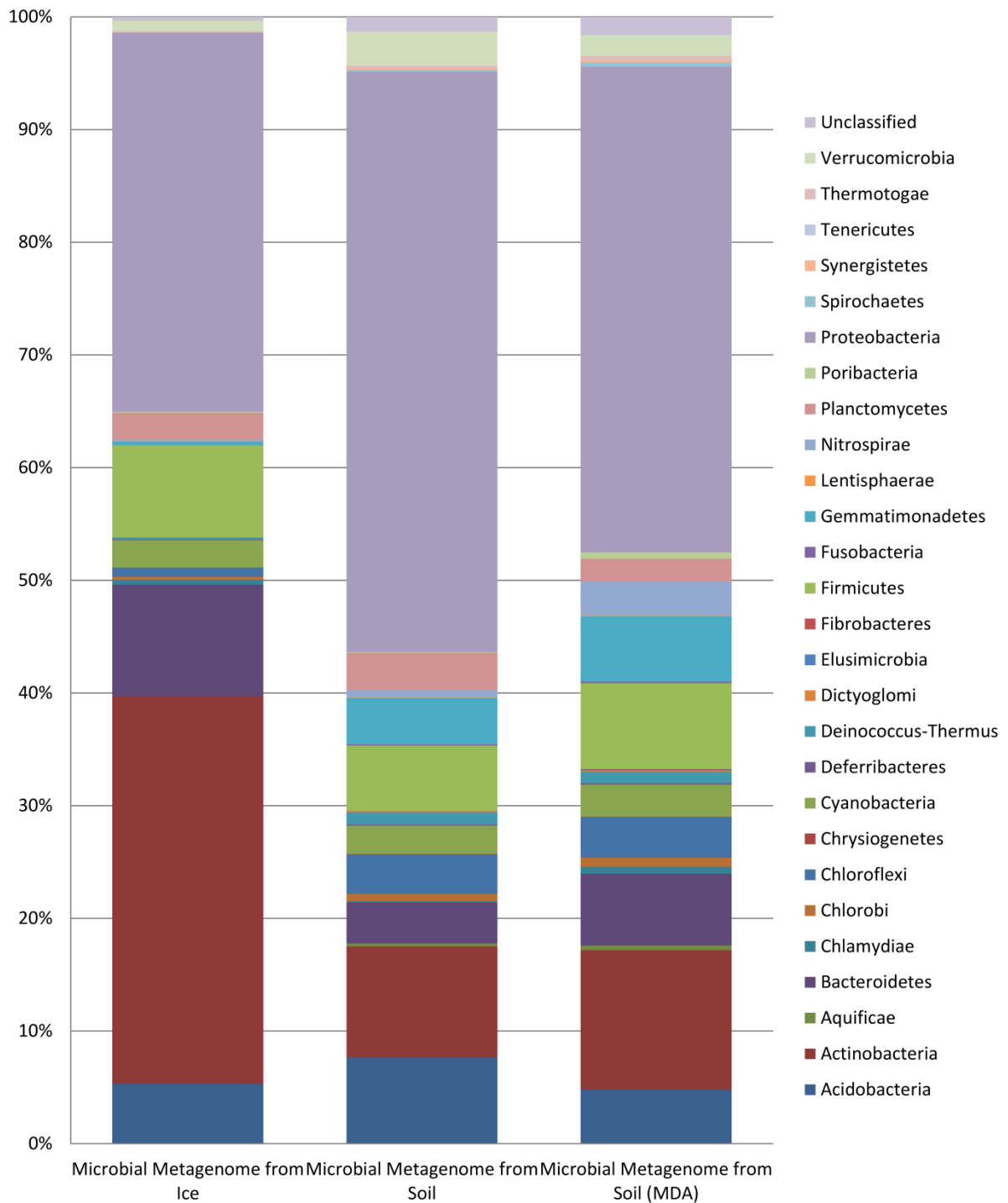


Figure 16. MG-RAST best hit taxonomical classification for the bacterial domain. Data represented as relative abundances.

The principal component analyzes (PCAs) done by MG-RAST (using normalized values and Bray-Curtis distance) based on organism abundance (best hit classification) and based on functional abundance (hierarchical classification) are shown in Figures 17 and 18 respectively. The results of the taxonomic PCA indicate that 62% of the variance is

explained by the first component which groups the samples per environment (*i.e.* ice or soil). While an additional 20% of the variance is explained by the second component which groups the samples per DNA origin (*i.e.* microbial or viral). As for the functional PCA, the first component explains 57% of the variance and it groups the sample per DNA origin (*i.e.* microbial or viral). However, this analysis does not group together the samples based on their environment. Therefore, it seems that although there is a taxonomic difference between the communities from the ice and from the soil, they are not functionally distinct. On the other hand, there are both taxonomic and functional differences among the microbial and viral fractions of the samples.

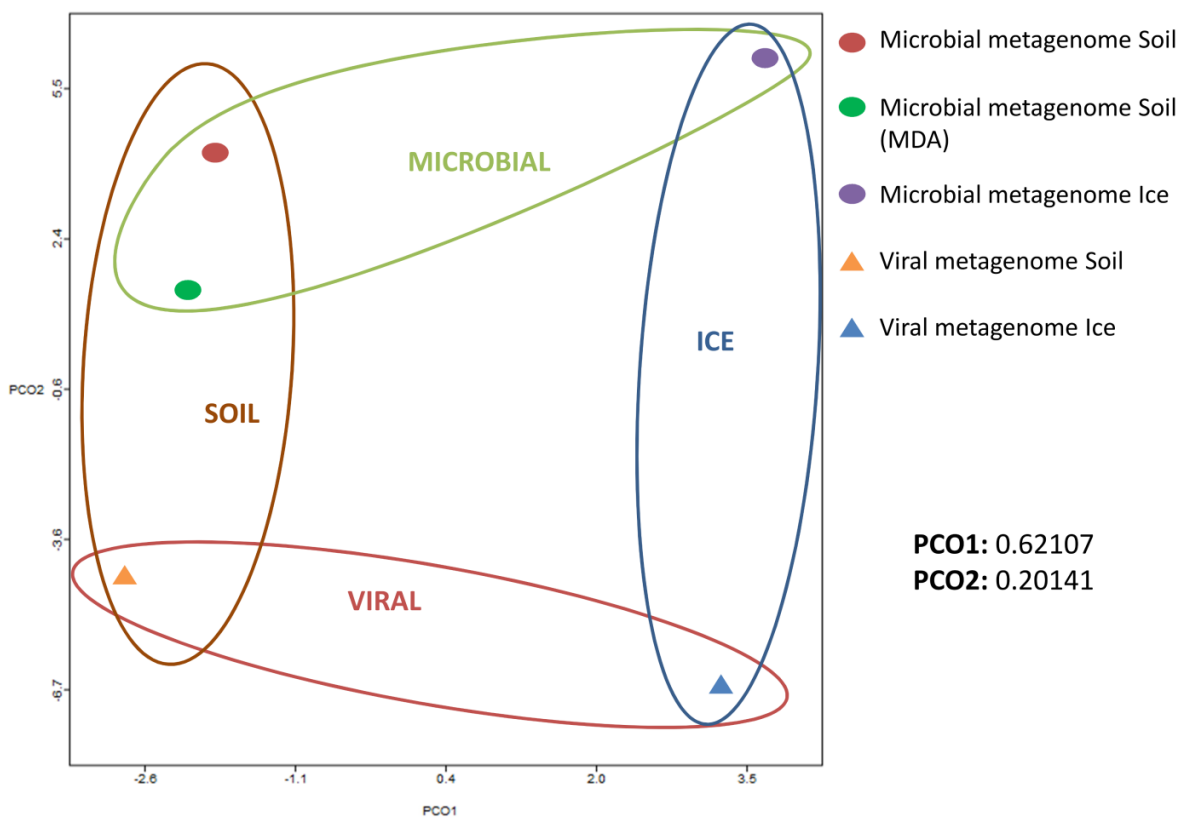


Figure 17. Principal Component Analysis from MG-RAST based on organism abundance on a best hit classification basis.

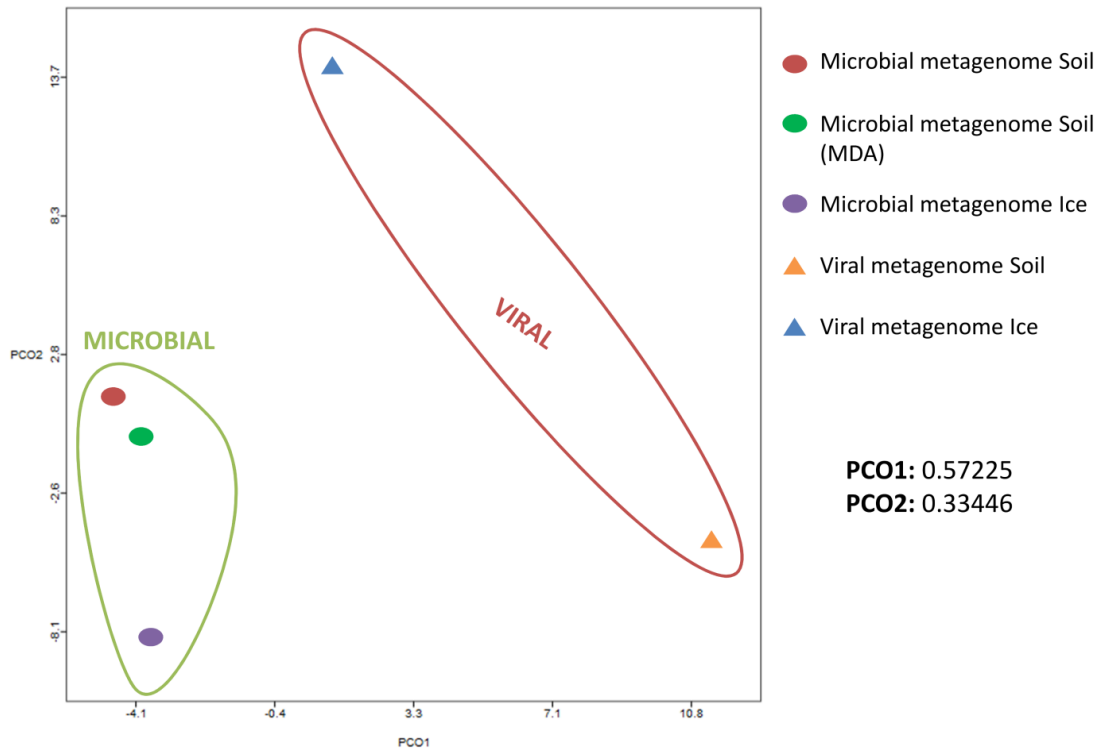


Figure 18. Principal Component Analysis from MG-RAST based on functional abundance.

Given the poor annotation of the viral metagenomes by the MG-RAST pipeline, they were also submitted to the Metavir web server for analysis. Viral sequences were assembled into contigs as described in the Material and Methods section and were also submitted to the Metavir server to see how their assignment compared to the raw reads assignment. One circular contig (429bp) was detected in the ice viral metagenome although it was mapped entirely to hypothetical proteins. The taxonomic annotations of the viral metagenome sequences, both reads and contigs, are presented as a heatmap in Table 5. The ratio of affiliated sequences was 8.88 for raw reads from the ice and 9.33 for soil and 26.39 for contigs from the ice and 22.21 from the soil. There is a higher abundance of single stranded DNA viruses in the ice relative to the soil, especially in the case of *Inoviridae* viruses. Rarefaction curves showed a higher diversity in the soil viral community as compared to the ice viral community. Thus, based on MG-RAST results, both viral and bacterial diversity would be lower in ice than in soil. This was in accordance with the viral diversity calculations that were discussed in the previous section. The Shannon index for ice was 3.36 and 5.88 for soil. Translated into effective number of species, this represents 28.79 viral species in the ice versus 357.8 viral species in the soil. It should be noted that the viral diversity index

for ice is much lower than any of the viral diversity indexes calculated in section 4.1.1. Therefore, arctic glacial ice might have a much lower viral diversity than alpine glaciers or nearby soils.

Table 5. Assignment of the viral reads (a) and viral contigs (b) computed by the server Metavir. Raw reads hits ratios and contigs best hit ratios. Threshold: 50 on Score.

a)

	Ice	Soil
Retro-transcribing viruses	0.23	0
Satellites	0	0.26
dsDNA viruses, no RNA stage	49.13	85
Adenoviridae	0	0.02
Ascoviridae	0.19	0.06
Asfarviridae	0	0.08
Baculoviridae	0.1	0.14
Bicaudaviridae	0.08	0.02
Caudovirales	28.21	61.39
Herpesvirales	0.21	0.43
Iridoviridae	0.53	0.58
Lipothrixviridae	0	0.04
Mimiviridae	3.18	5.18
Nimaviridae	0.02	0.01
Phycodnaviridae	8.98	7.03
Plasmaviridae	0	0.03
Polydnaviridae	0.04	0.04
Polyomaviridae	0.4	0
Poxviridae	0.1	0.6
Rudiviridae	0	0.01
Tectiviridae	0.02	0.05
unclassified phages	2.48	2.84
unclassified viruses	4.61	6.48
ssDNA viruses	49.61	11.91
Circoviridae	0.02	1.57
Geminiviridae	0	0.6
Inoviridae	39.19	0.06
Microviridae	0	5.72
Nanoviridae	0.21	0.32
unclassified viruses	13.19	3.65
ssRNA viruses	0.06	0.11
unassigned viruses	0	0.01
unclassified archaeal viruses	0	0.03
unclassified phages	0.97	2.67
unclassified virophages	0	0.02

b)

	Ice	Soil
dsDNA viruses, no RNA stage	76.42	68.09
Ascoviridae	0.81	0
Bicaudaviridae	0.81	0
Caudovirales	54.47	47.23
Iridoviridae	2.44	0.43
Mimiviridae	1.63	2.98
Phycodnaviridae	8.94	10.21
unclassified phages	1.63	2.98
unclassified viruses	5.69	4.26
ssDNA viruses	20.33	28.94
Circoviridae	0	4.26
Geminiviridae	0	2.13
Inoviridae	18.7	0.43
Microviridae	0	6.81
Nanoviridae	0.81	0.85
unclassified viruses	0.81	14.47
ssRNA viruses	0.81	0.43
unassigned viruses	0	2.13
unclassified phages	2.44	0.43

This metagenomic exploration through web based servers permitted a general view of the data before undertaking more specific tasks and analyzes. The eukaryotic representation in the ice microbial metagenome suggests that glacial ice harbors a complex community with several trophic levels similar to what has been observed in other studies mentioned in section 1.3.4. Nematodes were previously described in several glacier related environments (Vincent et al., 2004; Somerfield et al., 2006; Zawierucha et al., 2014). The sequences assigned to the *Chordata* phyla belonged mainly to the Mammalia class (order *Rodentia*). Although the Svalbard archipelago is thought not to host any small rodents, voles do inhabit it. These rodents are thought to have been accidentally introduced in Svalbard by supply ships from Russia in the first half of the 20th century (Fredga et al., 1990). There are also some differences between the ice and soil ecosystems regarding the bacterial communities in terms of relative abundance. There is a higher abundance of *Actinobacteria* and *Bacteroidetes* in the ice bacterial community similarly as what was seen in alpine glacial ice (Simon et al., 2009), and higher abundance of *Chloroflexi* and *Gemmatimonadetes* in the soil. *Gemmatimonadetes* is a bacterial phylum that is often present in soils, making up an average of 2% of the soil bacterial community (Janssen, 2006). It has also been suggested that they are best adapted to low moisture soils (DeBruyn et al., 2011). Given that the arctic soils in Ny-Alesund are frozen for most part of the year (Roth and Boike, 2001), the water content can be low (Kaštovská et al., 2005) making them a suitable habitat for organisms within this phylum. Indeed *Gemmatimonadetes* made up between 4% and 6% of the arctic soil community, well over the average 2%. These taxonomical differences between the ice and soil microbial communities group the environments apart in the PCA analysis. Yet, concerning function, communities seem to be making use of similar strategies in both environments.

Annotation of viral sequences remains a challenge given the limited size of viral databases. MG-RAST annotation results for the viral metagenomes is a good example of this bottleneck. It points to the need for enlarging the viral databases, but also to the need for alternative methods that would enable us to gather information on the ecology of prokaryotic viruses. The Metavir web server offers us some information on our viral metagenomes, even though it does not solve the problem of the database

size and thus its inferences are based on a scant number of sequences. Viral community distributions from ice and soil based on Metavir classifications seem to be similar with respect to double stranded DNA viruses. There is, however, a higher abundance of single stranded DNA viruses (especially from the *Inoviridae* family) in ice than in soil. This could be a bias due to MDA techniques which have been shown to preferentially amplify single stranded DNA viruses over double stranded DNA viruses (Kim and Bae, 2011), as will be discussed below. Nevertheless, the role of some viruses from this family could be of special relevance, a recent study suggests that they could help the host survive in its environment (sea-ice) (Yu et al., 2014). Web-based independent analyses were also carried out to estimate the viral diversity of the metagenomes. Contrary to what was seen in the analysis of publically available metagenomes, viral diversity was higher in soil than in ice.

A common problem when working with samples from extreme environments and viruses is low DNA yields. MDA techniques provide the means to overcome this problem and enable downstream applications such as sequencing. However, preferential amplification is a limitation of this technique which can induce biases in the results, making quantitative inferences unreliable (Yilmaz et al., 2010). The metagenomic exploration through MG-RAST (best hit classification) facilitated the examination of the possible biases induced by MDA techniques. Soil microbial DNA subject to MDA showed a higher percentage of sequences assigned to viruses than the soil microbial DNA not subjected to MDA, 0.33% and 0.03% respectively. MDA could selectively enhance the amplification of certain viral sequences, increasing their relative abundance in the pool of DNA. Indeed, as mentioned above, MDA techniques have previously been reported to have a preference in the amplification of single stranded DNA viruses versus double stranded DNA viruses (Kim and Bae, 2011). It thus might also be the case that they have a preference of viruses over bacterial sequences. However, no sequences from the ice microbial metagenome were assigned to viruses even though this metagenome was also subjected to MDA. The use of other DNA amplification techniques such as Linker Amplification (LA) (Duhaime et al., 2012), could help us create a more accurate picture of the viral community. The fact that there are more sequences assigned to viruses in the soil than in the ice might also be due to an

increase adherence of the viral particles to soil. Viruses in the ice might have passed through the microbial filter (pore size 0.2 μ m) with more ease while a portion of the viruses in soil might have been attached to the soil and thus trapped in the microbial filter. All of these open questions remain very relevant to the work carried out in this thesis and should be taken into account throughout.

4.2.2 Creating the Network

The impact of prokaryotic viruses on the ecology and evolution of microbial communities is tightly linked to their infectivity potential and their host range (Weinbauer and Rassoulzadegan, 2003; Koskella and Meaden, 2013). In order to address their impact, we need to link viruses to their microbial host(s) in the environment. Data on environmental viruses is scarce and high throughput methods for tracking their interactions with prokaryotes are in demand. Clustered regularly interspaced short palindromic repeats (CRISPRs), which contain viral sequences in bacterial genomes, could be used as a tool to link viruses with their hosts. We therefore designed CRISPR-based identification workflows (explained in detail in the material and methods section) which were used to identify the possible bacterial and viral origin of direct repeats (DRs) and spacers, respectively.

CRISPRs were extracted from the ice and soil metagenomes using the metaCRT software. The CRISPR arrays yielded can be found in Appendix 3. The CRISPRs from the ice (blue arrows) and soil (brown arrows) microbial metagenomes were used to define possible bacterial hosts (rectangles representing the DRs from the bacteria) and the infecting viruses (triangles) (Figure 19). While this figure depicts the interactions that could be taking place in the environment, this network is a simplified diagram of arctic viral-host interactions, as each DR could be common to several bacteria and every spacer could be common to different viruses. So while the network provides the possible connection between DRs and spacers from ice and soil, the number of putative bacteria and viruses involved is unknown.

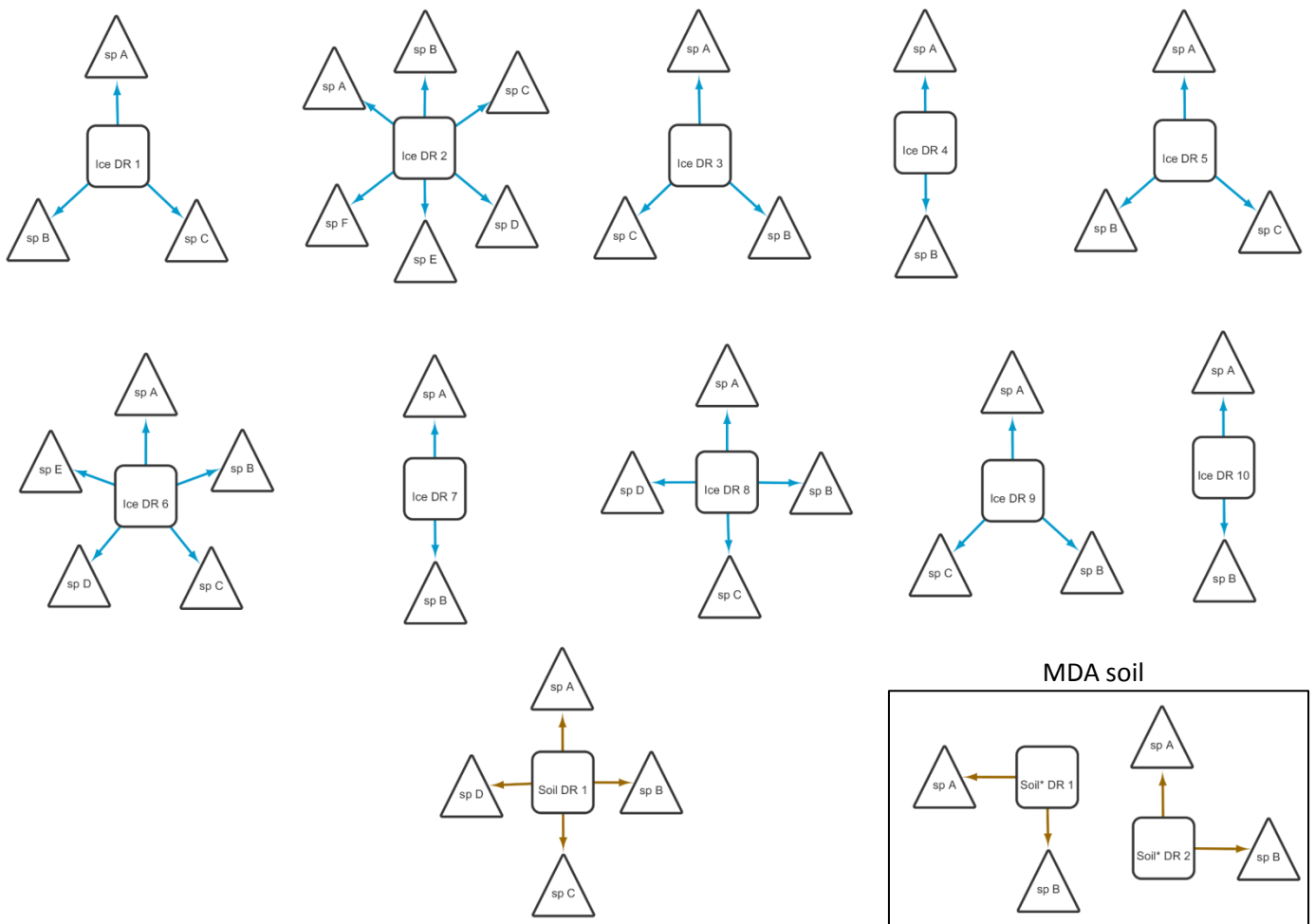


Figure 19. Virus-host networks inferred from CRISPRs recovered from ice (blue arrows) and soil (brown arrows). Triangles represent the viruses (spacers) and rectangles the microbial host (DRs). Networks in the lower right corner were inferred from CRISPRs recovered from arctic soil subjected to MDA.

Since CRISPRs contain sequences derived from bacteria and archaea (in the direct repeats) (Jansen et al., 2002) and spacers from viruses, we constructed the networks between DRs and spacers based on the CRISPRs from the different metagenome datasets that we sequenced from arctic glacial ice and soil (Figure 19). With similar quantity of metagenomic sequences from ice and soil (42078 reads for soil and 47626 reads for ice), we found more CRISPRs in ice than in soil and all CRISPR sequences found were unique. To take into account the possible influence of MDA (Multiple Displacement Amplification) on the number of CRISPRs found, a control (soil DNA subject to MDA) was evaluated along with the other samples. Two CRISPRs were found in the MDA control compared to the one CRISPR found in the soil (not subjected to

MDA), so the MDA did not seem to make a significant impact in the number of CRISPRs detected. As was discussed in the 4.1 section CRISPRs abundances can give us information on viral dynamics and host ranges in an environment. We hypothesize that a higher number of unique CRISPR sequences in a metagenome might represent higher viral diversity and interactions with their hosts through the infection of different microorganisms. Thus, the relatively more abundant unique CRISPRs in ice might be an indication of the high viral diversity, infectivity rates and broad host ranges as suggested in the literature (Anesio et al., 2007; Sävström, Granéli, et al., 2007; López-Bueno et al., 2009). However, this would be in disagreement with the viral diversity indexes calculates in section 4.2.1, where viral diversity was higher in soil than in ice. The low number of CRISPRs found in soil might also be due to a limited depth of coverage in the more diverse soil metagenome dataset. Due to the use of MDA techniques to create the metagenomes, these viral diversity indexes might not be accurate. CRISPRs from glacial ice had a range of 2 to 6 spacers with a mean of 3 spacers, but ascribing an ecological significance to the number of spacers in each CRISPR could be unreliable because the main driver behind the number of CRISPRs would probably be read length. CRISPRs from the soil metagenomes did not have any duplicate spacers while CRISPRs from the ice metagenome had one duplicate spacer. However, this duplicate spacer belonged to the same CRISPR, therefore no conclusions on host ranges can be drawn from this result. In any case, ice appears to be a good model ecosystem for this initial CRISPR application study.

One of the objectives of the thesis was to demonstrate the potential for using CRISPRs as a tool to link viruses with their hosts in environmental samples. Given that this demonstration is a proof of concept, we were quite stringent in the workflows that we designed to accomplish this goal. In both workflows (Figures 8 and 9 in section 3.4.9) we restrained the possible bacterial and archaeal hosts to those that had their genomes fully sequenced and reported in the NCBI database and to the viruses represented in the NCBI viral sequences. These prokaryotic genomes were used to identify potential originators of the direct repeats (DRs) from the CRISPRs found in the ice and soil metagenomes, but only if similar 16S rRNA genes were also identified in

the pyrotag libraries obtained for each environmental sample, which was another stringency parameter in both of our workflows.

The viral identification procedure was the same for both workflows. Its first step (a blast of the spacers against a viral database) resulted in a selection of 406 and 283 viral sequences as putative sources of the ice and soil spacers, respectively, out of the 12.954 nucleotide sequences that the database contained. Finally, the blast of the viral metagenomes against the putative sources of the spacers left us with 13 different phages identified as sources of the ice spacers and 84 different phages identified as sources of the soil spacers. On the other hand, the number of bacterial genomes identified as sources of the CRISPRs by the first workflow was 615 for the ice data and 449 for the soil data out of the 5193 full genomes in the database. As for the second and further curated workflow, the initial blast step issued 638 bacterial genomes as putative sources of the ice CRISPRs and 469 bacterial genomes as putative sources of the soil CRISPRs. None of the putative soil bacteria matched the 16S rRNA sequences in our pyrotag data from the arctic soil, while 47 genomes had hits to the 16S rRNA genes from the ice.

The second workflow was designed to further curate the results obtained by the first workflow, thus another stringency parameter was added as an intermediate step. Bacterial genomes, identified as potential sources of CRISPRs from ice and soil, were analyzed to see if they had CRISPRs. Respectively, 75% and 64% of these genomes were found to host CRISPRs within them. However, only 19% and 10% of the remaining genomes were kept as putative sources of the CRISPRs after blasting the DRs issued of the arctic ice and soil metagenomes against the DRs found in the candidate bacterial genomes. This added requirement in the second workflow significantly reduced the number of putative bacterial genomes identified as sources of the CRISPRs. This could mean that the DRs in the initial blast against the prokaryotic genomes database might not be matching DR sequences only but also other areas and repeated regions of the genomes. Nevertheless, it could also be due to the metaCRT software not finding all the CRISPRs present in the bacterial genomes database. In the first scenario, the first blast would introduce noise into the search. In the second scenario, we would be losing information by requiring that metaCRT finds the CRISPRs in the genomes of the

database. Both scenarios are probable and might actually be taking place at the same time in our search, there will be hits against other areas of the genomes which are not DRs and metaCRT will fail to identify some of the CRISPRs present in the genomes. However, deleting the first step of the workflows would entail the use of a DRs database instead and would therefore limit the identification to organisms with already described CRISPRs. The CRISPRdb (Grissa et al., 2007b) contains 1302 strains (779 species) having “convincing CRISPR structures” in August 2015. The database is continually being updated and its size has increased since the beginning of this thesis, however it still remains quite limited. Blasting against the prokaryotic genomes database might introduce false results, as was pointed out before, but by searching the selected genomes for CRISPRs with a software like metaCRT and re-blasting the arctic DRs against the DRs found in the CRISPRs of the prokaryotic genomes we curated the results in addition to identifying the source of the DRs even if their CRISPRs haven’t been described before.

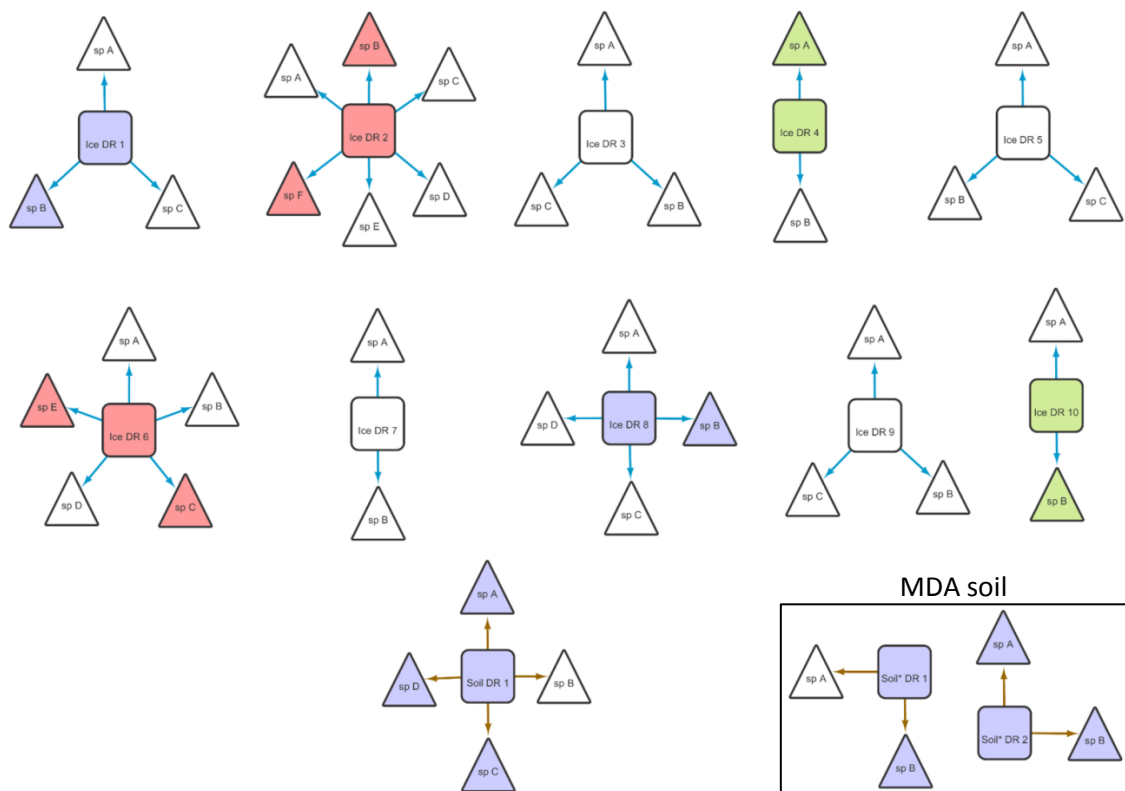


Figure 20. Virus-host networks inferred from CRISPRs recovered from ice (blue arrows) and soil (brown arrows). Triangles represent the viruses (spacers) and rectangles the microbial host (DRs). Networks in the lower right corner were inferred from CRISPRs recovered from arctic soil subjected to MDA. Shapes were shaded red when both workflows yielded putative identities, green when only workflow one did and grey when only workflow two did.

For those networks where the first workflow (figure 8) yielded a potential identity for the DR and at least one of the spacers, the nodes were shaded in blue in Figure 20. If the identities were yielded by the second curated workflow (figure 9) the nodes were shaded in green. If identities were proposed both by the first and second workflow the nodes were shaded red. For further analysis, these sub-sets of networks were recreated with their possible bacterial and viral identities. Figure 21 shows the sub-set of networks where identities were yielded through the use of the first workflow, while Figure 22 shows the sub-set of networks where identities were yielded by the curated workflow. In both of them, circles represent the microbial host, while hexagons are the microbial viruses. In Figure 22 the roman numerals within the circles relate to the legends above the network, where the putative bacteria identified with the DRs are listed. The lower case letters within the hexagons relate to the legends below the network, where the putative viruses identified with the spacers are listed. For Figure 21 the putative bacteria identified with the DRs and the putative viruses identified with the spacers are listed in Table 6. The first and second workflow shared the same process for the viral identification of the spacers. Yet the results were transferred to the networks only when the workflow offered identities for the DRs and at least one of the spacers of a same CRISPR. Only DR 2 and DR 6 were identified by both workflows as can be seen in Figure 20, and thus only their spacers share the same identifications. The stringency parameters implanted in the workflows reduced the number of CRISPR networks from ten CRISPRs from the ice and one or two from the soil (Figure 19) to four CRISPRs in ice and none to two in soil Figure 21 and Figure 22.

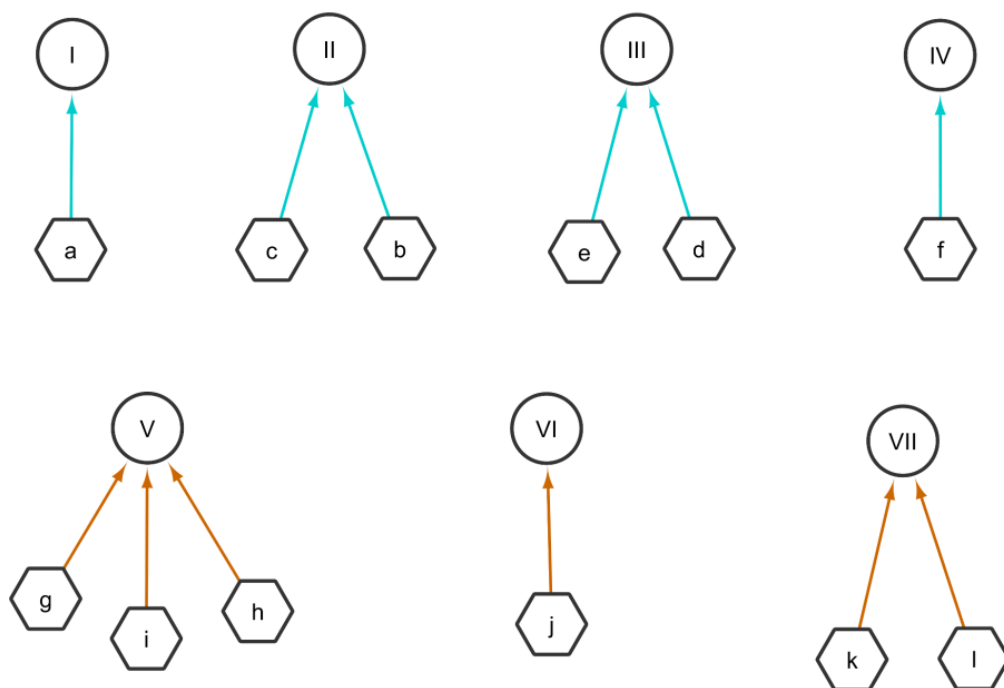


Figure 21. Networks with both the origin of the DR and at least one related spacer were putatively identified by the first workflow. Networks with blue arrows were issued from the analysis of the ice metagenome while networks with brown arrows were issued from the analysis of the soil metagenome. The identified prokaryotes (roman numerals) and the identified viruses (letters) are shown in Table 6.

Table 6. The identified prokaryotes (roman numerals) and the identified viruses (letters) from Figure 21.

	Genera	Phylum		Virus
I	Bacteroides	<i>Cytophaga-Flavobacteria-Bacteroides</i> (CFB)	a	Environmental Halophage
	Riemerella			
	Ornithobacterium			
	Parabacteroides			
	Mannheimia			
II	Bacteroides	<i>Cytophaga-Flavobacteria-Bacteroides</i> (CFB)	b	Synechococcus phage
	Ornithobacterium			
	Riemerella		c	Synechococcus phage
	Mannheimia			Cyanophage
III	Halothermothrix	<i>Firmicutes</i>	d	Cellulophaga phage
	Thermoanaerobacter			Pseudomonas phage
	Persephonella		e	Mycobacterium phage
	Xylella			
IV	Cytophaga	<i>Cytophaga-Flavobacteria-Bacteroides</i> (CFB)	f	Pseudomonas phage
	Kosmotoga			

	Genera			Virus	
V	Agrobacterium	Modestobacter	g	Edwardsiella phage	
	Alkalilimnicola	Mycobacterium		Enterobacteria phage	
	Amycolatopsis	Polymorphum			
	Bradyrhizobium	Pseudomonas			
	Burkholderia	Pseudoxanthomonas			
	Catenulispora	Pyrobaculum			
	Caulobacter	Ralstonia			
	Cupriavidus	Rhodanobacter			
	Desulfobacca	Rhodopseudomonas			
	Desulfovibrio	Rhodothermus	h		Brevibacillus phage
	Enterococcus	Saccharopolyspora			
	Frankia	Solibacter			
	Geobacter	Sphingobium			
	Gluconobacter	Starkeya			
	Herminiimonas	Streptomyces			
	Hyphomicrobium	Streptosporangium			
	Kineococcus	Synechococcus		i	Burkholderia cepacia phage
	Kribbella	Thiobacillus			
	Mesorhizobium	Thioflavicoccus			
Methylocystis	Tistrella				
Micromonospora	Vibrio	Brucella phage			
VI	Bacillus	Helicobacter	j	Mycobacterium phage	
	Colwellia	Hydrogenobacter		Caulobacter phage	
	Cyanothece	Mycoplasma		Prochlorococcus phage	
	Enterococcus	Nostoc			
	Fervidobacterium	Staphylococcus			
VII	Acinetobacter		k	Salmonella phage	
	Agrobacterium			Enterobacteria phage	
	Candidatus Korarchaeum			Alteromonas phage	
	Chloroflexus			Ralstonia phage RSK1 DNA	
	Helicobacter			Pseudomonas phage	
	Rhodothermus			Escherichia phage	
	Serratia			l	Mycobacterium phage
	Shewanella				Escherichia phage
	Syntrophobotulus				Salmonella phage

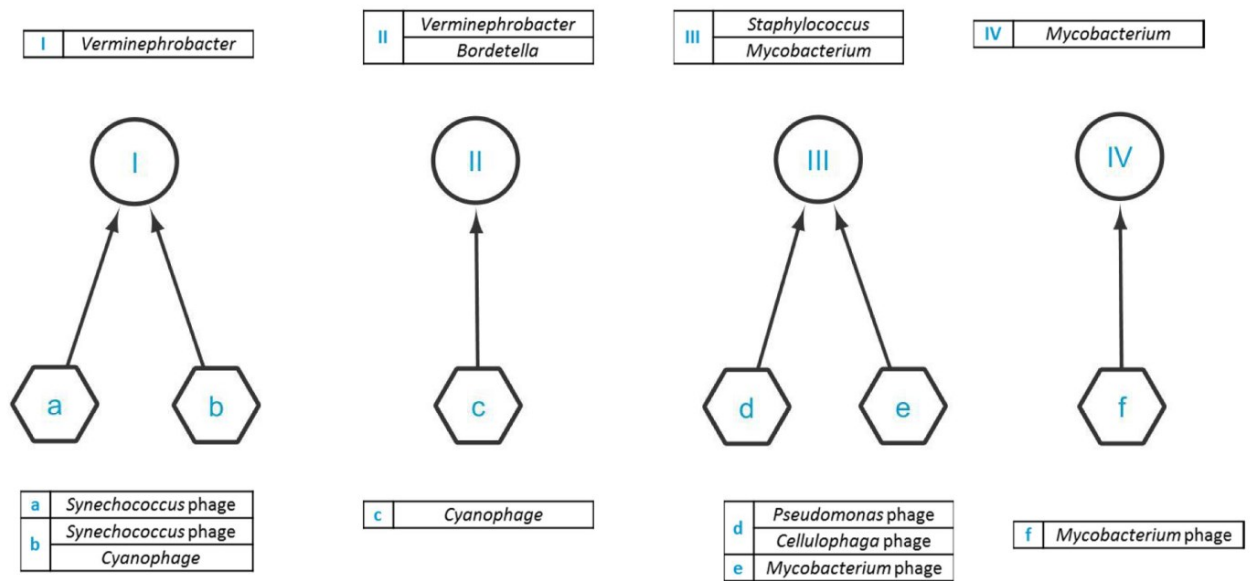


Figure 22. Networks with both the origin of the DR and at least one related spacer were putatively identified by the second workflow. All networks were issued from the analysis of the ice metagenome. Above and below each network, legends show the identified prokaryotes (roman numerals) and the identified viruses (letters).

Because of the less stringent procedure, the first workflow yielded between five to two possible bacterial genera as sources for each of the ice DRs, while for the soil, it yielded up to 42 different genera as possible sources of the DRs. The increased number of proposed genera for the soil DRs might be a consequence of the databases that have more representatives from this ecosystem as opposed to ice. Because of the long list of possible identities for the soil microbial hosts (V, VI and VI in Figure 21), they will not be commented here. As for the microbial identities proposed for the DRs from the ice CRISPRs, it is worth noting that several of the genera belong to the phylum *Cytophaga-Flavobacteria-Bacteroides* (CFB) and to the *Firmicutes* phylum, which are commonly found in glacial ice and snow samples (Table 6) (Miteva and Brenchley, 2005; Amato et al., 2007; Bidle et al., 2007; Zhang et al., 2008; Y. X. Zeng et al., 2013). However, the identified genera from the Firmicutes phylum (microbial hosts for III) were *Halothermothrix* and *Thermoanaerobacter* whose members have usually been described as thermophiles, although *Thermoanaerobacter* has been detected before in psychrophilic environments (Gaidos et al., 2009). The genus *Mannheimia*, possible microbial host for I and II, is found in animal hosts, some of which inhabit arctic regions such as polar bears and reindeers (Fisk et al., 2005; Josefsen et al., 2007). Other

identified genera have rarely been detected in polar environments, for example *Persephonella* and *Xylella*, which were identified as possible microbial hosts for III. *Xylella* is a phytopathogen mostly described in temperate environments (Simpson et al., 2000), while *Persephonella* members are deep-sea dwellers (Gotz et al., 2002). The *Kosmotoga* genera, possible identity for the microbial host IV, has been seen to tolerate a broad range of temperatures, but these are in the thermophilic and mesophilic spectra at the moment (DiPippo et al., 2009; Nesbø et al., 2010). However, the small size of viral nucleotide databases limit identification to a few, better studied viruses. At the same time, the less stringent process used to identify the microbial hosts might also be yielding inaccurate results.

The higher stringency level applied to the second workflow resulted in a more restricted list of identities for the microbial hosts with one or two bacterial genera per DR. The genus *Verminophrobacter* was identified as a possible microbial host in I and II (Figure 22). In the case of II, the microbial host could also be *Bordetella*. However, viruses infecting I and II were identified to be *Cyanobacterial* phages. While three *Bordetella* phage genomes were present in the prokaryotic viruses' database, there were no known *Verminophrobacter* phages. Thus, viruses a, b and c might be *Verminophrobacter* phages, although their closest match in the viral database are *Cyanobacteria* phages. Microbial host III was identified as a member of either *Staphylococcus* or *Mycobacterium*. Microbial host IV was also identified as a *Mycobacterium* and viruses e and f were identified as *Mycobacterium* phages. Phage d was assigned to either *Cellulophaga* phage or *Pseudomonas* phage. Further interrogation of the ice metagenomes through MG-RAST (Meyer et al., 2008) confirmed the assignment of reads to all identified hosts. Likewise, viruses identified from the spacers were represented in the viral ice metagenomes based on Metavir's (Roux et al., 2011) contig and/or read assignments, supporting data output from the CRISPR network system.

The analysis of ice CRISPRs shed some light on the communities inhabiting this environment and their possible viral relationships. *Verminophrobacter* (possible host I and II in Figure 22) and *Bordetella* (possible host II in Figure 22), have been described as a nephridia symbionts (Davidson et al., 2013), suggesting the presence of annelids in

the ice. No functional data for annelids was present in our metagenome dataset based on analysis with MG-RAST. However, given that samples were pre-filtered to remove debris, it is possible that we lost these in the process. Presence of ice worms has been confirmed in a range of glaciers (Hartzell et al., 2005) as was discussed in the introduction and representatives of *Verminephrobacter* were also found in Lake Vostoc ice (Shtarkman et al., 2013). Recently, a study described the presence of bacterial microbiota in ice worms (Murakami et al., 2015). *Bordetella* (possible host II) was also found to be present in ice cores from Pony Lake, Antarctica (Foreman et al., 2011). Although *Staphylococcus* and *Mycobacterium* (possible host III and possible host III and IV, respectively) are genera often associated with humans, they are ubiquitous and have been found inhabiting a wide range of environments (Kloos, 1980; Falkinham, 2002), including glacial ice (Christner et al., 2000; Miteva et al., 2004; Xiang et al., 2005). When looking at the taxonomic assignment of the 16S rRNA gene sequences, only *Mycobacterium* was present at the genus level (0.14% of assigned reads). No other bacteria identified through our network were represented in the taxonomic classification of the 16S rRNA gene sequences, which could point to a low abundance of these bacteria in their environment. It could also mean that these bacteria are the closest members present in the database, but that the bacteria actually inhabiting the ice do not have representatives in the 16S rRNA gene database, thus limiting the taxonomic assignment. With respect to the viruses identified in our network, all of them, with the exception of the *Cellulophaga* phage, were described by Cottrell and Kirchman, 2012 in their analysis of viral genes present in arctic bacteria. This data supports the assumption that these viruses are infecting arctic bacteria as indicated by our results. In addition, *Cellulophaga* phages seem to be ubiquitous in aquatic environments based on the study by Holmfeldt et al., 2013.

Both workflows proposed identifications for DRs 2 and 6 (Figure 20), this offers an opportunity to see how both methods compare. At the genera level these identifications do not share common points. However, at the phylum level, both workflows suggested *Proteobacteria* as a possible identity for DR 2 and both suggested *Firmicutes* as a possible identity for DR 6. Identities for the spacers related to these DRs are identical given that the viral identification process for both networks was the

same. The results of the second workflow, as compared to those of the first workflow, not only seem more congruent at the level of viral–host pairing, but they are also consistent with the different analysis of the metagenomic datasets and existing literature. Results of the first workflow are only similar at the phylum level for the microbial identities. Therefore, the added requirement in the second workflow that makes sure that the DRs were not matching other areas of the genomes seems to have strengthened the accuracy of the microbial identifications. Further stringency would probably increase the robustness of the analysis; however it would also result in the loss of information, as was mentioned above. The stringency parameters applied in the data interpretation have already eliminated possible bacterial/viral interactions when both the bacteria and the virus were unknown or the bacterial presence was not confirmed by 16S rRNA sequencing. The implementation of higher stringency parameters would therefore lead to more data loss. A good example is the use of DR databases as was discussed before. The blast of the arctic ice and soil DRs from the 13 different CRISPRs against the CRISPRdb (Grissa et al., 2007a) only yielded two hits, both from the same CRISPR, (maximum *e*-value 10, lower values yielded no hits) to the same organism *Taylorella equigenitalis*, a horse pathogen. The alignment was very poor and the identification would have probably not succeeded to pass the subsequent steps of the workflow. However, as part of the method validation, the CRISPRdb was blasted against the prokaryotic genomes database. For an *e*-value of 10^{-5} or lower the correct identity would be among the hits. The limited size of the DR database and the underrepresentation of environmental sequences make it unsuitable at the moment for the identification of DRs sourced from environmental CRISPRs, especially when working with data from understudied extreme environments. A compromise between stringency and data loss was therefore made in the development of the workflows.

CRISPR related databases and viral databases among many others will undoubtedly become larger in the following years with the increase of high throughput sequencing of environmental samples. However, DRs from soil CRISPRs proved to be more difficult to identify than the DRs from ice CRISPRs. A possible explanation is that databases have more representatives from soil ecosystems and the identification becomes

harder. In this case bigger databases might require the implementation of higher stringency steps into the workflow in order to identify viruses and hosts through CRISPRs. CRISPRs are also quite abundant in Archaea, but the number of fully sequenced archaeal genomes is considerably smaller than for bacteria. In a similar way, deeper microbial sequencing could also increase the number of CRISPR sequences found in metagenomes and make the identification of the microbial originator of the CRISPR a more likely occurrence in environments with high microbial diversity. On the other hand, our CRISPR-based network discovered viral/microbial interactions among unknown viruses and Bacteria/Archaea. The further identification of these microorganisms would help clarify the extent of these interactions. Despite working with small-sized databases and a relatively unexplored ecosystem whose inhabitants may not be well represented by the fully sequenced genomes in the database, we were able to describe several viral-host interactions taking place in the arctic glacial ice through the analysis of CRISPRs obtained from metagenomic data.

4.3 Transduction Events

If icy environments have communities of diverse viruses, with possible broad host ranges and even high infectivity, transduction could be a major driver of adaptation in these environments. However, not only the rates of transduction are relevant for bacterial adaptation through this mechanism, but also the content of the transducing material. In order to learn about the DNA that could be being transduced in the ice and its adaptive significance for the recipient cell a series of analysis were carried out.

In specialized transduction, the virus encapsidizes its own genome with part of the host microbial genome (Morse et al., 1956; Frost et al., 2005). If these events had taken place in the microbial communities we sampled in the Arctic, we would expect to find viral genomes with fragments of microbial DNA. In order to look for traces of specialized transduction events in our viral metagenomes, we selected a set of genes that we thought could be of adaptive interest to the Arctic microbial community. As the arctic ice and soil are subjected to low temperatures and mercury from the snowpack (Larose, Prestat, et al., 2013), we selected cold shock genes and mercury resistance genes as traits that might be an asset for microorganisms living in the Arctic.

Additionally, we selected a set of genes, such as efflux pumps genes, that might provide an adaptive advantage regardless of the ecosystem. We browsed our unassembled and assembled viral metagenomes for hybrid reads or contigs and subsequently identified the transduction agent through the viral part of the sequence. The viruses that could be potential transducing agents for the three selected genes based on the analysis of the viral reads are presented in Table 7 a, b and c. The viruses that could be potential transducing agents for the three selected genes based on the analysis of the viral contigs are presented in Tables 8 a, b and c.

Tables 7. Reads from the ice and soil metagenomes with both microbial and viral sequences. **a.** Reads with cold shock protein sequences and their viral assignment. **b.** Reads with *mer* genes' protein sequences and their viral assignment. **c.** Reads with efflux pumps' protein sequences and their viral assignment. * (viral accession numbers identified as transduction agents).

a.

COLD SHOCK PROTEINS in ice		COLD SHOCK PROTEINS in soil	
GenBank accession	Definition	GenBank accession	Definition
GU396103.1	Aeromonas phage	KF296717.1	Bacillus phage
DQ115859.1	Cyanobacteria phage		

b.

<i>mer</i> GENE PROTEINS in ice		<i>mer</i> GENE PROTEINS in soil	
GenBank accession	Definition	GenBank accession	Definition
KF887906.1	Ralstonia phage	AY349011.3*	Burkholderia cepacia phage
AB434711.1	Ralstonia phage	JX163858.1*	Caulobacter phage
AB259123.2	Ralstonia phage	JX100814.1	Caulobacter phage
		JX100813.1*	Caulobacter phage
		JX100812.1*	Caulobacter phage
DQ115859.1	Cyanobacteria phage	JX100811.1*	Caulobacter phage
		JX100809.1*	Caulobacter phage

c.

EFFLUX PUMPS PROTEINS in ice		EFFLUX PUMPS PROTEINS in soil	
GenBank accession	Definition	GenBank accession	Definition
AB828698.1	Ralstonia phage	HQ633071.1	Synechococcus phage
KF887906.1	Ralstonia phage		
AB434711.1	Ralstonia phage		
AB259123.2	Ralstonia phage		
GU911303.1	Burkholderia phage		
J01735.1	Enterobacteria phage	KF296717.1	Bacillus phage
AF234172.1	Enterobacteria phage		
AF234173.1	Enterobacteria phage		
DQ115859.1	Cyanobacteria phage		

Tables 8. Contigs from the ice and soil metagenomes with both microbial and viral sequences. **a.** Contigs with cold shock protein sequences and their viral assignment. **b.** Contigs with *mer* genes' protein sequences and their viral assignment. **c.** Contigs with efflux pumps' protein sequences and their viral assignment.

a.

COLD SHOCK PROTEINS		
Contig	GenBank accession	Definition
contig00022	AB259123.2	<i>Ralstonia</i> phage
contig00022	KF887906.1	<i>Ralstonia</i> phage
contig00022	AB434711.1	<i>Ralstonia</i> phage
contig00006	AB276040.1	<i>Ralstonia</i> phage
contig00500	GU396103.1	<i>Aeromonas</i> phage
contig00157	GU396103.1	<i>Aeromonas</i> phage

b.

<i>mer</i> GENE PROTEINS		
Contig	GenBank accession	Definition
contig00022	AB259123.2	<i>Ralstonia</i> phage
contig00022	KF887906.1	<i>Ralstonia</i> phage
contig00022	AB434711.1	<i>Ralstonia</i> phage

c.

EFFLUX PUMPS PROTEINS		
Contig	GenBank accession	Definition
contig00006	AB276040.1	<i>Ralstonia</i> phage
contig00007	JN662425.1	<i>Burkholderia</i> phage
contig00007	FJ937737.2	<i>Burkholderia</i> phage
contig00007	JX104231.1	<i>Burkholderia</i> phage
contig00022	AB259123.2	<i>Ralstonia</i> phage
contig00022	KF887906.1	<i>Ralstonia</i> phage
contig00022	AB434711.1	<i>Ralstonia</i> phage
contig00035	AB828698.1	<i>Ralstonia</i> phage
contig00035	GU911303.1	<i>Burkholderia</i> phage
contig00035	J01735.1	Bacteriophage lambda
contig00035	AF234172.1	<i>Enterobacteria</i> phage
contig00035	AF234173.1	<i>Enterobacteria</i> phage
contig00157	GU396103.1	<i>Aeromonas</i> phage

The potential transduction agents in the ice identified through the analysis of the reads were highly similar to those identified through the analysis of the contigs. A total of five phages had putative viral and microbial markers matching those in the microbial and the viral databases. *Cyanobacteria* phage, which was identified through the read analysis as a transduction agent for all three genes was however missing in the contigs analysis. All other four phages, *Aeromonas* phage, *Ralstonia* phage, *Burkholderia* phage and *Enterobacteria* phage were present in both analysis and more or less for the same set of genes. *Ralstonia* phage was identified by the contigs analysis as a transducing agent for all three microbial proteins studied here. The reads analysis identified it as a potential transduction agent of mercury resistance genes and efflux pumps genes. No soil contigs contained both a viral and a microbial marker based on our stringency parameters (see section 3.4.10). However, this was not the case for the soil reads were four phages were identified as putative transduction agents. Several of these phages were present in the CRISPR networks presented above. Red stars in Table 7 b mark the viral accession numbers which had been identified in this study as sources of the CRISPRs and as transduction agents. We were therefore able to map the transduction results onto our infection network (Figure 23). This means that the *Burkholderia* (i in Figure 23) and *Caulobacter* phages (j in Figure 23) could be acting as transduction agents in soil, potentially exchanging mercury resistance genes among the different bacteria identified in the CRISPR network (V and VI in Table 6). Although hybrid reads were detected, no hybrid contig was found in the soil virome while 6 hybrid contigs were found in the ice virome. Given that the overall number of viral ice contigs was roughly a third of the contigs from the soil, transduction in ice might be more predominant than in the soil, where other mechanisms of gene transfer might compete with transduction.

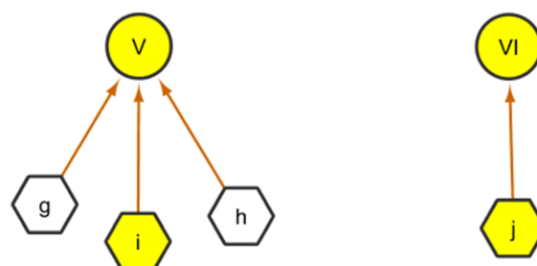


Figure 23. Networks extracted from Figure 21, shapes shaded yellow stand for the potential transduction agents and their hosts.

Our interest in linking phages and their host in the environment originated from the hypothesis that transduction has a significant role in microbial adaptation. We therefore tried to identify molecular tracers of transduction in our viral metagenomes. We found hybrid reads and contigs in our virome that had both viral and microbial sequences in the same sequence. Based on this result, up to five different phages in arctic glacial ice appear to be carrying pieces of various microbial genes that could be (or have been) transferred to their host(s). *Ralstonia* phages were frequent candidates in our ice viromes. Individual *Ralstonia* phages appear to be capable of infecting several different strains of *Ralstonia solanacearum* (Kawasaki et al., 2009) and possibly spreading important environmental genes throughout the population. Moreover, *Cupriavidus metallidurans*, formerly *Ralstonia metallidurans* (Vandamme and Coenye, 2004), is a bacteria known to resist a range of different metals including mercury by carrying the genes responsible for this resistance in a *mer* operon (von Rozycki and Nies, 2009). Given that we could find *Cupriavidus metallidurans* in our MG-RAST functional assignment data (Meyer et al., 2008), *Ralstonia* phages found in this study as transduction agents of mercury resistance genes might have *Cupriavidus metallidurans* as a host.

The potential impact of *Ralstonia* phages in glacial ice was further evaluated by blasting a set of proteins from the phage against the assembled microbial metagenome from ice. Proteins from the *Ralstonia* phage matched 13 different contigs (maximum *e*-value 9×10^{-10}). Based on these results, 3.4% of the assembled bacterial sequences from ice carry coding DNA from *Ralstonia* phages. However, all the *Ralstonia* phages identified as transduction agents belong to the *Inoviridae* family (ss DNA viruses), with the exception of AB276040.1 which belongs to the *Myoviridae* family (ds DNA viruses). As was seen in the 4.2.1 section MDA preferentially amplifies single stranded DNA viruses over double stranded DNA viruses. Therefore the repeated appearance of this phage in our analysis might be due to an over representation of its sequences in the metagenomic pool. The 13 contigs were then blasted against the prokaryote genomes database in order to identify them. Only contigs 6 and 7 had hits against the prokaryotes genomes database with *e*-values less than 9×10^{-5} . With 96% identity over 86 base pairs (*e*-value 1×10^{-25}), contig 6 was

assigned to *Brachy bacterium faecium*. *Brachy bacterium faecium* has been previously described in arctic sea-ice from Spitzbergen (Groudieva et al., 2004) and in Vostok accretion ice (Christner et al., 2005). Contig 7 was assigned to *Kocuria rhizophila* with 84% identity over 243 base pairs (e-value 9×10^{-25}). Members of the *Kocuria* genus have been detected in ice cores from glaciers and ice sheets all around the world (Miteva et al., 2004; Xiang et al., 2005; Zhang et al., 2008, 2010). Moreover, *Kocuria rhizophila* was detected in the sediments of a subglacial lake in Antarctica (Pearce et al., 2013).

These analyses enabled us to detect transduction candidates which could be playing an important role in their environment. This analysis could be carried out prior to the design of laboratory experiments to test for transduction in the environment. Moreover, the results yielded in this section can be combined with the infection network created through CRISPR sequences. In this fashion, not only do we gain information on phages that could be exchanging bacterial genes in the ice, but also which bacteria could be the receptors of these genes. Mapping the identification of transduction agents with infection networks could help us understand the magnitude of this phenomenon and its role on bacterial adaptation.

Overall, the results point to the presence of transduction events in glacial ice and maybe also in arctic soil. More interestingly, these transduction events would be promoting the exchange of genes of potential interest to the microbial community. *Ralstonia* phages seem to be transducing agents for the three genes we selected, and thus could be shaping microbial adaptation to toxic solutes in veins and cold temperatures. However, *Ralstonia* might have been detected in this study because of an over representation of its reads in the viral metagenome. Other phages could be leading similar processes and have not been detected because of a lower metagenomic representation or because they are transducing other sets of genes.

5. General Discussion

The main hypothesis on which this thesis is based is that viruses can drive microbial adaptation and evolution in the environment. The work done to test this hypothesis in our model environment, glacial ice, indicated viral diversity to be low and no conclusions could be drawn on host ranges. However, the number of CRISPRs in the ice metagenome was high compared to nearby soil, which points to increased microbial defenses against phage infection. The reason for this higher CRISPR abundance in ice could be higher infection rates, as have been described in the literature for other cold environments (S awstr om, Gran eli, et al., 2007). The CRISPR-based infection networks from the ice metagenome pointed to a complex environment where different niches were exploited by the microbial and viral community, such as annelids' nephridia. Phages infecting microbial symbionts in animals are believed to control the microbial population providing innate immunity to the animal (Carda-Diequez et al., 2015) and in thus, a perdurable niche for the microbial community. Moreover, transduction agents seem to be present in glacial ice, providing exchange of genetic material among cells. The nature of this genetic material supports the idea of these events favoring microbial adaptation through viruses. Therefore, the results yielded by this thesis are consistent with viruses being drivers of microbial adaptation, especially in cold environments.

Hereafter, the different aspects and bottlenecks encountered during the work of this thesis will be discussed, as well as the main conclusions drawn.

5.1 Viral Populations in Environments with a Range of Temperatures

The first objective of this thesis was to analyze if there was any evidence suggesting that environments with different temperatures hosted different sorts of viral communities. We thus analyzed 30 publicly available metagenomes from 10 different environments subjected to various temperatures and calculated Shannon indexes to compare viral diversity among them. The results pointed to a correlation between temperature and viral diversity, with cold environments having higher diversity than temperate environments. We must take into account that the number of metagenomes analyzed might not be big enough for the results to be reliable, for example, we are missing data from the hyperthermophilic spectra which could change

the conclusions drawn from the results. It could be that a higher viral diversity is a hallmark of extreme environments (Rice et al., 2001; Rachel et al., 2002), rather than just of cold ecosystems. However, what we can see is that glacial ice has similar Shannon indexes for viral diversity to those of freshwater lakes (Albufera), or marine systems (Puget Sound or Baltic sea), in accordance with what has been seen so far in the literature (López-Bueno et al., 2009; Bellas and Anesio, 2013; Senčilo et al., 2014). These results suggest that cold environments can be as rich as environments which have traditionally been considered highly diverse from the viral perspective (Suttle, 2005; Angly et al., 2006; Breitbart, 2012).

We also analyzed the CRISPRs found in these 30 metagenomes to see if the results we obtained matched those derived from traditional sequence analysis. CRISPRs could be a tool to approach ecological questions such as viral diversity, host ranges or viral infectivity. Spacers, as markers of past infections could be a reflection of the viral population in a given environment (He and Deem, 2010). Assessing the number of different spacers present in a metagenome might be a measure of the different active viruses. It should be noted that several different spacers could actually match one sole phage (Paez-Espino et al., 2013) and this could make viral diversity inferences through spacers unreliable. Nevertheless, Baltic sea water, Antarctic lake water and glacial ice were (in decreasing order) the environments with higher numbers of spacer clusters in accordance with the previous results for viral diversity. Baltic sea ice and glacial ice showed high viral diversities based on the Shannon index comparisons and one of the three metagenomes from Antarctic lake water also yielded a very high Shannon index. The number of different spacer sequences found in a metagenome could therefore be an indicator for viral diversity. Moreover, to be represented among the spacers, the viruses must have infected a microbial cell hosting a CRISPR system. Hence, conclusions drawn from spacer analysis could be a measure of the diversity of active viruses rather than the diversity of the total viral population. One study proposed that spacer diversity could be correlated to CRISPR activity, *i.e.* the ability to integrate new spacers in response to exposure to foreign DNA (Horvath et al., 2008). From our point of view, the rate of exposure to foreign DNA is partly determined by the rate of

infection by different viruses for which a new spacer is needed. Hence, the higher the diversity of infective viruses, the more active the CRISPR systems need to be.

If spacers represent the portion of the viral population which is infectively active, they could also help us address viral specificity ranges. If a given spacer is shared by several different cells it could point to the presence of a virus with broad host ranges. On the other hand, several studies have noted a “trailer-end conservation” among the spacers across entire populations (*i.e.* spacers farthest from the leader sequence were shared population-wide) (Díez-Villaseñor et al., 2010; He and Deem, 2010; Weinberger et al., 2012). It was modeled that these “old” spacers would actually confer immunity against dormant or lysogenic phages which bloomed or reemerged every certain time (Weinberger et al., 2012). Thus, duplicated spacers could belong to this group of trailer-end spacers. However, in order for these spacers to be shared across entire populations, either the dormant phages have broad host ranges or the primitive CRISPRs were laterally transferred across the population. In the first scenario, duplicate spacers would still point to broad host ranges. In the second case, they would be markers of environments with a higher rate of dormant or lysogenic viruses. Duplicity of spacers has already been used to assess viral specificity in the past (Brodthorn et al., 2011). The results of duplicate spacers per ecosystem varied depending on the software used for the recovery of CRISPRs. However, what was common for all three programs was that human gut and glacial ice had relatively high values of duplicated spacers. Some previous studies have pointed to the possible presence of phages with broad host ranges in the human gut (Modi et al., 2013; Waller et al., 2014) and glacial systems (Anesio et al., 2007), although phage specificity data of the healthy human gut and glacial ice is largely missing. Other reasons could also explain these results, for example if there was a high abundance of one specific bacterial species in the community with one associated phage a high proportion of cells would be infected by the same phage and we would therefore see high duplicate spacer numbers. This might be the case for the human gut where viral diversity has been described as low (Reyes et al., 2012), yet glacial ice had high viral diversity in our sequence analysis. Spacer duplicates as an approach to measuring viral specificity is difficult to evaluate given the lack of information about viral host ranges in general.

Abundance of CRISPR sequences could also have ecological implications related to viral dynamics and their interaction with the microbial community. Viruses with high infection rates or diverse viral communities could lead to increased CRISPR sequences in an environment as was discussed in the 4.1.2 section. However, this is contrary to the hypothesis put forward by Breitbart, 2012, where they suggest that CRISPRs are highly specific defense mechanisms and their presence would increase in environments with high viral density yet low viral diversity. In these environments there would be high contact rates, but from a limited set of phages. In their analysis, they saw that freshwater and hypersaline samples had the highest numbers of CRISPRs. Yet these environments have been described as hosting high viral diversities (López-Bueno et al., 2009; Santos et al., 2010; Emerson et al., 2012; Roux et al., 2012; Ventosa et al., 2015). In our analysis, glacial ice had the highest proportion of CRISPR sequences with two out of the three softwares used to search for CRISPRs. Based on the Breitbart, hypothesis this would point to high contact rates and low viral diversity, although alpine glacial ice had relatively high Shannon indexes in our results (4.8-5).

Overall, the results yielded in the analysis of the metagenomes from different environments with varying temperatures support the hypothesis that cold environments host a distinct viral community. The diversity of this community in cold environments could be as rich as other water systems considered to have high diversities. The analysis of the CRISPRs found within these metagenomes pointed to what could be broad host ranges and high infection rates. The results issued from this preliminary analysis encouraged us to use glacial ice as a model ecosystem to search for virus-host interactions.

5.2 CRISPR-based Infection Networks

The preliminary exploration of viral and microbial metagenomes from arctic ice and soil enabled us to draw some conclusions on the general ecological traits of these environments and their communities. Nevertheless, sequence similarity analyses are highly dependent on the content of the databases. Very little is known about microbial communities within ice, their diversity, composition or function and this lack of knowledge is reflected in the databases. Accordingly, in the MG-RAST pipeline, only

20% of the ice microbial reads were annotated to known proteins, while for soil, this value ranged from 36% to 39% of the microbial reads. Moreover, in this thesis we worked with viruses from arctic ice and soil. Environmental viruses are also underrepresented in databases, this was reflected in the high percentage of viral reads classified as unknown in the MG-RAST pipeline (*ca.* 90%). Even in the more virus-specific server Metavir, the ratios of affiliated sequences were low with glacial ice having less than 9% of the reads assigned. It thus follows that environmental viruses from glacial ice cores are largely unknown.

An unexpected result was the low viral diversity encountered in arctic glacial ice as opposed to what had been seen in the analysis of public metagenomes. The Shannon index computed for arctic glacial ice was lower than any of the Shannon indexes computed for the publicly available metagenomes, suggesting that arctic glacial ice does not host a highly diverse community of viruses. The size of the viral metagenome from ice was also one order of magnitude smaller than that of soil and the public metagenomes analyzed, which could explain this difference in viral diversity. Nevertheless, there was a higher proportion of CRISPRs in ice than in soil, therefore the model hypothesized by Breitbart, 2012 where a high proportion of CRISPRs stand as an indicator for low viral diversity and high viral density might hold for glacial ice. Extracted DNA quantities are consistent with this idea. Viral DNA from soil represented 12% of the total extracted DNA while viral DNA from the ice represented 20% of the total extracted DNA. This high viral density would be congruent with previous works that suggested high contact rates in ice channels (Wells and Deming, 2006) and could lead to high infection rates. The high number of CRISPRs found also suggests that glacial ice was a good model ecosystem for this study.

Inferences from the CRISPR-based infection network drawn for glacial ice pointed to the presence of nephridia symbionts as bacteria present in the ice and being infected by viruses. Although members of *Verminephrobacter* and *Bordetella* have so far only been described in the nephridia of earthworms, little investigation has been undertaken to describe symbionts in the nephridia of other annelids (Davidson et al., 2013). Moreover, little is known about the physiology of ice worms, yet members of the *Mesenchytraeus* genus have been described to have four pairs of nephridia (Chen

and Xie, 2015). Likewise, no phages have been described that infect nephridia symbionts, although the CRISPR finder database also hosts three CRISPRs from *Verminephrobacter* and several questionable structures for *Bordetella*. Given that there are no phages of nephridia symbionts in the database, the phages proposed by the network (*Cyanophage* and *Synechococcus* phage) must have been the closest match. Two CRISPRs from the ice were assigned to nephridia symbionts. Thus, CRISPR-based infection networks are able to link phages to their hosts even when the phages have never been described and their existence is unknown. Phages that regulate the populations of bacterial hosts in the nephridia of annelids could be essential to their health and ecology (Lund et al., 2010). This is a good example of how this approach offers a new perspective of the microbial community's ecology.

One of the main objectives in this thesis was to construct infection networks based on the information that can be retrieved through CRISPRs outsourced from metagenomic data. In this thesis a CRISPR-based workflow to link viruses with their host(s) in the environment was developed. This approach also allowed us to gain information on the microbial community of glacial ice, an understudied habitat which, as we have seen, is difficult to explore through traditional approaches such as sequence similarity. CRISPR-based infection networks facilitated the gathering of information on virus-host relationships in glacial ice. This new perspective could help us further understand and compare viral ecology in different ecosystems, giving us a more complete view of viral-microbial communities, their interactions, and consequently, their influence on global processes.

5.3 Transduction in Glacial Ice

Testing the hypothesis that prokaryotic viruses could drive microbial adaptation to environmental conditions through the transfer of genetic information was one of the major objectives of this thesis. Glacial ice was selected as a model environment because there was evidence in the literature and in the analyses carried out in this thesis that it is an environment with high infection rates and broad host ranges. These parameters could have an important impact on the rate of transduction and its relevance in microbial adaptation.

Although identification of transduction agents in arctic soil enabled their mapping onto the infection network, this was not achieved for glacial ice. However, it serves as a proof of concept. If we can identify phages that act as transduction agents and map them onto our CRISPR-based infection network, we can see which bacteria receive the microbial DNA they are carrying. In this study we identified several phage genomes, belonging to the *Caulobacter* phages and *Burkholderia* phages, as agents of transfer of mercury resistance genes in arctic soil. Unfortunately, the bacteria that these phages infected was not clear, given the difficulty in identifying DRs in soil. Transducing phages for *Caulobacter* and *Burkholderia* have been previously described (Ely and Johnson, 1977; Nzula et al., 2000; Ely et al., 2015). The application of this approach to different environments could eventually help assess the impact of transduction in microbial communities.

An important conclusion drawn from our analysis of transduction events in glacial ice was the identification of *Ralstonia* phages as transduction agents for all the microbial genes tested. *Ralstonia* phages could have an increased relevance in transduction, and thus in microbial adaptation in glacial ice. It is important to note that our approach to search for transduction agents was based on the analysis of reads from the viral metagenome which shared microbial and viral DNA. Microbial contamination of the viral fraction from which the viral metagenomes were derived cannot be ruled out. Therefore, it could be that the reads sharing microbial and viral DNA belonged to a microbial cell rather than to a viral particle indicating the presence of a prophage rather than a transduction hallmark. If this is the case, then integration sites of the prophage could probably be identified in the contig. It could also be that *Ralstonia* phages are the closest match in the database to the phage (or phages) that are actually present in the ice. However, we did find *Ralstonia* phage genes within the microbial fraction, pointing to presence of *Ralstonia* prophages which have indeed been previously described in the literature (Askora et al., 2009). This calls for further investigation in order to determine the role of these phages in glacial ice and their possible transducing nature.

Microbial contamination is a generalized problem when working with viral sequences and assessing transduction events. One study tested the publicly available viral

metagenomes for microbial contamination and concluded that many of the datasets were most probably contaminated (Roux et al., 2013). PCR amplification of 16S rRNA genes in the viral enriched fraction has been used to test for contamination prior to sequencing (Thurber et al., 2009). Yet, the 16S rRNA PCR is carried out once the viral DNA has been extracted. Given that generalized transduction could by chance pack a 16S rRNA gene within a phage, it is possible that an uncontaminated metagenome would seem contaminated by this test. A slightly different approach would be to do the PCR before the desencapsidation of the phages. Some groups have used this technique before to test for contamination of the viral sample before determining transduction events (Martinez-Castillo et al., 2013; Colomer-Lluch et al., 2014).

6. Conclusions

The main conclusions derived from this thesis are:

- Cold environments have higher viral diversities and higher abundance of CRISPRs than temperate environments.
- CRISPRs can be used as a tool to explore the nature of viral communities of environmental samples and link viruses to their hosts uncovering un-described relationships.
- Arctic glacial ice has low viral diversity (as opposed to alpine glacier ice), yet seems to have high density of viruses and possibly high infection rates.
- CRISPR-based infection networks could help elucidate the relevance of transduction in natural environments.
- Transduction could be taking place in glacial ice and *Ralstonia* phages could have an important role in this gene exchange.

7. Perspectives

The work carried out in this thesis has mainly dealt with the analysis and interpretation of data through bioinformatics. It has been based on CRISPR-based approaches to undertake ecological exploration from a new perspective and thus has identified relationships which had never been described. Yet, its reliability needs to be further tested and ascertained through further sequence analysis and empirical experimentation.

CRISPR-based infection networks could serve as an ecosystem modeling tool. It can probe the environment to find possible virus-host interactions which can be later further investigated through other means such as viral tagging (Deng et al., 2012). This method can be applied to environmental samples (Deng et al., 2014), it involves the fluorescent labeling of DNA from environmental viruses which are then combined with a pre-selected host labeled with isotopically heavy DNA. Infected cells are recovered through flow cytometry and viral DNA is parted from the host DNA with the help of a density gradient. Viral DNA can then be amplified and sequenced. In this fashion, you can test which viruses are actually capable of infecting a given organism. Moreover, this kind of evidence could ratify the hypothesis drawn with CRISPR-based infection networks.

Our analysis on transduction was done based on the presence of bacterial and viral DNA within a contig. This kind of analysis runs the risk of mistaking a transduction hallmark with a chimeric contig or microbial contamination. These possibilities lower the reliability of our analysis since we cannot be sure that the whole contig actually belonged to a virus. New sequencing technologies provide the opportunity of sequencing long fragments of DNA yielding very long reads. Nanopore technology, for example, could generate reads of up to 50.000bp (Branton et al., 2008). Thus, one read could actually comprise the whole genome of a phage (Schneider and Dekker, 2012). Having the whole genome within one read would not only omit biases due to read assembly, but also provide data which can be analyzed for transduction without microbial contamination. Implementation and affordability of these sequencing technologies will most probably lead to more reliable measures of the impact of transduction in natural environments. Moreover, the improvement in sequencing

technologies that require less DNA and the development of new methodologies, such as single genome sequencing or alternative DNA amplification methods, might also help avoid biases induced by MDA and create a more accurate picture of viral communities.

After their discovery in the first half of the 20th century and a short period of intense attention, prokaryotic viruses were mostly overlooked until the beginning of the 21st century, when their incredible abundance in the environment brought them back into the scientific scope. Since then, the in depth study of prokaryotic viruses, their physiology, their ecology and their impact on microbial populations and on global processes has become a leading subject in environmental, medical and engineering sciences. Among these, virus-host interactions in the environment are now a quickly evolving domain in science. Single cell genomics, bioinformatics tools, proteomics and other techniques are being exploited to learn about virus-host interactions (Navratil et al., 2009; Mendez-Rios and Uetz, 2010; Calderone et al., 2015; Labonté et al., 2015; Roux et al., 2015). The search of systems and mechanisms triggered by these interactions will probably produce new discoveries in the very near future. Already, a phage defense mechanism named BREX (Bacteriophage Exclusion) has been uncovered in the genomes of prokaryotes (Goldfarb et al., 2015). Other related topics such as communication among cells or population structure and their role in determining the phage defense mechanism deployed are being surveyed (Iranzo et al., 2015; Tan et al., 2015). Understanding how phages shape and regulate microbial communities is key, not only from an ecological perspective but also from an engineering and medical interest. The use of phages in medicine is prone to become a leading subject in medical sciences, especially since we approach an antibiotics crisis (Matsuzaki et al., 2014). Hence, it is expected that the investigation of the many unknowns surrounding prokaryotic viruses will urge the development of new methods and technologies, will lead to the discovery of mechanisms and systems with medical and commercial interest, and will help us further understand environmental and evolutionary microbiology.

8. References

- Abedon, S.T. and Murray, K.L. (2013) Archaeal viruses, not archaeal phages: An archaeological dig. *Archaea* **2013**.
- Achberger, A.M., Brox, T.I., Skidmore, M.L., and Christner, B.C. (2011) Expression and partial characterization of an ice-binding protein from a bacterium isolated at a depth of 3,519 m in the Vostok ice core, Antarctica. *Front. Microbiol.* **2**: 1–8.
- Ackermann, H.W. and Prangishvili, D. (2012) Prokaryote viruses studied by electron microscopy. *Arch. Virol.* **157**: 1843–1849.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–10.
- Amato, P., Hennebelle, R., Magand, O., Sancelme, M., Delort, A.-M., Barbante, C., et al. (2007) Bacterial characterization of the snow cover at Spitzberg, Svalbard. *FEMS Microbiol. Ecol.* **59**: 255–64.
- Anderson, R.E., Brazelton, W.J., and Baross, J.A. (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* **77**: 120–33.
- Anesio, A.M. and Bellas, C.M. (2011) Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends Microbiol.* **19**: 52–7.
- Anesio, A.M., Mindl, B., Laybourn-Parry, J., Hodson, A.J., and Sattler, B. (2007) Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J. Geophys. Res.* **112**: G04S31.
- Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. a, Carlson, C., et al. (2006) The marine viromes of four oceanic regions. *PLoS Biol.* **4**: e368.
- Arrigo, K.R. (2014) Sea ice ecosystems. *Ann. Rev. Mar. Sci.* **6**: 439–67.
- Askora, A., Kawasaki, T., Usami, S., Fujie, M., and Yamada, T. (2009) Host recognition and integration of filamentous phage phiRSM in the phytopathogen, *Ralstonia solanacearum*. *Virology* **384**: 69–76.
- Ball, M.M., Gómez, W., Magallanes, X., Rosales, R., Melfo, A., and Yarzabal, L.A. (2014) Bacteria recovered from a high-altitude, tropical glacier in Venezuelan Andes. *World J. Microbiol. Biotechnol.* **30**: 931–941.
- Bamber, J. (2012) Shrinking glaciers under scrutiny. *Nature* **482**: 482–483.
- Barkay, T. and Smets, B.F. (2005) Horizontal Gene Flow in Microbial Communities. **71**: 412–419.
- Barletta, R.E., Priscu, J.C., Mader, H.M., Jones, W.L., and Roe, C.H. (2012) Chemical analysis of ice vein microenvironments: II. Analysis of glacial samples from Greenland and Antarctica. *J. Glaciol.* **58**: 1109–1118.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–12.
- Barry, R.G., Jania, J., and Birkenmajer, K. (2011) Review article “A. B. Dobrowolski – the first cryospheric scientist – and the subsequent development of cryospheric

- science." *Hist. Geo. Space Sci.* **2**: 75–79.
- Bellas,C.M. and Anesio,A.M. (2013) High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles* **17**: 861–70.
- Bellas,C.M., Anesio,A.M., and Barker,G. (2015) Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* **6**: 656.
- Bellas,C.M., Anesio,A.M., Telling,J., Stibal,M., Tranter,M., and Davis,S. (2013) Viral impacts on bacterial communities in Arctic cryoconite. *Environ. Res. Lett.* **8**: 045021.
- Bidle,K.D., Lee,S., Marchant,D.R., and Falkowski,P.G. (2007) Fossil genes and microbes in the oldest ice on earth. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 13455–60.
- Björnsson,H., Gjessing,Y., Hamran,S.-E., Hagen,J.O., Liestol,O., Palsson,F., and Erlingsson,B. The thermal regime of sub-polar glaciers mapped by multi-frequency radio-echo sounding. *J. Glaciol.* **42**: 23–32.
- Bland,C., Ramsey,T.L., Sabree,F., Lowe,M., Brown,K., Kyrpides,N.C., and Hugenholtz,P. (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.
- Bolotin,A., Quinquis,B., Sorokin,A., and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551–61.
- Bondy-Denomy,J. and Davidson,A.R. (2014) When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52**: 235–242.
- Boyd,E.S., Hamilton,T.L., Havig,J.R., Skidmore,M.L., and Shock,E.L. (2014) Chemolithotrophic primary production in a subglacial ecosystem. *Appl. Environ. Microbiol.* **80**: 6146–53.
- Bramucci,A., Han,S., Beckers,J., Haas,C., and Lanoil,B. (2013) Composition, Diversity, and Stability of Microbial Assemblages in Seasonal Lake Ice, Miquelon Lake, Central Alberta. *Biology (Basel).* **2**: 514–532.
- Branton,D., Deamer,D.W., Marziali,A., Bayley,H., Benner,S. a, Butler,T., et al. (2008) The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**: 1146–1153.
- Breitbart,M. (2012) Marine Viruses: Truth or Dare. *Ann. Rev. Mar. Sci.* **4**: 425–448.
- Breitbart,M., Felts,B., Kelley,S., Mahaffy,J.M., Nulton,J., Salamon,P., and Rohwer,F. (2004) Diversity and population structure of a near-shore marine-sediment viral community. *Proc. Biol. Sci.* **271**: 565–74.
- Breitbart,M. and Rohwer,F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**: 278–84.
- Brodts,A., Lurie-Weinberger,M.N., and Gophna,U. (2011) CRISPR loci reveal networks of gene exchange in archaea. *Biol. Direct* **6**: 65.
- Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., et al. (2008) Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.)*. **321**: 960–964.

- Brown, J.R., Seymour, J.D., Brox, T.I., Skidmore, M.L., Wang, C., Christner, B.C., et al. (2014) Recrystallization inhibition in ice due to ice binding protein activity detected by nuclear magnetic resonance. *Biotechnol. Reports* **3**: 60–64.
- Bruce, T., Meirelles, P.M., Garcia, G., Paranhos, R., Rezende, C.E., de Moura, R.L., et al. (2012) Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS One* **7**:
- Brum, J.R., Hurwitz, B.L., Schofield, O., Ducklow, H.W., and Sullivan, M.B. (2015) Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* 1–13.
- Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulier, G., Acinas, S.G., Alberti, A., et al. (2015) Patterns and ecological drivers of ocean viral communities. *Science* (80-). **348**: 1261498–1261498.
- Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and Sensitive Protein Alignment using DIAMOND. *Under Rev.* **12**:
- Buckling, A. and Rainey, P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. Biol. Sci.* **269**: 931–6.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., et al. (1996) Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science* (80-). **273**: 1058–1073.
- Calderone, A., Licata, L., and Cesareni, G. (2015) VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.* **43**: D588–92.
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., and Brüssow, H. (2003) Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**: 417–424.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2011) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**: 335–336.
- Cappa, F., Suci, N., Trevisan, M., Ferrari, S., Puglisi, E., and Sandro, P. (2014) Science of the Total Environment Bacterial diversity in a contaminated Alpine glacier as determined by culture-based and molecular approaches. *Sci. Total Environ.* **497-498**: 50–59.
- Carda-Diequez, M., Mizuno, C.M., Ghai, R., Rodriguez-Valera, F., and Amaro, C. (2015) Replicating phages in the epidermal mucosa of the eel (*Anguilla anguilla*). *Front. Microbiol.* **6**: 1–10.
- Carroll, L. (1871) *Through the Looking-Glass*.
- Del Casale, A., Flanagan, P. V., Larkin, M.J., Allen, C.C.R., and Kulakov, L. a. (2011) Analysis of transduction in wastewater bacterial populations by targeting the phage-derived 16S rRNA gene sequences. *FEMS Microbiol. Ecol.* **76**: 100–108.
- Charpentier, E., Richter, H., van der Oost, J., and White, M.F. (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**: 428–441.
- Chen, J. and Xie, Z. (2015) *Mesenchytraeus asymmetriauritus*, a new enchytraeid (Annelida, Clitellata) from northeastern China. **128**: 1–7.

- Christner,B., Mosley-thompson,E., Thompson,L.G., and Reeve,J.N. (2005) Recovery and identification of bacteria from polar and non-polar glacial ice. In, In S. O. Rogers and J. Castello (eds) (ed), *Life in Ancient Ice*. Princeton University Press, Princeton, New Jersey, pp. 209–227.
- Christner,B.C., Mosley-thompson,E., Thompson,L.G., Zagorodnov,V., Sandman,K., and Reeve,J.N. (2000) Recovery and Identification of Viable Bacteria Immured in Glacial Ice. *Icarus* **144**: 479–485.
- Christner,B.C., Royston-Bishop,G., Foreman,C.M., Arnold,B.R., Tranter,M., Welch,K. a., et al. (2006) Limnological conditions in Subglacial Lake Vostok, Antarctica. *Limnol. Oceanogr.* **51**: 2485–2501.
- Chu,V.W. (2013) Greenland ice sheet hydrology: A review.
- Chylinski,K., Le Rhun,A., and Charpentier,E. (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**: 726–737.
- Clokic,M.R., Millard,A.D., Letarov,A. V, and Heaphy,S. (2011) Phages in nature. *Bacteriophage* **1**: 31–45.
- Colangelo-Lillis,J.R. and Deming,J.W. (2013) Genomic analysis of cold-active Colwelliophage 9A and psychrophilic phage-host interactions. *Extremophiles* **17**: 99–114.
- Cole,J.R., Wang,Q., Fish,J.A., Chai,B., McGarrell,D.M., Sun,Y., et al. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**: D633–42.
- Collins,R.E. and Deming,J.W. (2011) Abundant dissolved genetic material in Arctic sea ice Part I: Extracellular DNA. *Polar Biol.* **34**: 1819–1830.
- Colomer-Lluch,M., Calero-Cáceres,W., Jebri,S., Hmaied,F., Muniesa,M., and Jofre,J. (2014) Antibiotic resistance genes in bacterial and bacteriophage fractions of Tunisian and Spanish wastewaters as markers to compare the antibiotic resistance patterns in each population. *Environ. Int.* **73**: 167–75.
- Cottrell,M. and Kirchman,D. (2012) Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquat. Microb. Ecol.* **66**: 107–116.
- d’Herelle,F. (1929) Studies Upon Asiatic Cholera. *Yale J. Biol. Med.* **1**: 195–219.
- d’Herelle,F. (1917) Sur un microbe invisible antagonistedes bacilles dysenteriques. *Les Comptes rendus del’Académie des Sci.* **165**: 373–5.
- d’Hérelle,F. and Smith,G.H. (1926) The bacteriophage and its behavior Am Assoc Immnol.
- Dani,K.G.S., Mader,H.M., Wolff,E.W., and Wadham,J.L. (2012) Modelling the liquid-water vein system within polar ice sheets as a potential microbial habitat. *Earth Planet. Sci. Lett.* **333-334**: 238–249.
- Davidson,S.K., Powell,R., and James,S. (2013) A global survey of the bacteria within earthworm nephridia. *Mol. Phylogenet. Evol.* **67**: 188–200.
- DeBruyn,J.M., Nixon,L.T., Fawaz,M.N., Johnson,A.M., and Radosevich,M. (2011) Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl. Environ. Microbiol.* **77**: 6295–6300.

- Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602–7.
- Deming, J.W. (2009) Sea Ice Bacteria and Viruses. In, Thomas, D.N. and Dieckmann, G.S. (eds), *Sea Ice: An Introduction to Its Physics, Chemistry, Biology, and Geology*. Wiley-Blackwell, Oxford, UK, pp. 247–282.
- Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012) Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. **3**: 1–8.
- Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B. (2014) Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature*.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., et al. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**: 1390–400.
- Díez-Villaseñor, C., Almendros, C., García-Martínez, J., and Mojica, F.J.M. (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology* **156**: 1351–61.
- DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N.K., and Noll, K.M. (2009) *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. *Int. J. Syst. Evol. Microbiol.* **59**: 2991–3000.
- Dommergue, A., Larose, C., Faïn, X., Clarisse, O., Hintelmann, H., Schneider, D., and Ferrari, C.P. (2010) Deposition of mercury species in the Ny-Alesund area (79 degrees N) and their transfer during snowmelt. *Environ. Sci. Technol.* **44**: 901–7.
- Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kasthuber, E.R., Livshits, G., et al. (2015) Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**: 390–4.
- Dowdeswell, J. a. and Siegert, M.J. (2003) The physiography of modern Antarctic subglacial lakes. *Glob. Planet. Change* **35**: 221–236.
- Doyle, S., Montross, S., Skidmore, M., and Christner, B. (2013) Characterizing Microbial Diversity and the Potential for Metabolic Function at –15 °C in the Basal Ice of Taylor Glacier, Antarctica. *Biology (Basel)*. **2**: 1034–1053.
- Dubey, G.P. and Ben-Yehuda, S. (2011) Intercellular nanotubes mediate bacterial communication. *Cell* **144**: 590–600.
- Duhaime, M.B., Deng, L., Poulos, B.T., and Sullivan, M.B. (2012) Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: A rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* **14**: 2526–2537.
- Dziewit, L., Grzesiak, J., Ciok, A., Nieckarz, M., Zdanowski, M.K., and Bartosik, D. (2013) Sequence determination and analysis of three plasmids of *Pseudomonas* sp. GLE121, a psychrophile isolated from surface ice of Ecology Glacier (Antarctica). *Plasmid* **70**: 254–262.
- Ebina, H., Misawa, N., Kanemura, Y., and Koyanagi, Y. (2013) Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* **3**: 2510.

- Edgar,R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**: 18.
- Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–1.
- Edwards,A., Pachebat,J. a, Swain,M., Hegarty,M., Hodson,A.J., Irvine-Fynn,T.D.L., et al. (2013) A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. *Environ. Res. Lett.* **8**: 035003.
- Edwards,R.A. and Rohwer,F. (2005) Viral Metagenomics. *Handb. Mol. Microb. Ecol. II Metagenomics Differ. Habitats* **3**: 3–13.
- Ely,B., Gibbs,W., Diez,S., and Ash,K. (2015) The Caulobacter crescentus Transducing Phage Cr30 is a Unique Member of the T4-Like Family of Myophages. *Curr. Microbiol.* **70**: 854–858.
- Ely,B. and Johnson,R.C. (1977) GENERALIZED TRANSDUCTION IN CAULOBACTER CRESCENTUS. *Genetics* **87**: 391–399.
- Emerson,J.B., Thomas,B.C., Andrade,K., Allen,E.E., Heidelberg,K.B., and Banfield,J.F. (2012) Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl. Environ. Microbiol.* **78**: 6309–20.
- Erdmann,S. and Garrett,R. a (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol. Microbiol.* **85**: 1044–56.
- Falkinham,J.O.I. (2002) Nontuberculous mycobacteria in the environment. *Clin. Chest Med.* **23**: 529–551.
- Fassler,J. and Cooper,P. (2011) BLAST Glossary.
- Filée,J., Tétart,F., Suttle,C. a, and Krisch,H.M. (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 12471–6.
- Fineran,P.C. and Charpentier,E. (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**: 202–9.
- Fisk,A.T., De Wit,C. a., Wayland,M., Kuzyk,Z.Z., Burgess,N., Letcher,R., et al. (2005) An assessment of the toxicological significance of anthropogenic contaminants in Canadian arctic wildlife. *Sci. Total Environ.* **351-352**: 57–93.
- Fleming,A. (1929) On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *Br. J. Exp. Pathol.* **10**: 226.
- Foreman,C.M., Cory,R.M., Morris,C.E., SanClements,M.D., Smith,H.J., Lisle,J.T., et al. (2013) Microbial growth under humic-free conditions in a supraglacial stream system on the Cotton Glacier , Antarctica. *Environ. Res. Lett.* **8**: 1–11.
- Foreman,C.M., Dieser,M., Greenwood,M., Cory,R.M., Laybourn-Parry,J., Lisle,J.T., et al. (2011) When a habitat freezes solid: microorganisms over-winter within the ice column of a coastal Antarctic lake. *FEMS Microbiol. Ecol.* **76**: 401–12.
- Fredga,K., Jaarola,M., Ims,R.A., Steen,H., and Yoccoz,N.G. (1990) The “common vole” in Svalbard identified as *Microtus epiroticus* by chromosome analysis. *Polar Res.* **8**: 283–290.

- Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**: 722–32.
- Fuhrman, J. a (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.
- Gaidos, E., Marteinson, V., Thorsteinsson, T., Jóhannesson, T., Rúnarsson, A.R., Stefansson, A., et al. (2009) An oligarchic microbial assemblage in the anoxic bottom waters of a volcanic subglacial lake. *ISME J.* **3**: 486–97.
- Galtier, N., Tourasse, N., and Gouy, M. (1999) A non hyperthermophilic common ancestor to extant life forms. *Science* **283**: 220–1.
- Garcia-Heredia, I., Martin-Cuadrado, A.-B., Mojica, F.J.M., Santos, F., Mira, A., Antón, J., and Rodriguez-Valera, F. (2012) Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS One* **7**: e33802.
- Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D. a, Barrangou, R., Boyaval, P., et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67–71.
- Ghai, R., Hernandez, C.M., Picazo, A., Mizuno, C.M., Ininbergs, K., Díez, B., et al. (2012) Metagenomes of Mediterranean Coastal Lagoons. *Sci. Rep.* **2**: 1–13.
- Giovannoni, S., Temperton, B., and Zhao, Y. (2013) Giovannoni et al. reply. *Nature* **499**: E4–5.
- GLIMS, and National Snow and Ice Data Center (2012) *GLIMS Glacier Database*. Boulder, Colorado USA: National Snow and Ice Data.
- Godde, J.S. and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **62**: 718–729.
- Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., et al. (2015) BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**: 169–83.
- Goodman, D. (1971) *Ecological Investigations of Ice Worms on Casement Glacier, Southeastern Alaska*.
- Gotz, D., Banta, A., Beveridge, T.J., Rushdi, A.I., Simoneit, B.R.T., and Reysenbach, A.L. (2002) *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* **52**: 1349–1359.
- Gowing, M., Garrison, D., Gibson, A., Krupp, J., Jeffries, M., and Fritsen, C. (2004) Bacterial and viral abundance in Ross Sea summer pack ice communities. *Mar. Ecol. Prog. Ser.* **279**: 3–12.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007a) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**: W52–7.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007b) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

- Groudieva,T., Kambourova,M., Yusef,H., Royter,M., Grote,R., Trinks,H., and Antranikian,G. (2004) Diversity and cold-active hydrolytic enzymes of culturable bacteria associated with Arctic sea ice, Spitzbergen. *Extremophiles* **8**: 475–88.
- Gunde-Cimerman,N., Sonjak,S., Zalar,P., Frisvad,J.C., Diderichsen,B., and Plemenitaš, a. (2003) Extremophilic fungi in arctic ice: a relationship between adaptation to low temperature and water activity. *Phys. Chem. Earth, Parts A/B/C* **28**: 1273–1278.
- Haft,D.H., Selengut,J., Mongodin,E.F., and Nelson,K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**: e60.
- Hagen,J. (1988) Glacier mass balance investigations in the balance year 1986-87. *Polar Res.* **6**: 205–209.
- Hanna,E., Navarro,F.J., Pattyn,F., Domingues,C.M., Fettweis,X., Ivins,E.R., et al. (2013) Ice-sheet mass balance and climate change. *Nature* **498**: 51–9.
- Hartzell,P.L., Nghiem,J. V, Richio,K.J., and Shain,D.H. (2005) Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus rainierensis*). *Can. J. Zool.* **83**: 1206–1213.
- He,J. and Deem,M.W. (2010) Heterogeneous Diversity of Spacers within CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). *Phys. Rev. Lett.* **105**: 128102.
- Heler,R., Samai,P., Modell,J.W., Weiner,C., Goldberg,G.W., Bikard,D., and Marraffini,L.A. (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**: 199–202.
- Hermans,P.W., van Soolingen,D., Bik,E.M., de Haas,P.E., Dale,J.W., and van Embden,J.D. (1991) Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* **59**: 2695–2705.
- Hodson,A., Anesio,A.M., Tranter,M., Fountain,A., Osborn,M., Priscu,J.C., et al. (2008) Glacial Ecosystems. *Ecol. Monogr.* **78**: 41–67.
- Holmfeldt,K., Solonenko,N., Shah,M., Corrier,K., Riemann,L., Verberkmoes,N.C., and Sullivan,M.B. (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 12798–803.
- Horvath,P., Romero,D.A., Coûté-Monvoisin,A.-C., Richards,M., Deveau,H., Moineau,S., et al. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**: 1401–12.
- Hurwitz,B.L., Hallam,S.J., and Sullivan,M.B. (2013) Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* **14**: R123.
- Iranzo,J., Lobkovsky,A.E., Wolf,Y.I., and Koonin,E. V (2015) Immunity, suicide or both? Ecological determinants for the combined evolution of anti-pathogen defense systems. *BMC Evol. Biol.* **15**: 43.
- Irvine-Fynn,T.D.L., Hodson,A.J., Moorman,B.J., Vatne,G., and Hubbard,A.L. (2011) Polythermal glacier hydrology : A review. *Rev. Geophys.* **49**: 1–37.
- Ishino,Y., Shinagawa,H., Makino,K., Amemura,M., and Nakata, a (1987) Nucleotide

- sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**: 5429–5433.
- Jackson, R.N., Golden, S.M., van Erp, P.B.G., Carter, J., Westra, E.R., Brouns, S.J.J., et al. (2014) Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **1473**: 11–15.
- Jacob, T., Wahr, J., Pfeffer, W.T., and Swenson, S. (2012) Recent contributions of glaciers and ice caps to sea level rise. *Nature* **482**: 514–518.
- Jacques, N., Zenouche, A., Gunde-Cimerman, N., and Casaregola, S. (2015) Increased diversity in the genus *Debaryomyces* from Arctic glacier samples. *Antonie Van Leeuwenhoek* **107**: 487–501.
- Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**: 1598–602.
- Jansen, R., Embden, J.D. a Van, Gaastra, W., and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**: 1565–75.
- Janssen, P.H. (2006) Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes MINIREVIEWS Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Appl. Environ. Microbiol.* **72**: 1719–1728.
- Jiang, S.C. and Paul, J.H. (1998) Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* **64**: 2780–7.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.)*. **337**: 816–821.
- Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**: 181–196.
- Josefsen, T.D., Sørensen, K.K., Mørk, T., Mathiesen, S.D., and Ryeng, K. a (2007) Fatal inanition in reindeer (*Rangifer tarandus tarandus*): pathological findings in completely emaciated carcasses. *Acta Vet. Scand.* **49**: 27.
- Jungblut, A.D., Vincent, W.F., and Lovejoy, C. (2012) Eukaryotes in Arctic and Antarctic cyanobacterial mats. *FEMS Microbiol. Ecol.* **82**: 416–428.
- Karl, D., Bird, D.F., Björkman, K., Houlihan, T., Shackelford, R., and Tupas, L. (1999) Microorganisms in the Accreted Ice of Lake Vostok , Antarctica. *Science (80-.)*. **286**: 2144–2147.
- Kaštovská, K., Elster, J., Stibal, M., and Šantrůčková, H. (2005) Microbial Assemblages in Soil Microbial Succession After Glacial Retreat in Svalbard (High Arctic). *Microb. Ecol.* **50**: 396–407.
- Kawasaki, T., Shimizu, M., Satsuma, H., Fujiwara, A., Fujie, M., Usami, S., and Yamada, T. (2009) Genomic characterization of *Ralstonia solanacearum* phage phiRSB1, a T7-like wide-host-range phage. *J. Bacteriol.* **191**: 422–7.

- Kenzaka,T., Tani,K., and Nasu,M. (2010) High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J.* **4**: 648–659.
- Kerr,R.A. (2013) Glaciology. Melting glaciers, not just ice sheets, stoking sea-level rise. *Science* **340**: 798.
- Kim,K.H. and Bae,J.W. (2011) Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. Environ. Microbiol.* **77**: 7663–7668.
- Kloos,W.E. (1980) Natural populations of the genus *Staphylococcus*. *Annu. Rev. Microbiol.* **34**: 559–592.
- Konopka,A. (2009) What is microbial community ecology? *ISME J.* **3**: 1223–1230.
- Koskella,B. and Brockhurst,M. a. (2014) Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**: 916–31.
- Koskella,B. and Meaden,S. (2013) Understanding bacteriophage specificity in natural microbial communities. *Viruses* **5**: 806–23.
- Kunin,V., Sorek,R., and Hugenholtz,P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**: R61.
- Kyrke-Smith,T.M., Katz,R.F., and Fowler,A.C. (2014) Subglacial hydrology and the formation of ice streams. *Proc. R. Soc. A* **470**: 20130494.
- Labonté,J.M. and Suttle,C.A. (2013) Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* **7**: 2169–77.
- Labonté,J.M., Swan,B.K., Poulos,B., Luo,H., Koren,S., Hallam,S.J., et al. (2015) Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.*
- Lang,A.S., Zhaxybayeva,O., and Beatty,J.T. (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* **10**: 472–82.
- Langford,H.J., Irvine-Fynn,T.D.L., Edwards, a., Banwart,S. a., and Hodson, a. J. (2014) A spatial investigation of the environmental controls over cryoconite aggregation on Longyearbreen glacier, Svalbard. *Biogeosciences Discuss.* **11**: 3423–3463.
- Larose,C., Berger,S., Ferrari,C., Navarro,E., Dommergue,A., Schneider,D., and Vogel,T.M. (2010) Microbial sequences retrieved from environmental samples from seasonal arctic snow and meltwater from Svalbard, Norway. *Extremophiles* **14**: 205–12.
- Larose,C., Dommergue,A., and Vogel,T. (2013) The Dynamic Arctic Snow Pack: An Unexplored Environment for Microbial Diversity and Activity. *Biology (Basel).* **2**: 317–330.
- Larose,C., Prestat,E., Cecillon,S., Berger,S., Malandain,C., Lyon,D., et al. (2013) Interactions between Snow Chemistry, Mercury Inputs and Microbial Population Dynamics in an Arctic Snowpack. *PLoS One* **8**: e79972.
- Lauro,F.M., DeMaere,M.Z., Yau,S., Brown,M. V, Ng,C., Wilkins,D., et al. (2011) An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J.* **5**: 879–95.

- Laybourn-Parry, J., Hodson, A.J., and Tranter, M. (2012) The Ecology of Snow and Ice Environments.
- Leeson, A.A., Shepherd, A., Briggs, K., Howat, I., Fettweis, X., Morlighem, M., and Rignot, E. (2015) Supraglacial lakes on the Greenland ice sheet advance inland under warming climate. *Nat. Clim. Chang.* **5**: 51–55.
- Lehman, N. (2013) Origin of life: Cold-hearted RNA heats up life. *Nat. Chem.* **5**: 987–9.
- Liao, H.-K., Gu, Y., Diaz, A., Marlett, J., Takahashi, Y., Li, M., et al. (2015) Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells. *Nat. Commun.* **6**: 6413.
- Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–9.
- Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11013–8.
- Llosa, M., Gomis-Rüth, F.X., Coll, M., and De la Cruz, F. (2002) Bacterial conjugation: A two-step mechanism for DNA transport. *Mol. Microbiol.* **45**: 1–8.
- Llubes, M., Lanseau, C., and Rémy, F. (2006) Relations between basal condition, subglacial hydrological networks and geothermal flux in Antarctica. *Earth Planet. Sci. Lett.* **241**: 655–662.
- López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcamí, A. (2009) High diversity of the viral community from an Antarctic lake. *Science* **326**: 858–61.
- López-García, P. and Moreira, D. (2012) Viruses in Biology. *Evol. Educ. Outreach* **5**: 389–398.
- Lund, M.B., Holmstrup, M., Lomstein, B.A., Damgaard, C., and Schramm, A. (2010) Beneficial effect of Verminephrobacter nephridial symbionts on the fitness of the earthworm *Aporrectodea tuberculata*. *Appl. Environ. Microbiol.* **76**: 4738–43.
- Luria, S.E. and Anderson, T.F. (1942) The Identification and Characterization of Bacteriophages with the Electron Microscope. *Proc. Natl. Acad. Sci. U. S. A.* **28**: 127–130.1.
- Lutz, S., Anesio, A.M., Jorge Villar, S.E., and Benning, L.G. (2014) Variations of algal communities cause darkening of a Greenland glacier. *FEMS Microbiol. Ecol.* **89**: 402–14.
- Ma, L.-J., Rogers, S.O., Catranis, C.M., and Starmer, W.T. (2000) Detection and characterization of ancient fungi entrapped in glacial ice. *Mycologia* **92**: 286–295.
- Ma, Y., Zhang, L., and Huang, X. (2014) Genome modification by CRISPR/Cas9. *FEBS J.* **281**: 5186–93.
- MacDonell, S. and Fitzsimons, S. (2008) The formation and hydrological significance of cryoconite holes. *Prog. Phys. Geogr.* **32**: 595–610.
- Mader, H.M., Pettitt, M.E., Wadham, J.L., Wolff, E.W., and Parkes, R.J. (2006) Subsurface ice as a microbial habitat. *Geology* **34**: 169.

- Majewski, J. (2001) Sexual isolation in bacteria. *FEMS Microbiol. Lett.* **199**: 161–169.
- Makarova, K.S., Brouns, S.J.J., Horvath, P., Sas, D.F., and Wolf, Y.I. (2012) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**: 467–477.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., et al. (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**: 722–736.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., et al. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823–6.
- Mansutti, D., Bucchignani, E., Otero, J., and Glowacki, P. (2015) Modeling and numerical sensitivity study on the conjecture of a subglacial lake at Amundsenisen, Svalbard Elsevier Inc.
- Maranger, R., Bird, D.F., and Juniper, S.K. (1994) Viral and bacterial dynamics in Arctic sea ice during the spring algal bloom near Resolute, NWT, Canada. *Mar. Ecol. Prog. Ser.* **111**: 121–128.
- Margesin, R. and Miteva, V. (2011) Diversity and ecology of psychrophilic microorganisms. *Res. Microbiol.* **162**: 346–61.
- Margesin, R., Zacke, G., and Schinner, F. (2002) Characterization of Heterotrophic Microorganisms in Alpine Glacier Cryoconite. *Arctic, Antarct. Alp. Res.* **34**: 88–93.
- Marshall, S.J. (2011) The cryosphere Princeton University Press.
- Martinez-Castillo, A., Quirós, P., Navarro, F., Miró, E., and Muniesa, M. (2013) Shiga toxin 2-encoding bacteriophages in human fecal samples from healthy individuals. *Appl. Environ. Microbiol.* **79**: 4862–8.
- Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., et al. (2015) Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat. Med.* **21**: 256–62.
- Matilla, M.A. and Salmond, G.P.C. (2014) Bacteriophage ϕ MAM1, a viunlikevirus, is a broad-host-range, high-efficiency generalized transducer that infects environmental and clinical isolates of the enterobacterial genera *Serratia* and *Kluyvera*. *Appl. Environ. Microbiol.* **80**: 6446–57.
- Matsuzaki, S., Uchiyama, J., Takemura-Uchiyama, I., and Daibata, M. (2014) Perspective: The age of the phage. *Nature* **509**: S9.
- McMillan, M., Nienow, P., Shepherd, A., Benham, T., and Sole, A. (2007) Seasonal evolution of supra-glacial lakes on the Greenland Ice Sheet. *Earth Planet. Sci. Lett.* **262**: 484–492.
- Mendes, L.W., Kuramae, E.E., Navarrete, A. a, van Veen, J. a, and Tsai, S.M. (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J.* 1577–1587.
- Mendez-Rios, J. and Uetz, P. (2010) Global approaches to study protein-protein interactions among viruses and hosts. *Future Microbiol.* **5**: 289–301.
- Menor-Salván, C. and Marín-Yaseli, M.R. (2012) Prebiotic chemistry in eutectic solutions at the water-ice matrix. *Chem. Soc. Rev.* **41**: 5404–15.

- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.
- Miteva, V. (2008) Bacteria in Snow and Glacier Ice. In, Margesin, RosaSchinner, F., Marx, J.-C., and Gerday, C. (eds), *Psychrophiles: from Biodiversity to Biotechnology*.
- Miteva, V., Burlingame, C., Sowers, T., and Brenchley, J. (2014) Comparative evaluation of the indigenous microbial diversity vs. drilling fluid contaminants in the NEEM Greenland ice core. *FEMS Microbiol. Ecol.* **89**: 238–256.
- Miteva, V., Teacher, C., Sowers, T., and Brenchley, J. (2009) Comparison of the microbial diversity at different depths of the GISP2 Greenland ice core in relationship to deposition climates. *Environ. Microbiol.* **11**: 640–56.
- Miteva, V.I. and Brenchley, J.E. (2005) Detection and Isolation of Ultrasmall Microorganisms from a 120,000-Year-Old Greenland Glacier Ice Core. *Appl. Environ. Microbiol.* **71**: 7806–7818.
- Miteva, V.I., Sheridan, P.P., and Brenchley, J.E. (2004) Phylogenetic and Physiological Diversity of Microorganisms Isolated from a Deep Greenland Glacier Ice Core. *Appl. Environ. Microbiol.* **70**: 202–213.
- Miyakawa, S., Cleaves, H.J., and Miller, S.L. (2002) THE COLD ORIGIN OF LIFE: HYDROGEN CYANIDE AND FORMAMIDE. *Orig. Life Evol. Biosph.* **32**: 195–208.
- Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., and Ghai, R. (2013) Expanding the marine virosphere using metagenomics. *PLoS Genet.* **9**: e1003987.
- Modi, S.R., Lee, H.H., Spina, C.S., and Collins, J.J. (2013) Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**: 219–22.
- Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–40.
- Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**: 174–82.
- Mojica, F.J.M., Díez-Villasenor, C., Soria, E., and Juez, G. (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**: 244–246.
- Mojica, F.J.M., Ferrer, C., and Juez, G. (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloforax mediterranei* and *Haloforax volcanii* and could be involved in replicon partitioning. **17**: 85–93.
- Mojica, K.D. a and Brussaard, C.P.D. (2014) Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiol. Ecol.* **89**: 495–515.
- Møller, A.K., Barkay, T., Hansen, M. a, Norman, A., Hansen, L.H., Sørensen, S.J., et al. (2014) Mercuric reductase genes (*merA*) and mercury resistance plasmids in High Arctic snow, freshwater and sea-ice brine. *FEMS Microbiol. Ecol.* **87**: 52–63.

- Morse, M.L., Lederberg, E.M., and Lederberg, J. (1956) Transduction in *Escherichia coli* k-12. *Genetics* **41**: 142–156.
- Mosier, A.C., Murray, A.E., and Fritsen, C.H. (2007) Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS Microbiol. Ecol.* **59**: 274–288.
- Murakami, T., Segawa, T., Bodington, D., Dial, R., Takeuchi, N., Kohshima, S., and Hongoh, Y. (2015) Census of bacterial microbiota associated with the glacier ice worm *Mesenchytraeus solifugus*. *FEMS Microbiol. Ecol.* **91**: fiv003.
- Navratil, V., de Chasse, B., Meyniel, L., Delmotte, S., Gautier, C., André, P., et al. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.* **37**: D661–8.
- Nesbø, C.L., Kumaraswamy, R., Dlutek, M., Doolittle, W.F., and Foght, J. (2010) Searching for mesophilic thermotogales bacteria: “Mesotogas” in the wild. *Appl. Environ. Microbiol.* **76**: 4896–4900.
- Ninetyeth Annual Meeting of the British Medical Association (1922) *Br. Med. J.* **2**: 289–309.
- Norman, A., Hansen, L.H., and Sørensen, S.J. (2009) Conjugative plasmids: vessels of the communal gene pool. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364**: 2275–89.
- Nzula, S., Vandamme, P., and Govan, J.R.W. (2000) Sensitivity of the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* to transducing bacteriophages. *FEMS Immunol. Med. Microbiol.* **28**: 307–312.
- Obleitner, F. and Lehning, M. (2004) Measurement and simulation of snow and superimposed ice at the Kongsvegen glacier, Svalbard (Spitzbergen). *J. Geophys. Res.* **109**: D04106.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., et al. (2015) vegan: Community Ecology Package.
- van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **12**: 479–492.
- Ousterout, D.G., Kabadi, A.M., Thakore, P.I., Majoros, W.H., Reddy, T.E., and Gersbach, C.A. (2015) Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* **6**: 6244.
- Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K.-I., Stahl, B., et al. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat. Commun.* **4**: 1430.
- Palmer, S.J., Dowdeswell, J. a., Christoffersen, P., Young, D. a., Blankenship, D.D., Greenbaum, J.S., et al. (2013) Greenland subglacial lakes detected by radar. *Geophys. Res. Lett.* **40**: 6154–6159.
- Panikov, N.S., Flanagan, P.W., Oechel, W.C., Mastepanov, M. a., and Christensen, T.R. (2006) Microbial activity in soils frozen to below -39°C. *Soil Biol. Biochem.* **38**: 785–794.
- Paterson, H. and Laybourn-Parry, J. (2012) Antarctic sea ice viral dynamics over an

- annual cycle. *Polar Biol.* **35**: 491–497.
- Payet, J.P. and Suttle, C.A. (2013) To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol. Oceanogr.* **58**: 465–474.
- Pearce, D., Hodgson, D., Thorne, M., Burns, G., and Cockell, C. (2013) Preliminary Analysis of Life within a Former Subglacial Lake Sediment in Antarctica. *Diversity* **5**: 680–702.
- Peel, M.C., Finlayson, B.L., and McMahon, T.A. (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**: 1633–1644.
- Pimentel, S. and Flowers, G.E. (2011) A numerical study of hydrologically driven glacier dynamics and subglacial flooding. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **467**: 537–558.
- Popa, O. and Dagan, T. (2011) Trends and barriers to lateral gene transfer in prokaryotes. *Curr. Opin. Microbiol.* **14**: 615–623.
- Port, J. a., Wallace, J.C., Griffith, W.C., and Faustman, E.M. (2012) Metagenomic Profiling of Microbial Composition and Antibiotic Resistance Determinants in Puget Sound. *PLoS One* **7**.
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653–63.
- Prangishvili, D. (2013) The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.* **67**: 565–85.
- Prangishvili, D., Forterre, P., and Garrett, R. a (2006) Viruses of the Archaea: a unifying view. *Nat. Rev. Microbiol.* **4**: 837–848.
- Prangishvili, D., Garrett, R.A., and Koonin, E. V (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.* **117**: 52–67.
- Price, P.B. (2000) A habitat for psychrophiles in deep Antarctic ice. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 1247–51.
- Price, P.B. (2009) Microbial genesis, life and death in glacial ice. *Can. J. Microbiol.* **55**: 1–11.
- Price, P.B. (2007) Microbial life in glacial ice and implications for a cold origin of life. *FEMS Microbiol. Ecol.* **59**: 217–231.
- Price, P.B., Nagornov, O. V, Bay, R., Chirkin, D., He, Y., Miocinovic, P., et al. (2002) Temperature profile for glacial ice at the South Pole: implications for life in a nearby subglacial lake. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 7844–7.
- Price, P.B. and Sowers, T. (2004) Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 4631–6.
- R Core Team (2014) R: A Language and Environment for Statistical Computing.
- Rachel, R., Bettstetter, M., Hedlund, B.P., Häring, M., Kessler, A., Stetter, K.O., and Prangishvili, D. (2002) Remarkable morphological diversity of viruses and virus-like

- particles in hot terrestrial environments. *Arch. Virol.* **147**: 2419–29.
- Raymond, J. a., Christner, B.C., and Schuster, S.C. (2008) A bacterial ice-binding protein from the Vostok ice core. *Extremophiles* **12**: 713–717.
- Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012) Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* **10**: 607–17.
- Rho, M., Wu, Y.-W., Tang, H., Doak, T.G., and Ye, Y. (2012) Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* **8**: e1002441.
- Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001) Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 13341–5.
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasić, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009) Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**: 828–36.
- Rodriguez-Valera, F., Mizuno, C.M., and Ghai, R. (2014) Tales from a thousand and one phages. *Bacteriophage* **4**: e28265.
- Rohde, R. a and Price, P.B. (2007) Diffusion-controlled metabolism for long-term survival of single isolated microorganisms trapped within ice crystals. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 16592–7.
- Rohwer, F. and Thurber, R.V. (2009) Viruses manipulate the marine environment. *Nature* **459**: 207–12.
- Roth, K. and Boike, J. (2001) Quantifying the thermal dynamics of a permafrost site near Ny-??lesund, Svalbard. *Water Resour. Res.* **37**: 2901–2914.
- Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., et al. (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One* **7**: e33641.
- Roux, S., Faubladier, M., Mahul, A., Paulhe, N., Bernard, A., Debroas, D., and Enault, F. (2011) Metavir: a web server dedicated to virome analysis. *Bioinformatics* **27**: 3074–5.
- Roux, S., Hallam, S.J., Woyke, T., and Sullivan, M.B. (2015) Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *Elife* **4**: 1–20.
- Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013) Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol.* **3**: 130160.
- Roux, S., Tournayre, J., Mahul, A., Debroas, D., and Enault, F. (2014) Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* **15**: 76.
- von Rozycki, T. and Nies, D.H. (2009) Cupriavidus metallidurans: evolution of a metal-resistant bacterium. *Antonie Van Leeuwenhoek* **96**: 115–39.
- Ruska, H. (1940) Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. *Naturwissenschaften* **28**: 45–46.

- Russel,M. (1995) Moving through the membrane with filamentous phages. *Trends Microbiol.* **3**: 223–228.
- Saito,R., Smoot,M.E., Ono,K., and Ruscheinski,J. (2012) A travel guide to Cytoscape plugins. *Nat. Methods* **9**: 1069–1076.
- Sano,E., Carlson,S., Wegley,L., and Rohwer,F. (2004) Movement of Viruses between Biomes. *Appl. Environ. Microbiol.* **70**: 5842–5846.
- Santos,F., Yarza,P., Parro,V., Briones,C., and Antón,J. (2010) The metavirome of a hypersaline environment. *Environ. Microbiol.* **12**: 2965–76.
- Säwström,C., Anesio,M.A., Granéli,W., and Laybourn-Parry,J. (2007) Seasonal Viral Loop Dynamics in Two Large Ultraoligotrophic Antarctic Freshwater Lakes. *Microb. Ecol.* **53**: 1–11.
- Säwström,C., Granéli,W., Laybourn-Parry,J., and Anesio,A.M. (2007) High viral infection rates in Antarctic and Arctic bacterioplankton. *Environ. Microbiol.* **9**: 250–5.
- Schneider,G.F. and Dekker,C. (2012) DNA sequencing with nanopores. *Nat. Biotechnol.* **30**: 326–8.
- Seed,K.D., Lazinski,D.W., Calderwood,S.B., and Camilli,A. (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**: 489–491.
- Senčilo,A., Luhtanen,A.-M., Saarijärvi,M., Bamford,D.H., and Roine,E. (2014) Cold-active bacteriophages from the Baltic Sea ice have diverse genomes and virus-host interactions. *Environ. Microbiol.*
- Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84–7.
- Shtarkman,Y.M., Koçer,Z. a, Edgar,R., Veerapaneni,R.S., D’Elia,T., Morris,P.F., and Rogers,S.O. (2013) Subglacial Lake Vostok (Antarctica) accretion ice contains a diverse set of sequences from aquatic, marine and sediment-inhabiting bacteria and eukarya. *PLoS One* **8**: e67221.
- Siegert,M.J. (2005) Lakes beneath the ice sheet: The occurrence, analysis, and future exploration of Lake Vostok and other Antarctic subglacial lakes. *Annu. Rev. Earth Planet. Sci.* **33**: 215–245.
- Simon,C., Wiezer,A., Strittmatter,A.W., and Daniel,R. (2009) Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* **75**: 7519–7526.
- Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., et al. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* **406**: 151–9.
- Singh,P., Hanada,Y., Singh,S.M., and Tsuda,S. (2013) Antifreeze protein activity in Arctic cryoconite bacteria. *FEMS Microbiol. Lett.* 1–9.
- Skenneron,C.T., Imelfort,M., and Tyson,G.W. (2013) Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res.* 1–10.

- Skidmore, M., Anderson, S.P., Sharp, M., Foght, J., and Lanoil, B.D. (2005) Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Appl. Environ. Microbiol.* **71**: 6986–97.
- De Smet, W.H. and Van Rompu, E.A. (1994) Rotifera and Tardigrada from some cryoconite holes on a Spitsbergen (Svalbard) glacier. *Belgian J. Zool.* **124**: 27–37.
- Smith, L.C., Chu, V.W., Yang, K., Gleason, C.J., Pitcher, L.H., and Rennermalm, A.K. (2015) Efficient meltwater drainage through supraglacial streams and rivers on the southwest Greenland ice sheet. 1–6.
- Snyder, J.C., Bateson, M.M., Lavin, M., and Young, M.J. (2010) Use of Cellular CRISPR (Clusters of Regularly Interspaced Short Palindromic Repeats) Spacer-Based Microarrays for Detection of Viruses in Environmental Samples. *Appl. Environ. Microbiol.* **76**: 7251–7258.
- Solheim, H.T., Sekse, C., Urdahl, A.M., Wasteson, Y., and Nesse, L.L. (2013) Biofilm as an environment for dissemination of stx genes by transduction. *Appl. Environ. Microbiol.* **79**: 896–900.
- Somerfield, P.J., Cochrane, S.J., Dahle, S., and Pearson, T.H. (2006) Free-living nematodes and macrobenthos in a high-latitude glacial fjord. *J. Exp. Mar. Bio. Ecol.* **330**: 284–296.
- Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**: 237–66.
- Souchez, R., Petit, J.R., Tison, J.L., Jouzel, J., and Verbeke, V. (2000) Ice formation in subglacial Lake Vostok, Central Antarctica. *Earth Planet. Sci. Lett.* **181**: 529–538.
- Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K.E. (2008) Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res. Microbiol.* **159**: 349–357.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* **26**: 335–40.
- Stern, A., Mick, E., Tirosh, I., Sagy, O., and Sorek, R. (2012) CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* **22**: 1985–1994.
- Stern, A. and Sorek, R. (2011) The phage-host arms-race: Shaping the evolution of microbes. *Bioessays* **33**: 43–51.
- Stibal, M., Hasan, F., Wadham, J.L., Sharp, M.J., and Anesio, A.M. (2012) Prokaryotic diversity in sediments beneath two polar glaciers with contrasting organic carbon substrates. *Extremophiles* **16**: 255–265.
- Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.
- Suttle, C. a (2007) Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**: 801–12.
- Suttle, C. a (2005) Viruses in the sea. *Nature* **437**: 356–61.
- Takeuchi, N. (2001) Structure, formation, and darkening process of albedo-reducing

- material cryoconite on a Himalayan glacier: a granular algal mat growing on the glacier. *Arctic, Antarct. Alp. Res.* **33**: 115–122.
- Tan,D., Svenningsen,S. Lo, and Middelboe,M. (2015) Quorum Sensing Determines the Choice of Antiphage Defense Strategy in *Vibrio anguillarum*. *MBio* **6**: e00627.
- Tatusova,T., Ciufu,S., Fedorov,B., O’Neill,K., and Tolstoy,I. (2014) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* **42**: D553–9.
- Terns,M.P. and Terns,R.M. (2011) CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**: 321–7.
- Thingstad,T.F. and Lignell,R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**: 19–27.
- Thurber,R. V, Haynes,M., Breitbart,M., Wegley,L., and Rohwer,F. (2009) Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **4**: 470–83.
- Thureborn,P., Lundin,D., Plathan,J., Poole,A.M., Sjöberg,B.M., and Sjöling,S. (2013) A Metagenomics Transect into the Deepest Point of the Baltic Sea Reveals Clear Stratification of Microbial Functional Capacities. *PLoS One* **8**.
- Torres,R., Martin,M.C., Garcia,A., Cigudosa,J.C., Ramirez,J.C., and Rodriguez-Perales,S. (2014) Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**: 3964.
- Torsvik,T. and Dundas,I.D. (1974) Bacteriophage of *Halobacterium salinarium*. *Nature* **248**: 680–681.
- Trinks,H., Schröder,W., and Biebricher,C.K. (2005) Ice and the origin of life. *Orig. Life Evol. Biosph.* **35**: 429–445.
- Tung,H.C., Bramall,N.E., and Price,P.B. (2005) Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 18292–6.
- Tuorto,S.J., Darias,P., McGuinness,L.R., Panikov,N., Zhang,T., Häggblom,M.M., and Kerkhof,L.J. (2014) Bacterial genome replication at subzero temperatures in permafrost. *ISME J.* **8**: 139–49.
- Tveit,A., Schwacke,R., Svenning,M.M., and Urich,T. (2013) Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *ISME J.* **7**: 299–311.
- Twort,F.W. (1915) AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *Lancet* **186**: 1241–1243.
- Uetake,J., Naganuma,T., Hebsgaard,M.B., Kanda,H., and Kohshima,S. (2010) Communities of algae and cyanobacteria on glaciers in west Greenland. *Polar Sci.* **4**: 71–80.
- Van Valen,L. (1974) Molecular evolution as predicted by natural selection. *J. Mol. Evol.* **3**: 89–101.
- Vandamme,P. and Coenye,T. (2004) Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int. J. Syst. Evol. Microbiol.* **54**: 2285–9.

- Varin,T., Lovejoy,C., Jungblut,A.D., Vincent,W.F., and Corbeil,J. (2012) Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Appl. Environ. Microbiol.* **78**: 549–59.
- Varin,T., Lovejoy,C., Jungblut,A.D., Vincent,W.F., and Corbeil,J. (2010) Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnol. Oceanogr.* **55**: 1901–1911.
- Ventosa,A., de la Haba,R.R., Sánchez-Porro,C., and Papke,R.T. (2015) Microbial diversity of hypersaline environments: a metagenomic approach. *Curr. Opin. Microbiol.* **25**: 80–87.
- Vincent,W.F. (2010) Microbial ecosystem responses to rapid climate change in the Arctic. *ISME J.* **4**: 1087–1090.
- Vincent,W.F., Mueller,D.R., and Bonilla,S. (2004) Ecosystems on ice: The microbial ecology of Markham Ice Shelf in the high Arctic. *Cryobiology* **48**: 103–112.
- Wais,A.C., Kon,M., MacDonald,R.E., and Stollar,B.D. (1975) Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H. halobium*.
- Waldor,M.K. and Mekalanos,J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910–4.
- Waller,A.S., Yamada,T., Kristensen,D.M., Kultima,J.R., Sunagawa,S., Koonin,E. V, and Bork,P. (2014) Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* **8**: 1391–402.
- Wang,H., Yang,H., Shivalila,C.S., Dawlaty,M.M., Cheng,A.W., Zhang,F., and Jaenisch,R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**: 910–8.
- Wania,F. (1997) Modelling the fate of non-polar organic chemicals in an ageing snow pack. *Chemosphere* **35**: 2345–2363.
- Weinbauer,M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**: 127–81.
- Weinberger,A.D., Sun,C.L., Pluciński,M.M., Denef,V.J., Thomas,B.C., Horvath,P., et al. (2012) Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* **8**: e1002475.
- Weitz,J.S., Poisot,T., Meyer,J.R., Flores,C.O., Valverde,S., Sullivan,M.B., and Hochberg,M.E. (2013) Phage-bacteria infection networks. *Trends Microbiol.* **21**: 82–91.
- Wells,L.E. (2008) Cold-Active Viruses. In, *Psychrophiles: from Biodiversity to Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 157–173.
- Wells,L.E. and Deming,J.W. (2006) Brief report Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ. Microbiol.* **8**: 1115–1121.
- Whitman,W.B., Coleman,D.C., and Wiebe,W.J. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 6578–6583.
- Wilke,A., Glass,E., Bischof,J., and Braithwaite,D. (2013) MG-RAST Technical report and manual for version 3.3. 6–rev.

- Wilke,A., Harrison,T., Wilkening,J., Field,D., Glass,E.M., Kyrpides,N., et al. (2012) The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **13**: 141.
- Williams,G., Layman,K.L., and Stefan,H.G. (2004) Dependence of lake ice covers on climatic, geographic and bathymetric variables. *Cold Reg. Sci. Technol.* **40**: 145–164.
- Williamson,K.E., Corzo,K. a., Drissi,C.L., Buckingham,J.M., Thompson,C.P., and Helton,R.R. (2013) Estimates of viral abundance in soils are strongly influenced by extraction and enumeration methods. *Biol. Fertil. Soils* **49**: 857–869.
- Willis,M.J., Herried,B.G., Bevis,M.G., and Bell,R.E. (2015) Recharge of a subglacial lake by surface meltwater in northeast Greenland. *Nature*.
- Wommack,K., Ravel,J., Hill,R., and Colwell,R. (1999) Hybridization analysis of chesapeake bay virioplankton. *Appl. Environ. Microbiol.* **65**: 241–50.
- Wommack,K.E., Nasko,D.J., Chopyk,J., and Sakowski,E.G. (2015) Counts and sequences, observations that continue to change our understanding of viruses in nature. *J. Microbiol.* **53**: 181–92.
- Woodward,J., Murray,T., and McCaig,A. (2002) Formation and reorientation of structure in the surge-type glacier Kongsvegen, Svalbard. *J. Quat. Sci.* **17**: 201–209.
- Wright,A. and Siegert,M. (2012) A fourth inventory of Antarctic subglacial lakes. *Antarct. Sci.* 1–6.
- Wylie,K.M., Truty,R.M., Sharpton,T.J., Mihindukulasuriya,K.A., Zhou,Y., Gao,H., et al. (2012) Novel bacterial Taxa in the human microbiome. *PLoS One* **7**: e35294.
- Xiang,S., Yao,T., An,L., Xu,B., and Wang,J. (2005) 16S rRNA sequences and differences in bacteria isolated from the Muztag Ata glacier at increasing depths. *Appl. Environ. Microbiol.* **71**: 4619–27.
- Yilmaz,S., Allgaier,M., and Hugenholtz,P. (2010) Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**: 943–944.
- Yu,Z.-C., Chen,X.-L., Shen,Q.-T., Zhao,D.-L., Tang,B.-L., Su,H.-N., et al. (2014) Filamentous phages prevalent in *Pseudoalteromonas* spp. confer properties advantageous to host survival in Arctic sea ice. *ISME J.* **9**: 871–881.
- Zarsky,J.D., Stibal,M., Hodson,A., Sattler,B., Schostag,M., Hansen,L.H., et al. (2013) Large cryoconite aggregates on a Svalbard glacier support a diverse microbial community including ammonia-oxidizing archaea. *Environ. Res. Lett.* **8**: 035044.
- Zawierucha,K., Kolicka,M., Takeuchi,N., and Kaczmarek,Ł. (2014) What animals can live in cryoconite holes ? A faunal review. *J. Zool.* 1–11.
- Zeng,Y.X., Yan,M., Yu,Y., Li,H.R., He,J.F., Sun,K., and Zhang,F. (2013) Diversity of bacteria in surface ice of Austre Lovénbreen glacier, Svalbard. *Arch. Microbiol.* **195**: 313–322.
- Zeng,Y.-X., Yan,M., Yu,Y., Li,H.-R., He,J.-F., Sun,K., and Zhang,F. (2013) Diversity of bacteria in surface ice of Austre Lovénbreen glacier, Svalbard. *Arch. Microbiol.*

195: 313–22.

- Zhang,F., Wen,Y., and Guo,X. (2014) CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* **23**: 1–7.
- Zhang,S.H., Hou,S.G., Yang,G.L., and Wang,J.H. (2010) Bacterial community in the East Rongbuk Glacier, Mt. Qomolangma (Everest) by culture and culture-independent methods. *Microbiol. Res.* **165**: 336–345.
- Zhang,X.F., Yao,T.D., Tian,L.D., Xu,S.J., and An,L.Z. (2008) Phylogenetic and physiological diversity of bacteria isolated from Puruogangri ice core. *Microb. Ecol.* **55**: 476–488.
- Zhao,Y., Temperton,B., Thrash,J.C., Schwalbach,M.S., Vergin,K.L., Landry,Z.C., et al. (2013) Abundant SAR11 viruses in the ocean. *Nature* **494**: 357–360.
- Zuckermann,M., Hovestadt,V., Knobbe-Thomsen,C.B., Zapatka,M., Northcott,P.A., Schramm,K., et al. (2015) Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nat. Commun.* **6**: 7391.

9. Appendix

9.1 Appendix 1: Diversity and CRISPR Abundance in Public Metagenomes

30 publically available metagenomes were analyzed in the section 4.1 to assess the distribution of CRISPRs in natural environments. Data issued from this analysis is presented in Table 9.

Table 9. The 30 metagenomes used in the analyses in section 4.1 with the environment they came from, the temperature of the samples, the proportion of CRISPRs with each of the different software used and their weighted average percentage, the viral and bacterial diversity estimates (Shannon Indexes) and the study they were derived from.

Environment	Metagenome	T°	Proportion of CRISPRs			Weighted average percentage			Diversity (Shannon Index)		Reads assigned to viruses (%)	References
			CRASS	metaCRT	PILER	CRASS	metaCRT	PILER	Viral Div.	Bacterial Div.		
Human Gut	Human Gut 86	37	3.42134E-05	0.00018148	8.92524E-05	0.002477958	0.012441413	0.005420533	4.096375	10.137360	0.39	(Wylie et al., 2012)
	Human Gut 31	37	8.06267E-06	9.19144E-05	2.25755E-05				4.763205	9.876925	0.33	
	Human Gut 32	37	3.10229E-05	9.61708E-05	4.80854E-05				4.54525	10.305888	0.35	
Amazonian Soil	Amazonia Soil 57	28	2.07028E-05	0.000103514	2.76037E-05	0.00233617	0.013783404	0.004205106	4.304865	9.152988	0.20	(Mendes et al., 2014)
	Amazonia Soil 90	28	2.7665E-05	5.533E-05	1.38325E-05				4.122091	9.137084	0.19	
	Amazonia Soil 94	28	2.16519E-05	0.000259823	8.66076E-05				4.084359	9.075249	0.18	
Coral Reef	Timbebas	24	0	4.00711E-05	0	0.001126157	0.036712704	0.001126157	4.118599	9.561749	0.88	(Bruce et al., 2012)
	Pab	24	2.36703E-05	0.000418176	2.36703E-05				3.865421	9.354171	1.16	
	California	24	1.19394E-05	0.000620847	1.19394E-05				4.067922	6.835692	1.04	
Albufera	Albufera1	19.5	2.36911E-05	0.000280344	3.15881E-05	0.002200188	0.022001877	0.003813659	5.013849	10.049506	2.04	(Ghai et al., 2012)
	Albufera2	19.5	1.52486E-05	0.000167734	2.6685E-05				4.978668	10.304998	1.03	
	Albufera3	19.5	3.00879E-05	0.000210615	6.61933E-05				4.742663	9.973289	0.71	
Puget Sound	Puget Sound P1	11.5	1.00629E-05	0.00027673	0	0.000868346	0.022204855	0.000496198	4.810329	10.059123	0.71	(Port et al., 2012)
	Puget Sound P28	11.5	5.76239E-06	0.000230496	0				4.805967	10.007949	0.27	
	Puget Sound P5	11.5	9.21997E-06	0.000193619	9.21997E-06				4.796785	10.635706	1.74	
Arctic Soil	Arctic Soil 125	6	9.67436E-06	8.70692E-05	1.93487E-05	0.002047826	0.009946581	0.003218012	4.187622	9.827383	0.19	(Tveit et al., 2013)
	Arctic Soil 126	4	1.965E-05	9.82502E-05	1.965E-05				4.112746	9.817104	0.24	
	Arctic Soil 127	2	2.92657E-05	0.000109746	5.12149E-05				4.103029	10.069858	0.22	
Baltic sea wáter	Balctic Sea 1	4.1	5.38187E-05	0.001420814	1.61456E-05	0.004615588	0.063387411	0.003795039	4.39557	5.069984	0.23	(Thureborn et al., 2013)
	Balctic Sea 2	4.7	6.48048E-05	0.00071567	3.94464E-05				4.836351	4.611013	0.33	
	Balctic Sea 3	5.7	2.76347E-05	0.00023029	4.60579E-05				4.871031	5.504196	0.22	

Polar Microbial Mats	McMurdo	7.4	1.2009E-05	7.20539E-05	3.60269E-05				4.130426	9.586140	0.31	(Varin et al., 2010)
	WHI	0.9	2.38304E-05	0.000247241	0.000116173	0.004291052	0.024414604	0.011541449	4.640296	10.643713	0.36	
	MIS	1.8	7.78668E-05	0.000295894	0.00014016				4.562522	10.538262	0.26	
Antarctica Lake	Antarctica 104	2.71	2.39128E-05	0.000314546	2.39128E-05				5.679599	10.782557	0.69	(Lauro et al., 2011)
	Antarctica 106	0.42	1.6853E-05	0.000187069	5.05591E-06	0.002635238	0.032883185	0.001489482	4.543684	10.044361	4.62	
	Antarctica 115	1.2	3.77937E-05	0.000479816	1.6432E-05				4.691972	9.989399	4.38	
Glacial ice	Glacier1	-1	6.68045E-05	0.000289486	0.000181857				5.027731	8.202311	0.35	(Simon et al., 2009)
	Glacier2	-1	5.87073E-05	0.00026614	0.000144811	0.007059661	0.028981765	0.017370481	4.945844	7.351873	0.37	
	Glacier3	-1	8.8606E-05	0.000311893	0.0001843				4.859832	7.601995	0.34	

9.2 Appendix 2: Rarefaction Curves

Rarefaction curves for all the microbial and viral metagenomes studied in this study were generated in order to assess the different degree of OTU sampling among metagenomes.

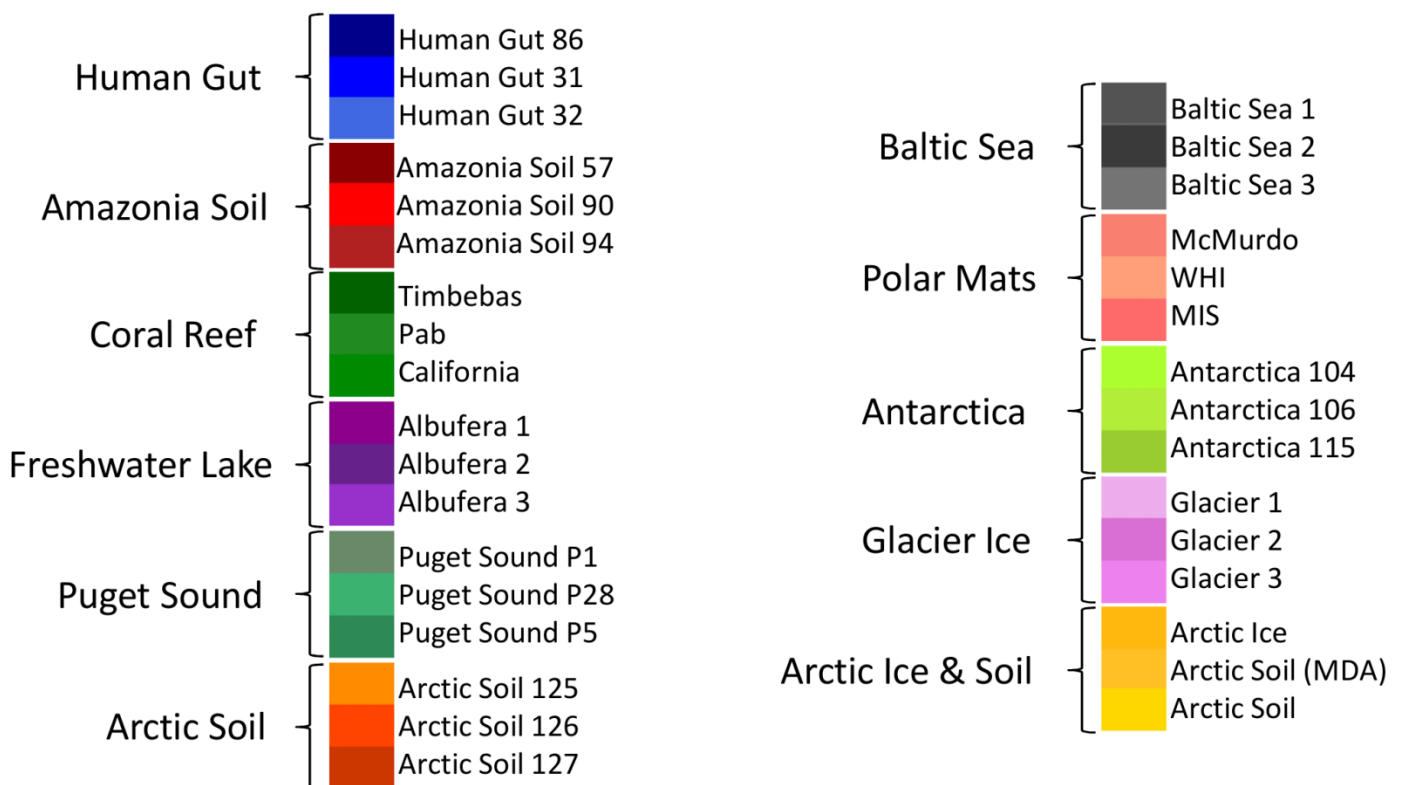
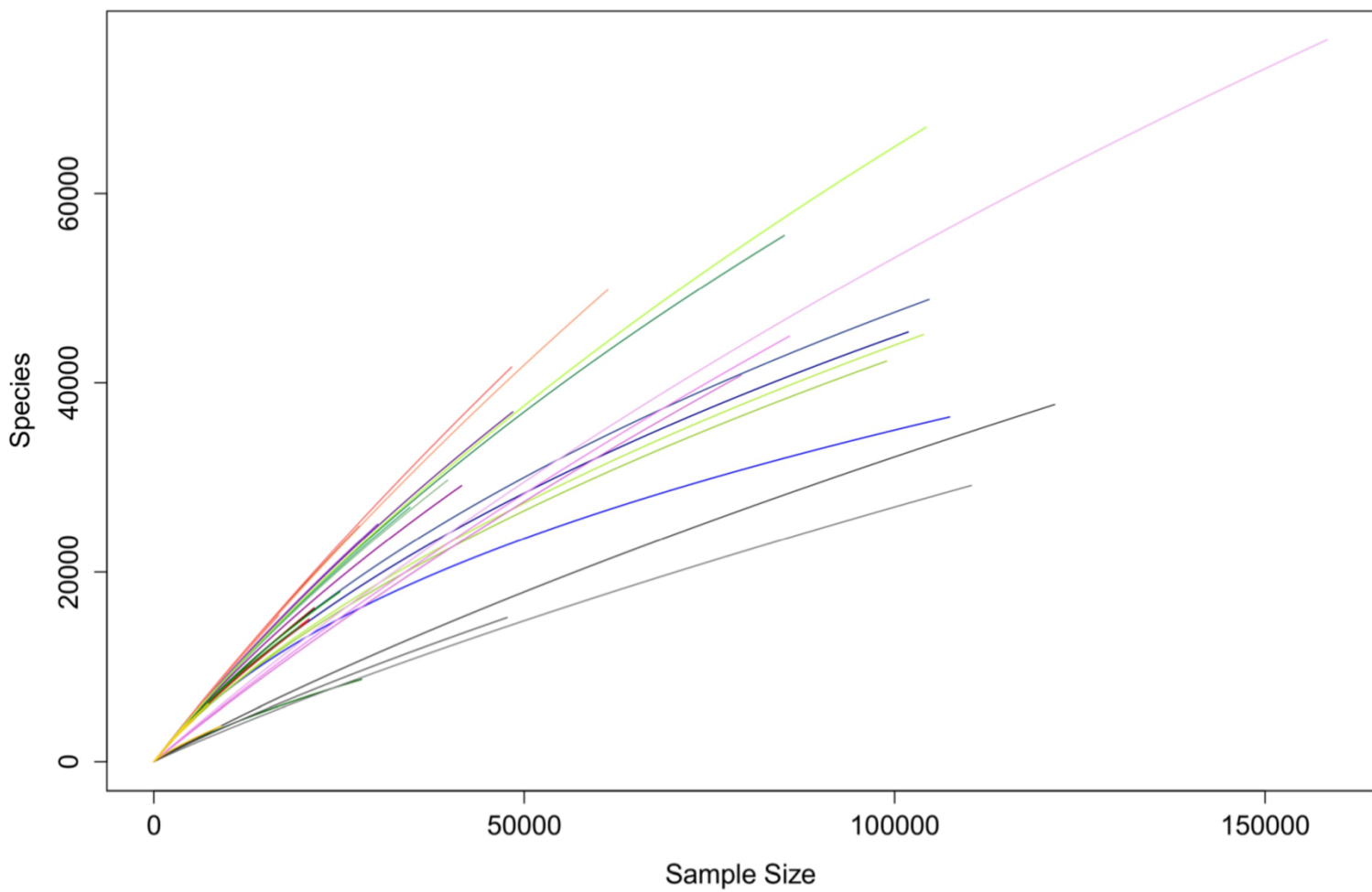


Figure 24. Rarefaction curves for all the microbial metagenomes used in this study. These plots depict the number of species (OTUs) as a function of the number of samples.

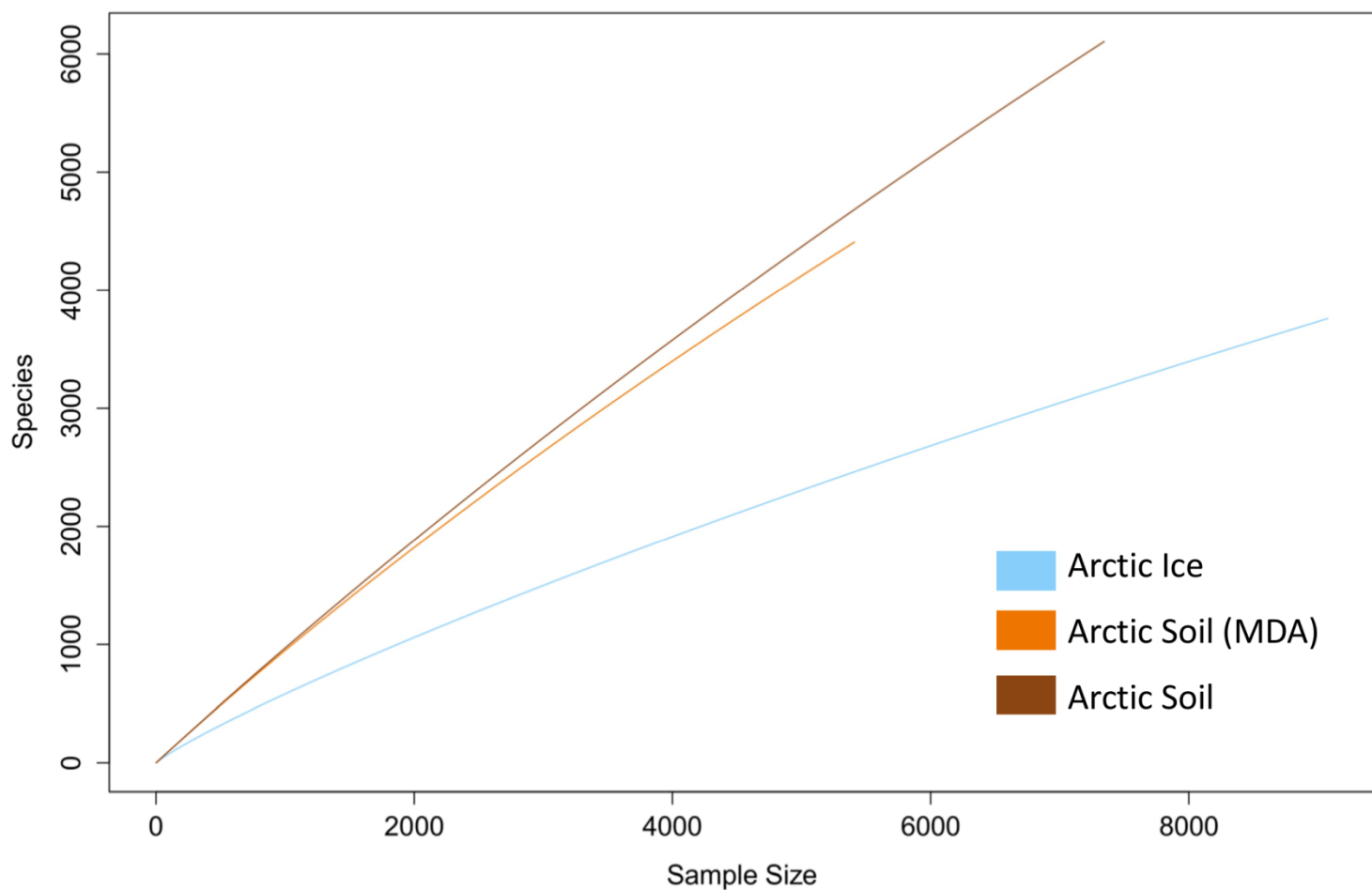


Figure 25. Rarefaction curves for the arctic ice and soil microbial metagenomes generated in this study. These plots depict the number of species (OTUs) as a function of the number of samples.

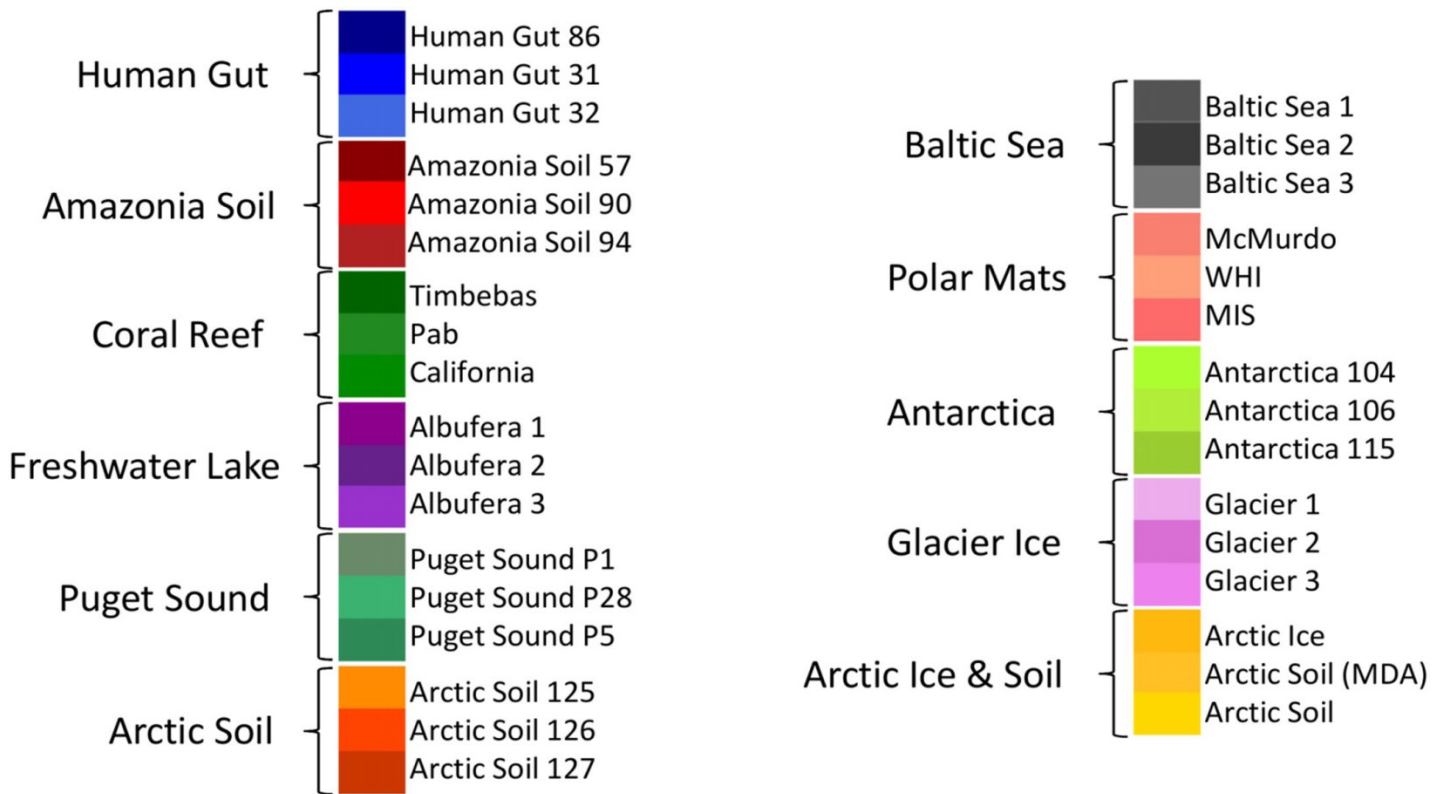
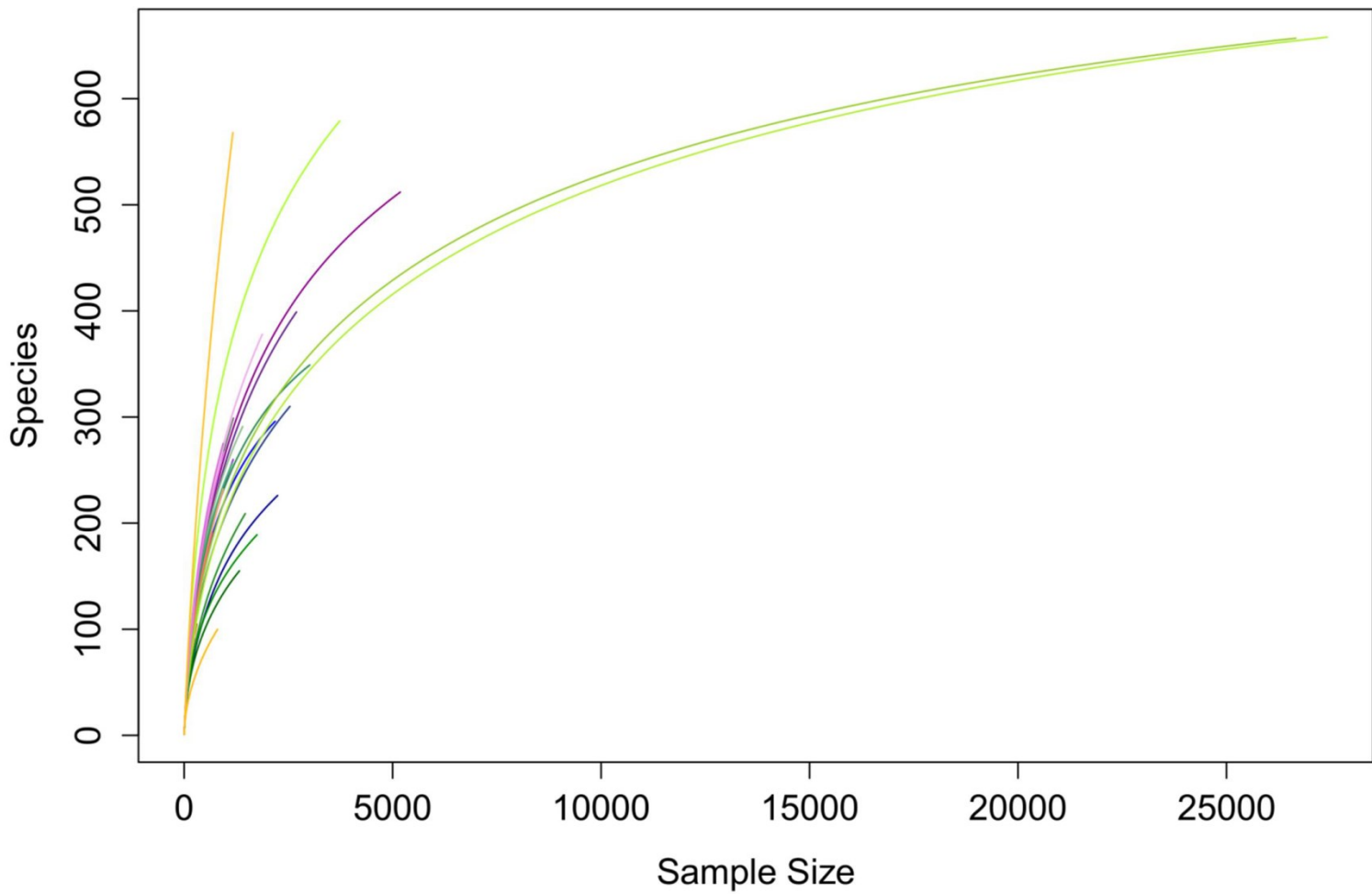


Figure 26. Rarefaction curves for all the viral metagenomes used in this study. These plots depict the number of species (OTUs) as a function of the number of samples.

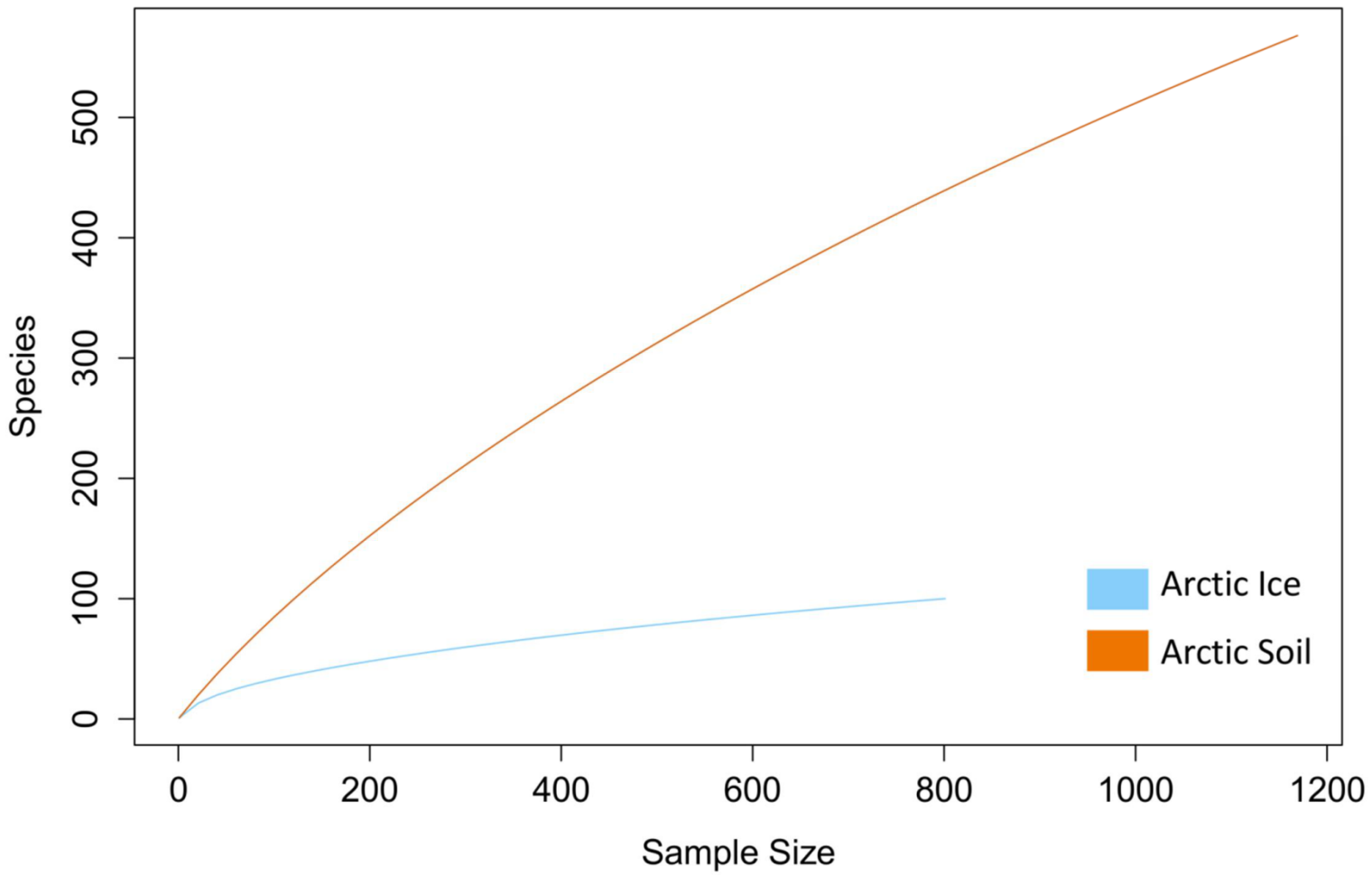


Figure 27. Rarefaction curves for the arctic ice and soil viral metagenomes generated in this study. These plots depict the number of species (OTUs) as a function of the number of samples.

9.3 Appendix 3: CRISPR Sequences in Arctic Ice and Soil Metagenomes

CRISPR sequences from arctic glacial ice and soil metagenomes were extracted using the software metaCRT. These CRISPRs were used to construct the CRISPR-based infection networks in section 4.2.2. The CRISPR sequences yielded by the software metaCRT are shown in Tables 10 (a, b and c).

Table 10. CRISPR sequences found by metaCRT in **a.** the glacial ice metagenome, **b.** the arctic soil metagenome, **c.** the arctic soil metagenome subjected to MDA. The sub-tables within a, b and c, show the name of the read the CRISPR was found in, the length of the CRISPR, how many CRISPRs within the read it has found, the position within the read of the CRISPR, the CRISPR sequence separated into the DRs and the spacers, the number of DRs and their consensus sequence and the average length of the DRs and the spacers. At the end of each table, the size of the metagenome, the total number of CRISPRs found and the time the software took to execute the search are shown.

Table 10.a **CRISPRs found by metaCRT in the microbial metagenome from arctic glacial ice**

SEQ: IECDFUC02H5308 length=221 xy=3232_3930 region=2 run=R_2013_08_13_14_34_32_		
Bases: 221		
Total CRISPR: 1		
CRISPR 1 Range: 40		
- 216		
POSITION	REPEAT	SPACER
-----	-----	
40	AAACACACACACACAAACACGCACA	CATACATACATACATACACACATACGCAC
94	ACACACACACATACACACACGCACA	CACGCGCACACGCTCATGC
138	AAACACACTCATACACACACATACA	CACACATACACACACATATACATACATAA
192	ACACATACACATACACACACACATA	
-----	-----	
Repeat-con	AAACACACACATACACACACACACA	
Repeats: 4	Average Length: 25	Average Length: 25

SEQ: IECDFUC02INGQ5 length=425 xy=3430_2559 region=2 run=R_2013_08_13_14_34_32_		
Bases: 425		
Total CRISPR: 1		

CRISPR 1 Range: 8 -
407

POSITION	REPEAT	SPACER
-----	-----	
	8 ACACACACACACGCACACAC	TTACTCTTACACATGTAAACACATATGCAAACCCCCACACACATACATCCACACAC AAACGCACACACAA
	98 ATACATAAACACACACAGAC	AAAACTGGCACACACACACATATACACACACACATACACACACAAAAC
	168 ACACACACACACACACAAAC	GTAAATAAACACACACATACACACACAAACATGCATGCACAT
	230 TTACACACACACACACACAC	ATAAGAGTAAACACACATACTCTCACACACTC
	282 ATACACACACAGACATACTC	TCACACACTCACACTCATACTATACATACACAAAAACACACAGAAGT
	348 AAACACACACATACACACAC	ACATTCACACACTCATAAC
	388 AAACACACACANACACACAA	
-----	-----	
Repeat-con	ATACACACACACACACAC	
Repeats: 7	Average Length: 20	Average Length: 43

SEQ: IECDFUC02H4OUI length=208 xy=3216_3128 region=2 run=R_2013_08_13_14_34_32_

Bases: 208
Total CRISPR: 1

CRISPR 1 Range: 11
- 185

POSITION	REPEAT	SPACER
----------	--------	--------

```

11 CAGACAACAGACAGACAGACAGACATAC   ATACAGACACTTTACAACAACAGA
63 CAGACAACAGACAGACAGACAGACAGAC   AGACATAGTAACAGACAGT
110 CAGACAACAGACAGACAGACAGACAGAC  AGACAGACAGACAGACAGAC
158 AGACAGACAGACAGACAGACAGACAGAC

```

```

-----
Repeat-con      CAGACAACAGACAGACAGACAGACAGAC
Repeats: 4      Average Length: 28           Average Length: 21

```

SEQ: IECDFUC02GC0UE length=185 xy=2491_2292 region=2 run=R_2013_08_13_14_34_32_

Bases: 185
Total CRISPR: 1

CRISPR 1 Range: 56
- 188

POSITION	REPEAT	SPACER
-----	-----	
	56 CACACACACACATACACACA	CATACACACGCAAACACAAATACACAAATGCGTAAAAAACT
	118 CACACACACACACACACT	CTCCTCCCCCANACCTCCCCANAAATAAATA
	169 CACACACACACACACGT---	
-----	-----	

```

Repeat-con      CACACACACACACACACA
Repeats: 3      Average Length: 20           Average Length: 36

```

SEQ: IECDFUC02H26A6 length=491 xy=3199_2080 region=2 run=R_2013_08_13_14_34_32_

Bases: 491

Total CRISPR: 1

CRISPR 1 Range: 243 - 403

POSITION	REPEAT	SPACER
-----	-----	
	243 ACGCCCACTCAGACACCGCC	ATCGGATAAATACTGCGCCCTATAGT
	291 ACCCATCACTCAGACACCACC	ATCGGATAAATACTGCGCCCTATAGT
	339 ACGCCATCCACTCAGACACCACC	GAGCGATTATACGCCAACA
	381 ACGCCCTCACTCAGACACCACC	
-----	-----	
Repeat-con	ACGCCATCACTCAGACACCACC	
Repeats: 4	Average Length: 23	Average Length: 23

SEQ: IECDFUC02HUH4Y length=335 xy=3100_3008 region=2 run=R_2013_08_13_14_34_32_

Bases: 335

Total CRISPR: 1

CRISPR 1 Range: 74

- 314

POSITION	REPEAT	SPACER
-----	-----	
	74 TGTCTATCTGTATGTCTGTAT	GTCTGTTTGTATGTCTGTA
	114 TGCATGTCTGTATGTCTGTAT	GTCTATTTGTATGTCTGTATGCATGTC
	162 TGTATGTCTGTATGTCTGTCT	CTCTGTCTGTCTCTCCGACAGTATGTCTACC
	214 TGTATGTCTGTATGTCTGTAT	GTTTGTATGTCTGAATGTC

	254 TGAATGTCTGTATGTCTGTAT	GTCTGTATGTATGTCTGTC
	294 TGTCTGTCTGTCTGTCTGTCT	
-----	-----	
Repeat-con	TGTATGTCTGTATGTCTGTAT	
Repeats: 6	Average Length: 21	Average Length: 23

SEQ: IECDFUC02HZ7MQ length=490 xy=3165_3088 region=2 run=R_2013_08_13_14_34_32_

Bases: 490
Total CRISPR: 1

CRISPR 1 Range: 317 - 476

POSITION	REPEAT	SPACER
-----	-----	
	GGGAGTCCGCCGCGTCCACCGGGGAGTCCG	
317	CCGCGTCC	CCAACGCCACGTCTCCACCG
	GGGAGTCCGCCGCGTCCGCCAGGGAGTCCG	
376	CCGCGTCC	GCCAGGGAGTCCGCCGCGTCCGCCA
	GGGCGTCCGCCGCGTCCGCCGGGAAGTCCG	
439	CCGCGTCC	
-----	-----	
Repeat-con	GGGAGTCCGCCGCGTCCGCCGGGAGTCCG	
Repeats: 3	CCGCGTCC	
	Average Length: 38	Average Length: 23

SEQ: IECDFUC02G8CQO length=442 xy=2848_1774 region=2 run=R_2013_08_13_14_34_32_

Bases: 442
Total CRISPR: 1

CRISPR 1 Range: 137 - 329

POSITION	REPEAT	SPACER
-----	-----	
137	CAAATGAACGCAATACGTCG	GATTGAAATCAGGGTCGAAC
177	CGATGCAAGGCAATACGTCG	ACGATGATGAACGGAAAGTGCAACCGAAA
226	CGATGCAAGGCAATACGTCG	GTTGAACGATGAGATCCGAC
266	GATTGCAAGGCAATACGTCG	AACAGAAGAGACGGAGGGTCGAAA
310	CGATGCAAGGCAATACGTCG	
-----	-----	
Repeat-con	CGATGCAAGGCAATACGTCG	
Repeats: 5	Average Length: 20	Average Length: 23

SEQ: IECDFUC02F1U6C length=480 xy=2364_1922 region=2 run=R_2013_08_13_14_34_32_

Bases: 480
Total CRISPR: 1

CRISPR 1 Range: 194 - 380

POSITION	REPEAT	SPACER
-----	-----	
194	TGTCTTTTTGTCTGTCTGTCTGTCTGT	CTGCCTGTTTGTCTGGCTCTGTCTA
250	TCTGTTTGTCTGGCTCTGTCTGTCTGC	CTGCCTGTTTGTCTGTCTGTCTGCCTCTCG
308	TTTGTTTGTCTGGCTCTTTCTGTCTGT	CTGTCTATCTGTTTGTCTG

```

354 TCTGTCTGTCTGGCTCTGTCTGTCTGTAT
-----
Repeat-con TCTGTTTGTCTGGCTCTGTCTGTCTGT
Repeats: 4 Average Length: 27 Average Length: 26

```

```

SEQ: IECDFUC02FMX82 length=501 xy=2194_2388 region=2 run=R_2013_08_13_14_34_32_
Bases: 501
Total CRISPR: 1

CRISPR 1 Range: 390 - 510
POSITION REPEAT SPACER
-----
390 CATCTTGATCAAACCTGCCT CGATCGTAGCCCGCGTAGGCACTACAACCC
440 CATCTTGATCAAACCTGCCT CAAATACATGAGGCCTACCGGCCCGAAGGAT
491 CATCTTGATCC-----
-----
Repeat-con CATCTTGATCAAACCTGCCT
Repeats: 3 Average Length: 20 Average Length: 30

```

Total sequences checked: 47626
 Total bases checked: 17811035
 Total sequences with predicted CRISPR: 10
 Time used: 4199 ms

Table 10.b

CRISPR found by metaCRT in the microbial metagenome from arctic soil

SEQ: IECDFUC02GPUED length=397 xy=2637_2451 region=2 run=R_2013_08_13_14_34_32_		
Bases: 397		
Total CRISPR: 1		
CRISPR 1 Range: 40 - 361		
POSITION	REPEAT	SPACER
-----	-----	
	40 GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	CGAAGCGCCGTAAAACTCGCACGCCAAGCATGTT
	111 GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	GCACGGTGACGGCGAGGGGCTCGCTATTGTTTCGAA
	182 GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	GTCAAGCGTAGACGTAAAAGGGCTAAGTAGTACAAT
	254 GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	ATGAAGCGGGATGTCGAAGACATAAAGGACCGCGTT
	326 GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	
-----	-----	
Repeat-con	GCTTCAATGTGGCCGCGGCTGATCAGCCGCGGAAGG	
Repeats: 5	Average Length: 36	Average Length: 35

Total sequences checked: 42078

Total bases checked: 14352412

Total sequences with predicted CRISPR: 1

Time used: 3229 ms

Table 10.c

CRISPR found by metaCRT in the microbial metagenome from arctic soil subjected to MDA

SEQ: IECDFUC02JCCYE length=255 xy=3714_0772 region=2 run=R_2013_08_13_14_34_32_		
Bases: 255		
Total CRISPR: 1		
CRISPR 1 Range: -11 - 89		
POSITION	REPEAT	SPACER
-----	-----	
	-11 -----CAAGAACA	CGAACGTACTAATAACCTTAAT
	31 CAACAAAAGCAACAAGAACA	CCAACAAGTACTACGGCAT
	70 CAACAAAAACAACAAGAACA	
-----	-----	
Repeat-con	CAACAAAAACAACAAGAACA	
Repeats: 3	Average Length: 20	Average Length: 20

SEQ: IECDFUC02JOURA length=520 xy=3856_2084 region=2 run=R_2013_08_13_14_34_32_		
Bases: 520		
Total CRISPR: 1		
CRISPR 1 Range: 94 - 252		
POSITION	REPEAT	SPACER
-----	-----	
	94 GAGGTTAAGCGTGGTCAAATTCGTCAG	ACCAATCAGAACGGAGAAGTCTTTCAGACGATTCTCCCC
	160 GAGGTTAAGCGTGGTCAGATTTGTCAG	GCCCGCCGGAAGGAGAAGTCGGTCAGTCCATTGCGGTC

	226	GAGGATAAGCGTGGTCAGATTCGTCAG	
-----		-----	
Repeat-con		GAGGTTAAGCGTGGTCAGATTCGTCAG	
Repeats: 3		Average Length: 27	Average Length: 39

Total sequences checked: 33241

Total bases checked: 13078370

Total sequences with predicted CRISPR: 2

Time used: 2994 ms

9.3 Appendix 4: Publications

Sanguino,L., Franqueville,L., Vogel,T.M., and Larose,C. (2015) Linking environmental prokaryotic viruses and their host through CRISPRs. FEMS Microbiol. Ecol.

Maccario, L., Sanguino, L., Vogel,T.M., and Larose,C. (in press) Snow and Ice ecosystems: Not so extreme. Research in Microbiology.



RESEARCH ARTICLE

Linking environmental prokaryotic viruses and their host through CRISPRs

Laura Sanguino, Laure Franqueville, Timothy M. Vogel and Catherine Larose*

Environmental Microbial Genomics, Laboratoire Ampère, CNRS UMR 5005, Ecole Centrale de Lyon, Université de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully, France

*Corresponding author: Environmental Microbial Genomics, Laboratoire Ampère, CNRS UMR 5005, Ecole Centrale de Lyon, Université de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully, France. Tel: +33-4-72-18-62-37; Fax +33-4-78-43-37-17; E-mail: catherine.larose@ec-lyon.fr

One sentence summary: We developed a method to recover data on virus-host interactions from environmental samples using CRISPR sequences. We applied this workflow to Arctic ice and soil metagenomes.

Editor: Julie Olson

ABSTRACT

The ecological pressure that viruses place on microbial communities is not only based on predation, but also on gene transfer. In order to determine the potential impact of viruses and transduction, we need a better understanding of the dynamics of interactions between viruses and their hosts in the environment. Data on environmental viruses are scarce, and methods for tracking their interactions with prokaryotes are needed. Clustered regularly interspaced short palindromic repeats (CRISPRs), which contain viral sequences in bacterial genomes, might help document the history of virus-host interactions in the environment. In this study, a bioinformatics network linking viruses and their hosts using CRISPR sequences obtained from metagenomic data was developed and applied to metagenomes from Arctic glacial ice and soil. The application of our network approach showed that putative interactions were more commonly detected in the ice samples than the soil which would be consistent with the ice viral-bacterial interactions being more dynamic than those in soil. Further analysis of the viral sequences in the CRISPRs indicated that *Ralstonia* phages might be agents of transduction in the Arctic glacial ice.

Keywords: virus-host interactions; CRISPRs; transduction; glacial ice

INTRODUCTION

Viruses outnumber prokaryotes and are the most abundant biological entities in a range of ecosystems (Fuhrman 1999; Williamson *et al.* 2013) and possibly in the entire biosphere. Through infection and microbial lysis, they can regulate microbial populations as well as generate and maintain microbial diversity (Thingstad 2000; Rodriguez-Valera *et al.* 2009), which in turn has an impact on food webs and biogeochemical cycles (Weinbauer 2004; Suttle 2007). Furthermore, viruses can be agents of horizontal gene transfer through transduction by packaging microbial genetic material into their capsid during

the lytic cycle and transferring it to another host (Canchaya *et al.* 2003). The rate of gene transfer by transduction in oceans has been calculated to be 10^{24} genes per year (Jiang and Paul 1998; Rohwer and Thurber 2009), which was concluded to be an important mechanism for microbial adaptation. Examples of specific viral transduction events have also been reported: marine viruses carry a range of carbon metabolism genes (Lindell *et al.* 2005; Hurwitz, Hallam and Sullivan 2013); transduction of toxin genes has been seen to transform bacteria into pathogens, thus extending the niches of their hosts (Waldor and Mekalanos 1996). Viruses might be drivers of microbial adaptation to

Received: 19 January 2015; Accepted: 16 April 2015

© FEMS 2015. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

environmental conditions by enabling the acquisition of whole segments of genomic information that can recombine or integrate into the host genome.

In order to evaluate the impact of viruses on microbial adaptation and possibly evolution, we need further information about environmental virus-host interactions. However, at least 50% of the sequences retrieved from publically available viral metagenomes have no similarities with specific viral sequences in databases (Rosario and Breitbart 2011; Ray et al. 2012), rendering the study of environmental viruses through high-throughput methods challenging. Moreover, relative host-species specificities in uncultured bacteria and viruses remain unknown. Clustered regularly interspaced short palindromic repeats (CRISPRs), which play a role in the defense of prokaryotes against viruses, have been suggested as a potential tool to identify virus-host interactions in complex environmental communities (Anderson, Brazelton and Baross 2011). The system is heritable and widespread (Fineran and Charpentier 2012) and nearly all archaea and about half of the bacteria have been found to have CRISPR-Cas systems (Jansen et al. 2002; Terns and Terns 2011). CRISPRs are made of a series of direct repeats (DRs) of microbial origin interspaced by short sequences called spacers of viral origin (among others) (Mojica et al. 2005). After an infection, short sequences from the virus (proto-spacers) are inserted into the CRISPR. These short specific sequences enable the host to recognize the virus and protect itself against future infections (Barrangou et al. 2007) by silencing the foreign DNA through the Cas enzymatic machinery (Terns and Terns 2011). In order to test whether we could link viruses and their hosts through CRISPRs, we used Arctic glacier ice and soil as model ecosystems. Studies about viruses in polar environments support the idea of high viral diversity (López-Bueno et al. 2009), high infection rates (Såwström et al. 2007; Bellas et al. 2013) and broad host ranges (Anesio et al. 2007; Colangelo-Lillis and Deming 2013) that would enable them to infect different bacteria within the same environment and possibly bacteria from different environments (Sano et al. 2004; Anesio et al. 2007; Bellas and Anesio 2013). These characteristics become especially relevant when studying viral transduction and explain why this environment is an interesting model ecosystem for this study. Furthermore, we hypothesize that an environment with a high diversity of viruses with broad host ranges would harbor a higher number of unique CRISPR sequences. This hypothesis is based on the idea that to protect itself from a viral infection, a microorganism has to have a spacer for each different virus that could (or did) infect it (Barrangou et al. 2007). If the diversity of viruses is high and they can infect several different microbes, the number of spacer combinations in CRISPR arrays would also be high, giving place to an increase in the number of different CRISPRs. Therefore, we hypothesize that the relative number of unique CRISPRs found in a metagenome could be an indicator for viral diversity and for the dynamism of the interactions with their hosts.

We developed a method to construct infection networks using spacers and DRs outsourced from CRISPR sequences from two different microbial and viral metagenomes. We applied this workflow to Arctic ice and soil metagenomes. This workflow identified putative virus-host interactions and provided the basis for comparing soil and ice bacterial/viral interaction dynamics. Gene transfer events were also evaluated in an effort to further define the nature of these interactions in Arctic ice.

MATERIAL AND METHODS

Study site

Ice samples were recovered from Kongsvegen glacier in Svalbard (N78°45'19.8", E13°20'11"), Norway, in the spring of 2007. A 15 m ice core was extracted from the accumulation area of the glacier, at an altitude of 670 m over sea level, using a manual drill (PICO corer) without any drilling fluids. The ice was stored in clean, sterile bags and frozen (−20°C) until further use. The core was composed of ice and firn layers; only ice layers were used in this study. On the other hand, soil was collected at the outskirts of Ny-Alesund (N78°45'20.16", E13°20'11.4") and was also stored frozen (−20°C). The soil was covered by snow and had a lichen crust over it. Samples were recovered from the first 5–10 cm and were found to be very rich in clay and organic matter.

Filtration, DNA extraction and sequencing

Microbial and viral DNA from ice

A total of 9 l of ice was melted at room temperature and filtrated simultaneously through 10 µm 47 mm (Millipore) filters and through 0.2 µm 47 mm (Millipore) to recover the bacterial fraction. Filtrate was then filtered through 0.02 µm 25 mm (Anodisc, Anopore Inorganic Membranes, Whatman) to recover viral fraction. The 0.02 µm filters were left overnight in PBS (without Ca or Mg, autoclaved and filtered through 0.02 µm) and then put in the shaker for 2 h before discarding the filter and proceeding with the extraction protocol. To assess the presence of viral particles and/or microbial cells, aliquots of the samples were filtered through 0.02 µm filters, stained with SYBER Gold (Life Technologies) and examined through epifluorescence microscopy as described by Thurber et al. (2009).

Viral DNA was extracted following the protocol from Thurber et al. (2009), yielding 0.502 ng µl^{−1} of DNA based on Qubit DNA HS fluorometric quantitation (Life Technologies). Multiple displacement amplification (MDA) was performed for viral DNA (illustra GenomiPhi HY DNA Amplification Kit, GE Healthcare Life Sciences, NJ, USA) following manufacturer instructions in order to have enough DNA for sequencing. DNA was then cleaned with GeneClean Turbo Kit (MP Biomedicals, France). Microbial DNA was extracted from two 10 µm and two 0.2 µm filters as described by (Larose et al. (2010)), DNA concentrations were of the order of 2 ng µl^{−1} and samples were stored at −20°C until further use. DNA from viral and microbial extractions was sent for sequencing using GS FLX System (Roche-454 Life Science, Beckman Coulter Inc., MA, USA).

Microbial and viral DNA from soil

A total of 900 g of soil was used for DNA extraction, following a modified protocol from Thurber et al. (2009). Briefly, 20 ml of PBS (without Ca or Mg in order to decrease the adherence of cells to soil particles, autoclaved and filtered through 0.02 µm filters) was added to every 50 g of soil and then blended with a Waring blender. The mixture was then centrifuged for 10 min at 2500 g and the supernatant was retrieved with sterile syringes. The supernatant was filtered through 0.45 µm 47 mm filters (Millipore) and then through 0.2 µm 47 mm filters. Microbial DNA was recovered from the filters while viral DNA was recovered from the filtrate. Microbial extraction of soil was done along with microbial extraction of ice as described by Larose et al. (2010). The extraction yielded 28 ng µl^{−1} of microbial DNA. Examination of the viral concentrate through epifluorescence microscopy was done

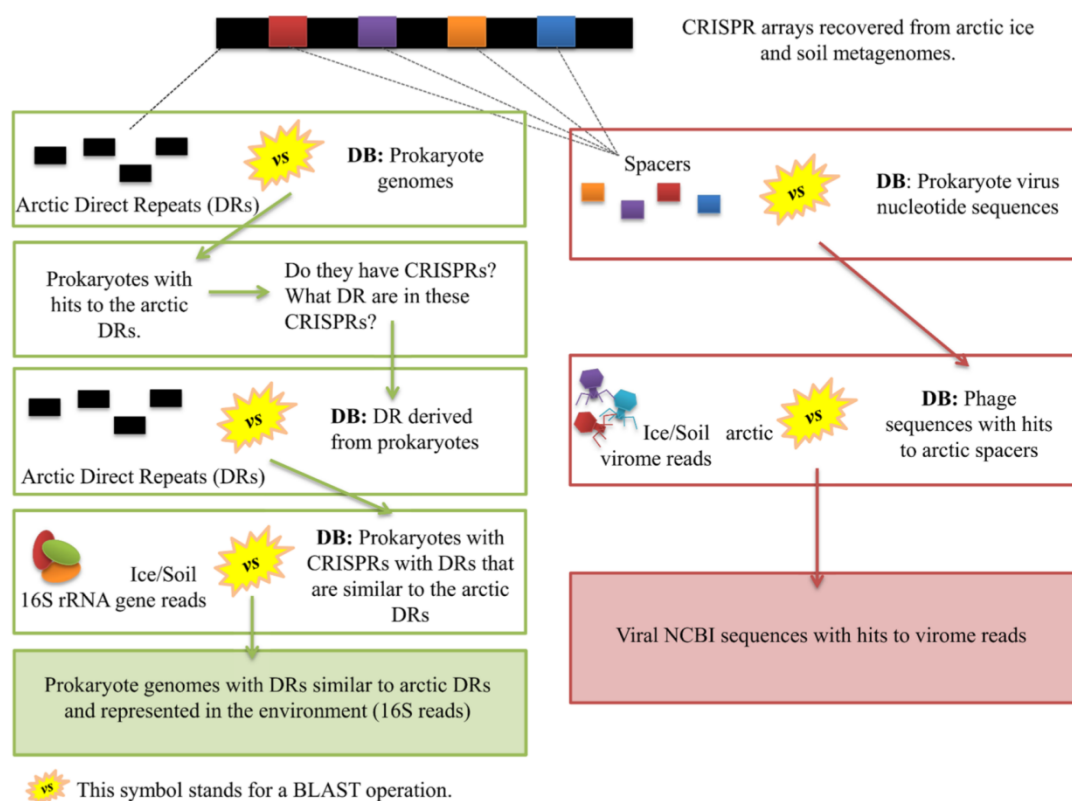


Figure 1. Diagram explaining the workflow used to identify the putative origin of the DRs (left) and the putative origin of spacers (right). This information was used to produce the networks linking phages to their hosts (Fig. 2). DB stands for database.

as described above. Viral DNA was extracted following the protocol from Thurber *et al.* (2009), DNA concentrations were $3.83 \text{ ng } \mu\text{L}^{-1}$, and the samples were stored at -20°C until further use.

16S rRNA sequencing from ice and soil

The 454 Pyrotag sequencing was done for regions V4 to V6 from the 16S rRNA genes from soil and ice microbial DNA extracts. Barcoded amplicons were prepared using MID-primers; each 50 μL PCR reaction contained 37 μL water, 5 μL Titanium Taq PCR Buffer, 1 μL of each MID-primer pair, 1 μL of dNTP Mix, 1 μL Titanium Taq DNA Polymerase (Clontech) and 4 μL (10 ng) of template DNA. The PCR cycling conditions were: 1 min at 95°C for denaturation followed by 30 cycles of 30 s denaturation at 95°C , 3 min annealing and elongation at 68°C and one cycle of 3 min at 68°C . The amplified DNA was size selected by excising the appropriate gel bands and purified using Illustra GFX PCR DNA and gel band purification kit. The DNA was quantified using Qubit DNA HS assay (Life Technologies). Amplicons were pooled in equimolar concentrations. The amplicon libraries were sequenced by Beckman Coulter Inc. (MA, USA) using 454 platforms. Sequencing data were analyzed following QIIME pipeline (Caporaso *et al.* 2011) for *de novo* OTU picking and diversity analysis using 454 data. Briefly, OTUs were picked based on sequence similarity and taxonomy was assigned through uclust consensus taxonomy classifier.

CRISPR infection network development

CRISPRs were browsed in the ice and soil metagenomes using metaCART (Rho *et al.* 2012). It was seen that the assembly of the

metagenomes into contigs significantly reduced the number of CRISPRs found; thus, only the individual reads were used for this purpose. The resulting CRISPR arrays were decomposed into a DR file and a spacer file. The workflow followed is summarized in Fig. 1. Ten CRISPRs were found in the ice metagenome while one CRISPR was found in the soil metagenome (Fig. 2). There were no repeated CRISPR sequences.

DRs have been seen to be conserved among taxa (Mojica, Ferrer and Juez 1995; Kunin, Sorek and Hugenholtz 2007), so they could serve to identify the organisms to whom the different CRISPR arrays belonged to by blasting the DR against an archaeal and bacterial genome database. However, there is also the possibility that a DR is shared among different bacteria. The database we used was a GenBank database (5193 full genomes), downloaded from their site in May 2014. The organisms retrieved through the blast of the DRs against the database were considered candidate sources of the CRISPRs (default stringency parameters: maximum *e*-value 10). We then verified if these candidate bacteria had CRISPR arrays in their genomes and made a database of the DR found in these CRISPRs. The DRs we had retrieved from our metagenomic CRISPRs were blasted against this database (default stringency parameters: maximum *e*-value 10). This left us with a list of prokaryotic genomes with CRISPRs whose DRs were similar to the DR we had outsourced from our metagenomic CRISPRs. We then checked to see if these prokaryotic genomes were represented in our 16S rRNA gene libraries by doing a blast of the 16S rRNA gene sequences against the prokaryotic genomes we had left at this point as putative originators of the metagenomics CRISPRs (97% identity over a minimum of 200 bp). While we had no hits for the soil, there were 47 genomes represented in our 16S rRNA gene library that had

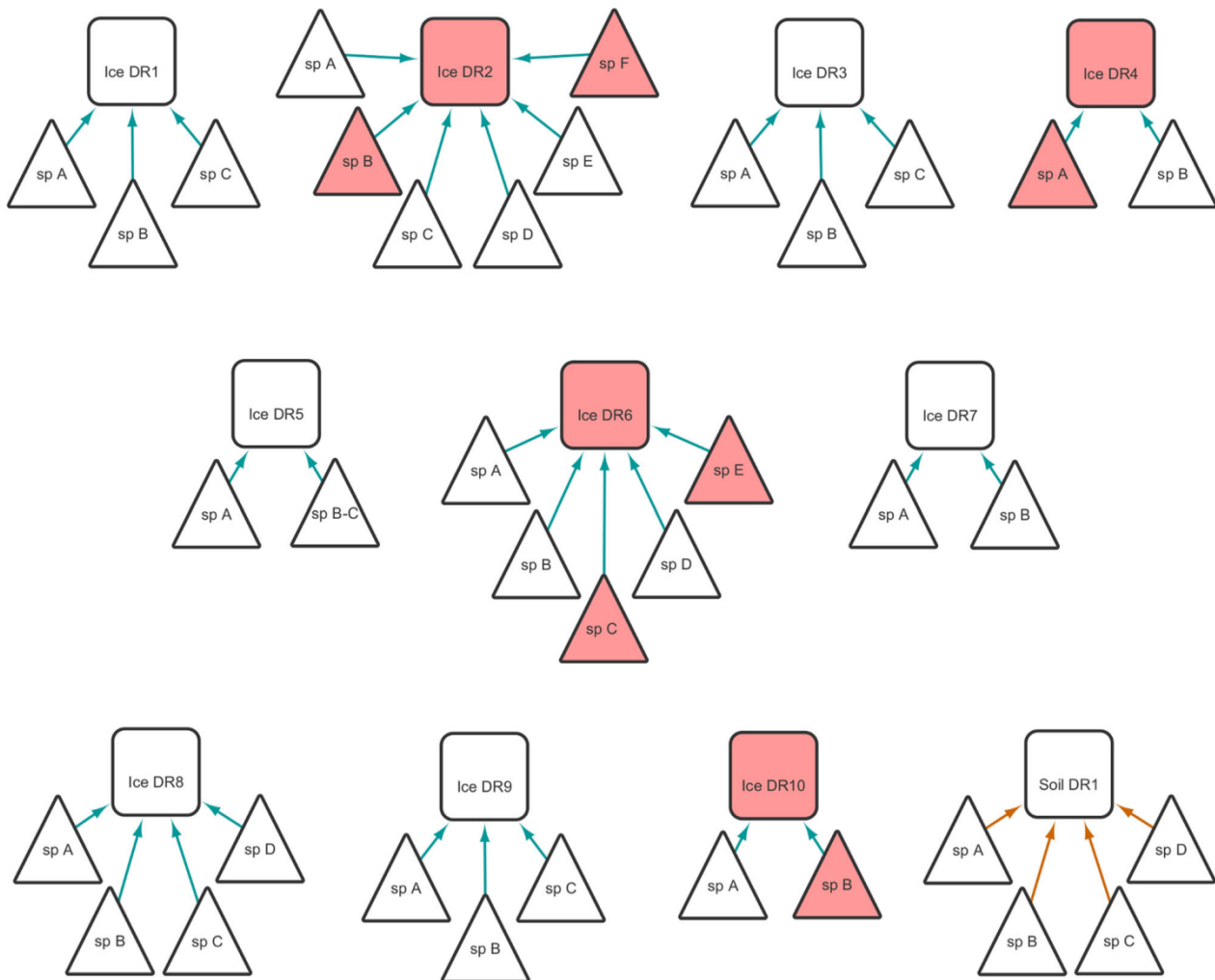


Figure 2. Set of networks representing the CRISPRs found in ice (blue arrows) and soil (brown arrows). Rectangles represent the DRs and triangles represent the spacers. Where identification of the microbial origin of the DR and the viral origin of the spacer was possible, nodes are shaded in red.

CRISPRs with DR similar to the DR of five different CRISPRs from the ice metagenomes. These 47 genomes belonged to 11 different prokaryotic species spread through six genera.

The spacers are sequences that can belong to viruses, and therefore, we used them to putatively identify potential viral infectors of microbial cells. Given that these are small viral DNA fragments, the sequences might be found in more than one virus. Spacers outsourced from the CRISPRs from the ice were blasted against a prokaryotic viruses' GenBank database sourced from NCBI in May 2014 (12,954 nucleotide sequences). Given that we were unable to identify the microbial source of the DR for the soil samples, only the spacers from the ice samples were analyzed. The viruses recovered (default stringency parameters: maximum e -value 10) were considered to be putative sources of the spacers. In a similar way as for the DR, we checked if these viruses were represented in our metagenomes. Since there is no analogous sequence to the 16S rRNA gene for viruses, we decided to do a blast of the viral metagenomes against the putative sources (maximum e -value 9×10^{-10} to account for the small size of the database). A total of 15 viruses were represented in our data and were considered as potential sources of the spacers from the ice samples, outsourced from four different CRISPRs. The 15 viruses belonged to one of these

seven types of viruses: *Mycobacterium* phage, *Cellulophaga* phage, Cyanophage, *Synechococcus* phage, *Pseudomonas* phage, environmental Halophage and *Geobacillus* virus.

Finally, the infection network was built using the software Cytoscape (Saito et al. 2012). The four CRISPRs (I to IV, Fig. 3) for which we had identified both the origin of the spacers and of the DR were represented in networks (Fig. 3).

Searching for transduction events

Transduction can take place through two different mechanisms. Generalized transduction is when a virus encapsidates pieces of microbial DNA instead of its own genome. On the other hand, when a temperate virus takes an adjacent piece of microbial DNA along with its viral genome, it is called specialized transduction. Thus, sequences sharing microbial and viral information could stand as a hallmark for specialized transduction; therefore, we looked for such hallmarks in our viral contigs. Viral 454 reads were assembled into contigs using the software provided by Roche Diagnostics Corporation, GS De Novo Assembler. Sequence redundancy had been previously reduced using CD-HIT-454 (Fu et al. 2012). Assembly of the sequences from the ice yielded 689 contigs, with a mean length of 1423 bp. In the

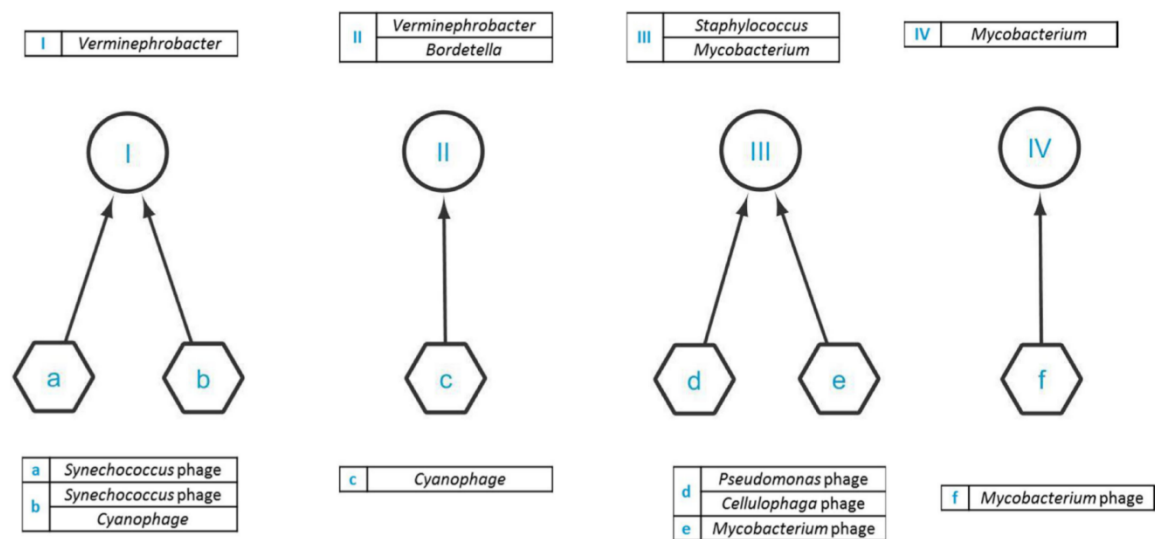


Figure 3. Networks with both the origin of the DR and at least one related spacer were putatively identified. Above and below each network, legends show the identified prokaryotes (roman numerals) and the identified viruses (letters).

case of viral soil sequences, assembly yielded 2076 contigs with a mean length of 711 bp. GenBank protein sequences for mercury resistance, cold shock and efflux pumps genes were downloaded from NCBI in May 2014. Each of these sets of sequences was blasted against the ice and the soil viral contigs. Contigs with hits to the protein sequences (maximum blast e -value 9×10^{-10}) were taken as contigs with microbial DNA within them. In order to test if these contigs also had viral DNA within them, we blasted them against the prokaryotic virus database. The contigs with hits to the prokaryotic virus database (maximum blast e -value 9×10^{-10}) were taken as possible hallmarks of transduction since they shared both microbial and viral information. The viral part of the contig enabled us to identify the possible transduction agents.

RESULTS

CRISPR infection network

The aim of this study was to develop a method to use CRISPRs to decipher virus-host interactions in complex environments. Our CRISPR-based identification workflow (explained in detail in the section 'Material and Methods') was used to determine the possible bacterial and viral origin of DRs and spacers, respectively. CRISPRs extracted with the help of the software metaCRT were then analyzed through our bioinformatics pipeline. The CRISPRs from the ice (blue arrows) and soil (brown arrows) microbial metagenomes were used to define possible bacterial hosts (rectangles representing the DRs from the bacteria) and the infecting viruses (triangles) (Fig. 2). While this figure depicts the interactions that could be taking place in the environment, this network is a simplified diagram of Arctic virus-host interactions as each DR could be common to several bacteria and every spacer could be common to different viruses. So while the network provides the possible connection between DRs and spacers from ice and soil, the number of putative bacteria and viruses involved is unknown. For further analysis, the networks (shaded in red in Fig. 2) with a DR that was found in the bacterial genome database (5193 full genomes) and with a spacer also found in the viral

database were recreated with their possible bacterial and viral identities (Fig. 3). Thus, circles stand for the microbial host and hexagons stand for the microbial viruses (Fig. 3). The roman numerals within the circles relate to the legends above the network, where the putative bacteria identified with the DRs are listed. The lowercase letters within the hexagons relate to the legends below the network, where the putative viruses identified with the spacers are listed. An added stringency to the data presented in Fig. 3 was the requirement that the putative bacteria from the 5193 full genomes had a matching 16S rRNA gene sequence in the pyrotags sequenced from our ice and soil samples. The networks in Fig. 3 belong exclusively to ice virus-host interactions as none of the putative soil bacteria matched the 16S rRNA sequences in our pyrotag data from the Arctic soil.

The genus *Verminephrobacter* was identified as a possible microbial host in I and II (Fig. 3). In the case of II, the microbial host could also be *Bordetella*. However, viruses infecting I and II were identified to be Cyanobacterial phages. While three *Bordetella* phage genomes were present in the prokaryotic viruses' database, there were no known *Verminephrobacter* phages. Thus, viruses a, b and c might be *Verminephrobacter* phages, although their closest match in the viral database are Cyanobacteria phages. Microbial host III was identified as a member of either *Staphylococcus* or *Mycobacterium*. Microbial host IV was also identified as a *Mycobacterium*; congruently viruses e and f were identified as *Mycobacterium* phages. Phage d was assigned to either *Cellulophaga* phage or *Pseudomonas* phage. Further interrogation of the ice metagenomes through MG-RAST (Meyer et al. 2008) confirmed the assignment of reads to all identified hosts. Likewise, viruses identified from the spacers were represented in the viral ice metagenomes based on Metavir's (Roux et al. 2011) contig and/or read assignments, supporting data output from the CRISPR network system.

Potential transduction agents in ice and soil

In specialized transduction, the virus encapsidates its own genome with part of the host microbial genome (Morse,

Table 1. Hybrid contigs with both microbial and viral sequences, taken as hallmarks of specialized transduction. **a.** Contigs with cold shock protein sequences and their viral assignment. **b.** Contigs with mer genes' protein sequences and their viral assignment. **c.** Contigs with efflux pumps' protein sequences and their viral assignment.

Contig	GenBank accession	Definition
a. COLD SHOCK PROTEINS		
contig00022	AB259123.2	Ralstonia phage
contig00022	KF887906.1	Ralstonia phage
contig00022	AB434711.1	Ralstonia phage
contig00006	AB276040.1	Ralstonia phage
contig00500	GU396103.1	Aeromonas phage
contig00157	GU396103.1	Aeromonas phage
b. mer GENE PROTEINS		
contig00022	AB259123.2	Ralstonia phage
contig00022	KF887906.1	Ralstonia phage
contig00022	AB434711.1	Ralstonia phage
c. EFFLUX PUMP PROTEINS		
contig00006	AB276040.1	Ralstonia phage
contig00007	JN662425.1	Burkholderia phage
contig00007	FJ937737.2	Burkholderia phage
contig00007	JX104231.1	Burkholderia phage
contig00022	AB259123.2	Ralstonia phage
contig00022	KF887906.1	Ralstonia phage
contig00022	AB434711.1	Ralstonia phage
contig00035	AB828698.1	Ralstonia phage
contig00035	GU911303.1	Burkholderia phage
contig00035	J01735.1	Bacteriophage lambda
contig00035	AF234172.1	Enterobacteria phage
contig00035	AF234173.1	Enterobacteria phage
contig00157	GU396103.1	Aeromonas phage

Lederberg and Lederberg 1956; Frost et al. 2005). If these events had taken place in the microbial communities we sampled in the Arctic, we would expect to find viral genomes with fragments of microbial DNA. In order to look for traces of specialized transduction events in our viral metagenomes, we selected a set of genes that we thought could be of adaptive interest to the Arctic microbial community. As the Arctic ice and soil are subjected to low temperatures and mercury from the snowpack (Larose et al. 2013), we selected cold shock genes and mercury resistance genes as traits that might be an asset for microorganisms living in the Arctic. Additionally, we selected a set of genes, such as efflux pumps genes, that might provide an adaptive advantage regardless of the ecosystem. We browsed our assembled viral metagenomes for hybrid contigs and subsequently identified the transduction agent through the viral part of the contig. The viruses that could be potential transducing agents for the three selected genes are presented in Table 1 a, b and c. No soil contigs contained both a viral and microbial marker based on our stringency parameters (see the section 'Material and Methods'). While in the case of the ice metagenome, there were six contigs that had putative viral and microbial markers matching those in the microbial and the viral databases. *Ralstonia* phage was identified as a transducing agent for all three microbial proteins studied here (Table 1 a, b and c).

DISCUSSION

The impact of prokaryotic viruses on the ecology and evolution of microbial communities is tightly linked to their infectivity

potential and their host range (Weinbauer and Rassoulzadegan 2003; Koskella and Meaden 2013). In order to address their impact, we need to link viruses to their microbial host(s) in the environment. The aim of this paper was to demonstrate the potential of using CRISPRs as a tool to link viruses with their hosts in environmental samples. Given that this demonstration is a proof of concept, we were quite stringent in our workflow in order to clearly show both the application to unknown viruses and bacteria as well as with bacterial and viral sequences of putative origin. Since CRISPRs contain sequences derived from bacteria and archaea (in the DRs) (Jansen et al. 2002) and spacers from viruses, we initially constructed the networks between DRs and spacers based on identified CRISPRs from the different metagenome datasets that we sequenced from Arctic glacial ice and soil (Fig. 2). With similar quantity of metagenomic sequences from ice and soil (42 078 reads for soil and 47 626 reads for ice), we found more CRISPRs in ice than in soil and all CRISPR sequences found were unique. To take into account the possible influence of MDA on the number of CRISPRs found, a control (soil DNA subject to MDA) was evaluated along with the other samples. Two CRISPRs were found in the MDA control compared to the one CRISPR found in the soil (not subjected to MDA), so the MDA did not significantly change the number of CRISPRs detected. Further work on the environmental aspects of CRISPRs needs to be carried out to evaluate the different CRISPR abundances in different environments and its consequences. However, we hypothesize that a higher number of unique CRISPR sequences in a metagenome might represent higher viral diversity and interactions with their hosts through the infection of different microorganisms. Thus, the relatively more abundant CRISPRs in ice might be an indication of the high viral diversity, infectivity rates and broad host ranges as suggested in the literature (Anesio et al. 2007; Säwström et al. 2007; López-Bueno et al. 2009). When we restrained the possible bacterial and archaeal hosts to those that had their genomes fully sequenced and reported in the NCBI database and to the viruses represented in the NCBI viral sequences, the number of networks was significantly reduced (Fig. 3). These prokaryotic genomes were used to identify potential originators of the DRs from the CRISPRs found in the ice and soil metagenomes, but only if similar 16S rRNA genes were also identified in the pyrotag libraries obtained for each environmental sample. Viral databases are, for the moment, very scant, making it difficult to identify viruses through DNA sequences such as spacers, especially in the environment. However, despite working with a relatively unexplored ecosystem whose inhabitants may not be well represented by the fully sequenced genomes in the database, we were able to describe several virus-host interactions taking place in the Arctic glacial ice through the analysis of CRISPRs obtained from metagenomic data.

With the requirement that the DRs be found in fully sequenced bacterial genomes in the NCBI database and that our 16S rRNA gene pyrotags from the Arctic match the 16S rRNA sequences from those fully sequenced genomes, our CRISPR network (10 CRISPRs from the ice and 1 from the soil) was reduced to four CRISPRs in ice and none in soil (Fig. 3). The low number of CRISPRs found in soil might be due to a limited depth of coverage in the more diverse soil metagenome dataset. Therefore, ice appears to be a good model ecosystem for this initial CRISPR application study. The analysis of ice CRISPRs shed some light on the communities inhabiting this environment and their possible viral relationships. *Verminephrobacter* (possible host I and II in Fig. 3) has been described as a nephridia symbiont, suggesting the presence of annelids in the ice. No functional data

for annelids were present in our metagenome dataset based on an analysis with MG-RAST. However, given that samples were pre-filtered to remove debris, it is possible that we lost these in the process. The presence of ice worms has been confirmed in a range of glaciers (Hartzell et al. 2005), and representatives of *Verminephrobacter* were also found in Lake Vostok ice (Shtarkman et al. 2013). Recently, a study described the presence of bacterial microbiota in ice worms (Murakami et al. 2015). *Bordetella* (possible host II) was also found to be present in ice cores from Pony Lake, Antarctica (Foreman et al. 2011). Although *Staphylococcus* and *Mycobacterium* (possible host III and possible host III and IV, respectively) are genera often associated with humans, they are ubiquitous genera that have been found inhabiting a wide range of environments (Kloos 1980; Falkinham 2002), including glacial ice (Christner et al. 2000; Miteva, Sheridan and Brenchley 2004; Xiang et al. 2005). When looking at the taxonomic assignment of the 16S rRNA gene sequences, only *Mycobacterium* was present at the genus level (0.14% of assigned reads). No other bacteria identified through our network were represented in the taxonomic classification of the 16S rRNA gene sequences, which could point to a low abundance of these bacteria in their environment. It could also mean that these bacteria are the closest members present in the database, but that the bacteria actually inhabiting the ice do not have representatives in the 16S rRNA gene database, thus limiting the taxonomic assignment. With respect to the viruses identified in our network, all of them, with the exception of the *Cellulophaga* phage, were described by Cottrell and Kirchman (2012) in their analysis of viral genes present in Arctic bacteria. These data support the assumption that these viruses are infecting Arctic bacteria as indicated by our results. In addition, *Cellulophaga* phages seem to be ubiquitous in aquatic environments based on the study by Holmfeldt et al. (2013).

Our interest in linking phages and their host in the environment originated from the hypothesis that transduction has a significant role in microbial adaptation. We, therefore, tried to identify molecular tracers of transduction in our viral metagenomes. We found hybrid contigs in our assembled virome that had both viral and microbial sequences in the same contig. From this search for transducing agents, *Ralstonia* phages were frequent candidates in our Arctic viromes. No hybrid contig was found in the soil virome while six hybrid contigs were found in the ice virome. Given that the overall number of viral ice contigs was roughly a third of the contigs from the soil, transduction in ice might be more predominant than in the soil, where other mechanisms of gene transfer might compete with transduction. Based on this result, *Aeromonas* and *Ralstonia* phages in Arctic glacial ice appear to be carrying pieces of various microbial genes that could be (or have been) transferred to their host(s). 16S rRNA genes assigned to *Aeromonas* were found in viral particles issued from wastewater (Sander and Schmieger 2001). In addition, individual *Ralstonia* phages appear to be capable of infecting several different strains of *Ralstonia solanacearum* (Kawasaki et al. 2009) and possibly spreading important environmental genes throughout the population. Moreover, *Cupriavidus metallidurans*, formerly *Ralstonia metallidurans* (Vandamme and Coenye 2004), is a bacteria known to resist a range of different metals including mercury by carrying the genes responsible for this resistance in a *mer* operon (von Rozycki and Nies 2009). Given that we could find *Cupriavidus metallidurans* in our MG-RAST functional assignment data (Meyer et al. 2008), *Ralstonia* phages found in this study as transduction agents of mercury resistance genes might have *Cupriavidus metallidurans* as a host.

The stringent nature of the data interpretation here has eliminated possible bacterial/viral interactions when both the bac-

teria and the virus were unknown or the bacterial presence was not confirmed by 16S rRNA sequencing. Therefore, this reduced output results in part from the rather limited databases available. The increase in viruses included in viral databases as well as bacterial genomes in genomic databases will improve the effectiveness of CRISPRs as a tool to link known viruses to their known hosts. CRISPRs are also quite abundant in archaea, but the number of fully sequenced archaeal genomes is considerably smaller than that for bacteria. In a similar way, deeper microbial sequencing could also increase the number of CRISPR sequences found in metagenomes and make the identification of the microbial originator of the CRISPR a more likely occurrence in environments with high microbial diversity. On the other hand, our CRISPR-based network discovered viral/microbial interactions among unknown viruses and bacteria/archaea. The further identification of these microorganisms would help clarify the extent of these interactions.

The results described here are consistent with Arctic ice being an ecosystem with virus-host interactions and that these interactions seem to be relatively (on a per mass of DNA basis) more productive than the ones taking place in nearby Arctic soil. In addition, the transduction of the genes studied here might be more prevalent in ice than in soil as a consequence of more dynamic interactions. This would be consistent with previous work that suggested virus-host interactions in cold environments could be highly dynamic (Wells and Deming 2006; Anesio and Bellas 2011). Our study has developed a CRISPR-based tool to link viruses with their host(s) in the environment through metagenomic data. In doing so we were able to provide a new perspective and gather information on virus-host relationships. This new perspective could help us further understand and compare viral ecology in different ecosystems, giving us a more complete view of microbial communities, their interactions, and consequently, their influence on global processes.

ACKNOWLEDGEMENTS

This research is part of the European project Trainbiodiverse and was funded by Marie Curie Actions. We would also like to acknowledge the support of IPEV for field work (Chimerpol III project).

FUNDING

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. [289949].

Conflict of interest. None declared.

REFERENCES

- Anderson RE, Brazelton WJ, Baross JA. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* 2011;77:120–33.
- Anesio AM, Bellas CM. Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends Microbiol* 2011;19:52–7.
- Anesio AM, Mindl B, Laybourn-Parry J, et al. Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J Geophys Res* 2007;112:G04S31.

- Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007;**315**:1709–12.
- Bellas CM, Anesio AM. High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles* 2013;**17**:861–70.
- Bellas CM, Anesio AM, Telling J, et al. Viral impacts on bacterial communities in Arctic cryoconite. *Environ Res Lett* 2013;**8**:045021.
- Canchaya C, Fournous G, Chibani-Chennoufi S, et al. Phage as agents of lateral gene transfer. *Curr Opin Microbiol* 2003;**6**: 417–24.
- Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2011;**7**:335–6.
- Christner BC, Mosley-thompson E, Thompson LG, et al. Recovery and identification of viable bacteria immured in glacial ice. *Icarus* 2000;**144**:479–85.
- Colangelo-Lillis JR, Deming JW. Genomic analysis of cold-active Colwelliophage 9A and psychrophilic phage-host interactions. *Extremophiles* 2013;**17**:99–114.
- Cottrell M, Kirchman D. Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquat Microb Ecol* 2012;**66**:107–16.
- Falkinham JOI. Nontuberculous mycobacteria in the environment. *Clin Chest Med* 2002;**23**:529–51.
- Fineran PC, Charpentier E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 2012;**434**:202–9.
- Foreman CM, Diesler M, Greenwood M, et al. When a habitat freezes solid: microorganisms over-winter within the ice column of a coastal Antarctic lake. *FEMS Microbiol Ecol* 2011;**76**:401–12.
- Frost LS, Lepplae R, Summers AO, et al. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;**3**:722–32.
- Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;**28**:3150–2.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature* 1999;**399**:541–8.
- Hartzell PL, Nghiem JV, Richio KJ, et al. Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus rainierensis*). *Can J Zool* 2005;**83**:1206–13.
- Holmfeldt K, Solonenko N, Shah M, et al. Twelve previously unknown phage genera are ubiquitous in global oceans. *P Natl Acad Sci USA* 2013;**110**:12798–803.
- Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* 2013;**14**:R123.
- Jansen R, Embden JD, Gaastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002;**43**:1565–75.
- Jiang SC, Paul JH. Gene transfer by transduction in the marine environment. *Appl Environ Microb* 1998;**64**:2780–87.
- Kawasaki T, Shimizu M, Satsuma H, et al. Genomic characterization of *Ralstonia solanacearum* phage phiRSB1, a T7-like wide-host-range phage. *J Bacteriol* 2009;**191**:422–7.
- Kloos WE. Natural populations of the genus *Staphylococcus*. *Annu Rev Microbiol* 1980;**34**:559–92.
- Koskella B, Meaden S. Understanding bacteriophage specificity in natural microbial communities. *Viruses* 2013;**5**: 806–23.
- Kunin V, Sorek R, Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 2007;**8**:R61.
- Larose C, Berger S, Ferrari C, et al. Microbial sequences retrieved from environmental samples from seasonal arctic snow and meltwater from Svalbard, Norway. *Extremophiles* 2010;**14**: 205–12.
- Larose C, Prestat E, Cecillon S, et al. Interactions between snow chemistry, mercury inputs and microbial population dynamics in an Arctic snowpack. *PLoS One* 2013;**8**:e79972.
- Lindell D, Jaffe JD, Johnson ZI, et al. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 2005;**438**:86–9.
- López-Bueno A, Tamames J, Velázquez D, et al. High diversity of the viral community from an Antarctic lake. *Science* 2009;**326**:858–61.
- Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008;**9**:386.
- Miteva VI, Sheridan PP, Brenchley JE. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl Environ Microb* 2004;**70**: 202–13.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005;**60**: 174–82.
- Mojica FJM, Ferrer C, Juez G. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. 1995;**17**:85–93.
- Morse ML, Lederberg EM, Lederberg J. Transduction in *Escherichia coli* k-12. *Genetics* 1956;**41**:142–56.
- Murakami T, Segawa T, Bodington D, et al. Census of bacterial microbiota associated with the glacier ice worm *Mesenchytraeus solifugus*. *FEMS Microbiol Ecol* 2015;**91**:1–10.
- Ray J, Dondrup M, Modha S, et al. Finding a needle in the virus metagenome haystack—micro-metagenome analysis captures a snapshot of the diversity of a bacteriophage armoire. *PLoS One* 2012;**7**:e34238.
- Rho M, Wu Y-W, Tang H, et al. Diverse CRISPRs evolving in human microbiomes. *PLoS Genet* 2012;**8**:e1002441.
- Rodríguez-Valera F, Martín-Cuadrado A-B, Rodríguez-Brito B, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* 2009;**7**:828–36.
- Rohwer F, Thurber RV. Viruses manipulate the marine environment. *Nature* 2009;**459**:207–12.
- Rosario K, Breitbart M. Exploring the viral world through metagenomics. *Curr Opin Virol* 2011;**1**:289–97.
- Roux S, Faubladier M, Mahul A, et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics* 2011;**27**:3074–5.
- Saito R, Smoot ME, Ono K, et al. A travel guide to Cytoscape plugins. *Nat Methods* 2012;**9**:1069–76.
- Sander M, Schmiegner H. Method for host-independent detection of generalized transducing bacteriophages in natural habitats. *Appl Environ Microb* 2001;**67**:1490–3.
- Sano E, Carlson S, Wegley L, et al. Movement of Viruses between Biomes. *Appl Environ Microb* 2004;**70**:5842–6.
- Sävström C, Granéli W, Laybourn-Parry J, et al. High viral infection rates in Antarctic and Arctic bacterioplankton. *Environ Microbiol* 2007;**9**:250–5.
- Shtarkman YM, Koçer ZA, Edgar R, et al. Subglacial Lake Vostok (Antarctica) accretion ice contains a diverse set of sequences

- from aquatic, marine and sediment-inhabiting bacteria and eukarya. *PLoS One* 2013;**8**:e67221.
- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 2007;**5**:801–12.
- Terns MP, Terns RM. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* 2011;**14**:321–7.
- Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* 2000;**45**:1320–8.
- Thurber RV, Haynes M, Breitbart M, et al. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 2009;**4**:470–83.
- Vandamme P, Coenye T. Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Micr* 2004;**54**:2285–9.
- von Rozycki T, Nies DH. *Cupriavidus metallidurans*: evolution of a metal-resistant bacterium. *Antonie Van Leeuwenhoek* 2009;**96**:115–39.
- Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 1996;**272**:1910–4.
- Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 2004;**28**:127–81.
- Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* 2003;**6**:1–11.
- Wells LE, Deming JW. Brief report Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ Microbiol* 2006;**8**:1115–21.
- Williamson KE, Corzo KA, Drissi CL, et al. Estimates of viral abundance in soils are strongly influenced by extraction and enumeration methods. *Biol Fert Soils* 2013;**49**:857–69.
- Xiang S, Yao T, An L, et al. 16S rRNA sequences and differences in bacteria isolated from the Muztag Ata glacier at increasing depths. *Appl Environ Microb* 2005;**71**:4619–27.



Institut Pasteur

Research in Microbiology 166 (2015) 782–795



www.elsevier.com/locate/resmic

Snow and ice ecosystems: not so extreme

Lorrie Maccario¹, Laura Sanguino¹, Timothy M. Vogel, Catherine Larose*

Environmental Microbial Genomics, Laboratoire Ampère, CNRS UMR 5005, Université de Lyon, Ecole Centrale de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully, France

Received 27 April 2015; accepted 11 September 2015

Available online 25 September 2015

Abstract

Snow and ice environments cover up to 21% of the Earth's surface. They have been regarded as extreme environments because of their low temperatures, high UV irradiation, low nutrients and low water availability, and thus, their microbial activity has not been considered relevant from a global microbial ecology viewpoint. In this review, we focus on why snow and ice habitats might not be extreme from a microbiological perspective. Microorganisms interact closely with the abiotic conditions imposed by snow and ice habitats by having diverse adaptations, that include genetic resistance mechanisms, to different types of stresses in addition to inhabiting various niches where these potential stresses might be reduced. The microbial communities inhabiting snow and ice are not only abundant and taxonomically diverse, but complex in terms of their interactions. Altogether, snow and ice seem to be true ecosystems with a role in global biogeochemical cycles that has likely been underestimated. Future work should expand past resistance studies to understanding the function of these ecosystems.

© 2015 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Snow; Ice; Extreme; Microbial ecology

1. Introduction

Extreme environments are defined as areas where one or more of the abiotic parameters, i.e. pH, temperature, water content, nutrient levels, solar irradiation, pressure, etc., reach extreme values [1]. These harsh conditions are considered to be limiting for the development of life. Nevertheless, the range of values that are considered extreme are often regarded from a macro-organism point of view, whereas the limit for the development of microbial life on Earth has not yet been fully established. The parameters cited above influence all organisms at the individual scale and could, to different extents, cause cell physiological stress. However, microorganisms harbor a remarkable capacity to adapt to a wide range of habitats. To date, no place on Earth has been shown to be sterile, and even the most “inhospitable” environments, such

as deep sea hydrothermal vents, acid mine drainage and hot springs, contain many viable microorganisms. In the case of hot springs, the occurrence of a viable and active microbial community was recognized early [2]. Many studies have focused on its ecology both for fundamental and application purposes: as a model for the development of life under putative primitive Earth conditions and for enzymes with industrial interest due to their activity at high temperatures.

At the opposite end of the temperature spectrum, snow and ice have received much less attention. Over the past twenty years, microorganisms inhabiting the cryosphere have been increasingly studied for the potential discovery of enzymes with biotechnological interest and for fundamental research on the ecology of “extreme” environments [3–7]. Snow and ice have unexpectedly high microbial abundance and diversity. As main components of the cryosphere, they cover over 10^8 km² of the Earth's surface. Snow in the winter can cover up to 12% of the Earth's surface [8], and approximately 10% of the planet's land surface is covered by glacial ice in the form of ice caps, ice sheets or glaciers, storing 75% of the world's fresh

* Corresponding author.

E-mail address: catherine.larose@ec-lyon.fr (C. Larose).

¹ Both authors contributed equally to this work.

water [9]. Given their coverage at a global scale, snow and ice could have a major and underestimated role in global biogeochemical cycling (Fig. 1). Therefore, the relevant knowledge and knowledge gaps concerning the microbial ecology of snow and ice ecosystems need to be evaluated for future research.

In this article, we will identify and evaluate the physical and chemical conditions in snow and ice environments and review the mechanisms used by microorganisms to survive under these conditions. We will also address the abundance and diversity of the snow and ice microbial communities, as well as the interactions that take place within the community that help them to successfully inhabit snow and ice. Given that an ecosystem is defined as the compendium of the organisms present, the physical system that hosts them and the interactions of the organisms with abiotic factors as well as the interactions between organisms [10], the available data support the idea that snow and ice are true ecosystems and actually not so extreme for these microorganisms.

2. Snow and ice environmental constraints: definition, variability and individual cell responses

Snow and ice habitats are the only ones on Earth whose matrices are partly composed of frozen water, and thus impose specific conditions on microbial life. The driving forces for microbial community structure and function might be the result of these physical–chemical conditions: low temperatures, scarcity of nutrients and water as well as high UV

radiation. In this section, we compare the different abiotic attributes of snow and ice ecotypes and the related microbial adaptations (Fig. 2) in order to understand how the microbial community actively interacts with its physical–chemical environment. The objective is then to investigate whether these environmental constraints define snow and ice as extreme habitats for microorganisms.

2.1. Cold temperature and low water availability

Snow and ice are found in the coldest places on the planet and fall within group E of the Köppen climate classification or in polar and alpine climates, which are characterized by monthly mean temperatures below 10 °C [11]. The term “cryosphere”, first introduced by Dobrowolski in 1923 [12], derives from the Greek *cryos*, which means “cold” or “ice” and refers to areas where water is in a frozen state [5]. Mean temperatures observed in snow and ice environments can be highly variable at different depths, sites or seasons. For example, surfaces directly exposed to wind are influenced by air temperature (as low as –50 °C and –70 °C during the winter in the Arctic and Antarctic, respectively, and can be as warm as 0 °C during the melt period in summer) and therefore exhibit the coldest and most variable temperature. Top snow layers work as insulators leading to an increase in temperatures in the layers below. The low temperatures can be deleterious for cells due to increases in membrane rigidity that will limit substrate exchange, reduce enzyme activity and damage cells with cytoplasmic water crystal formation. Many studies on microbial life in the cryosphere have focused on understanding how microorganisms can adapt to the cold. The term “psychrophiles” refers to cold-adapted microorganisms with optimal growth temperatures below 15 °C, and “psychrotolerants” are organisms able to survive below 0 °C, but with optimal growth temperatures between 20 and 25 °C [13]. They use strategies at different levels of cell physiology, including increased membrane fluidity (unsaturated fatty acid synthesis, desaturases), freeze protection (compatible solutes, anti-freeze and ice-binding proteins synthesis), and maintenance of protein catalytic efficiency (reduction of internal hydrophobic interaction, proline and arginine content, increased length of external loops and accessibility to active sites) (see detailed reviews on psychrophilic adaptations [14,15]), and recently highlighted by sequencing technologies [16]. Sequences associated with cold-adapted organisms have been systematically detected in snow and ice sequence datasets [17–20]. The psychrophilic lifestyle of some snow and ice microorganisms has been confirmed via isolates cultivated at low temperature [21,22]. In addition, the possibility that this icy lifestyle could have a remote or ancient origin cannot be discarded. Although the origin of life and the emergence of microorganisms on Earth remain unknown, some proposals for a cold origin of life have been made [23,24]. This hypothesis is supported by various studies including assays on prebiotic polymerization in an icy matrix [25–28] and bacterial phylogeny [29]. If this were to be the case, then native conditions at the origin of microbial development should not be considered as extreme.

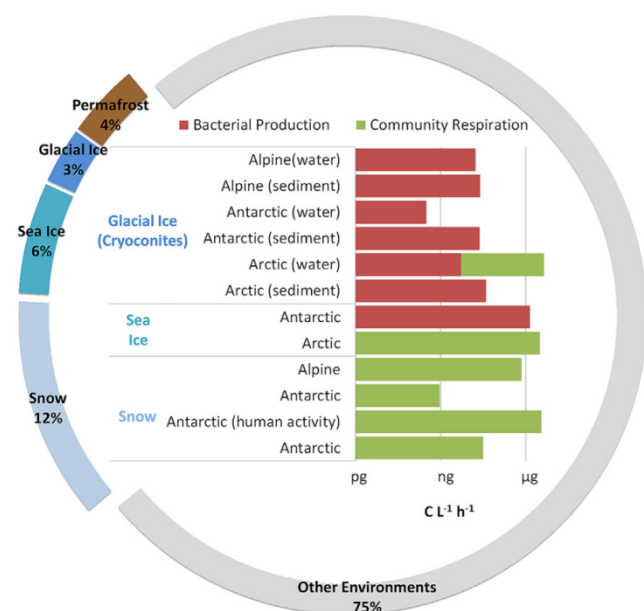


Fig. 1. Coverage and microbial productivity of snow and ice. The graph on the outer ring depicts the percentage of the surface of the Earth covered by the different environments of the cryosphere. The bar chart shows bacterial production and community respiration rates (in carbon per liter per h) found in the literature for these environments. Sediment or water represent the different parts of cryoconite. Human activity refers to snow samples taken near a scientific station and potentially impacted by human activity.

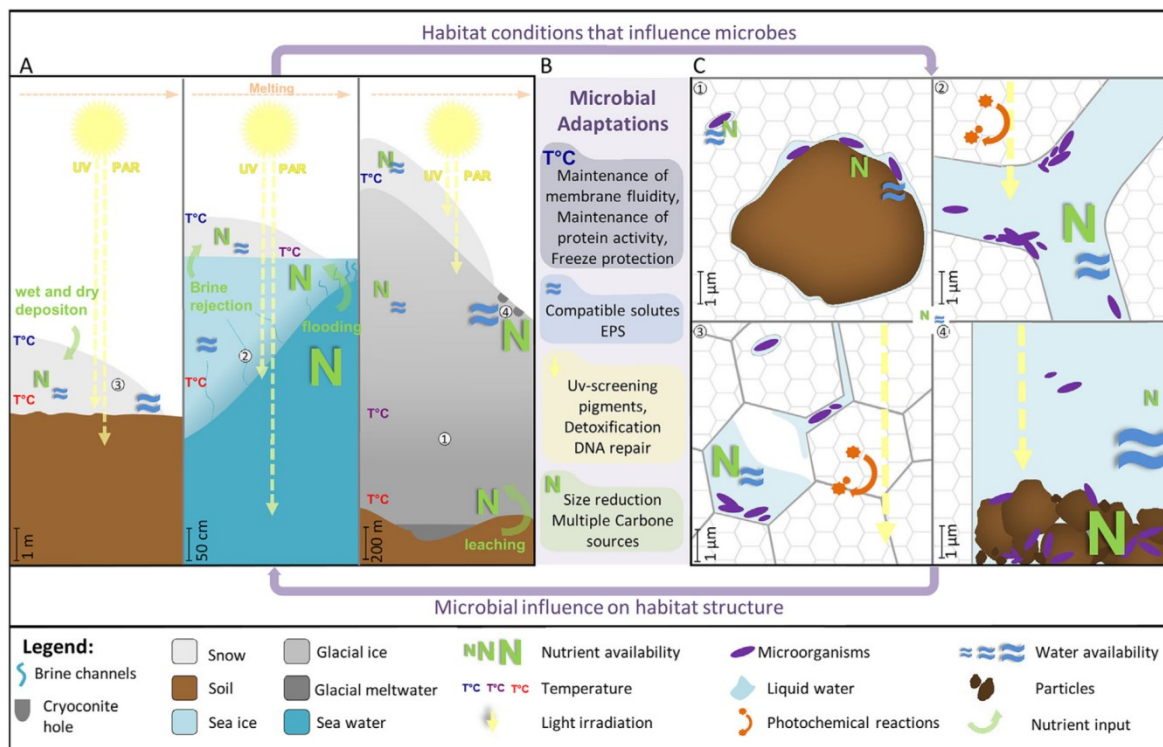


Fig. 2. Microbial interactions with abiotic conditions in the different habitats within snow and ice environments. A: Habitat structure and abiotic conditions at macroscopic level. B: Microbial adaptation in response to potential stress caused by low temperature, water and nutrient availability and high UV-radiation. C: Examples of niches at a micro-scale (1: particle or microbial cell trapped within glacial ice, 2: sea-ice brine channel, 3: snow, 4: cryoconite hole).

One physical consequence of cold temperatures is the low liquid water activity, since frozen water is not available for most chemical reactions. As the seawater freezes, liquid water containing all the elements excluded from ice crystals is concentrated in brine channels and pockets. Channel sizes can vary from several to hundreds of micrometers in diameter [30]. Temperatures determine solute concentrations, and salinity varies from 30 ppt at freezing point temperature of $-2\text{ }^{\circ}\text{C}$ (equivalent to seawater salinity) and up to 230 ppt at $-30\text{ }^{\circ}\text{C}$ [31]. As glacial ice is formed by the accumulation of snow, impurities are excluded from the ice, thus creating hypersaline and hyper-acidic vein networks [32,33]. As the pressure exerted by the upper layers increases, ice crystals grow, while air spaces and vein size decrease. Several detailed reviews of the different types of glacial ice and their hydrology have been published recently [34,35]. The thermal structure of a glacier regulates the distribution of liquid water and thus might define the different microbial habitats that can be found in glaciers. The snowpack is a porous medium formed by the accumulation of deposited snowflakes. Its water content can vary from less than 0.5% in dry snow to above 15% in soaked snow [36] depending in part on temperature. This variation appears to affect water distribution at the microscale. At low water content, water forms isolated patches in concave areas such as at grain boundaries. When water content reaches above 10%, the water film is continuous, with no interface between the ice crystal and interstitial air [37]. The exact location of microorganisms within the snowpack and the size

and configuration of their microscale habitat within the water film are largely unknown. Microorganisms in snow and ice might undergo desiccation and high osmotic pressure that would result in cytoplasmic water loss and intracellular ionic imbalance. Mechanisms of microbial response to low water activity in snow matrices have not been demonstrated. Given the large variety of compatible solutes (free amino acids, sugars and polyol) used by a wide range of microorganisms (Archaea, Bacteria and Eukarya) [38–40], further research could help determine how important osmoprotection is in snow ecosystems. Potential microbial adaptations to low water availability have been detected in sea-ice and glacial ice, such as the synthesis and uptake of compatible solutes acting as osmolytes. Genes encoding the synthesis of some osmoprotectants, such as glycine, betaine, choline, sarcosine and glutamate, have been detected in alpine glacier ice core metagenomes [17]. Glycine-betaine (and proline), were shown to offer protection against fluctuating temperatures and salinity in two sea-ice model organisms, *Colwellia psychrerythraea* strain 34H and *Psychrobacter* P7E [41]. Moreover, proteomic comparisons from three sea-ice halophilic *Gamma-Proteobacteria* *Alteromonadales* against non-halophilic and extremely halophilic organisms showed a higher ratio of acidic membrane proteins, which may prevent protein precipitation under high salt conditions [42]. Extracellular polymeric substances (EPS) can also help with tolerating osmotic stress as sodium chelator or diffusion barriers between the cell wall and the environment. EPS production during salt stress has been

observed for the sea ice diatom *Fragilariopsis cylindrus* [43] and for a sea-ice isolate from the *Pseudoalteromonas* genera [44]. *Chlamydomonas* sp. ARC, isolated from Arctic sea-ice, tolerates a broad range of salinities from 2.5 to 100 ppt [45]. Transforming its vacuoles into tubular complexes called spongiomes, the algae expel excess water when exposed to decreasing salinity correlated with sea-ice melt [45]. Using selective hyperosmotic (sugar or salt) media, up to 9.10^3 CFU of xerophilic fungi identified as black yeast, yeast or a large variety of *Penicillium* were isolated from an Arctic glacial ice core [46]. However, some snow and ice niches benefit from increased water availability such as in water ponds and streams and cryoconite holes. As the spring season progresses, the increase in temperature and light radiation leads to the growth of wet areas and the dilution of solutes on snow and ice surfaces. Metagenomic datasets of samples from microbial mats that formed on Arctic and Antarctic ice-sheet ponds appear to have numerous genes involved in compatible solute synthesis and uptake. Antarctic mats metagenome datasets had a higher representation of genes encoding a general stressor regulon Sigma B, which has been suggested to reflect greater osmotic pressure at freeze-up in the more saline waters from Antarctic ponds compared to Arctic ponds [18]. In addition, the combination of friction created by the movement of the glaciers and the geothermal heat from the ground beneath them can melt the ice at the base of the glaciers, creating a layer of liquid water between the ice and the bedrock [47–49].

2.2. High light radiation and low nutrient concentrations

Snow and ice located in the polar zones are subjected to continuous light during the summer, while alpine ecosystems are exposed to high light levels due to altitude. The optical properties of snow and ice are intimately related to their physical structure (air, water and dust content), but generally snow and ice are highly scattering media with a highly reflective surface (high albedo) [50]. Due to recent ozone depletion, polar regions have been subjected to increased UV exposure. Snowpacks are especially reactive to UV light and have been described as natural photochemical bioreactors due to a multitude of ice grains resulting in an ice–air interface increasing the scattering effect [51,52]. The sequestered chemical compounds are photolyzed and reactive trace gases are released in the snow boundary layer. The processes are increased by higher gas diffusion due to low temperatures. These photochemically-induced reactions may result in the accumulation of reactive species within the snowpack, and thus a hyper-oxidative stressed habitat. However, UV-B transmittance is efficiently attenuated by the snow (e.g., measured levels are one order of magnitude lower at a depth of 8 cm [50], thus protecting underlying communities [53]). Uncovered sea-ice has been shown to be an efficient medium for photochemical reactions, with 85% occurring in the first meter of the sea-ice column [54]. UV radiation is known to be deleterious for cell physiology [55] and the subsequent damage and microbial resistance are wavelength-dependent and

variable at the species level. Generally, UV-B is involved in direct damage to absorbing biomolecules, such as DNA, by forming pyrimidine dimers that impede replication and transcription, whereas UV-A has indirect effects by producing reactive oxidative species (ROS) that cause oxidative damage to lipids, DNA and proteins [56]. The production of UV-screening compounds, mycosporine-like amino acids (MAA), was observed in several ice samples exposed to high levels of UV radiation in the Arctic first year sea-ice [57], the surface layers of Baltic sea-ice [58] and in Antarctic sea-ice [59]. Astaxanthin, a secondary carotenoid with multifunctional stress response characteristics, also acts as a detoxifier of ROS produced by photochemical reactions and UV-screen in various algae [60–63]. Other algae, such as members of the *Zygnematophyceae* class that inhabit glacial surfaces, do not harbor any mycosporine-like compounds or secondary carotenoids, but accumulate a brownish UV-absorbing pigment identified as derived from purpurogallin [64]. Some algae, such as *Mesotaenium berggrenii*, reduce or even lose the pigment-containing vacuoles when exposed to lower irradiance levels [65]. These examples illustrate that specific and unique strategies to adapt to UV radiation might exist in snow and ice algal communities, but the total community response to UV is probably not restricted to pigment in photosynthetic algae. Metabolic stress, with cell activity reduction, has been observed in Antarctic sea ice bacteria after initial UV-B exposure, but the possible resistance mechanism was not elucidated [66]. Although metagenomes from arctic snowpack [20], arctic glacier ice core [17] and microbial mats from arctic and Antarctic ice-sheets ponds [18] highlighted the occurrence of some genetic traits for their response to photo-oxidative stress, UV-radiation and photo-oxidative stress response mechanisms specific to snow and ice communities as a whole remain largely undescribed. A variety of defense mechanisms against oxidative and photo-oxidative stress are known for microorganisms, including heterotrophic representatives from various habitats. For example, melanin and mannitol are involved as an antioxidant and in UV protection in fungi [67,68]. Detoxification systems (OxyR and SoxRS regulons) involving enzymatic oxidant elimination and DNA repair were described in a wide range of bacteria [69]. DNA repair mechanisms are also widespread in all microbial kingdoms based on nucleotide excision [70]. Searching for these mechanisms in snow and ice should be addressed in the future, as microbial communities could have developed a wide diversity of adaptation mechanisms given the levels of UV-light in these environments.

On the other side of the light spectra, photosynthetically active radiation (PAR) might be an important factor in snow and ice microbial community establishment. Indeed, snow and ice environments are generally oligotrophic with low bulk dissolved organic carbon (DOC), so inorganic carbon fixation and autotrophic energy production by primary producers might be important processes [32,71,72]. Photosynthetic activity might be important in several cryosphere ecosystems such as colored snow colonized by snow algae [63,73], ice sheet surfaces [74], ice lake covers [75], sea-ice [76] and

cryoconite holes [77]. Non-photochemical autotrophic processes could also be involved in nutrient dynamics within these habitats. The importance of this lifestyle was supported by the presence of a significant number of genes involved in carbon fixation in glacial ice metagenomes [17]. In subglacial areas, water, organic carbon and minerals sustain the development of microbial communities where chemolithotrophic primary production could be an important process [78–80]. Another source of organic compounds for microorganisms in snow and ice might be carbon released from dead cells and allochthonous carbon from local or remote origins [81], and thus, certain niches in snow and ice ecosystems could actually be nutrient-rich environments. Since nutrients and particles are excluded from the ice crystal structure and are concentrated in pore spaces, concentrations of nitrate and sulfate in ice veins from Greenland and Antarctica were 10^3 – 10^5 times more concentrated than in bulk ice [32]. These levels of nutrients are similar to what can be found in rich growth media [82] and are likely sufficient to support microbial activity in deep glacial ice [83,84]. Cryoconite holes might also be nutrient-rich areas as these semi-stagnant aquatic habitats have been considered to be eutrophic [85,86]. In sea-ice, DOC concentration is often correlated to algal biomass and can be higher than the underlying sea-water [211,213]. In addition, organisms could be benefiting from tidal or wave induced transport of DOC within the ice-brine channel network. As snowflakes fall, they scavenge atmospheric particles including microorganisms, nutrients, and anthropogenic pollutants [87] and thus, snow has been defined as a nutrient and microbial reservoir [88]. For example, due to wet and dry atmospheric deposition events, snowpack inhabitants are exposed to occasional inputs of nitrogen compounds that microorganisms might be able to metabolize. Nitrogen cycling pathway genes were apparently increased in microbial communities during spring and were correlated with different nitrogen compounds [89]. Fluxes of nitrogen species in snow during polar night (i.e. in the absence of photochemical reactions) were observed and attributed to microbial activity [90]. Numerous genes detected in glacier ice core metagenomes related to xenobiotics, biopolymers and other carbon sources suggest that glacial ice microorganisms have the potential to degrade a wide range of substrates [17]. Microbial degradation of xenobiotic compounds from distant allochthonous sources has been observed on the Greenland ice sheet [91]. Microbial preferences for different carbon classes were also studied in Antarctic snow and results showed a higher rate of carbon uptake when snow microcosms were amended with a combination of simple and complex carbon sources [92]. In the same study, snow isolates were capable of oxidizing a broad spectrum of low and high molecular weight carbon sources including amino acids, amines, amides, carboxylic acids, carbohydrates, and complex polymers. Altogether, these results highlight the potential for high metabolic versatility of microorganisms in snow and ice habitats with variable concentrations of many different carbon sources.

Although snow and ice share multiple characteristics due to their ice crystal backbone structure and their homogenous appearance at a macroscopic level, they are composed of a

multitude of niches under different conditions. At a microscopic level for each micro-habitat, stresses, such as low nutrient or water availability might not be as extreme as expected. Moreover, individual cells could use a wide range of strategies to resist the potential stresses in snow and ice. Snow and ice microorganisms are not only able to face the physico-chemical conditions of their habitats, but they are also shaping their environment. Pigment producers develop at surfaces and form watermelon snow or brown, red and gray ice, which drastically decreases surface albedo and increases local temperature and water content [35,93]. In glaciers and lake ice, the production of ice binding proteins by bacteria can modify ice vein structure by inhibiting recrystallization and potentially reducing diffusion in veins [94,95]. Likewise, EPS production by sea-ice algae and bacteria modifies the structure of the ice by reducing permeability and is thus considered essential for the creation of connected and complex brine channels within sea-ice [31,96]. In both cases, microbial activity leads to retention of molecules and microorganisms, and thus actively promotes the formation of a stable microbial community.

3. Snow and ice inhabitants in complex communities

In the previous section, we discussed individual cell mechanisms involved in resistance to the potentially harsh conditions of snow and ice habitats. Here we would like to examine the complex communities of snow and ice inhabitants in order to investigate how their interactions can impact snow and habitability, and whether they constitute extreme communities.

3.1. Abundance, richness and variability

The various niches in snow and ice are composed of microbial communities that strongly vary in size, composition and diversity (Fig. 3). The estimated bacterial diversity varied among the different ecosystems studied, but can occasionally be similar to that of temperate soils ($2.4 < H' < 3.6$) [97]. In general, microbial cell counts registered in the snowpack range from 10^3 to 10^4 cells per ml of melted snow collected over soil and ice in Arctic, Antarctic and Alpine environments. A few reports described counts up to 10^5 cells per ml in some Alpine sites [98,99]. In debris-rich sea and glacier ice, the microbial abundance varies between 10^4 and 10^7 cells per ml of melted bulk ice. The lowest microbial abundance was observed in Antarctic snowpack samples from Concordia, which had only hundreds of cells per ml of melted snow. At the bottom of the sea-ice, the concentration of microorganisms, especially algae, appears to increase. Also, while microbial abundance is expressed by total volume of melted samples and therefore averaged over the sample size, cells are concentrated into the liquid phase of snow and ice. Models of particle exclusion in glacial ice veins predict that the concentration of cells could be a thousand times higher in the veins than the bulk melt water [82].

Bacteria seem to be ubiquitous in snow and ice, as they have been observed systematically in all snow and ice

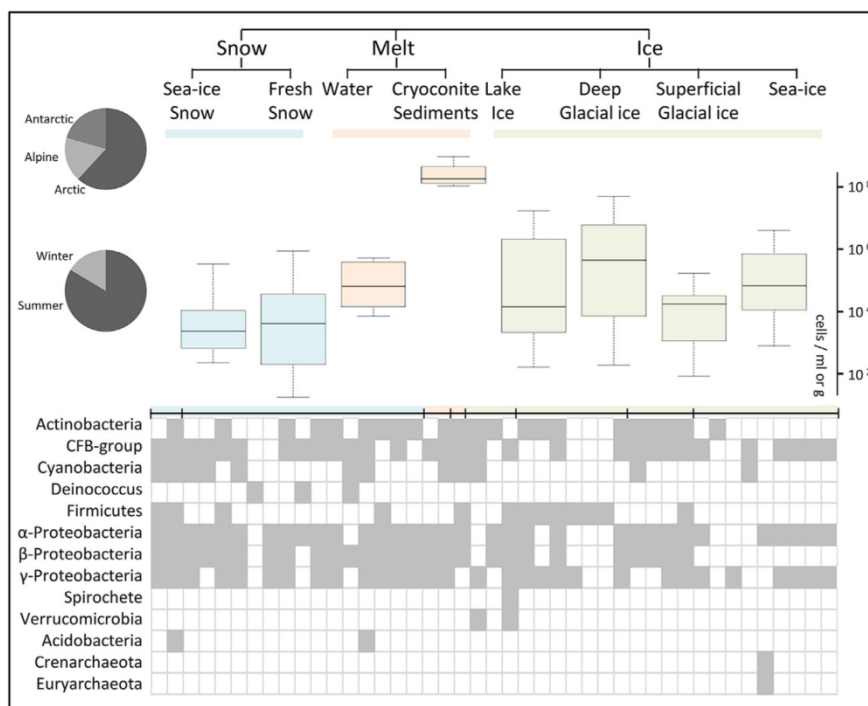


Fig. 3. Snow and ice prokaryotic inhabitants. Abundance and community composition data were extracted from available literature. Pie charts depict geographical distribution (Arctic, Antarctic or Alpine) and sampling season (summer/spring or winter/fall) of the data used. Data were grouped into 8 snow and ice ecotypes: snow over sea-ice, fresh snow over soil or glacier, melt water (ponds, streams, cryoconite hole water), lake ice, deep glacial ice, superficial glacial ice and sea-ice. Prokaryotic abundance is illustrated by box plots of cell counts per ml of melted snow or ice (or g for cryoconite sediments) for each ecotype. Community composition is assessed by a heatmap of presence (in gray) of most detected taxa at phylum level in the following studies. Data from Refs. [17–22,82,98–101,105,106,108,110,111,130,136,156,158,165,171,174–206].

environments studied, regardless of the methodological approaches used (culture dependent or not) (Fig. 3). The bacteria identified from these studies belong to numerous taxa, although mostly from Proteobacteria (Alpha-, Beta- and Gamma-), the Cytophaga–Flexibacter–Bacteroides group, Actinobacteria and Cyanobacteria. Unlike Bacteria, Archaea are only rarely observed. Sequences associated with the Archaeal domain (*Thaumarchaeota* and *Euryarchaeota*) were detected with a relative abundance below 1% in Arctic spring snow samples over soil, sea-ice and ice-sheets [20,100,101]. Archaea were estimated at up to 6.6% of the microbial community in Antarctic sea ice [102]. *Thaumarchaeota* were also detected at low levels in cryoconite holes from Arctic and Antarctic glaciers [103,104]. Snow and ice also provide habitats for a wide range of eukaryotic microorganisms such as fungi, microalgae or heterotrophic protists (see review by Arrigo, 2014 [7]). In sea ice/ice sheet snow cover, the eukaryotic community is largely dominated by fungi (*Ascomycota* and *Basidiomycota*) followed by *Alveolata* (*Dinoflagellata*) [101,105,106] and in terrestrial snow, fungi sequences can represent, at times, up to 70% of the total metagenomic sequences [20]. However, fungi are rarely detected in sea ice, which is dominated by microalgae, the most abundant taxa being diatoms (*Bacillariophyceae*) in both in Arctic and Antarctic samples. Algae can be found within cryoconite holes or on the surface of glaciers [107]. Little is known about the biodiversity in supraglacial lakes and streams

and how their ephemeral nature might affect their habitability. Nevertheless, the presence of *Proteobacteria*, *Bacteroidetes* and algal phototrophs was described in a stream system of Antarctica with low dissolved organic matter content [108]. Another study reported the presence of green algae in two supraglacial lakes of Antarctica, although biomass was considered to be low [109].

This high variability in snow and ice communities is also correlated with dynamic changes in physical and chemical conditions of these systems. For example, the light regime changes that occur during seasonal transitions might have a major impact on polar snow and ice microbial communities. In one high Canadian Arctic winter study on sea-ice, most of the 16S rRNA gene clone library sequences were identified as SAR11 clade *Alphaproteobacteria* and marine group I.1a *Crenarchaeota* [110] that are rarely found during other seasons. Snow community composition has been shown to vary with depth [100] and time during snow warming and melting in spring [20,111]. Snow chemistry, such as pH, organic acids and mercury, has been correlated with bacterial community dynamics and higher relative abundance of associated genes [20,112].

These examples illustrate how snow and ice communities are not only able to resist harsh conditions, but might also respond dynamically to modifications in their surrounding physical and chemical environment. While extreme organisms would be restricted to a stringent range of conditions and be

highly stable, snow and ice communities seem to be diverse and dynamic. A truly extreme environment might also sustain only highly specialized microorganisms, perfectly adapted to their specific environment and unable to exist elsewhere. This does not seem to be the case for all microorganisms in snow and ice ecosystems. Cold adaptation, the most frequently described adaptation in icy environments, is a good example of lack of knowledge about extreme ecology. While cold-adapted microorganisms or cold adaptations seem to be ubiquitous in snow and ice habitats, close relatives to psychrophilic or psychrotolerant microorganisms have also been detected in non-cold environments. For example, strains isolated from a shallow estuary in Portugal with mean temperatures around 20 °C formed an operational taxon unit (OTU) whose cluster consensus sequence showed 100% similarity with the partial 16S rRNA gene sequence of *Psychrobacter glacincola*, which was originally isolated from an ice core from Antarctic sea ice and ice shelf [113]. Although all strains of *Psychrobacter* were able to grow at 4 and 28 °C, most of them also grew at 37 °C, suggesting a higher optimal growth temperature than for the genera described in colder environments [114]. Metagenomic comparisons also indicated that cold adaptation mechanisms were detected in similar relative abundance in various environments, including cold and temperate ones [20]. These findings suggest that cold adaptation or adapted microorganisms have not been restricted to cold environments. In addition, based on 16S rRNA gene annotations, many of the microorganisms identified in snow and ice have not been described as psychrophiles. This might suggest that these microorganisms, including mesophiles, are trapped in ice and snow in a vegetative non-adapted state until release upon snow and ice melting [115]. Secondly, this could also be explained by a bias towards mesophilic representatives in genome databases. At the time of writing, 194 genomes, i.e. less than 2% of the gold genome database (updated in real time), corresponded to cold-adapted microorganisms [116]. Therefore, sequences obtained from an unknown psychrophile or psychrotolerant might be matched against genomes of a closely related mesophile, leading to an erroneous physiological classification. Microbial genome plasticity that enables temperature adaptation could be another hypothesis. Experiments have shown that after exposing mesophiles, including *Escherichia coli*, to temperatures of –1.8 °C for forty-eight-days, their optimal growth temperature decreased and they even lost their ability to form colonies at 37 °C [117]. On the other hand, the application of supra-optimal temperatures induced differentiation to heat-tolerant phenotypes in psychrophilic prokaryotes and eukaryotes [118,119]. Although highly specialized microorganisms exist, their actual proportion in active snow and ice communities, their roles and whether they are essential to the entire community in snow and ice habitats are unknown.

3.2. Complex interactions

The majority of microorganisms that inhabit any given ecosystem are rarely alone, as they often live in communities

and share their niche with other organisms. Biological interactions between the members of a community can shape and partly define an ecosystem [120]. In the cryosphere, these relationships could be an important asset in the colonization of the different habitats. Horizontal gene transfer (HGT) between the members of the community might promote adaptation to environmental conditions [121]. Some studies suggest that conjugation events could be taking place in the cryosphere, for example with the occurrence of UV-tolerance operons and mercury resistance genes carried by plasmids in bacteria in glacier, sea-ice and snow [122,123]. In addition, ice-binding protein genes might be transferred from bacteria to algae through horizontal gene transfer in sea ice and snow. This putative HGT within the snow and ice microbial community still needs confirmatory experiments in order to firmly establish the causation. Viral-host interactions and transduction events might also take place in snow and ice. In the last twenty years, a variety of reports have pointed to an unforeseen high abundance and high diversity of viruses in cold environments [124–129]. Abundance and diversity values seem to be highly variable, with cryoconite hole sediments having the highest viral-like particle (VLP) numbers (2.5×10^9 VLP g⁻¹) [130]. Several studies have found newly undescribed viral groups for sea ice and supra-glacial systems [127,131]. In sea ice, siphovirus and myoviruses were identified as phages of *Flavobacterium* sp. and *Shewanella* sp., respectively [132]. In glacial and sea ice, with most bacteria and viruses in the veins of liquid water, contact rates between viruses and bacteria might be higher than estimated in bulk ice [133]. A recent study suggested that viral interactions with their host in Arctic glacial ice were more dynamic than in nearby soil [134], while high concentrations of extracellular DNA and viruses have been described in sea-ice [135]. Thus, based on high viral abundance, diversity, infection rates [136] and broad host ranges [137], viruses have been promoted as drivers of evolution in glaciers and, in general, in cold environments [138]. Viruses also influence nutrient and organic carbon recycling through the lysis of microbial cells in what has been named the “viral shunt” [139]. This process might have an increased relevance in oligotrophic environments such as snow and ice. Although no study has attempted to determine the viral diversity in snow, prophages were induced from bacteria (*Paenibacillus* spp.) recovered from the snow and all bacteria tested had up to 3 morphologically different inducible phages [140]. Viral sequences accounting for less than 0.1% of the total sequences have also been detected in public snow metagenomes [20]. In snow, algal–bacterial interactions have been described where bacteria benefit from the organic carbon excreted by the algae [141]. This “commensalistic” relationship has also been reported in sea ice [142]. For example, bacteria were seen to consume the toxic hydrogen peroxide produced by diatoms during photosynthesis [143]. A large diversity of non-microbial organisms has also been observed in snow and ice. Meiofauna identified in glacial and lake ice as well as in sea ice includes nematodes, copepods, rotifers and tardigrades, among others [144–146]. Even macrofauna, including several annelid species of the genus

Mesenchytraeus, (snow or ice worms), have been recurrently found inhabiting glacial ice and snow, and actually live their entire lives in these systems [147]. They graze on microbial algae [148] and maybe also bacteria, and thus, control microbial populations. *Mesenchytraeus solifugus* exhibit bacteria associated with their gut walls in what seems to be a symbiotic relationship [149]. Thus, the macrofauna also provide yet another niche for microorganisms in snow and ice. A hypothetical network of trophic interactions in snow and ice ecosystems is presented in Fig. 4. As depicted above, interactions between organisms can promote the settlement of communities and can assist in the adaptation of its members. So community interactions can reinforce the capacity of microbial inhabitants to face the stresses of snow and ice ecosystems.

3.3. Activity

In laboratory experiments, the coldest temperature where bacterial growth has been observed is $-15\text{ }^{\circ}\text{C}$ [150] and activity was detected at temperatures as low as $-40\text{ }^{\circ}\text{C}$ [151]. Comparison of temperature dependence of metabolic rates for microbial communities from diverse environments suggests that, at temperatures far below the freezing point, liquid water inside ice is available for metabolism [152]. However, only a few data are available on in situ microbial activity in frozen water habitats. Microbial activity in snowpack, as measured by labeled thymidine and leucine incorporation, can be variable depending on temperature, location and the proximity of anthropogenic activities [153], and ranges from $4.2\text{ fmol L}^{-1}\text{ h}^{-1}$ in fresh fallen snow to more than $160\text{ pmol L}^{-1}\text{ h}^{-1}$ in Antarctic sites with clear anthropogenic effects [154–156]. Due to variations in temperature and water content in laboratory measurements, these measurements

might not provide evidence of in situ metabolism for extremely low temperature habitats such as Antarctic plateau snow fields [154,157]. However, microbial activity in the snow slush layer of an alpine lake has been shown to be much higher than the underlying lake water [153]. Actively respiring cells, labeled using fluorescent electron transport system-specific reagent CTC, were observed within brine in winter sea-ice section at $-20\text{ }^{\circ}\text{C}$. Carbon bacterial production measured in the brine of thick Antarctic pack ice was as high as $0.4\text{ }\mu\text{g of C liter}^{-1}\text{ h}^{-1}$ and was forty times higher than the earlier open water [158,211]. Within glacial ice, ice veins might not be the only habitat able to support microbial life. Microbes might be also present outside veins, like on the surface of trapped minerals [23]. Laboratory ice formation simulation suggested that while larger particles ($5\text{ }\mu\text{m}$ and $10\text{ }\mu\text{m}$) get trapped in the bulk ice, microbial cells and smaller particles are excluded from the ice and go into the veins [82]. However, microorganisms might also survive trapped directly in ice crystals by a coat (1 nm) of water that would be sufficient for nutrient exchange [84]. Microorganisms might also be associated with mineral particles and may extract energy from redox reactions with putative methane production based on observations from ice cores, where methanogenic archaea have been detected in correlation with methane excess [159]. Within deep glacial ice, microbial activity is presumed to be very low, but sufficient to maintain cell components and methanogen metabolism [77].

The glacial ice surface is an important component of glaciers and ice-sheet with occasional high microbial activity [77,160,161]. For example, cryoconite holes, already mentioned in this review, are unique niches on glacier surfaces. From the Greek *kruos* (ice) and *konis* (dust), cryoconites are small aggregate granules of mineral particles, organic matter and microorganisms [162]. These granules absorb light,

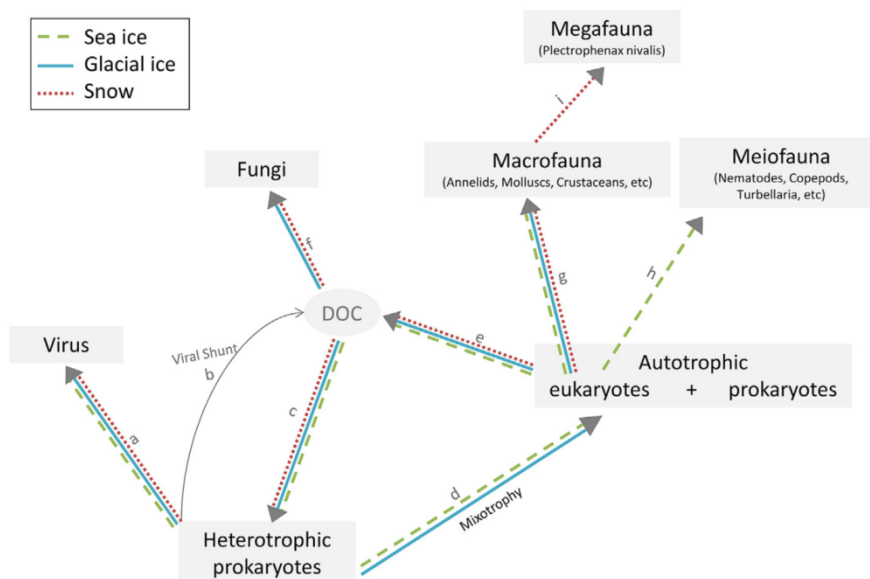


Fig. 4. Food web in snow and ice environments. Green dashed arrows represent the interactions present in sea ice, while blue continuous arrows present the interactions in glacial ice and red-dotted arrows represent those in snow. The gray arrow represents theoretical assumptions not yet described in these environments. Data from a. [131,136,140]; b. [139]; c. [7,86,105]; d. [106,207]; e. [64,73,76]; f. [86,106]; g. [147,208]; h. [209]; i. [210].

leading to the melting of snow and ice and the creation of cryoconite holes that are filled with liquid water and a bottom layer of sediment. This unique environment characterized by high water availability and nutrient concentration is extensively colonized by microorganisms. The greatest abundance of microbial cells registered in cryoconite hole sediments was up to 10^8 cells/g, which is almost equivalent to that of soil communities in temperate regions [163]. Filamentous cyanobacteria, along with algal and bacterial EPS production, determine the size and stability of these aggregates [164]. These aggregates trap mineral particles inside for potential use as nutrient sources and photosynthetic primary production sustains heterotrophic bacteria [85] along with fungi and archaea [19,104]. Microbial activity measured as several μg of carbon per gram of cryoconite aggregates per day for photosynthesis and respiration during summer is similar to that in soils and sediments in temperate regions [77]. High concentrations of snow algae (*Chlamydomonas nivalis* and related species) can be found in snowpacks from various locations (from arctic pack ice snow [165,166] to Alpine systems like the Rocky Mountains [167] or the Austrian Alps [73]) producing the red snow or watermelon snow phenomenon. Highly active cells of *Ancylonema nordenskiöldii* algae were isolated from arctic glacial ice [64], which is consistent with these photosynthetic microbes being relevant in carbon input in snow and ice habitats. However, these niches with high microbial abundance and activity are periodic and ephemeral; i.e., most cryoconites freeze again during the cold period [168,169] and algal are probably restrained to specific conditions. Polar snow and ice microorganisms are subjected to total absence of light during winter. Photosynthetic organisms and activity in deeper layers of glacier ice cores are rarely detected, likely due to lower levels of irradiation [17]. How the phototrophic members of the community survive through winter and are influenced by this low light regime is still largely unknown. The occurrence of mixotrophic algae, able to sustain their metabolism through bacterivory, has been described in Arctic sea ice samples at the end of the polar night as an alternative pathway [106].

We already mentioned in the previous section that the proportion of highly specialized microbes in the total snow community remains unclear. Likewise, their activity in situ and their contribution to total microbial transformations and fluxes in snow and ice are unknown. More generally, we can wonder about the discrepancy between the total present microbial community and its active members in these types of ecosystems, whether it is higher as compared to more conventional environments.

4. How to study these non-conventional habitats?

The field of research in cryospheric microbial ecology is directly influenced by the effect of extreme conditions on humans doing the sampling and analyses, due to technical and logistical issues more than to the microorganisms themselves. Access to the sampling site and the sampling itself can be complicated and difficult. This is even more critical during the

polar winter, as illustrated by the strong bias towards spring and summer microbial characterization of snow and ice polar habitats (Fig. 3). Sample management can also greatly interfere with natural community condition observations and these potential biases are major concerns for cryosphere microbial ecologists. Deep ice core studies can be limited by drilling difficulties such as contamination. Recently, fluorescent microspheres were developed as tracers for assessing potential microbial contamination from indigenous microorganisms in glacial ice during drilling [170,171]. The study of frozen environments is dependent on, and probably highly biased by, melting procedures. First, it destroys the 3D structure of samples, even though this structure might be a major factor in snow and ice microbial distribution and function. Non-destructive methods, such as intact ice section microscopy, have shown that the majority of active cells are attached to surfaces within the brine channel network of winter arctic sea ice [158,212]. Accessing the microscopic structure of snow and ice habitats would certainly help understand the ecology of their microbial inhabitants. Secondly, one strategy to cope with low nutrients is to reduce cell size, since small sizes increase their surface to volume ratio and thereby improve the ability to efficiently scavenging for nutrients [172]. Both a dominance of cells with a size inferior to $0.1 \mu\text{m}$ and numerous isolates that were closely related to the oligotrophic ultramicrobacterium *Sphingopyxis alaskensis* were observed in a Greenland glacial ice core [173]. Due to the potential small cell size of cryosphere microorganisms, the microbial abundance of frozen water ecosystems like snow and ice might be underestimated, since methods used for cell count measurements generally involve filtration, and the entire community may not be retained by the filters [173,174]. Technological and methodological improvements, in particular, in order to determine active members, would increase the extent to which we can probe the snow and ice and their biologically diverse communities.

5. Conclusion

Here we focused on the characterization of snow and ice ecosystems and especially how microorganisms interact with their complex abiotic habitat under specific conditions in terms of irradiation, water activity, salinity and nutrient content. Although many adaptive mechanisms remain to be described, snow and ice inhabitants possess a large set of tools to deal with these variable and intense conditions, which might not be so extreme from an adapted microbial point of view. Snow and ice microbial communities do not seem limited by their specific conditions, at least in superficial environments, and have numerous and diverse members involved in a complex network of interactions. Although their production rates might be low compared to other major biosphere compartments, such as ocean and soil, snow and ice ecosystems might sustain an underestimated part of global biogeochemical activity. Future research focusing on snow and ice ecology should expand past resistance studies to functional studies of these ecosystems.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

LM was supported in part by the Region Rhone-Alpes (France). LS was funded from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement no. [289949]. This research is part of the European project Trainbiodiverse and was funded by Marie Curie Actions.

References

- [1] Rothschild LJ, Mancinelli RL. Life in extreme environments. *Nature* 2001;409:1092–101.
- [2] Brock TD. Life at high temperatures. *Science* 1985;230:132–8.
- [3] Margesin R, Miteva V. Diversity and ecology of psychrophilic microorganisms. *Res Microbiol* 2011;162:346–61.
- [4] Deming JW. Sea ice bacteria and viruses. In: Thomas DN, Dieckmann GS, editors. *Sea ice – an introd. to its physics, chemistry, biol. geol.* Oxford, UK: Wiley-Blackwell; 2009. p. 247–82.
- [5] Miteva V. Bacteria in snow and glacier ice. In: Margesin Rosa, Schinner F, Marx J-C, Gerday C, editors. *Psychrophiles from biodivers. to biotechnol.*; 2008.
- [6] Larose C, Dommergue A, Vogel T. The dynamic arctic snow pack: an unexplored environment for microbial diversity and activity. *Biology (Basel)* 2013;2:317–30.
- [7] Arrigo KR. Sea ice ecosystems. *Ann Rev Mar Sci* 2014;6:439–67.
- [8] Marshall SJ. *The cryosphere*. Princeton University Press; 2011.
- [9] GLIMS, National Snow and Ice Data Center. *GLIMS Glacier Database*. Boulder, Colorado, USA: National Snow and Ice Data; 2012.
- [10] Tansley A. The use and abuse of vegetational concepts and terms. *Ecology* 1935;16:284–307.
- [11] Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen–Geiger climate classification. *Hydrol Earth Syst Sci* 2007;11:1633–44.
- [12] Barry RG, Jania J, Birkenmajer K. Review article “A. B. Dobrowolski – the first cryospheric scientist – and the subsequent development of cryospheric science.” *Hist Geo Space Sci* 2011;2:75–9.
- [13] Morita RY. Psychrophilic bacteria. *Bacteriol Rev* 1975;39:144–67.
- [14] Feller G, Gerday C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* 2003;1:200–8.
- [15] De Maayer P, Anderson D, Cary C, Cowan DA. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* 2014;15:508–17.
- [16] Casanueva A, Tuffin M, Cary C, Cowan DA. Molecular adaptations to psychrophily: the impact of “omic” technologies. *Trends Microbiol* 2010;18:374–81.
- [17] Simon C, Wiezer A, Strittmatter AW, Daniel R. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl Environ Microbiol* 2009;75:7519–26.
- [18] Varin T, Lovejoy C, Jungblut AD, Vincent WF, Corbeil J. Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Appl Environ Microbiol* 2012;78:549–59.
- [19] Edwards A, Pachebat JA, Swain M, Hegarty M, Hodson AJ, Irvine-Fynn TDL, et al. A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. *Environ Res Lett* 2013;8:035003.
- [20] Maccario L, Vogel TM, Larose C. Potential drivers of microbial community structure and function in Arctic spring snow. *Front Microbiol* 2014;5:413.
- [21] Junge K, Imhoff F, Staley T, Deming JW. Phylogenetic diversity of numerically important Arctic sea-ice bacteria cultured at subzero temperature. *Microb Ecol* 2002;43:315–28.
- [22] Amato P, Hennebelle R, Magand O, Sancelme M, Delort A-M, Barbante C, et al. Bacterial characterization of the snow cover at Spitzberg, Svalbard. *FEMS Microbiol Ecol* 2007;59:255–64.
- [23] Price PB. Microbial life in glacial ice and implications for a cold origin of life. *FEMS Microbiol Ecol* 2007;59:217–31.
- [24] Price PB. Microbial genesis, life and death in glacial ice. *Can J Microbiol* 2009;55:1–11.
- [25] Miyakawa S, Cleaves HJ, Miller SL. The cold origin of life: hydrogen cyanide and formamide. *Orig Life Evol Biosph* 2002;32:195–208.
- [26] Trinks H, Schröder W, Biebricher CK. Ice and the origin of life. *Orig Life Evol Biosph* 2005;35:429–45.
- [27] Menor-Salván C, Marín-Yaseli MR. Prebiotic chemistry in eutectic solutions at the water–ice matrix. *Chem Soc Rev* 2012;41:5404–15.
- [28] Lehman N. Origin of life: cold-hearted RNA heats up life. *Nat Chem* 2013;5:987–9.
- [29] Galtier N, Tourasse N, Gouy M. A non hyperthermophilic common ancestor to extant life forms. *Science* 1999;283:220–1.
- [30] Krembs C, Gradinger R, Spindler M. Implications of brine channel geometry and surface area for the interaction of sympagic organisms in Arctic sea ice. *J Exp Mar Biol Ecol* 2000;243:55–80.
- [31] Ewert M, Deming J. Sea ice microorganisms: environmental constraints and extracellular responses. *Biology (Basel)* 2013;2:603–28.
- [32] Barletta RE, Priscu JC, Mader HM, Jones WL, Roe CH. Chemical analysis of ice vein microenvironments: II. Analysis of glacial samples from Greenland and Antarctica. *J Glaciol* 2012;58:1109–18.
- [33] Dani KGS, Mader HM, Wolff EW, Wadham JL. Modelling the liquid–water vein system within polar ice sheets as a potential microbial habitat. *Earth Planet Sci Lett* 2012;333–334:238–49.
- [34] Irvine-fynn TDL, Hodson AJ, Moorman BJ, Vatne G, Hubbard AL. Polythermal glacier hydrology: a review. *Rev Geophys* 2011;49:1–37.
- [35] Hodson A, Anesio AM, Tranter M, Fountain A, Osborn M, Priscu JC, et al. Glacial ecosystems. *Ecol Monogr* 2008;78:41–67.
- [36] Techel F, Pielmeier C. Point observations of liquid water content in wet snow – investigating methodical, spatial and temporal aspects. *Cryosphere* 2011;5:405–18.
- [37] Colbeck SC. An overview of seasonal snow metamorphism. *Rev Geophys* 1982;20:45.
- [38] Roberts MF. Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst* 2005;1:5.
- [39] Brown AD. Compatible solutes and extreme water stress in eukaryotic micro-organisms. *Adv Microb Physiol* 1978;17:181–242.
- [40] Kirst GO. Osmotic adjustment in phytoplankton and macroalgae. In: Keller MD, Kiene RP, Kirst GO, Visscher PT, editors. *Biol. environ. chem. DMSF relat. sulfonium compd.* Springer, Plenum Press; 1998. p. 121–9.
- [41] Ewert M, Deming JW. Bacterial responses to fluctuations and extremes in temperature and brine salinity at the surface of Arctic winter sea ice. *FEMS Microbiol Ecol* 2014;89:476–89.
- [42] Saum SH, Pfeiffer F, Palm P, Rapp M, Schuster SC, Müller V, et al. Chloride and organic osmolytes: a hybrid strategy to cope with elevated salinities by the moderately halophilic, chloride-dependent bacterium *Halobacillus halophilus*. *Environ Microbiol* 2013;15:1619–33.
- [43] Aslam SN, Cresswell-Maynard T, Thomas DN, Underwood GJC. Production and characterization of the intra- and extracellular carbohydrates and polymeric substances (EPS) of three sea-ice diatom species, and evidence for a cryoprotective role for EPS. *J Phycol* 2012;48:1494–509.
- [44] Liu S-B, Chen X-L, He H-L, Zhang X-Y, Xie B-B, Yu Y, et al. Structure and ecological roles of a novel exopolysaccharide from the arctic sea ice bacterium *Pseudoalteromonas* sp. strain SM20310. *Appl Environ Microbiol* 2012;79:224–30.
- [45] Eddie B, Krembs C, Neuer S. Characterization and growth response to temperature and salinity of psychrophilic, halotolerant *Chlamydomonas* sp. ARC isolated from Chukchi Sea ice. *Mar Ecol Prog Ser* 2008;354:107–17.

- [46] Gunde-Cimerman N, Sonjak S, Zalar P, Frisvad JC, Diderichsen B, Plemenitaš A. Extremophilic fungi in arctic ice: a relationship between adaptation to low temperature and water activity. *Phys Chem Earth Parts A/B/C* 2003;28:1273–8.
- [47] Llubes M, Lanseau C, Rémy F. Relations between basal condition, subglacial hydrological networks and geothermal flux in Antarctica. *Earth Planet Sci Lett* 2006;241:655–62.
- [48] Kyrke-Smith TM, Katz RF, Fowler AC. Subglacial hydrology and the formation of ice streams. *Proc R Soc A* 2014;470:20130494.
- [49] Dowdeswell JA, Siegert MJ. The physiography of modern Antarctic subglacial lakes. *Glob Planet Change* 2003;35:221–36.
- [50] Perovich D. Ultraviolet radiation and the optical properties of sea ice and snow. In: Hessen DO, editor. *UV radiat. Arct. ecosyst.*, vol. 153. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002. p. 73–89.
- [51] Grannas AM, Jones AE, Dibb J, Ammann M, Anastasio C, Beine HJ, et al. An overview of snow photochemistry: evidence. *Mech Impacts* 2007;4329–73.
- [52] Domine F, Albert M, Huthwelker T, Jacobi H, Kokhanovsky AA, Lehning M, et al. Snow physics as relevant to snow photochemistry 2008;10:171–208.
- [53] Cockell CS, Rettberg P, Horneck G, Wynn-Williams DD, Scherer K, Gugg-Helminger A. Influence of ice and snow covers on the UV exposure of terrestrial microbial communities: dosimetric studies. *J Photochem Photobiol B Biol* 2002;68:23–32.
- [54] King MD, France JL, Fisher FN, Beine HJ. Measurement and modelling of UV radiation penetration and photolysis rates of nitrate and hydrogen peroxide in Antarctic sea ice: an estimate of the production rate of hydroxyl radicals in first-year sea ice. *J Photochem Photobiol A Chem* 2005;176:39–49.
- [55] Moan J, Peak MJ. Effects of UV radiation on cells. *J Photochem Photobiol B Biol* 1989;4:21–34.
- [56] Santos AL, Oliveira V, Baptista I, Henriques I, Gomes NCM, Almeida A, et al. Wavelength dependence of biological damage induced by UV radiation on bacteria. *Arch Microbiol* 2013;195:63–74.
- [57] Mundy CJ, Gosselin M, Ehn JK, Belzile C, Poulin M, Alou E, et al. Characteristics of two distinct high-light acclimated algal communities during advanced stages of sea ice melt. *Polar Biol* 2011;34:1869–86.
- [58] Uusikivi J, Vähätalo AV, Granskog MA, Sommaruga R. Contribution of mycosporine-like amino acids and colored dissolved and particulate matter to sea ice optical properties and ultraviolet attenuation. *Limnol Oceanogr* 2010;55:703–13.
- [59] Ryan KG, McMinn A, Mitchell KA, Trenerry L. Mycosporine-like amino acids in Antarctic sea ice algae, and their response to UVB radiation. *Z Naturforsch C* 2002;57:471–7.
- [60] Lemoine Y, Schoefs B. Secondary ketocarotenoid astaxanthin biosynthesis in algae: a multifunctional response to stress. *Photosynth Res* 2010;106:155–77.
- [61] Duval B, Shetty K, Thomas WH. Phenolic compounds and antioxidant properties in the snow alga *Chlamydomonas nivalis* after exposure to UV light 2000;559–66.
- [62] Gorton HL, Vogelmann TC. Ultraviolet radiation and the snow alga *Chlamydomonas nivalis* (Bauer) Wille. *Photochem Photobiol* 2003;77:608–15.
- [63] Remias D, Wastian H, Lütz C, Leya T. Insights into the biology and phylogeny of *Chloromonas polyptera* (Chlorophyta), an alga causing orange snow in Maritime Antarctica. *Antarct Sci* 2013;25:648–56.
- [64] Remias D, Holzinger A, Aigner S, Lütz C. Ecophysiology and ultrastructure of *Ancylonema nordenskiöldii* (Zygnematales, Streptophyta), causing brown ice on glaciers in Svalbard (high Arctic). *Polar Biol* 2012;35:899–908.
- [65] Remias D, Schwaiger S, Aigner S, Leya T, Stuppner H, Lütz C. Characterization of an UV- and VIS-absorbing, purpurogallin-derived secondary pigment new to algae and highly abundant in *Mesotaenium berggrenii* (Zygnematophyceae, Chlorophyta), an extremophyte living on glaciers. *FEMS Microbiol Ecol* 2011;79:638–48.
- [66] Martin A, Hall J, Ryan K. Low salinity and high-level UV-B radiation reduce single-cell activity in Antarctic sea ice bacteria. *Appl Environ Microbiol* 2009;75:7570–3.
- [67] Gessler NN, Egorova AS, Belozerskaya TA. Melanin pigments of fungi under extreme environmental conditions (Review) *Appl Biochem Microbiol* 2014;50:105–13.
- [68] Solomon PS, Waters ODC, Oliver RP. Decoding the mannitol enigma in filamentous fungi. *Trends Microbiol* 2007;15:257–62.
- [69] Ziegelhoffer EC, Donohue TJ. Bacterial responses to photo-oxidative stress 2009;7:856–63.
- [70] Sinha RP, Häder DP. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* 2002;1:225–36.
- [71] Thomas DN, Papadimitriou S. Biogeochemistry of sea ice. In: Thomas DN, Dieckmann GS, editors. *Sea ice – an intro. to its physics, chem. biol. geol.* Blackwell Publishing; 2008. p. 267–302.
- [72] Larose C, Dommergue A, De Angelis M, Cossa D, Averty B, Maruszczak N, et al. Springtime changes in snow chemistry lead to new insights into mercury methylation in the Arctic. *Geochim Cosmochim Acta* 2010;74:6263–75.
- [73] Remias D, Lütz-meindl U, Lütz C, Lu U. Photosynthesis, pigments and ultrastructure of the alpine snow alga *Chlamydomonas nivalis*. *Eur J Phycol* 2005;40:259–68.
- [74] Yallop ML, Anesio AM, Perkins RG, Cook J, Telling J, Fagan D, et al. Photophysiology and albedo-changing potential of the ice algal community on the surface of the Greenland ice sheet. *ISME J* 2012;6:2302–13.
- [75] Fritsen CH, Priscu JC. Cyanobacterial assemblages in permanent ice covers on Antarctic lakes: distribution, growth rate, and temperature response of photosynthesis. *J Phycol* 1998;34:587–97.
- [76] Koh EY, Martin AR, McMinn A, Ryan KG. Recent advances and future perspectives in microbial phototrophy in Antarctic sea ice. *Biology (Basel)* 2012;1:542–56.
- [77] Anesio AM, Hodson AJ, Fritz A, Psenner R, Sattler B. High microbial activity on glaciers: importance to the global carbon cycle. *Glob Chang Biol* 2009;15:955–60.
- [78] Skidmore M, Anderson SP, Sharp M, Foght J, Lanoil BD. Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Appl Environ Microbiol* 2005;71:6986–97.
- [79] Boyd ES, Hamilton TL, Havig JR, Skidmore ML, Shock EL. Chemolithotrophic primary production in a subglacial ecosystem. *Appl Environ Microbiol* 2014;80:6146–53.
- [80] Stibal M, Hasan F, Wadham JL, Sharp MJ, Anesio AM. Prokaryotic diversity in sediments beneath two polar glaciers with contrasting organic carbon substrates. *Extremophiles* 2012;16:255–65.
- [81] Stibal M, Šabacká M, Žárský J. Biological processes on glacier and ice sheet surfaces. *Nat Geosci* 2012;5:771–4.
- [82] Mader HM, Pettitt ME, Wadham JL, Wolff EW, Parkes RJ. Subsurface ice as a microbial habitat. *Geology* 2006;34:169.
- [83] Price PB. A habitat for psychrophiles in deep Antarctic ice. *Proc Natl Acad Sci U S A* 2000;97:1247–51.
- [84] Rohde RA, Price PB. Diffusion-controlled metabolism for long-term survival of single isolated microorganisms trapped within ice crystals. *Proc Natl Acad Sci U S A* 2007;104:16592–7.
- [85] Takeuchi N. Structure, formation, and darkening process of albedo-reducing material cryoconite on a Himalayan glacier: a granular algal mat growing on the glacier. *Arct Antarct Alp Res* 2001;33:115–22.
- [86] Margesin R, Zacke G, Schinner F. Characterization of heterotrophic microorganisms in Alpine glacier cryoconite. *Arct Antarct Alp Res* 2002;34:88–93.
- [87] Barrie LA. Snow formation and processes in the atmosphere that influence its chemical composition. In: Davies TD, Tranter M, Jones HG, editors. *Seas. snowpacks SE – 1*, vol. 28. Springer Berlin Heidelberg; 1991. p. 1–20.
- [88] Jones HG. The ecology of snow-covered systems: a brief overview of nutrient cycling and life in the cold. *Hydrol Process* 1999;13:2135–47.
- [89] Larose C, Dommergue A, Vogel TM. Microbial nitrogen cycling in Arctic snowpacks. *Environ Res Lett* 2013;8:035004.
- [90] Amoroso A, Domine F, Esposito G, Morin S, Savarino J, Nardino M, et al. Microorganisms in dry polar snow are involved in the exchanges

- of reactive nitrogen species with the atmosphere. *Environ Sci Technol* 2010;44:714–9.
- [91] Stibal M, Bælum J, Holben WE, Sørensen SR, Jensen A, Jacobsen CS. Microbial degradation of 2,4-dichlorophenoxyacetic acid on the Greenland ice sheet. *Appl Environ Microbiol* 2012;78:5070–6.
- [92] Antony R, Mahalinganathan K, Krishnan KP, Thamban M. Microbial preference for different size classes of organic carbon: a study from Antarctic snow. *Environ Monit Assess* 2012;184:5929–43.
- [93] Vincent WF, Mueller DR, Bonilla S. Ecosystems on ice: the microbial ecology of Markham Ice Shelf in the high Arctic. *Cryobiology* 2004;48:103–12.
- [94] Brown JR, Seymour JD, Brox TI, Skidmore ML, Wang C, Christner BC, et al. Recrystallization inhibition in ice due to ice binding protein activity detected by nuclear magnetic resonance. *Biotechnol Rep* 2014;3:60–4.
- [95] Achberger AM, Brox TI, Skidmore ML, Christner BC. Expression and partial characterization of an ice-binding protein from a bacterium isolated at a depth of 3,519 m in the Vostok ice core, Antarctica. *Front Microbiol* 2011;2:1–8.
- [96] Krembs C, Eicken H, Deming JW. Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic. *Proc Natl Acad Sci U S A* 2011;108:3653–8.
- [97] Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* 2006;103:626–31.
- [98] Segawa T, Miyamoto K, Ushida K, Agata K, Okada N, Kohshima S. Seasonal change in bacterial flora and biomass in mountain snow from the Tateyama Mountains, Japan, analyzed by 16S rRNA gene sequencing and real-time PCR 2005;71:123–30.
- [99] Liu Y, Yao T, Jiao N, Kang S, Xu B, Zeng Y, et al. Bacterial diversity in the snow over Tibetan Plateau Glaciers. *Extremophiles* 2009;13:411–23.
- [100] Møller AK, Søborg DA, Al-soud WA, Sørensen SJ, Kroer N. Bacterial community structure in High-Arctic snow and freshwater as revealed by pyrosequencing of 16S rRNA genes and cultivation. *Polar Res* 2013;1:1–11.
- [101] Cameron KA, Hagedorn B, Dieser M, Christner BC, Choquette K, Sletten R, et al. Diversity and potential sources of microbiota associated with snow on western portions of the Greenland Ice Sheet. *Environ Microbiol* 2015;17:594–609.
- [102] Cowie ROM, Maas EW, Ryan KG. Archaeal diversity revealed in Antarctic sea ice. *Antarct Sci* 2011;23:531–6.
- [103] Cameron KA, Hodson AJ, Osborn AM. Structure and diversity of bacterial, eukaryotic and archaeal communities in glacial cryoconite holes from the Arctic and the Antarctic. *FEMS Microbiol Ecol* 2012;82:254–67.
- [104] Zarsky JD, Stibal M, Hodson A, Sattler B, Schostag M, Hansen LH, et al. Large cryoconite aggregates on a Svalbard glacier support a diverse microbial community including ammonia-oxidizing archaea. *Environ Res Lett* 2013;8:035044.
- [105] Harding T, Jungblut AD, Lovejoy C, Vincent WF. Microbes in high arctic snow and implications for the cold biosphere. *Appl Environ Microbiol* 2011;77:3234–43.
- [106] Bachy C, López-García P, Vereshchaka A, Moreira D. Diversity and vertical distribution of microbial eukaryotes in the snow, sea ice and seawater near the North Pole at the end of the polar night. *Front Microbiol* 2011;2:1–12.
- [107] Lutz S, Anesio AM, Jorge Villar SE, Benning LG. Variations of algal communities cause darkening of a Greenland glacier. *FEMS Microbiol Ecol* 2014;89:402–14.
- [108] Foreman CM, Cory RM, Morris CE, SanClements MD, Smith HJ, Lisle JT, et al. Microbial growth under humic-free conditions in a supraglacial stream system on the Cotton Glacier, Antarctica. *Environ Res Lett* 2013;8:1–11.
- [109] Keskitalo J, Leppäranta M, Arvola L. First records of primary producers of epiglacial and supraglacial lakes in western Dronning Maud Land, Antarctica. *Polar Biol* 2013;36:1441–50.
- [110] Collins RE, Rocap G, Deming JW. Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ Microbiol* 2010;12:1828–41.
- [111] Hell K, Edwards A, Zarsky J, Podmirseg SM, Girdwood S, Pachebat JA, et al. The dynamic bacterial communities of a melting High Arctic glacier snowpack. *ISME J* 2013;7:1814–26.
- [112] Larose C, Prestat E, Cecillon S, Berger S, Malandain C, Lyon D, et al. Interactions between snow chemistry, mercury inputs and microbial population dynamics in an Arctic snowpack. *PLoS One* 2013;8:e79972.
- [113] Bowman JP, Nichols DS, McKeekin TA. *Psychrobacter glacincolia* sp. nov., a halotolerant, psychrophilic bacterium isolated from Antarctic sea ice. *Syst Appl Microbiol* 1997;20:209–15.
- [114] Azevedo JSN, Correia A, Henriques S. Molecular analysis of the diversity of genus *Psychrobacter* present within a temperate estuary. *FEMS Microbiol Ecol* 2013;84:451–60.
- [115] Cowan DA, Tow LA. Endangered antarctic environments. *Annu Rev Microbiol* 2004;58:649–90.
- [116] Reddy TBK, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* 2014;43:D1099–106.
- [117] Smith JJ, Howington JP, McFeters GA. Survival, physiological response and recovery of enteric bacteria exposed to a polar marine environment. *Appl Environ Microbiol* 1994;60:2977–84.
- [118] Preyer JM, Oliver JD. Starvation-induced thermal tolerance as a survival mechanism in a psychrophilic marine bacterium. *Appl Environ Microbiol* 1993;59:2653–6.
- [119] Possmayer M, Berardi G, Beall BFN, Trick CG, Hüner NPA, Maxwell DP. Plasticity of the psychrophilic green alga *Chlamydomonas raudensis* (UWO 241) (Chlorophyta) to supraoptimal temperature stress. *J Phycol* 2011;47:1098–109.
- [120] Konopka A. What is microbial community ecology? *ISME J* 2009;3:1223–30.
- [121] Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;3:722–32.
- [122] Dziejwit L, Grzesiak J, Ciok A, Nieckarz M, Zdanowski MK, Bartosik D. Sequence determination and analysis of three plasmids of *Pseudomonas* sp. GLE121, a psychrophile isolated from surface ice of Ecology Glacier (Antarctica). *Plasmid* 2013;70:254–62.
- [123] Moller AK, Barkay T, Hansen MA, Norman A, Hansen LH, Sorensen SJ, et al. Mercuric reductase genes (merA) and mercury resistance plasmids in High Arctic snow, freshwater and sea-ice brine. *FEMS Microbiol Ecol* 2014;87:52–63.
- [124] Kepner RL, Wharton RA, Suttle CA. Viruses in Antarctic lakes. *Limnol Oceanogr* 1998;43:1754–61.
- [125] López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. High diversity of the viral community from an Antarctic lake. *Science* 2009;326:858–61.
- [126] Laybourn-Parry J, Marshall WA, Madan NJ. Viral dynamics and patterns of lysogeny in saline Antarctic lakes. *Polar Biol* 2007;30:351–8.
- [127] Bellas CM, Anesio AM. High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles* 2013;17:861–70.
- [128] Maranger R, Bird DF, Juniper SK. Viral and bacterial dynamics in Arctic sea ice during the spring algal bloom near Resolute, NWT, Canada. *Mar Ecol Prog Ser* 1994;111:121–8.
- [129] Zavar-Reza P, Argüello-Astorga GR, Kraberger S, Julian L, Stainton D, Broady PA, et al. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol* 2014;26:132–8.
- [130] Bellas CM, Anesio AM, Telling J, Stibal M, Tranter M, Davis S. Viral impacts on bacterial communities in Arctic cryoconite. *Environ Res Lett* 2013;8:045021.
- [131] Senčilo A, Luhtanen A-M, Saarijärvi M, Bamford DH, Roine E. Cold-active bacteriophages from the Baltic Sea ice have diverse genomes and virus-host interactions. *Environ Microbiol* 2014.

- [132] Luhtanen A-M, Eronen-Rasimus E, Kaartokallio H, Rintala J-M, Autio R, Roine E. Isolation and characterization of phage–host systems from the Baltic Sea ice. *Extremophiles* 2014;18:121–30.
- [133] Wells LE, Deming JW. Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. Brief report *Environ Microbiol* 2006;8:1115–21.
- [134] Sanguino L, Franqueville L, Vogel TM, Larose C. Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiol Ecol* 2015;91(5).
- [135] Collins RE, Deming JW. Abundant dissolved genetic material in Arctic sea ice part I: extracellular DNA. *Polar Biol* 2011;34:1819–30.
- [136] Sävström C, Granéli W, Laybourn-Parry J, Anesio AM. High viral infection rates in Antarctic and Arctic bacterioplankton. *Environ Microbiol* 2007;9:250–5.
- [137] Anesio AM, Mindl B, Laybourn-Parry J, Hodson AJ, Sattler B. Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J Geophys Res* 2007;112:G04S31.
- [138] Anesio AM, Bellas CM. Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends Microbiol* 2011;19:52–7.
- [139] Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea aquatic food webs. *Bioscience* 1999;49:781–8.
- [140] Van Houdt R, Deghorain M, Vermeersch M, Provoost A, Lo Giudice A, Leys N, et al. Characterization of culturable *Paenibacillus* spp. from the snow surface on the high Antarctic Plateau (DOME C) and their dissemination in the Concordia research station. *Extremophiles* 2013;17:565–73.
- [141] Thomas WH, Duval B. Sierra Nevada, California, U.S.A., snow algae: snow albedo changes, algal–bacterial interrelationships, and ultraviolet radiation effects. *Arct Alp Res* 1995;27:389–99.
- [142] Stewart FJ, Fritsen CH. Bacteria–algae relationships in Antarctic sea ice. *Antarct Sci* 2004;16:143–56.
- [143] Hünken M, Harder J, Kirst GO. Epiphytic bacteria on the Antarctic ice diatom *Amphiprora kufferathii* Manguin cleave hydrogen peroxide produced during algal photosynthesis. *Plant Biol* 2008;10:519–26.
- [144] Marquardt M, Kramer M, Carnat G, Werner I. Vertical distribution of sympagic meiofauna in sea ice in the Canadian Beaufort Sea. *Polar Biol* 2011;34:1887–900.
- [145] Shtarkman YM, Koçer ZA, Edgar R, Veerapaneni RS, D'Elia T, Morris PF, et al. Subglacial Lake Vostok (Antarctica) accretion ice contains a diverse set of sequences from aquatic, marine and sediment-inhabiting bacteria and eukarya. *PLoS One* 2013;8:e67221.
- [146] Zawierucha K, Koliccka M, Takeuchi N, Kaczmarek L. What animals can live in cryoconite holes? A faunal review. *J Zool* 2014;1–11.
- [147] Hartzell PL, Nghiem JV, Richio KJ, Shain DH. Distribution and phylogeny of glacier ice worms (*Mesenchytraeus solifugus* and *Mesenchytraeus solifugus rainierensis*). *Can J Zool* 2005;83:1206–13.
- [148] Goodman D. Ecological investigations of ice worms on Casement Glacier, Southeastern Alaska. 1971. vol. Report No.
- [149] Murakami T, Segawa T, Bodington D, Dial R. Census of bacterial microbiota associated with the glacier ice worm *Mesenchytraeus solifugus*. *FEMS Microbiol Ecol* 2015:1–28.
- [150] Mykytczuk NCS, Foote SJ, Omelon CR, Southam G, Greer CW, Whyte LG. Bacterial growth at $-15\text{ }^{\circ}\text{C}$: molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J* 2013;7:1211–26.
- [151] Panikov NS, Flanagan PW, Oechel WC, Mastepanov MA, Christensen TR. Microbial activity in soils frozen to below $-39\text{ }^{\circ}\text{C}$. *Soil Biol Biochem* 2006;38:785–94.
- [152] Price PB, Sowers T. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc Natl Acad Sci U S A* 2004;101:4631–6.
- [153] Felip M, Sattler B, Psenner R, Catalan J. Highly active microbial communities in the ice and snow cover of high mountain lakes. *Appl Environ Microbiol* 1995;61:2394–401.
- [154] Carpenter EJ, Lin S, Capone DG. Bacterial activity in South Pole snow. *Appl Environ Microbiol* 2000;66:4514–7.
- [155] Sattler B, Puxbaum H, Psenner R. Bacterial growth in supercooled cloud droplets. *Geophys Res Lett* 2001;28:239–42.
- [156] Lopatina A, Krylenkov V, Severinov K. Activity and bacterial diversity of snow around Russian Antarctic stations. *Res Microbiol* 2013;164:949–58.
- [157] Warren SG, Hudson SR. Bacterial activity in South Pole snow is questionable. Letters to the Editor 2003;69:6340–2.
- [158] Junge K, Eicken H, Deming JW. Wintertime sea ice bacterial activity at -2 to $-20\text{ }^{\circ}\text{C}$ in Arctic wintertime sea ice 2004;70.
- [159] Tung HC, Bramall NE, Price PB. Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proc Natl Acad Sci U S A* 2005;102:18292–6.
- [160] Anesio AM, Laybourn-Parry J. Glaciers and ice sheets as a biome. *Trends Ecol Evol* 2012;27:219–25.
- [161] Stibal M, Telling J, Cook J, Mak KM, Hodson A, Anesio AM. Environmental controls on microbial abundance and activity on the Greenland ice sheet: a multivariate analysis approach. *Microb Ecol* 2012;63:74–84.
- [162] MacDonell S, Fitzsimons S. The formation and hydrological significance of cryoconite holes. *Prog Phys Geogr* 2008;32:595–610.
- [163] Whitman WB, Coleman DC, Wiebe WJ. Perspective prokaryotes: the unseen majority 1998;95:6578–83.
- [164] Langford HJ, Irvine-Fynn TDL, Edwards A, Banwart SA, Hodson AJ. A spatial investigation of the environmental controls over cryoconite aggregation on Longyearbreen glacier. *Svalbard Biogeosci Discuss* 2014;11:3423–63.
- [165] Gradinger R, Nurnberg D. Snow algal communities on Arctic pack ice floes dominated by *Chlamydomonas nivalis* (Bauer) Wille 1996:35–43.
- [166] Stibal M, Elster J, Sabacká M, Kastovská K. Seasonal and diel changes in photosynthetic activity of the snow alga *Chlamydomonas nivalis* (Chlorophyceae) from Svalbard determined by pulse amplitude modulation fluorometry. *FEMS Microbiol Ecol* 2007;59:265–73.
- [167] Williams WE, Gorton HL, Vogelmann TC. Surface gas-exchange processes of snow algae. *Proc Natl Acad Sci U S A* 2003;100:562–6.
- [168] Leeson AA, Shepherd A, Briggs K, Howat I, Fettweis X, Morlighem M, et al. Supraglacial lakes on the Greenland ice sheet advance inland under warming climate. *Nat Clim Chang* 2015;5:51–5.
- [169] McMillan M, Nienow P, Shepherd A, Benham T, Sole A. Seasonal evolution of supra-glacial lakes on the Greenland Ice Sheet. *Earth Planet Sci Lett* 2007;262:484–92.
- [170] Miteva V, Sowers T, Brenchley J. Penetration of fluorescent microspheres into the NEEM (North Eemian) Greenland ice core to assess the probability of microbial contamination. *Polar Biol* 2013;37:47–59.
- [171] Miteva V, Burlingame C, Sowers T, Brenchley J. Comparative evaluation of the indigenous microbial diversity vs. drilling fluid contaminants in the NEEM Greenland ice core. *FEMS Microbiol Ecol* 2014:1–19.
- [172] Christner BC, Skidmore ML, Priscu JC, Tranter M, Foreman CM. Bacteria in subglacial environments. 2008. p. 51–71.
- [173] Miteva VI, Brenchley JE. Detection and isolation of ultrasmall microorganisms from a 120,000-year-old Greenland glacier ice core. *Appl Environ Microbiol* 2005;71:7806–18.
- [174] Karl D, Bird DF, Björkman K, Houlihan T, Shackelford R, Tupas L. Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* (80-) 1999;286:2144–7.
- [175] Doyle S, Montross S, Skidmore M, Christner B. Characterizing microbial diversity and the potential for metabolic function at $-15\text{ }^{\circ}\text{C}$ in the basal ice of Taylor glacier, Antarctica. *Biology (Basel)* 2013;2:1034–53.
- [176] Miteva VI, Sheridan PP, Brenchley JE. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl Environ Microbiol* 2004;70:202–13.
- [177] Miteva V, Teacher C, Sowers T, Brenchley J. Comparison of the microbial diversity at different depths of the GISP2 Greenland ice core in relationship to deposition climates. *Environ Microbiol* 2009;11:640–56.
- [178] Irvine-Fynn TDL, Edwards A, Newton S, Langford H, Rassner SM, Telling J, et al. Microbial cell budgets of an Arctic glacier surface quantified using flow cytometry. *Environ Microbiol* 2012;14:2998–3012.
- [179] Zeng YX, Yan M, Yu Y, Li HR, He JF, Sun K, et al. Diversity of bacteria in surface ice of Austre Lovénbreen glacier. *Svalbard Arch Microbiol* 2013;195:313–22.

- [180] Cappa F, Cuciù N, Trevisan M, Ferrari S, Puglisi E, Sandro P. Science of the total environment bacterial diversity in a contaminated Alpine glacier as determined by culture-based and molecular approaches. *Sci Total Environ* 2014;497–498:50–9.
- [181] Ball MM, Gómez W, Magallanes X, Rosales R, Melfo A, Yarzabal LA. Bacteria recovered from a high-altitude, tropical glacier in Venezuelan Andes. *World J Microbiol Biotechnol* 2014;30:931–41.
- [182] Zhang XF, Yao TD, Tian LD, Xu SJ, An LZ. Phylogenetic and physiological diversity of bacteria isolated from Puruogangri ice core. *Microb Ecol* 2008;55:476–88.
- [183] Priscu JC, Adams EE, Lyons WB, Voytek MA, Mogk DW, Brown RL, et al. Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* (80-) 1999;286:2141–4.
- [184] Foreman CM, Dieser M, Greenwood M, Cory RM, Laybourn-Parry J, Lisle JT, et al. When a habitat freezes solid: microorganisms overwinter within the ice column of a coastal Antarctic lake. *FEMS Microbiol Ecol* 2011;76:401–12.
- [185] Jungblut AD, Vincent WF, Lovejoy C. Eukaryotes in Arctic and Antarctic cyanobacterial mats. *FEMS Microbiol Ecol* 2012;82:416–28.
- [186] Werner I, Ikävalko J, Schünemann H. Sea-ice algae in Arctic pack ice during late winter. *Polar Biol* 2007;30:1493–504.
- [187] Fiala M, Kuosa H, Kopczyńska EE, Oriol L, Delille D. Spatial and seasonal heterogeneity of sea ice microbial communities in the first-year ice of Terre Adélie area (Antarctica). *Aquat Microb Ecol* 2006;43:95–106.
- [188] Booth BC, Horner RA. Microalgae on the arctic ocean section, 1994: species abundance and biomass. *Deep Sea Res Part II Top Stud Oceanogr* 1997;44:1607–22.
- [189] Møller AK, Barkay T, Abu Al-Soud W, Sørensen SJ, Skov H, Kroer N. Diversity and characterization of mercury-resistant bacteria in snow, freshwater and sea-ice brine from the High Arctic. *FEMS Microbiol Ecol* 2011;75:390–401.
- [190] Carpenter EJ, Lin S, Capone DG. Bacterial activity in South Pole snow 2000;66.
- [191] Hauptmann AL, Stibal M, Bælum J, Sicheritz-Pontén T, Brunak S, Bowman JS, et al. Bacterial diversity in snow on North Pole ice floes. *Extremophiles* 2014;18(6):945–51.
- [192] Ewert M, Carpenter SD, Colangelo-Lillis J, Deming JW. Bacterial and extracellular polysaccharide content of brine-wetted snow over Arctic winter first-year sea ice. *J Geophys Res Ocean* 2013;118:726–35.
- [193] Bowman JS, Rasmussen S, Blom N, Deming JW, Rysgaard S, Sicheritz-Pontén T. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J* 2012;6:11–20.
- [194] Bowman JP, McCammon SA, Brown MV, Nichols DS, McMeekin TA. Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl Environ Microbiol* 1997;63:3068–78.
- [195] Petri R, Imhoff JF. Genetic analysis of sea-ice bacterial communities of the Western Baltic Sea using an improved double gradient method. *Polar Biol* 2014;24:252–7.
- [196] Brinkmeyer R, Knittel K, Jürgens J, Weyland H, Amann R, Helmke E. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl Environ Microbiol* 2003;69:6610–9.
- [197] Brown MV, Bowman JP. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol* 2001;35:267–75.
- [198] Larose C, Berger S, Ferrari C, Navarro E, Dommergue A, Schneider D, et al. Microbial sequences retrieved from environmental samples from seasonal arctic snow and meltwater from Svalbard, Norway. *Extremophiles* 2010;14:205–12.
- [199] Zhang S, Yang G, Wang Y, Hou S. Abundance and community of snow bacteria from three glaciers in the Tibetan Plateau. *J Environ Sci* 2010;22:1418–24.
- [200] Michaud L, Lo Giudice A, Mysara M, Monsieurs P, Raffa C, Leys N, et al. Snow surface microbiome on the high Antarctic plateau (DOME C). *PLoS One* 2014;9:e104505.
- [201] González-Toril E, Amils R, Delmas RJ, Petit J-R, Komárek J, Elster J. Diversity of bacteria producing pigmented colonies in aerosol, snow and soil samples from remote glacial areas (Antarctica, Alps and Andes). *Biogeosci Discuss* 2008;5:1607–30.
- [202] Chuvochina MS, Marie D, Chevaillier S, Petit J-R, Normand P, Alekhina IA, et al. Community variability of bacteria in Alpine snow (Mont Blanc) containing Saharan dust deposition and their snow colonisation potential. *Microbes Environ* 2011;26:237–47.
- [203] Naff CS, Darcy JL, Schmidt SK. Phylogeny and biogeography of an uncultured clade of snow chytrids. *Environ Microbiol* 2013;15:2672–80.
- [204] Collins RE, Carpenter SD, Deming JW. Spatial heterogeneity and temporal dynamics of particles, bacteria, and pEPS in Arctic winter sea ice. *J Mar Syst* 2008;74:902–17.
- [205] Bramucci A, Han S, Beckers J, Haas C, Lanoil B. Composition, diversity, and stability of microbial assemblages in seasonal lake ice, Miquelon Lake, Central Alberta. *Biology (Basel)* 2013;2:514–32.
- [206] Ma L-J, Rogers SO, Catranis CM, Starmer WT. Detection and characterization of ancient fungi entrapped in glacial ice. *Mycologia* 2000;92:286–95.
- [207] Miezcán T, Górniak D, Swiatecki A, Zdanowski M, Tarkowska-Kukuryk M, Adamczuk M. Vertical microzonation of ciliates in cryoconite holes in Ecology Glacier, King George Island. *Polish Polar Res* 2013;34:201–12.
- [208] Brown TA, Belt ST. Identification of the sea ice diatom biomarker IP25 in Arctic benthic macrofauna: direct evidence for a sea ice diatom diet in Arctic heterotrophs. *Polar Biol* 2012;35:131–7.
- [209] Gradinger R, Friedrich C, Spindler M. Abundance, biomass and composition of the sea ice biota of the Greenland Sea pack ice. *Deep Res Part II Top Stud Oceanogr* 1999;46:1457–72.
- [210] Shain DH, Mason TA, Farrell AH, Michalewicz LA. Distribution and behavior of ice worms (*Mesenchytraeus solifugus*) in south-central Alaska. *Can J Zool* 2001;79:1813–21.
- [211] Grossmann S, Dieckmann GS. Bacterial standing stock, activity, and carbon production during formation and growth of sea ice in the Weddell Sea, Antarctica. *Appl Environ Microbiol* 1994;60:2746–53.
- [212] Junge K, Krembs C, Deming J, Stierle A, Eicken H. A microscopic approach to investigate bacteria under in situ conditions in sea-ice samples. *Ann Glaciol* 2001;33:304–10.
- [213] Vancoppenolle M, Meiners KM, Michel C, Bopp L, Brabant F, Carnat G, et al. Role of sea ice in global biogeochemical cycles: emerging views and challenges. *Quat Sci Rev* 2013;79:207–30.

