TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

The specificity, adaptability, and resilience of microbial community in extreme environments

Tianli Ma

Vollständiger Abdruck der von der TUM School of Life Sciences

der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Priv.-Doz. Dr. Klaus Neuhaus

Prüfer der Dissertation: 1. TUM Junior Fellow Li Deng, Ph.D.

2. Prof. Dr. Wolfgang Liebl

Diese Dissertation wurde am 22.09.2021 bei der Technischen Universität München eingereicht

und durch die TUM School of Life Sciences am 24.01.2022 angenommen.

Table of Contents

Zusammenfassung

Mikrobiota (hauptsächlich Bakterien und Viren) ist eine Sammlung ökologischer Gemeinschaften, einschließlich kommensalen, symbiotischen und pathogenen Mikroorganismen. Im Allgemeinen fördern mikrobielle Gemeinschaften in Naturumgebungen den biogeochemischen Kreislauf, indem sie am Elementkreislauf und Energiefluss teilnehmen. In menschenassoziierten Umgebungen sind mikrobielle Gemeinschaften eng mit der menschlichen Gesundheit verbunden, indem sie dem menschlichen Körper neue metabolische Fähigkeiten verleihen und die Entwicklung und Regulierung des Immunsystems unterstützen, und das Ungleichgewicht der mikrobiellen Gemeinschaft könnte sich wiederum auf die menschliche Gesundheit auswirken. Die Forschung zu mikrobiellen Gemeinschaften in extremen Umgebungen ist jedoch begrenzt. Basierend auf der Untersuchung von drei unterschiedlichen Arten von extremen Umgebungen, darunter oligotrophe Alpenseen mit häufigen Wetterstörungen, oligotrophes tiefes Grundwasser mit relativ stabilen Umgebungen, und Darmumgebungen von Patienten mit ösophagealen Erkrankungen, könnte die vorliegende Arbeit die Struktur und das Verhalten mikrobieller Gemeinschaften unter diesen Bedingungen aufdecken.

Erstens beherbergen extreme Umgebungen spezifische mikrobielle Gemeinschaften. Im tiefen Grundwasser und im Darm von Patienten mit Barrett-Ösophagus (BE) und Ösophagus-Adenokarzinom (EAC) veränderten sich der Reichtum (Ace) und die Diversität (Shannon) der Virusgemeinschaften sowie die Diversität und relative Häufigkeit ihrer vorhergesagten bakteriellen Wirte im Vergleich mit ihren normalen Gegenparts. Zweitens können sich mikrobielle Gemeinschaften an die Veränderungen extremer Umgebungen anpassen, indem sie ihre Gemeinschaftsstruktur, ihren Lebensstil und ihre individuellen genetischen Merkmale ändern. In alpinen Seen variierte die relative Häufigkeit von Proteobakterien, Aktinobakterien und Bacteroidetes mit

Umweltveränderungen. Auf Gattungsebene wuchsen *Acinetobacter* mit hoher Toleranz gegenüber niedrigen Temperaturen während der Kühlperiode schnell, während *Limnohabitans* und *Sediminibacterium* mit schwacher Anpassungsfähigkeit und Wettbewerbsfähigkeit relativ gering waren. Mit dem Steigen der Temperatur zeigte die Änderung ihrer relativen Häufigkeit einen gegenteiligen Trend. Im tiefen Grundwasser verringerten sich mit abnehmender Nährstoffversorgung die dominanten Bestandteile der Kernvirusgruppe. Der Anteil der reichlich vorhandenen Viren und der seltenen Viren veränderte sich dementsprechend. In human-assoziierten Umgebungen, mit der Entwicklung von Ösophaguserkrankungen von BE zu EAC, zeigten reichlich vorhandene Darmviren und seltene Darmviren einen statistisch signifikanten Unterschied. Außerdem neigten Viren dazu, den lysogenen Lebenszyklus im tiefen Grundwasser und im Darm von BE- und EAC-Patienten aufrechtzuerhalten. Darüber hinaus wurden beim Screening der genetischen Merkmale einzelner Viren metabolische Auxiliargene (AMGs) gefunden, die sie von ihren bakteriellen Wirten erhalten hatten, die ihre Überlebenschancen und ihr Gedeihen erhöhen könnten, indem sie Bakterienstämmen helfen, Anpassungsmerkmale unter verschiedenen Umweltbedingungen zu erwerben. Im tiefen Grundwasser wurden tiefenspezifische AMGs gefunden, darunter *coff*, *cysH*, *luxR* und *spo0A*, die sich auf den Bakterienmetabolismus und die Regulierung der Bakteriendichte beziehen. Im Darm von BE- und EAC-Patienten wurde eine etwas höhere Häufigkeit von *spyA* (das mit bakteriellem Exotoxin in Zusammenhang steht) gefunden. Die mikrobielle Gemeinschaft ist widerstandsfähig und könnte in extremen Umgebungen in ihren Zustand vor der Störung zurückkehren. In Alpenseen ist es offensichtlich, dass sich Betaproteobakterien und Gammaproteobakterien mit dem Auftreten und Verschwinden der plötzlichen Abkühlungsperiode schnell verschieben. Unterdessen wiesen die Zusammensetzungen der Bakteriengemeinschaften auf der Ebene der operationalen taxonomischen Einheiten (OTU) zu Beginn und am Ende der Vegetationsperiode eine hohe Ähnlichkeit auf. Darüber hinaus könnten seltene mikrobielle Gemeinschaften die Widerstandsfähigkeit aufgrund geringer Häufigkeit und verschiedener Genotypen widerspiegeln. Im tiefen Grundwasser traten hohe Konzentrationen seltener Viren im oberen Grundwasserleiter auf, was darauf hindeuten kann, dass seltene Viren dazu beitragen, die Widerstandsfähigkeit extremer Ökosysteme zu erhöhen, die anfällig für Störungen sind.

Insgesamt zeigte diese Studie die Veränderungen, Anpassungsfähigkeit und Widerstandsfähigkeit mikrobieller Gemeinschaften sowie die voneinander abhängige Überlebensstrategie von Bakterien und Viren in Alpenseen, tiefem Grundwasser und dem Darm von BE- und EAC-Patienten auf und ergänzt damit das vorhandene Wissen über die mikrobiellen Gemeinschaft und ihre entscheidende Funktion in extremen Umgebungen.

Summary

Microbiota (mainly bacteria and viruses) is a collection of ecological communities, including commensal, symbiotic, and pathogenic microorganisms. Generally, in natural environments, microbial communities promote the biogeochemical cycle by participating in the element cycle and energy flow. In human-associated environments, microbial communities are closely related to human health by providing the human body with new metabolic capabilities and aiding the development and regulation of the immune system. The imbalance of microbial community could, in turn, affect human health. However, the research on microbial communities in extreme environments is limited. Based on studies of three distinct types of extreme environments, including an oligotrophic alpine lake with frequent weather disturbances, oligotrophic deep groundwater with relatively stable environments, and intestinal environments of patients with esophageal-related diseases, the current thesis could reveal the structure and behaviors of microbial communities in these conditions.

Firstly, extreme environments harbor specific microbial communities. In deep groundwater and the gut of Barrett Esophagus (BE) and Esophageal Adenocarcinoma (EAC) patients, the richness (Ace) and diversity (Shannon) of viral communities, as well as the diversity and relative abundance of their predicted bacterial hosts, changed compared with their normal counterparts. Secondly, microbial communities can adapt to extreme environments by altering their community structure, lifestyles, and individual genetic characteristics. In the alpine lake, the relative abundance of Proteobacteria, Actinobacteria, and Bacteroidetes varied with environmental changes. Specifically, genus *Acinetobacter* with high tolerance to low temperature grew fast during the cooling period, while *Limnohabitans and Sediminibacterium* with weak adaptability and competitiveness were relatively low. When the temperature rose, the alteration of their relative abundance showed an opposite trend.

In deep groundwater, as the nutrient supply decreased, the dominant components of the core viral group reduced. Meanwhile, the proportion of abundant viruses and rare viruses changed accordingly. In human-associated environments, with the development of esophageal diseases from BE to EAC, abundant and rare gut viruses showed statistically significant differences. Besides, viruses tended to maintain the lysogenic life cycle in deep groundwater and gut of BE and EAC patients. Moreover, when screening the genetic traits of individual viruses, auxiliary metabolic genes (AMGs) obtained from their bacterial hosts were found, increasing their chances of survival and prosper by aiding bacterial strains to acquire adaptative traits under different environmental characteristics. In deep groundwater, depth-specific AMGs were found, including *cofF*, *cysH*, *luxR,* and *spo0A,* which are related to bacterial metabolism and regulation of bacterial density. In the gut of BE and EAC patients, a slightly higher abundance of *spyA* (related to bacterial exotoxin) was found. Finally, the microbial community is resilient and could return to its pre-disturbance state in extreme environments. In the alpine lake, it was apparent that Betaproteobacteria and Gammaproteobacteria rapidly shifted with the occurrence and disappearance of the sudden cooling period. Meanwhile, at the operational taxonomic units (OTU) level, bacterial community compositions were similar at the beginning and end of the growing season. In addition, the rare microbial community could reflect the resilience due to low abundance and various genotypes. In deep groundwater, high levels of rare viruses appeared in the upper aquifer, indicating rare viruses contribute to enhancing the resilience of extreme ecosystems that are susceptible to disturbance.

Overall, this study revealed the changes, adaptability, and resilience of microbial communities, as well as the interdependent survival strategy of bacteria and viruses in the alpine lake, deep groundwater, and the gut of BE and EAC patients, further supplementing the existing knowledge about microbial communities and their crucial function in extreme environments.

Chapter 1

1 Introduction

1.1 Microbiota

The microbiota was first proposed by Joshua Lederberg, describing it as a collection of ecological communities including commensal, symbiotic, and pathogenic microorganisms (Lederberg and McCray 2001; Peterson et al. 2009). It has been found that microbiota plays critical roles in a variety of areas (Diaz Carrasco et al. 2019; Fan et al. 2019; Van Elsas and Boersma 2011), for example (1) Microbiota promotes material circulation and energy flow in nature environments (Jiao et al. 2019; Shen et al. 2019); (2) In agricultural production, microbiota facilitates the increase of effective nutrients in the soil through nitrogen fixation and mineral transformation, etc., thereby improving soil fertility and supporting high crop yields (Bonanomi et al. 2020; Wu et al. 2021); (3) In industrial production, microbiota and their metabolites or metabolic activities are widely used to produce necessities of human life, such as the use of microbial fermentation to make wine (Bokulich et al. 2016; Huang et al. 2019); (4) Microbiota could degrade pollutants in environments. The biodegradation of microbiota has been used to treat sewage, poison, etc. (Yadav and Yadav 2019); (5) Microbiota is closely related to the health and disease of humans and livestock. Microbiota can synthesize vitamins and amino acids to provide nutrients for humans and livestock, while pathogenic microbiota can induce them to develop various diseases (Bäckhed et al. 2004; Berger and Mainou 2018).

1.1.1 Constituents of microbiota

Microbiota contains bacteria, archaea, protists, fungi, and viruses. Among them, bacteria and viruses are two of the most abundant and genetically diverse entities on earth. Specifically, bacteria are single-celled prokaryotes without a membrane-bound nucleus, whose size range from 0.2 µm

(e.g., *Mycoplasma pneumoniae*) to 750 µm (e.g., *Thiomargarita namibiensis*) (Schulz and Jørgensen 2001). Their genetic material is usually a circular chromosome located in the cytoplasm (Thanbichler et al. 2005). Some bacteria also contain plasmids, which are another type of genetic material that usually carry genes related to resistance or virulence (Kado 2015). According to the structure and composition of the cell wall, bacteria can be divided into Gram-positive and Gram-negative bacteria (Takeuchi et al. 1999). The cell walls of Gram-positive bacteria are thick (20-80 nm) and usually contain peptidoglycan, phosphorus poly(acid)ester, polysaccharides, protein, and lipopolysaccharides (LPS), such as *Staphylococcus*, *Streptococcus*, and *Bacillus subtilis*. In comparison, the cell walls of Gram-negative bacteria are relatively thin (<10 nm), including protein and LPS, such as *Escherichia coli*, *Pseudomonas aeruginosa*, and *Acinetobacter* (Mai-Prochnow et al. 2016). Viruses are organisms without cell structure, with a small size ranging from tens to hundreds of nanometers (Grgacic and Anderson 2006). Most viruses are composed of genetic material (DNA or RNA) and a protein coat. A few viruses have protein coats encapsulated by a layer of lipids acquired from host cells (Gelderblom 1996). Moreover, viruses only replicate and reproduce within their host cells. According to the host types, viruses are divided into eukaryotic viruses (e.g., human immunodeficiency virus, norovirus, hepatitis B virus) and prokaryotic viruses (including bacteriophages and archaeal viruses) (Gelderblom 1996). Furthermore, the survival of all microbiotas is linked to their habitats.

1.1.2 Habitats of microbiota

Microbiota existed in a wide range of natural environments, including groundwater, lake, soil, etc. For example, it has been estimated that there are about 10^{30} bacterial cells and 10^{31} viral particles (mainly bacteriophages or phages) in aquatic environments (Flemming and Wuertz 2019; Suttle 2007; Wigington et al. 2016; Wommack and Colwell 2000). Meanwhile, the microbiota can be found in various parts of the human body such as skin (Findley et al. 2013), mouth (Anesti et al. 2005), stomach (Bik et al. 2006), lung (Dickson et al. 2014), and intestinal (Duncan et al. 2007; Zoetendal et al. 2001). It is reported that about 90% of cells in the human body are bacteria (Savage 1977), and the number of viral particles (most of which are composed of bacteriophages) is about ten times than that of bacterial cells (Cadwell 2015; Carding et al. 2017; Dalmasso et al. 2014; Lepage et al. 2013; Mills et al. 2013; Wylie et al. 2012). Thus, it is concluded that microbiota widely appeared in different habitats with high abundance. Meanwhile, the different components

of microbiota could influence each other to compete for nutrition and space or achieve a dynamic balance to adapt to the environmental characteristics of habitats (Lupp and Finlay 2005).

1.1.3 Lifestyle and interplay of bacteria and viruses

The relationship between bacteria and viruses (dominated by bacteriophages or phages) has been well investigated (Carding et al. 2017; Wylie et al. 2012). Bacteria utilize nutrients provided by environments to metabolize and proliferate, including autotrophic bacteria that use carbon dioxide as the sole carbon source and heterotrophic bacteria that use organic matter as nutrients (Ettwig et al. 2010; Sellner 1997). While phages rely on bacterial hosts for survival and reproduction. Generally, phages could sustain lytic (virulent phages), lysogenic (temperature phages), chronic and pseudolysogenic cycles (Figure 1). The lytic and lysogenic cycle was the main life cycle of phages and was widely discribed. The lytic-lysogenic decision is mediated by the phage-encoded DNA recombinases (such as integrase and excisionase), which occur at specific attachment sites (*attB*) of the bacterial genome (Nash 1981). When phage sustains a lytic infection strategy, the phage genome is injected into the bacterial host, and then complete replication, packaging, and phage particle assembly, resulting in the lysis of the bacterial host cell and the release of the progeny phage. When the phage sustains a lysogenic infection strategy, the phage genome is inserted into the bacterial chromosome (prophage) and replicated with the bacterial host chromosome (Erez et al. 2017; Grgacic and Anderson 2006). However, stimuli (such as reactive oxygen species) may induce the prophage to be excised from the bacterial chromosome, followed by the expression of the lytic gene, thereby prompting the phage to enter the lytic cycle (Rigottier-Gois 2013). During the chronic cycle, the phage replicates in bacterial host and is released without killing its host. In pseudolysogeny, the phage genome neither integrates nor propagates. It has been revealed that bacteria and phages maintain an interdependent relationship. Normally, bacteria provide places and resources for the replication of phages. Phages affect the bacterial community by killing bacterial hosts and mainly follow the Lotka-Volterra (or 'kill the winner') dynamic predation relationship in environments. The main features include top-down control of the bacterial communities, rapid bacterial and phage population shifts, and the Red Queen coevolution relationship (Reyes et al. 2010). When phages and bacteria keep temperate coexistence relationships, phages as prophages benefit bacterial fitness under different environmental conditions. For example, an experiment to evaluate the effect of prophage on the bacterial

physiology by deleting the prophage of *Escherichia coli* (*E.coli*) showed that prophage promoted bacterial cell growth, increased the resistance to antibiotics, and aided bacterial cells to resist environmental stresses (such as osmotic, oxidative and acid stresses) (Wang et al. 2010). In turn, bacterial fitness could increase the survival and propagation chances of phages. Meanwhile, during infecting bacterial hosts, phages could acquire some AMGs, expanding the bacterial metabolism. Besides, phage also acts as a vehicle of horizontal gene transfer (HGT) to facilitate the exchange of genetic materials between bacteria populations and then contribute to bacterial genetic diversity, further promoting bacterial evolution and adaptivity (Touchon et al. 2017). Overall, bacteria and phages are mutually reliant, playing critical roles in the stability and function of their habitats.
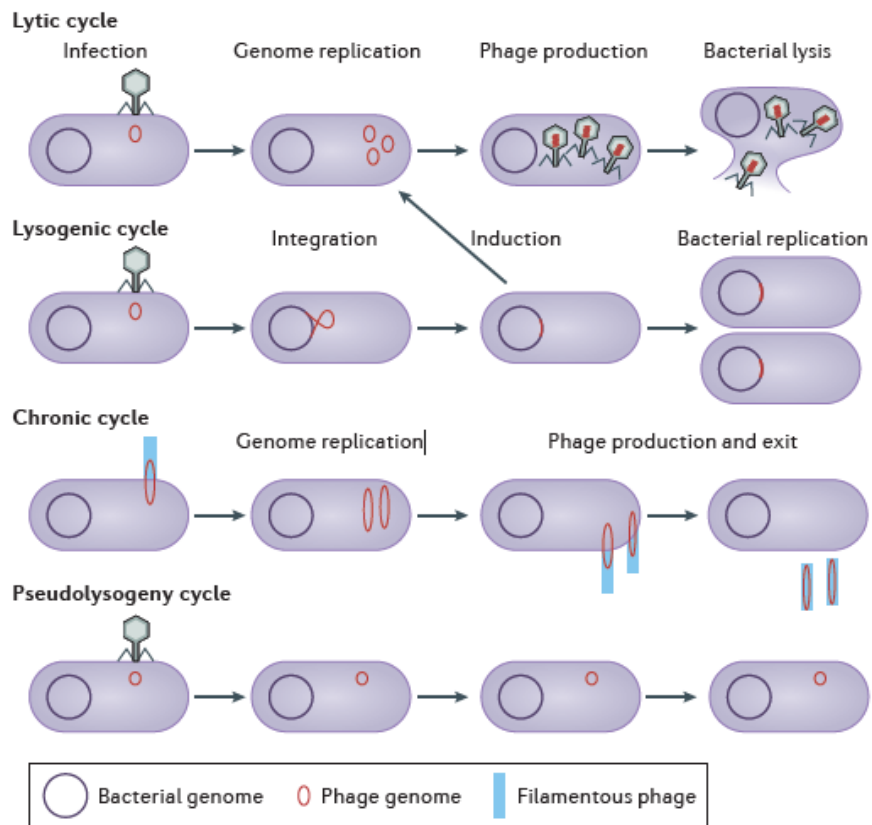

Figure 1 The life cycle of bacteriophages (Mirzaei et al. 2017)

1.2 Function of microbiota in habitats

Generally, microbiota contributes to maintaining the ecological processes and stability of the ecosystem. In natural environments, microbiota promotes the production and decomposition of nutrients, the circulation of elements, and the flow of energy, which ultimately influence the

biogeochemical cycle of the natural environments (Figure 2) (Newton et al. 2011). In human-associated environments, the microbiota is linked to human health and disease (Young 2017). The microbiota performs functions related to nutrient processing, metabolic capabilities, and immune regulation to maintain human health (Hooper and Gordon 2001). However, diseases such as inflammatory bowel disease (IBD), colorectal cancer (CRC) were also reported to be associated with the imbalance of microbiota (Figure 3) (Manor et al. 2020; Münch et al. 2019; Ren et al. 2019; Snider et al. 2019).

1.2.1 The crucial roles of bacteria and viruses in natural environments

In natural environments, bacteria could directly utilize substrates in the habitats for their metabolism, further contributing to the circulation of carbon, nitrogen, sulfur, and other elements. For example, cyanobacteria and methanotrophs could use inorganic carbon to grow and reproduce, thereby promoting the carbon cycle of lakes (Bastviken et al. 2003; Bullerjahn and Post 2014). Some bacteria can help degrade organic pollutants in aquatic and soil environments. The *Pseudomonas putida* KT2440 can help degrade benzene, toluene, ethylbenzene, and xylenes to purify the aquatic environments and maintain the function of ecosystems (Hernández-Sánchez et al. 2016). Thus, bacterial communities are critical for maintaining the stability of natural environments. While the effects of bacteriophages on biogeochemical fluxes of matter and energy are mainly achieved by influencing the dynamics and physiology of bacteria (Rohwer and Thurber 2009; Suttle 2005; Thingstad 2000). Specifically, phages reconstruct the bacterial community to influence the function of natural environments. The most well-known model is the 'kill the winner' predation relationship (Fuhrma and Suttle 1993), which aims to control the abundance of the most abundant bacterial populations, thereby redistributing nutrients to most bacteria to ensure the regular cycle of nutrients and energy. Meanwhile, during the viral lysis of bacterial cells, the cell content (such as carbon, nitrogen, phosphorus) is released and participates in the elemental cycle (Weinbauer 2004; Wilhelm and Suttle 1999). It is estimated that 6% to 26% of the organic carbon produced by bacterial photosynthesis will eventually enter the environment due to the viral lysis of cells (Weinbauer 2004). Besides, phages can be used as a reservoir to store and transfer genetic information of their bacterial host, affecting the fitness and metabolism of bacterial host and then influencing the ecosystem function (Sullivan et al. 2006). For example, the photosynthesis-related genes including *psbA*, *hli*, *talc*, and *nrd* were found in the Cyanophage P-SSP7 of marine (Lindell

et al. 2007), which further expanded the photosynthesis process and influenced the biogeochemical cycle of the aquatic environment. Collectively, bacteria and viruses jointly promote element circulation and energy flow of natural environments.
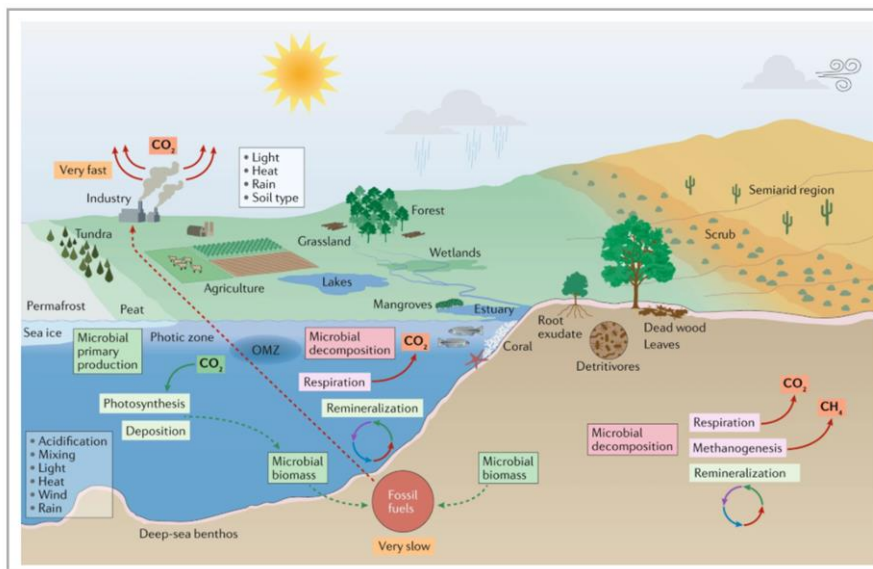
Figure 2   The impact of microbiota in carbon cycle (Cavicchioli et al. 2019)

1.2.2 Contribution of bacterial and viral communities in human health and diseases

In human-associated environments, bacteria utilize the nutrients in the human body for metabolism, which is usually beneficial for human physiology. For example, intestinal bacteria contribute to the absorption and digestion of nutrients (Imhann et al. 2016; Zhernakova et al. 2016) and produce vitamins and oligo-elements, which are helpful to maintain nutrition and energy balance of the human body (Bäckhed et al. 2004). Short-chain fatty acids, the metabolites of bacteria, can be used as energy sources for intestinal epithelial cells to regulate the production of cytokine production and induce the expansion of regulatory T cells, enhancing the tolerance and immune cell homeostasis of the human body (Lee and Hase 2014; Schirmer et al. 2016). However, some bacteria such as *Helicobacter pylori* were reported to be linked to gastric carcinoma, thus increasing the risk of disease occurrence (Correa 1995; Kusters et al. 2006). Phages have been shown to impact human health by affecting bacterial community and physiology (Carding et al. 2017; Wylie et al. 2012). It has been reported that phages could kill bacterial pathogens such as *Staphylococcus aureus* and *Pseudomonas aeruginosa* to treat cystic fibrosis patients and improve

their health (Kvachadze et al. 2011). While some phages could produce enzymes to modify the O-antigen component of LPS, thereby changing the antigenicity of their bacterial hosts and potentially influencing human health (Ogilvie et al. 2012; Rakhuba et al. 2010).

Furthermore, prophage inserted into the bacterial genomes can supply specific genes for bacteria to increase bacterial fitness (Mirzaei et al. 2020). For example, the virulence gene encoding enterotoxin A was revealed to have been transferred to *Staphylococcus aur*eus by phage-mediated HGT, further affecting human health (COLEMAN et al. 1989; Saunders et al. 2001). Additionally, phages could directly enter the epithelial cell layer via transcytosis and activate the underlying immune cells, potentially causing diseases (Breitbart et al. 2003; Sinha and Maurice 2019). For example, the interplay of *E.coli* phages and the immune system has been linked to Type I Diabetes autoimmunity (Tetz et al. 2019). Apart from bacteriophages, virus-related pathogens can also affect human health by infecting human cells. For example, norovirus is the main cause of gastroenteritis (Berger and Mainou 2018). Taken together, bacteria and viruses are linked to human health and disease.
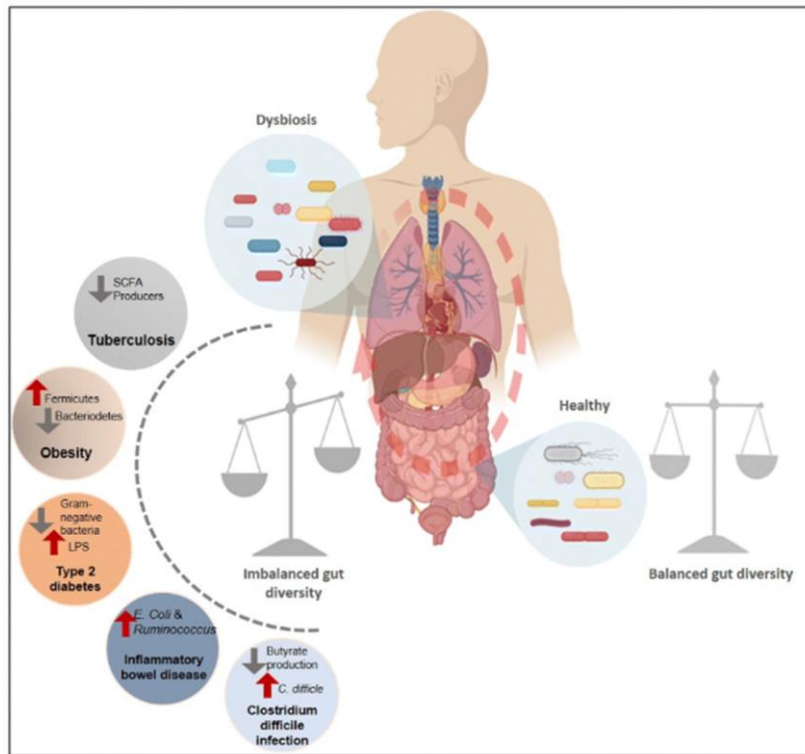
Figure 3   The impact of microbiota on human health and disease (Singhvi et al. 2020)

1.2.3 Consequences of microbial transfer across habitats

The natural environments and human-associated environments are connected since humans are exposed to natural environments. Microbiota could be transferred between these two different environmental types through contact, food, inhalation, and other means. It has been reported that pathogenic bacteria (e.g., *Salmonella, Cryptosporidium*) (Craun et al. 2006) and viruses (noroviruses, adenoviruses) (Abbaszadegan et al. 2003; Borchardt et al. 2003) could enter the human body with drinking water, which could cause a series of diseases, such as gastroenteritis paralysis, fevers, and respiratory illness (Kitajima and Gerba 2015; Lewis et al. 2009), potentially threatening human health (Bosch et al. 2008; Rigotto et al. 2011). In turn, humans could also transfer microbiota to natural environments through feces, breathing, and other means (Szymczak et al. 2020). Meanwhile, the microbiota can also be transferred between different natural environments and different human individuals, such as coronavirus can spread in various human individuals and further cause the pandemic of COVID-19 (Yuki et al. 2020). Besides, the transfer of microbiota is also a treatment for diseases such as, Fecal microbial transplantation (FMT), that is, transplanting the feces of healthy individuals into the guts of patients for the treatment of gastrointestinal tract (GI) disease (Brandt 2013). FMTs have been successfully applied to treat GI tract disease caused by *Clostridium difficile* infection (Fuentes et al. 2014; Moelling and Broecker 2016). Therefore, the transfer of microbiota may be beneficial or harmful to target habitats.

In addition to the effect of microbiota on habitats, environmental factors and their changes in habitats also affect the community and function of microbiota.

1.3 Influence of habitats on microbial community structure and function

In general, microbiota and their habitats maintain a coexisting relationship. The habitat supplies nutrients and space for the growth and reproduction of microbiota. In natural environments, the habitat provides microbiota with the necessary light, nutrients, space, etc., for its growth. In human-associated environments, the human body offers nutrient-rich and diverse habitats. However, some disturbances or stimuli suffered by these habitats may cause the microbial community change, affecting the stability of habitats.

1.3.1 The impact of natural environments on bacteria and viruses

Numerous studies have demonstrated the link between bacteria and their habitats and how the environmental conditions of habitats impact bacterial community (Costello et al. 2009; Lawes et al. 2016). In the natural environments, the habitat supplies necessary substances and relevant environmental factors for bacterial growth and metabolism (Camacho et al. 2001; Li et al. 2020; Van Berkum and Eardly 2002). However, when environmental conditions change, such as pH (Kawagoshi et al. 2005; Xiong et al. 2012), temperature (Kawagoshi et al. 2005; Liu et al. 2013), nutrient availability (Newton and McMahon 2011; Van der Gucht et al. 2005), the bacterial community composition altered. In an experiment simulating organic matter change in an alpine lake, the bacterial community changed rapidly (Rofner et al. 2017). However, the change of environmental conditions may not influence the dominant bacteria at the phylum level. Research that collected a large number of bacterial data in freshwater lakes discovered that the major bacterial phyla are common, including Actinobacteria, Bacteroidetes, Cyanobacteria, Proteobacteria, and Verrucomicrobia (Díaz-Muñoz and Koskella 2014). When further explored the recovery of microbial communities after disturbance, it was found that in nutrient-rich lowland lakes, the bacterial communities have returned to their original states after experiencing human intervention, typhoons, and other disturbances (Jones et al. 2008; Shade et al. 2011). The relationship between viruses (mainly bacteriophages) and habitats is related to the bacterial community. The habitats supply nutrients for the growth and reproduction of bacterial hosts of phages, further leading to the proliferation of phages. When environmental conditions change, the phage community and their life cycle also vary accordingly. In natural environments, the propagation of bacteria caused by high temperatures and high concentrations of dissolved organic carbon could lead to viral proliferation (Filippini et al. 2008). In addition, environmental factors such as pH, UV light, trophic status could induce the phage replication cycle to switch between lytic and lysogenic cycles. Low nutrient availability facilitates phage tend to lysogenic cycle (Weinbauer 2004).

1.3.2 The impact of human-associated environments on bacteria and viruses

In the human-associated environments, the human body provides bacteria with various habitats (mainly GI tract) and abundant substrates for their growth and reproduction, such as carbohydrates, protein, lipid, vitamins (Morowitz et al. 2011). For different individuals, bacterial communities have a large variation caused by many factors such as the human environment, human lifestyle,

human physiology, and human genotype (Turnbaugh et al. 2007). For the same individuals, external stimuli (such as radiation) (Li et al. 2009) and changes in the internal environments (such as pH, oxygen concentration, water availability) (Duncan et al. 2009; Espey 2013; Ridlon et al. 2014; Sutton and Hill 2019) may cause the dysbiosis of the bacterial community, leading to host dysfunction, such as immune disorders and abnormal inflammation, and ultimately develop into various diseases including cancer (Banerjee and Robertson 2019; Rajagopala et al. 2017). A study investigating the effect of pH on the development of human colonic bacteria revealed that pH mainly affected the competition between phylogenetically and functionally different groups of bacteria in the microbial community, potentially threatening human health (Duncan et al. 2009). In the human body, the bacteriophage community altered under different physical conditions (Lim et al. 2015). Disease-specific bacteriophages were revealed in IBD (Norman et al. 2015). Moreover, the virulence bacteriophages are dominant in the healthy human body (Gogokhia et al. 2019). In contrast, in the disease condition (such as Crohn's disease), the virulence bacteriophages were replaced by temperate phages (Clooney et al. 2019).

Overall, in natural environments and human-associated environments, microbiota contributes to the stability and function of habitats, while environmental factors of habitats influence microbial community structure and function. However, under the extreme conditions of these environments (that is, extreme environments), the understanding of the interaction between microbiota and habitats is limited in former studies due to technical problems.

1.4 The interaction of microbial community and extreme environments

Extreme environments are habitats that make it difficult for microbiota to perform fundamental metabolic functions. Extreme natural environments are mainly related to high temperature (McMahon et al. 2012), low temperature (Fedorov et al. 2019), alkalinity (Chao et al. 2017), acidity (Korzhenkov et al. 2019), hypersaline (Pal et al. 2020), high pressure (Miettinen et al. 2015), intense radiation (Chanal et al. 2006), low oxygen concentration (Coskun et al. 2019), poor nutrient level (Mehrshad et al. 2020), presence of toxic substances (Coryell et al. 2018), etc. After the long-term evolution and natural selection, it has been found that specific microbiota has adapted to these extreme conditions, such as alkaliphilic bacteria (Sarethy et al. 2011), anaerobic bacteria (Beatty et al. 2005), oligotrophic bacteria (Ohhata et al. 2007), etc. Alpine lakes and deep groundwater belong to typical oligotrophic environments. It has been reported that changes in temperature and

nutrient availability caused by climate change are the main factors affecting the microbial community in alpine lakes (Salmaso et al. 2012). In an alpine lake (Kalakuli Lake), the dominant bacteria shifted from Proteobacteria to Actinobacteria with the increasing carbon-to-nitrogen ratio (Liu et al. 2017). In deep groundwater, the availability of nutrients is the key factor affecting the bacterial community structure (Hubalek et al. 2016; Yan et al. 2020). Meanwhile, bacteria have developed various physiological features to utilize nutrients effectively (Ho et al. 2017). For example, different carbon source levels and types in deep groundwater led to bacteria developing various carbon assimilation mechanisms (Major et al. 2010; Nyyssönen et al. 2014; Sokatch 2014). Although studies have been conducted on the microbial communities in these two oligotrophic areas, our understanding of them, such as the virome community, is not comprehensive. This is mainly due to two limitations (1) Sampling is difficult because of the remote locations and sampling technologies; (2) Detecting, quantifying, and characterizing virome communities is challenging. Metagenomics based on nucleic acid extraction is the state-of-art method to study the virome community due to the lack of shared genetic elements of viruses (Rohwer and Edwards 2002). However, it is hard to obtain sufficient amounts of microbial materials to extract nucleic acid directly from the water samples of alpine lakes and deep groundwater with low microbial abundance. Therefore, the water samples need to be concentrated before nucleic acid extraction. Currently, $FeCl_3$ flocculation (Hurwitz et al. 2013) and Monolithic adsorption filtration (Hjelmsø et al. 2017) are two efficient methods to concentrate viruses from water samples. In this thesis, the $FeCl_3$ flocculation method was used to concentrate water samples for virome DNA extraction.

Human-associated extreme environments are related to the abnormal state of the internal environment of the human body, such as hypothermia, hypoxia, hyperglycemia, anemia, immunosuppressive conditions, agranulocytosis, and lymphocytopenia (Koksal et al. 2009; Mock and Olsen 2003; Watanabe et al. 2010). These abnormal states are accompanied by various diseases such as IBD, CRC, esophageal-related diseases and are extreme environments for microbiota growth. Numerous studies have revealed bacterial communities under different diseases conditions (Feng and Weinberg 2006; Koliarakis et al. 2019). For example, unique bacterial communities were found in the biopsy samples of Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC) patients. Compared with the normal esophagus, the Gram-positive bacteria (Firmicutes) were gradually replaced by Gram-negative bacteria (Bacteroidetes,

Proteobacteria, Fusobacteria, and Spirochaetes) in BE. As the disease developed into EAC, Gram-negative bacteria *E.coli* and *Fusobacterium nucleatum* became more dominant (Yang et al. 2012).

Meanwhile, virome communities were also widely explored in various diseases such as GI tract disease (e.g., IBD, CRC), Parkinson's disease, and Type I diabetes (Nakatsu et al. 2018; Norman et al. 2015; Tetz et al. 2018; Tetz et al. 2019). However, previous studies on the role of viruses in GI tract disease have mainly focused on lower GI tract disease (e.g., IBD, CRC). Less is known about the relationship between gut viruses and upper GI tract disease. This is mainly because stool samples could directly reflect the microbial communities related to the lower GI tract disease, while biopsy samples are the best choice to reveal the microbial community at the lesion related to upper GI tract disease (Pei et al. 2004; Pei and Yang 2005). However, the biopsy sampling method is invasive and time-consuming, may potentially induce potential complications in the human body (Fillon et al. 2012). Moreover, biopsy samples could not provide enough microbial materials for sequencing (Lim et al. 2009), while the traditional amplification step may lead to deviations from the actual results. Therefore, in this thesis, the stool samples were used to study the gut virome community related to upper GI tract disease.

1.5 Aims of the thesis

Microbiota (mainly bacteria and viruses) plays critical roles in natural environments, as well as human health and disease. However, our understanding of microbial communities in extreme environments related to them is still in its infancy. Therefore, this thesis selected three different types of extreme environments to study microbial communities, including (1) Lake Unterer Giglachsee, which belongs to alpine lakes that are located at high altitudes, far away from human activities and other industrial pollution, thus its ecosystem remains relatively pristine and unproductive (Weckström et al. 2016). Moreover, due to simple and oligotrophic conditions, the microbiota living there are sensitive to climate changes (e.g., increasing temperature). Therefore, Lake Unterer Giglachsee is also considered essential sentinels of climate change (Thompson et al. 2005). (2) Deep groundwater of Äspö Hard Rock Laboratory (Äspö HRL), which is in a deep biosphere and lacks photosynthesis and nutrient input, also belongs to an oligotrophic environment (Danielopol et al. 2000; Griebler and Lueders 2009). In addition, the ecosystem is relatively stable due to fewer human activities and weather variations. Thus, the ecological strategies of bacteria surviving here are especially significant due to the geochemically-stable and low-energy

conditions (Mehrshad et al. 2020). (3) Intestinal environments of patients with esophageal-related diseases. BE is characterized by the metaplastic replacement of the normal squamous epithelium with columnar epithelium. It is strongly linked to the development of EAC, a condition with a high mortality rate in which cancerous cells develop in esophageal tissues (Wiethaler et al. 2019). BE and EAC indicate the deterioration of the internal environment in the human body, which is challenging for the growth of microbiota (Koksal et al. 2009; Mock and Olsen 2003; Watanabe et al. 2010). In this thesis, the structure and behavior of microbial communities in these three types of extreme environments were revealed to supplement the gap of former studies. The main aspects of the thesis are as follows:

- Reveal the impact of extreme environments on microbial communities.
- Illustrate the adaptation of microbial communities to extreme environments.
- Investigate the resilience of microbial communities in extreme environments.
- Clarify the dynamic relationship between viruses (bacteriophages) and bacteria in extreme environments and the potential impact on habitats.
- Provide theoretical support for a more profound understanding of microbial function in both natural environments and human-associated environments in the future.

Chapter 2

2 Materials and Methods

2.1 Research objects description and sampling

2.1.1 Oligotrophic aquatic environments

Lake Unterer Giglachsee (47.28°N, 13.65°E) is in a catchment area of the Niedere Tauern Alps in Austria, dominated by carbonaceous bedrock (Luoto and Nevalainen 2013). This area is remotely located, and its high altitude (1,922 m) and catchment characteristics make it far from human disturbance, thus forming an oligotrophic and transparent lake. The detailed information can be found in a previous study (Weckström et al. 2016). The Zentralanstalt für Meteorologie und Geodynamik contains three nearby weather monitoring stations, namely Obertauern Station (1,772m), Schmittenhöhe station (1,956m), and Rudolfshütte Station (2,317m), which are used to monitor meteorological parameters including air temperature, average wind speed, precipitation, sunshine hours, and snow layer. Besides, the water temperature was monitored every 4 hours at a depth of 2.5m of the deepest parts of the lake by Water thermistors (MINILOG, Vemco Ltd). During the summer growing season of 2010 (including July 19, 21, 27, and 31 and August 01, 04, 09, and 18), the water samples (1.5-2 L depth-integrated) were collected from the deepest part of the lake.

Äspö HRL is located on the east coast of the Baltic Sea in Sweden (Stanfors et al. 1999), which was excavated at a depth of 450m below Äspö island and comprised a 3.6 km long tunnel spiraling down to a granitic bedrock (Kyle et al. 2008). The detailed information is given by Stanfors et al. 1999. The boreholes in the tunnel provide the possibility to collect deep groundwater. Due to the lack of nutrient input, deep groundwater is a relatively oligotrophic environment (Danielopol et al. 2000; Griebler and Lueders 2009). In August 2013, three borehole sites were conducted sampling:

507B (namely Shallow Aquifer, short SA, 71m), which aged from months to years (Banwart et al. 1996), 1327B (namely Intermediate Aquifer, short IA,196m), which is more than 20 years old (Wu et al. 2016), TASF (namely Deep Aquifer, short DA, 450m), which is approximately 7,000 years old (Laaksoharju et al. 1999; Wu et al. 2016). Additionally, Baltic Sea water (BS) was also collected because of the significant influence on the groundwater aquifer (Hengsuwan et al. 2015). The groundwater (2 L) samples were collected at each sampling site.

2.1.2 Intestinal environments of patients with esophageal-related diseases

Human stool samples were obtained from patients with esophageal-related diseases and healthy individuals, supplied by Prof. Dr. Michael Quante from II. Medizinische Klinik, Klinikum Rechts der Isar, Technische Universität München, Germany. In detail, the stool samples, which were collected using Stool Collection Tubes with Stool DNA Stabilizer (STRATEC Molecular GmbH, Berlin, Germany), were collected from six BE patients, four EAC patients, and six CT from the German BarrettNET registry (Wiethaler et al. 2019). The sampling procedure was carried out mainly at home or in the clinic. After collecting, samples were immediately shipped to the clinic human sample biobank and stored at -80°C.

2.2 The pre-treatment of samples

2.2.1 Pre-treatment of samples for bacteriome DNA extraction

The collected water samples from Lake Unterer Giglachsee were pre-filtered to remove most eukaryotic algae and particle-associated bacteria through glass fiber filters. The filtrate was then filtered through nitrocellulose (NC) membranes (pore diameter 0.2 µm), remaining the bacteria on the membranes. Subsequently, membranes were immediately transferred into Eppendorf tubes and were stored at -20°C until further bacterial DNA extraction.

2.2.2 Pre-treatment of samples for virome DNA extraction

The water samples from deep groundwater of Äspö HRL were first pre-filtered through 0.22 µm filters (PES Membrane, Merck Millipore, Lot No. ROCB29300, Ireland) to remove most eukaryotic algae and particle-associated bacteria. Then the viruses in the groundwater were concentrated by $FeCl_3$ flocculation described in the previous description with minor modifications (John et al. 2011). Briefly, after the addition of 20 mg Fe L^{-1} to the groundwater, the mixture was

incubated for 1 hour. Subsequently, the mixture was filtered by 0.22 µm filters (PES Membrane, Merck Millipore, Lot No. ROCB29300, Ireland), which was pressured by a vacuum pump, remaining the Fe-virus concentrates in the filter surface. Then Fe-virus concentrates were resuspended in the Ascorbate-EDTA buffer until dissolved completely (the recipe referred to John et al. 2011). Finally, the solution was concentrated to less than 50 µL by Amicon® Ultra Centrifugal Filters (10kDA, Merck Millipore, Lot No. R9EA18187, Ireland) for further virome DNA extraction.

The stool samples from esophageal disease patients were vortexed vigorously for 4 hours at 4℃, then centrifuged for 30 mins at 4,000 g to obtain supernatant. The supernatant was then passed through 0.22 µm filters (PES Membrane, Merck Millipore, Lot No. ROCB29300, Ireland) to remove bacterial-associated particles, and the volume was concentrated to less than 50 µL using Amicon® Ultra Centrifugal Filters (10kDA, Merck Millipore, Lot No. R9EA18187, Ireland) for further virome DNA extraction.

2.3 DNA extraction and sequencing

2.3.1 Bacteriome DNA extraction and sequencing

NC membranes that retain bacteria were loaded into a bead beater tube to break the bacterial cells physically. The bacterial DNA was extracted following the instructions of the NucleoSpin® Soil DNA extraction kit (Macherey-Nagel, Germany) and then eluted in 40 µl elution buffer. The quality and quantity of extracted bacterial DNA were analyzed by NanoDrop®ND-1000 spectrophotometer (Thermo Fisher Scientific). Subsequently, bacterial genomic DNA was amplified using the primer pairs 338F/1046R (338F: 5'-CGTATCGCCTCCCTCGCGCCA TCAG ACGAGTGCGT ACTCCTACGGGAGGCAGCAG-3') and the reverse primer (1046R: 5'-CTATGCGCCTTGCCAGCCCGC TCAG ACGAGTGCGT CGACAGCCATGCANCACCT-3') for the V3-V6 hypervariable regions of 16sDNAs. The adaptor, key, and barcode sequences (underlined and separated by a space) were included in the primer pairs, and the specific sequence binding to bacterial rDNA is indicated at the 3' end (Huse et al. 2008). Subsequently, the 50 µl PCR reaction mix was prepared, including 10 µl of 5×buffer HF, 1µl of 10 mM of each deoxynucleotide triphosphate, 1 unit of a proofreading polymerase (Phusion Hot Start High-Fidelity DNA Polymerase; Thermo Fisher Scientific), and 2.5 µl of forward and reverse primers

(10 pmol µl⁻¹). The PCR program was as follows: annealing temperature at 67.8°C, elongation for 30 s, and 20 cycles. PCR products were tested in agarose gel to test the quality and specificity. Afterward, PCR products were cut out and purified by the gel purification kit (Qiagen, Hilden, Germany). Four independent PCR amplicons were pooled and sequenced on the GS-FLX platform with 454 titanium chemistry. Four amplicons were then sequenced in four regions separated by four-region gaskets, loading 3,100,000 amplicon-coated beads each run and yielding 800,000 sequence tags.

2.3.2 Virome DNA extraction and sequencing

A sample of less than 50 µL was mixed with 1/5 volume of chloroform and centrifuged for 3 mins at 14,000 g, retaining the upper phase to remove proteins. Following that, DNaseI (1U/ µL, Invitrogen, USA, Lot No. 1158858) was added and incubated at 37°C for 1 h to remove bacterial DNA fragments. Lysis buffer was then added (700 µL KOH stock (0.43g/10ml), 430 µL DDT stock (0.8g/10ml), 370 µL H_2O, pH=12) to incubated at room temperature for 10 mins, followed by frozen at -80°C for 2 hours. Afterward, the samples were incubated at 55°C for 5 mins following by adding 1 µL Proteinase K (20mg/ml, Invitrogen, USA, Lot No. 1112907) with incubating for 30 mins at 55°C. Subsequently, AMPure beads (Agencourt, Beckman Coulter, USA) were used to clean the samples. In detail, the AMPure beads were added to samples and co-incubated for 15 mins to adsorption DNA. Then DNA was eluted from beads with 35 µL Tris buffer (10mM, pH=9.8) and stored at -80°C for further sequencing. Viral sequencing was performed on an Illumina Novoseq 6000 instrument using chemistry for 2×150 bp reads.

2.4 Bioinformatic analysis

2.4.1 Bacteriome sequence processing

The 16S rDNA sequences were processed by QIIME 19.0 (Caporaso et al. 2010b). Based on the different barcodes of sequences, multiplex reads were assigned to their original samples. The low-quality reads (Phred quality score > 2) were filtered from raw reads, and the remaining reads were clustered into various OTUs based on 97% sequence similarity. Afterward, the phylotypes of bacterial OTUs were identified by UCLUST (Edgar 2010) consensus taxonomy classifier, and the corresponding representative sequences were aligned using PYNAST (Caporaso et al. 2010a). In

each OTU cluster, the most abundant sequence was chosen as the representative sequence and was annotated the taxonomy by the Greengenes database (DeSantis et al. 2006).

2.4.2 Virome sequence processing

Low-quality bases and adaptors of raw sequences were removed using fastp (v0.20.1) (Chen et al. 2018). The remaining reads were deduplicated by dedupe.sh from bbmap suite (v38.76) (Bushnell 2019), and clean reads were assembled into contigs by metaSPAdes (v3.14.0) with default settings (Nurk et al. 2017). Then contigs longer than 1kbp were pooled from all samples, and redundancy contigs were filtered using dedupe.sh. Afterward, viral contigs were predicted with the combination of VirSorter (v1.0.6) (Roux et al. 2015), CAT (v5.0.4) (von Meijenfeldt et al. 2019), and DeepVirFinder (v1.0) (Ren et al. 2020). If the predicted viral contigs shared more than 95% identity over 80% of the contig length, the viral contigs were clustered by CD-HIT (Fu et al. 2012), and the longest viral contigs in each cluster were used as representative viral contigs for further analysis. ORFs of representative viral contigs were predicted by Prodigal (v2.6.3) (Hyatt et al. 2010) and provided to vConTACT2 (Jang et al. 2019), CAT, and Demovir script (https://github.com/feargalr/Demovir) with default parameters and database for taxonomy annotation. Clean reads were mapped to viral contigs using bbmap.sh from bbmap suite (v38.76) to calculate the relative abundance of viruses in each sample, and the contig coverage was assessed by CoverM (v0.4.0). Feature Counts (v2.0.0) (Liao et al. 2014) were used to estimate the number of reads that mapped to each gene. Predicted viral proteins were fed into VIBRANT (Kieft et al. 2020) to identify the lytic and lysogenic cycle of phages and to do functional annotation by searching viral proteins against KEGG (Kanehisa et al. 2017), VOGDB, and PFAM databases. Meanwhile, the function of AMGs was also identified. The BPROM software was used to predicted promoters in the genome of viral contigs (Salamov and Solovyevand 2011; Umarov and Solovyev 2017).

2.5 Data analysis

2.5.1 Statistical data analysis

Non-metric multi-dimensional scaling (NMDS) based on ''Bray–Curtis'' similarities was conducted in R (version 3.2, package vegan), and Analysis of similarities (ANOSIM) was further used to test the significant difference. Venn diagram analysis was conducted at

http://www.interactivenn.net/. Other graphs were created by Origin Lab 8. Redundancy analysis (RDA) with the time dependence of consecutive samples as a covariate was performed on CANOCO software for Windows (version 4.5). (Manuscript 1)

Viral richness (Ace) and diversity (Shannon) were calculated using Qiime2 (https://qiime2.org). Principal Coordinates Analysis (PCoA) that based on "Bray-Curtis" similarities was conducted on R (Version 3.2, package vegan). The significant difference was determined by Permutational Multivariate Analysis of Variance (PERMANOVA). Statistical analyses were performed in Prism 9- GraphPad for two-way analysis of variance [ANOVA] with Tukey's post hoc test, and R (Version 4.0.2, stats package) for the Kruskal-Wallis with Dunn's post hoc test. The Jonckheere trend test was performed in IBM SPSS Statistics (Version 27.0). Meanwhile, multiple testing corrections were used to modify the P value using the "Bonferroni Holm" method. Graphs were generated in Prism 9- GraphPad, Origin 2020b, Microsoft Excel, and R (Version 3.3.3, ggplot2 package). (Manuscript 2)

Rarefaction curves for the viral richness (Ace) and diversity (Shannon) were constructed with 20,000 sequences as a sampling depth. Venn diagram and Heatmap were generated using R (Version 3.3.3) with packages VennDiagram (Version 1.6.20) and pheatmap (Version 1.0.12), respectively. All Histograms and line charts were generated in Origin 2020b. Pie charts were generated in Microsoft Excel. (Manuscript 3)

2.5.2 Phage-host analysis

The interactions between viruses (phages) and bacteria (host) were predicted by VirHostMatcher-Net (Wang et al. 2020). In the current thesis, viral contigs with a length of more than 10kb were used to predict bacterial hosts. Only those with a score of more than 95% had higher accuracy and were retained for further investigation. (Manuscript 2, Manuscript 3)

2.5.3 Viral sources prediction

Viral sources were predicted by blasting with the database of Integrated MICROBIAL GENOME/VIRUS (IMG.VR), which are mainly divided into three sources: environmental sources (terrestrial, sediment, marine, and freshwater), engineered sources (wastewater, solid waste, lab

enrichment, built environment, and bioreactor), and host-associated sources (human, plants, and other). (Manuscript 3)

2.5.4 The definition of core viruses, abundant viruses, moderate viruses, and rare viruses

Core viruses in similar habitats were identified as overlapping areas of circles in the Venn diagram (Shade and Handelsman 2012) (Manuscript 3). Abundant viruses were defined by the contigs with the relative abundance of more than 1% in total contigs. Moderate viruses were defined by the contigs with the relative abundance of more than 0.1% and less than 1% in total contigs. Rare viruses were defined by the contigs with the relative abundance of less than 0.1% in total contigs. (Manuscript 2, Manuscript 3)

2.5.5 Phylogenetic tree

The phylogenetic tree was built by MEGA X (Kumar et al. 2018), based on the Maximum Likelihood method (Jones et al. 1992). The evolutionary history of the amino acid sequence of the examined gene was shown by the bootstrap consensus tree derived from 1,000 replicates (Felsenstein 1985) is taken to represent the evolutionary history of the taxa analyzed (Zuckerkandl and Pauling 1965). The phylogenetic trees were displayed using Interactive Tree of Life (https://itol.embl.de/). (Manuscript 3)

Chapter 3

3 Manuscript overview

This thesis contains three manuscripts, here listed a summary, the publication status, and the contribution of authors is given.

I.   Manuscript 1

Ma, T.; Jiang, Y.; Elbehery, A.H.; Blank, S.; Kurmayer, R.; Deng, L. Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake. *Hydrobiologia* **2020**, 847, 535-548.

II.   Manuscript 2

Ma, T.; Ru, J.; Xue, J.; Schulz, S.; Mirzaei, M.K.; Janssen, K.-P.; Quante, M.; Deng, L. Differences in Gut Virome Related to Barrett Esophagus and Esophageal Adenocarcinoma. *Microorganisms* **2021**, 9, 1701.

III.   Manuscript 3

Ma, T.; Xue, J.; Ru, J.; Deng, L. Depth-specific virome reveals the survival strategy of aquifer microbiota adapting to the oligotrophic environments. Submitted

Manuscript 1: Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake

Authors: Tianli Ma, Yiming Jiang, Ali H.A. Elbehery, Stephan Blank, Rainer Kurmayer&Li Deng

This manuscript explored the change of planktonic bacterial community structure during the short-term weather deterioration in an alpine lake. The influence of a short-term cooling period during growing season on the planktonic bacterial community structure of an alpine lake was studied using 16S rDNA pyrosequencing. It was found that Proteobacteria, Actinobacteria, and Bacteroidetes were the most abundant phyla. A rapid cooling period with high precipitation occurred throughout the sampling period (from July to August 2010), as indicated by decreased conductivity, calcium, and dissolved organic carbon concentration due to the increased runoff. The relative abundance of Actinobacteria, Betaproteobacteria, and Cyanobacteria decreased during this short-term cooling period. Instead, a rapid shift from Betaproteobacteria to Gammaproteobacteria occurred, mainly caused by an increase of *Acinetobacter rhizosphaerae*. Soon after the short-term cooling period, warmer weather conditions got re-established, and Betaproteobacteria recovered and became again dominant. Non-metric multidimensional scaling analysis and Venn diagrams indicated a planktonic bacterial community composition with high similarity at the beginning and end of the growing season. The relative abundance of observed variation in the operational taxonomic unit (OTU) was significantly correlated with air temperature and precipitation. It is concluded that a distinct planktonic bacterial OTU community developed to respond to the short-term cooling period. It rapidly diminished, however, as summer conditions became re-established, implying the recovery of the original bacterial community structure.

Remark: LD and RK initiated study concept and design. SB conducted the experimental part. TM analyzed the data. TM, YJ, AE, SB, RK and LD constructed the manuscript.

Manuscript 2: Differences in Gut Virome Related to Barrett Esophagus and Esophageal Adenocarcinoma

Authors: Tianli Ma, Jinlong Ru, Jinling Xue, Sarah Schulz, Mohammadali Khan Mirzaei, Klaus-Peter Janssen, Michael Quante and Li Deng

This manuscript investigated the gut virome (dominated by bacteriophages or phages) community changes in Barrett's esophagus (BE) patients and esophageal adenocarcinoma (EAC) patients compared to healthy controls (CT). The relationship between phages and lower gastrointestinal (GI) tract diseases has been studied, however, the relationship between gut phages and upper GI tract diseases, such as esophageal diseases, which mainly include BE and EAC, is poorly described. This study aimed to reveal the gut phage community and their behaviors in the development of esophageal diseases. In total, the gut phage community of sixteen samples was analyzed from patients with esophageal diseases as well as healthy controls. Among three groups, differences were found in the community composition of abundant and rare phages. Besides, the auxiliary metabolic genes (AMGs) related to bacterial exotoxin and virulence factors such as lipopolysaccharides (LPS) biosynthesis proteins were observed to be abundant in the genome of rare phages from BE and EAC samples compared to CT. These results suggested that the gut phage community composition and functional characteristics encoded by them varied in two stages of esophageal diseases. However, the findings of the current study need to be validated with larger sample sizes in the future.

Remark: KPJ, MQ, and LD initiated study concept and design. TM conducted the experimental part and analyzed the data. JR performed the bioinformatic analysis. TM, JR, SS, JX, MKM, KPJ, MQ, and LD constructed the manuscript.

Manuscript 3: Depth-specific virome reveals the survival strategy of aquifer microbiota adapting to the oligotrophic environment

Author: Tianli Ma, Jinling Xue, Jinlong Ru, Li Deng

Submitted

This study revealed the virome community in different depths of deep groundwater. Viruses (dominated by bacteriophages) play critical roles in driving the development of bacterial community and the circulation of elements in biosphere. However, the exploration of viral communities in deep groundwater, an important part of the deep biosphere, is limited. With the help of metagenomic sequencing, the present study revealed the structure and behavior of viral communities in deep groundwater. Environmental variation caused by increasing groundwater depth showed a distinct preference for the component of core viral groups. Whereas viruses in deep groundwater have a consistent tendency to switch to lysogenic life cycle. In addition, abundant viruses that perform main ecological functions dominated in deeper aquifer, while rare viruses that enhance ecological resilience dominated in upper aquifer. Furthermore, viral traits associated with auxiliary metabolic genes (AMGs) appeared depth-specific patterns. AMGs related to methane metabolism (*cofF*), sulfur metabolism (*cysH*), and quorum sensing (*LuxR* and *spo0A*) were found in the genome of viruses, and their distribution at different depths correlated with environmental characteristics and bacterial metabolism there. Considering former studies on microbiome, abundant viruses may contribute to the stability of simple ecosystems by keeping a "Red Queen coevolution" relationship with abundant bacteria, while rare viruses may aid the resilience of complex ecosystems by maintaining a temperate coexistence relationship with rare bacteria. Moreover, viral AMGs further support the adaptation of bacterial hosts to the environment of different depths and jointly prompt the biogeochemical cycle of the ecosystem. Collectively, viruses performed specific survival strategies at different depths of deep groundwater.

Remark: LD initiated study concept and design. TM conducted the experimental part and analyzed the data. JR performed the bioinformatic analysis. TM, JX, JR and LD constructed the manuscript.

Chapter 4

4 Discussion

Extreme natural environments mainly represent habitats containing extreme environmental factors such as oligotrophic, low oxygen concentration, cold temperature (Chen et al. 2017; Ducey et al. 2010; Yin et al. 2013). Human-associated extreme environments represent the abnormal states of the internal environment (e.g., the change of pH or oxygen concentration) in the human body (Kellum 2000; Maltepe and Saugstad 2009), which are frequently accompanied by diseases. Together, extreme environments cannot provide suitable living conditions for most microbiota. However, in the long-term natural selection and microbial evolution, microbial communities have gradually adapted to extreme environments through changing their community structure, physiological ecology, and adaptive mechanisms (Ma et al. 2020; Yadav et al. 2016). For example, in patients with BE and EAC, the bacterial community in the esophagus is dominated by Gram-negative bacteria (such as Proteobacteria, Bacteroidetes), which is different from the bacterial community (mainly Gram-positive bacteria) in the normal esophagus (Yang et al. 2012). Besides, studies investigating the physiological characteristics of psychrophilic bacteria which isolated from cold alpine habitats (Gangwar et al. 2009; Liu et al. 2007) showed that psychrophilic bacteria have highly cold-active enzymes such as lipase (Feller et al. 1990), amylase (Ramesh et al. 2001), and protease (Kuddus and Ramteke 2012). Here, this thesis supplements the existing knowledge by studying the microbial communities (mainly including bacteria and viruses) in three different types of extreme environments, including an alpine lake (Unterer Giglachsee), which belongs to an oligotrophic aquatic environment and is frequently disturbed by extreme weather; deep groundwater (Äspö HRL) which belongs to an oligotrophic aquatic environment and is relatively stable with less disturbance; intestinal environments of patients with esophageal-related diseases (BE and EAC) which is a deteriorated condition of human physical condition. In detail, this thesis

mainly revealed the response of microbial communities to extreme environments from four aspects (1) the structure changes of microbial communities in extreme environments compared to their normal counterparts (Manuscript 2, Manuscript 3); (2) the adaptation strategy of microbial communities to extreme environments (Manuscript 1, Manuscript 2, Manuscript 3); (3) the resilience of microbial communities to the disturbance in extreme environments (Manuscript 1, Manuscript 3); (4) the interplay between bacteria and viruses in extreme environments and potential influence on habitats (Manuscript 2, Manuscript 3).

4.1 The structure changes of microbial communities in extreme environments compared to their normal counterparts

Extreme environments provide distinctive and narrow niche range for microbial communities, thus forming specific microbial communities (González-Toril et al. 2003; Singh et al. 2019). Through investigating the virome community in human-associated environments and natural environments, the richness (Ace) and diversity (Shannon) of the virome in extreme environments (the intestinal environments of patients with BE and EAC, deep groundwater) changed compared with those in control groups (the intestinal environments of healthy individuals, nutrient-rich sea), which mainly affected by ecological factors (such as oligotrophic status) (Yin et al. 2013) and biotic factors (mainly bacterial community) (Norman et al. 2015). Meanwhile, the predicted bacterial hosts have also changed. In deep groundwater, the range of predicted bacterial hosts reduced compared with nutrient-rich sea. In the gut of BE and EAC patients, the relative abundance of predicted bacteria hosts showed obvious variation. For example, compared with CT, the relative abundance of *Bacteroidia*, *Bacilli*, and *Erysipelotrichia* predicted by gut viruses in BE and EAC patients were lower, while the relative abundance of *Clostridia* was higher. The differences in bacterial hosts can also indirectly reflect the change of bacterial communities in these two extreme environments. Collectively, these results indicated that microbial community structures altered in extreme environments compared with their normal counterparts.

4.2 The adaptation strategy of microbial community to extreme environments

Extreme environments threaten the fundamental survival and metabolic function of microbiota. Meanwhile, it has been observed that microbial communities in extreme environments are more

sensitive to different disturbances because of various growth-limiting factors such as nutrients, temperature (Ma et al. 2020). Therefore, along with the long-term evolution and natural selection, microbiota developed a series of processes to adapt the extreme environments and disturbances (Bang et al. 2018), for example (1) Changes in the microbial community, mainly include the expansion or reduction of the population size of specific microbial members, or the introduction of new members and loss of some members. It was reported that disease-specific virome communities were found in IBD (Norman et al. 2015). (2) Alteration in gene-expression patterns. It has been found that *Bacilli* that inhabits alpine lakes could adapt to low temperatures by activating hydrolytic enzymes and regulating membrane fluidity (Yadav et al. 2016). (3) Variation of genetic characteristics. Microbiota can increase their fitness to extreme environments through the exchange of genetic materials. For example, phage-mediated HGT can transfer antibiotic resistance genes between bacterial hosts, leading to the ineffectiveness of antibiotics (Weinbauer 2004). Overall, the changes of microbial communities and their new metabolic traits could facilitate the adaptation to extreme environments.

In alpine lakes, extreme weather frequently occurs (Anneville et al. 2010) and is considered the primary disturbance type (Gallina et al. 2011; Kasprzak et al. 2017). Our sampling period belongs to the growing season with relatively high temperatures (15.1°C±0.12°C). However, a short-term cooling period appeared, which was characterized by low temperature (6.7°C±0.31°C) and rainfall. It was observed that bacterial communities mainly adapt to extreme weather disturbance by increasing or decreasing specific bacterial populations. During cooling period, the relative abundance of Gammaproteobacteria (belong to Proteobacteria) rose, mainly due to the rapid growth of *Acinetobacter rhizosphaerae* with higher tolerance to low temperature. While the relative abundance of Betaproteobacteria dropped, among its genera, the relative abundance of *Limnohabitans,* which had less adaptability and competitiveness in lower temperature and dissolved organic carbon (DOC) concentration, was observed to reduce.

Furthermore, the decreased relative abundance of Actinobacteria and Bacteroidetes was related to low temperature and low DOC concentration during the cooling period. However, when the warmer weather returned, Betaproteobacteria regained dominance, and Limnohabitans grew fast because of their rapid development in favorable DOC concentrations and higher temperatures. The relative abundance of Actinobacteria and Bacteroidetes also rose with increasing temperature and

DOC concentration. In addition to temperature, nutrient availability can also cause changes in the microbial community (Bonilla-Rosso et al. 2012). In deep groundwater, the supply of nutrients is the key factor that affects the microbial community here (Hubalek et al. 2016; Yan et al. 2020). It was shown that virome communities in deep groundwater changed to adapt to different levels of nutrient supply. In general, the viral diversity (Shannon) and richness (Ace) decreased with increasing depth, accompanied by low nutrient input. When we further revealed the composition of the core viral group, *Myoviridae*, *Siphoviridae,* and *Unclassified Caudovirales* were major components. However, different depths had distinct preferences for the core viral group components, and the number of dominant components of the core viral group decreased as the depth increased. This might be because increased environmental restrictions have reduced available niche dimensions, which further restricts microbial colonization and reproduction in the deeper groundwater (Mehrshad et al. 2021; Yan et al. 2020).

Moreover, the changes were observed in the proportion of abundant viruses and rare viruses at different depths, which was also strategies of virome communities to adapt to environmental variations. Especially, abundant viruses that perform main ecological functions dominated in the deeper aquifer, contributing to the stability of simple ecosystems, whereas rare viruses that can enhance ecological resilience dominated in the upper aquifer, aiding the resilience of complex ecosystems. In the human-associated extreme environments, with the development of the esophageal disease from BE to EAC, the richness (Ace) and diversity (Shannon) of the gut viral community decreased. Although PCoA analysis revealed no significant differences in gut viral communities of BE and EAC patients, comparatively abundant viruses with a higher relative abundance and rare viruses with a higher number of viral contigs did. Overall, extreme environments support distinct microbial communities, which are sensitive to environmental variations.

In addition to alterations in microbial communities, microbiota also has specific lifestyles in extreme environments. It was shown that viruses tended to sustain the lysogenic cycle in deep groundwater and the gut of BE and EAC patients. This temperate life cycle caused by low nutrient levels and bacterial densities of deep groundwater could potentially enhance the survival chances of bacteria and viruses, allowing them to thrive again under appropriate conditions (Boras et al. 2009; Herndl 1991). In the gut of BE and EAC patients, the lysogenic cycle allowed phage to insert

its genome, which included virulence-related genes, into the bacterial genomes, potentially increasing the chance of viral replication by improving the fitness of bacterial hosts under deteriorating physical conditions (Mirzaei et al. 2020). Collectively, the lysogenic cycle is beneficial to the survival and propagation of viruses in extreme environments.

Moreover, when screening the viral genetic characteristics, it was found that viruses could transfer AMGs across their bacterial hosts and serve as genetic revisors, assisting bacterial strains in obtaining adaptative features to different environmental conditions, which in turn is favorable to viral development (Keeling and Palmer 2008). In deep groundwater, it has been shown that AMGs were closely related to their surrounding environmental characteristics and bacterial metabolism. Depth-specific AMGs including *cofF*, *cysH*, *luxR,* and *spo0A* were found in the genome of viruses. Specifically, *cofF,* which encodes the alpha-L-glutamate ligases of methanogenic coenzyme F420 (Li et al. 2003), is involved in methane metabolism. The gene *cofF* was found mainly in the viral genome of BS, which belongs to an aerobic environment and serves as a habitat for aerobic bacteria, implying that viruses might help the metabolism of the aerobic methanogens that live here (Bogard et al. 2014). *CysH* encodes phosphoadenosine phosphosulfate reductase, which aids in sulfate assimilation (Berendt et al. 1995). Multiple copies of *cysH* were observed mainly in the genome of viruses of DA, where sulfur bedrock is widely distributed, indicating virus-carrying *cysH* could expand the sulfur metabolism function of bacteria and further enhance the adaptability of bacteria. *LuxR* and *spo0A* are genes related to quorum sensing (Ahmer et al. 1998) that appeared in the viral genome of deep groundwater, where bacterial competition for poor nutrient resources is fierce. *LuxR* and *spo0A* could help bacteria regulate bacterial population density to improve bacterial survival ability in the oligotrophic environment (Bassler 2002). Taken together, the differences of the major AMGs at different depths were strongly related to the environmental factors. They might explain the metabolic types and survival strategies of bacteria residing there. In the gut of BE and EAC patients, AMGs related to bacterial exotoxin (*spyA*, *tccC*, *entB*, *entD*) and LPS biosynthesis proteins (*lpxD*, *kdsC*, *gmnB*) were found in the genome of gut viruses. Specifically, *spyA* was shown to be slightly more abundant in the gut viruses of BE and EAC patients, possibly facilitating the production of bacterial exotoxins, which alter cytoskeletal structures and promote pathogenic bacterial colonization (Coye and Collins 2004). The relatively high abundance of AMGs associated with LPS biosynthesis proteins was also observed in the gut viruses of BE and EAC patients, suggesting Gram-negative bacterial dominance and the possible inflammatory

consequences of phage-bacteria interactions. Viruses that carrying these AMGs may accelerate the severity of esophageal-related diseases. Collectively, the acquisition of viral AMGs could help the survival strategy of bacteria in adapting to extreme environments, ultimately resulting in viral propagation.

Overall, microbiota could alter their community structure, lifestyle, genetic traits, etc., to adapt extreme environments, ultimately benefiting the future microbial community development.

4.3 The resilience of microbial community to disturbance in extreme environments

The microbial community is generally resilient and quickly returns to its pre-disturbance condition because of its rapid growth rate and physiological flexibility (Allison and Martiny 2008). It has been reported that the bacterial community returned to its pre-disturbance state after experiencing disturbances such as typhoons and human intervention (Downing et al. 2008; Jones et al. 2008). In the alpine lake study, the resilience of bacterial communities was evident in the alteration of Betaproteobacteria and Gammaproteobacteria. In detail, when a sudden short-term cooling period happened, a rapid shift from Betaproteobacteria to Gammaproteobacteria was observed. However, when warmer weather conditions were restored, Betaproteobacteria regained dominance. When the bacterial community composition at OTU level was examined, a high similarity was found between the beginning and end of the growing season. It might indicate that the bacterial community in an alpine lake was resilient to short-term weather deterioration throughout the growing season.

Additionally, rare microbiota reflects the ability of the microbial community to recover their original state after being disturbed due to their low abundance and various genotypes (Galand et al. 2009). In deep groundwater, it has been found that higher number and higher relative abundance of rare viruses decreased with the increasing depth, indicating that rare viruses contribute to enhance the resilience of complex ecosystems that are susceptible to disturbance. Overall, the microbial communities may have higher resilience to respond to disturbance in fluctuated extreme environments such as the alpine lake and upper aquifer of deep groundwater.

4.4 The interplay of bacteria and viruses in extreme environments and potential influence on habitats

In extreme environments, the relationship between bacteria and viruses could, in turn, affect their survival and prosperity. Generally, viruses change the bacterial community by maintaining the lytic cycle to kill the bacterial host (Reyes et al. 2010). However, in extreme environments (deep groundwater, the gut of patients with esophageal-related diseases), viruses tended to perform a lysogenic cycle that inserts their genome into the bacterial genome as prophage to maintain a temperate coexistence with bacteria, which could increase the bacterial fitness. It has been reported that the prophage in the genome of *E.coli* promoted the growth of *E.coli* cells and enhanced *E.coli* cells to resist environmental stresses (such as osmotic, oxidative, and acid stresses) (Wang et al. 2010). Moreover, during the infection, viruses could obtain beneficial AMGs. Through HGT, the viral AMGs were introduced to the genome of their bacterial hosts to help prompt bacterial fitness and expand bacterial metabolism function in extreme environments (deep groundwater, the gut of patients with esophageal-related diseases). In deep groundwater, multiple HGTs appeared in the *cysH* related to sulfur metabolism. This might be because sulfur works as an elector acceptor to sustain the fundamental survival metabolism of bacteria in deep groundwater with an anoxic environment, allowing bacteria to gain adaptability in oxygen-deficient environments (Thomas and Nielsen 2005). In turn, bacterial adaptability could benefit the development of viruses in an extreme environment. Meanwhile, the interplay of bacteria and viruses also has an impact on the function of extreme environments. For example, the HGT of *cysH* in deep groundwater could prompt the sulfur metabolism in deep groundwater, further contributing to the biogeochemical cycle. While in the gut of BE and EAC patients, the higher abundance of *spyA* may contribute to the production of bacterial exotoxins, leading to further deterioration of the human body microenvironments (Coye and Collins 2004). Overall, the interplay between bacteria and viruses formed in extreme environments, in turn, enhances their resistance to extreme conditions, ultimately benefit to their survival and prosperity. Besides, the interplay of bacteria and viruses influences the habitats as well. In this thesis, their interplay is critical for the function of deep groundwater and affects the health of the human body.

Chapter 5

5 Conclusion and outlook

The present study revealed the microbial community (mainly bacteria and viruses) in extreme natural environments and human-associated environments. Overall, extreme environments harbor unique microbial communities, and the microbial communities are sensitive to environmental variations. Moreover, microbial communities have developed various survival strategies, including changing community structure, lifestyles, and individual genetic characteristics to adapt to extreme environments. Meanwhile, when the disturbance is experienced, the microbial community is resilient and quickly returns to its pre-disturbance state in extreme environments. In conclusion, this study revealed the changes, adaptability, and resilience of microbial communities in extreme environments, further supplementing the existing knowledge about the microbial community and function in extreme environments.

The exploration of microbial communities in extreme environments could enlarge the human understanding of unknown microbiota. The analysis of their genomes and potential functions is conducive to microbial application in industries, agriculture, and medicine. For example, cold-adaptation enzymes expressed by microbiota provide an opportunity to study the life adaptation of cold habitats and the development of biotechnology (Yadav et al. 2016). Besides, through revealed the microbial communities and their functional traits in human-associated environments, the indicator might be found to predict the upper gastrointestinal diseases and might have an impact on early diagnosis of esophageal diseases.

Acknowledgments

The completion of the thesis heralded the end of my five-year doctoral career. Looking back on the past few years in Germany, my heart is full of gratitude. I thank the teachers, friends, and family who encouraged me to study abroad because it gave me an unforgettable experience in my lifetime. I thank China Scholarship Council for providing my living and study expenses abroad. I thank Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt and the Institute of Virology for supplying the academic environment to improve my professional skills.

I would like to thank my supervisor Dr. Li Deng give me the chance to study in her group "Virus in Nature and Health." I am very grateful that she always guides and supports me to think more about my projects. Her rigorous scientific attitude and active exploration spirit have profoundly affected me. I would like to thank my second supervisor Dr. Jinling Xue. When I am confused about experiments or academic issues, she always explained patiently and gave me valuable advice.

I would like to thank Dr. Judith Kaletta, she told me the laboratory rules and taught me the details of the experiment when I first came to the laboratory. I would like to thank Jinlong Ru. He kindly helped me a lot in my data analysis.

I would like to thank Dr. Jinling Xue, Shiqi Luo, Wanqi Huang. Their company and help in daily life keep me from feeling lonely abroad. I would like to thank the other members of our group Dr. Khan Mirzaei Mohammadali, Dr. Nguyen Fabian, Dr. Costa Rita, Dr. Würstle Silvia, Schulz Sarah, Pfeifer-Nigisch Sarah-Irina, Smith Sophie, Unterer Magdalena, and Tiamani Kawtar, for their valuable suggestions during my work.

Finally, I would like to thank my family for their continued encouragement and support. They have always been my solid backing and my driving force.

References

Abbaszadegan, M., M. Lechevallier & C. Gerba, 2003. Occurrence of viruses in US groundwaters. Journal-American Water Works Association 95(9):107-120.

Ahmer, B. M., J. Van Reeuwijk, C. D. Timmers, P. J. Valentine & F. Heffron, 1998. Salmonella typhimurium encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid. Journal of bacteriology 180(5):1185.

Allison, S. D. & J. B. Martiny, 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105(Supplement 1):11512-11519.

Anesti, V., I. R. McDonald, M. Ramaswamy, W. G. Wade, D. P. Kelly & A. P. Wood, 2005. Isolation and molecular detection of methylotrophic bacteria occurring in the human mouth. Environmental Microbiology 7(8):1227-1238.

Anneville, O., J. C. Molinero, S. Souissi & D. Gerdeaux, 2010. Seasonal and interannual variability of cladoceran communities in two peri-alpine lakes: uncoupled response to the 2003 heat wave. Journal of Plankton Research 32(6):913-925.

Bäckhed, F., H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich & J. I. Gordon, 2004. The gut microbiota as an environmental factor that regulates fat storage. Proceedings of the national academy of sciences 101(44):15718-15723.

Banerjee, S. & E. S. Robertson, 2019. Future Perspectives: Microbiome, Cancer and Therapeutic Promise Microbiome and Cancer. Springer, 363-389.

Bang, C., T. Dagan, P. Deines, N. Dubilier, W. J. Duschl, S. Fraune, U. Hentschel, H. Hirt, N. Hülter & T. Lachnit, 2018. Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? Zoology 127:1-19.

Banwart, S., E.-L. Tullborg, K. Pedersen, E. Gustafsson, M. Laaksoharju, A.-C. Nilsson, B. Wallin & P. Wikberg, 1996. Organic carbon oxidation induced by large-scale shallow water

intrusion into a vertical fracture zone at the Äspö Hard Rock Laboratory (Sweden). Journal of Contaminant Hydrology 21(1-4):115-125.

Bassler, B. L., 2002. Small talk: cell-to-cell communication in bacteria. Cell 109(4):421-424.

Bastviken, D., J. Ejlertsson, I. Sundh & L. Tranvik, 2003. Methane as a source of carbon and energy for lake pelagic food webs. Ecology 84(4):969-981.

Beatty, J. T., J. Overmann, M. T. Lince, A. K. Manske, A. S. Lang, R. E. Blankenship, C. L. Van Dover, T. A. Martinson & F. G. Plumley, 2005. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proceedings of the National Academy of Sciences 102(26):9306-9310.

Berendt, U., T. Haverkamp, A. Prior & J. D. Schwenn, 1995. Reaction mechanism of thioredoxin: 3′-phospho-adenylylsulfate reductase investigated by site-directed mutagenesis. European journal of biochemistry 233(1):347-356.

Berger, A. K. & B. A. Mainou, 2018. Interactions between enteric bacteria and eukaryotic viruses impact the outcome of infection. Viruses 10(1):19.

Bik, E. M., P. B. Eckburg, S. R. Gill, K. E. Nelson, E. A. Purdom, F. Francois, G. Perez-Perez, M. J. Blaser & D. A. Relman, 2006. Molecular analysis of the bacterial microbiota in the human stomach. Proceedings of the National Academy of Sciences 103(3):732-737.

Bogard, M. J., P. A. Del Giorgio, L. Boutet, M. C. G. Chaves, Y. T. Prairie, A. Merante & A. M. Derry, 2014. Oxic water column methanogenesis as a major component of aquatic CH 4 fluxes. Nature communications 5(1):1-9.

Bokulich, N. A., T. S. Collins, C. Masarweh, G. Allen, H. Heymann, S. E. Ebeler & D. A. Mills, 2016. Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. MBio 7(3):e00631-16.

Bonanomi, G., F. De Filippis, M. Zotti, M. Idbella, G. Cesarano, S. Al-Rowaily & A. Abd-ElGawad, 2020. Repeated applications of organic amendments promote beneficial microbiota, improve soil fertility and increase crop yield. Applied Soil Ecology 156:103714.

Bonilla-Rosso, G., M. Peimbert, L. D. Alcaraz, I. Hernández, L. E. Eguiarte, G. Olmedo-Alvarez & V. Souza, 2012. Comparative metagenomics of two microbial mats at Cuatro Ciénegas

Basin II: community structure and composition in oligotrophic environments. Astrobiology 12(7):659-673.

Boras, J. A., M. M. Sala, E. Vázquez-Domínguez, M. G. Weinbauer & D. Vaqué, 2009. Annual changes of bacterial mortality due to viruses and protists in an oligotrophic coastal environment (NW Mediterranean). Environmental microbiology 11(5):1181-1193.

Borchardt, M. A., P. D. Bertz, S. K. Spencer & D. A. Battigelli, 2003. Incidence of enteric viruses in groundwater from household wells in Wisconsin. Appl Environ Microbiol 69(2):1172-1180.

Bosch, A., S. Guix, D. Sano & R. M. Pinto, 2008. New tools for the study and direct surveillance of viral pathogens in water. Current Opinion in Biotechnology 19(3):295-301.

Brandt, L. J., 2013. American Journal of GastroenterologyLecture: Intestinal Microbiota and the Role of Fecal Microbiota Transplant (FMT) in Treatment ofC. difficileInfection. Official journal of the American College of Gastroenterology| ACG 108(2):177-185.

Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon & F. Rohwer, 2003. Metagenomic analyses of an uncultured viral community from human feces. Journal of bacteriology 185(20):6220-6223.

Bullerjahn, G. S. & A. F. Post, 2014. Physiology and molecular biology of aquatic cyanobacteria. Frontiers in microbiology 5:359.

Bushnell, B., 2019. BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); 2014.

Cadwell, K., 2015. The virome in host health and disease. Immunity 42(5):805-813.

Camacho, A., J. Erez, A. Chicote, M. Florín, M. M. Squires, C. Lehmann & R. Backofen, 2001. Microbial microstratification, inorganic carbon photoassimilation and dark carbon fixation at the chemocline of the meromictic Lake Cadagno (Switzerland) and its relevance to the food web. Aquatic Sciences 63(1):91-106.

Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen & R. Knight, 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26(2):266-267.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich & J. I. Gordon, 2010b. QIIME allows analysis of high-throughput community sequencing data. Nature methods 7(5):335-336.

Carding, S. R., N. Davis & L. Hoyles, 2017. The human intestinal virome in health and disease. Alimentary pharmacology & therapeutics 46(9):800-815.

Cavicchioli, R., W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd & A. T. Classen, 2019. Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology 17(9):569-586.

Chanal, A., V. Chapon, K. Benzerara, M. Barakat, R. Christen, W. Achouak, F. Barras & T. Heulin, 2006. The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. Environmental microbiology 8(3):514-525.

Chao, J.-B., H.-J. Wang, Y.-B. Zhang, Z.-Q. Li, Y.-H. Liu, F.-J. Huo, C.-X. Yin, Y.-W. Shi & J.-J. Wang, 2017. A single pH fluorescent probe for biosensing and imaging of extreme acidity and extreme alkalinity. Analytica chimica acta 975:52-60.

Chen, J., A. Hanke, H. E. Tegetmeyer, I. Kattelmann, R. Sharma, E. Hamann, T. Hargesheimer, B. Kraft, S. Lenk & J. S. Geelhoed, 2017. Impacts of chemical gradients on microbial community structure. The ISME journal 11(4):920-931.

Chen, S., Y. Zhou, Y. Chen & J. Gu, 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17):i884-i890.

Clooney, A. G., T. D. Sutton, A. N. Shkoporov, R. K. Holohan, K. M. Daly, O. O'Regan, F. J. Ryan, L. A. Draper, S. E. Plevy & R. P. Ross, 2019. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. Cell host & microbe 26(6):764-778. e5.

COLEMAN, D. C., D. J. SULLIVAN, R. J. RUSSEL, J. P. ARBUTHNOTT, B. F. CAREY & H. M. POMEROY, 1989. Staphylococcus aureus bacteriophages mediating the simultaneous lysogenic conversion of β-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. Microbiology 135(6):1679-1697.

Correa, P., 1995. Helicobacter pylori and gastric carcinogenesis. The American journal of surgical pathology 19:S37-43.

Coryell, M., M. McAlpine, N. V. Pinkham, T. R. McDermott & S. T. Walk, 2018. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. Nature communications 9(1):1-9.

Coskun, Ö. K., V. Özen, S. D. Wankel & W. D. Orsi, 2019. Quantifying population-specific growth in benthic bacterial communities under low oxygen using H 2 18 O. The ISME journal 13(6):1546-1559.

Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon & R. Knight, 2009. Bacterial community variation in human body habitats across space and time. science 326(5960):1694-1697.

Coye, L. H. & C. M. Collins, 2004. Identification of SpyA, a novel ADP-ribosyltransferase of Streptococcus pyogenes. Molecular microbiology 54(1):89-98.

Craun, M. F., G. F. Craun, R. L. Calderon & M. J. Beach, 2006. Waterborne outbreaks reported in the United States. Journal of water and health 4(S2):19-30.

Dalmasso, M., C. Hill & R. P. Ross, 2014. Exploiting gut bacteriophages for human health. Trends in microbiology 22(7):399-405.

Danielopol, D. L., P. Pospisil & R. Rouch, 2000. Biodiversity in groundwater: a large-scale view. Trends in Ecology & Evolution 15(6):223-224.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu & G. L. Andersen, 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology 72(7):5069-5072.

Díaz-Muñoz, S. L. & B. Koskella, 2014. Bacteria–phage interactions in natural environments. Advances in applied microbiology 89:135-183.

Diaz Carrasco, J. M., N. A. Casanova & M. E. Fernández Miyakawa, 2019. Microbiota, gut health and chicken productivity: what is the connection? Microorganisms 7(10):374.

Dickson, R. P., J. R. Erb-Downward, H. C. Prescott, F. J. Martinez, J. L. Curtis, V. N. Lama & G. B. Huffnagle, 2014. Cell-associated bacteria in the human lung microbiome. Microbiome 2(1):1-10.

Downing, A. L., K. M. DeVanna, C. N. Rubeck-Schurtz, L. Tuhela & H. Grunkemeyer, 2008. Community and ecosystem responses to a pulsed pesticide disturbance in freshwater ecosystems. Ecotoxicology 17(6):539-548.

Ducey, T. F., M. B. Vanotti, A. D. Shriner, A. A. Szogi & A. Q. Ellison, 2010. Characterization of a microbial community capable of nitrification at cold temperature. Bioresource technology 101(2):491-500.

Duncan, S. H., P. Louis & H. J. Flint, 2007. Cultivable bacterial diversity from the human colon. Letters in applied microbiology 44(4):343-350.

Duncan, S. H., P. Louis, J. M. Thomson & H. J. Flint, 2009. The role of pH in determining the species composition of the human colonic microbiota. Environmental microbiology 11(8):2112-2122.

Edgar, R. C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460-2461.

Erez, Z., I. Steinberger-Levy, M. Shamir, S. Doron, A. Stokar-Avihail, Y. Peleg, S. Melamed, A. Leavitt, A. Savidor & S. Albeck, 2017. Communication between viruses guides lysis–lysogeny decisions. Nature 541(7638):488-493.

Espey, M. G., 2013. Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota. Free Radical Biology and Medicine 55:130-140.

Ettwig, K. F., M. K. Butler, D. Le Paslier, E. Pelletier, S. Mangenot, M. M. Kuypers, F. Schreiber, B. E. Dutilh, J. Zedelius & D. de Beer, 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature 464(7288):543-548.

Fan, L., Z. Wang, M. Chen, Y. Qu, J. Li, A. Zhou, S. Xie, F. Zeng & J. Zou, 2019. Microbiota comparison of Pacific white shrimp intestine and sediment at freshwater and marine cultured environment. Science of the Total Environment 657:1194-1204.

Fedorov, A. Y., D. A. Rychkov, E. A. Losev, T. N. Drebushchak & E. V. Boldyreva, 2019. Completing the picture of tolazamide polymorphism under extreme conditions: a low-temperature study. Acta Crystallographica Section C: Structural Chemistry 75(5):598-608.

Feiner, R., T. Argov, L. Rabinovich, N. Sigal, I. Borovok & A. A. Herskovits, 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nature Reviews Microbiology 13(10):641-650.

Feller, G., M. Thiry, J.-L. Arpigny, M. Mergeay & C. Gerday, 1990. Lipases from psychrotropic Antarctic bacteria. FEMS microbiology letters 66(1-3):239-243.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. evolution 39(4):783-791.

Feng, Z. & A. Weinberg, 2006. Role of bacteria in health and disease of periodontal tissues. Periodontology 2000 40(1):50-76.

Filippini, M., N. Buesing & M. O. Gessner, 2008. Temporal dynamics of freshwater bacterio-and virioplankton along a littoral–pelagic gradient. Freshwater biology 53(6):1114-1125.

Fillon, S. A., J. K. Harris, B. D. Wagner, C. J. Kelly, M. J. Stevens, W. Moore, R. Fang, S. Schroeder, J. C. Masterson & C. E. Robertson, 2012. Novel device to sample the esophageal microbiome—the esophageal string test. PloS one 7(9):e42938.

Findley, K., J. Oh, J. Yang, S. Conlan, C. Deming, J. A. Meyer, D. Schoenfeld, E. Nomicos, M. Park & H. H. Kong, 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498(7454):367-370.

Flemming, H.-C. & S. Wuertz, 2019. Bacteria and archaea on Earth and their abundance in biofilms. Nature Reviews Microbiology 17(4):247-260.

Fu, L., B. Niu, Z. Zhu, S. Wu & W. Li, 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150-3152.

Fuentes, S., E. Van Nood, S. Tims, I. Heikamp-de Jong, C. J. Ter Braak, J. J. Keller, E. G. Zoetendal & W. M. De Vos, 2014. Reset of a critically disturbed microbial ecosystem: faecal transplant in recurrent Clostridium difficile infection. The ISME journal 8(8):1621-1633.

Fuhrma, J. A. & C. A. Suttle, 1993. Viruses in marine planktonic systems. Oceanography 6(2):51-63.

Galand, P. E., E. O. Casamayor, D. L. Kirchman & C. Lovejoy, 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences 106(52):22427-22432.

Gallina, N., O. Anneville & M. Beniston, 2011. Impacts of extreme air temperatures on cyanobacteria in five deep peri-Alpine lakes. Limnology and Oceanography 70(2):186-196.

Gangwar, P., S. I. Alam, S. Bansod & L. Singh, 2009. Bacterial diversity of soil samples from the western Himalayas, India. Canadian journal of microbiology 55(5):564-577.

Gelderblom, H. R., 1996. Structure and classification of viruses. Medical Microbiology 4th edition.

Gogokhia, L., K. Buhrke, R. Bell, B. Hoffman, D. G. Brown, C. Hanke-Gogokhia, N. J. Ajami, M. C. Wong, A. Ghazaryan & J. F. Valentine, 2019. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. Cell host & microbe 25(2):285-299. e8.

González-Toril, E., E. Llobet-Brossa, E. Casamayor, R. Amann & R. Amils, 2003. Microbial ecology of an extreme acidic environment, the Tinto River. Applied and environmental microbiology 69(8):4853-4865.

Grgacic, E. V. & D. A. Anderson, 2006. Virus-like particles: passport to immune recognition. Methods 40(1):60-65.

Griebler, C. & T. Lueders, 2009. Microbial biodiversity in groundwater ecosystems. Freshwater Biology 54(4):649-677.

Hengsuwan, M., C. Heim, K. Simon & B. T. Hansen, 2015. Long-term investigation of Sr-isotope and rare earth element fractionation processes within three major aquifers in the Äspö Hard Rock Laboratory (Sweden). Geomicrobiology Journal 32(3-4):243-254.

Hernández-Sánchez, V., L. Molina, J. L. Ramos & A. Segura, 2016. New family of biosensors for monitoring BTX in aquatic and edaphic environments. Microbial biotechnology 9(6):858-867.

Herndl, G. J., 1991. Microbial biomass dynamics along a trophic gradient at the Atlantic Barrier Reef off Belize (Central America). Marine Ecology 12(1):41-51.

Hjelmsø, M. H., M. Hellmér, X. Fernandez-Cassi, N. Timoneda, O. Lukjancenko, M. Seidel, D. Elsässer, F. M. Aarestrup, C. Löfström & S. Bofill-Mas, 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLoS One 12(1).

Ho, A., D. P. Di Lonardo & P. L. Bodelier, 2017. Revisiting life strategy concepts in environmental microbial ecology. FEMS microbiology ecology 93(3):fix006.

Hooper, L. V. & J. I. Gordon, 2001. Commensal host-bacterial relationships in the gut. Science 292(5519):1115-1118.

Huang, Z.-R., W.-L. Guo, W.-B. Zhou, L. Li, J.-X. Xu, J.-L. Hong, H.-P. Liu, F. Zeng, W.-D. Bai & B. Liu, 2019. Microbial communities and volatile metabolites in different traditional fermentation starters used for Hong Qu glutinous rice wine. Food research international 121:593-603.

Hubalek, V., X. Wu, A. Eiler, M. Buck, C. Heim, M. Dopson, S. Bertilsson & D. Ionescu, 2016. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. The ISME journal 10(10):2447-2458.

Hurwitz, B. L., L. Deng, B. T. Poulos & M. B. Sullivan, 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. Environmental microbiology 15(5):1428-1440.

Huse, S. M., L. Dethlefsen, J. A. Huber, D. M. Welch, D. A. Relman & M. L. Sogin, 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet 4(11):e1000255.

Hyatt, D., G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer & L. J. Hauser, 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics 11(1):1-11.

Imhann, F., M. J. Bonder, A. V. Vila, J. Fu, Z. Mujagic, L. Vork, E. F. Tigchelaar, S. A. Jankipersadsing, M. C. Cenit & H. J. Harmsen, 2016. Proton pump inhibitors affect the gut microbiome. Gut 65(5):740-748.

Jang, H. B., B. Bolduc, O. Zablocki, J. H. Kuhn, S. Roux, E. M. Adriaenssens, J. R. Brister, A. M. Kropinski, M. Krupovic & R. Lavigne, 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nature biotechnology 37(6):632-639.

Jiao, S., Y. Xu, J. Zhang, X. Hao & Y. Lu, 2019. Core microbiota in agricultural soils and their potential associations with nutrient cycling. Msystems 4(2):e00313-18.

John, S. G., C. B. Mendez, L. Deng, B. Poulos, A. K. M. Kauffman, S. Kern, J. Brum, M. F. Polz, E. A. Boyle & M. B. Sullivan, 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environmental microbiology reports 3(2):195-202.

Jones, D. T., W. R. Taylor & J. M. Thornton, 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8(3):275-282.

Jones, S. E., C.-Y. Chiu, T. K. Kratz, J.-T. Wu, A. Shade & K. D. McMahon, 2008. Typhoons initiate predictable change in aquatic bacterial communities. Limnology and Oceanography 53(4):1319-1326.

Kado, C. I., 2015. Historical events that spawned the field of plasmid biology. Plasmids: Biology and Impact in Biotechnology and Discovery:1-11.

Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato & K. Morishima, 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 45(D1):D353-D361.

Kasprzak, P., T. Shatwell, M. O. Gessner, T. Gonsiorczyk, G. Kirillin, G. Selmeczy, J. Padisák & C. Engelhardt, 2017. Extreme weather event triggers cascade towards extreme turbidity in a clear-water lake. Ecosystems 20(8):1407-1420.

Kawagoshi, Y., N. Hino, A. Fujimoto, M. Nakao, Y. Fujita, S. Sugimura & K. Furukawa, 2005. Effect of inoculum conditioning on hydrogen fermentation and pH effect on bacterial community relevant to hydrogen production. Journal of Bioscience and Bioengineering 100(5):524-530.

Keeling, P. J. & J. D. Palmer, 2008. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9(8):605-618.

Kellum, J. A., 2000. Determinants of blood pH in health and disease. Critical care 4(1):1-9.

Kieft, K., Z. Zhou & K. Anantharaman, 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8(1):1-23.

Kitajima, M. & C. Gerba, 2015. Aichi virus 1: environmental occurrence and behavior. Pathogens 4(2):256-268.

Koksal, Y., U. Caliskan & E. Unal, 2009. Hypothermia in a child with Hodgkin disease. Journal of pediatric hematology/oncology 31(2):136-138.

Koliarakis, I., I. Messaritakis, T. K. Nikolouzakis, G. Hamilos, J. Souglakos & J. Tsiaoussis, 2019. Oral bacteria and intestinal dysbiosis in colorectal cancer. International journal of molecular sciences 20(17):4146.

Korzhenkov, A. A., S. V. Toshchakov, R. Bargiela, H. Gibbard, M. Ferrer, A. V. Teplyuk, D. L. Jones, I. V. Kublanov, P. N. Golyshin & O. V. Golyshina, 2019. Archaea dominate the microbial community in an ecosystem with low-to-moderate temperature and extreme acidity. Microbiome 7(1):1-14.

Kuddus, M. & P. W. Ramteke, 2012. Recent developments in production and biotechnological applications of cold-active microbial proteases. Critical reviews in microbiology 38(4):330-338.

Kumar, S., G. Stecher, M. Li, C. Knyaz & K. Tamura, 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution 35(6):1547-1549.

Kusters, J. G., A. H. Van Vliet & E. J. Kuipers, 2006. Pathogenesis of Helicobacter pylori infection. Clinical microbiology reviews 19(3):449-490.

Kvachadze, L., N. Balarjishvili, T. Meskhi, E. Tevdoradze, N. Skhirtladze, T. Pataridze, R. Adamia, T. Topuria, E. Kutter & C. Rohde, 2011. Evaluation of lytic activity of

staphylococcal bacteriophage Sb-1 against freshly isolated clinical pathogens. Microbial biotechnology 4(5):643-650.

Kyle, J. E., H. S. Eydal, F. G. Ferris & K. Pedersen, 2008. Viruses in granitic groundwater from 69 to 450 m depth of the Äspö hard rock laboratory, Sweden. The ISME Journal 2(5):571-574.

Laaksoharju, M., E.-L. Tullborg, P. Wikberg, B. Wallin & J. Smellie, 1999. Hydrogeochemical conditions and evolution at the Äspö HRL, Sweden. Applied Geochemistry 14(7):835-859.

Lawes, J. C., B. A. Neilan, M. V. Brown, G. F. Clark & E. L. Johnston, 2016. Elevated nutrients change bacterial community composition and connectivity: high throughput sequencing of young marine biofilms. Biofouling 32(1):57-69.

Lederberg, J. & A. T. McCray, 2001. Ome SweetOmics--A genealogical treasury of words. The scientist 15(7):8-8.

Lee, W.-J. & K. Hase, 2014. Gut microbiota–generated metabolites in animal health and disease. Nature chemical biology 10(6):416-424.

Lepage, P., M. C. Leclerc, M. Joossens, S. Mondot, H. M. Blottière, J. Raes, D. Ehrlich & J. Doré, 2013. A metagenomic insight into our gut's microbiome. Gut 62(1):146-158.

Lewis, P. F., M. A. Schmidt, X. Lu, D. D. Erdman, M. Campbell, A. Thomas, P. R. Cieslak, L. D. Grenz, L. Tsaknardis & C. Gleaves, 2009. A community-based outbreak of severe respiratory illness caused by human adenovirus serotype 14. The Journal of infectious diseases 199(10):1427-1434.

Li, F., M. A. Hullar, Y. Schwarz & J. W. Lampe, 2009. Human gut bacterial communities are altered by addition of cruciferous vegetables to a controlled fruit-and vegetable-free diet. The Journal of nutrition 139(9):1685-1691.

Li, H., H. Xu, D. E. Graham & R. H. White, 2003. Glutathione synthetase homologs encode α-L-glutamate ligases for methanogenic coenzyme F420 and tetrahydrosarcinapterin biosyntheses. Proceedings of the National Academy of Sciences 100(17):9785-9790.

Li, J., P. Liu, J. Wang, A. P. Roberts & Y. Pan, 2020. Magnetotaxis as an Adaptation to Enable Bacterial Shuttling of Microbial Sulfur and Sulfur Cycling Across Aquatic Oxic-Anoxic Interfaces. Journal of Geophysical Research: Biogeosciences 125(12):e2020JG006012.

Liao, Y., G. K. Smyth & W. Shi, 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923-930.

Lim, E. H., S.-L. Zhang, J.-L. Li, W.-S. Yap, T.-C. Howe, B.-P. Tan, Y.-S. Lee, D. Wong, K.-L. Khoo & K.-Y. Seto, 2009. Using whole genome amplification (WGA) of low-volume biopsies to assess the prognostic role of EGFR, KRAS, p53, and CMET mutations in advanced-stage non-small cell lung cancer (NSCLC). Journal of Thoracic Oncology 4(1):12-21.

Lim, E. S., Y. Zhou, G. Zhao, I. K. Bauer, L. Droit, I. M. Ndao, B. B. Warner, P. I. Tarr, D. Wang & L. R. Holtz, 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nature medicine 21(10):1228-1234.

Lindell, D., J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen & W. R. Hess, 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449(7158):83-86.

Liu, K., Y. Liu, N. Jiao, B. Xu, Z. Gu, T. Xing & J. Xiong, 2017. Bacterial community composition and diversity in Kalakuli, an alpine glacial-fed lake in Muztagh Ata of the westernmost Tibetan Plateau. FEMS microbiology ecology 93(7).

Liu, Y., T. Yao, N. Jiao, X. Liu, S. Kang & T. Luo, 2013. Seasonal dynamics of the bacterial community in Lake Namco, the largest Tibetan lake. Geomicrobiology Journal 30(1):17-28.

Liu, Y., T. Yao, S. Kang, N. Jiao, Y. Zeng, S. Huang & T. Luo, 2007. Microbial community structure in major habitats above 6000 m on Mount Everest. Chinese Science Bulletin 52(17):2350-2357.

Luoto, T. P. & L. Nevalainen, 2013. Climate change impacts on zooplankton and benthic communities in Lake Unterer Giglachsee (N iedere Tauern Alps, Austria). International review of hydrobiology 98(2):80-88.

Lupp, C. & B. B. Finlay, 2005. Intestinal microbiota. Current Biology 15(7):R235-R236.

Ma, T., Y. Jiang, A. H. Elbehery, S. Blank, R. Kurmayer & L. Deng, 2020. Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake. Hydrobiologia 847(2):535-548.

Mai-Prochnow, A., M. Clauson, J. Hong & A. B. Murphy, 2016. Gram positive and Gram negative bacteria differ in their sensitivity to cold plasma. Scientific reports 6(1):1-11.

Major, T. A., Y. Liu & W. B. Whitman, 2010. Characterization of energy-conserving hydrogenase B in Methanococcus maripaludis. Journal of bacteriology 192(15):4022-4030.

Maltepe, E. & O. D. Saugstad, 2009. Oxygen in health and disease: regulation of oxygen homeostasis-clinical implications. Pediatric research 65(3):261-268.

Manor, O., C. L. Dai, S. A. Kornilov, B. Smith, N. D. Price, J. C. Lovejoy, S. M. Gibbons & A. T. Magis, 2020. Health and disease markers correlate with gut microbiome composition across thousands of people. Nature communications 11(1):1-12.

McMahon, J. M., M. A. Morales, C. Pierleoni & D. M. Ceperley, 2012. The properties of hydrogen and helium under extreme conditions. Reviews of modern physics 84(4):1607.

Mehrshad, M., M. Lopez-Fernandez, J. Sundh, E. Bell, D. Simone, M. Buck, R. Bernier-Latmani, S. Bertilsson & M. Dopson, 2020. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. BioRxiv.

Mehrshad, M., M. Lopez-Fernandez, J. Sundh, E. Bell, D. Simone, M. Buck, R. Bernier-Latmani, S. Bertilsson & M. Dopson, 2021. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. Nature Communications 12(1):1-12.

Miettinen, H., R. Kietäväinen, E. Sohlberg, M. Numminen, L. Ahonen & M. Itävaara, 2015. Microbiome composition and geochemical characteristics of deep subsurface high-pressure environment, Pyhäsalmi mine Finland. Frontiers in microbiology 6:1203.

Mills, S., F. Shanahan, C. Stanton, C. Hill, A. Coffey & R. P. Ross, 2013. Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. Gut microbes 4(1):4-16.

Mirzaei, M. K., J. Xue, R. Costa, J. Ru, S. Schulz, Z. E. Taranu & L. Deng, 2020. Challenges of studying the human virome–relevant emerging technologies. Trends in Microbiology.

Mock, V. & M. Olsen, Current management of fatigue and anemia in patients with cancer. In: Seminars in oncology nursing, 2003. vol 19. Elsevier, p 36-41.

Moelling, K. & F. Broecker, 2016. Fecal microbiota transplantation to fight Clostridium difficile infections and other intestinal diseases. Bacteriophage 6(4):e1251380.

Morowitz, M. J., E. M. Carlisle & J. C. Alverdy, 2011. Contributions of intestinal bacteria to nutrition and metabolism in the critically ill. Surgical Clinics 91(4):771-785.

Münch, N. S., H.-Y. Fang, J. Ingermann, H. C. Maurer, A. Anand, V. Kellner, V. Sahm, M. Wiethaler, T. Baumeister & F. Wein, 2019. High-fat diet accelerates carcinogenesis in a mouse model of barrett's esophagus via interleukin 8 and alterations to the gut microbiome. Gastroenterology 157(2):492-506. e2.

Nakatsu, G., H. Zhou, W. K. K. Wu, S. H. Wong, O. O. Coker, Z. Dai, X. Li, C.-H. Szeto, N. Sugimura & T. Y.-T. Lam, 2018. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. Gastroenterology 155(2):529-541. e5.

Nash, H. A., 1981. Integration and excision of bacteriophage λ: the mechanism of conservative site specific recombination. Annual review of genetics 15(1):143-167.

Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon & S. Bertilsson, 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews 75(1):14-49.

Newton, R. J. & K. D. McMahon, 2011. Seasonal differences in bacterial community composition following nutrient additions in a eutrophic lake. Environmental Microbiology 13(4):887-899.

Norman, J. M., S. A. Handley, M. T. Baldridge, L. Droit, C. Y. Liu, B. C. Keller, A. Kambal, C. L. Monaco, G. Zhao & P. Fleshner, 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell 160(3):447-460.

Nurk, S., D. Meleshko, A. Korobeynikov & P. A. Pevzner, 2017. metaSPAdes: a new versatile metagenomic assembler. Genome research 27(5):824-834.

Nyyssönen, M., J. Hultman, L. Ahonen, I. Kukkonen, L. Paulin, P. Laine, M. Itävaara & P. Auvinen, 2014. Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. The ISME journal 8(1):126-138.

Ogilvie, L. A., J. Caplin, C. Dedi, D. Diston, E. Cheek, L. Bowler, H. Taylor, J. Ebdon & B. V. Jones, 2012. Comparative (meta) genomic analysis and ecological profiling of human gut-specific bacteriophage φB124-14. PloS one 7(4):e35053.

Ohhata, N., N. Yoshida, H. Egami, T. Katsuragi, Y. Tani & H. Takagi, 2007. An extremely oligotrophic bacterium, Rhodococcus erythropolis N9T-4, isolated from crude oil. Journal of bacteriology 189(19):6824-6831.

Pal, S., R. Biswas, A. Misra, A. Sar, S. Banerjee, P. Mukherjee & B. Dam, 2020. Poorly known microbial taxa dominate the microbiome of hypersaline Sambhar Lake salterns in India. Extremophiles 24(6):875-885.

Pei, Z., E. J. Bini, L. Yang, M. Zhou, F. Francois & M. J. Blaser, 2004. Bacterial biota in the human distal esophagus. Proceedings of the National Academy of Sciences 101(12):4250-4255.

Pei, Z. & L. Yang, 2005. Bacterial biota in reflux esophagitis and Barrett's esophagus. World journal of gastroenterology 11(46):7277.

Peterson, J., S. Garges, M. Giovanni, P. McInnes, L. Wang, J. A. Schloss, V. Bonazzi, J. E. McEwen, K. A. Wetterstrand & C. Deal, 2009. The NIH human microbiome project. Genome research 19(12):2317-2323.

Rajagopala, S. V., S. Vashee, L. M. Oldfield, Y. Suzuki, J. C. Venter, A. Telenti & K. E. Nelson, 2017. The human microbiome and cancer. Cancer Prevention Research 10(4):226-234.

Rakhuba, D., E. Kolomiets, E. S. Dey & G. Novik, 2010. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. Pol J Microbiol 59(3):145-55.

Ramesh, B., P. R. M. Reddy, G. Seenayya & G. Reddy, 2001. Effect of various flours on the production of thermostable β-amylase and pullulanase by Clostridium thermosulfurogenes SV2. Bioresource technology 76(2):169-171.

Ren, J., K. Song, C. Deng, N. A. Ahlgren, J. A. Fuhrman, Y. Li, X. Xie, R. Poplin & F. Sun, 2020. Identifying viruses from metagenomic data using deep learning. Quantitative Biology:1-14.

Ren, Z., A. Li, J. Jiang, L. Zhou, Z. Yu, H. Lu, H. Xie, X. Chen, L. Shao & R. Zhang, 2019. Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. Gut 68(6):1014-1023.

Reyes, A., M. Haynes, N. Hanson, F. E. Angly, A. C. Heath, F. Rohwer & J. I. Gordon, 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466(7304):334-338.

Ridlon, J. M., D. J. Kang, P. B. Hylemon & J. S. Bajaj, 2014. Bile acids and the gut microbiome. Current opinion in gastroenterology 30(3):332.

Rigottier-Gois, L., 2013. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. The ISME journal 7(7):1256-1261.

Rigotto, C., K. Hanley, P. Rochelle, R. De Leon, C. Barardi & M. Yates, 2011. Survival of adenovirus types 2 and 41 in surface and ground waters measured by a plaque assay. Environmental science & technology 45(9):4145-4150.

Rofner, C., H. Peter, N. Catalán, F. Drewes, R. Sommaruga & M. T. Pérez, 2017. Climate-related changes of soil characteristics affect bacterial community composition and function of high altitude and latitude lakes. Global change biology 23(6):2331-2344.

Rohwer, F. & R. Edwards, 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. Journal of bacteriology 184(16):4529-4535.

Rohwer, F. & R. V. Thurber, 2009. Viruses manipulate the marine environment. Nature 459(7244):207-212.

Roux, S., F. Enault, B. L. Hurwitz & M. B. Sullivan, 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.

Salamov, V. S. A. & A. Solovyevand, 2011. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture, biomedicine and environmental studies Hauppauge: Nova Science Publishers:61-78.

Salmaso, N., F. Buzzi, L. Garibaldi, G. Morabito & M. Simona, 2012. Effects of nutrient availability and temperature on phytoplankton development: a case study from large lakes south of the Alps. Aquatic Sciences 74(3):555-570.

Sarethy, I. P., Y. Saxena, A. Kapoor, M. Sharma, S. K. Sharma, V. Gupta & S. Gupta, 2011. Alkaliphilic bacteria: applications in industrial biotechnology. Journal of Industrial Microbiology and Biotechnology 38(7):769.

Saunders, J. R., H. Allison, C. E. James, A. J. McCarthy & R. Sharp, 2001. Phage-mediated transfer of virulence genes. Journal of Chemical Technology & Biotechnology 76(7):662-666.

Savage, D. C., 1977. Microbial ecology of the gastrointestinal tract. Annual review of microbiology 31(1):107-133.

Schirmer, M., S. P. Smeekens, H. Vlamakis, M. Jaeger, M. Oosting, E. A. Franzosa, R. Ter Horst, T. Jansen, L. Jacobs & M. J. Bonder, 2016. Linking the human gut microbiome to inflammatory cytokine production capacity. Cell 167(4):1125-1136. e8.

Schulz, H. N. & B. B. Jørgensen, 2001. Big bacteria. Annual Reviews in Microbiology 55(1):105-137.

Sellner, K. G., 1997. Physiology, ecology, and toxic properties of marine cyanobacteria blooms. Limnology and oceanography 42(5part2):1089-1104.

Shade, A. & J. Handelsman, 2012. Beyond the Venn diagram: the hunt for a core microbiome. Environmental microbiology 14(1):4-12.

Shade, A., J. S. Read, D. G. Welkie, T. K. Kratz, C. H. Wu & K. D. McMahon, 2011. Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. Environmental microbiology 13(10):2752-2767.

Shen, M., Q. Li, M. Ren, Y. Lin, J. Wang, L. Chen, T. Li & J. Zhao, 2019. Trophic status is associated with community structure and metabolic potential of planktonic microbiota in plateau lakes. Frontiers in microbiology 10:2560.

Singh, P., K. Jain, C. Desai, O. Tiwari & D. Madamwar, 2019. Microbial community dynamics of extremophiles/extreme environment Microbial diversity in the genomic era. Elsevier, 323-332.

Singhvi, N., V. Gupta, M. Gaur, V. Sharma, A. Puri, Y. Singh, G. P. Dubey & R. Lal, 2020. Interplay of human gut microbiome in health and wellness. Indian journal of microbiology 60(1):26-36.

Sinha, A. & C. F. Maurice, 2019. Bacteriophages: uncharacterized and dynamic regulators of the immune system. Mediators of inflammation 2019.

Snider, E. J., G. Compres, D. E. Freedberg, H. Khiabanian, Y. R. Nobel, S. Stump, A.-C. Uhlemann, C. J. Lightdale & J. A. Abrams, 2019. Alterations to the esophageal microbiome associated with progression from Barrett's esophagus to esophageal adenocarcinoma. Cancer Epidemiology and Prevention Biomarkers 28(10):1687-1693.

Sokatch, J. R., 2014. Bacterial physiology and metabolism. Academic Press.

Stanfors, R., I. Rhén, E.-L. Tullborg & P. Wikberg, 1999. Overview of geological and hydrogeological conditions of the Äspö hard rock laboratory site. Applied Geochemistry 14(7):819-834.

Sullivan, M. B., D. Lindell, J. A. Lee, L. R. Thompson, J. P. Bielawski & S. W. Chisholm, 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4(8):e234.

Suttle, C. A., 2005. Viruses in the sea. Nature 437(7057):356-361.

Suttle, C. A., 2007. Marine viruses—major players in the global ecosystem. Nature reviews microbiology 5(10):801-812.

Sutton, T. D. & C. Hill, 2019. Gut bacteriophage: current understanding and challenges. Frontiers in endocrinology 10:784.

Szymczak, W. A., D. Y. Goldstein, E. P. Orner, R. A. Fecher, R. T. Yokoda, K. A. Skalina, M. Narlieva, I. Gendlina & A. S. Fox, 2020. Utility of stool PCR for the diagnosis of COVID-19: comparison of two commercial platforms. Journal of Clinical Microbiology 58(9):e01369-20.

Takeuchi, O., K. Hoshino, T. Kawai, H. Sanjo, H. Takada, T. Ogawa, K. Takeda & S. Akira, 1999. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity 11(4):443-451.

Tetz, G., S. M. Brown, Y. Hao & V. Tetz, 2018. Parkinson's disease and bacteriophages as its overlooked contributors. Scientific reports 8(1):1-11.

Tetz, G., S. M. Brown, Y. Hao & V. Tetz, 2019. Type 1 diabetes: an association between autoimmunity, the dynamics of gut amyloid-producing E. coli and their phages. Scientific reports 9(1):1-11.

Thanbichler, M., S. C. Wang & L. Shapiro, 2005. The bacterial nucleoid: a highly organized and dynamic structure. Journal of cellular biochemistry 96(3):506-521.

Thingstad, T. F., 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnology and Oceanography 45(6):1320-1328.

Thomas, C. M. & K. M. Nielsen, 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature reviews microbiology 3(9):711-721.

Thompson, R., C. Kamenik & R. Schmidt, 2005. Ultra-sensitive Alpine lakes and climate change. Journal of Limnology 64(2):139.

Touchon, M., J. A. M. De Sousa & E. P. Rocha, 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Current opinion in microbiology 38:66-73.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight & J. I. Gordon, 2007. The human microbiome project. Nature 449(7164):804-810.

Umarov, R. K. & V. V. Solovyev, 2017. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PloS one 12(2):e0171410.

Van Berkum, P. & B. D. Eardly, 2002. The aquatic budding bacterium Blastobacter denitrificans is a nitrogen-fixing symbiont of Aeschynomene indica. Applied and Environmental Microbiology 68(3):1132-1136.

Van der Gucht, K., T. Vandekerckhove, N. Vloemans, S. Cousin, K. Muylaert, K. Sabbe, M. Gillis, S. Declerk, L. De Meester & W. Vyverman, 2005. Characterization of bacterial communities in four freshwater lakes differing in nutrient load and food web structure. FEMS Microbiology Ecology 53(2):205-220.

Van Elsas, J. & F. Boersma, 2011. A review of molecular methods to study the microbiota of soil and the mycosphere. European Journal of Soil Biology 47(2):77-87.

von Meijenfeldt, F. B., K. Arkhipova, D. D. Cambuy, F. H. Coutinho & B. E. Dutilh, 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome biology 20(1):1-14.

Wang, W., J. Ren, K. Tang, E. Dart, J. C. Ignacio-Espinoza, J. A. Fuhrman, J. Braun, F. Sun & N. A. Ahlgren, 2020. A network-based integrated framework for predicting virus–prokaryote interactions. NAR genomics and bioinformatics 2(2):lqaa044.

Wang, X., Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino & T. K. Wood, 2010. Cryptic prophages help bacteria cope with adverse environments. Nature communications 1(1):1-9.

Watanabe, M., K. Miyajima, I. Matsui, C. Tomiyama-Miyaji, E. Kainuma, M. Inoue, H. Matsumoto, Y. Kuwano & T. Abo, 2010. Internal environment in cancer patients and proposal that carcinogenesis is adaptive response of glycolysis to overcome adverse internal conditions. Health 2(07):781.

Weckström, K., J. Weckström, K. Huber, C. Kamenik, R. Schmidt, W. Salvenmoser, M. Rieradevall, T. Weisse, R. Psenner & R. Kurmayer, 2016. Impacts of climate warming on Alpine lake biota over the past decade. Arctic, Antarctic, and Alpine Research 48(2):361-376.

Weinbauer, M. G., 2004. Ecology of prokaryotic viruses. FEMS microbiology reviews 28(2):127-181.

Wiethaler, M., J. Slotta-Huspenina, A. Brandtner, J. Horstmann, F. Wein, T. Baumeister, N. Radani, S. Gerland, A. Anand & S. Lange, 2019. BarrettNET—a prospective registry for risk estimation of patients with Barrett's esophagus to progress to adenocarcinoma. Diseases of the Esophagus 32(8):doz024.

Wigington, C. H., D. Sonderegger, C. P. Brussaard, A. Buchan, J. F. Finke, J. A. Fuhrman, J. T. Lennon, M. Middelboe, C. A. Suttle & C. Stock, 2016. Re-examination of the relationship between marine virus and microbial cell abundances. Nature microbiology 1(3):1-9.

Wilhelm, S. W. & C. A. Suttle, 1999. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. Bioscience 49(10):781-788.

Wommack, K. E. & R. R. Colwell, 2000. Virioplankton: viruses in aquatic ecosystems. Microbiology and molecular biology reviews 64(1):69-114.

Wu, X., K. Holmfeldt, V. Hubalek, D. Lundin, M. Åström, S. Bertilsson & M. Dopson, 2016. Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. The ISME journal 10(5):1192-1203.

Wu, X., Y. Liu, Y. Shang, D. Liu, W. Liesack, Z. Cui, J. Peng & F. Zhang, 2021. Peat-vermiculite alters microbiota composition towards increased soil fertility and crop productivity. Plant and Soil:1-14.

Wylie, K. M., G. M. Weinstock & G. A. Storch, 2012. Emerging view of the human virome. Translational Research 160(4):283-290.

Xiong, J., Y. Liu, X. Lin, H. Zhang, J. Zeng, J. Hou, Y. Yang, T. Yao, R. Knight & H. Chu, 2012. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. Environmental microbiology 14(9):2457-2466.

Yadav, A. N., S. G. Sachan, P. Verma, R. Kaushik & A. K. Saxena, 2016. Cold active hydrolytic enzymes production by psychrotrophic Bacilli isolated from three sub-glacial lakes of NW Indian Himalayas. Journal of basic microbiology 56(3):294-307.

Yadav, N. & A. Yadav, 2019. Biodegradation of biphenyl compounds by soil microbiomes. Biodivers Int J 3:37-40.

Yan, L., M. Herrmann, B. Kampe, R. Lehmann, K. U. Totsche & K. Küsel, 2020. Environmental selection shapes the formation of near-surface groundwater microbiomes. Water research 170:115341.

Yang, L., F. Francois & Z. Pei, 2012. Molecular pathways: pathogenesis and clinical implications of microbiome alteration in esophagitis and Barrett esophagus. Clinical Cancer Research 18(8):2138-2144.

Yin, Q., B. Fu, B. Li, X. Shi, F. Inagaki & X.-H. Zhang, 2013. Spatial variations in microbial community composition in surface seawater from the ultra-oligotrophic center to rim of the South Pacific Gyre. PLoS One 8(2):e55148.

Young, V. B., 2017. The role of the microbiome in human health and disease: an introduction for clinicians. Bmj 356.

Yuki, K., M. Fujiogi & S. Koutsogiannaki, 2020. COVID-19 pathophysiology: A review. Clinical immunology 215:108427.

Zhernakova, A., A. Kurilshikov, M. J. Bonder, E. F. Tigchelaar, M. Schirmer, T. Vatanen, Z. Mujagic, A. V. Vila, G. Falony & S. Vieira-Silva, 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352(6285):565-569.

Zoetendal, E. G., A. D. Akkermans, W. M. Akkermans-van Vliet, J. A. G. de Visser & W. M. de Vos, 2001. The host genotype affects the bacterial community in the human gastronintestinal tract. Microbial ecology in health and disease 13(3):129-134.

Zuckerkandl, E. & L. Pauling, 1965. Evolutionary divergence and convergence in proteins Evolving genes and proteins. Elsevier, 97-166.

Appendix A

Manuscript 1-3

PRIMARY RESEARCH PAPER

Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake

Tianli Ma · Yiming Jiang · Ali H. A. Elbehery · Stephan Blank · Rainer Kurmayer ⬢ · Li Deng

Abstract The disturbing effect of a short-term cooling period during summer on planktonic bacterial community structure of an alpine lake was investigated using 16S rDNA pyrosequencing. Proteobacteria, Actinobacteria, and Bacteroidetes constituted the most abundant phyla. During the sampling period (from July to August 2010), a sudden cooling period with high precipitation occurred, as indicated by a decrease in conductivity, calcium, and dissolved organic carbon concentration resulting from increased runoff. The relative abundance of Actinobacteria, Betaproteobacteria, and Cyanobacteria decreased during this short-term cooling period. Instead, a rapid shift from Betaproteobacteria to Gammaproteobacteria occurred, which was mainly caused by an increase of *Acinetobacter rhizosphaerae*. Soon after the short-term cooling period, warmer weather conditions got re-established and Betaproteobacteria recovered and became again dominant. Non-metric multi-dimensional scaling analysis and Venn diagrams revealed a planktonic bacterial community composition with high similarity at the beginning and the end of the growing season. Air temperature and precipitation were significantly correlated with the observed variation in operational taxonomic unit (OTU) relative abundance. It is concluded that, in response to the short-term cooling period, a distinct planktonic bacterial OTU community developed. It rapidly diminished, however, as summer conditions became re-established, implying the recovery of the original bacterial community structure.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10750-019-04118-8) contains supplementary material, which is available to authorized users.

T. Ma · Y. Jiang · A. H. A. Elbehery · L. Deng (✉)
Institute of Virology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
e-mail: li.deng@helmholtz-muenchen.de

T. Ma · Y. Jiang · A. H. A. Elbehery · L. Deng
Institute of Virology, Technical University of Munich, Trogerstrasse 30, 81675 Munich, Germany

S. Blank · R. Kurmayer (✉)
Research Department for Limnology, Mondsee, University of Innsbruck, Mondseestrasse 9, 5310 Mondsee, Austria
e-mail: rainer.kurmayer@uibk.ac.at

Introduction

Bacterial communities play crucial roles in decomposing organic matter and nutrient cycling in aquatic ecosystems (Salcher et al., 2010), and their activities are driven by multiple ecological factors such as temperature (Ren et al., 2013), pH (Stepanauskas et al., 2003), and nutrient concentration (Van Der Gucht et al., 2001; Yannarell & Triplett, 2004). In lake ecosystems, local and climatic factors can affect the bacterial community by abiotic and biotic drivers (Battarbee et al., 2002; Castro et al., 2010). Although numerous studies have explored the effect of climatic changes on bacterial communities in lowland lakes (Allgaier & Grossart, 2006; Comte et al., 2006; Bertilsson et al., 2007), fewer studies have considered remote lakes located in the alpine vegetation zone. Compared to lowland lakes, these lakes are considered relatively pristine due to less intensive anthropogenic activities taking place on or around them (Weckström et al., 2016). In addition, alpine lakes are considered sensitive to disturbance, e.g., through nutrient introduction (Heiri and Lotter, 2003), water temperature rises depending on the altitude range where the reduction in snow cover duration is most pronounced (Thompson et al., 2005), and predator effects by introduced fish (Magnea et al., 2013).

In recent years, studies have paid attention to the response of bacterial communities in alpine lakes to sudden changes in weather conditions. For example, the bacterial community structure in an arctic lake showed a rapid change over a short time period after snow melting or rain (Crump et al., 2003), mainly because transient bacteria were washed into the lake and some rare bacteria profited from terrestrial organic matter entering the lake via terrestrial runoff (Whalen & Cornwell, 1985; O'Brien et al., 1997). In an experiment that simulated changes in soil characteristics induced by climate change, the availability of soil-derived phosphorus and carbon induced a change in the bacterial community structure within 3 days (Rofner et al., 2017). In addition to these indirect factors, temperature is one of the most important direct factors affecting the bacterial community structure (Lindström et al. 2005). Bacteria inhabiting in alpine lakes can adapt to local temperature by modulating membrane fluidity and gene expression (Yadav et al. 2016). Temperature can also indirectly mediate the release of nutrients from sediment and trophic transfer

and further affect bacterial growth and diversity (Rasconi et al., 2017). The resilience of bacterial communities to disturbance in general has been studied widely (e.g., Shade et al., 2012), and it has been generally argued that microbial composition is resilient and quickly return to its pre-disturbance state because of fast growth rate and physiological flexibility (e.g., Allison & Martiny, 2008). In lake ecosystems, studies have shown that the planktonic bacterial community (as well as phytoplankton or zooplankton) indeed return to their pre-disturbance state after experiencing pesticide input (Downing et al., 2008), disturbance of the water column stability caused by typhoon (Jones et al. 2008), or human intervention (Shade et al., 2011). Fewer studies have been performed in remote unproductive and relatively cold alpine lakes where the role of climatic disturbance might be more influential compared to warmer and more productive lowland regions. In a recent study, the average water temperature after ice break in spring until the sampling date has been observed as a limiting factor during the earlier growing season in alpine lakes (Jiang et al., 2019).

This study aimed investigate the changes in the bacterial community structure during the summer growing season occurring in a remote alpine lake, Unterer Giglachsee in the Niedere Tauern at 1922 m asl, in response to a local cooling period disturbance linked to rain and snowfall. For this purpose, we (i) monitored the local weather and environmental conditions, (ii) described the bacterial community composition turnover at the phylum, class, and genus levels, (iii) analyzed the effect size of meteorological conditions in comparison to chemicophysical parameters, and (iv) compared the bacterial community composition at the beginning and the end of the observation period to describe the resilience after the disturbance by the short-term cooling period.

Materials and methods

Site description and sampling

Unterer Giglachsee (47.28° N, 13.65° E) is an oligotrophic alpine lake, which is located in a catchment area dominated by carbonaceous bedrock (limestone and dolomite) at the tree and timberline. The altitude, length, width, depth, and area are 1922 m asl,

1 km, 40 to 280 m, 18 m, and 16.5 ha, respectively. Further detailed information is given by Weckström et al. (2016). Meteorological parameters for this area (including air temperature, average wind speed, precipitation, sunshine hours, and snow layer) were obtained from Zentralanstalt für Meteorologie und Geodynamik, who run three weather monitoring stations located more closely in the alpine region: Obertauern Station (1772 m), Rudolfshütte Station (2317 m), and Schmittenhöhe Station (1956 m). Water thermistors (MINILOG, Vemco Ltd.) were installed at 2.5 m depth in the deepest parts of the lake and used to monitor water temperature every 4 h. Depth-integrated water samples (1.5–2 l) were collected at the deepest part of the lake on July 19, 21, 27, and 31 and August 01, 04, 09, and 18 of year 2010. The water samples were first pre-filtered in the field through glass fiber filters using a hand vacuum pump (GF/C) to remove most of the eukaryotic algae and particle-associated bacteria. Then, the filtrate was filtered through nitrocellulose (NC) membranes (pore diameter 0.2 µm). NC membranes were transferred into Eppendorf tubes and were stored at − 20°C until DNA extraction.

DNA extraction, 16S rDNA amplification, and sequencing

NC membranes were loaded into a bead beater tube to physically break the cells, and DNA was subsequently isolated and purified using a NucleoSpin® Soil DNA extraction kit (Macherey–Nagel, Germany) and eluted in 40 µl elution buffer. The quality and quantity of extracted DNA were estimated using a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific). The extracted bacterial genomic DNA (100 ng) was amplified in a 50-µl reaction mix containing 10 µl of 5 × buffer HF, 1 µl of 10 mM of each deoxynucleotide triphosphate, 1 unit of a proofreading polymerase (Phusion Hot Start High-Fidelity DNA Polymerase; Thermo Fisher Scientific), and 2.5 µl of forward and reverse primers (10 pmol µl^{-1}) binding to conserved regions (V3 and V6) within the 16S rDNA. The forward primer (338F: 5′-CGTATCGCCTCCCT CGCGCCA TCAG ACGAGTGCGT ACTCCTA CGGGAGGCAGCAG-3′) and the reverse primer (1046R: 5′-CTATGCGCCTTGCCAGCCCGC TCAG ACGAGTGCGT CGACAGCCATGCANCACCT-3′) included the sequences representing the adaptor, key,

and barcode sequence (underlined and separated by a space), whereas the specific sequence binding to bacterial rDNA is indicated at the 3′ end (Huse et al., 2008). Polymerase chain reaction (PCR) products (709 bp) were obtained under the following cycling conditions: annealing temperature at 67.8°C, elongation for 30 s, and 20 cycles. After running the gel, PCR products were cut out from agarose gel and purified using the gel purification kit (Qiagen, Hilden, Germany). Four independent PCR amplicons were pooled. Purified amplicons were sequenced using a GS-FLX platform with 454 titanium chemistry. Four amplicons were sequenced in four regions separated by four-region gaskets, loading approximately 3,100,000 amplicon-coated beads per run and recovering a total of 800,000 sequence tags.

16S rDNA sequence processing

The sequences were processed using QIIME 19.0 (Caporaso et al., 2010). The multiplex reads were assigned to the original samples based on their barcode sequences. The raw reads were filtered by quality (Phred quality score > 2) and the sequences were clustered into operational taxonomic units (OTUs) at 97% sequence similarity. The UCLUST (Edgar, 2010) consensus taxonomy classifier was used to identify bacterial phylotypes from assigned OTUs. Representative sequences from each phylotype were aligned using PYNAST (Caporaso et al., 2009) and the sequence that was found to be most abundant in each OTU cluster was selected as the representative sequence. The Greengenes database (DeSantis et al., 2006) was used to identify the taxonomy of each phylotype. The raw sequencing data have been deposited in the National Center for Biotechnology Information (NCBI) Raw Sequence Read Archive under the following accession numbers: SAMN0990 6023, SAMN09906024, SAMN09906025, SAMN09 906026, SAMN09906027, SAMN09906028, SAMN 10743546, and SAMN10743550.

Statistical data analysis

Non-metric multi-dimensional scaling (NMDS) based on "Bray–Curtis" similarities was used to compare the bacterial community composition at the OTU level between samples (including all OTUs > 0.1% in total

reads). Analysis was performed using R (version 3.2, package vegan). Analysis of similarities (ANOSIM) was used to test whether there was a significant difference between groups of samples as identified by NMDS. Venn diagram analysis was performed at http://www.interactivenn.net/ and used to illustrate the OTUs shared between samples. For clarity, we only chose the most abundant OTUs ($> 1\%$ in total reads) in each sample. All other graphs were generated with OriginLab 8.

To study the relationship between bacterial community composition at the OTU level and meteorological variables, detrended correspondence analysis was applied to test whether a unimodal or a linear relationship between OTUs and variables was appropriate. The length of gradient varied from 3 to 4, implying that both unimodal and linear models could be used. The linear redundancy analysis (RDA) method was chosen (Lepš & Šmilauer, 2003), including the time dependence of consecutive samples as a covariate. As deduced from variance inflation factors, the variables air temperature (at noon) and sunshine hours showed high correlation. Thus, the latter were excluded from further analysis and the meteorological variables (air temperature, precipitation, and average wind speed) were tested using the forward selection procedure ($P < 0.05$). A Monte Carlo permutation test (499 randomized data sets) was used to test whether the influence of axis 1 was statistically significant at $P < 0.05$. Ordination analysis was performed using CANOCO software for Windows (version 4.5).

Results

Weather and environmental conditions during the study period

In general, relatively warm (average $\pm$ SE air temperature at noon, $15.1°C \pm 0.12°C$) and good weather during July 2010 was followed by a distinct cooling period at the end of July, as indicated by a decrease in air and water temperature (Fig. S1). The average $\pm$ SE air temperature during the cooling period was $6.7°C \pm 0.31°C$. On July 27, conductivity, calcium, and dissolved organic carbon (DOC) values all declined, indicating a significant runoff of rain and melting snow (Fig. 1). As this change was observed for a short time only, it was concluded that runoff

influence declined again in the second part of the study period, i.e., during August 2010. Nevertheless, water temperature remained relatively low compared to the summer period before the cooling period.

Sequencing output

A total of 104,612 raw sequences were obtained. After the removal of linkers, barcodes, and primers, as well as low-quality or ambiguous reads, a total of 89,508 high-quality reads were identified. For individual sampling dates, read numbers ranged from 2,041 to 20,611 (Table 1). In total, 2,484 OTUs were selected by applying the 97% sequence similarity threshold. Low abundance reads ($< 0.1\%$ in total reads) were removed for further analysis, resulting in 436 OTUs. The value of "Goods_coverage" indicated that, for all samples, the sequencing depth was sufficient to represent the actual microbial diversity.

Bacterial community composition at the phylum and class levels

Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Chloroflexi, Firmicutes, Thermi (Deinococcus–Thermus), and Verrucomicrobia constituted the most abundant phyla in the bacterial community composition (Fig. 2a). On average, Proteobacteria, Actinobacteria, and Bacteroidetes accounted for $> 90\%$ of all OTUs, with Proteobacteria constituting the most abundant phylum. Only on August 09, which was after the cooling period, the relative abundance of Actinobacteria reached 43.1% and exceeded that of Proteobacteria. Within the Proteobacteria, Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, and Gammaproteobacteria constituted the four major classes (Fig. 2b). In general, the relative abundance of Deltaproteobacteria was the lowest ($< 1\%$). Alphaproteobacteria ranged in relative abundance from 5.0% to 16.8%. However, during the cooling period (on July 27 and 31), the relative abundance of Gammaproteobacteria reached maxima of 55.9% and 46.8%, respectively. Furthermore, during the cooling period, the relative abundance of Betaproteobacteria was lower on July 27 (14.2%) and July 31 (14.6%) and lowest on August 09 (13.1%) compared to other sampling dates (30.2–47.8%). On average, the relative abundance of Actinobacteria was much lower during the cooling period (10.3%, 13.0%,

Fig. 1 a Water chemicophysical parameters and **b** air temperature and water temperature in the alpine lake Unterer Giglachsee during July and August 2010. The short-term cooling period is indicated by the shaded area
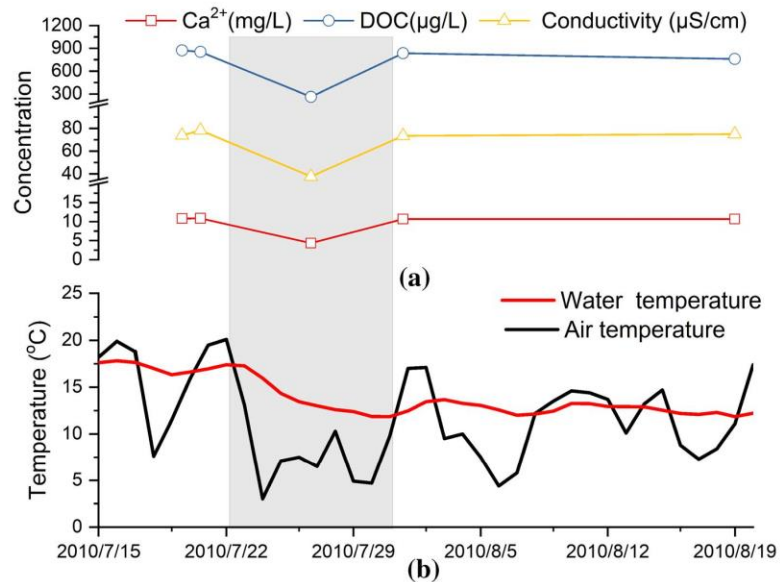
Table 1 Sequencing output for samples from alpine lake Unterer Giglachsee analyzed in the course of a cooling period in July/August 2010

Sampling (date)	Total reads	Total OTUs	Goods_coverage	Chao1	Fisher	Shannon	Simpson
July 19	5,647	375	0.967	747	99.81	5.38	0.92
July 21	10,308	390	0.984	670.78	84.88	5.69	0.96
July 27	11,759	328	0.985	766.9	66.49	3.49	0.7
July 31a	20,611	419	0.991	553.04[a]	50.56[a]	4.09[a]	0.82[a]
July 31b	7,227	245	0.983				
August 01	2,260	280	0.942	504.4	88.99	6.22	0.97
August 04	19,027	489	0.988	904.64	96.99	5.31	0.94
August 09	10,628	366	0.985	630.39	77.07	5.74	0.96
August 18	2,041	366	0.897	817.21	154.93	6.31	0.95

[a]July31a and July31b were used as technical replicates and were combined subsequently

and 6.4% on July 27, July 31, and August 01, respectively) compared to warmer weather conditions before and after. Similarly, the relative abundance of Bacteroidetes decreased on July 27 and 31 (9.2% and 6.3%, respectively) compared to warmer weather conditions before and after. The relative abundance of Cyanobacteria ranged from 1.6% to 5.4% and showed a similar trend of decline during the cooling period, with results for July 27 and 31 of 1.6% and 2.2%, respectively. On average, the other phyla (Nitrospirae, Gemmatimonadetes, Chlorobi, and Acidobacteria) contributed < 1% of all OTUs. Thus,

whereas the Gammaproteobacteria class seemed to increase during the short-term cooling period, other phyla and classes regained abundance once warmer weather conditions were re-established.

Short-term changes in the bacterial community composition at the genus level

In all samples, the genera that were taxonomically classified belonged to the phyla Proteobacteria, Actinobacteria, Firmicutes, Cyanobacteria, Bacteroidetes, and Thermi (Fig. 3). The relative abundance of the
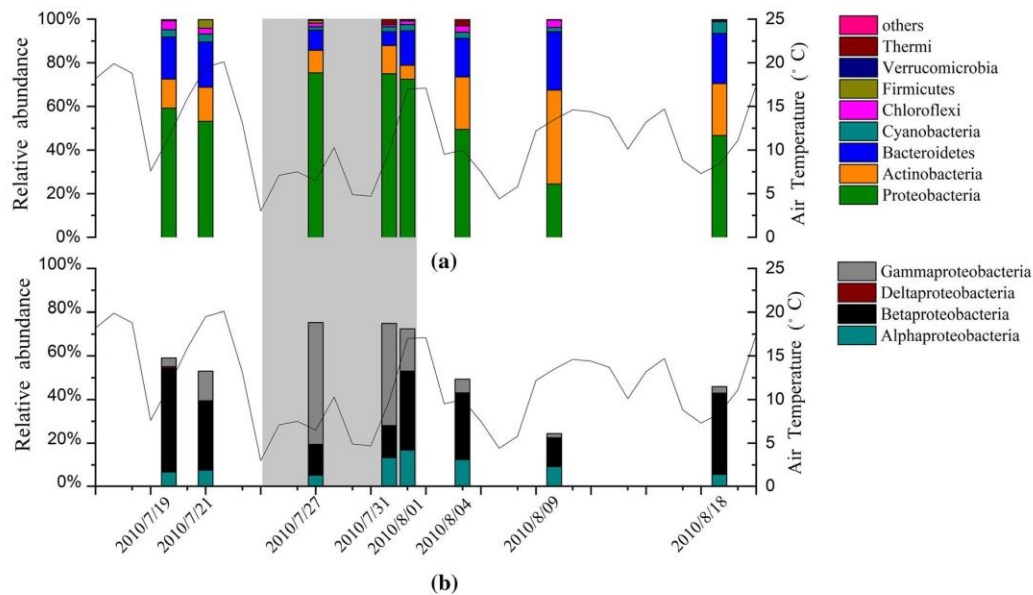
Fig. 2 Relative abundance of **a** bacterial phyla and **b** classes of Proteobacteria in the alpine lake Unterer Giglachsee during July and August 2010. The air temperature at noon (12:00 midday) is indicated by the straight gray line. The deterioration in weather conditions is indicated by the shaded area

assigned genera was highest on July 27 and 31, i.e., 64.5% and 62.8% and < 50% on all other sampling dates. Eight genera assigned to Proteobacteria were identified, with *Acinetobacter*, *Limnohabitans*, *Sphingomonas*, and *Polynucleobacter* found to be the most abundant (Fig. 4a). Several genera showed a significant increase in relative abundance in coincidence with the short-term cooling period. During the study period, *Acinetobacter* (17 OTUs) increased in relative abundance and occurred in highest number during the short-term cooling period, i.e., on July 27 and 31 (54.7% and 45.5%, respectively). *Sphingomonas* was not detected on July 19 and August 18, whereas it occurred at high relative abundance on July 31, August 01, and August 04 (3.2%, 7.4%, and 3.0%, respectively). However, the abundance of *Limnohabitans* (13 OTUs) declined to a minimum during the cooling period (0.3% and 0.4% on July 27 and 31, respectively), whereas its highest relative abundance was observed on July 21 (10.2%) followed by August 01 and 09 (7.7% and 4.0%, respectively). Similarly, the relative abundance of *Polynucleobacter* was higher at the beginning (July 19; 9.1%) and at the end (August

18; 6.8%) of the study period but declined during the short-term cooling period.

Within Actinobacteria, three genera with low abundance were identified, including *Propionibacterium* (1 OTU), *Rhodococcus* (2 OTUs), and *Mycobacterium* (3 OTUs; Fig. 4b). Within Cyanobacteria, the relative abundance of *Synechococcus* (7 OTUs) ranged from 0.1% (August 09) to 5.1% (August 18) and decreased during the short-term cooling period on July 27 (1.5%) and July 31 (2.1%; Fig. 4c). Within Bacteroidetes, *Flavobacterium* (7 OTUs), *Fluviicola* (15 OTUs), and *Sediminibacterium* (14 OTUs) were identified (Fig. 4d). The relative abundance of *Flavobacterium* ranged from 0.04% on July 31 to 1.3% on July 19. *Fluviicola* decreased in relative abundance during the cooling period (0.4% and 0.6% on July 27 and 31, respectively) and increased at the end of the study period (3.0%, 7.3%, and 3.0% on August 04, 09, and 18, respectively). Similarly, the relative abundance of *Sediminibacterium* decreased on July 27 and 31 (0.7% and 0.5%, respectively), whereas it occurred with higher relative abundance at the beginning and end of the observation period, i.e., on August 04 and 09 (1.3% and 7.3%,
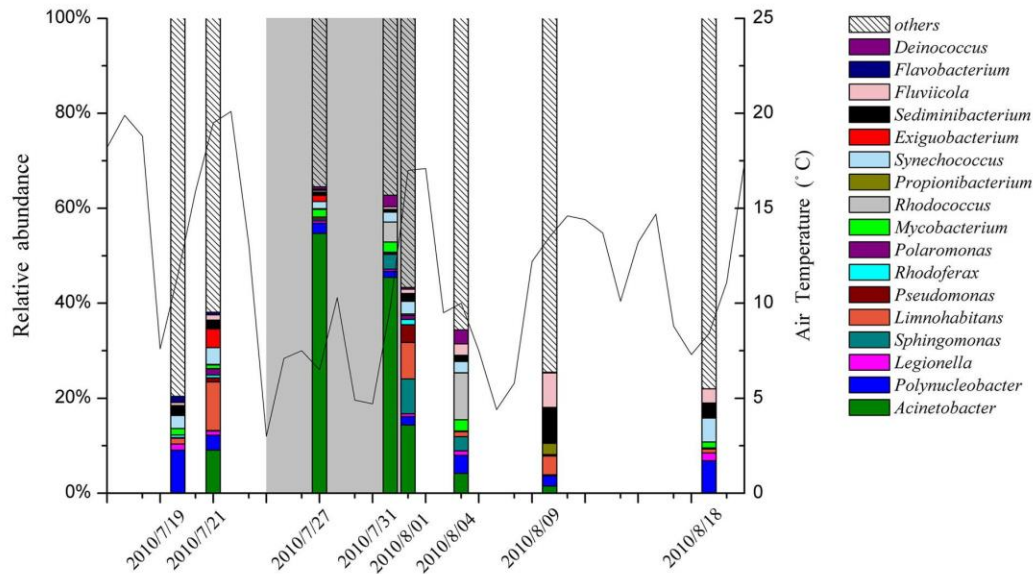
Fig. 3 Relative abundance of bacterial genera in the alpine lake Unterer Giglachsee during July and August 2010. The air temperature at noon (12:00 midday) is indicated by the straight gray line. The deterioration in weather conditions is indicated by the shaded area

respectively). In summary, these results showed that a distinct change in the bacterial community composition coincided with the short-term cooling period.

Similarity of the bacterial community structure at the OTU level

Accordingly, with changes in community composition, a lower richness and diversity were observed on July 27 and 31, coinciding with the short-term weather change (Table 1). Three groups of samples were identified using NMDS analysis (including OTUs > 0.1% in total reads). Group 1 included OTUs on July 27 and 31, which was during the short-term weather deterioration. Two technical replicates of sequencing (July 31a and 31b) were found to be most similar, indicating that there was good reproducibility. Group 2 included OTUs in the beginning and the end of the study period, i.e., July 19, August 04, and August 18. The OTUs obtained on July 21, August 01, and August 09 formed Group 3 (Fig. S2). ANOSIM revealed marginally significant differences between two groups (Groups 1 and 2, Groups 2 and 3, and Groups 1 and 3; $R = 1$, $P = 0.1$). Indeed, the differences among the three groups were significantly greater than the differences

within groups ($R = 0.89$, $P < 0.01$). The bacterial community structure was compared using Bray–Curtis similarity indices among three dates representing the beginning of the study period (July 19), the cooling period (July 31), and the end of the study period (August 18), respectively. Notably, bacterial community structure on July 19 and August 18 (similarity index = 0.79) was found more similar compared to July 19 and 31 (similarity index = 0.67), implying that the bacterial community structure recovered under nice weather conditions.

In addition to NMDS analysis, the most abundant OTUs (> 1% in total reads; Table S1) were analyzed using Venn diagrams to illustrate which OTUs were shared between samples (15–22 OTUs per sample). Venn diagrams showed that three separate groups were formed (Fig. S3): Group 1 comprised OTUs from July 21, August 01, and August 09 only, sharing 5 OTUs belonging to the orders Acidimicrobiales, Cytophagales, Burkholderiales, and Pseudomonadales. In contrast, Group 2 comprised OTUs from the cooling period (July 27 and 31) only, sharing 9 OTUs belonging to the orders Pseudomonadales, Burkholderiales, Sphingomonadales, Rhizobiales, Synechococcales, Cytophagales, and the families
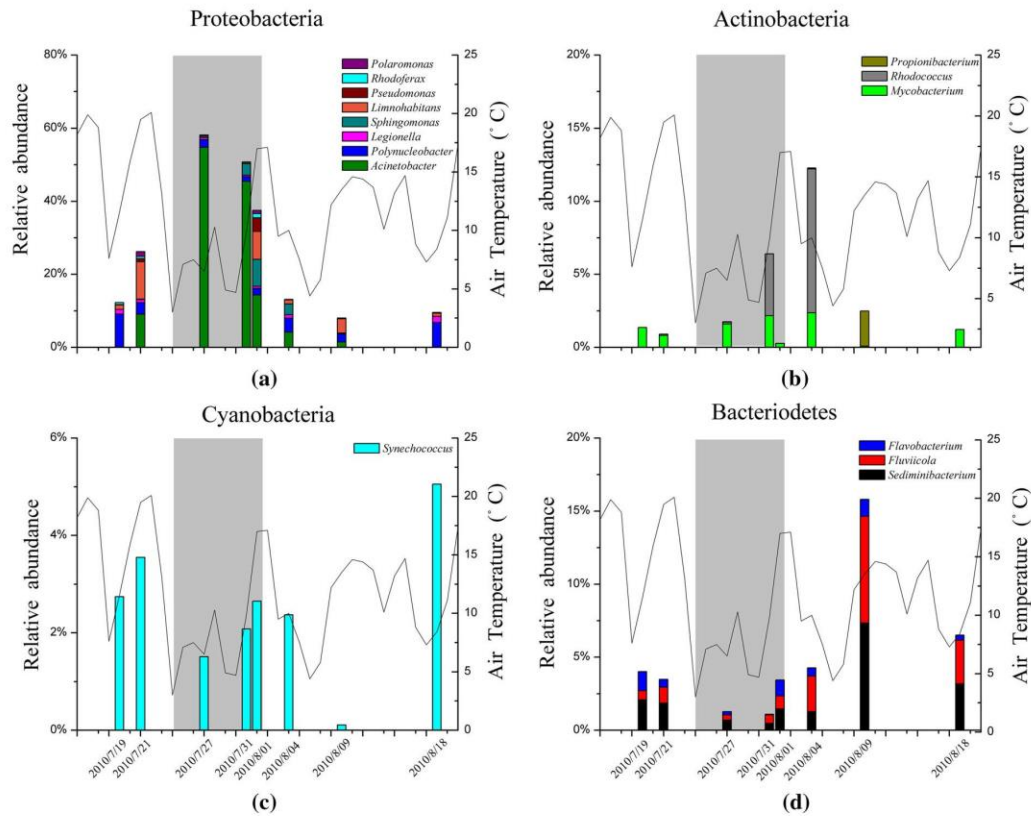
Fig. 4 Relative abundance of abundant bacterial genera from different phyla **a** Proteobacteria, **b** Actinobacteria, **c** Cyanobacteria, and **d** Bacteroidetes in the alpine lake Unterer Giglachsee during July and August 2010. The air temperature at noon (12:00 midday) is indicated by the straight gray line. The deterioration in weather conditions is indicated by the shaded area

Mycobacteriaceae, C111. Group 3 comprised OTUs occurring at the beginning and the end of the study period (July 19, August 04, and August 18), sharing 8 OTUs belonging to the orders Burkholderiales, Sphingomonadales, Cytophagales, and the families Mycobacteriaceae, C111. In summary, the OTUs shared between the sampling dates at the beginning and the end of the study period supported the conclusion that the bacterial community became re-established after the short-term cooling period.

Relationship between the bacterial community composition and meteorological variables

RDA was used to explore the relationship between the bacterial community composition at the OTU level (> 1% in total reads) and meteorological variables.

Using the most abundant taxa (75 OTUs), two variables were identified by forward selection (using time dependence as a covariate). Air temperature and precipitation significantly explained the observed variation in the bacterial community composition at OTU level ($P = 0.012$). The two factors were found to be relatively unrelated to each other as they were plotted roughly orthogonally (Fig. 5). The canonical axis 1 explained 33% of the total variability in the OTU data, implying a distinct gradient of air temperature, whereas the canonical axis 2 explained 15% of the variability related to precipitation. Variance partitioning revealed that the explanatory effect of air temperature only was still significant ($P = 0.026$), although the total variability explained decreased from 53 to 29%. The influence of precipitation was found
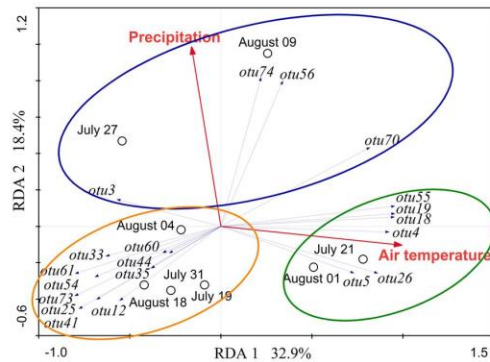
Fig. 5 Redundancy analysis (RDA) of bacterial OTUs (> 1% in total reads) and environmental variables. The two canonical axes explained 47% of the microbial community differentiation, *P* value = 0.012. For clarity, only the twenty most abundant OTUs are shown. Taxa represented by OTUs 1–75 are listed in Table S1

marginally significant (*P* = 0.07), and the total variability explained declined to 34%. It is concluded that both variables represented non-redundant gradients determining bacterial OTU composition. Three groups of OTUs were identified. One group was positively correlated with precipitation, including the genera *Acinetobacter* and *Sediminibacterium* and unknown genera from families C111 and Cerasicoccaceae. The other two groups were correlated negatively or positively with air temperature. One group was related positively to air temperature, including unknown genera of families Comamonadaceae and Cytophagaceae and the genera *Acinetobacter* and *Limnohabitans*. The other group was related negatively to air temperature, including unknown genera from the orders Sphingomonadales, Roseiflexales, the families C111, Cytophagaceae, Comamonadaceae, and Methylobacteriaceae and the genera *Rhodococcus*, *Mycobacterium*, *Synechococcus*, and *Polynucleobacter*. It is concluded that, in response to the short-term cooling period, a rather distinct aquatic bacterial OTU community developed but rapidly diminished when summer conditions got re-established.

Among the most abundant bacteria, *Acinetobacter rhizosphaerae* (OTU No3) was positively correlated with precipitation and negatively correlated with air temperature. Compared to other samples, OTU No3 had a pronounced increase in relative abundance (July 27 at 53.05% and July 31 at 40.47%) during the

cooling period (Table S1). The sequence of OTU No3 was blasted against the NCBI database, and the metadata of the hits indicated that these bacteria originated from soil (Chanika et al., 2011; Kasana 2017) and the plant root system (Jossi 2008; Marasco et al., 2013).

Discussion

Influence of weather deterioration on lake planktonic bacterial community

During the cooling period, there was a notable drop in conductivity and Ca^{2+} concentration, which indicated a terrestrial runoff influence. The Ca^{2+} concentration was reduced by almost one-third within a few days, implying that slow-growing species were washed out due to a lacking competitiveness (Hibbing et al., 2010). In addition, terrestrial runoff can increase organic matter transport and input to the lake (Tranvik & Jansson, 2002; Hongve et al., 2004; Worrall et al., 2018) leading to resource diversification, i.e., organic matter is carried into lakes from terrestrial plants and soils through catchment runoff (Crump et al., 2003). In this study, however, DOC concentration declined during weather deterioration, implying that increased precipitation resulted in lake water dilution rather than enrichment with organic matter from terrestrial runoff. In summary, the cooling period led to a detectable disturbance of the alpine planktonic habitat, which became reversed as soon as the source of disturbance disappeared. It has been argued that every ecosystem reacts to environmental changes in a relatively predictable manner depending on its biological capacity. In general, extreme environments such as arctic/alpine ecosystems are known to be more sensitive to various kinds of disturbance because of various growth-limiting factors. Nevertheless, it has been suggested that disturbance and climate change might interact, e.g., early successional ecosystems may be more sensitive to climate change influence compared to later successional states, thus resulting in state shifts only when disturbed (Kröel-Dulay et al., 2015). Aquatic ecosystems have long been studied for regime and state shifts induced by non-linear ecosystem behavior in relatively short periods (Scheffer and Carpenter 2003). Such regime shifts would have significant consequences on the ecosystem level,

e.g., high algal biomass production during summer and oxygen consumption during the ice cover period resulting in a cascade of changes in the whole ecosystem. Modeling experiments have revealed that such regime shifts can be foreseen already through long-term monitoring by statistical anomalies, i.e., through the increased variance in residues from dependent variables in linear regression models (Seekell et al., 2011). In other words, weather-induced disturbances might actually increase the variance in dependent variables to a potential extent, which possibly increases the likelihood of state shifts. In a related study, runoff events induced by precipitation were compared in five alpine lakes (Jiang et al., 2019). Notably, the evidence that richness or diversity may be influenced by runoff through rainfall or snow melting during two summer periods in 2010 and 2011 was found relatively low, implying that the pronounced cooling period observed in this study was rather the exception than the rule. In summary, a potential interaction between weather-induced disturbances and climate change effects should be considered, in particular for alpine systems known to be more sensitive to temperature rise effects than lowland aquatic ecosystems.

Changes in the bacterial community composition at the phylum level

In general, five phyla have been frequently reported in lakes, including Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, and Verrucomicrobia (Newton et al., 2011). In this study, the two main classes of Proteobacteria, namely Betaproteobacteria and Gammaproteobacteria alternated in relative abundance with dominance of Gammaproteobacteria during the cooling phase. Only a few studies showed that Gammaproteobacteria occurred in high abundance in mountain lakes (Power et al., 2005), although this class usually was found to be dominant under deepwater conditions only (Bel'kova et al., 1996). Some authors also concluded that this group may be transiently washed into lakes with surface runoff (Lindström & Leskinen, 2002). In our study, the high relative abundance of Gammaproteobacteria during the cooling period was possibly caused by the rapid growth of *A. rhizosphaerae*. In general, Betaproteobacteria are typically found to be dominant in the euphotic zone of a lake (Bel'kova et al., 1996) and are

known to respond quickly to changes in nutrient availability (Hornák et al., 2006; Posch et al., 2007; Nelson, 2009). When nutrient concentrations are high, Betaproteobacteria tend to be fast growing (Newton et al., 2011). In our study, after the cooling period, Betaproteobacteria rapidly regained dominance probably because of their fast growth rate under more favorable DOC concentrations. In addition, the relative abundance of Cyanobacteria during the entire study period was relatively low. This low relative abundance might be a result of the overall oligotrophic conditions as higher abundance of Cyanobacteria typically occurs in more eutrophic freshwater (Eiler & Bertilsson, 2004). In addition, Cyanobacteria are sometimes favored by increased temperature (Paerl & Huisman, 2008). Similar to Betaproteobacteria, the growth of Bacteroidetes is also related to organic matter resulting from a phytoplankton bloom (Newton et al., 2011) and frequently occurs during periods with high external DOC loading and algae-derived DOC production (Eiler & Bertilsson, 2004; Kolmonen et al., 2004). Accordingly, in our study, the increased relative abundance of Bacteroidetes coincided with the period of higher DOC and higher Cyanobacteria relative abundance. Actinobacteria are ubiquitous in freshwater lakes (Haukka et al., 2006; Humbert et al., 2009) but favor less eutrophic conditions (Haukka et al., 2006). Previous studies confirmed that their abundance decreased with decreasing oxygen concentration (Taipale et al., 2009). As Actinobacteria are typically small and slow-growing, they might have been disfavored by the cooling period in our study. The relative abundance of Verrucomicrobia was generally low in all samples, which can be explained by their overall dependence on more eutrophic conditions (Newton et al., 2011). Other phyla occurred in lowest relative abundance, which might be caused by specific local conditions, including lake type, local weather conditions, nutrient availability, physical and chemical properties, catchment characteristics, and others (Crump et al., 2003; Corno, 2006; Bertilsson et al., 2007; Corno et al., 2009).

Short-term succession of the bacterial community related to the cooling period

It was obvious that *Acinetobacter*, affiliated to Proteobacteria, had a successional development along with temperature decrease. Previous studies have

reported that many members of *Acinetobacter*, isolated from oligotrophic conditions, also had a higher tolerance to low temperatures and can grow well in cold environments (Huang et al., 2013; Yao et al., 2013). In our study, *A. rhizosphaerae* was frequently identified during the cooling period and probably originated from rhizospheric soil (Kasana, 2017). The known higher growth rate of *A. rhizosphaerae* under low-temperature conditions (3.0°C–10.3°C) might have supported the net bacterial growth even under higher flow-through conditions as indicated by the dilution of Ca^{2+} and DOC concentrations. However, another genus, *Limnohabitans*, which is also affiliated to Proteobacteria, is known to have a high growth and substrate uptake rate as well as a high mortality rate (Kasalický et al., 2010). It was found to be positively related to air temperature. During the cooling period, the decrease in water temperature and DOC concentration might have actually reduced the adaptability and competitiveness of *Limnohabitans* compared to other bacteria. The relative rapid increase in their relative abundance may have been due to their rapid absorption of nutrients. *Synechococcus* was the only genus affiliated to Cyanobacteria. Typically, it is found to be abundant in oligotrophic environments under well-illuminated conditions in the euphotic zone (Ruber et al. 2016). The relatively low growth rates might have been exceeded by higher flow-through as observed during the cooling period. *Sediminibacterium*, affiliated to Bacteroidetes, has been isolated from sediment (Qu & Yuan, 2008), soil (Kim et al., 2013), and activated sludge (Ayarza et al., 2014). Nutrient supply and temperature were most closely related to *Sediminibacterium* net production and growth (Sander & Kalff, 1993). Compared to the cooling period, the relative abundance of *Sediminibacterium* was higher under warmer weather conditions and higher DOC concentrations. It is concluded that the two meteorological factors, air temperature and precipitation, were directly and indirectly related to the observed change in the bacterial community composition.

Conclusion

In this study, the bacterial community composition changed substantially during a short-term deterioration in weather conditions as revealed at the taxonomic level of phylum, class, and genus. This significant change seemed to be caused directly and indirectly by the decreased air temperature and increased precipitation, resulting in the cooling of lake water by terrestrial runoff and increased flow-through. Notably, the planktonic bacterial community structure returned to the previous state, implying a re-installation of the original environmental conditions. We conclude that, in comparison with lowland lakes, changes in weather conditions during the summer growing season can have a more direct impact on planktonic bacterial community structure. Thus, the change in weather conditions can increase variability in the planktonic bacterial community structure and can overrule other more constant factors such as nutrient availability. Moreover, the reestablishment of the bacterial planktonic community structure to a previous state at the start of the study period implies rather robust planktonic ecological conditions.

References

Allgaier, M. & H.-P. Grossart, 2006. Seasonal dynamics and phylogenetic diversity of free-living and particle-associated bacterial communities in four lakes in northeastern Germany. Aquatic Microbial Ecology 45: 115–128.

Allison, S. & J. Martiny, 2008. Resistance, resilience, and redundancy in microbial communities. In Avise, J. C., S. Hubbell & F. J. Ayala (eds), In the Light of Evolution: Volume II: Biodiversity and Extinction. National Academies Press (US), Washington, DC: 149–165.

Ayarza, J. M., E. L. Figuerola & L. Erijman, 2014. Draft genome sequences of type strain *Sediminibacterium salmoneum* NJ-44 and *Sediminibacterium* sp. strain C3, a novel strain isolated from activated sludge. Genome Announcements 2: e01073-13.

Battarbee, R. W., J.-A. Grytnes, R. Thompson, P. G. Appleby, J. Catalan, A. Korhola, H. Birks, E. Heegaard & A. Lami, 2002. Comparing palaeolimnological and instrumental evidence of climate change for remote mountain lakes over the last 200 years. Journal of Paleolimnology 28: 161–179.

Bel'kova, N., L. Denisova, E. Manakova, E. Zaĭchikov & M. Grachev, 1996. Species diversity of deep water microorganisms in Lake Baikal, detected by 16S rRNA sequences. Doklady Akademii nauk 348: 692.

Bertilsson, S., A. Eiler, A. Nordqvist & N. O. G. Jørgensen, 2007. Links between bacterial production, amino-acid utilization and community composition in productive lakes. The ISME Journal 1: 532.

Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen & R. Knight, 2009. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26: 266–267.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich & J. I. Gordon, 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335.

Castro, H. F., A. T. Classen, E. E. Austin, R. J. Norby & C. W. Schadt, 2010. Soil microbial community responses to multiple experimental climate change drivers. Applied and Environmental Microbiology 76: 999–1007.

Chanika, E., D. Georgiadou, E. Soueref, P. Karas, E. Karanasios, N. G. Tsiropoulos, E. A. Tzortzakakis & D. G. Karpouzas, 2011. Isolation of soil bacteria able to hydrolyze both organophosphate and carbamate pesticides. Bioresource technology 102(3): 3184–3192.

Comte, J., S. Jacquet, S. Viboud, D. Fontvieille, A. Millery, G. Paolini & I. Domaizon, 2006. Microbial community structure and dynamics in the largest natural French lake (Lake Bourget). Microbial Ecology 52: 72–89.

Corno, G., 2006. Effects of nutrient availability and *Ochromonas* sp. predation on size and composition of a simplified aquatic bacterial community. FEMS Microbiology Ecology 58: 354–363.

Corno, G., B. E. Modenutti, C. Callieri, E. G. Balseiro, R. Bertoni & E. Caravatia, 2009. Bacterial diversity and morphology in deep ultraoligotrophic Andean lakes: the role of UVR on vertical distribution. Limnology and Oceanography 54: 1098–1112.

Crump, B. C., G. W. Kling, M. Bahr & J. E. Hobbie, 2003. Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. Applied and Environmental Microbiology 69: 2253–2268.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu & G. L. Andersen, 2006. Greengenes, a chimera-checked 16S

rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72: 5069–5072.

Downing, A. L., K. M. DeVanna, C. N. Rubeck-Schurtz, L. Tuhela & H. Grunkemeyer, 2008. Community and ecosystem responses to a pulsed pesticide disturbance in freshwater ecosystems. Ecotoxicology 17(6): 539–548.

Edgar, R. C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461.

Eiler, A. & S. Bertilsson, 2004. Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environmental Microbiology 6: 1228–1243.

Haukka, K., E. Kolmonen, R. Hyder, J. Hietala, K. Vakkilainen, T. Kairesalo, H. Haario & K. Sivonen, 2006. Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial Ecology 51: 137–146.

Heiri, O. & A. F. Lotter, 2003. 9000 years of chironomid assemblage dynamics in an Alpine lake: long-term trends, sensitivity to disturbance, and resilience of the fauna. Journal of Paleolimnology 30(3): 273–289.

Hibbing, M. E., C. Fuqua, M. R. Parsek & S. B. Peterson, 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology 8: 15.

Hongve, D., G. Riise & J. F. Kristiansen, 2004. Increased colour and organic acid concentrations in Norwegian forest lakes and drinking water–a result of increased precipitation? Aquatic Sciences 66: 231–238.

Hornák, K., J. Jezbera, J. Nedoma, J. M. Gasol & K. Simek, 2006. Effects of resource availability and bacterivory on leucine incorporation in different groups of freshwater bacterioplankton, assessed using microautoradiography. Aquatic Microbial Ecology 45: 277–289.

Huang, X., W. Li, D. Zhang & W. Qin, 2013. Ammonium removal by a novel oligotrophic *Acinetobacter* sp. Y16 capable of heterotrophic nitrification–aerobic denitrification at low temperature. Bioresource Technology 146: 44–50.

Humbert, J. F., U. Dorigo, P. Cecchi, B. Le Berre, D. Debroas & M. Bouvy, 2009. Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. Environmental Microbiology 11: 2339–2350.

Huse, S. M., L. Dethlefsen, J. A. Huber, D. M. Welch, D. A. Relman & M. L. Sogin, 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genetics 4: e1000255.

Jiang, Y., H. Huang, T. Ma, J. Ru, S. Blank, R. Kurmayer & L. Deng, 2019. Temperature response of planktonic microbiota in remote alpine lakes. Frontiers in microbiology 10: 1714.

Jones, S. E., C.-Y. Chiu, T. K. Kratz, J.-T. Wu, A. Shade & K. D. McMahon, 2008. Typhoons initiate predictable change in aquatic bacterial communities. Limnology and Oceanography 53(4): 1319–1326.

Jossi, M., 2008. Rhizosphere bacterial communities associated with Lolium perenne. Université de Neuchâtel.

Kasalický, V., J. Jezbera, K. Šimek & M. W. Hahn, 2010. *Limnohabitans planktonicus* sp. nov. and *Limnohabitans parvus* sp. nov., planktonic betaproteobacteria isolated from a freshwater reservoir, and emended description of

the genus Limnohabitans. International Journal of Systematic and Evolutionary Microbiology 60: 2710–2714.

Kasana, R. C., 2017. Bacterial Diversity in Cold Environments of Indian Himalayas Mining of Microbial Wealth and MetaGenomics. Springer, Berlin: 83–99.

Kim, Y.-J., N.-L. Nguyen, H.-Y. Weon & D.-C. Yang, 2013. *Sediminibacterium ginsengisoli* sp. nov., isolated from soil of a ginseng field, and emended descriptions of the genus *Sediminibacterium* and of *Sediminibacterium salmoneum*. International Journal of Systematic and Evolutionary Microbiology 63: 905–912.

Kolmonen, E., K. Sivonen, J. Rapala & K. Haukka, 2004. Diversity of cyanobacteria and heterotrophic bacteria in cyanobacterial blooms in Lake Joutikas, Finland. Aquatic Microbial Ecology 36: 201–211.

Kröel-Dulay, G., J. Ransijn, I. K. Schmidt, C. Beier, P. De Angelis, G. De Dato, J. S. Dukes, B. Emmett, M. Estiarte & J. Garadnai, 2015. Increased sensitivity to climate change in disturbed ecosystems. Nature Communications 6: 6682.

Lepš, J. & P. Šmilauer, 2003. Multivariate Analysis of Ecological Data Using CANOCO. Cambridge University Press, Cambridge.

Lindström, E. & E. Leskinen, 2002. Do neighboring lakes share common taxa of bacterioplankton? Comparison of 16S rDNA fingerprints and sequences from three geographic regions. Microbial Ecology 44: 1–9.

Lindström, E. S., M. P. Kamst-Van Agterveld & G. Zwart, 2005. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. Applied and Environmental Microbiology 71(12): 8201–8206.

Magnea, U., R. Sciascia, F. Paparella, R. Tiberti & A. Provenzale, 2013. A model for high-altitude alpine lake ecosystems and the effect of introduced fish. Ecological Modelling 251: 211–220.

Marasco, R., E. Rolli, M. Fusi, A. Cherif, A. Abou-Hadid, U. El-Bahairy, S. Borin, C. Sorlini & D. Daffonchio, 2013. Plant growth promotion potential is equally represented in diverse grapevine root-associated bacterial communities from different biopedoclimatic environments. BioMed research international. https://doi.org/10.1155/2013/491091.

Nelson, C. E., 2009. Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. The ISME Journal 3: 13.

Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon & S. Bertilsson, 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75: 14–49.

O'Brien, W. J., M. Bahr, A. E. Hershey, J. E. Hobbie, G. W. Kipphut, G. W. Kling, H. Kling, M. McDonald, M. C. Miller & P. Rublee, 1997. The Limnology of Toolik Lake Freshwaters of Alaska. Springer, Berlin: 61–106.

Paerl, H. W. & J. Huisman, 2008. Blooms like it hot. Science 320: 57–58.

Posch, T., B. Mindl, K. Horňák, J. Jezbera, M. M. Salcher, B. Sattler, B. Sonntag, J. Vrba & K. Šimek, 2007. Biomass reallocation within freshwater bacterioplankton induced by manipulating phosphorus availability and grazing. Aquatic Microbial Ecology 49: 223–232.

Power, M. L., J. Littlefield-Wyer, D. M. Gordon, D. A. Veal & M. B. Slade, 2005. Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. Environmental Microbiology 7: 631–640.

Qu, J.-H. & H.-L. Yuan, 2008. *Sediminibacterium salmoneum* gen. nov., sp. nov., a member of the phylum Bacteroidetes isolated from sediment of a eutrophic reservoir. International Journal of Systematic and Evolutionary Microbiology 58: 2191–2194.

Rasconi, S., K. Winter & M. J. Kainz, 2017. Temperature increase and fluctuation induce phytoplankton biodiversity loss–Evidence from a multi-seasonal mesocosm experiment. Ecology and Evolution 7: 2936–2946.

Ren, L., D. He, J. Zeng & Q. L. Wu, 2013. Bacterioplankton communities turn unstable and become small under increased temperature and nutrient-enriched conditions. FEMS Microbiology Ecology 84: 614–624.

Rofner, C., H. Peter, N. Catalán, F. Drewes, R. Sommaruga & M. T. Pérez, 2017. Climate-related changes of soil characteristics affect bacterial community composition and function of high altitude and latitude lakes. Global Change Biology 23: 2331–2344.

Ruber, J., F. R. Bauer, A. D. Millard, U. Raeder, J. Geist & K. Zwirglmaier, 2016. Synechococcus diversity along a trophic gradient in the Osterseen Lake District, Bavaria. Microbiology 162(12): 2053–2063.

Salcher, M. M., J. Pernthaler & T. Posch, 2010. Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. Limnology and Oceanography 55: 846–856.

Sander, B. C. & J. Kalff, 1993. Factors controlling bacterial production in marine and freshwater sediments. Microbial Ecology 26: 79–99.

Scheffer, M. & S. R. Carpenter, 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in Ecology & Evolution 18(12): 648–656.

Seekell, D. A., S. R. Carpenter & M. L. Pace, 2011. Conditional heteroscedasticity as a leading indicator of ecological regime shifts. The American Naturalist 178(4): 442–451.

Shade, A., J. S. Read, D. G. Welkie, T. K. Kratz, C. H. Wu & K. D. McMahon, 2011. Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. Environmental Microbiology 13(10): 2752–2767.

Shade, A., H. Peter, S. Allison, D. Baho, M. Berga, H. Buergmann, D. Huber, S. Langenheder, J. Lennon, J. Martiny, K. Matulich, T. Schmidt & J. Handelsman, 2012. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology 3: 417.

Stepanauskas, R., M. A. Moran, B. A. Bergamaschi & J. T. Hollibaugh, 2003. Covariance of bacterioplankton composition and environmental variables in a temperate delta system. Aquatic Microbial Ecology 31: 85–98.

Taipale, S., R. I. Jones & M. Tiirola, 2009. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquatic Microbial Ecology 55: 1–16.

Thompson, R., C. Kamenik & R. Schmidt, 2005. Ultra-sensitive Alpine lakes and climate change. Journal of Limnology 64(2): 139–152.

Tranvik, L. & M. Jansson, 2002. Climate change (Communication arising): terrestrial export of organic carbon. Nature 415: 861.

Van Der Gucht, K., K. Sabbe, L. De Meester, N. Vloemans, G. Zwart, M. Gillis & W. Vyverman, 2001. Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. Environmental Microbiology 3: 680–690.

Weckström, K., J. Weckström, K. Huber, C. Kamenik, R. Schmidt, W. Salvenmoser, M. Rieradevall, T. Weisse, R. Psenner & R. Kurmayer, 2016. Impacts of climate warming on Alpine lake biota over the past decade. Arctic, Antarctic, and Alpine Research 48: 361–376.

Whalen, S. & J. Cornwell, 1985. Nitrogen, phosphorus, and organic carbon cycling in an arctic lake. Canadian Journal of Fisheries and Aquatic Sciences 42: 797–808.

Worrall, F., N. J. Howden, T. P. Burt & R. Bartlett, 2018. Declines in the dissolved organic carbon (DOC) concentration and flux from the UK. Journal of Hydrology 556: 775–789.

Yadav, A. N., S. G. Sachan, P. Verma, R. Kaushik & A. K. Saxena, 2016. Cold active hydrolytic enzymes production by psychrotrophic Bacilli isolated from three subglacial lakes of NW Indian Himalayas. Journal of Basic Microbiology 56(3): 294–307.

Yannarell, A. C. & E. W. Triplett, 2004. Within-and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. Applied and Environmental Microbiology 70: 214–223.

Yao, S., J. Ni, T. Ma & C. Li, 2013. Heterotrophic nitrification and aerobic denitrification at low temperature by a newly isolated bacterium, Acinetobacter sp. HA2. Bioresource Technology 139: 80–86.

Supplementary data to the Article:

Contents

Table S1: The most abundant OTUs in different samples

Figure S1: Meteorological variables and water temperature recorded in alpine lake Unterer Giglachsee during July and August 2010. Meteorological data were used from relevant meteorological stations, including Obertauern station (OBE, 1772m), Rudolfshütte station (RUD, 2317m) and Schmittenhoche station (SCH, 1956m). For air temperature, we chose the data at noon (12:00 am) since it was found most indicative of weather change. For water temperature, the readings at noon have been used (12:00 am).

Figure S2: NMDS analysis of bacterial OTUs (>0.1% in total reads) recorded from different sampling dates in the alpine lake Unterer Giglachsee during July and August 2010. Jul31a + Jul31b indicate technical replicates. A stress value of <0.05 indicates overall statistically significant difference.

Figure S3: Venn diagrams showing the number of bacterial OTUs shared between sampling dates (only OTUs ≥1% in total reads) were analyzed.

Table S1

Table S1 The most abundant OTUs (>1%) in different samples

#OTUID	P Jul19	P Jul21	P Jul27	P Jul31	P Aug01	P Aug04	P Aug09	P Aug18	taxonomy
otu1	0.00%	0.00%	0.00%	0.00%	1.48%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu2	0.00%	0.00%	0.00%	0.00%	1.48%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu3	0.00%	0.00%	53.05%	40.47%	0.00%	3.62%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__rhizosphaerae
otu4	0.00%	3.98%	0.00%	0.00%	2.03%	0.00%	1.08%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__
otu5	0.00%	4.31%	0.00%	0.00%	8.82%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__
otu6	0.00%	0.00%	0.00%	1.28%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__
otu7	0.00%	0.00%	0.00%	3.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter
otu8	1.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.47%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Legionellales; f__Legionellaceae; g__Legionella; s__
otu9	2.27%	0.00%	0.00%	0.00%	1.05%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Legionellales; f__Legionellaceae; g__
otu10	0.00%	2.08%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae
otu11	7.09%	0.00%	1.19%	0.00%	0.00%	0.00%	0.00%	5.16%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas; s__
otu12	0.00%	1.85%	0.00%	0.00%	0.00%	3.25%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Polymuckobacter; s__
otu13	1.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Polymuckobacter; s__
otu14	0.00%	2.41%	0.00%	0.00%	0.00%	0.00%	1.45%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Polymuckobacter; s__
otu15	0.00%	0.00%	0.00%	1.12%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__; s__
otu16	0.00%	0.00%	0.00%	0.00%	1.22%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Enhydrobacter; s__
otu17	0.00%	1.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Polaromonas; s__
otu18	0.00%	3.56%	0.00%	0.00%	2.57%	0.00%	1.63%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__
otu19	0.00%	3.36%	0.00%	0.00%	2.32%	0.00%	1.42%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__
otu20	0.00%	2.85%	0.00%	0.00%	2.32%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__
otu21	0.00%	0.00%	0.00%	0.00%	1.52%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu22	3.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
otu23	2.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.26%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
otu24	0.00%	1.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
otu25	23.00%	0.00%	8.98%	8.47%	0.00%	18.58%	0.00%	17.94%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
otu26	0.00%	7.77%	0.00%	0.00%	5.70%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
otu27	2.81%	0.00%	0.00%	0.00%	0.00%	1.84%	0.00%	1.30%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Alcaligenaceae; g__; s__
otu28	0.00%	1.53%	0.00%	0.00%	1.05%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Alcaligenaceae; g__; s__
otu29	0.00%	0.00%	0.00%	0.00%	4.68%	0.00%	0.00%	0.00%	k__Bacteria; p__Verrucomicrobia; c__[Methylacidiphilae]; o__[Methylacidiphilales]; f__LD19; g__; s__
otu30	0.00%	0.00%	0.00%	2.89%	0.00%	2.50%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Sphingomonas; s__
otu31	0.00%	0.00%	0.00%	0.00%	1.48%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu32	0.00%	2.44%	0.00%	0.00%	3.50%	0.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__
otu33	1.11%	0.00%	1.45%	1.60%	0.00%	3.83%	0.00%	1.17%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__
otu34	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.77%	0.00%	k__Bacteria; p__Verrucomicrobia; c__[Methylacidiphilae]; o__[Methylacidiphilales]; f__LD19; g__; s__
otu35	0.00%	0.00%	1.12%	5.58%	0.00%	1.32%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Methylobacteriaceae; g__; s__

OTU								Taxonomy
otu36	1.04%	0.00%	0.00%	1.36%	0.00%	2.02%	0.00%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__
otu37	0.00%	2.24%	0.00%	0.00%	2.57%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__
otu38	0.00%	0.00%	1.27%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__[Exiguobacteraceae]; g__Exiguobacterium; s__
otu39	0.00%	3.82%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__[Exiguobacteraceae]; g__Exiguobacterium; s__
otu40	0.00%	3.43%	0.00%	0.00%	2.53%	0.00%	0.00%	k__Bacteria; p__Cyanobacteria; c__Synechococcophycideae; o__Synechococcales; f__Synechococcaceae; g__Synechococcus; s__
otu41	2.56%	0.00%	1.47%	2.05%	0.00%	2.31%	4.12%	k__Bacteria; p__Cyanobacteria; c__Synechococcophycideae; o__Synechococcales; f__Synechococcaceae; g__Synechococcus; s__
otu42	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.67%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu43	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.06%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas; s__maltophilia
otu44	3.60%	0.00%	1.15%	0.00%	0.00%	2.35%	0.00%	k__Bacteria; p__Chloroflexi; c__Chloroflexi; o__[Roseiflexales]; f__; g__; s__
otu45	0.00%	2.08%	0.00%	0.00%	1.48%	0.00%	0.00%	k__Bacteria; p__Chloroflexi; c__Chloroflexi; o__[Roseiflexales]; f__; g__; s__
otu46	1.47%	0.00%	0.00%	0.00%	0.00%	0.00%	3.42%	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__
otu47	0.00%	0.00%	0.00%	0.00%	0.00%	1.59%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__; s__
otu48	0.00%	0.00%	0.00%	0.00%	0.00%	1.25%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__rhizosphaerae
otu49	0.00%	0.00%	0.00%	0.00%	0.00%	1.07%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__johnsonii
otu50	1.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__
otu51	0.00%	0.00%	0.00%	0.00%	0.00%	5.60%	0.00%	k__Bacteria; p__Proteobacteria; c__TA18; o__PHOS-HD29; f__; g__; s__
otu52	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.39%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__; g__; s__
otu53	0.00%	0.00%	0.00%	0.00%	2.00%	0.00%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu54	11.09%	0.00%	6.31%	3.98%	0.00%	9.72%	5.68%	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
otu55	0.00%	14.20%	0.00%	9.32%	0.00%	7.05%	0.00%	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
otu56	0.00%	1.02%	0.00%	0.00%	0.00%	5.91%	0.00%	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__Sediminibacterium; s__
otu57	1.42%	0.00%	0.00%	0.00%	0.00%	0.00%	2.04%	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__Sediminibacterium; s__
otu58	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.39%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu59	0.00%	0.00%	0.00%	0.00%	0.00%	2.32%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu60	0.00%	0.00%	0.00%	4.12%	0.00%	9.18%	0.00%	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardiaceae; g__Rhodococcus; s__fascians
otu61	1.27%	0.00%	1.57%	2.14%	0.00%	2.31%	1.08%	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Mycobacteriaceae; g__Mycobacterium; s__
otu62	0.00%	0.00%	0.00%	0.00%	0.00%	1.84%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu63	1.36%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__ACK-M1; g__; s__
otu64	0.00%	0.00%	0.00%	0.00%	0.00%	1.50%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Enhydrobacter; s__
otu65	0.00%	0.00%	0.00%	0.00%	0.00%	1.33%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__rhizosphaerae
otu66	0.00%	1.20%	0.00%	0.00%	0.00%	3.79%	0.00%	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__ACK-M1; g__; s__
otu67	0.00%	0.00%	0.00%	0.00%	0.00%	1.89%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu68	0.00%	0.00%	0.00%	0.00%	0.00%	2.21%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu69	0.00%	0.00%	0.00%	0.00%	0.00%	2.05%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu70	0.00%	7.21%	0.00%	2.07%	0.00%	6.52%	0.00%	k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__
otu71	0.00%	0.00%	0.00%	0.00%	0.00%	2.47%	0.00%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Thiotrichales; f__; g__; s__
otu72	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.25%	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
otu73	5.16%	0.00%	3.94%	3.60%	0.00%	1.49%	9.10%	k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__
otu74	0.00%	0.00%	0.00%	0.00%	0.00%	5.54%	8.14%	k__Bacteria; p__Verrucomicrobia; c__Opitutae; o__[Cerasicoccales]; f__[Cerasicoccaceae]; g__; s__
otu75	0.00%	0.00%	2.05%	0.00%	0.00%	2.76%	0.00%	k__Bacteria; p__[Thermi]; c__Deinococci; o__Deinococcales; f__Deinococcaceae; g__Deinococcus; s__

Figure S1
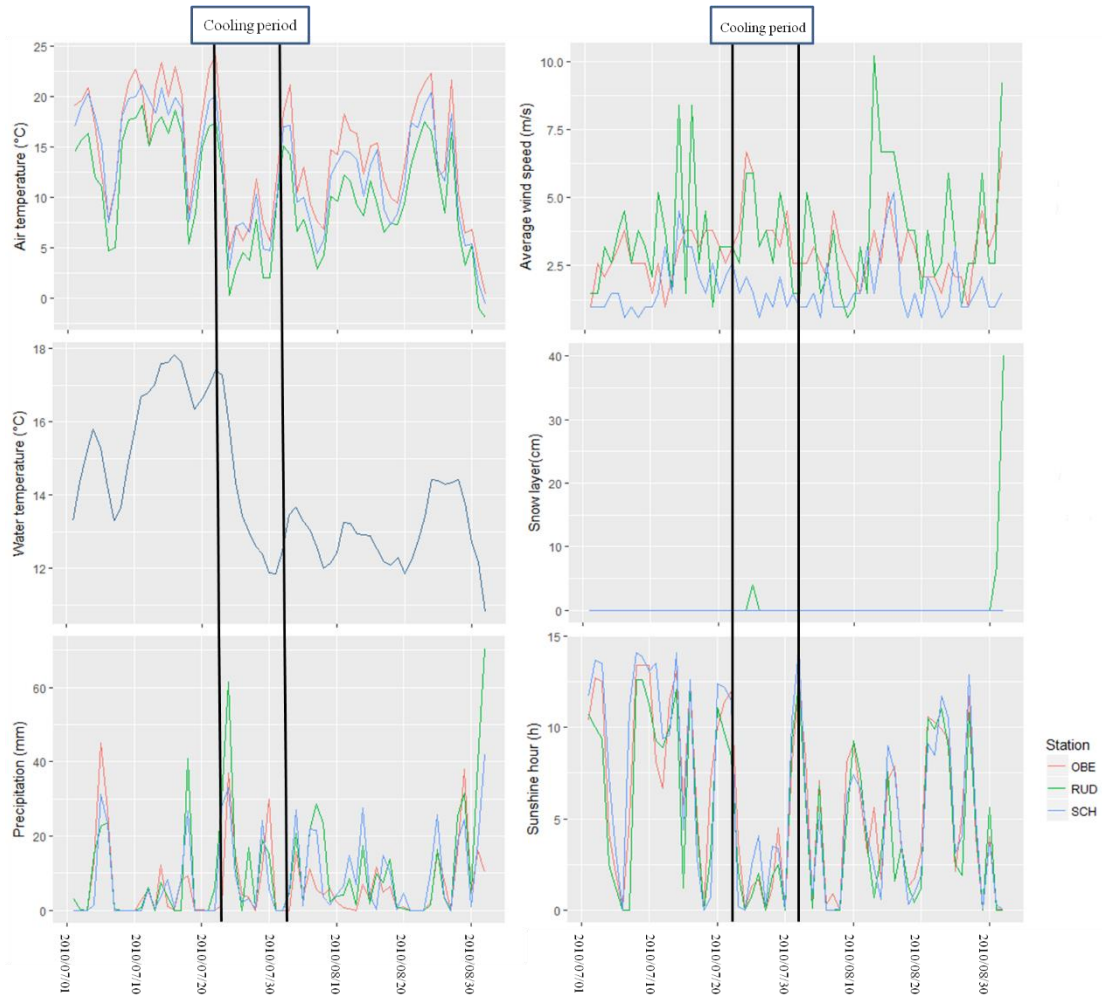
Figure S2
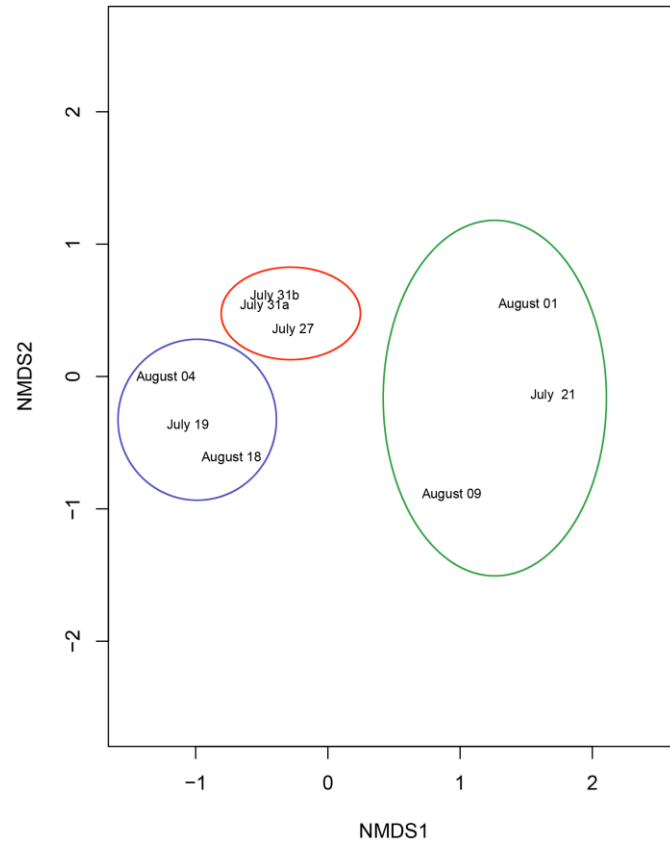
Figure S3
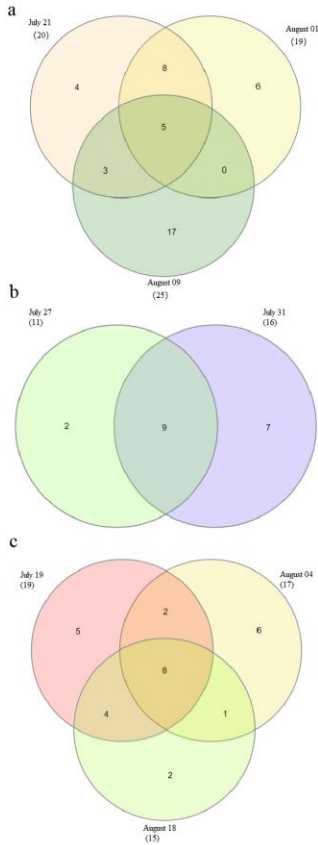
a: Shared OTUs among Jul21, Aug01, and Aug09 (n = 5)
k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__
k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Limnohabitans; s__
k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__

b: Shared OTUs between Jul27 and Jul31 (n = 9)
k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__rhizosphaerae
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Polynucleobacter; s__
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__
k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Methylobacteriaceae; g__; s__
k__Bacteria; p__Cyanobacteria; c__Synechococcophycideae; o__Synechococcales; f__Synechococcaceae; g__Synechococcus; s__
k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Mycobacteriaceae; g__Mycobacterium; s__
k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__

c: Shared OTUs among Jul19, Aug04 and Aug18 (n = 8)
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__Polynucleobacter; s__
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Alcaligenaceae; g__; s__
k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__; g__; s__
k__Bacteria; p__Cyanobacteria; c__Synechococcophycideae; o__Synechococcales; f__Synechococcaceae; g__Synechococcus; s__
k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Mycobacteriaceae; g__Mycobacterium; s__
k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__

Article

Differences in Gut Virome Related to Barrett Esophagus and Esophageal Adenocarcinoma

Tianli Ma [1,2,†], Jinlong Ru [1,2,†], Jinling Xue [1,2], Sarah Schulz [1,2], Mohammadali Khan Mirzaei [1,2], Klaus-Peter Janssen [3], Michael Quante [4,5,*] and Li Deng [1,2,*]

[1] Helmholtz Centre Munich—German Research Center for Environmental Health, Institute of Virology, 85764 Neuherberg, Germany; tianli.ma@helmholtz-muenchen.de (T.M.); jinlong.ru@helmholtz-muenchen.de (J.R.); jinling.xue@helmholtz-muenchen.de (J.X.); sarah.schulz@helmholtz-muenchen.de (S.S.); m.khanmirzaei@helmholtz-muenchen.de (M.K.M.)
[2] Institute of Virology, Technical University of Munich, 81675 Munich, Germany
[3] Department of Surgery, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany; klaus-peter.janssen@tum.de
[4] II. Medizinische Klinik, Klinikum Rechts der Isar, Technische Universität München, 81675 Munich, Germany
[5] Innere Medizin II, Universitätsklinik Freiburg, Universität Freiburg, 79106 Freiburg, Germany
*   Correspondence: michael.quante@uniklinik-freiburg.de (M.Q.); li.deng@helmholtz-muenchen.de (L.D.)
†   These authors contributed equally to this work.

Abstract: The relationship between viruses (dominated by bacteriophages or phages) and lower gastrointestinal (GI) tract diseases has been investigated, whereas the relationship between gut bacteriophages and upper GI tract diseases, such as esophageal diseases, which mainly include Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC), remains poorly described. This study aimed to reveal the gut bacteriophage community and their behavior in the progression of esophageal diseases. In total, we analyzed the gut phage community of sixteen samples from patients with esophageal diseases (six BE patients and four EAC patients) as well as six healthy controls. Differences were found in the community composition of abundant and rare bacteriophages among three groups. In addition, the auxiliary metabolic genes (AMGs) related to bacterial exotoxin and virulence factors such as lipopolysaccharides (LPS) biosynthesis proteins were found to be more abundant in the genome of rare phages from BE and EAC samples compared to the controls. These results suggest that the community composition of gut phages and functional traits encoded by them were different in two stages of esophageal diseases. However, the findings from this study need to be validated with larger sample sizes in the future.

1. Introduction

Barrett's esophagus (BE) is the only known precursor for the development of esophageal adenocarcinoma (EAC) with a five-year survival rate of less than 20%. The incidence of these diseases is on the rise globally [1,2]. Early diagnosis of patients at risk could prevent the progression of BE to EAC, and effectively reduce the development of EAC. However, as only 0.3–0.5% of BE patients develop EAC, endoscopic biopsy surveillance, while linked to higher survival rates, is only recommended for at-risk patients [3]. In addition, endoscopies are often discomforting, and sometimes lead to inconclusive results [4]. Thus, noninvasive diagnostics with higher accuracy are sought after. The human gut is home to trillions of microorganisms, including bacteria, viruses, fungi, and protozoa. These microorganisms and their human host maintain a symbiotic relationship, in which the host provides a nutrient-rich habitat, and the microbiota supplies key metabolic capabilities, protects against pathogen invasion, and trains the immune system [5]. In addition, an imbalance in gut microbiota, termed dysbiosis, is associated with several human diseases or conditions,

including inflammatory bowel disease (IBD), and colorectal cancer (CRC). These microbial communities have shown disease-specific community structure, suggesting that they can be used as signatures for diagnosing some dysbiosis-associated diseases [6–9].

Both BE and EAC biopsy samples have been found to harbor a unique bacterial community. Compared to the normal esophagus, Gram-positive bacteria (Firmicutes) were gradually replaced by Gram-negative bacteria (Bacteroidetes, Proteobacteria, Fusobacteria, and Spirochaetes) in BE [10]. As the disease progressed from BE to EAC, the Gram-negative bacteria *Escherichia coli* (*E.coli*) and *Fusobacterium nucleatum* became more dominant [11]. These changes are important as LPS, the outer membrane component of Gram-negative bacteria, could promote the secretion of pro-inflammatory cytokines through activating the Toll-like receptor (TLR) and the downstream NF- κB pathway in different cell types, contributing to the severity of esophageal diseases [11]. In human and mice models with BE, elevated levels of pro-inflammatory cytokines and activated TLR were observed in the gastroesophageal junction [12]. The resulting chronic inflammation could induce systemic immune responses, which further promote the development of GI tract diseases [13]. In the BE mouse model, the chemokines IL-1b and IL-8, secreted by epithelial cells in the esophagus and forestomach squamous epithelium, facilitated the progression of BE to EAC [6]. Moreover, the gut microbiome was associated with this process, as germ-free L2-1L1B mice did not develop dysplasia while the shift of the gut microbiome resulted in different speeds of developing esophageal dysplasia and tumor [6]. The above evidence further shows that these alterations of the bacterial community associated with inflammation can accelerate the development of esophageal diseases.

However, this is not limited to the gut bacteria as viruses, which outnumber bacterial cells by about tenfold in the gut, also contribute to human health and diseases [14–19]. In addition to the widely reported eukaryotic viruses [20–23], mounting data suggests that phages play a critical role in human health by affecting the bacterial community and function [19,24,25]. For example, bacterial-cell lysis caused by phage infection can lead to the release of nucleic acids, proteins, and lipids, which may trigger an inflammation response [26,27]. In addition, prophages that are integrated in bacterial genomes could supply them with virulence-associated genes that can increase their fitness under specific conditions [28]. Under stimulus (such as, DNA damage [29]), the prophages may switch to the lytic cycle [30], which can lead to gene exchange between bacteria, increasing their pathogenicity [31]. For example, the virulence gene that encodes the enterotoxin A was transferred to *Staphylococcus aureus* by phage-mediated horizontal gene transfer (HGT) [32,33]. Furthermore, phages can also obtain AMGs from bacteria to modulate bacterial metabolism [34]. These phage behaviors that regulate bacterial physiology could further indirectly influence human health, such as the occurrence of GI tract and non-GI tract diseases including IBD, CRC, Parkinson's disease, and Type I diabetes [27,35–37].

Former studies that investigated the role of phages in GI tract diseases have mainly focused on the phage community related to lower GI tract diseases, several studies have already described the disease-specific phage community that has been revealed in inflammation-induced diseases such as Crohn's disease and ulcerative colitis [27]. In a mouse model of intestinal colitis, it was reported that the bacteriophage community structure correlated with the disease status, and the presence of some phages during colitis was associated with an increase in pathobiontic host bacteria (*Escherichia-Shigella*, *Salmonella*, *Mycobacterium*) that was linked to the intestinal inflammation response [38]. However, the role of phages in the upper GI tract remains poorly described and limited to a few studies that have explored the viral community of the oral cavity [39,40]. The research related to the role of the phage community in esophageal diseases is also limited to one study that has used metagenomic data from the whole microbial community without isolating the viral like particles (VLPs) before sequencing [41]. Profiling the community composition of gut phages in esophageal diseases such as BE and EAC can provide some further insight into the role of phages in upper GI tract diseases.

Order and Family level taxonomic annotations were predicted using Demovir script (https://github.com/feargalr/Demovir; accessed on 27 July 2019) with default parameters and database. To calculate the relative abundances of viruses in each sample, clean reads from each sample were mapped to viral contigs using bbmap.sh from bbmap suite (v38.76). CoverM (v0.4.0) (https://github.com/wwood/CoverM; accessed on 20 February 2020) was used to estimate contig coverage. Feature Counts (v2.0.0) [51] was then used to estimate the number of reads that mapped to each gene. Viral proteins predicted in the previous step were fed into VIBRANT (v1.2.1) [52] to identify lytic and lysogenic phages and the function was annotated using protein mode with default parameters. VIBRANT annotates viral proteins by searching viral proteins against KEGG [53], VOGDB and PFAM databases, which include function annotation of protein sequences and AMGs. The virus (phage)-bacteria (host) interactions were predicted by VirHostMatcher-Net, which is a method based on the combination of features: virus-virus similarity, virus-host alignment-free similarity, virus-host shared CRISPR spacers and virus-host alignment-based matches [54]. Bacterial hosts were predicted for contigs with a length greater than 10 kb and score higher than 95% according to VirHostMatcher-Net.

2.4. Statistics Analysis

Alpha diversity of phage community was measured using qiime2 (https://qiime2.org; accessed on 29 January 2020). Principal Coordinates Analysis (PCoA) based on "Bray-Curtis" similarities was performed using R (v3.2, package vegan, The R Foundation, Vienna, Austria, 2016). Permutational Multivariate Analysis of Variance (PERMANOVA) was used to test the significant difference. All data performed statistical analyses, which were conducted in Prism 9- GraphPad (v9.0.0, GraphPad Software, San Diego, CA, USA, 2020) for the two-way analysis of variance [ANOVA], Tukey's post hoc test, and R (v4.0.2, stats package, The R Foundation, Vienna, Austria, 2020) for the Kruskal–Wallis and Dunn's post hoc test. The Jonckheere trend test was conducted in IBM SPSS Statistics (v27.0, IBM Corporation, Armonk, NY, USA, 2020). Meanwhile, multiple testing correction were performed to adjust the p value based on the "Bonferroni Holm" method. Only significant differences were shown in figures. Graphs were generated using Prism 9- GraphPad (v9.0.0, GraphPad Software, San Diego, CA, USA, 2020), Origin (v2020b, OriginLab Corporation, Northampton, MA, USA, 2020), Microsoft Excel (v365, Microsoft Corporation, Redmond, WA, USA), and R (v3.3.3, ggplot2 package, The R Foundation, Vienna, Austria, 2017). The data in results are provided as average $\pm$ SE.

3. Results

3.1. Gut Bacteriophage Community Structure Differed for BE and EAC Compared to Their Healthy Counterparts

On average, $43 \pm 2\%$ of all reads generated through sequencing were from viruses. In total, $854 \pm 50, 1136 \pm 19, 920 \pm 33$ viral contigs were obtained from sequences identified as viruses for CT, BE, and EAC, respectively. On average, from these contigs, over 95% of sequences were assigned to phages. The order of *Caudovirales*, which included *Herelleviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, and *Unclassified Caudovirales*, were the most abundant phages, accounting for more than 50% of total sequences in all three groups (Figure 1a, Figure S1). Among those phage families, the relative abundance of *Herelleviridae* was lower than 1% in three groups, the relative abundance of *Myoviridae* (1.12–41.97% in CT, 7.19–18.61% in BE, 1.37–34.36% in EAC), *Podoviridae* (2.03–31.68% in CT, 5.72–18.44% in BE, 3.72–11.01% in EAC) and *Siphoviridae* (8.28–79.60% in CT, 36.89–57.19% in BE, 41.48–75.69% in EAC) showed great variation within each group ($p > 0.05$). Some viral contigs were assigned to other phage or viral families including *Inoviridae*, *Microviridae*, *Tectiviridae*, *Herpesvirales*, *Marseilleviridae*, and *Pithoviridae* with a relative abundance of less than 1%. Meanwhile, the large difference in specific viral taxa between individuals was observed in the same group, which may be attributable to multiple factors such as age, gender, diet, or drug usage (Table S1). We next determined the dominant phage replication cycle

This study aimed to investigate the alteration of gut phages in different stages of esophageal diseases. For this purpose, we (1) determined the composition of the isolated bacteriophage community in BE patients, EAC patients, and healthy controls (CT); (2) predicted the bacterial host ranges of the gut phages in all three groups; (3) identified the metabolic pathways encoded by these phages.

2. Materials and Methods

2.1. Sample Collection

Sixteen samples were selected from the German BarrettNET registry including six BE patients, four EAC patients, and six CT for virome analysis. The clinical data are shown in Table S1, and additional information can be found in a previous study [42]. Stool samples were collected using Stool Collection Tubes with Stool DNA Stabilizer (STRATEC Molecular GmbH, Berlin, Germany). The sampling procedure was conducted mostly at home or in the clinic if the patients were on outpatient visits. Samples were shipped to the clinic human sample biobank and stored at $-80\ ^\circ$C until further virome DNA extraction.

2.2. Virome DNA Extraction

The stool samples were vortexed vigorously for 4 h at 4 $^\circ$C, then centrifuged at 4000 g for 30 min to collect supernatant. The supernatant was passed through 0.22 μm filters (PES Membrane, Lot No. ROCB29300, Merck Millipore, Co., Cork, Ireland) to remove bacterial-associated particles, and the volume was subsequently concentrated to less than 50 μL by Amicon® Ultra Centrifugal Filters (10 kDA, Lot No. R9EA18187, Merck Millipore, Co., Cork, Ireland). Then 1/5 volume of chloroform was mixed with the samples and centrifuged at 14,000 g for 3 min, retaining the upper phase followed by a DNAse I (1 U/μL, Lot No. 1158858, Invitrogen, Carlsbad, CA, USA) treatment for 1 h at 37 $^\circ$C to remove non-phage DNA. DNase I was inactivated by adding EDTA (0.1 M). Subsequently, lysis buffer (700 μL KOH stock (0.43 g/10 mL), 430 μL DDT stock (0.8 g/10 mL), and 370 μL H_2O, pH = 12) was added to the reaction and incubated at room temperature for 10 min followed by 2 h incubation at $-80\ ^\circ$C, and 5 min at 55 $^\circ$C. Lysed VLPs were then treated for 30 min at 55 $^\circ$C with Proteinase K (20 mg/mL, Lot No. 1112907, Invitrogen, Carlsbad, CA, USA) to digest remaining viral capsid and extract the virome DNA. AMPure beads (Agencourt, Beckman Coulter, Brea, CA, USA) were added to the extracted DNA and incubated for 15 min at room temperatureF. DNA was eluted from beads by 35 μL Tris buffer (10 mM, pH = 9.8) and stored at $-80\ ^\circ$C until it was sent for sequencing. Sequencing was performed on an Illumina HiSeq-PE150 platform.

2.3. Bioinformatic Analysis

On average, 9,358,935 $\pm$ 169,389 reads per samples were generated. Raw reads were processed with fastp (v0.20.1) [43] to remove adaptors and low-quality bases. Remaining reads were deduplicated using dedupe.sh from bbmap suite (v38.76) (https://sourceforge.net/projects/bbmap/; accessed on 29 January 2020). Then the obtained reads were assembled into contigs using metaSPAdes (v3.14.0) [44] with default parameters retaining only contigs longer than 1 kb. Redundant contigs were removed by dedupe.sh. Remaining contigs were used to predict viral sequences by the combination of VirSorter (v1.0.6) [45], CAT (v5.0.4) [46] and DeepVirFinder (v1.0) [47]. Contigs predicted as category 1 and 2 by Virsorter, or predicted as viruses by CAT, were classified as viruses. Contigs also were classified as viruses if they were predicted as category 3 by VirSorter or could not be classified to taxonomy by CAT but were predicted as a virus by DeepVirFinder with q value < 0.01. Predicted viral contigs were clustered using CD-HIT [48] if they shared >95% identity over 80% of the contig length, the longest contigs in each cluster were retained as a representative for downstream analysis.

For each representative viral contig, ORFs were predicted using Prodigal (v2.6.3) [49] and provided to vConTACT (v2.0) [50] for taxonomy annotation. For contigs that could not be assigned a taxonomy by vConTACT, CAT annotations were used. Otherwise,

(lytic versus lysogenic cycle). On average, EAC samples had more temperate phages (lysogenic cycle) than BE and CT ($p > 0.05$), 11.97% ± 2.43% in CT, 13.47% ± 1.15% in BE, 19.13% ± 4.90% in EAC (Figure S2).
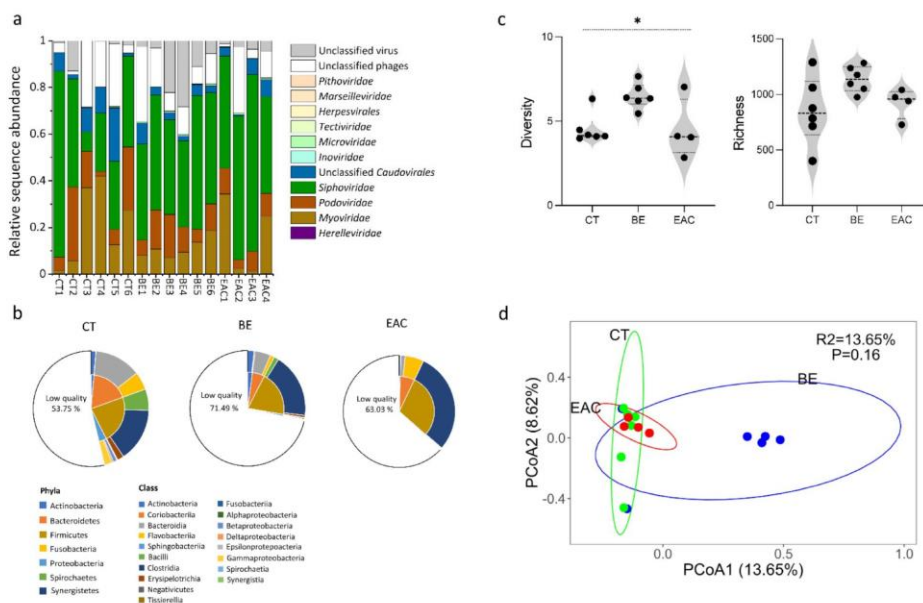
Figure 1. Composition of CT, BE, and EAC VLPs. (**a**) Relative abundance of viral families in CT, BE, and EAC; (**b**) The percentage of predicted bacterial hosts in CT, BE, and EAC. The inner cycle represents bacterial hosts at the phylum level, the outer cycle represents bacterial hosts at the class level. The low quality represents bacterial hosts predicted by contigs with a length lower than 10 kb and the score was lower than 95%; (**c**) Viral alpha diversity including richness (Ace) and diversity (Shannon) in samples from CT, BE, and EAC; (**d**) PCoA plot of the viral community composition based on the Bray–Curtis distances in CT, BE, and EAC samples. CT represents stool samples from healthy controls; BE represents stool samples from Barrett Esophagus patients; EAC represents stool samples from Esophageal Adenocarcinoma patients. Error bars indicate the average ± SE. Statistical significance was determined by Kruskal–Wallis, Dunn's post hoc test, asterisk indicates $p < 0.05$.

We next predicted the bacterial host range of the viral contigs from different groups in the study (Figure 1b). We observed that the bacterial hosts mainly spanned the phyla *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*, which were common across all three groups. In addition, we found that less than 0.1% of the phages were predicted to infect *Fusobacteria*, *Spirochaetes*, and *Synergistetes*. When the predicted bacterial host in class level was further compared, their relative abundance showed more obvious variation among the different groups, but these results were not statistically significant. For *Actinobacteria*, the relative abundance in CT (1.33% ± 0.28%) and BE (1.77% ± 0.21%) was higher than EAC (0.37% ± 0.11%) ($p > 0.05$). For *Flavobacteriia*, the relative abundance in CT (5.02% ± 1.45%) and EAC (5.38% ± 1.45%) was higher than BE (1.14% ± 0.16%) ($p > 0.05$). Notably, the classes *Betaproteobacteria*, *Deltaproteobacteria*, and *Gammaproteobacteria* were more abundant in CT compared with BE and EAC. Moreover, the relative abundance of *Bacteroidia* (13.08% ± 2.34% in CT, 4.38% ± 0.45% in BE, 1.25% ± 0.22% in EAC), *Bacilli* (5.97% ± 1.51% in CT,1.33% ± 0.14% in BE, <0.1% in EAC), and *Erysipelotrichia* (1.86% ± 0.54% in CT, 0.65% ± 0.093% in BE, <0.1% in EAC) were lower in BE and EAC compared to CT, while the relative abundance of *Clostridia* (15.06% ± 0.52% in CT, 18.04% ± 0.90% in BE, 29.20% ± 5.60% in EAC) was higher in BE and EAC com-

pared to CT. However, there was no significant difference (Jonckheere trend test, $p > 0.05$). Furthermore, the remaining classes had a lower relative abundance (0.0001%–0.31%) across the three groups.

We further examined how the changes in phages community composition affected the overall diversity. For the alpha diversity, a significant difference in phage diversity (Shannon) was found among the three groups ($p = 0.036$), while no significant difference was observed in phage richness (Ace) ($p > 0.05$) (Figure 1c). Furthermore, the alpha diversity showed differences among BE and EAC compared to CT samples ($p > 0.05$). Specifically, in both BE (1136.17 ± 19.48) and EAC (920.50 ± 33.87), the richness (Ace) was higher compared with that in CT (854.00 ± 50.73). However, only in BE (6.50 ± 0.11), the diversity (Shannon) was higher compared with that in CT (4.53 ± 0.15). Furthermore, BE had a higher level of richness (Ace) and diversity (Shannon) than EAC. In addition, no significant difference was detected ($p > 0.05$) in beta diversity (PCoA) among the three groups (Figure 1d).

3.2. Abundant and Rare Phage Communities in the Gut May Contribute to the Progress of Esophageal Carcinogesis

We used a sorting approach commonly applied in ecological study that classifies microbes into three groups based on their abundance [55,56], aiming to explore the role of less abundant microbes in different ecosystems. Using this approach, the contribution of rare, less abundant, bacterial Operational Taxonomic Units (OTUs) to some of the key ecological functions was revealed in the environment [57], which was previously overlooked. We believe this approach can be beneficial for studying phages in the gut. To this end, we divided phage contigs into abundant phages (relative abundance was more than 1% in total viral contigs), moderate phages (relative abundance was more than 0.1% and less than 1% in total viral contigs), and rare phages (relative abundance was less than 0.1% in total viral contigs). At these three relative abundance levels, members of the order *Caudovirales* (*Myoviridae*, *Siphoviridae*, and *Podoviridae*) showed the highest relative abundance in all three groups (Figure 2a). Subsequently, we observed that abundant phages presented significantly higher relative abundance (79.54% ± 2.28% in CT, 54.28% ± 2.19% in BE, and 72.25% ± 4.06% in EAC) when compared with moderate (14.79% ± 1.83% in CT, 34.38% ± 1.68% in BE, and 21.19% ± 3.57% in EAC) and rare phages (4.51% ± 0.52% in CT, 11.34% ± 0.85% in BE, and 6.56% ± 0.52% in EAC) in all three groups (abundant vs. moderate $p < 0.001$, abundant vs rare $p < 0.001$) (Figure 2b,c), while the highest number of contigs was from rare phages (788 ± 48 in CT, 994 ± 18 in BE, and 836 ± 28 in EAC), exceeding abundant (13 ± 1 in CT, 17 ± 1 in BE, and 11 ± 1 in EAC) and moderate (54 ± 8 in CT, 126 ± 7 in BE, and 74 ± 15 in EAC) phages in all three groups (Figure 2b,c). The highest relative abundance of abundant phages and the highest number of contigs of rare phages may suggest their different behaviors in relation to the gut bacterial community and esophageal diseases. Moreover, a significant difference was observed in beta-diversity on abundant ($p = 0.004$) and rare phages ($p = 0.003$) (Figure S3), which may imply that these two groups of phages showed higher sensitivity to the changes in the upper GI tract through esophageal disease progression. In addition, we found that the abundance of temperate phages that displayed a lysogenic replication cycle increased with the development of esophageal diseases. This may suggest a higher occurrence of HGT in these samples.

To further evaluate the importance of rare phages in HGT, we compared these three groups of phages to the number of bacterial hosts they infect. On the class level, we observed small differences between phage groups from different health conditions, rare phages infected 18 different bacterial classes whereas abundant phages infected 14 (Figure 1c). However, when bacterial hosts were compared on the genus level, both diversity and abundance showed large differences, 84 for rare versus 46 for abundant phages (Table S2). In particular, contigs belonging to rare phages showed similar characteristics regarding the number of hosts they infect over three groups, showing a broader bacterial host range compared to moderate and abundant phages. For example, the contigs from rare phages were able to infect 6 or 7 different bacterial hosts at the genus level (Table S3),

which was relatively higher than the bacterial hosts predicted for the contigs from abundant and moderate phages. The broader bacterial host range and higher number of contigs (Figure 2b, Tables S2 and S3) of rare phages could potentially lead to storing more AMGs in their genomes and, in turn, expand the frequency of HGT between gut bacteria.
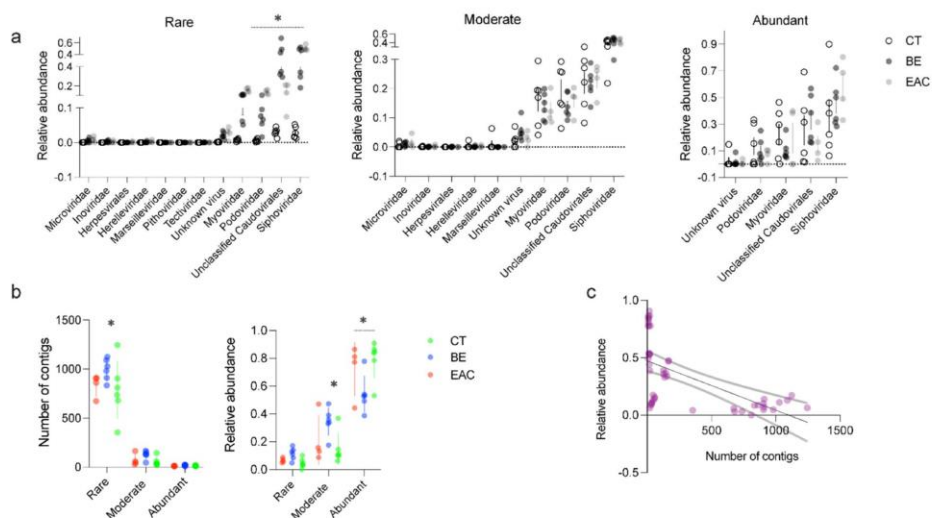
Figure 2. Composition of the rare, moderate, and abundant gut viruses in CT, BE and EAC samples. Rare, moderate, and abundant viruses were categorized based on the viral contig level. Abundant viruses represent viral contigs whose relative abundance was more than 1% in total contigs, moderate viruses represent viral contigs whose relative abundance was more than 0.1% and less than 1% in total contigs, and rare viruses represent viral contigs whose relative abundance was less than 0.1% in total contigs. (**a**) The relative abundance of viral families; (**b**) Number of contigs generated each viral contig category, rare, moderate, and abundant, on left and relative abundance of them on right. (**c**) Negative correlation between number of contigs, from rare, moderate, and abundant phages, and their relative abundance. CT represents stool samples from healthy controls; BE represents stool samples from Barrett Esophagus patients; EAC represents stool samples from Esophageal Adenocarcinoma patients. Statistical significance was determined by two–way analysis of variance [ANOVA], Tukey's post hoc test, asterisk indicates $p < 0.05$.

3.3. AMGs Found in Rare Bacteriophages Showed Increment in Esophageal Diseases

After annotation of the viral contigs, viruses were found to be involved in most of the microbial functions related to metabolism, cellular processes, genetic information processing, environment information processing, organismal system, and human disease (Figures 3a and S4). Significant differences were found for genes related to metabolism of cofactors and vitamins ($p = 0.0083$) and genes related to the prokaryotic defense system among the three groups ($p = 0.0202$) (Figure 3a). Genes involved in metabolism of cofactors and vitamins were found to be most abundant in CT phages, whereas genes related to the prokaryotic defense system were more abundant in EAC phages, suggesting a stronger arms race between phages and bacteria in this disease (Figure 3a). Notably, AMGs encoding bacterial toxins were found to be more abundant in the genome of rare bacteriophages including the *spyA* gene, *tccC* gene, *entB* gene and *entD* gene, which are involved in microbial cellular processes. The *spyA* gene, which encodes a C3 family ADP-ribosyltransferase (bacterial exotoxin) [58], showed a slightly higher level of relative abundance in BE and EAC ($p > 0.05$) compared with the other three AMGs (Figure 3b). Moreover, the *spyA* gene level was relatively higher in BE (0.00040 ± 0.00011) and EAC (0.0027 ± 0.0012) compared with CT (0.00031 ± 0.000012) ($p > 0.05$). Other AMGs that relate to LPS biosynthesis pro-

teins were also found in the genome of rare phages including the *lpxD* gene, *kdsC* gene and *gmnB* gene, which are involved in microbial metabolism (Figure 3b). The *lpxD* gene only presented in BE with a relative abundance of 0.00031 ± 0.000113. The *kdsC* gene presented in BE (0.000089 ± 0.000036) and CT (0.0000024 ± 0.00000097). For the *gmnB* gene, it was relatively higher in EAC (0.00064 ± 0.00029) and BE (0.00024 ± 0.000094) compared with CT (0.00015 ± 0.000044) ($p > 0.05$). The higher abundance of these genes in phages from BE and EAC compared to CT may have resulted from the increase of pathogenic bacteria, mainly Gram-negatives, in the esophageal diseases, leading to a higher chance of obtaining AMGs, which are related to LPS biosynthesis proteins encoded by phages. We next explored the appearance of these genes in the Gut Phages Database (GPD) containing 142,809 non-redundant globally distributed phage genomes. We found many phages encoding these genes in GPD with one exception, *tccC*, showing these AMGs are ubiquitous in the human gut (Figure S5). Toxin complex (Tc) is a multisubunit toxin consisting of three components (TcA, B, and C) encoded by pathogenic bacteria infecting both insects and humans. TcAs that make functional pores combine with TcB-TcC subunits to create active chimeric holotoxins. Tc toxins are encoded by human pathogens like *Yersinia pestis*, *Y. pseudotuberculosis*, and *Morganella morganii* and are believed to significantly contribute to these bacteria's pathogenicity. Yet, their role in EAC remains to be revealed [59]. The increase of these genes in phages from BE and EAC may contribute to the severity of these diseases through exchanging genes that are involved in bacterial exotoxin production and LPS biosynthesis in esophageal carcinogenesis. This warrants further investigation.
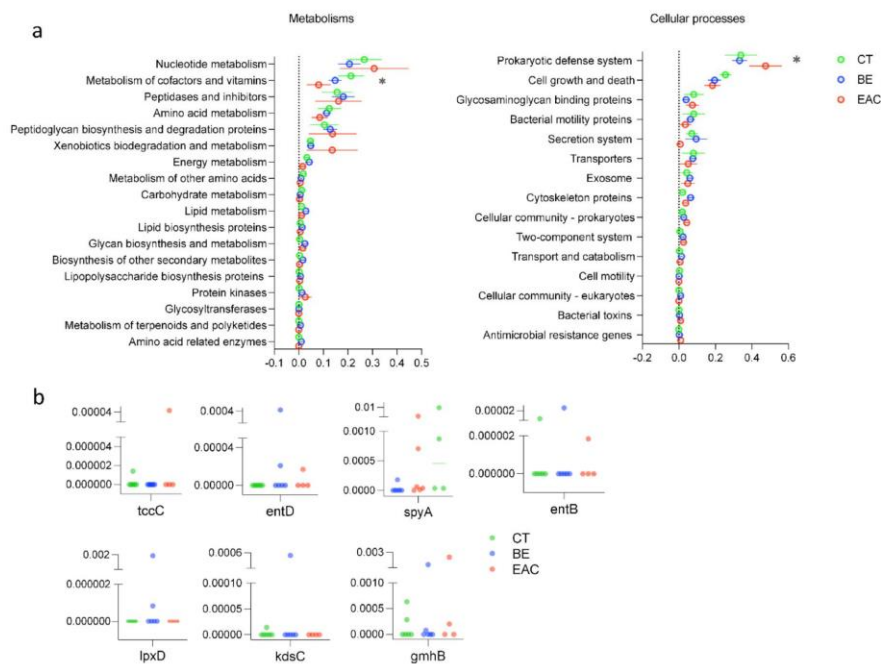
Figure 3. Viral functional traits. (**a**) The relative abundance of different functional traits in viral sequences; (**b**) The relative abundance of genes encoding four different bacterial toxins with higher abundance in BE and EAC samples compared with CT on the top, and genes encoding the LPS biosynthesis proteins on the bottom. Error bars indicate the average $\pm$ SE. Statistical significance was determined by two–way analysis of variance [ANOVA], Tukey's post hoc test, asterisk indicates $p < 0.05$.

4. Discussion

Barrett's esophagus (BE) is a condition caused by the metaplastic replacement of the normal squamous epithelium by columnar epithelium. BE is closely associated with the development of esophageal adenocarcinoma (EAC), a disease in which cancerous cells develop in the tissues of the esophagus with a high mortality rate [42]. It has been recently shown that gut dysbiosis can activate oncogenic signaling pathways, leading to the production of tumor-promoting metabolites, and further influence the esophageal mucosal inflammation and tumorigenesis [60]. For example, gut bacteria regulate bile acid (BA) metabolism. Under stimulation such as a high-fat diet, the gut bacteria changed, and the level of BA increased accordingly [61]. The reflux of BA to the esophagus caused esophageal damage, leading to BE and subsequent EAC. In an animal experiment simulating BA reflux, overexpression of the inflammatory cells, IL-6 and TNF- α, was found [62]. This indicated that gut bacterial alterations could indirectly induce the esophageal mucosal inflammation and carcinogenesis [62–64]. Despite a wealth of data on the role of gut bacteria in GI tract disease, we have only recently recognized the association of gut viruses with some GI tract diseases, including CRC in which the diversity of the gut viruses is significantly increased in stool samples from CRC patients, suggesting a disease-specific signature that can be used to differentiate CRC samples from controls [37]. The CRC-associated virome includes primarily temperate bacteriophages belonging to *Siphoviridae* and *Myoviridae* families [65]. The impact of phages on gut homeostasis is not restricted to their interactions with gut bacteria as phages can directly interact with the human host. In vitro studies have demonstrated that phages can cross the epithelial cell layer through transcytosis, thereby stimulating the underlying immune cells [22,66–69]. For example, the interaction between *E.coli* phages and the immune system has been associated with Type I Diabetes autoimmunity [36]. It has been reported that phages can activate IFN-γ produced by CD4+ T cells via the nucleotide-sensing receptor TLR9, which accelerates intestinal inflammation and colitis, leading to a systemic inflammation response [70]. The consistent disease-specific signature of gut viruses [27,37], suggests a potential association between gut viruses and human disease.

Studies that investigated the esophageal virome, using metagenomic data of whole microbial communities rather than profiling the isolated viral communities, have identified a range of phages, including *Streptococcus*, *Campylobacter*, *Lactococcus*, and γ-Proteobacteria phages [71]. The aforementioned and those that only explored the bacterial community of the esophagus have mainly used biopsy samples for virome and bacterium analysis [10,72,73]. Although, biopsies could directly reflect the disease-associated microbial signature at the lesion, the sampling procedure is invasive, time-consuming, costly, and may induce potential complications [74]. Moreover, biopsy samples often have limited microbial materials, with a lower probability of successful sequencing and downstream analysis [75]. Thus, an amplification step (e.g., whole genome amplification) is necessary, which might introduce biases to study results. On the contrary, stool samples collected by non-invasive methods often supply sufficient materials for research purposes [76].

Here we explored stool samples from BE, EAC, and CT phages community composition in esophageal diseases. Our in-depth gut virome analysis during esophageal carcinogenesis provided some evidence of gut phage community changes between different stages of esophageal diseases. Consistent with previous studies that have explored the gut viruses, mainly in the lower GI tract diseases such as IBD and CRC [27,65], phages from the order *Caudovirales* were the most dominant phages in the samples from esophageal diseases. Compared with CT, the alpha diversity has changed with the esophageal diseases progress, and a relatively higher alpha diversity was observed in BE samples compared to CT and EAC. This was not reflected in the beta diversity as no significant differences were observed among three groups. Using a common sorting approach in microbial ecology, we identified disease-associated differences in diversity and abundance of rare phages, suggesting a potential link between these phages and esophageal diseases. In addition, consistent with previous studies on diseases like IBD [77] and CRC [65], we

observed changes in the proportion of lytic/lysogenic replication cycles of phages, and more temperate phages were observed in esophageal carcinogenesis. These results further support earlier studies that reported the dominance of virulent phages (lytic cycle) in the healthy human gut replaced by temperate phages in Crohn's disease and ulcerative colitis [23,24]. Furthermore, the relatively higher percentage of temperate phages in samples from esophageal diseases may imply more influence on the bacterial physiology through phage mediated HGT in those groups. However, we did not study the bacterial community of these samples, the community structure of the predicted bacterial hosts for the phages identified in the study may suggest a complex relationship between bacteria and bacteriophage community in esophageal diseases. Earlier studies on lower GI tract diseases such as CRC have observed that the effect of phages resulted from their interactions with the whole bacterial community, rather than the bacterial taxa directly contributing to the disease severity [65]. However, there was no direct correlation between bacterial diversity and phage diversity [27,37].

In addition, we found several AMGs in the genome of the rare phages, further emphasizing the potential role of phages in regulating bacterial physiology by supplying their host with beneficial genes. Specifically, a slightly higher abundance of *spyA* ($p > 0.05$) was observed in BE and EAC, potentially contributing to the production of bacterial exotoxins, which disrupt cytoskeletal structures and promote colonization of pathogenic bacteria [58]. The relatively higher abundance of AMGs related to LPS biosynthesis proteins were also found in BE and EAC, which may indicate the dominance of Gram-negative bacteria and the potential inflammatory effects of phage–bacteria interactions. Phages that carry these AMGs can introduce these genes to the genome of gut bacteria via integration, which may contribute to the severity of the esophageal diseases through lysogenic conversion. This could further induce gut inflammation through expression of the phage-derived virulence genes and deteriorate esophageal disease. Intestinal permeability caused by phage-mediated changes of gut microbiota could also lead to systemic inflammatory responses [78]. Given the high variability of the microbiome between individuals and the limited number of samples analyzed, it is difficult to identify significant differences in viral community structure between different groups in the current study. Thus, our findings should be further pursued with a larger sample size.

5. Conclusions

In summary, this study provides further evidence of potential relationship between gut phages and esophageal diseases. Interestingly, the distinct gut phage community structure was identified in two different stages of esophageal diseases, and these differences were mainly found in abundant and rare bacteriophages. Notably, rare phages and HGT mediated by them have been found to be more related to esophageal diseases. Specially, the rare phages contributed to enriching AMGs related to bacterial exotoxin and LPS biosynthesis proteins, and the possible upregulated level of these genes. These, in turn, may contribute to changes in the gut bacterial composition and inflammation, which lead to the development of esophageal diseases, as previously suggested [6]. However, given the small sample size in our study, the potential diagnostic importance of AMGs and disease-specific viral signature identified should be experimentally validated in further studies.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/microorganisms9081701/s1, Figure S1: Gene-sharing taxonomic network of viral sequences in this study, including viral RefSeq viruses v85. Figure S2: The proportion of lytic/lysogenic replication cycles predicted for the viral contigs from three groups. Figure S3: PCoA plot of the viral community composition based on the Bray-Curtis distances in CT, BE, and EAC samples. Figure S4: The relative abundance of different functional traits in viral sequences. Figure S5: The number of phages that contained the identified AMGs of this study in the Gut Phage Database (GPD). Table S1: Clinical information of individuals from three groups. Table S2: The relative abundance of bacterial host at genus level for abundant, moderate, and rare bacteriophages. Table S3: The percentage of contig

relative abundance in different number of predicted bacterial genus types for abundant, moderate, and rare bacteriophages. Table S4: The relative abundance of identified AMGs.

References

1. He, Y.; Li, D.; Shan, B.; Liang, D.; Shi, J.; Chen, W.; He, J. Incidence and mortality of esophagus cancer in China, 2008–2012. *Chin. J. Cancer Res.* **2019**, *31*, 426–434. [CrossRef] [PubMed]
2. Pennathur, A.; Gibson, M.K.; Jobe, B.A.; Luketich, J.D. Oesophageal carcinoma. *Lancet* **2013**, *381*, 400–412. [CrossRef]
3. Martinucci, I.; De Bortoli, N.; Russo, S.; Bertani, L.; Furnari, M.; Mokrowiecka, A.; Malecka-Panas, E.; Savarino, V.; Savarino, E.; Marchi, S. Barrett's esophagus in 2016: From pathophysiology to treatment. *World J. Gastrointest. Pharmacol. Ther.* **2016**, *7*, 190–206. [CrossRef]
4. Pohl, H.; Koch, M.; Khalifa, A.; Papanikolaou, I.; Scheiner, K.; Wiedenmann, B.; Rösch, T. Evaluation of endocytoscopy in the surveillance of patients with Barrett's esophagus. *Endoscopy* **2007**, *39*, 492–496. [CrossRef]
5. Hooper, L.V.; Gordon, J.I. Commensal host-bacterial relationships in the gut. *Science* **2001**, *292*, 1115–1118. [CrossRef]
6. Münch, N.S.; Fang, H.-Y.; Ingermann, J.; Maurer, H.C.; Anand, A.; Kellner, V.; Sahm, V.; Wiethaler, M.; Baumeister, T.; Wein, F.; et al. High-fat diet accelerates carcinogenesis in a mouse model of barrett's esophagus via interleukin 8 and alterations to the gut microbiome. *Gastroenterology* **2019**, *157*, 492–506. [CrossRef]
7. Snider, E.J.; Compres, G.; Freedberg, D.E.; Khiabanian, H.; Nobel, Y.R.; Stump, S.; Uhlemann, A.-C.; Lightdale, C.J.; Abrams, J.A. Alterations to the esophageal microbiome associated with progression from Barrett's esophagus to esophageal adenocarcinoma. *Cancer Epidemiol. Prev. Biomark.* **2019**, *28*, 1687–1693. [CrossRef]
8. Ren, Z.; Li, A.; Jiang, J.; Zhou, L.; Yu, Z.; Lu, H.; Xie, H.; Chen, X.; Shao, L.; Zhang, R.; et al. Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. *Gut* **2019**, *68*, 1014–1023. [CrossRef]
9. Manor, O.; Dai, C.L.; Kornilov, S.A.; Smith, B.; Price, N.D.; Lovejoy, J.C.; Gibbons, S.M.; Magis, A.T. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat. Commun.* **2020**, *11*, 5206. [CrossRef] [PubMed]
10. Yang, L.; Lu, X.; Nossa, C.W.; Francois, F.; Peek, R.M.; Pei, Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. *Gastroenterology* **2009**, *137*, 588–597. [CrossRef]
11. Yang, L.; Francois, F.; Pei, Z. Molecular pathways: Pathogenesis and clinical implications of microbiome alteration in esophagitis and Barrett esophagus. *Clin. Cancer Res.* **2012**, *18*, 2138–2144. [CrossRef]
12. Zaidi, A.H.; Kelly, L.A.; Kreft, R.E.; Barlek, M.; Omstead, A.N.; Matsui, D.; Boyd, N.H.; Gazarik, K.E.; Heit, M.I.; Nistico, L.; et al. Associations of microbiota and toll-like receptor signaling pathway in esophageal adenocarcinoma. *BMC Cancer* **2016**, *16*, 52. [CrossRef] [PubMed]
13. Okereke, I.; Hamilton, C.; Wenholz, A.; Jala, V.; Giang, T.; Reynolds, S.; Miller, A.; Pyles, R. Associations of the microbiome and esophageal disease. *J. Thorac. Dis.* **2019**, *11*, S1588–S1593. [CrossRef]
14. Wylie, K.M.; Weinstock, G.M.; Storch, G.A. Emerging view of the human virome. *Transl. Res.* **2012**, *160*, 283–290. [CrossRef] [PubMed]
15. Carding, S.R.; Davis, N.; Hoyles, L. The human intestinal virome in health and disease. *Aliment. Pharmacol. Ther.* **2017**, *46*, 800–815. [CrossRef] [PubMed]
16. Cadwell, K. The virome in host health and disease. *Immunity* **2015**, *42*, 805–813. [CrossRef] [PubMed]
17. Lepage, P.; Leclerc, M.C.; Joossens, M.; Mondot, S.; Blottière, H.M.; Raes, J.; Ehrlich, D.; Doré, J. A metagenomic insight into our gut's microbiome. *Gut* **2013**, *62*, 146–158. [CrossRef] [PubMed]

18. Mills, S.; Shanahan, F.; Stanton, C.; Hill, C.; Coffey, A.; Ross, R.P. Movers and shakers: Influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* **2013**, *4*, 4–16. [CrossRef] [PubMed]

19. Dalmasso, M.; Hill, C.; Ross, R.P. Exploiting gut bacteriophages for human health. *Trends Microbiol.* **2014**, *22*, 399–405. [CrossRef]

20. Ungaro, F.; Massimino, L.; Furfaro, F.; Rimoldi, V.; Peyrin-Biroulet, L.; D'alessio, S.; Danese, S. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* **2019**, *10*, 149–158. [CrossRef]

21. Tetz, G.; Tetz, V. Prion-like domains in eukaryotic viruses. *Sci. Rep.* **2018**, *8*, 8931. [CrossRef] [PubMed]

22. Breitbart, M.; Hewson, I.; Felts, B.; Mahaffy, J.M.; Nulton, J.; Salamon, P.; Rohwer, F. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **2003**, *185*, 6220–6223. [CrossRef] [PubMed]

23. Reyes, A.; Haynes, M.; Hanson, N.; Angly, F.E.; Heath, A.C.; Rohwer, F.; Gordon, J.I. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **2010**, *466*, 334–338. [CrossRef]

24. Sabino, J.; Hirten, R.P.; Colombel, J.F. bacteriophages in gastroenterology—From biology to clinical applications. *Aliment. Pharmacol. Ther.* **2020**, *51*, 53–63. [CrossRef]

25. De Sordi, L.; Lourenço, M.; Debarbieux, L. The battle within: Interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host Microbe* **2019**, *25*, 210–218. [CrossRef]

26. Łusiak-Szelachowska, M.; Weber-Dąbrowska, B.; Jończyk-Matysiak, E.; Wojciechowska, R.; Górski, A. Bacteriophages in the gastrointestinal tract and their implications. *Gut Pathog.* **2017**, *9*, 44. [CrossRef]

27. Norman, J.M.; Handley, S.A.; Baldridge, M.T.; Droit, L.; Liu, C.Y.; Keller, B.C.; Kambal, A.; Monaco, C.L.; Zhao, G.; Fleshner, P.; et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **2015**, *160*, 447–460. [CrossRef] [PubMed]

28. Mirzaei, M.K.; Xue, J.; Costa, R.; Ru, J.; Schulz, S.; Taranu, Z.E.; Deng, L. Challenges of studying the human virome–relevant emerging technologies. *Trends Microbiol.* **2021**, *29*, 171–181. [CrossRef]

29. Lin, E.C.; Lynch, A.S. *Regulation of Gene Expression in Escherichia coli*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.

30. Feiner, R.; Argov, T.; Rabinovich, L.; Sigal, N.; Borovok, I.; Herskovits, A.A. A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* **2015**, *13*, 641–650. [CrossRef] [PubMed]

31. Brüssow, H.; Canchaya, C.; Hardt, W.-D. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 560–602. [CrossRef] [PubMed]

32. Saunders, J.R.; Allison, H.; James, C.E.; McCarthy, A.J.; Sharp, R. Phage-mediated transfer of virulence genes. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 662–666. [CrossRef]

33. Coleman, D.C.; Sullivan, D.J.; Russel, R.J.; Arbuthnott, J.P.; Carey, B.F.; Pomeroy, H.M. Staphylococcus aureus bacteriophages mediating the simultaneous lysogenic conversion of β-lysin, staphylokinase and enterotoxin A: Molecular mechanism of triple conversion. *Microbiology* **1989**, *135*, 1679–1697. [CrossRef]

34. Crummett, L.T.; Puxty, R.J.; Weihe, C.; Marston, M.F.; Martiny, J.B. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **2016**, *499*, 219–229. [CrossRef]

35. Tetz, G.; Brown, S.M.; Hao, Y.; Tetz, V. Parkinson's disease and bacteriophages as its overlooked contributors. *Sci. Rep.* **2018**, *8*, 10812. [CrossRef]

36. Tetz, G.; Brown, S.M.; Hao, Y.; Tetz, V. Type 1 diabetes: An association between autoimmunity, the dynamics of gut amyloid-producing E. coli and their phages. *Sci. Rep.* **2019**, *9*, 9685. [CrossRef]

37. Nakatsu, G.; Zhou, H.; Wu, W.K.K.; Wong, S.H.; Coker, O.O.; Dai, Z.; Li, X.; Szeto, C.-H.; Sugimura, N.; Lam, T.Y.-T.; et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology* **2018**, *155*, 529–541. [CrossRef] [PubMed]

38. Duerkop, B.A.; Kleiner, M.; Paez-Espino, D.; Zhu, W.; Bushnell, B.; Hassell, B.; Winter, S.E.; Kyrpides, N.C.; Hooper, L.V. Murine colitis reveals a disease-associated bacteriophage community. *Nat. Microbiol.* **2018**, *3*, 1023–1031. [CrossRef]

39. Ly, M.; Abeles, S.R.; Boehm, T.K.; Robles-Sikisaka, R.; Naidu, M.; Santiago-Rodriguez, T.; Pride, D.T. Altered oral viral ecology in association with periodontal disease. *MBio* **2014**, *5*, e01133-14. [CrossRef]

40. Abeles, S.R.; Robles-Sikisaka, R.; Ly, M.; Lum, A.G.; Salzman, J.; Boehm, T.K.; Pride, D.T. Human oral viruses are personal, persistent and gender-consistent. *ISME J.* **2014**, *8*, 1753–1767. [CrossRef]

41. Deshpande, N.P.; Riordan, S.M.; Castaño-Rodríguez, N.; Wilkins, M.R.; Kaakoush, N.O. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome* **2018**, *6*, 227. [CrossRef] [PubMed]

42. Wiethaler, M.; Slotta-Huspenina, J.; Brandtner, A.; Horstmann, J.; Wein, F.; Baumeister, T.; Radani, N.; Gerland, S.; Anand, A.; Lange, S.; et al. BarrettNET—A prospective registry for risk estimation of patients with Barrett's esophagus to progress to adenocarcinoma. *Dis. Esophagus* **2019**, *32*, doz024. [CrossRef]

43. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]

44. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [CrossRef] [PubMed]

45. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985. [CrossRef] [PubMed]

46. von Meijenfeldt, F.B.; Arkhipova, K.; Cambuy, D.D.; Coutinho, F.H.; Dutilh, B.E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* **2019**, *20*, 217. [CrossRef]

47. Ren, J.; Song, K.; Deng, C.; Ahlgren, N.A.; Fuhrman, J.A.; Li, Y.; Xie, X.; Poplin, R.; Sun, F. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **2020**, *8*, 64–77. [CrossRef] [PubMed]

48. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]

49. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]

50. Jang, H.B.; Bolduc, B.; Zablocki, O.; Kuhn, J.H.; Roux, S.; Adriaenssens, E.M.; Brister, J.R.; Kropinski, A.M.; Krupovic, M.; Lavigne, R.; et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **2019**, *37*, 632–639. [CrossRef]

51. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef] [PubMed]

52. Kieft, K.; Zhou, Z.; Anantharaman, K. VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **2020**, *8*, 90. [CrossRef]

53. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef]

54. Wang, W.; Ren, J.; Tang, K.; Dart, E.; Ignacio-Espinoza, J.C.; Fuhrman, J.A.; Braun, J.; Sun, F.; Ahlgren, N.A. A network-based integrated framework for predicting virus–prokaryote interactions. *NAR Genom. Bioinform.* **2020**, *2*, lqaa044. [CrossRef]

55. Ji, M.; Kong, W.; Stegen, J.; Yue, L.; Wang, F.; Dong, X.; Cowan, D.A.; Ferrari, B.C. Distinct assembly mechanisms underlie similar biogeographical patterns of rare and abundant bacteria in Tibetan Plateau grassland soils. *Environ. Microbiol.* **2020**, *22*, 2261–2272. [CrossRef]

56. Sjöstedt, J.; Koch-Schmidt, P.; Pontarp, M.; Canbäck, B.; Tunlid, A.; Lundberg, P.; Hagström, Å.; Riemann, L. Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl. Environ. Microbiol.* **2012**, *78*, 1361–1369. [CrossRef]

57. Vigneron, A.; Cruaud, P.; Alsop, E.; de Rezende, J.R.; Head, I.M.; Tsesmetzis, N. Beyond the tip of the iceberg; a new view of the diversity of sulfite-and sulfate-reducing microorganisms. *ISME J.* **2018**, *12*, 2096–2099. [CrossRef]

58. Coye, L.H.; Collins, C.M. Identification of SpyA, a novel ADP-ribosyltransferase of Streptococcus pyogenes. *Mol. Microbiol.* **2004**, *54*, 89–98. [CrossRef] [PubMed]

59. Leidreiter, F.; Roderer, D.; Meusch, D.; Gatsogiannis, C.; Benz, R.; Raunser, S. Common architecture of Tc toxins from human and insect pathogenic bacteria. *Sci. Adv.* **2019**, *5*, eaax6497. [CrossRef]

60. Deng, Y.; Tang, D.; Hou, P.; Shen, W.; Li, H.; Wang, T.; Liu, R. Dysbiosis of gut microbiota in patients with esophageal cancer. *Microb. Pathog.* **2021**, *150*, 104709. [CrossRef] [PubMed]

61. Schwabe, R.F.; Jobin, C. The microbiome and cancer. *Nat. Rev. Cancer* **2013**, *13*, 800–812. [CrossRef] [PubMed]

62. Sun, D.; Wang, X.; Gai, Z.; Song, X.; Jia, X.; Tian, H. Bile acids but not acidic acids induce Barrett's esophagus. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 1384. [PubMed]

63. Schmidt, M.; Ankerst, D.P.; Chen, Y.; Wiethaler, M.; Slotta-Huspenina, J.; Becker, K.-F.; Horstmann, J.; Kohlmayer, F.; Lehmann, A.; Linkohr, B.; et al. Epidemiologic Risk Factors in a Comparison of a Barrett Esophagus Registry (BarrettNET) and a Case–Control Population in Germany. *Cancer Prev. Res.* **2020**, *13*, 377–384. [CrossRef]

64. Elliott, D.R.F.; Perner, J.; Li, X.; Symmons, M.F.; Verstak, B.; Eldridge, M.; Bower, L.; O'Donovan, M.; Gay, N.J.; Fitzgerald, R.C. Impact of mutations in Toll-like receptor pathway genes on esophageal carcinogenesis. *PLoS Genet.* **2017**, *13*, e1006808.

65. Hannigan, G.D.; Duhaime, M.B.; Ruffin, M.T.; Koumpouras, C.C.; Schloss, P.D. Diagnostic potential and interactive dynamics of the colorectal cancer virome. *MBio* **2018**, *9*, e02248-18. [CrossRef]

66. Sinha, A.; Maurice, C.F. Bacteriophages: Uncharacterized and dynamic regulators of the immune system. *Mediat. Inflamm.* **2019**, *2019*, 3730519. [CrossRef] [PubMed]

67. Nguyen, S.; Baker, K.; Padman, B.S.; Patwa, R.; Dunstan, R.A.; Weston, T.A.; Schlosser, K.; Bailey, B.; Lithgow, T.; Lazarou, M.; et al. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *MBio* **2017**, *8*, e01874. [CrossRef] [PubMed]

68. Bichet, M.C.; Chin, W.H.; Richards, W.; Lin, Y.-W.; Avellaneda-Franco, L.; Hernandez, C.A.; Oddo, A.; Chernyavskiy, O.; Hilsenstein, V.; Neild, A. Bacteriophage uptake by mammalian cell layers represents a potential sink that may impact phage therapy. *Iscience* **2021**, *24*, 102287. [CrossRef]

69. Wahida, A.; Tang, F.; Barr, J.J. Rethinking phage-bacteria-eukaryotic relationships and their influence on human health. *Cell Host Microbe* **2021**, *29*, 681–688. [CrossRef]

70. Gogokhia, L.; Buhrke, K.; Bell, R.; Hoffman, B.; Brown, D.G.; Hanke-Gogokhia, C.; Ajami, N.J.; Wong, M.C.; Ghazaryan, A.; Valentine, J.F.; et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* **2019**, *25*, 285–299. [CrossRef]

71. Moonen, A.; Annese, V.; Belmans, A.; Bredenoord, A.J.; Des Varannes, S.B.; Costantini, M.; Dousset, B.; Elizalde, J.I.; Fumagalli, U.; Gaudric, M.; et al. Long-term results of the European achalasia trial: A multicentre randomised controlled trial comparing pneumatic dilation versus laparoscopic Heller myotomy. *Gut* **2016**, *65*, 732–739. [CrossRef]

72. Pei, Z.; Bini, E.J.; Yang, L.; Zhou, M.; Francois, F.; Blaser, M.J. Bacterial biota in the human distal esophagus. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4250–4255. [CrossRef]

73. Pei, Z.; Yang, L. Bacterial biota in reflux esophagitis and Barrett's esophagus. *World J. Gastroenterol.* **2005**, *11*, 7277–7283. [CrossRef] [PubMed]

74. Fillon, S.A.; Harris, J.K.; Wagner, B.D.; Kelly, C.J.; Stevens, M.J.; Moore, W.; Fang, R.; Schroeder, S.; Masterson, J.C.; Robertson, C.E.; et al. Novel device to sample the esophageal microbiome—The esophageal string test. *PLoS ONE* **2012**, *7*, e42938. [CrossRef]

75. Lim, E.H.; Zhang, S.-L.; Li, J.-L.; Yap, W.-S.; Howe, T.-C.; Tan, B.-P.; Lee, Y.-S.; Wong, D.; Khoo, K.-L.; Seto, K.-Y.; et al. Using whole genome amplification (WGA) of low-volume biopsies to assess the prognostic role of EGFR, KRAS, p53, and CMET mutations in advanced-stage non-small cell lung cancer (NSCLC). *J. Thorac. Oncol.* **2009**, *4*, 12–21. [CrossRef] [PubMed]

76. Bull, M.J.; Plummer, N.T. Part 1: The human gut microbiome in health and disease. *Integr. Med. A Clin. J.* **2014**, *13*, 17.

77. Clooney, A.G.; Sutton, T.D.; Shkoporov, A.N.; Holohan, R.K.; Daly, K.M.; O'Regan, O.; Ryan, F.J.; Draper, L.A.; Plevy, S.E.; Ross, R.P.; et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe* **2019**, *26*, 764–778. [CrossRef] [PubMed]

78. Tetz, G.; Tetz, V. Bacteriophage infections of microbiota can lead to leaky gut in an experimental rodent model. *Gut Pathog.* **2016**, *8*, 33. [CrossRef]

Supplementary data to the Article:

Contents

Table S1: Clinical information of individuals from three groups

Table S2: The relative abundance of bacterial host at genus level for abundant, moderate, and rare bacteriophages

Table S3: The percentage of viral contig relative abundance in different number of predicted bacterial genus types for abundant, moderate, and rare bacteriophages

Table S4: The relative abundance of identified AMGs

Figure S1: Gene-sharing taxonomic network of viral sequences in this study, including viral RefSeq viruses v85. RefSeq viruses are in triangle shape. Viruses from the *Siphoviridae* family are highlighted in red, *Podoviridae* in cyan, *Myoviridae* in light green, *Microviridae* in orange, *Inoviridae* in purple, *Leviviridae* in yellow, *Fuselloviridae* in dark green, others in black, and unclassified in grey. Each node represents a genome, whereas connecting edges show shared genes between genomes.

Figure S2: The proportion of lytic/lysogenic replication cycles predicted for the viral contigs from three groups; CT represents stool samples from healthy controls; BE represents stool samples from Barrett Esophagus patients; EAC represents stool samples from Esophageal Adenocarcinoma patients. Error bars indicate the average±SE. Statistical significance was determined by Kruskal-Wallis, Dunn's post hoc test, $p < 0.05$.

Figure S3: PCoA plot of the viral community composition based on the Bray-Curtis distances in CT, BE, and EAC samples. (**a**) PCoA plot of the abundant viral community composition (**b**) PCoA plot of the moderate viral community composition, and (**c**) PCoA plot of the abundant rare viral community composition. CT represents stool samples from healthy controls; BE represents stool samples from Barrett Esophagus patients; EAC represents stool samples from Esophageal Adenocarcinoma patients. Statistical significance was determined by the PERMANOVA test.

Table S1

Table S1 Clinical information of individuals from three groups

	Age	Gender	Height	Weight	Smoke	Alcohol	Region	Medicine	Antacid-Sodbr-medic	Omeprazole-Sodbr-medic	Fever	Surgical interventions	High blood pressure	CholesterolE
CT1	62	male	#	#	yes	yes	#	#	#	#	#	#	no	no
CT2	59	male	#	#	yes	yes	Western and Central Europe	#	#	#	no	yes	yes	no
CT3	49	male	#	#	yes	Yes	Western and Central Europe	Painkiller	#	#	no	yes	no	yes
CT4	76	male	#	#	no	yes	Western and Central Europe	Heart medicince, Sleep aid medicine	#	#	no	yes	#	#
CT5	33	Female	169cm	57kg	no	yes	South america	Heartburn medicine,Other medicine	#	Esomeprazole (NexumMUPS)	no	yes	no	no
CT6	37	Female	#	#	no	yes	Western and Central Europe	Other medicine	#	Pantoprazole (Pantozol)	no	yes	no	no
BE1	75	Female	169cm	66kg	no	yes		Hormone medicine	#	Esomeprazole (NexumMUPS)	no	yes	no	yes
BE2	#	#	#	#	#	#		#	#	#	#	#	#	#
BE3	63	male	#	#	yes	no	Western and Central Europe	Painkiller, Heartburn medicine	#	Omeprazole (AntraMUPS)	no	yes	yes	no
BE4	70	Female	#	#	no	yes	Western and Central Europe	Heartburn medicine, Other medicine	Talcid	Esomeprazole (NexumMUPS)	no	yes	no	yes
BE5	50	male	#	#	yes	yes	Western and Central Europe	Painkiller, Other medicine	#	#	no	yes	no	no
BE6	79	male	#	#	yes	yes	Western and Central Europe	Other medicine	#	#	no	yes	no	no
EAC1	47	male	#	#	yes	yes	Western and Central Europe	Other medicine	#	Pantoprazole (Pantozol)	no	yes	yes	yes
EAC2	69	Female	#	#	no	yes	Western and Central Europe	Other medicine	#	#	no	yes	no	no
EAC3	69	male	#	#	yes	yes	Western and Central Europe	#	#	#	no	yes	yes	yes
EAC4	62	male	#	#	no	#	Western and Central Europe	#	#	Pantoprazole (Pantozol)	no	yes	no	no

The information cannot be obtained

Table S2

Table S2 the relative abundance of bacterial host at genus level for abundant, moderate, and rare bacteriophages

Abundant

Genus	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
Achromobacter	0.005513	0.002251	0	0	0	0
Actinomyces	0.000985	0.000402	0	0	0	0
Afipia	0.000203	8.27E-05	0	0	0	0
Alistipes	0.000203	8.27E-05	0.000167	6.81E-05	0	0
Anaerobutyricum	0	0	0.000381	0.000156	0.001385	0.000692
Anaerostipes	0.000522	0.000213	0	0	0.002077	0.001039
Anaerotruncus	0.005523	0.002255	0	0	0	0
Bacteroides	0.096228	0.022958	0.015158	0.003637	0	0
Bifidobacterium	0.009196	0.002928	0.01038	0.002435	0	0
Bilophila	0.004305	0.001297	0.000501	0.000204	0	0
Blautia	0.002611	0.001066	0.007486	0.001939	0.000692	0.000346
Butyrivibrio	0	0	0	0	0.016142	0.006738
Chryseobacterium	0	0	0	0	0.000896	0.000448
Citrobacter	0.018488	0.007548	0	0	0	0
Clostridioides	0.052532	0.009371	0.012846	0.002675	0.097294	0.048365
Clostridium	0.007708	0.002012	0.026737	0.007141	0.073707	0.029832
Coprococcus	0.001044	0.000426	0.001359	0.000431	0	0
Delftia	0.002757	0.001125	0	0	0	0
Dorea	0	0	0.001646	0.000471	0.001385	0.000692
Dysgonomonas	0.00104	0.000424	0	0	0.001344	0.000672
Enhydrobacter	0.000246	0.000101	0.00054	0.00022	0	0
Erysipelatoclostridium	0.015803	0.005469	0.002318	0.000748	0	0
Escherichia	0.002054	0.000839	0	0	0	0
Eubacterium	0	0	0.000226	9.21E-05	0.05205	0.01975
Faecalibacterium	0.017932	0.006612	0.011779	0.001895	0	0
Flavonifractor	0.011924	0.004147	0.032813	0.006469	0.004581	0.00229
Hungatella	0.001627	0.000459	0	0	0.005387	0.002367
Intestinibacter	0.001057	0.000432	0.00182	0.000743	0	0
Leptotrichia	0	0	0.000496	0.000203	0	0
Mycobacterium	0.000246	0.000101	0	0	0	0
Myroides	0.043255	0.013614	0.002975	0.000848	0.047275	0.01281
Paenibacillus	0.001105	0.000451	0	0	0	0
Parabacteroides	0	0	0.000802	0.000328	0	0
Paraprevotella	0.002148	0.000877	0.000428	0.000175	0	0
Parvimonas	0	0	0.000496	0.000203	0	0
Peptoniphilus	0.000753	0.000307	0	0	0	0
Prevotella	0.013396	0.005469	0	0	0	0
Pseudoflavonifractor	0.00151	0.000535	0.004021	0.000954	0	0
Pyramidobacter	0.000203	8.27E-05	0	0	0	0
Roseburia	0.000522	0.000213	0.001644	0.000453	0	0
Ruminococcus	0	0	0.002364	0.000965	0	0
Slackia	0	0	0.001213	0.000495	0	0
Staphylococcus	0.054111	0.013793	0.006357	0.001312	0	0
Streptococcus	0.000465	0.00019	0	0	0	0
Subdoligranulum	0.00652	0.002206	0.00707	0.001948	0	0
Tyzzerella	0	0	0.000381	0.000156	0	0

Moderate

Genus	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
Achromobacter	0.001555	0.000604	0.00038	0.000155	0	0
Actinomyces	3.64E-05	1.49E-05	0.00072	0.000213	4.26E-05	2.1315E-05
Afipia	0.000457	0.000167	0.001604	0.000292	0.001698	0.000540311
Alistipes	0.001364	0.000557	0.002105	0.000333	0.001302	0.000620306
Anaerobutyricum	0	0	0.000343	6.89E-05	0	0
Anaerostipes	4.85E-05	1.98E-05	5.91E-05	2.41E-05	4.05E-05	2.02353E-05
Anaerotruncus	0.001662	0.000434	0.001491	0.00018	0.003216	0.0008435
Bacteroides	0.005706	0.001253	0.010875	0.001713	0.002607	0.000756292
Bifidobacterium	0.00118	0.000427	0.004175	0.000806	0.001354	0.000589975
Bilophila	0.000375	8.87E-05	0.001007	0.000208	0	0
Blautia	0.001659	0.000356	0.001625	0.000233	0.000389	0.000194729
Butyricicoccus	0.000234	5.09E-05	0.00085	0.000193	0.001495	0.000484112
Butyrivibrio	0.00033	0.000135	0.000514	8.35E-05	9.01E-05	4.50465E-05
Chryseobacterium	0	0	0.000469	8.86E-05	0	0
Clostridioides	0.005933	0.000705	0.008535	0.001067	0.001595	0.000495816
Clostridium	0.004076	0.000779	0.004009	0.000566	0.005594	0.001049737
Collinsella	3.64E-05	1.49E-05	0.00053	9.84E-05	0.000385	0.000141486
Coprococcus	0.000636	0.00015	0.000757	0.000255	0.000265	0.000132648
Corynebacterium	0	0	8.81E-05	2.86E-05	0	0
Cutibacterium	5.46E-05	2.23E-05	0.000101	4.14E-05	0.000296	0.000147873
Desulfitobacterium	0.000429	0.000111	0.00028	0.000114	0	0
Desulfovibrio	3.61E-05	1.47E-05	0	0	0	0
Dorea	0.000226	9.23E-05	0.002669	0.000236	0.000187	9.35529E-05
Dysgonomonas	0.000312	0.000127	0.001274	0.000174	0.000927	0.000463669
Enhydrobacter	0.000187	6.65E-05	0.000271	5.59E-05	0.00128	4.22382E-05
Enterobacter	0.000102	4.15E-05	0	0	8.18E-05	4.09123E-05
Enterococcus	5.09E-05	2.08E-05	0	0	0	0
Erysipelatoclostridium	0.002125	0.000378	0.003204	0.000381	5.04E-05	2.51803E-05
Escherichia	0	0	0.000505	0.000118	0.000111	5.53899E-05
Eubacterium	7.19E-05	0	0.001364	0.000199	0.00147	0.00054096
Faecalibacterium	0.002773	0.000548	0.01378	0.001083	0.004101	0.000991461
Finegoldia	4.81E-05	1.97E-05	0	0	8.18E-05	4.09123E-05
Flavonifractor	0.003463	0.000632	0.00608	0.000703	0.002115	0.000766266
Gordonibacter	5.83E-05	2.38E-05	0.000385	7.63E-05	0.000206	6.28178E-05
Helcococcus	0	0	0	0	0.00025	0.000124972
Helicobacter	0	0	0.000177	7.21E-05	0.00036	0.000179915
Holdemania	0	0	0.000366	5.4E-05	0	0
Hungatella	0.00107	0.000259	0.003545	0.00057	0.001724	0.000463064
Intestinibacter	0.000188	5.66E-05	1.7E-05	6.94E-06	0.000107	3.85867E-05
Johnsonella	0	0	0.000233	4.49E-05	0	0
Klebsiella	0	0	0.000442	0.000133	0	0
Lactobacillus	0.000372	9.84E-05	0.003012	0.000458	0	0
Leuconostoc	0	0	2.79E-05	1.14E-05	0	0
Marvinbryantia	0	0	0	0	0.000255	9.5157E-05
Microbacterium	0	0	3.94E-05	1.03E-05	0	0
Myroides	0.005783	0.001181	0.005357	0.000886	0.003768	0.001337244
NAmissing	0	0	7.37E-05	2.21E-05	0	0
Olsenella	1.82E-05	7.43E-06	0.000233	5.43E-05	0	0
Paenibacillus	9.34E-05	2.85E-05	0.000176	3.74E-05	0	0
Parabacteroides	0.001548	0.00048	0.002335	0.000711	0.001193	0.000596287
Paraprevotella	0.00013	5.31E-05	0.000306	9.71E-05	2.73E-05	1.36681E-05
Prevotella	0.002053	0.000539	0.003988	0.000675	0.001736	0.000867816
Pseudoflavonifractor	0.000447	7.57E-05	0.000662	7.61E-05	0.000196	9.77577E-05
Pyramidobacter	7.62E-05	3.11E-05	0	0	0.000134	6.70518E-05
Raoultella	0	0	0.000147	4.42E-05	0	0
Rhodococcus	8.49E-05	3.46E-05	0.000466	8.76E-05	0.000841	0.000420573
Roseburia	0.000715	0.000178	0.001112	0.000228	0.000412	0.000205966
Ruminococcus	0.001135	0.000188	0.00081	0.000233	0.002961	0.001330326
Schaalia	1.82E-05	7.43E-06	0.000274	8.21E-05	7.39E-05	3.69682E-05
Segniliparus	0	0	0	0	8.16E-05	2.36014E-05
Selenomonas	8.52E-05	3.48E-05	0.000339	0.000123	0.000976	0.000417346
Slackia	0.000118	4.83E-05	0.00035	0.000143	4.26E-05	2.1315E-05
Sphingobacterium	0	0	3.39E-05	1.39E-05	0	0
Staphylococcus	0.001607	0.000257	0.002387	0.000318	0.000123	6.13685E-05
Streptococcus	5.09E-05	2.08E-05	0.000318	9.41E-05	0	0
Subdoligranulum	0.000366	7.7E-05	0.00087	0.00016	0.000746	0.000239469
Sutterella	0.000232	9.47E-05	0.000278	8.37E-05	0.001132	0.000565848
Treponema	6.28E-05	2.56E-05	0.00042	8.62E-05	0	0
Tyzzerella	0.000247	6.39E-05	0.001483	0.00019	0	0

Rare

Genus	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
Achromobacter	6.74E-05	2.06E-05	0.000164	5.54E-05	0.000258	8.03E-05
Acinetobacter	4.2E-06	1.7E-06	5.83E-07	2.38E-07	7.03E-06	2.5E-06
Actinomyces	0.000335	5.54E-05	0.000414	3.12E-05	1.55E-05	2.42E-06
Afipia	0.000168	3.81E-05	0.00038	7.64E-05	0.000245	6.69E-05
Alistipes	0.001236	0.000338	0.000703	6.87E-05	0.000432	9.81E-05
Anaerobutyric	7.87E-05	1.45E-05	4.44E-05	1.06E-05	4.34E-05	1.49E-05
Anaerofustis	1.65E-06	5.51E-07	5.37E-07	2.19E-07	4.07E-07	1.89E-07
Anaerostipes	1.2E-06	2.24E-07	2.83E-05	4.28E-06	6.16E-06	2.61E-06
Anaerotruncus	0.000656	0.000121	0.000595	6.04E-05	0.000874	0.000206
Bacteroides	0.003541	0.000645	0.003567	0.000144	0.002225	9.14E-05
Bifidobacteriu	0.000962	0.000198	0.00839	9.56E-05	0.000746	0.000152
Bilophila	0.000212	4.11E-05	0.000156	1.9E-05	0.00013	3.03E-05
Blautia	0.000278	4.98E-05	0.000451	4.29E-05	0.000769	0.000111
Butyricicoccus	0.000123	1.91E-05	0.000271	3.05E-05	0.000241	5.19E-05
Butyrivibrio	8.76E-05	1.63E-05	0.000134	1.59E-05	2.46E-05	7.25E-06
Chryseobacter	2.49E-05	4.1E-06	5.94E-05	7.26E-06	5.99E-05	2.2E-05
Citrobacter	1.18E-07	4.8E-08	0	0	0	0
Clostridioides	0.000761	6.12E-05	0.001861	8.98E-05	0.000837	0.00011
Clostridium	0.001765	0.00027	0.002313	0.000199	0.001217	0.000174
Collinsella	3.42E-05	6.81E-06	0.000113	2.04E-05	6.43E-05	1.07E-05
Coprococcus	8.99E-05	1.46E-05	0.000421	6.52E-05	0.000109	2.04E-05
Corynebacteri	1.26E-05	2.36E-06	2.63E-05	5.93E-06	2.2E-06	7.92E-07
Cutibacterium	1.45E-05	5.3E-06	2E-05	6.3E-06	1.06E-05	4.96E-06
Delftia	1.77E-07	5.97E-08	2.55E-06	7.38E-07	0	0
Desulfitobacte	4.75E-05	9.46E-06	0.000306	5.34E-05	0.000156	4.5E-05
Desulfovibrio	1.83E-05	6.5E-06	1.32E-05	3.59E-06	1.02E-07	2.33E-08
Dialister	9.62E-06	2.24E-06	3.52E-06	6.72E-07	2.24E-05	1.09E-05
Dorea	0.000322	7.65E-05	0.000586	5.3E-05	0.000425	6.68E-05
Dysgonomona	0.000111	2.79E-05	0.000236	2.75E-05	2.15E-05	7.48E-06
Enhydrobacter	9.18E-05	1.54E-05	0.000134	1.15E-05	2.83E-05	1.18E-05
Enterobacter	1.03E-05	3.24E-06	2.95E-05	7.28E-06	4E-06	1.17E-06
Enterococcus	4.5E-06	1.65E-06	8.82E-06	2.99E-06	1.1E-06	5.48E-07
Erysipelatoclo	0.000562	0.000131	0.000577	5.87E-05	0.000309	5.29E-05
Escherichia	0.000152	2.33E-05	3.21E-05	4.06E-06	0.000241	5.67E-05
Eubacterium	0.000245	5.28E-05	0.000506	4.68E-05	4.48E-05	9.83E-06
Faecalibacteri	0.004274	0.001059	0.003134	0.000224	0.002177	0.000307
Finegoldia	4.28E-06	8.98E-07	6.6E-06	1.19E-06	8.52E-06	3.45E-06
Flavonifractor	0.002843	0.000534	0.002667	0.000271	0.001301	0.000136
Fusobacterium	0	0	0	0	1.15E-05	5.73E-06
Gemella	4.6E-07	1.09E-07	2.81E-06	6.08E-06	3.47E-06	1.17E-06
Gordonibacter	0.000166	3.66E-05	0.000176	2.27E-05	5.78E-05	1.25E-05
Helcococcus	3.36E-07	8.05E-08	9.53E-06	3.73E-06	1.95E-08	9.75E-09
Helicobacter	0	0	3.93E-05	1.18E-05	2.64E-06	1.21E-06
Holdemanella	6.47E-09	1.68E-09	3.22E-06	1.31E-06	4.63E-06	2.32E-06
Holdemania	7.27E-05	1.76E-05	5.44E-06	1.65E-06	2.75E-07	1.27E-07
Hungatella	0.000622	6.75E-05	0.001377	0.000141	0.000918	0.000101
Intestinibacter	1.22E-05	2.13E-06	0.000121	2.64E-05	7.9E-05	2.34E-05
Johnsonella	4.25E-06	7.36E-07	1.47E-05	2.55E-06	1.12E-06	2.71E-07
Klebsiella	0	0	3.27E-05	9.89E-06	0	0
Lactobacillus	0.00028	8.8E-05	0.000146	2.27E-05	4.63E-05	7.19E-06
Lautropia	2.3E-06	7.47E-07	1.31E-06	3.49E-07	0	0
Leptotrichia	6.11E-07	1.41E-07	0	0	0	0
Leuconostoc	1.06E-05	2.5E-06	3.56E-06	8.97E-07	1.46E-06	4.95E-07
Marvinbryanti	0	0	1.45E-05	5.3E-06	4.12E-06	1.94E-06
Microbacteriu	6.58E-06	1.77E-06	2.95E-06	1.18E-06	1.7E-05	3.98E-06
Mitsuokella	1.83E-05	5.8E-06	2.14E-05	4.38E-06	1.83E-06	7.82E-07
Mycobacteriu	3.07E-05	5.92E-06	1.81E-05	7.33E-06	3.38E-08	1.69E-08
Myroides	0.001151	0.000141	0.002559	0.000129	0.001829	0.000237
NAmissing	0	0	5.44E-06	1.65E-06	0	0
Olsenella	2.49E-05	4.16E-06	6.4E-05	1.87E-05	2.09E-05	4.77E-06
Paenibacillus	2.76E-05	5.95E-06	3.89E-05	6.08E-06	3.73E-05	1.56E-05
Parabacteroid	0.000116	2.25E-05	0.000803	0.000118	0.000259	6.12E-05
Paraprevotella	0.000305	0.000103	0.000179	3.32E-05	5.52E-05	1.9E-05
Parvimonas	6.11E-07	1.41E-07	0	0	0	0
Peptoniphilus	8.24E-08	3.36E-08	0	0	5.68E-06	2.84E-06
Prevotella	0.001368	0.000489	0.000829	0.000122	0.000351	9.7E-05
Pseudoflavoni	0.000295	6.22E-05	0.000271	3.13E-05	0.000155	2E-05
Pyramidobact	2.01E-05	6.36E-06	5.12E-06	8.62E-07	1.56E-06	7.16E-07
Raoultella	6.28E-07	1.94E-07	1.68E-05	3.53E-06	9.03E-07	4.51E-07
Rhodococcus	6.66E-05	1.46E-05	0.00011	2.43E-05	0.000232	5.31E-05
Roseburia	0.000213	5.08E-05	0.000255	3.49E-05	0.000121	2.61E-05
Ruminococcus	0.000365	5.27E-05	0.000526	7.29E-05	0.000468	6.4E-05
Schaalia	2.87E-05	1.02E-05	5.34E-05	1.2E-05	3.79E-06	1.62E-06
Segniliparus	2.52E-05	6.58E-06	1.4E-05	4.52E-06	1.42E-07	7.12E-08
Selenomonas	0.000113	1.62E-05	0.000134	1.89E-05	0.000131	2.1E-05
Serratia	1.88E-06	5.83E-07	1.78E-05	6.61E-06	2.71E-06	1.35E-06
Slackia	1.49E-05	3.37E-06	3.67E-05	6.31E-06	2.43E-05	5.43E-06
Sphingobacter	3.36E-06	1.35E-06	1.78E-06	2.81E-06	2.35E-05	1.15E-05
Staphylococcu	0.001513	0.000442	0.000731	4.7E-05	0.000216	4.04E-05
Streptococcus	4.34E-05	4.6E-06	2.48E-05	3.68E-06	7.39E-06	2.02E-06
Subdoligranul	0.000689	0.000146	0.000537	2.73E-05	0.000252	2.06E-05
Sutterella	1.14E-05	9.49E-07	7.58E-05	1.18E-05	3.21E-05	1.55E-05
Treponema	0.000168	5.01E-05	0	0	5.9E-05	1.96E-05
Tyzzerella	8.08E-05	1.21E-05	0.000217	3.17E-05	6.68E-05	1.7E-05

Table S3 The percentage of contig relative abudance in different number of predicted bacterial genus types for abundnant, moderate, and rare bacteriophages

Abundant

Number of Genus types	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
1	0.563243	0.056394	0	0	0.46021	0.117223
2	0.271195	0.064494	0	0	0.289584	0.100959
3	0.027843	0.011367	1	0	0	0
4	0.137719	0.028148	0	0	0.250207	0.124951

Abundant

Number of Genus types	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
1	0.368621	0.037396	0.463691	0.04072	0.243545	0.034582
2	0.301105	0.020224	0.259292	0.033916	0.212674	0.04067
3	0.1634	0.038109	0.09526	0.01813	0.176374	0.052464
4	0.116392	0.014472	0.107201	0.016908	0.25085	0.081153
5	0.050482	0.007002	0.074556	0.009785	0.116557	0.034432

Rare

Number of Genus types	CT Mean	CT SE	BE Mean	BE SE	EAC Mean	EAC SE
1	0.504299047	0.033764	0.455566	0.021368	0.441091	0.055248
2	0.2717794	0.024402	0.267355	0.016591	0.23722	0.026469
3	0.113266187	0.003514	0.131134	0.010079	0.174656	0.016142
4	0.048951746	0.007087	0.082065	0.009509	0.068529	0.010761
5	0.03161361	0.005112	0.045175	0.003549	0.038503	0.010688
6	0.028121617	0.010699	0.017804	0.003084	0.025082	0.007284
7	0.001968393	0.000355	0.000902	0.000195	0.01492	0.007431

Table S4

	Table S4 The relative abundance of identified AMGs						
	Bacterial toxin				**LPS biosynthesis proteins**		
	tccC	entD	spyA	entB	lpxD	kdsC	gmhB
CT1	0	0	0	0	0	0	0
CT2	0	0	0	5.48E-06	0	0	0.000634
CT3	1.43E-06	0	6.2E-06	0	0	1.43E-05	0
CT4	0	0	0	0	0	0	0
CT5	0	0	0.00018	0	0	0	0.000286
CT6	0	0	4.55E-06	0	0	0	5.05E-07
BE1	0	0	3.66E-05	0	0	0	0
BE2	0	0	0	0	8.38E-07	0	0
BE3	0	0.000437	0.000707	0	0	0	7.71E-05
BE4	0	0	1.32E-05	0	0.001867	0	0
BE5	0	2.1E-05	6.3E-05	0	0	0.000532	0.001387
BE6	0	0	0.001561	2.52E-05	0	0	0
EAC1	0	0	3.66E-05	0	0	0	0
EAC2	4.6E-05	0	3.59E-05	1.84E-06	0	0	0
EAC3	0	0	0.000873	0	0	0	0.002371
EAC4	0	1.7E-05	0.009725	0	0	0	0.000204

Figure S1
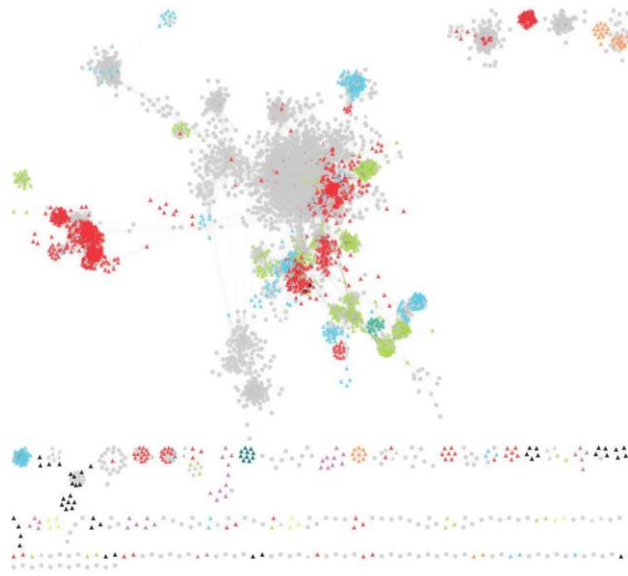
Figure S2
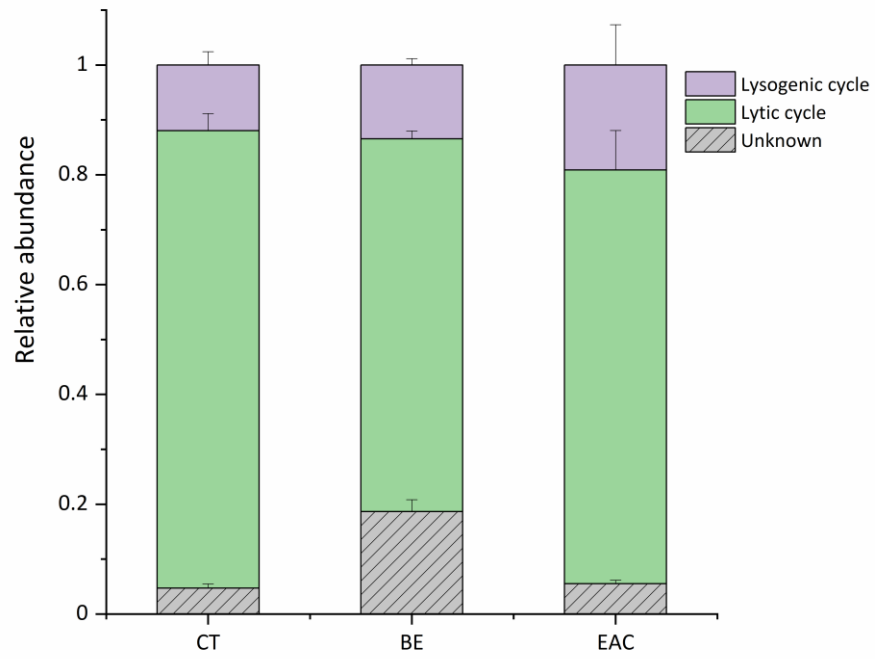
Figure S3
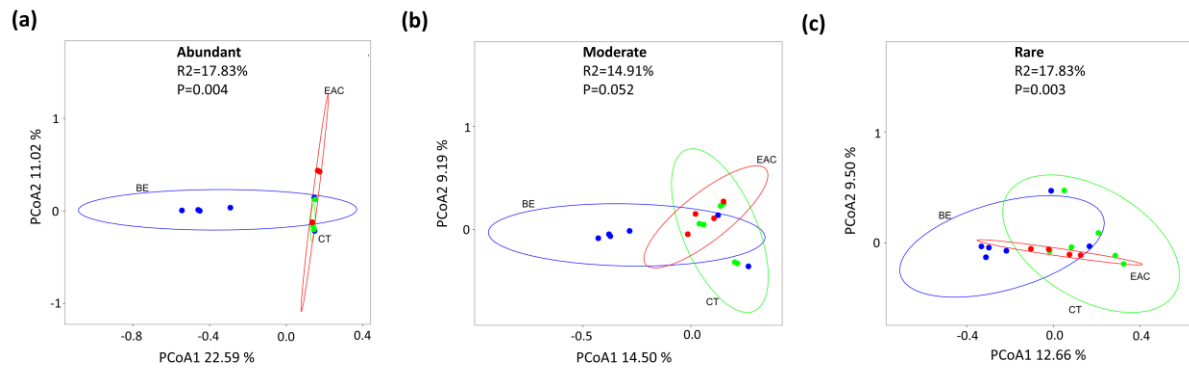
Title:

Depth-specific virome reveals the survival strategy of aquifer microbiota adapting to the oligotrophic environment

Tianli Ma[1,2], Jinling Xue[1,2], Jinlong Ru[1,2], Li Deng[1,2,*]

[1] Institute of Virology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[2] Institute of Virology, Technical University of München, Munich, Germany

Corresponding author:

Li Deng, PhD

Institute of Virology, Helmholtz Zentrum München,

German Research Center for Environmental Health,

Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

Phone: +49 89 3187 49195

e-mail: li.deng@helmholtz-muenchen.de

Conflicts of interest

The authors disclose no conflicts.

Abstract:

Viruses (dominated by bacteriophages) play critical roles in driving the development of bacterial community and the circulation of elements in biosphere. However, the exploration of viral communities in deep groundwater, an important part of the deep biosphere, is limited. With the help of metagenomic sequencing, the present study revealed the structure and behavior of viral communities in deep groundwater. Environmental variation caused by increasing groundwater depth showed a distinct preference for the component of core viral groups. Whereas viruses in deep groundwater have a consistent tendency to switch to lysogenic life cycle. In addition, abundant viruses that perform main ecological functions dominated in deeper aquifer, while rare viruses that enhance ecological resilience dominated in upper aquifer. Furthermore, viral traits associated with auxiliary metabolic genes (AMGs) appeared depth-specific patterns. AMGs related to methane metabolism (*cofF*), sulfur metabolism (*cysH*), and quorum sensing (*LuxR* and *spo0A*) were found in the genome of viruses, and their distribution at different depths correlated with environmental characteristics and bacterial metabolism there. Considering former studies on microbiome, abundant viruses may contribute to the stability of simple ecosystems by keeping a "Red Queen coevolution" relationship with abundant bacteria, while rare viruses may aid the resilience of complex ecosystems by maintaining a temperate coexistence relationship with rare bacteria. Moreover, viral AMGs further support the adaptation of bacterial hosts to the environment of different depths and jointly prompt the biogeochemical cycle of the ecosystem. Collectively, viruses performed specific survival strategies at different depths of deep groundwater.

Introduction

Deep groundwater is characterized by less photosynthesis and nutrient input, and complex surrounding hydrochemical environment, thus forming an oligotrophic and spatially heterogeneous deep biosphere (1). With the recent advancement of deep groundwater sampling as well as the microbial molecular research technology, microbial life detected has extended to a depth of several kilometers in groundwater (2-4), which has extensively motivated the exploration of microbial communities in deep groundwater and greatly expanded our knowledge of the deep biosphere (5-8). Complex hydrochemical characteristics and oligotrophic conditions of deep groundwater prompt the microbiota living there to optimize survival strategies to adapt to this setting, contributing to the biogeochemical cycle of the deep biosphere (1). To date, studies have shown that dominant bacterial communities differed under various environmental conditions of different depth groundwater, such as sulfate-reducing bacteria was found to be abundant in the area with higher sulfide concentration in deep granitic groundwater (9, 10). Furthermore, bacteria have evolved various physiological characteristics to regulate their metabolic mechanisms to achieve effective nutrient utilization strategies (11). It is reported that carbon source levels and types in different depths of groundwater have induced diverse carbon assimilation mechanisms of bacteria (9, 12, 13). Viruses, as the most abundant and genetically diverse components of microbial communities in the environment, their survival strategies could directly affect bacterial community and physiology, and further, affect ecological functions of the deep biosphere.

Generally, viral survival strategy in deep biosphere mainly depends on their interplay with bacterial hosts (14-16). Upon the lytic cycle, viruses enter the productive state through lysing bacterial cells, further controlling the size of bacterial populations and the dynamics of bacterial communities (17). In hydraulically fractured deep wells, viruses performed top-down control by reducing the density of the dominant bacteria *Halanaerobium*. Meanwhile, the bottom-up controls were exerted by releasing intracellular metabolites of lysed bacterial cells, which supported the metabolism of other microbiota in this ecosystem (6, 18). By contrast, upon the lysogenic cycle, viruses enter the dormant state by integrating their genome into the bacterial chromosome, thereby aiding bacterial gene diversity and fitness (19, 20). It is believed that the viral lysogenic cycle is beneficial under extreme conditions such as low productivity, high temperature, low pH, which could increase the survival chance of viruses and bacteria (21-23). Apart from the viral life cycle, the response of viral functional traits to environmental changes also reflects survival strategy of

viruses. Under different depths of mine tailings, the identified viral communities with diverse taxa exhibit distinct functional profiles (23-25). When the gene profiles of viral community were deeply analyzed, auxiliary metabolic genes (AMGs) were found in many viral genomes which were acquired from their bacterial host during the infection course (20, 24). In sulfur-rich and oxygen-depleting deep sea, many AMGs related to sulfur metabolism, such as *rdsr*, were identified in the viral genome, expanding bacterial metabolism of utilizing sulfate and in turn benefiting the propagation of viruses (23, 26, 27). It implies that viruses act as a genetic reservoir to help the sulfur cycle. All these aspects indicated that the survival strategy of viruses is conducive to the survival and reproduction of viruses and their bacterial hosts, potentially facilitating nutrient turnover and energy flow in deep biosphere. However, as the most important component of deep biosphere, less is know about viral community of deep groundwater.

Äspö Hard Rock Laboratory (Äspö HRL), which located on the island of Äspö near Oskarshamn, Sweden, comprises a 3.6km long tunnel that spirals down to a depth of 460m in a granitic bedrock (28). The geological formation of this area could track back 1.6 to 3.1 billion years old. Frequent geological activities have caused bedrock fracture events, allowing fluids of different Quaternary ages (such as marine, brackish or glacial water) to penetrate these fractures, thereby forming various aquifers (29, 30). The geological and hydrochemical data here have been well documented (31, 32). In the study of bacteria in this area, it has been pointed out that as the depth increased, the dominant bacteria changed and the bacterial diversity decreased (33), whereas the supply of nutrients was the crucial abiotic factor affecting the bacterial community patterns here (33, 34). Although viruses life have been observed by transmission electron microscopy (28), an in-depth and comprehensive investigation should be carried out to reveal survival strategy of viral community and their possible role in shaping bacterial communities and metabolic processes in the deep groundwater of Äspö HRL, which could further explain their contribution in maintaining the stability and function of deep groundwater ecosystem and supplement the knowledge of deep biosphere. To achieve this, in the present study, groundwater from different depths were collected in Äspö HRL and metagenomic sequencing was used to gain deeper insight into the changes of virome community. Firstly, the viral life cycle, functional viral groups (core viral groups, abundant viral group, rare viral group), and viral physiology characteristics were evaluated under environmental conditions of different depths. As the groundwater was directly or indirectly disturbed from the recharge area, the migrant viruses were exclusively predicted from

viral community to indicate the interferences of different depths and further infers their possible role in functional viral groups. Finally, combined with the predicted bacterial host communities, speculating the potential relationship between viruses and bacteria.

Methods

Sample collection

Three borehole sites of the Äspö HRL tunnel were conducted for sampling: 507B (namely Shallow Aquifer, short SA, 71m), which aged from months to years (in other words, water turnover time) (35), 1327B (namely Intermediate Aquifer, short IA,196m), which is more than 20 years old (36), TASF (namely Deep Aquifer, short DA, 450m), which is approximately 7 000 years old (31, 36). Additionally, Baltic Sea water (BS) was also collected because of the great influence on the groundwater aquifer (37). After collected 2L of groundwater in each sampling site, they were transported to laboratory and stored at 4 °C in dark immediately.

Concentration of groundwater viruses

The groundwater samples were first pre-filtered through 0.22 µm filters (PES Membrane, Merck Millipore, Lot No. ROCB29300, Ireland) to remove most eukaryotic algae and particle-associated bacteria. Then the viruses in the groundwater were concentrated by iron-based flocculation as shown in previous description with minor modifications (38). Briefly, after the addition of 20 mg Fe L^{-1} to the groundwater, the mixture was incubated for 1 hour and rigorously vortex during the incubation time. Subsequently, the mixture was filtered through 0.22 µm filters (PES Membrane, Merck Millipore, Lot No. ROCB29300, Ireland), remaining Fe-virus aggregates in the filter surface. Then Fe-virus aggregates were resuspended in Ascorbate-EDTA buffer (38) until dissolved completely. Finally, the solution was concentrated to less than 50 µL by Amicon® Ultra Centrifugal Filters (10kDA, Merck Millipore, Lot No. R9EA18187, Ireland) for further virome DNA extraction.

Virome DNA extraction

Virome DNA extraction method referred to former study (39). In brief, chloroform was first used to pre-treat samples, retaining the upper phase followed by adding DNase I (1U/ µL, Invitrogen, USA, Lot No. 1158858), lysis buffer (700µL KOH stock (0.43 g/10 ml), 430 µL DDT stock (0.8 g/10 ml), 370 µL H_2O, pH=12), and Proteinase K (20 mg/ml, Invitrogen, USA, Lot No. 1112907) for virome DNA purification. Afterwards, AMPure beads (Agencourt, Beckman Coulter, USA)

were added to adsorb DNA and the DNA was then eluted from beads by 35 µL Tris buffer (10 mM, pH=9.8). The obtained virome DNA was stored at -80°C until further sequencing, which was performed on an Illumina HiSeq-PE150 platform.

Bioinformatic analysis

11 773 140, 11 624 576, 11 607 046, 7 430 874 raw reads were obtained from BS, SA, IA, and DA, respectively. Detailed analysis methods and software used can be referred to former study (39). In brief, the low-quality bases and adaptors were first removed, followed by discarding the duplicated reads. The clean reads (11 386 580, 11 309 670, 11 311 258, 7 294 188 from BS, SA, IA, and DA, respectively) were assembled into contigs and only contigs longer than 1kb were retained. Afterward, the redundancy contigs were removed and the remaining contigs were used to predict viral contigs by the combination of VirSorter, CAT, and DeepVirFinder. Subsequently, the viral contigs were clustered and the longest contigs in each cluster remained as a representative contig, which was used to predicted ORFs for the taxonomy annotation through the combination of vConTACT2, CAT, and Demovir script. Next, the clean reads from each sample were mapped to the viral contigs and estimated the number of reads that mapped to each gene for further calculating the relative abundance of viruses in each sample. Viral proteins predicted in the previous step were fed into VIBRANT to identify the lytic/lysogenic cycle and annotated the function by searching viral protein against KEGG, VOGDB, and PFAM databases. Meanwhile, AMGs were screened and BLAST-based homology searches with viral references were conducted to identify the most related microbe and/or virus encoding similar metabolic and viral-like gens to validate the AMGs were encoded on viral sequence and functional annotation were correct. The raw sequencing data are deposited in NCBI Sequence Read Archive (SRA) under accession number SUB9050817.

Data analysis

Viral sources were predicted using Integrated MICROBIAL GENOME/VIRUS (IMG.VR). The database mainly includes three sources: environmental sources (terrestrial, sediment, marine, and freshwater), engineered sources (wastewater, solid waste, lab enrichment, built environment, and bioreactor), and host-associated sources (human, plants, and other). The virus (phage)-bacteria (host) interactions were predicted by VirHostMatcher-Net (40). In the present study, only predicted bacterial hosts with scores higher than 95% were remained. Rarefaction curves were

built for the viral richness (Ace) and diversity (Shannon), with 20000 sequences as a sampling depth. Phylogenetic tree was generated using MEGA X (41) based on the Maximum Likelihood method (42). The bootstrap consensus tree inferred from 1000 replicates (43) is taken to represent the evolutionary history of the taxa analyzed (44). The Interactive Tree of Life (https://itol.embl.de/) was used for phylogenetic tree display. The promoters in genome of viral contigs were predicted using the BPROM software (45, 46). Venn diagram and Heatmap were generated by R (Version 3.3.3) with packages VennDiagram (Version 1.6.20) and pheatmap (Version 1.0.12). All Histograms and line chart were generated in Origin 2020b. Pie chart were generated in Microsoft Excel.

Results and Discussion

Responses of overall virome structure to deep groundwater environment at different depths

Generally, 4 481, 2 238, 645, and 520 viral contigs were assembled in BS, SA, IA, and DA, respectively. During the processing, 64 viral contigs in DA were predicted to come from accidental human contamination, which may be caused by human activities in the laboratory here. In further analysis, these 64 viral contigs were excluded, remaining 456 viral contigs in DA. After taxonomy annotation, more than 60% of viral contigs were assigned to bacteriophages in four different depths (**Fig. 1a**). Among them, the order of *Caudovirales* (28.01% - 49.86%) was the main classified bacteriophages and most viral contigs belonged to families *Siphoviridae* (14.23%-39.23%), *Myoviridae* (1.97%-9.35%), and *Podoviridae* (4.16%-10.81%). In addition to *Caudovirales*, *Tectiviridae* has a relatively higher relative abundance (1.65%) in IA compared with BS, SA and DA. In-depth analysis about the virome community showed that the viral richness and diversity decreased with increasing depth (**Fig. 1b**). Among all viral contigs, 69 of viral contigs were shared by four different depths, and each group harbored its unique viral contigs (**Fig. 1c**). Furthermore, most viruses were predicted to sustain the lytic replication cycle (**Fig. 1d**). The ratio of lysogenic/lytic cycle rose as the depth increased, and only dropped in IA (**Fig. 1d**). Due to the proximity to recharge area and infiltration across geographic barriers, the identified viruses in four different depths consisted of original viruses and migrant viruses. According to the results of viral source prediction, the relative abundance of migrant viruses showed a downward trend with increasing depth (**Fig. 1e**), which coincided with the less external input in the deeper aquifer,

representing less interference and a relatively stable environment. Interestingly, the viruses from the engineered source had a higher percentage in SA (7.45%) and IA (4.91%) than those in BS (0.80%), the latter being imported with more migrant viruses. This may be attributed to the rapid water turnover, which allowed the viruses to spread and dilute, leading to a lower concentration in BS. In contrast, the slow water turnover allowed the viruses to accumulate in SA and IA. Through predicting the bacterial host range of viral contigs, it was found that bacterial hosts spanned fewer phyla with the increasing depth (**Fig. 1f, Table S1**). In four different depths, most viral contigs were predicted to infect Firmicutes (5.5%-20.86%, 0 in DA), Bacteroidetes (2.73%-16.27%), Proteobacteria (5.21%-32.58%), and Actinobacteria (<1%-5.14%). Less than 1% of viral contigs could infect other bacterial phyla. Only the viral contigs in SA accounted for a higher percentage (1.78%) to infect Acidobacteria.

Overall, depth-specific virome structure patterns were found, which may be caused by heterogeneous hydrochemical environments and distinct bacterial communities in four different depths. Moreover, according to studies on sulfidic mine tailings and freshwater, environmental factors may contribute more to the virome structure than bacterial communities (23, 47). Besides, the viral alpha diversity decreased as the depth increased. Obviously, the reduced supply of nutrients was the crucial abiotic factor promoting the decline of viral diversity here (33, 34). Meanwhile, fewer bacteria are supported to survive also further limit the viral proliferation (1). Although changes in virome structure have been observed, it is necessary to in-depth reveal the inherent survival strategy of viruses from the entire, partial, and individual levels to adapting different environmental characteristics.

Changes of core viral group structure and viral life cycle at different depths

In similar habitats, the difference of microbial communities induced by geographical variation was mainly due to environmental factors such as temperature, dissolved oxygen concentrations, and dissolved organic carbon (48). However, the core microbiome represents stable and consistent components across different microbial communities within the same type of habitat (49). In the current study, although alterations of overall virome structure were found with increasing depth, the determination of core viral group is of great significance for understanding the stable fraction and the subsequent studying of their function in groundwater. Here, the core viral group was defined as the overlapping areas of circles in Venn diagrams (50). A total of 102 viral contigs were

shared by three depths of deep groundwater (SA, IA, and DA) (**Fig. 1d, Fig. S1**) and were identified as the core viral group. The number of viral contigs belonging to the core viral group accounted for 4.55%, 15.79%, and 22.32% in SA, IA, and DA, and corresponding relative abundance were 21.72%, 24.87%, and 33.38%. Although there was no advantage in numbers (such as SA), the relative abundance of core viral group represented more than 20% of the overall viral community and increased with increasing depth, indicating they may have traits compatible with oligotrophic and low energy environments, thereby enhancing their adaptability.

Moreover, the core viral group was assigned to the classified order of *Caudovirales* including *Myoviridae*, *Podoviridae*, *Siphoviridae,* and Unclassified *Caudovirales* (**Fig. 2a**), which were the major families in the overall virome structure (**Fig. 1a**). Three depths had different preferences for the components of core virus group. Specifically, In SA, *Myoviridae* (21.88%), *Siphoviridae* (19.77%) and Unclassified *Caudovirales* (17.46%) were dominant. In IA, *Siphoviridae* (17.39%) and Unclassified *Caudovirales* (19.11%) were dominant. *Siphoviridae* (50.10%) was the only dominant viral family in DA. As the depth increased, the reduction of dominant components of the core viral group indicated that the selection of preferred components was more stringent, which was mainly determined by microbial niche driven by environmental factors (51). Intensified environmental constraints led to a reduction in the available niche dimensions, which further controls the colonization and reproduction of microbiota in the deeper groundwater (51). In a hillside flow system study about the spatial differences of microbial communities in aquifers at different depths (34), it was also found that environmental factors, the availability of oxygen and nitrogen compounds, were determinants of the core bacterial operational taxonomic units (OTUs). Meanwhile, the abundant bacterial OTUs were predominance in the core microbiota. Therefore, this study hypothesized abundant viruses (relative abundance was more than 1% in total contigs) were also dominant in the core viral group. The result showed that although numbers of abundant viral contigs (4-7) were low, their corresponding relative abundance was rather high (36.37%-93.47%) (**Fig. 2b**). This is because the core virus group is more competitive than other viruses, and therefore has more opportunities to survive and prosper. The gradually increased relative abundance in deeper groundwater suggested this competitiveness is more advantageous in a harsher environment.

Besides, viruses could change their replication cycle to adapt to environmental variations such as the trophic state. In this study, the proportion of lysogenic cycle in oligotrophic environments (SA, IA, DA) was higher than that in the nutrient-rich environment (BS) (**Fig. 1d**), which may indicate that viruses in oligotrophic environment tend to the lysogenic replication cycle. A previous study also reported that the lysogenic replication cycle increased during periods of low inorganic nitrogen and phosphate concentration in offshore water (52). Generally, lower nutrients accompanied by low bacterial densities promote phage selection for lysogenic replication cycle, which in turn increased the survival chances of bacteria and viruses, enabling them to thrive under appropriate conditions (53, 54). Moreover, the lysogenic cycle of viruses is beneficial to the competitiveness of bacterial cells (lysogens) under oligotrophic conditions, because lysogens prevent them from being infected by homologous phages, and meanwhile, they could utilize the contents of other lysed cells to maintain the metabolism of themselves (54).

From the perspective of the entire viral community, the core viral group occupied a major position at different depths with relatively higher proportions. However, depth-specific environmental characteristics induced changes in microbial niche dimensions, leading to differences of dominant components in core viral group at different depths. Moreover, the relative abundance of abundant viruses was higher in core viral group and increased with depth, indicating components of core viral group have the best competition in deeper aquifers with relatively harsh environments. Meanwhile, the viral replication cycle tended to shift to the lysogenic life cycle, suggesting the lysogenic life cycle is another strategy that beneficial to the survival and propagation of viruses in oligotrophic environment.

Abundant viral groups dominate in stable deeper aquifer while rare viral groups dominate in fluctuating upper aquifer

Abundant and rare microbiota are critical parts of microbial communities and play different roles in maintaining the function and stability of ecosystems (55, 56). The abundant microbiota is thought to be well adapted to their resident environment and makes the greatest contribution to the biogeochemical functions of ecosystems (55). As shown in **Fig. 3a**, in four different depths, there were 2-14 of viral contigs belonging to abundant viral groups and their relative abundance accounted for 4.03%-71.30%. After screening the taxonomy of abundant viral groups (**Fig. 3b**), they were assigned to classified families including *Siphoviridae, Myoviridae, Podoviridae, and*

Tectiviridae, which were belonged to major classified taxonomy in this study (**Fig. 1a**). Viral life cycle prediction suggested that in BA and IA, abundant viruses sustained lytic cycle (**Fig. 3c**), while in SA, three viral contigs sustained lysogenic cycle and their proportion of relative abundance exceeded viruses that sustained lytic cycle, and in DA, one viral contig sustained lysogenic cycle and had a lower proportion of relative abundance than those sustained lytic cycle (**Fig. 3c**). On average, the lytic replication cycle was dominant for abundant viruses. Viral source prediction results showed abundant viruses included original and migrant viruses (**Fig. 3d)**, which may suggest that migrant viruses also affected the main ecosystem functions of deep groundwater. Consistently, in an activated sludge ecosystem, abundant migrant bacteria were also functionally characterized and act on biogeochemical processes (such as carbon turnover) of the activated sludge (57). The coexistence of abundant original and migrant viruses may suggest that they jointly influence the main ecosystem functions of groundwater due to the long-term impact of recharge area. Longitudinally, it was obvious that the number of abundant viral contigs and their relative abundance were ascending with the increasing depths (**Fig. 3a**). Only in DA, the number of abundant viral contigs was slightly lower than those in IA. This might be because the highly variable environment of the upper aquifer is a challenge for selecting tolerant and well-adapted viruses to develop to be abundant population, while the bottom aquifer experienced lower disturbance, supplying a relatively stable environment to ensure some viruses can flourish to an abundant population (34). The result of virome structure was similar to bacterial pattern in groundwater in which the abundant bacteria were also dominant in the deeper aquifer (34), which could also directly prompt the reproduction of viruses. Surprisingly, abundant viruses were found to be unique in four different depths (**Fig. 3e**), showing that relatively independent abundant viral groups were formed at different depths. After screening the predicted bacterial hosts of abundant viruses, the result showed that bacterial hosts were almost unique in different depths (**Fig. 3f**). Generally, the bacterial hosts of abundant viruses spanned four phyla, including Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. But there was distinction between depths, especially at the genus level. The great variations of abundant viral groups and their bacterial hosts may imply that these four deep groundwater ecosystems provide different ecological niches for viruses inhabiting here, resulting in the emergence and reproduction of viruses that are differentiated but most suitable local ecological niches. Overall, abundant viral groups were found to mainly contribute to the stability of simple ecosystems. Moreover, considering the previous studies on

bacteria, the consistent changes in the relative abundance of abundant viruses and abundant bacteria may indicate that they sustain the model of Red Queen coevolution, where virus (phage)-bacteria (host) relationship was maintained by counter-adaptation (58, 59). The dominant lytic cycle of abundant viruses in current study also supported the hypothesis. Former research also reported that the abundant bacteria and abundant viruses persist overtime to maintain the stability of ecosystem in the stable environment (60).

Contrary to abundant microbiota, rare microbiota is characterized by low abundance and various genotypes. However, rare microbiota could become abundant to respond sudden environmental change (56, 61), reflecting the ability of microbial community to recover their original composition after being disturbed, that is resilience of microbial community (62). Thus, rare microbiota play an important role in restructuring microbial communities and restoring ecosystem functions. In the current study, there were 395-4379 of rare viral contigs (the relative abundance of each rare viral contig was less than 0.1% in total contigs) and their relative abundance accounted for 5.62%-76.33% in four different depths (**Fig. 4a**). Compared with abundant viral groups, the rare viruses were assigned to more taxonomy (**Fig. 4b**) including rare taxa, such as *Herelleviridae, Ackermannviridae, Lavidaviridae, Marseilleviridae, Mimiviridae, Inoviridae, and Phycodnaviridae*, which means that the rare viruses have more genotypes. Besides, although lytic cycle was still dominant in rare viruses (**Fig. 4c**), the overall proportion of lysogenic cycle was slightly higher than that of abundant viruses (**Fig. S2**). The possible explanation is that rare viruses have the potential to sustain the lysogenic cycle. Among those rare viruses, apart from original viruses (39.23%- 63.02%), most migrant viruses come from environmental sources (25.16% - 58.90%) belonging to aquatic ecosystems (**Fig. 4d**), which may be a hint that once the ecosystem is disturbed, rare viruses will serve as a backup force to assist the recovery of ecosystem functions of groundwater which also belong to aquatic ecosystem. Generally, rare viruses contain more genotypes and potentially have more function, which helps to restore the virome community when disturbed and further recover the ecological function of groundwater. Longitudinally, the number of rare viral contigs and their relative abundance decreased with increasing depth (**Fig. 4a**). Only in DA, the number of contigs and relative abundance have a slight increment compared with IA. The higher proportion of rare viruses in upper aquifer may be due to the greater disturbance of upper aquifer associated with surface recharge area, resulting in higher viral diversity (63). Moreover, more rare viruses also reflect the increased potential and resilience to respond to the

changeable environment (64) based on the characteristics of high dispersal and low loss rates of rare microbiota (62). Furthermore, the shared and unique rare viruses were found among four different depths, and BS contained more unique rare viruses, followed by SA, IA, and DA (**Fig. 4d**). The predicted bacterial hosts of rare viruses spanned 22 bacterial phyla in four depths, and the diversity of bacterial hosts decreased with the increasing depth (**Fig. 4f**). The more complex bacterial hosts of rare viruses than abundant viruses further confirmed the function of rare viruses existed in ecosystems as 'seed bank' (65), especially in complex ecosystems, through obtaining and transferring genes from their broad bacterial host to assist bacteria to behave sudden environmental challenges, further enhancing the resilience of complex ecosystems. Normally, the rare bacteria would not be lysed by viruses (66). Although the rare viruses in this study mainly maintain the lytic cycle, compared with abundant viruses, the increasing trend of lysogenic cycle leads us to propose that the rare viruses gradually tend to coexist with bacteria. Collectively, the changes in the proportion of abundant viruses and rare viruses at different depths are also survival strategies of virome communities to adapt to environmental variation.

In general, abundant viruses dominated in deeper aquifers while rare viruses dominated in upper aquifer, which may suggest that abundant viruses contribute to maintain the stability of stable simple ecosystems, while rare viruses aid to enhance the resilience of complex ecosystems. Considering the bacterial characteristics in previous studies, it was assumed that abundant viruses and abundant bacteria prone to keep a "Red Queen coevolution" relationship, while rare viruses and rare bacteria prefer to maintain a temperate coexistence relationship.

Viral AMGs benefit the fitness of their bacterial hosts under environmental variations

In addition to changes in functional viral groups, individual viral gene profiles also varied under different environmental conditions. Previous studies have reported that viral traits associated with auxiliary metabolic genes (AMGs) are closely related to their surrounding environmental characteristics and bacterial metabolism (67, 68). For example, in deep ocean or sulfide mine tailings where sulfur is widely distributed, AMGs related to sulfur metabolism, such as *rdsr,* was found in the genome of some viruses, which could further help surrounding sulfur metabolizing bacteria, thereby influencing the sulfur cycle in biosphere (23, 26). Thus, it was hypothesized that viral AMGs appeared distinct patterns in different depths. Through investigating viral gene profiles, depth-specific AMGs were found in the genome of viruses including *cofF*, *cysH*, *LuxR* and *spo0A*

(**Fig. 5a**). Specifically, *cofF* encodes the alpha-L-glutamate ligases of methanogenic coenzyme F420 (69), which is involved in methane metabolism. The gene *cofF* only existed in the genome of viruses in BS (0.018%) and SA (0.0015%) (**Fig. 5b**). Then the subsequent phylogenetic analysis was conducted to explore the origin of *cofF*. The *cofF* in this study was relatively far away from those in the genome of bacterial strains and isolated phages in terms of evolutionary relationship and has less support for those existed in the genome of other known viral contigs in IMR database, which may imply that *cofF* in current study has been acquired for a long time (**Fig. 5c**). However, it has been remained under long-term natural selection, indicating that in the process of evolution, it was beneficial to the survival of bacterial hosts and their methane metabolism. The higher level of *cofF* appeared in the viral genome of BS which belongs to an aerobic environment and provides habitat of aerobic bacteria, suggesting that viruses could help the metabolism of the aerobic methanogens living here (70). *CysH* encodes phosphoadenosine phosphosulfate reductase that contributes to assimilatory sulfate reduction (71). Multiple copies of *cysH* were observed and existed in the genome of viruses in SA (2.0%) and DA (17.0%) (**Fig 5b**). Moreover, C*ysHs* in this study were close to those in the genome of bacterial strains, which indicated that multiple HGT occurred between different bacterial strains (**Fig. 5c**). A higher level of *cysH* was found in the genome of viruses of DA, where sulfur bedrock is most widely distributed, indicating virus-carrying *cysH* genes could assist the bacterial sulfur metabolism and multiple HGT occurring may imply that the function of sulfur metabolism was further expanded. *LuxR* and *spo0A* are genes related to the quorum sensing. *LuxR* represents a family of transcriptional regulators that affect microbial behaviour such as bioluminescence, virulence gene expression, etc. (72). It was found that *LuxR* mainly appeared in the genome of viruses in IA (0.19%) (**Fig. 5b**). In this study, *LuxR* was rather closer to those in the genome of bacterial strains compared with those in the genome of previously isolated phage (**Fig. 5c**). It suggested that the HGT mainly occurred in bacterial strains. *Spo0A* encodes stage 0 sporulation protein A, which is a response regulator belonging to two-component system (73). It was mainly found in the genome of viruses in IA (0.52%), followed by DA (0.37%) and SA (0.16%) (**Fig. 5b**). The phylogenetic analysis showed *spo0A* in this study was really like an outgroup (**Fig. 5c**), implying that *spo0A* was acquired early, and seems to maintain its function under long-term natural selection. *Spo0A* and *LuxR* appeared in the genome of viruses in SA, IA, and DA, where belong to the nutrient-poor environments and bacterial competition for resources is fierce. The AMGs, *LuxR* and *spo0A*, could help bacteria to regulate the bacterial

population density to improve the survival ability of the bacterial population in the fragile and oligotrophic environment (74). Taken together, the difference between the main AMGs at different depths was closely related to the environmental factors there. It could also potentially explain the metabolic types and survival strategies of bacteria living there.

Apart from depth-specific AMGs, different HGT levels of these four genes revealed by evolutionary origin analysis (**Fig. 5c**) may also be explained by local environmental factors. It has been reported that HGT occurs across bacteria strains with similar factors such as oxygen tolerance (75). In the current study, multiple HGTs appeared in genes (*cysH*) related to sulfur metabolism. This may be due to the fact that sulfur acts as an elector acceptor to maintain the fundamental survival metabolism of bacteria in an anoxic environment, allowing bacteria to gain adaptability in an oxygen-deficient environment (76). Besides, microbial density also affects the occurrence of HGT (77). In the oligotrophic environment with low microbial density, lower HGT or no HGT were found (*LuxR* or *spo0A*). This is because decreased contact possibility of microbiota result in a reduced chance of HGT occurrence. Moreover, HGT does not confer a selection advantage in a stable environment without environmental stress (78), where the lower competition may lead to slow evolution speed of microbiota. Thus, evolutionary analysis showed that the gene (*spo0A*) that have acquired earlier still exist without being eliminated. In contrast, HGT could promote the faster adaptation of microbiota in fluctuation environments such as BS (79). However, it was found that *cofF* appeared in BS was obtained earlier and HGT did not occur. It may be because *cofF* related to the methanogens was found in an aerobic environment. However, most methanogens are anaerobic bacteria and exist in an anaerobic environment (80). Therefore, and aerobic environment has no selective advantage for *cofF*. Furthermore, recent studies suggested that HGT may appear detrimental to microbiota in fluctuating environments (81) since the adaptation speed are more important than microbial fitness under such situation, thus lower rates of HGT tend to be preferred (81). This may also result in the HGT of *cofF* not being found. Collectively, the acquisition of viral AMGs and the HGT of these genes are potentially affected by surrounding environment.

Generally, depth-specific AMGs of viruses play critical roles in genetic variation of bacteria, which could aid bacterial strains to acquire adaptative traits to different environmental characteristics and further expand their ecological niche (82), in turn, increasing the chances of

viral gene replication. Meanwhile, the acquisition of AMGs promotes viruses as a key agent to participate in the biogeochemical cycle of deep groundwater.

Conclusion

Overall, this study revealed the change of viral community at different groundwater depths and suggested the possible strategies of virus communities adapting to environmental variations. At the whole level, the relative abundance of the identified core viral groups exceeded 20%, indicating their significant central position in the groundwater ecosystem, and environmental variation influences the component preference of the core viral group. Meanwhile, compared with the nutrient-rich environment, temperate viruses have a higher percentage, implying temperate viruses were more adapted to the oligotrophic environment than virulence viruses. On the partial level, the abundant viruses and rare viruses in the viral community also appeared with different characteristics at different depths. As the depth increased, both the number of viral contigs and relative abundance of abundant increased, while rare viruses decreased. Abundant viruses that dominated in the deeper aquifer may help to maintain the stability of the simple ecosystem.

In contrast, rare viruses that dominated in the upper aquifer may be beneficial to enhance the resilience of the complex ecosystem. Considering the bacterial characteristics in the previous study, it was hypothesized that abundant viruses and abundant bacterial keep a "Red Queen coevolution" relationship. In contrast, rare viruses and rare bacteria maintain a temperate coexistence relationship. Finally, on the individual level, the viral genetic traits were distinct in different depths. The main AMGs of viruses at different depths related to environmental features there. These environmental characteristics drive the inhabitation of bacterial communities with different metabolic types, and the AMGs of viruses contributed to the bacterial metabolism and further participated in the biogeochemical cycle. In a further study, bacterial data should also be collected to follow the dynamic relationship between bacteria and viruses, which will be beneficial to in-depth understanding their role in the ecological functions of deep groundwater.

References

1.	Mehrshad M, Lopez-Fernandez M, Sundh J, Bell E, Simone D, Buck M, et al. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. bioRxiv. 2020.

2.	Amend JP, Teske A. Expanding frontiers in deep subsurface microbiology. Geobiology: Objectives, Concepts, Perspectives. 2005:131-55.

3.	Itävaara M, Salavirta H, Marjamaa K, Ruskeeniemi T. Geomicrobiology and metagenomics of terrestrial deep subsurface microbiomes. Advances in applied microbiology. 2016;94:1-77.

4.	Nyyssönen M, Bomberg M, Kapanen A, Nousiainen A, Pitkänen P, Itävaara M. Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in Olkiluoto, Finland. Geomicrobiology Journal. 2012;29(10):863-78.

5.	Hare DK, Helton AM, Johnson ZC, Lane JW, Briggs MA. Continental-scale analysis of shallow and deep groundwater contributions to streams. Nature communications. 2021;12(1):1-10.

6.	Daly RA, Roux S, Borton MA, Morgan DM, Johnston MD, Booker AE, et al. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. Nature microbiology. 2019;4(2):352-61.

7.	Haveman SA, Pedersen K. Distribution of culturable microorganisms in Fennoscandian Shield groundwater. FEMS Microbiology Ecology. 2002;39(2):129-37.

8.	Pedersen K. Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H 2. The ISME journal. 2013;7(4):839-49.

9.	Nyyssönen M, Hultman J, Ahonen L, Kukkonen I, Paulin L, Laine P, et al. Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. The ISME journal. 2014;8(1):126-38.

10.	Pedersen K, Arlinger J, Eriksson S, Hallbeck A, Hallbeck L, Johansson J. Numbers, biomass and cultivable diversity of microbial populations relate to depth and borehole-specific conditions in groundwater from depths of 4–450 m in Olkiluoto, Finland. The ISME journal. 2008;2(7):760-75.

11.	Ho A, Di Lonardo DP, Bodelier PL. Revisiting life strategy concepts in environmental microbial ecology. FEMS microbiology ecology. 2017;93(3):fix006.

12.   Major TA, Liu Y, Whitman WB. Characterization of energy-conserving hydrogenase B in Methanococcus maripaludis. Journal of bacteriology. 2010;192(15):4022-30.

13.   Sokatch JR. Bacterial physiology and metabolism: Academic Press; 2014.

14.   Kothari A, Roux S, Zhang H, Prieto A, Soneja D, Chandonia J-M, et al. Ecogenomics of groundwater viruses suggests niche differentiation linked to specific environmental tolerance. bioRxiv. 2020.

15.   Rohwer F, Prangishvili D, Lindell D. Roles of viruses in the environment. Environ Microbiol. 2009;11(11):2771-4.

16.   Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature. 2009;459(7244):207-12.

17.   Breitbart M. Marine viruses: truth or dare. 2011.

18.   De Smet J, Zimmermann M, Kogadeeva M, Ceyssens P-J, Vermaelen W, Blasdel B, et al. High coverage metabolomics analysis reveals phage-specific alterations to Pseudomonas aeruginosa physiology during infection. The ISME journal. 2016;10(8):1823-35.

19.   Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. Plos Biol. 2005;3(5):e144.

20.   Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. Current opinion in microbiology. 2016;31:161-8.

21.   Stewart FM, Levin BR. The population biology of bacterial viruses: why be temperate. Theoretical population biology. 1984;26(1):93-117.

22.   Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. The ISME journal. 2017;11(7):1511-20.

23.   Gao S-M, Schippers A, Chen N, Yuan Y, Zhang M-M, Li Q, et al. Depth-related variability in viral communities in highly stratified sulfidic mine tailings. Microbiome. 2020;8(1):1-13.

24.   Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the 'core'and 'flexible'Pacific Ocean Virome. The ISME journal. 2015;9(2):472-84.

25.   Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proceedings of the National Academy of Sciences. 2014;111(29):10714-9.

26.	Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. Science. 2014;344(6185):757-60.

27.	Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537(7622):689-93.

28.	Kyle JE, Eydal HS, Ferris FG, Pedersen K. Viruses in granitic groundwater from 69 to 450 m depth of the Äspö hard rock laboratory, Sweden. The ISME Journal. 2008;2(5):571-4.

29.	Soederbaeck B. Geological evolution, palaeoclimate and historical development of the Forsmark and Laxemar-Simpevarp areas. Site descriptive modelling SDM-Site. Swedish Nuclear Fuel and Waste Management Co.; 2008.

30.	Drake H, Tullborg E-L. Paleohydrogeological events recorded by stable isotopes, fluid inclusions and trace elements in fracture minerals in crystalline rock, Simpevarp area, SE Sweden. Applied Geochemistry. 2009;24(4):715-32.

31.	Laaksoharju M, Tullborg E-L, Wikberg P, Wallin B, Smellie J. Hydrogeochemical conditions and evolution at the Äspö HRL, Sweden. Applied Geochemistry. 1999;14(7):835-59.

32.	Stanfors R, Rhén I, Tullborg E-L, Wikberg P. Overview of geological and hydrogeological conditions of the Äspö hard rock laboratory site. Applied Geochemistry. 1999;14(7):819-34.

33.	Hubalek V, Wu X, Eiler A, Buck M, Heim C, Dopson M, et al. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. The ISME journal. 2016;10(10):2447-58.

34.	Yan L, Herrmann M, Kampe B, Lehmann R, Totsche KU, Küsel K. Environmental selection shapes the formation of near-surface groundwater microbiomes. Water research. 2020;170:115341.

35.	Banwart S, Tullborg E-L, Pedersen K, Gustafsson E, Laaksoharju M, Nilsson A-C, et al. Organic carbon oxidation induced by large-scale shallow water intrusion into a vertical fracture zone at the Äspö Hard Rock Laboratory (Sweden). Journal of Contaminant Hydrology. 1996;21(1-4):115-25.

36.	Wu X, Holmfeldt K, Hubalek V, Lundin D, Åström M, Bertilsson S, et al. Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. The ISME journal. 2016;10(5):1192-203.

37.	Hengsuwan M, Heim C, Simon K, Hansen BT. Long-term investigation of Sr-isotope and rare earth element fractionation processes within three major aquifers in the Äspö Hard Rock Laboratory (Sweden). Geomicrobiology Journal. 2015;32(3-4):243-54.

38.	John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, et al. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environmental microbiology reports. 2011;3(2):195-202.

39.	Ma T, Ru J, Xue J, Schulz S, Mirzaei MK, Janssen K-P, et al. Differences in Gut Virome Related to Barrett Esophagus and Esophageal Adenocarcinoma. Microorganisms. 2021;9(8):1701.

40.	Wang W, Ren J, Tang K, Dart E, Ignacio-Espinoza JC, Fuhrman JA, et al. A network-based integrated framework for predicting virus–prokaryote interactions. NAR genomics and bioinformatics. 2020;2(2):lqaa044.

41.	Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018;35(6):1547-9.

42.	Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8(3):275-82.

43.	Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. evolution. 1985;39(4):783-91.

44.	Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. Evolving genes and proteins: Elsevier; 1965. p. 97-166.

45.	Salamov VSA, Solovyevand A. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture, biomedicine and environmental studies Hauppauge: Nova Science Publishers. 2011:61-78.

46.	Umarov RK, Solovyev VV. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PloS one. 2017;12(2):e0171410.

47.	Kyle JE, Ferris FG. Geochemistry of virus–prokaryote interactions in freshwater and acid mine drainage environments, Ontario, Canada. Geomicrobiology Journal. 2013;30(9):769-78.

48.	Fillinger L, Hug K, Griebler C. Selection imposed by local environmental conditions drives differences in microbial community composition across geographically distinct groundwater aquifers. FEMS microbiology ecology. 2019;95(11):fiz160.

49.   Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. Genome research. 2009;19(7):1141-52.

50.   Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. Environmental microbiology. 2012;14(1):4-12.

51.   Mehrshad M, Lopez-Fernandez M, Sundh J, Bell E, Simone D, Buck M, et al. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. Nature Communications. 2021;12(1):1-12.

52.   Williamson S, Houchin L, McDaniel L, Paul J. Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. Applied and environmental microbiology. 2002;68(9):4307-14.

53.   Herndl GJ. Microbial biomass dynamics along a trophic gradient at the Atlantic Barrier Reef off Belize (Central America). Marine Ecology. 1991;12(1):41-51.

54.   Boras JA, Sala MM, Vázquez-Domínguez E, Weinbauer MG, Vaqué D. Annual changes of bacterial mortality due to viruses and protists in an oligotrophic coastal environment (NW Mediterranean). Environmental microbiology. 2009;11(5):1181-93.

55.   Zhang Y, Jiao N, Cottrell MT, Kirchman DL. Contribution of major bacterial groups to bacterial biomass production along a salinity gradient in the South China Sea. Aquatic Microbial Ecology. 2006;43(3):233-41.

56.   Gomez-Alvarez V, Pfaller S, Pressman J, Wahman D, Revetta R. Resilience of microbial communities in a simulated drinking water distribution system subjected to disturbances: role of conditionally rare taxa and potential implications for antibiotic-resistant bacteria. Environmental Science: Water Research & Technology. 2016;2(4):645-57.

57.   Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a core community of abundant organisms. The ISME journal. 2016;10(1):11-20.

58.   Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010;466(7304):334-8.

59.   Van V, Van Valen L. A new evolutionary law. 1973.

60.   Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, et al. Viral and microbial community dynamics in four aquatic environments. The ISME journal. 2010;4(6):739-51.

61.   Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, et al. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio. 2014;5(4):e01371-14.

62.   Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences. 2009;106(52):22427-32.

63.   Gobet A, Böer SI, Huse SM, Van Beusekom JE, Quince C, Sogin ML, et al. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. The ISME journal. 2012;6(3):542-53.

64.   Downing AS, van Nes EH, Mooij WM, Scheffer M. The resilience and resistance of an ecosystem to a collapse of diversity. PloS one. 2012;7(9):e46135.

65.   Finlay BJ. Global dispersal of free-living microbial eukaryote species. Science. 2002;296(5570):1061-3.

66.   Pedrós-Alió C. Dipping into the rare biosphere. SCIENCE-NEW YORK THEN WASHINGTON-. 2007;5809:192.

67.   Gazitúa MC, Vik DR, Roux S, Gregory AC, Bolduc B, Widner B, et al. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. The ISME journal. 2021;15(4):981-98.

68.   Ruiz-Perez CA, Tsementzi D, Hatt JK, Sullivan MB, Konstantinidis KT. Prevalence of viral photosynthesis genes along a freshwater to saltwater transect in Southeast USA. Environmental microbiology reports. 2019;11(5):672-89.

69.   Li H, Xu H, Graham DE, White RH. Glutathione synthetase homologs encode α-L-glutamate ligases for methanogenic coenzyme F420 and tetrahydrosarcinapterin biosyntheses. Proceedings of the National Academy of Sciences. 2003;100(17):9785-90.

70.   Bogard MJ, Del Giorgio PA, Boutet L, Chaves MCG, Prairie YT, Merante A, et al. Oxic water column methanogenesis as a major component of aquatic CH 4 fluxes. Nature communications. 2014;5(1):1-9.

71.   Berendt U, Haverkamp T, Prior A, Schwenn JD. Reaction mechanism of thioredoxin: 3′-phospho-adenylylsulfate reductase investigated by site-directed mutagenesis. European journal of biochemistry. 1995;233(1):347-56.

72.   Ahmer BM, Van Reeuwijk J, Timmers CD, Valentine PJ, Heffron F. Salmonella typhimurium encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid. Journal of bacteriology. 1998;180(5):1185.

73.   Castilla-Llorente V, Salas M, Meijer WJ. kinC/D-mediated heterogeneous expression of spo0A during logarithmical growth in Bacillus subtilis is responsible for partial suppression of φ29 development. Molecular microbiology. 2008;68(6):1406-17.

74.   Bassler BL. Small talk: cell-to-cell communication in bacteria. Cell. 2002;109(4):421-4.

75.   Jain R, Rivera MC, Moore JE, Lake JA. Horizontal gene transfer accelerates genome innovation and evolution. Molecular biology and evolution. 2003;20(10):1598-602.

76.   Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature reviews microbiology. 2005;3(9):711-21.

77.   Drudge CN, Warren LA. Prokaryotic horizontal gene transfer in freshwater lakes: implications of dynamic biogeochemical zonation. Journal of Environmental Protection. 2012;3(12):1634-54.

78.   Raz Y, Tannenbaum E. The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments. Genetics. 2010;185(1):327-37.

79.   Shakibaie M, Jalilzadeh K, Yamakanamardi S. Horizontal transfer of antibiotic resistance genes among gram negative bacteria in sewage and lake water and influence of some physico-chemical parameters of water on conjugation process. Journal of Environmental Biology. 2009;30(1):45-9.

80.   Baek G, Kim J, Lee C. A review of the effects of iron compounds on methanogenesis in anaerobic environments. Renewable and Sustainable Energy Reviews. 2019;113:109282.

81.   Goyal A, Gelbwaser-Klimovsky D, Gore J. Horizontal gene transfer becomes disadvantageous in rapidly fluctuating environments. bioRxiv. 2020.

82.   Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics. 2008;9(8):605-18.

Figures

Figure 1: Depth-specific virome structures. **(a)** Relative abundance of sequences assigned to the indicated viral taxa in four different depths. **(b)** Rarefaction curves of viral richness (Ace) and diversity (Shannon) in four different depths. **(c)** Venn diagram of viral contigs in four different depths. **(d)** The predicted ratio of lysogenic/lytic cycles for viral contigs from different depths. **(e)** The predicted relative abundance of viruses from different sources in four different depths. Original means viruses that only appeared in the current study. Migrant means viruses that also appeared in other sources. **(f)** The relative abundance of predicted bacterial hosts in four different depths. BS represents water samples from the Baltic Sea; SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

Figure 2: Characteristics of core viral group. **(a)** Relative abundance of sequences in core viral group assigned to the indicated viral taxa **(b)** The relative abundance of abundant viruses in core viral group. BS represents water samples from the Baltic Sea; SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

Figure 3: Abundant viral groups pattern of different depths. **(a)** Relative abundance and number of abundant viruses in four different depths. **(b)** Relative abundance of abundant viruses assigned to the indicated viral taxa in four different depths. **(c)** The predicted ratio of lysogenic/lytic cycles for abundant viruses from different depths. **(d)** Predicted relative abundance of abundant viruses from different sources in four depths. **(e)** Venn diagram of the abundant viral contigs in four different depths. **(f)** Predicted bacterial host of abundant viruses. Gray represents the absence of viral contigs, the other colors represent the presence of viral contigs. BS represents water samples from the Baltic Sea; SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

Figure 4: Rare viral groups pattern of different depths. **(a)** Relative abundance and number of rare viruses in four different depths. **(b)** Relative abundance of rare viruses assigned to the indicated viral taxa in four different depths. **(c)** The predicted ratio of lysogenic/lytic cycles predicted for the rare viruses from different depths. **(d)** Predicted relative abundance of rare viruses from different sources in four depths. **(e)** Venn diagram of the rare viral contigs in four different depths. **(f)** Predicted bacterial host of rare viruses. Gray represents the absence of viral contigs, the other colors represent the presence of viral contigs. BS represents water samples from the Baltic Sea; SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

**Figure 5:** Characterization, distribution, and evolutionary analysis of viral AMGs **(a)** Genome map of six viral contigs containing AMGs **(b)** Relative abundance of AMGs in four different depths **(c)**The phylogenetic tree of protein sequences encoded by viral AMGs *cofF*, *cysH*, *LuxR*, and *spo0A*. Red represents the protein sequences encoded by AMGs of the predicted bacterial host. Black represents the protein sequences encoded by AMGs of cultivable bacteria in NCBI. Black bold represents the protein sequences encoded by AMGs of cultivable phage in NCBI. Purple bold represents the protein sequences encoded by AMGs of viral contigs in IMG. Blue represents the protein sequences encoded by AMGs of viral contigs in the current study. The number represents the confidence of sequence similarity. BS represents water samples from the Baltic Sea; SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.
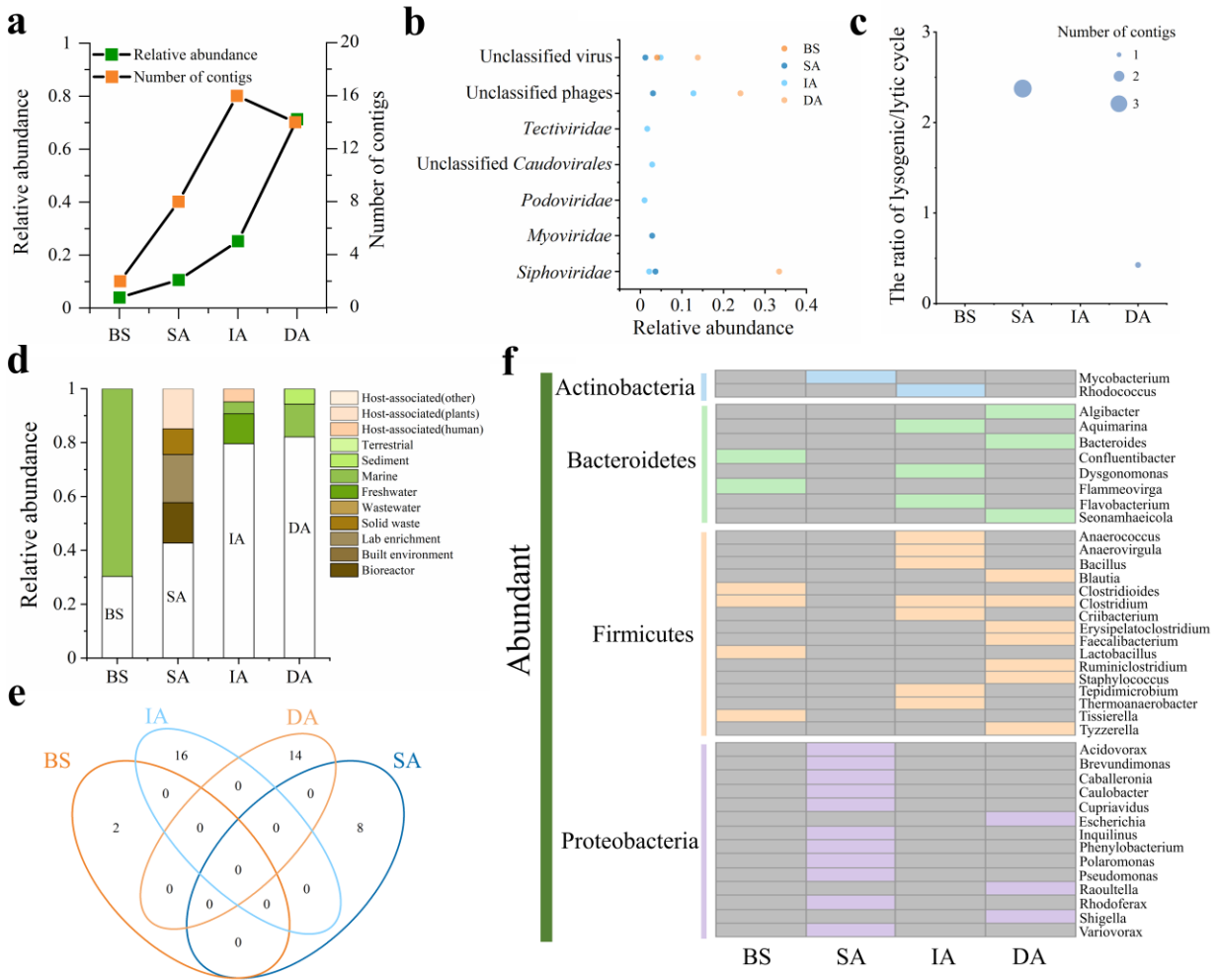
Figure 1
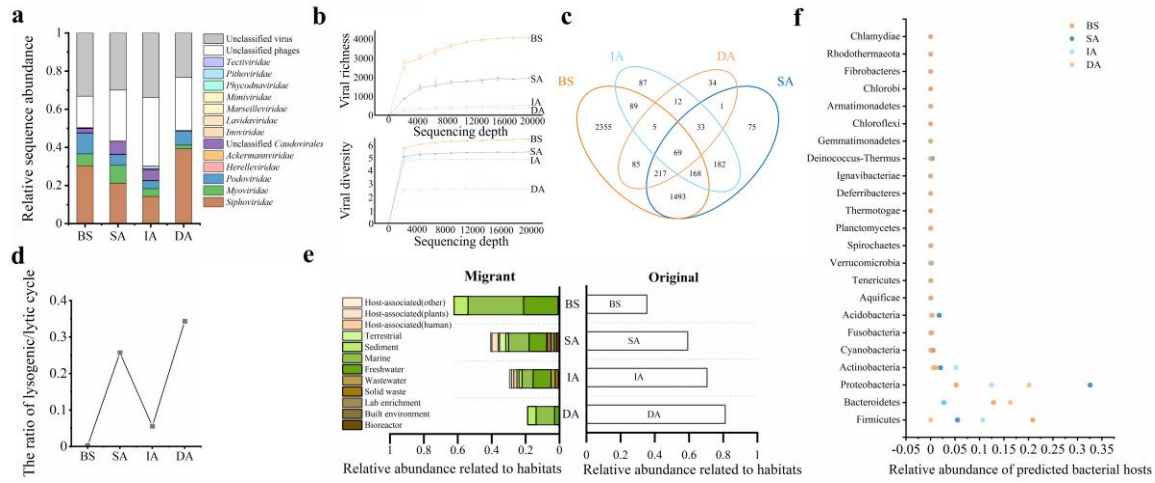
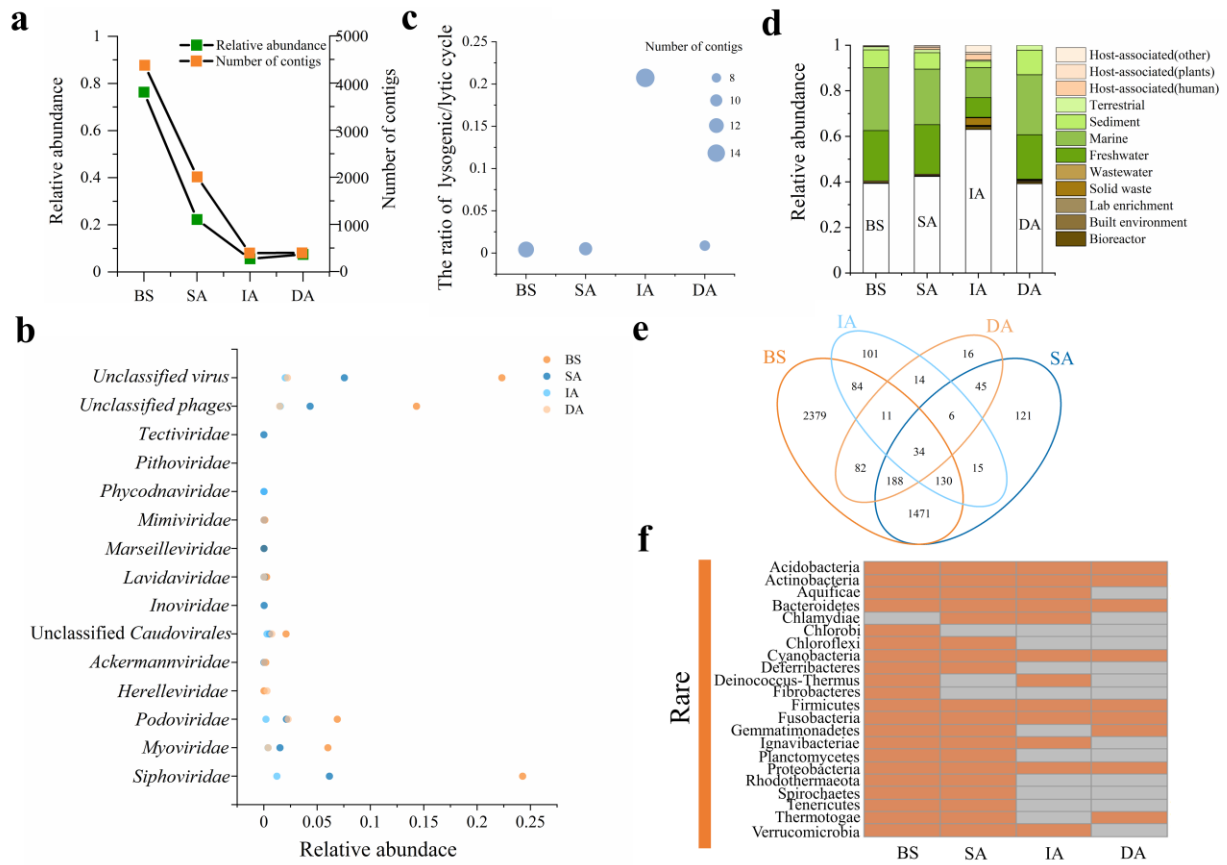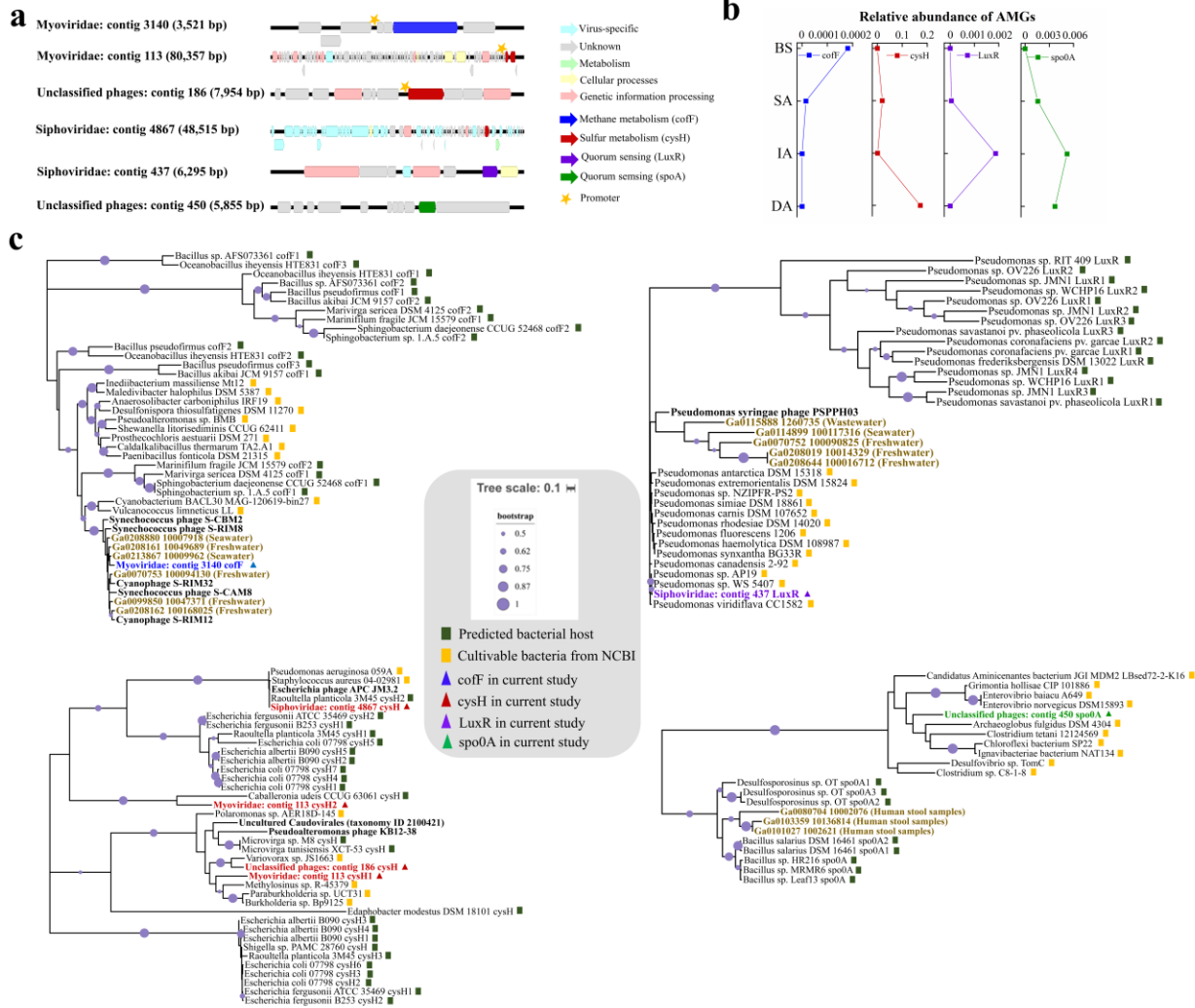**Figure 2**
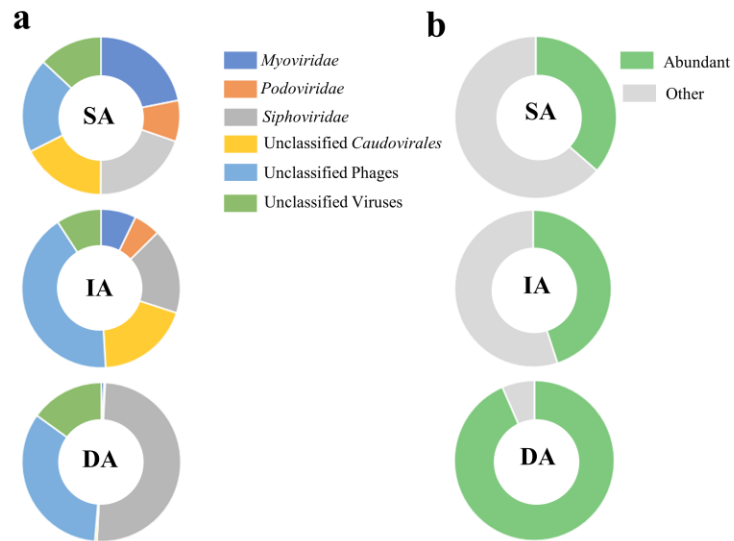
Figure 3

Figure 4

Figure 5

Supplementary data to the Article:

Contents

Table S1: The relative abundance of predicted bacterial hosts.

Figure S1: Venn diagram of the viral contigs in deep groundwater, including SA, IA, and DA. SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

Figure S2: The predicted ratio of lysogenic/lytic cycles for the abundant viruses and rare viruses in four depths. SA represents water samples from the shallow aquifer; IA represents water samples from the intermediate aquifer; DA represents water samples from the deep aquifer.

Table S1

| Table S1 The relative abundance of predicted bacterial hosts				
	BS	SA	IA	DA
Firmicutes	0.208635	0.054964	0.106634	0
Bacteroidetes	0.128197	0.027311	0.028378	0.162706
Proteobacteria	0.052146	0.325797	0.124398	0.201107
Actinobacteria	0.009923	0.020234	0.051372	0.005066
Cyanobacteria	0.005886	0.001861	0.000526	0.000294
Fusobacteria	0.002459	0.000724	0.001173	0.001212
Acidobacteria	0.001921	0.017788	0.00441	0.001771
Aquificae	0.000821	0.000223	7.46E-05	0
Tenericutes	0.000796	6.74E-05	0	0
Verrucomicrobia	0.000758	0.001159	0.002685	3.61E-05
Spirochaetes	0.000493	7.63E-05	0	0
Planctomycetes	0.000227	0.000413	0.000199	0
Thermotogae	0.000217	5.07E-05	0	4.84E-06
Deferribacteres	0.00021	1.49E-05	0	0
Ignavibacteriae	5.49E-05	4.69E-05	1.19E-05	0
Deinococcus-Thermus	3.66E-05	0.002913	4.26E-05	0
Gemmatimonadetes	2.11E-05	1.73E-05	0	5.94E-05
Chloroflexi	2.02E-05	1.03E-05	8.23E-06	0
Armatimonadetes	1.74E-05	0	0	0.00044
Chlorobi	1.65E-05	0	0	0
Fibrobacteres	8.19E-06	0	0	0
Rhodothermaeota	5.35E-06	1.03E-05	0.000166	0
Chlamydiae	0	5.01E-05	2.13E-05	0

Figure S1
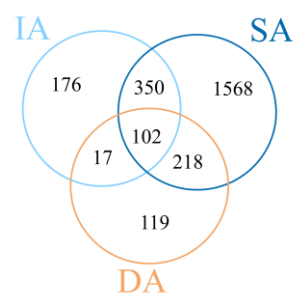
Figure S2
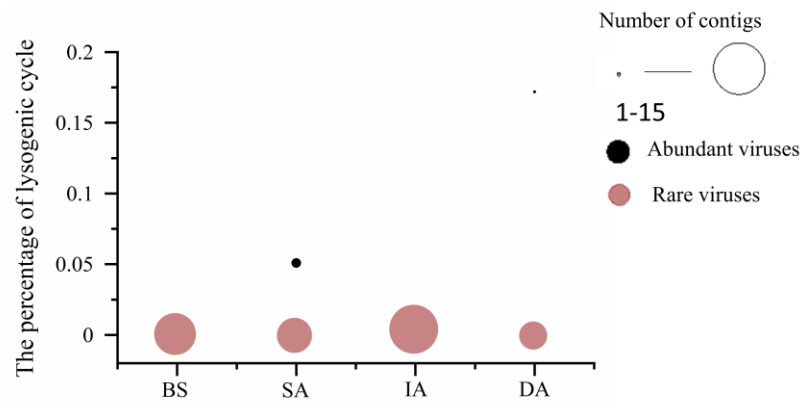
Appendix B

Curriculum Vitae

Tianli Ma

Personal Details

Name: Tianli Ma

Gender: Female

Date of birth: March 15, 1991

Place of Birth: Shanxi, China

Nationality: P.R. China

Address: Heimgartenstr 38b, Unterschleißheim, 85716, Germany

Email: tianli.ma@helmholtz-muenchen.de

Education Background

Sep 2016- Present	Ph.D. candidate	
	Helmholtz Zentrum München, Munich, Germany	
Sep 2012- Jul 2015	Master of Science	
	Northwest A&F University, Yangling, China.	
Sep 2008- Jul 2012	Bachelor of Science	
	Shanxi Agriculture University, Taigu, China.	

Publications

Ma, T.; Jiang, Y.; Elbehery, A.H.; Blank, S.; Kurmayer, R.; Deng, L. Resilience of planktonic bacterial community structure in response to short-term weather deterioration during the growing season in an alpine lake. *Hydrobiologia* **2020**, 847, 535-548.

Ma, T.; Ru, J.; Xue, J.; Schulz, S.; Mirzaei, M.K.; Janssen, K.-P.; Quante, M.; Deng, L. Differences in Gut Virome Related to Barrett Esophagus and Esophageal Adenocarcinoma. *Microorganisms* **2021**, *9*, 1701.

Ma, T.; Xue, J.; Ru, J.; Deng, L. Depth-specific virome reveals the survival strategy of aquifer microbiota adapting to the oligotrophic environments. Submitted