School of Doctoral Studies in Biological Sciences

University of South Bohemia in České Budějovice

Faculty of Science

Ecology of Gemmatimonadota

Ph.D. Thesis

Izabela Mujakić

Supervisor: Assoc. Prof. Michal Koblížek,
Laboratory of Anoxygenic Phototrophs, Center Algatech
Institute of Microbiology, The Czech Academy of Sciences

Advisor: Assoc. Prof. Kasia Piwosz,
Department of Fisheries, Oceanography and Marine Ecology
National Marine Fisheries Research Institute

České Budějovice 2023

This thesis should be cited as:

Mujakić, I., 2023: Ecology of Gemmatimonadota. Ph.D. Thesis Series, No.12, University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 2023, 151 pp.

Annotation

Phylum Gemmatimonadota is a common group present in many natural environments. Yet, this interesting bacterial group is rarely studied. Since its discovery 20 years ago, only six cultured species have been described. Abundances of Gemmatimonadota in various environments are usually low (≤ 1%), with the exception of soils, where they are one of the more abundant phyla. Probably for this reason, at the onset of this work, most of the knowledge about Gemmatimonadota came from studies of soil environments, while information about their ecology in freshwater lakes was missing. An interesting discovery relating to the unexplored diversity of Gemmatimonadota in freshwaters was the isolation of the first phototrophic member of this group, *Gemmatimonas phototrophica,* which was cultured from a shallow freshwater lake. To learn more about the ecology of Gemmatimonadota in freshwaters, the main focus of this thesis was the analysis of their distribution and diversity in several freshwater lakes, with emphasis on presence of photoheterotrophic Gemmatimonadota. The potential ecological roles of this group, metabolic capabilities and general genomic characteristics have also been addressed, thereby allowing a broader scope comparison with Gemmatimonadota from other environments such as soil, marine waters, or wastewaters.

Declaration

I hereby declare that I am the author of this dissertation and that I have used only those sources and literature detailed in the list of references.

Izabela Mujakić
České Budějovice
July 2023

This thesis originated from a partnership of the Faculty of Science, University of South Bohemia, and the Institute of Microbiology, Algatech Center, Academy of Sciences of the Czech Republic, supporting doctoral studies in the Hydrobiology study program.

Přírodovědecká fakulta
Faculty of Science

Jihočeská univerzita v Českých Budějovicích
University of South Bohemia in České Budějovice

MBÚ

ALGATECH
INSTITUTE OF MICROBIOLOGY

Financial support

The work was supported by the GAČR (Grant Agency of the Czech Republic) project PhotoGemm+ no. 19-28778X.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Michal Koblížek for the opportunity to do a PhD in his lab. I am extremely grateful for his continuous support, encouragement and guidance during my research and writing of the papers and the thesis. I greatly appreciate all the knowledge he shared with me, and all the opportunities he gave me during these years. My sincere gratitude also goes to my advisor Dr. Kasia Piwosz, who was an important part of my PhD study. I am deeply grateful to her for all the guidance, valuable advice, the knowledge she shared with me, help with lab work and writing, and continuous support during my studies.
I also wish to thank all the people who helped with my investigations, collaborated, and shared with me their knowledge which has been crucial for my work. Without them my PhD work would also not be possible. Many thanks go to Dr. Yonghui Zeng who I visited for one of my study stays, Dr. Adrian-Ştefan Andrei and Dr. Rohit Ghai who worked with me on my first paper and Dr. Pedro J. Cabello-Yeves and Dr. Antonio Camacho who accepted me for my PhD study stay in their laboratory and collaborated with me on the last chapter of my thesis. Thanks also go to all the members, past and present, of Laboratory of Anoxygenic phototrophs with whom I shared my working days, and nice times, and a special thank you goes to a good college Dr. Alastair T. Gardiner for always being available for advice, help with English, and prepared with good jokes ☺.
I am also grateful to all the friends I met during my PhD who made my life in a foreign country easier. Even though our paths have separated, they were an important part of my journey, and have made my days nicer and easier, especially when times were hard.
Finally, I would like to thank my parents Vesna and Izat for their unlimited support and unconditional love, as well as my brothers Alen and Dino and my sister Dina. Without them I would not be where I am today. Thank you for always encouraging me, and believing in me, even when I did not. I would also specially like to thank Cristian, my partner but also a colleague who continuously supports and cheers me on, both in life and work. Thank you for your love, comfort, advice, encouragement, as well as extra push during difficult and stressful days. I am so lucky I met you here in Třeboň and I will always cherish many adventures we had. Special thank you also goes to my best friend Dina, who I am fortunate to have in life, and who I can always rely and lean on even though many kilometers separate us, as well as Alen for being a good friend I can count on. Thank you both for continuous support, advice, and friendship, and for always believing in me.

Od srca hvala svima!

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. **Mujakić, I.**, Andrei, A-Ş., Shabarova, T., Fecskeová. L.K., Salcher, M.M., Piwosz, K., Ghai, R., Koblížek, M. (2021) Common presence of phototrophic Gemmatimonadota in temperate freshwater lakes. mSystems 6:e01241-20. (IF = 7.32)
https://doi.org/10.1128/mSystems.01241-20.
I.M. analyzed genomic data (50%), conducted phylogenetic analysis (50%), and wrote the manuscript (80%). Contribution of I.M. was 60%.

II. **Mujakić, I.**, Piwosz, K., Koblížek, M. (2022) Phylum Gemmatimonadota and Its Role in the environment. Microorganisms, 10, 151. (IF = 4.92)
https://doi.org/10.3390/microorganisms10010151
I.M. collected the available information (90%) and wrote the manuscript (80%). Contribution of I.M. was 80%.

III. Villena-Alemany, C., **Mujakić, I**., Porcal, P., Koblížek, M. & Piwosz, K. (2023) Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environmental Microbiology Reports*, 15(1), 60–71. (IF = 4.0)
https://doi.org/10.1111/1758-2229.13131
I.M. did PCR for sequencing (100%), statistical analysis (80%), partially phylogenetic analysis (10%) and participated in manuscript revision (30%). Contribution of I.M. was 15%.

IV. **Mujakić, I.**, Cabello-Yeves, P.J., Villena-Alemany, C., Piwosz, K., Rodriguez-Valera, F., Picazo, A., Camacho, A., Koblížek, M.: Multi-environment ecogenomics analysis of the cosmopolitan phylum Gemmatimonadota, *Manuscript under revision*, submitted to Microbiology Spectrum (IF=9.043)
I.M. participated in designing the study (50%), analyzed genomic data, conducted phylogenetic and metabolic analyses (90%), and wrote the manuscript (80%). Contribution of I.M. was 70%.

Co-author agreement

Doc. Michal Koblížek, the supervisor of this Ph.D. thesis and co-author of paper I-IV, fully acknowledges the stated contribution of Izabela Mujakić to these manuscripts.

.....
Doc. Michal Koblížek, Ph.D.

Doc. Kasia Piwosz, the advisor of this Ph.D. thesis and co-author of paper I-IV, fully acknowledges the stated contribution of Izabela Mujakić to these manuscripts.

.....
Doc. Kasia Piwosz, Ph.D.

Cristian Villena-Alemany, the first author of paper III, fully acknowledges the stated contribution of Izabela Mujakić as the second author of the paper and her contributions as stated above.

.....
Cristian Villena-Alemany, MSc.

Table of contents

1. Introduction **1**
1.1. Bacteria in the environment 2
1.2. Biogeochemical cycles 3
1.3. Culturing and classification of bacteria 4
1.4. Unveiling microbial diversity 5
1.5. Metagenomics – a new approach to study uncultivated bacteria 7
1.6. Anoxygenic Photosynthesis 9
1.7. Aerobic anoxygenic phototrophic (AAP) bacteria 10
1.8. Gemmatimonadota 12
2. Aims and objectives **15**
3. Results **17**
Paper I 19
3.1. Summary results derived from Paper No. 1 21
Common presence of phototrophic Gemmatimonadota in temperate freshwater lakes 23
Paper II 43
3.2. Summary results derived from Paper No. 2 45
Phylum Gemmatimonadota and its role in the environment 47
Paper III 65
3.3. Summary results from Paper No. 3 67
Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake 69
Paper IV 81
3.4. Summary results from Paper No. 4 83
Multi-environment ecogenomics analysis of the cosmopolitan phylum Gemmatimonadota 85
4. Conclusion and future prospects **129**
5. References **131**
Curriculum vitae 149

1. Introduction

1.1. Bacteria in the environment

Bacteria evolved approximately 3.5 billion years ago (Scott, 1984), and have since then colonized most natural environments, from water, air, soil, rock surfaces, to more extreme environments such as deep hydrothermal vents, glaciers, hot springs or volcanos (Pace, 1997; Dunlap, 2001). Moreover, many species of bacteria live in association with plants, animals, or other microorganisms. Bacteria have immense, for most parts unknown genomic and metabolic diversity (Dunlap, 2001). Their morphology also shows variations, with shapes ranging from rods, ovoids, spheres, spirals, vibrios, and many other (Kysela et al., 2016). Bacterial size is usually around one micrometer, with some exceptions, such as ubiquitous free-living marine bacterium *Pelagibacter ubique* with length of 0.37–0.89 µm (Rappé et al., 2002), or aquatic bacterium Candidatus *Actinomarina minuta* with an average diameter of 0.29 µm (Ghai et al., 2013). On the opposite side of spectrum examples include giant bacteria like *Thiomargarita namibiensis* (750 µm) (Schulz et al., 1999) or a centimeter long and visible by eye *Thiomargarita magnifica* (Volland et al., 2022). Furthermore, bacteria represent large reservoirs of key elements, including carbon, nitrogen and phosphorous (Whitman et al., 1998), and it was estimated that they account for 15% of all global biomass, second only after the plants (Bar-On et al., 2018).

Bacterial community structure and population densities vary depending on the environment they inhabit and are influenced by abiotic and biotic factors, such as seasons, temperature, nutrients, pH, trophic status or relationship with other organisms (Cole, 1982; Lombard et al., 2011; Hobbie and Hobbie, 2013; Nemergut et al., 2013; Delgado-Baquerizo et al., 2017). In aquatic environments, their densities generally range from 10^4–10^5 cells ml^{-1} in the oceans (Whitman et al., 1998; Sogin et al., 2006), and 10^5–10^6 cells ml^{-1} in freshwater environments (Porter and Feig, 1980; Pedrós-Alió and Brock, 1982; Whitman et al., 1998; Chróst et al., 2009), but they can reach up to 10^8 cells ml^{-1} in more nutrient-rich or eutrophic lakes (Jugnia et al., 1998; Porter et al., 2004). The highest densities are usually found in different types of soils or sediments where they can reach 10^8-10^{10} cells per gram of soil or sediment (Torsvik et al., 1990; Llobet-Brossa et al., 1998; Christensen et al., 1999; Musat et al., 2006; Raynaud and Nunan, 2014).

1.2. Biogeochemical cycles

Bacteria use diverse metabolic strategies to obtain and conserve energy in order to thrive under varying environmental conditions. Based on the source of carbon they need for growth, bacteria are classified as heterotrophs, which need organic carbon, or autotrophs that can use CO_2. Energy for metabolism can be obtained from three sources. The majority of bacteria oxidize various organic compounds and they are called chemoorganotrophs. Chemoorganotrophs produce reduced inorganic molecules as waste products of their metabolism, such as H_2, H_2S, or NH_3, which are then oxidized by chemolithotrophs (Madigan et al., 2010). Finally, phototrophic bacteria can use light and convert it into chemical energy needed for their metabolism.

Biogeochemical cycles of all major biogenic elements, such as carbon, nitrogen, phosphorus and sulfur are driven by bacteria (Falkowski et al., 2008). One of the most fundamental processes on Earth crucial to carbon cycling is photosynthesis (Hohmann-Marriott and Blankenship, 2011). Photoautotrophic bacteria (e.g. Cyanobacteria) and other primary producers such as algae fix inorganic carbon and produce new biomass (Falkowski et al., 1998). Part of the organic carbon fixed by photosynthetic organisms is released as dissolved organic matter to the aquatic ecosystems and used by heterotrophic bacteria to produce new biomass (secondary production) (Cole et al., 1982) making it available to upper trophic levels *via* microbial loop (Azam et al., 1983; Dodds and Whiles, 2020). The bacterial impact on energy flow and nutrient cycling in aquatic environments is especially important due to their key role in global carbon budget (Cole et al., 2007) through burial of organic carbon (Battin et al., 2009; Tranvik et al., 2009), and as source of CO_2 from microbial respiration (Berggren et al., 2012; Maberly et al., 2013).

Another important biogenic element is nitrogen (Vitousek et al., 2002), which is required for the synthesis of proteins and nucleic acids. Nitrogen represents a limiting nutrient in many ecosystems (LeBauer and Treseder, 2008). Despite the fact that gaseous dinitrogen is the most abundant element in the atmosphere, it is inaccessible to most organisms in this form. Only prokaryotic organisms called diazotrophs, which possess enzyme nitrogenase, are able to reduce N_2 into ammonia (Howard and Rees, 1996; Dekas et al., 2009). Diazotrophs are widely distributed in the environment and can be found in soils, oceans, lakes, sediments, and microbial mats, among many others (Zehr et al., 2003; Montoya et al., 2004; Coyne et al., 2020; Pierella Karlusich et al., 2021). Another important reaction is nitrification (Bristow et al., 2017). During this process, ammonia is oxidized to nitrite by aerobic ammonia-oxidizing bacteria belonging to several groups of Proteobacteria (French et al., 2012; Lehtovirta-Morley,

2018) as well as some Archaea species (Könneke et al., 2005; Lehtovirta-Morley, 2018). In the second step of nitrification, nitrite is oxidized to nitrate by nitrite-oxidizing bacteria belonging to Proteobacteria (Teske et al., 1994), some Chloroflexota (Sorokin et al., 2012) and the genus *Nitrospira* of the phylum Nitrospirae (Ehrich et al., 1995). Both ammonia-oxidizers and nitrite-oxidizers are ubiquitous in soils, biofilms, lakes and oceans (Kowalchuk and Stephen, 2001; Leininger et al., 2006; Wuchter et al., 2006; Yool et al., 2007; Beman et al., 2008; Bristow et al., 2017). Ammonia oxidation occurs also under anoxic conditions in the process called anammox (Bristow et al., 2017), conducted by bacteria belonging to Planctomycetes (Strous et al., 1999). During that process NO_2^- and NH_4^+ are converted into N_2 and water. In some environments, like oceans, this process is considered to significantly contribute to the loss of bioavailable nitrogen (Kuypers et al., 2005). Finally, nitrate can also be removed through denitrification by denitrifying bacteria. During this process intermediate forms of nitrogen like the greenhouse gas nitrous oxide, N_2O, are also produced.

The study of bacterial dynamics, functions and roles in key processes, as well as their ecology, composition and diversity, is an important topic which sheds light on how life on Earth is sustained.

1.3. Culturing and classification of bacteria

In the earliest times of microbiological studies, the knowledge of microorganisms came mostly from observations by light microscopy. A major breakthrough was the establishment of pure culture work (Pace, 1997; Austin, 2017). This was facilitated by the discovery of sterile solid agar media, which allowed simple and effective isolation of pure bacterial strains (Madigan et al., 2010; Austin, 2017). Still, the identification and classification of obtained organisms remained a major challenge, since microbiologists had to rely only on physiological, morphological, and chemical analyses to determine the phylogenetic relationship between the studied species.

The situation changed with the development of molecular methods. In 1965 Zuckerland and Pauling described various molecules that could be used for molecular phylogeny, calling them the "documents of evolutionary history" (Zuckerkandl and Pauling, 1965). Later, genes encoding small subunit rRNA (16S or 18S rRNA) were established as a common marker for identifying microbial species (Pace et al., 1986; Hugenholtz et al., 1998). Using this approach, Carl Woese divided all known organisms into three basic domains of life: Bacteria, Eukaryota and Archaea (Woese and Fox, 1977; Woese, 1987). Domains of Bacteria and Archaea are further divided into lower taxonomic ranks of phyla, classes and orders (Hugenholtz, 2002). Based on 16S rRNA gene sequences from cultured species Woese originally defined only 11 bacterial phyla: purple bacteria

(now part of Proteobacteria), Cyanobacteria, gram-positive bacteria, chlamydiae, planctomycetes and relatives, bacteroidetes-flavobacteria, green-sulfur bacteria, spirochetes, green non-sulfur bacteria, deinococci and relatives, and Thermotoga (Woese, 1987; Hugenholtz, 2002; Rappé and Giovannoni, 2003). Moreover, he also showed that some bacteria with different morphological and metabolic characteristics, such as Chloroflexota, which at that time consisted of four species with little phenotypic similarities, belonged to the same phylum (Woese, 1987; Rappé and Giovannoni, 2003).

1.4. Unveiling microbial diversity

While culture work provided information about bacterial ultrastructure, metabolism, and genomics, studying microbial diversity in nature using only culture-based techniques represented a serious constraint. Microbiologists realized that the number of cells determined in a natural environment by microscopy largely exceeded the colony numbers obtained by agar plating. This phenomenon called "the great plate count anomaly" (Jannasch and Jones, 1959; Staley, 1985) documents the fact that not every microbe present in the environment can be cultured under standard laboratory conditions (Staley, 1985; Torsvik et al., 1990; Pace, 1997; Rappé and Giovannoni, 2003; Austin, 2017). Indeed, classical cultivation reveals <1% of bacterial species living in a particular environment (Amann et al., 1995) resulting in a serious underestimation of microbial diversity and abundance (Hugenholtz et al., 1998).

The limitation of the culture-based studies was finally overcome by analyses of 16S rRNA genes present in environmental DNA samples. This approach ultimately revealed that the majority of the existing microorganisms remain uncultured (Giovannoni et al., 1990; Rappé and Giovannoni, 2003), and led to the description of novel bacterial groups, including high ranks such as phyla (Ludwig et al., 1997). One of the phyla identified from environmental 16S rRNA gene sequences was Gemmatimonadota (Li et al., 1999; Hugenholtz et al., 2001; Mummey and Stahl, 2003). Currently, there are 89 phyla in the SILVA database (Pruesse et al., 2007), and 148 phyla in the GTDB database (Parks et al., 2021). Most of them are candidate phyla without cultured representatives, recognized only from environmental sequences (Hug et al., 2016). Still, this is only a small part of the microbial diversity. It was estimated that there may exist more than 1000 bacterial phyla (Yarza et al 2014).

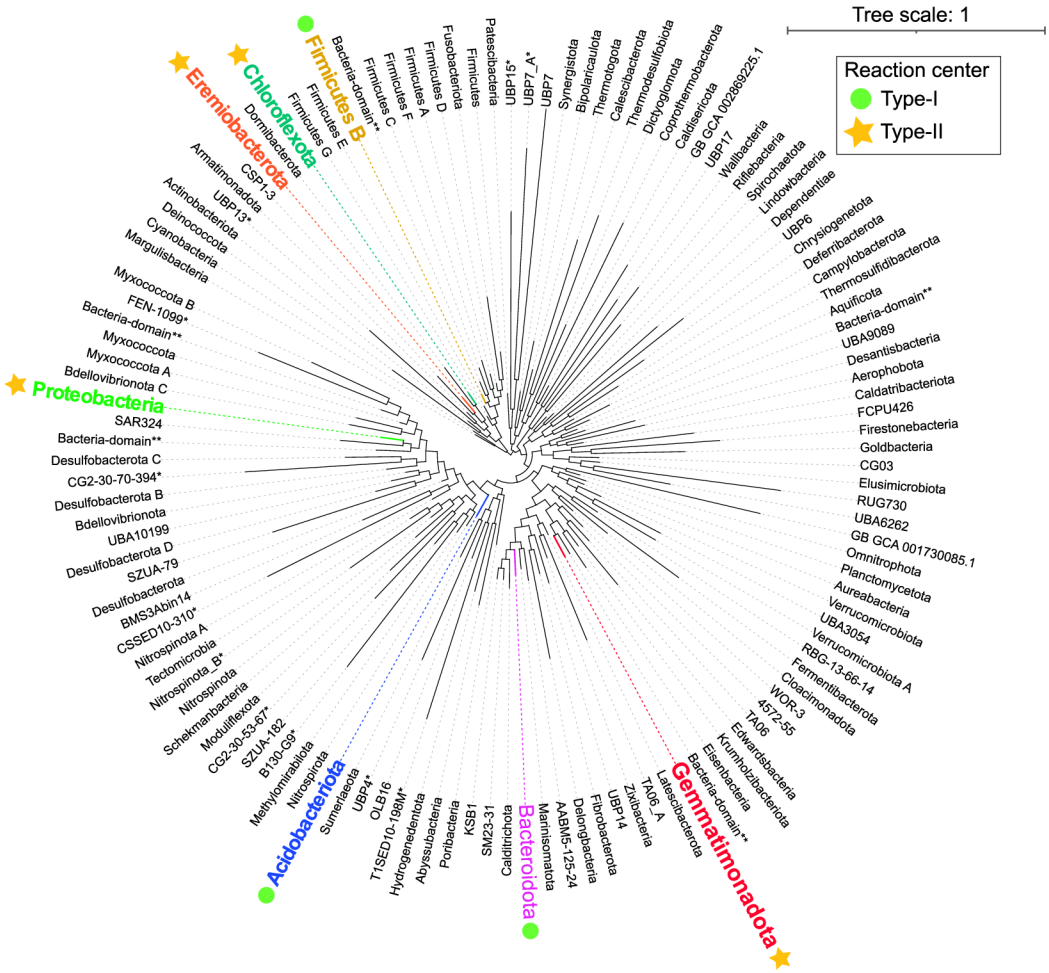

Figure 1. *The phylogenomic tree of all bacterial phyla present in the Genome Taxonomy Database (GTDB) (Release RS95) visualized using AnnoTree (version 1.2) (Mendler et al., 2019). Gemmatimonadota and six other phyla with members capable of anoxygenic phototrophy are marked in different colors, and the type of reaction center is indicated with a circle or a star (right legend). The double asterisk (**) means there is no corresponding taxonomy in GTDB for the genome, so a higher taxonomic level is shown. The tree was modified from Mujakić et al., 2022.*

A weak point of 16S rRNA gene sequencing can be the problematic quantification of observed organisms. This problem can be solved using fluorescence *in situ* hybridization (FISH), where bacteria are labelled on different phylogenetic levels with DNA or RNA oligonucleotide probes. This allows us to identify and enumerate them by fluorescence microscopy and provides estimates of their abundance (Amann et al., 1990, 2001; Moter and Göbel, 2000; Eickhorst and Tippkötter, 2008). However, FISH also has its technical limitations, especially when detecting small bacteria with low ribosome content or those with low abundances in the environment (Amann and Fuchs,

2008). For that reason, the original FISH protocol was improved by the application of probes labelled with enzyme horseradish peroxidase and subsequent catalyzed reported deposition of fluorescently labelled tyramides (CARD or TSA-FISH) (Schönhuber et al., 1997; Pernthaler et al., 2002; Amann and Fuchs, 2008; Kubota, 2013). The CARD-FISH method produces a much stronger signal, which is easier to detect by fluorescence microscopy. Both techniques have been successfully used in the description and identification of bacterial communities in various environments, such as sediments (Llobet-Brossa et al., 1998, 2002; Ishii et al., 2004), soil (Bertaux et al., 2007; Eickhorst and Tippkötter, 2008), or wastewaters (Gieseke et al., 2003; Maixner et al., 2006). Furthermore, in marine waters (Glöckner et al., 1999; Cottrell and Kirchman, 2000; Pernthaler et al., 2002), these methods helped to quantify and describe the diversity of the most abundant marine bacteria, SAR11 or *Pelagibacter* group (Herlemann et al., 2014), and aided in the culturing process of its members (Rappé et al., 2002). In fresh waters, these methods were used in the description of highly abundant Actinobacteria (Sekar et al., 2003; Neuenschwander et al., 2018), Alphaproteobacteria (Salcher et al., 2011, 2013; Neuenschwander et al., 2015), freshwater methylotrophic bacteria (Salcher et al., 2015), *Polynucleobacter* (Wu and Hahn, 2006), freshwater Chloroflexota (Mehrshad et al., 2018) and many others.

Culturing still has an important role in microbial ecology, however it is necessary to combine it with cultivation-independent methods like 16S rRNA gene amplicons or FISH and CARD-FISH when studying bacterial diversity. Furthermore, the combination of these methods with the newly emergent field of metagenomics, described in the next chapter, is becoming the new standard to describe uncultivated species of bacteria (Konstantinidis et al., 2017).

1.5. Metagenomics – a new approach to studying uncultivated bacteria

Despite of the significant progress in bacterial diversity achieved by the application of 16S RNA analyses, the taxonomic information does not provide information about the actual metabolic potential of uncultured species. Answers to such questions can be obtained through metagenomics, where bacterial communities are analyzed by sequencing all DNA present in an environment of interest (Wooley et al., 2010). Subsequently, by applying various computational tools, the environmental sequences are assembled into contigs and finally reconstructed into genomes called metagenome-assembled genomes (MAGs) present in that environment. Usage of shotgun metagenomics enables us to study bacterial diversity and uncover its richness within different environments, as well as their ecology and interactions, evolution, metabolic

pathways and functional diversity (Tringe et al., 2005; Hooper et al., 2008; Wooley et al., 2010; Shah et al., 2011; Escobar-Zepeda et al., 2015; Oulas et al., 2015). With the continuous advances in sequencing methods, an important contribution in genomics has come from long-read single-molecule DNA sequencing technologies like Pacific Biosciences (PacBio sequencing) and Oxford Nanopore Technologies, which can produce continuous long sequences (Logsdon et al., 2020). However, the use of these methods in metagenomics still has some challenges, such as DNA extraction for long molecules, lower sequencing depth, higher costs and higher error rates (Frank et al., 2016; Marx, 2023). On the other hand, as the short reads produced by Illumina sequencing can often limit the assembly of long contigs, the combination of these two methods is recommended for a successful reconstruction of bacterial genomes (Risse et al., 2015; Frank et al., 2016; Wick et al., 2023).

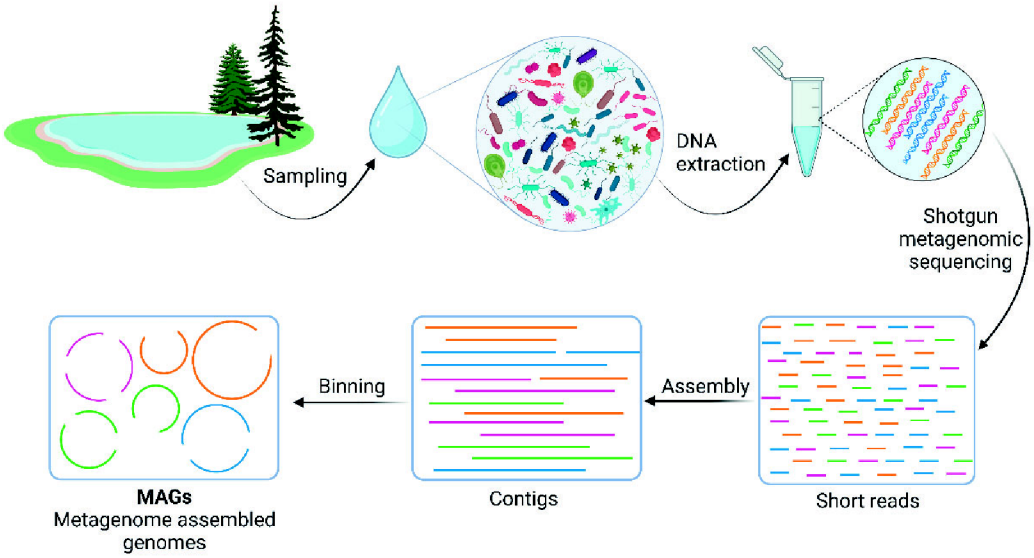

Figure 2. *Simplified scheme of metagenomic workflow from sampling of a freshwater lake to obtaining MAGs. The figure was created using BioRender.com*

The first environmental metagenomics studies focused on environments that were expected to have lower species diversity, such as acid mine drainage microbial biofilm (Tyson et al., 2004) or Sargasso Sea (Venter et al., 2004). However, with advances in high throughput sequencing methods and assembly software, it seems that there are no environments that cannot be analyzed. Metagenomic data unveiled novel genetic diversity with many draft or complete genomes of previously unknown groups of bacteria (Hug et al., 2016a; Parks et al., 2017) and expanded our understanding of their roles in biogeochemical processes (Hug et al., 2013, 2016b; Anantharaman et al., 2016). At the same time, microbial communities can be taxonomically identified and their metabolism can be analyzed (Hug et al., 2016b). Metagenomes are successfully used in

the description of bacterial communities of soils (Crits-Christoph et al., 2018; Pessi et al., 2022), wastewaters (Speth et al., 2016; Lawson et al., 2017; Chen et al., 2022), microbial mats (Lee et al., 2018; Mendes Monteiro et al., 2019), deep sediments and ground waters (Anantharaman et al., 2016; Hug et al., 2016b), marine waters (Rusch et al., 2007; Sunagawa et al., 2015; Haro-Moreno et al., 2018) as well as more extreme environments, like deepest ocean sediments and hydrothermal vents (Xie et al., 2011; Meier et al., 2017; Chen et al., 2021), or soda lakes (Vavourakis et al., 2018, 2019; Zorz et al., 2019). Metagenomic studies of freshwater environments are also numerous. They allow us to describe the metabolic potential of microbial communities in lakes through assemblies of new MAGs (Cabello-Yeves et al., 2017; He et al., 2017; Holland-Moritz et al., 2018; Mehrshad et al., 2018; Ward et al., 2019) and to understand bacterial roles in biogeochemical cycles (Ghai et al., 2014; Arora-Williams et al., 2018; Cabello-Yeves et al., 2018, 2020; Linz et al., 2018; Tran et al., 2021; Smith et al., 2022).

The advancement of metagenomics continuously expands our knowledge of the diversity and metabolic potential of bacteria in the natural environment. This also includes the bacterial phylum that is the main topic of this thesis, Gemmatimonadota, an interesting group with members capable of anoxygenic photosynthesis.

1.6. Anoxygenic Photosynthesis

The first phototrophic organisms lived under anaerobic conditions and did not produce oxygen (Olson, 2006; Hohmann-Marriott and Blankenship, 2011). In contrast to oxygenic Cyanobacteria, which use water as an electron donor and evolve oxygen, anoxygenic phototrophs do not evolve oxygen as they use different reduced electron donors such as hydrogen, sulfide, or ferrous ions (Hohmann-Marriott and Blankenship, 2011). Anoxygenic photosynthesis has scattered distribution across the Tree of Life and is currently known in seven bacterial phyla: Proteobacteria, Acidobacteria, Bacteroidota, Chloroflexota, Firmicutes, Eremiobacteriota, and Gemmatimonadota (Bryant and Liu, 2013; Zeng et al., 2014; Cardona, 2015; Parks et al., 2018; Ward et al., 2019). Anoxygenic phototrophs harvest light using different forms of bacteriochlorophylls (BChl) and carotenoids organized in light-harvesting complexes surrounding photosynthetic reaction centers (Yurkov and Csotonyi, 2009; Sener et al., 2016). There are two basic types of the reaction centers (Hohmann-Marriott and Blankenship, 2011). Type I reaction centers use ferredoxin as an electron acceptor and are present in anaerobic anoxygenic phototrophic Firmicutes, Chlorobi, and Acidobacteria (Cardona, 2016). In contrast, type II reaction centers use quinones as the final electron acceptors and are present in both aerobic and anaerobic anoxygenic phototrophic Proteobacteria, Chloroflexota, Eremiobacteriota and Gemmatimonadota

(Zeng et al., 2014; Ward et al., 2019).

Oxygenic Cyanobacteria appeared around 2.4 – 2.7 billion years ago and gradually transformed life on Earth through the oxygenation of the atmosphere and oceans (Nisbet and Fowler, 2003; Bekker et al., 2004; Holland, 2006; Lyons et al., 2014). This forced anaerobic phototrophs to retreat to the remaining anoxic environments (Hohmann-Marriott and Blankenship, 2011), where they express photosynthetic genes (Bauer et al., 2009) and perform photosynthesis under anaerobic conditions. Other phototrophic species partially or fully adapted to new aerobic conditions (Raymond and Segrè, 2006). For instance, aerobic anoxygenic phototrophic (AAP) bacteria harvest light energy under fully oxic conditions but have lost any carbon fixation capacity. With the pressure of new environmental conditions, some anoxygenic phototrophs transitioned to a heterotrophic lifestyle and lost their photosynthesis genes. Proteobacteria, for example (Woese, 1987; Nagashima and Nagashima, 2013) has a scattered distribution of phototrophy inside the phylum (Swingley et al., 2009; Koblížek et al., 2013; Kopejtka et al., 2017). The current diversity of phototrophic phyla in the Tree of Life is also explained by events of horizontal gene transfer (HGT) of phototrophic genes, a mechanism which is recorded in many groups of phototrophic bacteria (Igarashi et al., 2001; Raymond et al., 2002; Nagashima and Nagashima, 2013; Zeng et al., 2014; Brinkmann et al., 2018; Ward et al., 2019).

1.7. Aerobic anoxygenic phototrophic (AAP) bacteria

Aerobic anoxygenic phototrophic (AAP) bacteria are a functional group with members across several bacteria phyla that are able to obtain energy from light through BChl *a* containing type II reaction centers. AAPs are facultative photoheterotrophs that require a source of reduced organic carbon for growth, and light serves as an auxiliary source of energy (Yurkov and Beatty, 1998; Koblížek, 2015; Piwosz et al., 2018). They use cyclic photophosphorylation to generate a proton gradient that is used for ATP synthesis (Overmann and Garcia-Pichel, 2013), which provides a metabolic advantage over pure heterotrophs (Koblížek et al., 2005). To avoid the simultaneous presence of light and oxygen during bacteriochlorophyll synthesis, regulatory mechanisms restrict the expression of phototrophic genes and bacteriochlorophyll synthesis to dark periods (Nishimura et al., 1996; Yurkov and Beatty, 1998; Koblížek et al., 2005; Fecskeová et al., 2019). For protection from excessive light and oxygen, they contain a variety of carotenoids, which also give them vivid colors and serve as additional light-harvesting pigments (Yurkov and Csotonyi, 2009; Koblížek, 2015).

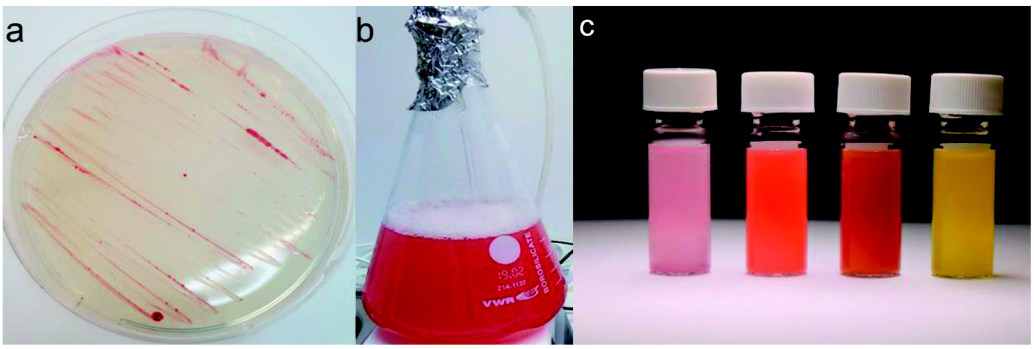

Figure 3. *Pictures showing cultures and different colors of AAP bacteria. (a) Gemmatimonas phototrophica, (b) Gemmatimonas groenlandica, (c) Roseobacter litoralis, Roseococcus thiosulfatophilus, Erythrobacter longus and Sphingomonas sp. AAP2. Picture is modified from (Koblížek, 2015; Zeng et al., 2021; Mujakić et al., 2022)*

AAP bacteria are diverse, highly abundant and important members of bacterial communities in marine waters (Kolber et al., 2001; Hojerová et al., 2011; Koblížek, 2015) as well as freshwater environments (Man et al., 2008; Mašín et al., 2012; Čuperová et al., 2013; Ruiz-González et al., 2013; Fauteux et al., 2015; Ferrera et al., 2017; Kolářová et al., 2019). AAPs are characterized by on average larger cell size and faster growth rates than heterotrophic bacteria (Koblížek et al., 2007; Ferrera et al., 2011; Cepáková et al., 2016; Garcia-Chaves et al., 2016; Piwosz et al., 2020). However, their bigger size makes them more vulnerable to protistan grazing, which indicates their important role in the microbial food webs and aquatic carbon cycle (Koblížek et al., 2007; Cepáková et al., 2016). Additionally, several studies showed that cultures of AAP bacteria exposed to light are able to reduce respiration rate by 75 %, and that way incorporate carbon more efficiently and increase their growth efficiency by 50 % (Hauruseu and Koblížek, 2012; Piwosz et al., 2018; Koblížek et al., 2020). The importance of AAP bacteria and their role in the carbon cycle was demonstrated also in nature (Piwosz et al., 2022). The activity of AAP bacteria in a freshwater lake was determined from microbial respiration and carbon compounds assimilation measured in dark and infrared light. Bacterial respiration of the whole microbial community was reduced by 15 % in infrared light, which selectively excites BChl *a*-containing photosystems. Additionally, AAP bacteria increased assimilation of carbon sources like glucose, pyruvate and leucine (Piwosz et al., 2022). As freshwater lakes, where AAP bacteria are commonly present (Cepáková et al., 2016; Fecskeová et al., 2019; Kolářová et al., 2019; Villena-Alemany et al., 2022), have an important role in the global carbon cycle (Tranvik et al., 2009; Berggren et al., 2012; Maberly et al., 2013), it is essential to take into account the potential impact that photoheterotrophy of AAP bacteria can have on it.

1.8. Gemmatimonadota

Gemmatimonadota is one of the bacterial phyla originally discovered using molecular phylogenetic methods. It was first established as BD/KS-B group based on five 16S rRNA gene sequences from various environments (Li et al., 1999; Hugenholtz et al., 2001; Madrid et al., 2001; Mummey and Stahl, 2003). The first cultured representative *Gemmatimonas aurantiaca* was isolated from the activated sludge of a wastewater treatment plant and gave the phylum its name Gemmatimonadota (Zhang et al., 2003). Later, three novel species were isolated from various soils, *Gemmatirosa kalamazoonesis* and *Roseisolibacter agrii* from agricultural soil (DeBruyn et al., 2013; Pascual et al., 2018) and *Longimicrobium terrae* from Mediterranean forest soil (Pascual et al., 2016). Finally, two isolates came from freshwater environments, *Gemmatimonas phototrophica* from a shallow freshwater lake (Zeng et al., 2015) and *Gemmatimonas groenlandica* from a stream (Zeng et al., 2021).

With the discovery of the photoheterotrophic *G. phototrophica*, Gemmatimonadota was recognized as one of the phyla containing anoxygenic phototrophic species (Zeng et al., 2014). The unique and efficient photosynthetic complex of *G. phototrophica,* which contains type II reaction centers surrounded by two concentric rings of light-harvesting antennae, was recently purified and characterized (Dachev et al., 2017; Qian et al., 2022). Genes encoding for reaction center and BChl *a* biosynthesis are organized into photosynthesis gene cluster (PGC) in all known phototrophic Gemmatimonadota, (Zeng et al., 2014, 2021). This organization is common for Proteobacteria (Nagashima and Nagashima, 2013), and since the PGC of Gemmatimonadota shows similar organization and close phylogenetic relationship to that of Proteobacteria, it was suggested they received it via HGT (Zeng et al., 2014). While HGT events in the same bacterial phylum are common (Igarashi et al., 2001; Nagashima and Nagashima, 2013; Brinkmann et al., 2018; Ward et al., 2018), this represents the first known example of HGT of complete PGC between different bacterial phyla (Zeng et al., 2014; Cardona, 2016).

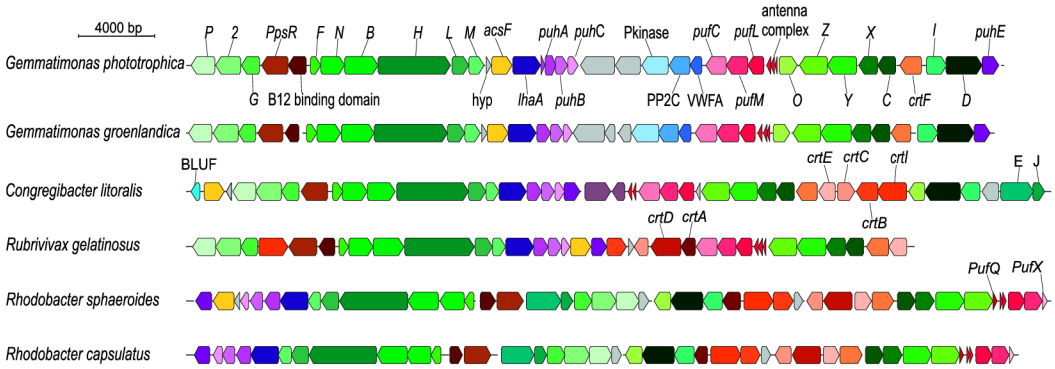

Figure 4. *Photosynthetic gene cluster of two photoheterotrophic cultures of Gemmatimonadota (G. phototrophica and G. groenlandica) and different photoheterotrophic Gammaproteobacteria (C. litoralis and R. gelatinosus) and Alphaproteobacteria (R. sphaeroides and R. capsulatus). Different colors indicate genes involved in bacteriochlorophyll biosynthesis (green), carotenoid biosynthesis (orange), genes encoding the reaction centre (puf operon-red), puh operon (pink/purple), other genes (brown and yellow), genes not involved in photosynthesis or hypothetical (blue/grey).*

BChl-containing Gemmatimonadota are facultative photoheterotrophs that cannot assimilate inorganic carbon and need a source of organic substrate (Zeng et al., 2014, 2015). The ability to use light provides them with additional energy, enables them to reduce respiration rate and increase the assimilation rate of organic compounds making them more efficient in carbon utilization and increasing their growth rate (Koblížek et al., 2020).

Gemmatimonadota are common in many natural environments (DeBruyn et al., 2011; Huang et al., 2016; Zeng et al., 2016). Until recently, Gemmatimonadota were mostly reported from soils, where they show higher abundancies then in other environments (DeBruyn et al., 2011; Delgado-Baquerizo et al., 2018). They were found in freshwaters (Morrison et al., 2017; Fecskeová et al., 2019), and often reported in freshwater sediments (Song et al., 2012; Zhang et al., 2015; Huang et al., 2016) but until now there have not been any studies that focused on the ecology, distribution and diversity of Gemmatimonadota in freshwater lakes.

2. Aims and objectives

The overarching aim of this work was to obtain more insights into the rarely studied bacterial phylum Gemmatimonadota and to reveal their distribution, diversity and potential ecological roles with the main focus on freshwater environments. Since some of the cultured members were capable of anoxygenic photosynthesis, a special interest of this study was also photoheterotrophy in this group.

One of the main ideas was to analyze the distribution of Gemmatimonadota in lakes and answer the question of their limnic character, that is, if they are truly present in the water column and not passively transferred to the lake from surrounding soil or sediments. Additionally, we also wanted to see if more photoheterotrophic members of Gemmatimonadota can be found in freshwater environments. To do so, we studied several European freshwater lakes located in the Czech Republic (Cep lake, Římov Reservoir and Jiřická pond) and Switzerland (Lake Zurich, Lake Constance and Lake Thun) and applied methods of amplicon sequencing, metagenomics and CARD-FISH. Finally, we have also chosen to explore the potential ecological roles and metabolic capabilities of Gemmatimonadota by using MAGs not only from freshwaters but also from several other environments like soil, marine waters, wastewaters or sediments. The objective of this part of the thesis was to obtain a global picture of the metabolic patterns and genomic differences of Gemmatimonadota across multiple environments.

3. Results

Paper I

Common presence of phototrophic Gemmatimonadota in temperate freshwater lakes

Mujakić, I., Andrei, A-Ş., Shabarova, T., Fecskeová, L.L., Salcher, M.M., Piwosz, K., Ghai, R., Koblížek, M. (2021)
mSystems 6:e01241-20

3.1. Summary results derived from Paper No. 1

We explored metagenomic data from five freshwater lakes with different trophic status that were collected over several years and seasons in order to answer the questions of distribution and diversity of Gemmatimonadota. Specifically, we wanted to see if they are truly limnic and not passively transferred to lakes via runoff water or by mixing of the sediments and to identify the most common freshwater heterotrophic and photoheterotrophic Gemmatimonadota.

Our results suggest that Gemmatimonadota in freshwater lakes represent only a small fraction of bacterial communities since their relative abundances were mostly around 1%. They were present at different depths of the lakes, both in epilimnion as well as hypolimnion, and their distribution appeared to follow seasonal patterns influenced by temperature and stratification of the lakes.

We reconstructed 45 freshwater Gemmatimonadota MAGs, almost half of which (19) contained photosynthetic genes. Based on phylogeny, we clustered the phototrophic Gemmatimonadota into three different groups. Group 1 (Pg1) probably represented novel photoheterotrophic genera. Group 2 (Pg2) clustered with culture *Gemmatimonas phototrophica* but was formed from MAGs, which, based on amino acid identity (AAI), represented several different species. Finally, group 3 (Pg3) also represented different species other than *G. phototrophica* but belonged to the same genus. The analysis also showed that the photosynthesis gene cluster (PGC) from limnic Gemmatimonadota has an identical or very similar organization as the PGC from *G. phototrophica*.

Since the majority of the Gemmatimonadota phylum consisted of uncultured environmental sequences, we used 16S rRNA gene sequences from newly assembled MAGs to define six new clusters and increase the taxonomic resolution of unclassified Gemmatimonadota.

Comparison of relative abundance of phototrophic and non-phototrophic Gemmatimonadota using fragment recruitment in metagenomes from Římov Reservoir showed the presence of phototrophic groups in the epilimnion and the highest contribution of non-phototrophic MAGs in summer in the hypolimnion.

Finally, we designed a CARD-FISH probe and for the first time visualized Gemmatimonadota cells from class *Gemmatimonadetes* in samples from the Římov reservoir. Small-sized free-living *Gemmatimonadetes* were mostly observed in the hypolimnion, whereas in the epilimnion Gemmatimonadota cells were often found associated with diatoms or cyanobacteria.

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Common Presence of Phototrophic *Gemmatimonadota* in Temperate Freshwater Lakes

Izabela Mujakić,[a,b] Adrian-Ştefan Andrei,[c] Tanja Shabarova,[c] Lívia Kolesár Fecskeová,[a] Michaela M. Salcher,[c] Kasia Piwosz,[a*] Rohit Ghai,[c] Michal Koblížek[a,b]

[a]Laboratory of Anoxygenic Phototrophs, Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czechia
[b]Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czechia
[c]Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Czechia

ABSTRACT Members of the bacterial phylum *Gemmatimonadota* are ubiquitous in most natural environments and represent one of the top 10 most abundant bacterial phyla in soil. Sequences affiliated with *Gemmatimonadota* were also reported from diverse aquatic habitats; however, it remains unknown whether they are native organisms or represent bacteria passively transported from sediment or soil. To address this question, we analyzed metagenomes constructed from five freshwater lakes in central Europe. Based on the 16S rRNA gene frequency, *Gemmatimonadota* represented from 0.02 to 0.6% of all bacteria in the epilimnion and between 0.1 and 1% in the hypolimnion. These proportions were independently confirmed using catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH). Some cells in the epilimnion were attached to diatoms (*Fragilaria* sp.) or cyanobacteria (*Microcystis* sp.), which suggests a close association with phytoplankton. In addition, we reconstructed 45 metagenome-assembled genomes (MAGs) related to *Gemmatimonadota*. They represent several novel lineages, which persist in the studied lakes during the seasons. Three lineages contained photosynthesis gene clusters. One of these lineages was related to *Gemmatimonas phototrophica* and represented the majority of *Gemmatimonadota* retrieved from the lakes' epilimnion. The other two lineages came from hypolimnion and probably represented novel photoheterotrophic genera. None of these phototrophic MAGs contained genes for carbon fixation. Since most of the identified MAGs were present during the whole year and cells associated with phytoplankton were observed, we conclude that they represent truly limnic *Gemmatimonadota* distinct from the previously described species isolated from soils or sediments.

IMPORTANCE Photoheterotrophic bacterial phyla such as *Gemmatimonadota* are key components of many natural environments. Its first photoheterotrophic cultured member, *Gemmatimonas phototrophica*, was isolated in 2014 from a shallow lake in the Gobi Desert. It contains a unique type of photosynthetic complex encoded by a set of genes which were likely received via horizontal transfer from *Proteobacteria*. We were intrigued to discover how widespread this group is in the natural environment. In the presented study, we analyzed 45 metagenome-assembled genomes (MAGs) that were obtained from five freshwater lakes in Switzerland and Czechia. Interestingly, it was found that phototrophic *Gemmatimonadota* are relatively common in euphotic zones of the studied lakes, whereas heterotrophic *Gemmatimonadota* prevail in deeper waters. Moreover, our analysis of the MAGs documented that these freshwater species contain almost the same set of photosynthesis genes identified before in *Gemmatimonas phototrophica* originating from the Gobi Desert.

KEYWORDS *Gemmatimonadota*, *Gemmatimonadetes*, anoxygenic phototrophs, aquatic bacteria, MAGs, metagenome, photosynthesis gene cluster, freshwater ecology, CARD-FISH

Citation Mujakić I, Andrei A-Ş, Shabarova T, Fecskeová LK, Salcher MM, Piwosz K, Ghai R, Koblížek M. 2021. Common presence of phototrophic *Gemmatimonadota* in temperate freshwater lakes. mSystems 6:e01241-20. https://doi.org/10.1128/mSystems.01241-20.

Editor Steven J. Hallam, University of British Columbia

Copyright © 2021 Mujakić et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Michal Koblížek, koblizek@alga.cz.

* Present address: Kasia Piwosz, Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, Gdynia, Poland.

Analysis of metagenome assembled genomes documented a widespread presence of phototrophic Gemmatimonadota in euphotic zones of freshwater lakes in central Europe.

Received 8 January 2021
Accepted 21 February 2021
Published 16 March 2021

Microorganisms conduct key biogeochemical processes involved in the main fluxes of matter and energy on Earth. Most microbial diversity remains uncultured, and only analyses of environmental DNA samples have made it possible to unravel existing microbial diversity and to identify the main species involved (1). Indeed, out of 112 known bacterial phyla in the Genome Taxonomy Database (GTDB) (2), more than one half are still recognized only from environmental sequences (3).

One of the phyla that was first identified using molecular phylogenetic methods was *Gemmatimonadota* (also called *Gemmatimonadetes* [4]), which was originally established as the so-called BD group based on five 16S rRNA gene sequences, which originated from deep-sea sediments, soils, and reactor sludge (5, 6). Independently, the group was also proposed as a candidate division KS-B based on three 16S rRNA gene sequences retrieved from coastal sediment samples from French Guiana (7).

The first cultured strain T27 belonging to the BD/KS-B group was isolated from a wastewater treatment plant in Japan. The isolate was named *Gemmatimonas aurantiaca* and established the new phylum *Gemmatimonadota*, along with its first class *Gemmatimonadetes* and genus *Gemmatimonas* (8). Subsequently, three more *Gemmatimonadota* genera (*Gemmatirosa*, *Longimicrobium*, and *Roseisolibacter*) with type strains were described from various soil environments (9–11). Apart from class *Gemmatimonadetes*, phylum *Gemmatimonadota* consists of four more class-level groups which include class *Longimicrobia*, two terrestrial groups (BD2-11 and S0134), and one marine benthic group (PAUC43f) (10, 12). The four cultured representatives from the phyla were all chemo-organoheterotrophs that require oxygen and grow under fully aerobic or semiaerobic conditions (8–11). An interesting metabolic potential and ecological role was reported for *G. aurantiaca*, as this species has the ability to reduce the greenhouse gas N_2O (13). However, with the discovery of *Gemmatimonas phototrophica*, which contains photosynthetic reaction centers (14, 15), *Gemmatimonadota* were added to several bacterial phyla containing anoxygenic phototrophic species alongside *Proteobacteria*, *Chlorobi* (now included as a class-level lineage in *Bacteroidota* [2]), *Chloroflexota*, *Firmicutes* (*Bacillota*), *Acidobacteriota*, and the newly discovered phylum "*Candidatus* Eremiobacterota" (WPS-2) (14, 16, 17). Anoxygenic phototrophs, such as *G. phototrophica*, are able to support their metabolism by harvesting light using bacteriochlorophylls; however, they require a supply of organic substrate for growth (18). Another characteristic found in *G. phototrophica* is the organization of its photosynthesis genes into a cluster called the photosynthesis gene cluster (PGC). Interestingly, the gene arrangement in the PGC of *G. phototrophica* is very similar to the one found in *Proteobacteria*, so it has been suggested that phototrophy in *Gemmatimonadota* originates from an ancient horizontal gene transfer of the entire PGC from *Proteobacteria* (14). As yet, this is the only known case of horizontal transfer of an entire set of photosynthesis genes between distant bacterial phyla (14, 19).

Information about the prevalence of *Gemmatimonadota* in different habitats is continuously growing, although information about their ecology is scarce. Members of this phylum were found in many natural environments (12, 20–22) and represent the eighth most abundant phylum in soils, accounting for about 1 to 2% of bacteria in soils worldwide (23). Their highest contributions are typically found in fertile agricultural and forest soils (20) but are also present in more unique soil environments, such as arid Antarctic Dry Valley soils (24, 25). It has been suggested that *Gemmatimonadota* may be relatively more abundant in dry soils (26). On the other hand, from the available data, it is known that they are also present in aquatic environments, such as freshwater lakes (27), sediments (22, 28–30), and estuaries (31, 32). In addition, *G. phototrophica* was isolated from a freshwater lake in Inner Mongolia (18). However, this organism does not grow in liquid culture and requires microaerophilic conditions, which are more typical for sediment-dwelling species. Thus, the data showing that *Gemmatimonadota* prefer dry environments does not seem to be universal. There is probably a large ecological and functional diversity

TABLE 1 Basic characteristics of the lakes studied

Characteristic	Římov Reservoir	Jiřická pond	Lake Zurich	Lake Constance	Lake Thun
Country[a]	CZ	CZ	CH	CH	CH
Latitude	48.50°N	48°36'N	47°18'N	47°32'N	46°41′N
Longitude	14.29°E	14°40'E	8°34′E	9°31'E	7°43′E
Altitude (m)	470	892	406	395	558
Area (km^2)	2.10	0.035	88.66	536	48.3
Volume (m^3)	34.5×10^6	6.59×10^3	3.3×10^9	48×10^9	6.5×10^9
Avg depth (m)	16.5	1.8	49	90	136
Maximum depth (m)	43	4.5	136	251	217
Mean hydraulic residence time	77 days	9 days	1.4 yrs	5 yrs	1.8 yrs
Trophic status	Mesoeutrophic	Dystrophic	Oligomesotrophic	Oligomesotrophic	Ultraoligotrophic
Mixing type	Dimictic	Polymictic	Monomictic	Monomictic	Monomictic
Sampling	2015−2017	2016−2017	2010−2019	2018	2018
No. of samples[b]	E = 10, H = 8	E = 5	E = 6, H = 4	E = 2, H = 2	E = 1, H = 1

[a]CZ, Czechia; CH, Switzerland.
[b]E refers to epilimnion, and H refers to hypolimnion. The exact depths for each of the lakes are provided in Materials and Methods.

among the members of *Gemmatimonadota*. The question remains whether *Gemmatimonadota*-related sequences identified in lakes originate from strictly limnic species, or perhaps they are just a passive component that enters the lakes along with runoff waters from surrounding soil.

Current progress in sequencing technologies and bioinformatics has circumvented the necessity for cultivated representatives and allowed biological and ecological inferences to be drawn by using genomic data recovered directly from microbial communities. Over the past few years, the usage of metagenome-assembled genomes (MAGs) has allowed the description of many novel bacterial divisions and unearthed large radiations in the prokaryotic tree of life (33). This approach has already led to the discovery of new phototrophic organisms belonging to the yet uncultured candidate phylum "*Ca*. Eremiobacterota" (16, 34). Also, two MAGs belonging to *Gemmatimonadota* were recovered from Lake Baikal. One MAG was more similar to *Gemmatirosa kalamazoonensis* found in soils and seemed more abundant at a depth of 5 m. The other MAG-encoded rhodopsin gene and was closely related to the phototrophic species *G. phototrophica* and showed a higher abundance at 20 m (35).

Therefore, in order to address the question whether there are any truly limnic *Gemmatimonadota* and to investigate their diversity, we analyzed metagenome data from five freshwater lakes in Czechia and Switzerland. The lakes were chosen based on their trophic status and included a representative mesoeutrophic Římov Reservoir, a dystrophic Jiřická pond, oligomesotrophic Lake Zurich and Lake Constance, and an ultraoligotrophic Lake Thun (Table 1). The metagenome sequences were collected over several years and seasons. The reconstructed MAGs were analyzed with the aim to identify the most common freshwater and photoheterotrophic *Gemmatimonadota* and to analyze their spatiotemporal variability. Using catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH), cells of *Gemmatimonadota* were visualized for the first time in their natural environment, and their association with other organisms was observed.

RESULTS

Abundance of *Gemmatimonadota*. The presence of *Gemmatimonadota* in freshwater lakes was first assessed by using relative 16S rRNA abundances extracted from our metagenomic data sets. In all studied lakes, *Gemmatimonadota* formed only a small part of the bacterial community with relative abundances typically below 1%. The highest relative abundance was found in Římov Reservoir, where *Gemmatimonadota* sequences were present over the entire sampling period (2015 to 2017). Interestingly, their relative abundance was higher in the hypolimnion (0.58% ± 0.23%; $n = 8$) than in the epilimnion (0.24% ± 0.21%; $n = 10$), with the highest numbers (1.03%) in August 2016. A similar pattern of higher contribution in the hypolimnion (0.44% ± 0.27%; $n = 4$) than in the

epilimnion (0.28% ± 0.25%; $n = 6$) was also observed in Lake Zurich, with the highest relative abundance occurring in spring (13 May 2013, 0.77%). The general pattern of higher abundance in the hypolimnion was observed also in both Lake Constance and Lake Thun and was recorded in summer with 0.81% and 0.58% relative abundance, respectively. Finally, the lowest contribution of *Gemmatimonadota* was found in Jiřická pond, where they represented less than 0.1% of the prokaryotic community, with a maximum of 0.09% recorded in summer, in August 2017.

Statistical analysis using distance-based linear models (DistLM) showed that the only environmental factor driving the *Gemmatimonadota* community in Římov Reservoir was water temperature, which explained 47% of the variability in the data set ($P = 0.0002$, pseudo-$F = 13.303$). A separate analysis of epilimnion and hypolimnion samples showed that temperature was also important in epilimnion ($P = 0.0079$, pseudo-$F = 5.7106$, and 41.6% of the explained variability in data set) but not in hypolimnion.

Metagenome-assembled genomes of *Gemmatimonadetes* and their distribution. To explore the diversity of aquatic *Gemmatimonadota*, we performed genome-resolved metagenomics analyses. The freshwater MAGs ($n = 45$) obtained in our study were analyzed together with those publicly available ($n = 226$) in March 2019 (see Table S1 in the supplemental material for a complete list). Most of the MAGs assembled by us were obtained from Římov Reservoir ($n = 27$), followed by Lake Zurich ($n = 10$), Lake Constance ($n = 5$), Lake Thun ($n = 2$), and Jiřická pond ($n = 1$). Publicly available MAGs originated from various environments: mostly from marine habitats ($n = 36$) and soils ($n = 37$), followed by sediments of soda lakes ($n = 38$), permafrost ($n = 23$), and other ($n = 11$). Only three MAGs from freshwater environments were available in the database at the time of the study.

Phylogenomically, based on Genome Taxonomy Database (GTDB) taxonomy (39), all MAGs recovered in this study clustered within class *Gemmatimonadetes* (Fig. 1). Although they were all recovered from a freshwater environment, six of them were more closely related to publicly available MAGs from environments like soils and sediments. MAGs CH-Jul18-bin44, ZH-3nov15-207, and TH-Jun18-bin75 clustered with MAGs from soils, and CH-Jul18-bin76, CH-Jul18-bin112, and ZH-3nov15-212 clustered with MAGs from sediments. Similarly, in 16S rRNA phylogeny (see Fig. S1 in the supplemental material), all freshwater MAGs with recovered 16S rRNA genes clustered within class *Gemmatimonadetes*, with the exception of two previously mentioned MAGs (CH-Jul18-bin112 and ZH-3nov15-212) that clustered within the BD2-11 terrestrial group based on SILVA SSU v138 database taxonomy (12).

16S rRNA gene diversity. The 16S rRNA gene fragments from metagenomes were classified taxonomically. As most of the *Gemmatimonadota* phylum consists of environmental sequences that remain uncultured, many of the fragments could not be classified. Therefore, in order to increase taxonomic resolution of unclassified *Gemmatimonadota*, we have used 16S rRNA gene sequences from our MAGs (Fig. S1) to define six new clusters (five from class *Gemmatimonadetes* and one from BD2-11 terrestrial group). In Římov Reservoir, the *Gemmatimonadota* community was solely composed of the class *Gemmatimonadetes* (Fig. 2). The genus *Gemmatimonas* was present in all seasons and both depths, but it was more abundant in the epilimnion, especially in summer (15-Aug-16). Epilimnion was dominated by the genus *Gemmatimonas* and unclassified *Gemmatimonadaceae*, with the exception of three samples (two in November and one in April) where other clusters were also present. In contrast, hypolimnion samples contained several different clusters, including phototrophic cluster PG1 that was one of the dominant clusters in every season. Similarly, clusters GRI1 and GSR7 were predominantly present in the hypolimnion. The presence of these three clusters in the epilimnion occurred only in November and April, when the reservoir was mixed. In Jiřícká pond, the *Gemmatimonadota* community was composed of three clusters from the class *Gemmatimonadetes*, i.e., genus *Gemmatimonas*, cluster GSR7, and cluster PG1, the latter being mostly dominant in summer. Additionally, the BD2-11 terrestrial group was also present in low

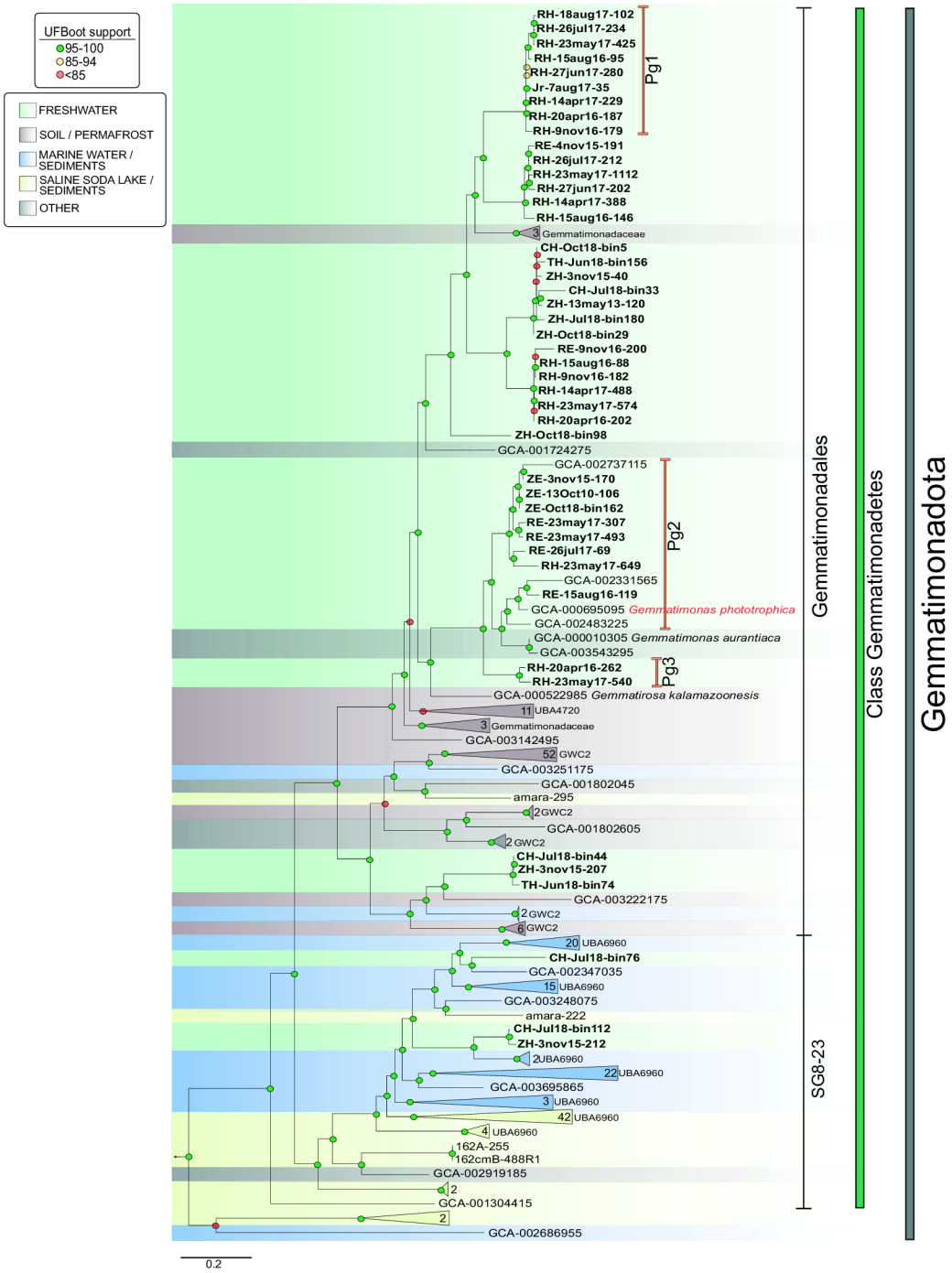

FIG 1 Maximum-likelihood tree (1,000 ultrafast bootstrap replicates, 121 conserved concatenated marker proteins) showing 45 MAGs of freshwater *Gemmatimonadota* as well as 226 MAGs and genomes of cultures collected from GTDB originating from different environments. The

(Continued on next page)

proportions in summer. The highest taxonomic diversity within *Gemmatimonadota* was observed in Swiss lakes, where three other class-level groups in addition to *Gemmatimonadetes* were recovered, i.e., representatives of classes *Longimicrobia*, BD2-11 and S0134, groups that were previously known only from terrestrial 16S rRNA gene sequences (12). Here, genus *Gemmatimonas* was also predominantly present in the epilimnion, whereas cluster GSR7 dominated the hypolimnion. In Lake Zurich, a large contribution of class *Longimicrobia* was recorded in a spring sample (13-May-13) in both the epilimnion and hypolimnion.

Photosynthesis genes in *Gemmatimonadota*. Further analyses of the assembled genomes of *Gemmatimonadota* revealed that 19 out of 45 freshwater MAGs contained phototrophic genes. In the phylogenetic tree, we could differentiate them into three distinct phototrophic groups (Fig. 1). Phototrophic group 1 contained eight MAGs from the hypolimnion of Římov Reservoir and one from Jiřická pond. Phototrophic group 2 was formed of nine MAGs (three from the epilimnion of Lake Zurich, five from the epilimnion of Římov Reservoir, and one from the hypolimnion of Římov Reservoir). Phototrophic group 2 also included the only cultured photoheterotrophic strain of the phylum, *G. phototrophica*, as well as the only three MAGs from freshwater environments that were publicly available. Phototrophic group 3 contained two MAGs from the hypolimnion of Římov Reservoir. All these MAGs were recovered from different time points (several years and different seasons).

Using representatives from each of these groups, we reconstructed the photosynthetic gene cluster and compared its structure and organization with the PGC found in *G. phototrophica* (14). For phototrophic groups 1 and 2, we reconstructed the entire PGC. In the case of phototrophic group 3, the PGC was missing several genes due to the incompleteness of the genomes (MAGs had completeness of 71.07. and 83.16%). From our findings, it appears that limnic *Gemmatimonadetes* have the same or very similar organization of phototrophic genes as *G. phototrophica* AP64 (Fig. 3 and Table S4). A distinct trait of *G. phototrophica* is a fragmented PGC divided by a set of hypothetical genes. The same split PGC was observed in three out of nine MAGs from phototrophic group 2, but it was absent in phototrophic groups 1 and 3. Furthermore, we observed the presence of the gene *frhB* (coenzyme F420-reducing hydrogenase, beta subunit) in the operon *bchP2G*, in the PGC of all members of phototrophic group 1, at the position where all other phototrophic *Gemmatimonadetes* have *bch2*. A full list of phototrophic genes for each MAG, as well as their distribution on contigs, is provided as Table S4.

As the presence of ribulose 1,5-bisphosphate carboxylase enzyme (RuBisCO) genes was previously reported in some sediment *Gemmatimonadota* MAGs, we also searched for them in our freshwater MAGs. We identified type IV RubisCO-like genes in eight phototrophic freshwater MAGs from Římov Reservoir (Fig. S3). Seven of these MAGs belonged to phototrophic group 1 recovered from hypolimnion, and one to phototrophic group 2 recovered from the epilimnion of Římov Reservoir.

Average amino acid identities and 16S rRNA and photosynthetic gene similarities of photoheterotrophic *Gemmatimonadetes*. We calculated average amino acid identities (AAI) between all the freshwater MAGs and the three genomes of cultured reference strains (*Gemmatimonas phototrophica*, *Gemmatimonas aurantiaca*, and *Gemmatirosa kalamazoonesis*) (Fig. S4). All MAGs of the phototrophic group 1 represent the same species, with AAI values of 99 and 100%, that was recovered in both Římov Reservoir and Jiřická pond, multiple times through different years and seasons. When phototrophic

FIG 1 Legend (Continued)
different environments are shown by different colors (left legend). The numbers shown at collapsed branches (i.e., 11 and 52) indicate the numbers of genomes (not shown) comprising the respective taxonomic categories. Genomes belonging to *Fibrobacterota* were used as an outgroup to root the tree. The annotations Pg1 to Pg3 define groups of photoheterotrophic *Gemmatimonadota*. Order, class-level, and phylum taxonomic labels are indicated through vertical delimiters (right part of the figure). The strength of support for internal nodes (assessed by ultrafast bootstrapping) is shown through colored circles (left legend). Details on all genomes can be found in Table S1 in the supplemental material.

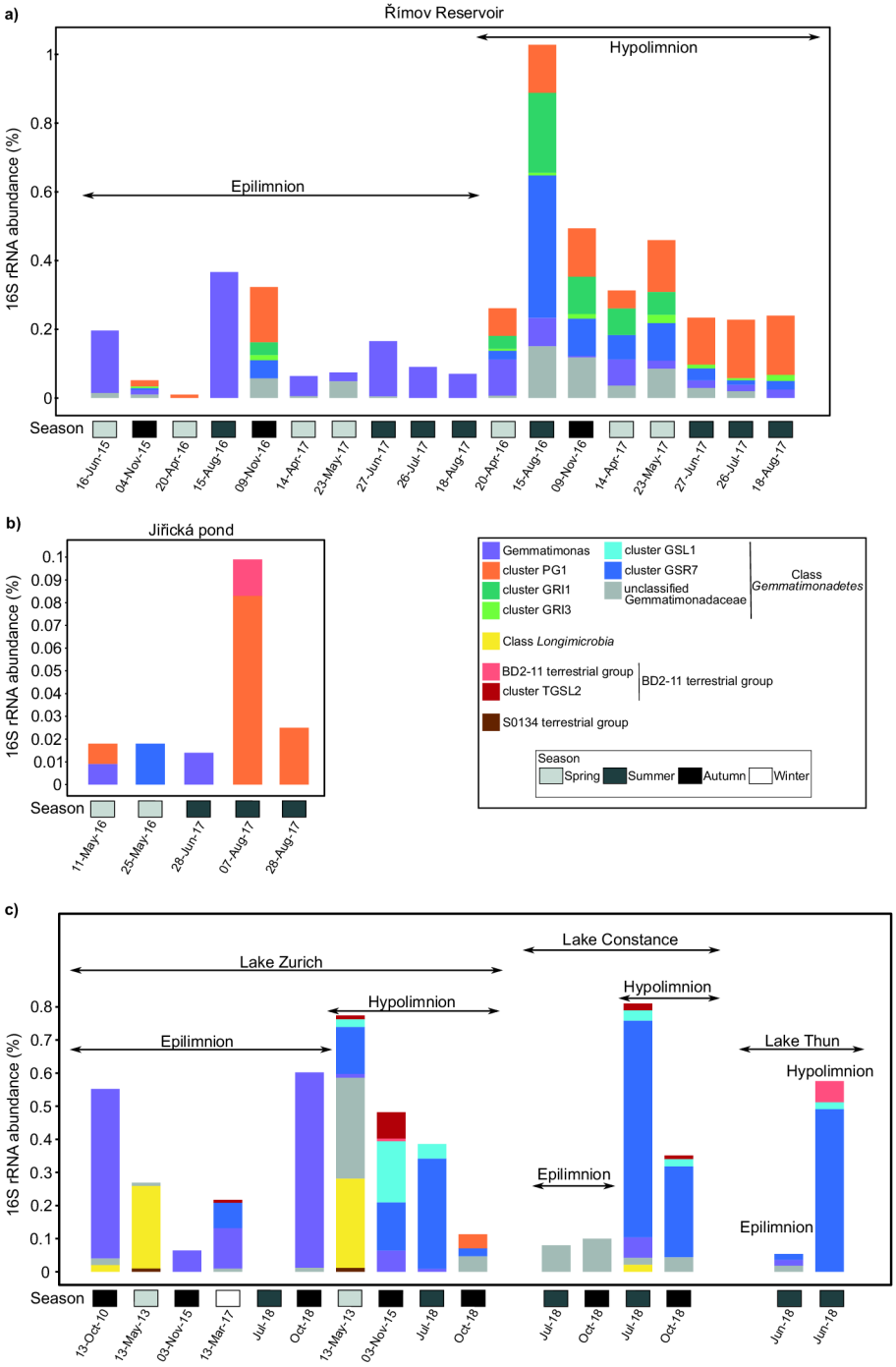

FIG 2 Relative abundance of *Gemmatimonadota* based on 16S rRNA gene fragments from metagenomes of five contrasting freshwater environments. These five freshwater environments are as follows: Římov Reservoir (a), Jiřická pond (b), and Lake

(Continued on next page)

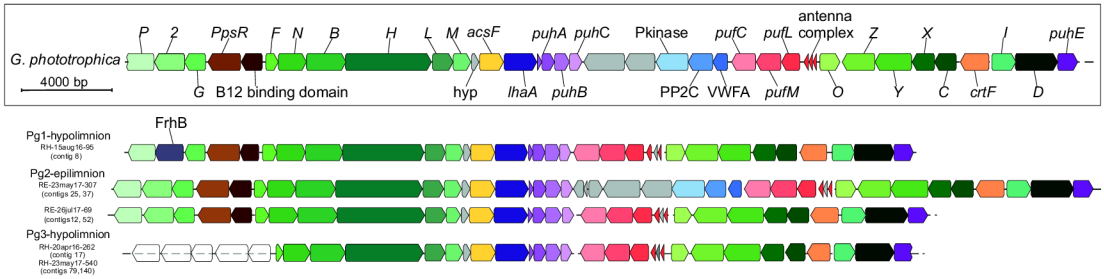

FIG 3 Structure and composition of photosynthetic gene clusters of *Gemmatimonas phototrophica* (top) and phototrophic MAGs of *Gemmatimonadota* that clustered into three photoheterotrophic groups. The names of the MAGs used to depict each photosynthetic group (photosynthetic group 1 [Pg1] to Pg3) are shown at the right part of the figure. *bch* genes involved in bacteriochlorophyll biosynthesis (green), *puf* operon encoding the reaction center (red), *puh* operon (pink/purple), carotenoid biosynthesis genes (orange), *ppsR* gene and B12 binding domain (brown), *acsF* (yellow), *lhaA* (dark purple), genes not involved in photosynthesis (blue), and hypothetical genes (gray) are indicated. A list of all genes is provided in Table S4.

group 1 was compared with phototrophic groups 2 and 3, AAI values were below 65%, suggesting that these groups represent different genera. Similarly, phototrophic group 1 showed AAI values below 65% (specifically 55 and 56%) compared with cultured representatives. In phototrophic group 2, the three MAGs from Lake Zurich represent the same species (AAI, 95 to 100%) recovered in different years, whereas other MAGs from Římov Reservoir in this group seem to belong to the same genus but are all different species (AAI, 77 to 82%). Phototrophic group 3 consists of two MAGs that represent the same species (AAI, 96 and 100%). Both phototrophic groups 2 and 3 (AAI, 66 to 68%) belong to the same genus as the cultured representatives *G. phototrophica* and *G. aurantiaca* (AAI 66%) but are different from *Gemmatirosa kalamazoonensis* (AAI, 56%).

From the phototrophic MAGs (phototrophic group 1), one MAG from the hypolimnion of Římov Reservoir (RH-18aug17-102) contained a 16S rRNA gene, which showed 92.8% identity with *G. phototrophica* and 91.5% and 89.6% identity with *G. aurantiaca* and *Gemmatirosa kalamazoonensis*, respectively (Fig. S1), and one MAG from Jiřická pond (Jr-7aug17-35) showed 92.1% 16S rRNA gene sequence identity with *G. phototrophica*.

To further analyze the differences among the identified phototrophic groups, we compared the *pufM* and *acsF* gene sequence identities between them and type species *G. phototrophica*. Both of these genes are markers for anoxygenic photoheterotrophic bacteria (22, 40). For the *pufM* gene (encoding the M subunit of the bacterial photosynthetic reaction center), phototrophic group 2 showed only 80 to 85% identity with *G. phototrophica*. Much lower identities were found for hypolimnion groups 3 and 1 (77 to 80% and 72%, respectively). The *acsF* gene (aerobic oxidative cyclase gene) gave similar results with only 71% identity between phototrophic group 1 and *G. phototrophica*. MAGs from phototrophic groups 2 and 3 had 79 to 89% and 78% identity of the *acsF* gene of *G. phototrophica*, respectively.

Relative abundance of MAGs of phototrophic and nonphototrophic *Gemmatimonadota*. In order to compare the relative abundance of phototrophic and nonphototrophic *Gemmatimonadota*, we used 28 metagenomes from Římov Reservoir from which 27 MAGs were recovered for fragment recruitment (Fig. 4). Nonphotosynthetic MAGs showed the highest contribution in the hypolimnion, especially in summer (15 August 2016); however, their abundance was not constant and changed over time. Similarly, photosynthetic group 1 was also present mostly in the hypolimnion, and its

FIG 2 Legend (Continued)

Zurich, Lake Constance, and Lake Thun (c). The figure depicts the classification of 16S rRNA gene fragments (as unassembled shotgun reads) retrieved from five freshwater data sets. The *x* axis shows the sampling dates, while the *y* axis indicates the percentage of *Gemmatimonadota* within the prokaryotic communities. The sample collection time, following a four-season breakdown, is indicated by the gray-scale-colored boxes arranged along the *x* axis.

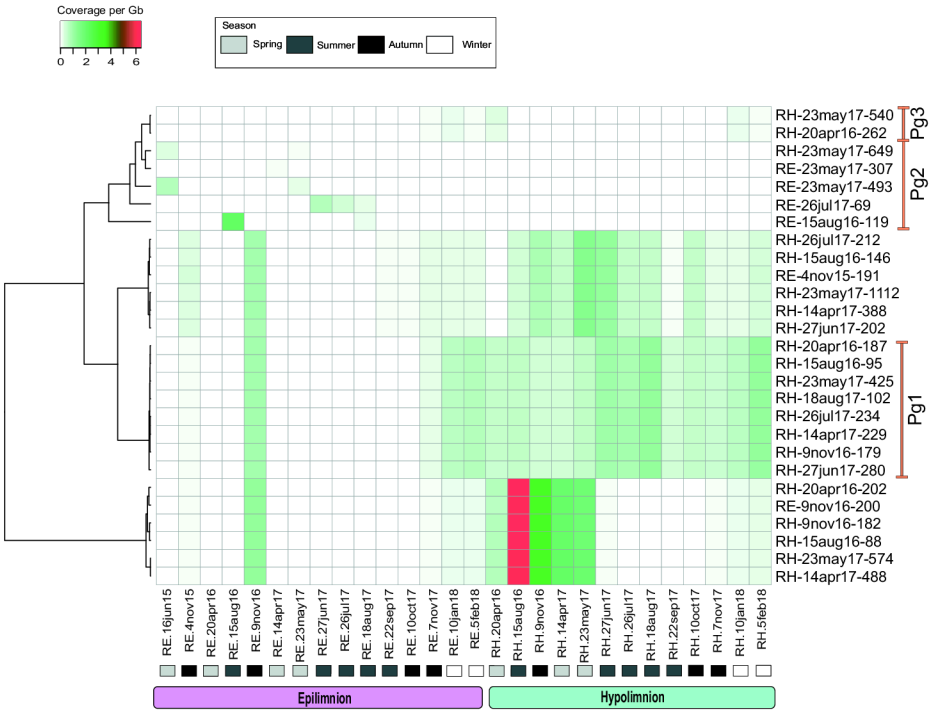

FIG 4 Heatmap of the relative abundance of 27 *Gemmatimonadota* MAGs gained from Římov Reservoir in 28 metagenomes from Římov Reservoir (Table S3). MAGs are clustered using complete linkage and Euclidian correlation. Columns are labeled with the sampling date, the depth of the sample (RE = epilimnion [0.5 m] and RH = hypolimnion [30 m]) and season. Three groups of MAGs (Pg1 to Pg3) containing photosynthetic gene clusters representing photoheterotrophic *Gemmatimonadota* are labeled with orange lines. Color key of abundance (coverage per gigabase) is shown in the top left corner, with red indicating the highest abundance and white indicating that a MAG is not present in a metagenome.

relative abundance varied during the season. The only occurrence of this group in the epilimnion was in late autumn (9 November 2016), at times of deep mixing of the reservoir (41). In contrast, photosynthetic groups 2 and 3 were present mostly in the epilimnion but remained at very low relative abundances in all seasons and at all depths.

CARD-FISH analyses. In order to visualize members of the class *Gemmatimonadetes* and access more information on their distribution and potential associations within freshwater environments, we designed the CARD-FISH probe Gemma_801 and applied it to a set of samples obtained from Římov Reservoir, comprising eight longitudinal transects collected during summer 2015. The hybridized cells were also counted; however, since the probe *in silico* matches only 35% of all *Gemmatimonadetes*, the presented numbers, while corresponding with 16S rRNA gene abundance from metagenomes, reflect only the detected, and not total *Gemmatimonadetes*, implying that the observed numbers can underestimate the absolute abundance of *Gemmatimonadetes* in the environments studied. The lowest relative (mean = 0.09%, minimum [min] = 0.00, maximum [max] = 0.19%) and absolute (mean = 3.25×10^3 cells ml^{-1}, min = 0.00, max = 7.9×10^3 cells ml^{-1}) abundances were observed at the river station of the reservoir (Fig. S2). However, at stations 2 and 3 (stations following the "river station" in the longitudinal transect), the contribution of hybridized planktonic cells had already reached up to 1.0% (4.3×10^4 cells ml^{-1}) during a bloom of *Fragilaria* sp. at the end of June. Besides free-living *Gemmatimonadetes*, we observed hybridized cells that seemed to be attached to diatoms (Fig. 5a). On some *Fragilaria* sp. colonies, hybridized *Gemmatimonadetes* contributed up to 22% of all bacterial cells detected on the diatom

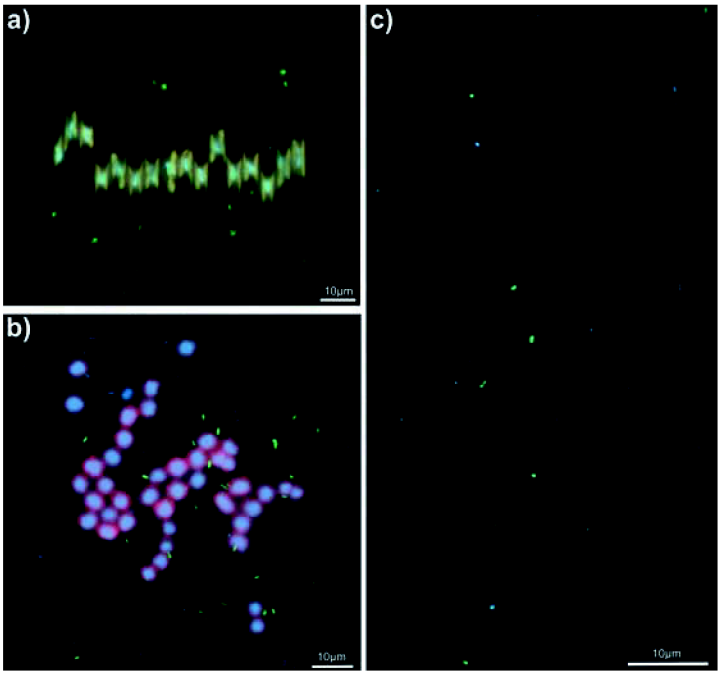

FIG 5 Overlay images of *Gemmatimonadetes* cells in Římov Reservoir visualized by CARD-FISH with probe Gemma_801. (a) *Gemmatimonadetes* associated with a *Microcystis* sp. colony. (b) *Gemmatimonadetes* observed on a colony of *Fragilaria* sp. (c) Free-living *Gemmatimonadetes* observed in the hypolimnion of Římov Reservoir. The probe signal is displayed in green, DAPI staining is displayed in blue, and autofluorescence of *Cyanobacteria* is displayed in red.

surface. In addition, *Gemmatimonadetes* seemed to be associated with cyanobacteria, as higher densities were detected with agglomerations of *Microcystis* sp. colonies (Fig. 5b and Fig. S5a). However, the highest numbers of hybridized cells were observed in the hypolimnion ($\geq$10 m) of Římov Reservoir (up to 4.43% and 2.75×10^4 cells ml^{-1}, respectively; Fig. 5c, Fig. 6, and Fig. S5b). These cells displayed different shapes and were smaller than those detected in the epilimnion.

DISCUSSION

In the presented work, we clearly demonstrated that *Gemmatimonadota* were permanently present in all of the studied lakes showing the ubiquitous nature of this phylum in contrasting freshwater lakes ranging from ultraoligotrophic to mesoeutrophic. The freshwater *Gemmatimonadota* community followed seasonal patterns with water temperature as the main driving variable, especially in the epilimnion. This observation further supports the idea that the studied *Gemmatimonadota* are truly limnic, since abundances of passively transferred microorganisms correlate with water mass movements rather than with temperature (42). The stratification of the lake also seemed to have a significant effect on the *Gemmatimonadota* community, as clusters inhabiting the epilimnion differed from those present in the hypolimnion, suggesting that they can be metabolically diverse and be specialized in different ecological roles. While *Gemmatimonadota* were relatively stable in the hypolimnion, their numbers in the epilimnion varied with higher numbers in late summer and autumn and lower numbers in spring and early summer. The stratification effect was previously studied in Grand Lake (OK, USA) where the *Gemmatimonadota* community was present in late summer in the free-living fraction within the thermocline and hypolimnion. The highest abundance in

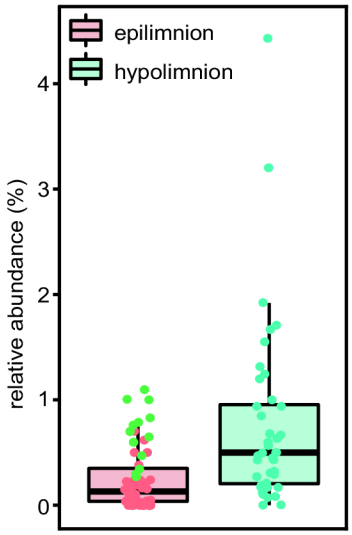

FIG 6 Relative abundance of planktonic *Gemmatimonadetes* in the epi- and hypolimnion of Římov Reservoir obtained by CARD-FISH analyses with the probe Gemma_801. Epilimnion samples where *Fragilaria* sp. and *Microcystis* sp. were observed are colored in green. The difference between the epilimnion and hypolimnion is significant (Mann-Whitney U test, [$P = <0.001$]).

the hypolimnion at this time seemed to be associated with the sedimentation of organic matter, turbidity, and a lack of oxygen that developed in the thermocline and hypolimnion (43). The *Gemmatimonadota* community in Grand Lake was mostly composed of the genus *Gemmatimonas*, which is suggested to participate in the degradation of organic matter after an algal bloom lysis (44). In contrast, we observed that the genus *Gemmatimonas* was present in both the epilimnion and hypolimnion, but it appeared with higher abundances in the epilimnion, while the hypolimnion community was dominated by other clusters, including phototrophic group 1. However, we could also correlate the highest abundance of the genus *Gemmatimonas* in the epilimnion (15-Aug-16) with the highest abundance of the cyanobacterial community in the same metagenome sample (41). The prokaryotic community of metagenomes from Římov Reservoir was analyzed previously (41) showing *Actinobacteria* as persistently abundant in Římov Reservoir through all seasons and in both the epilimnion and hypolimnion. Other more abundant groups in the epilimnion and the hypolimnion were *Alphaproteobacteria*, *Bacteroidetes*, and *Burkholderiales* (previously *Betaproteobacteriales*). Instead, while *Cyanobacteria* were abundant in the epilimnion, they were recovered in smaller numbers, which the authors attribute to the method of sampling where most of the filamentous *Cyanobacteria* are removed.

The clear difference between samples from the epilimnion and hypolimnion was also observed using epifluorescence microscopy. The highest number of small-sized hybridized cells was detected in the hypolimnion of Římov Reservoir. In contrast, the samples from the epilimnion contained many *Gemmatimonadota* cells attached or associated with photosynthetic organisms: diatoms (*Fragilaria* sp.) or cyanobacteria (*Microcystis* sp.) (Fig. 5a and b and Fig. 6). Since both phototrophic and heterotrophic *Gemmatimonadota* require an organic source of carbon, they may benefit from the input of organic carbon from algae and cyanobacteria, providing in exchange inorganic nutrients acquired through mineralizing organic substances (45). Additionally, cells from the epilimnion, where light is available, seemed to have a larger size. Some previous studies showed that aerobic anoxygenic photoheterotrophic bacteria can

often increase carbon assimilation in the presence of light, which allows them to have larger cells (46, 47). Subsequently, the distribution of *Gemmatimonadota* in the epilimnion could be dependent upon the distribution and occurrence of diatoms like *Fragilaria* sp. and cyanobacteria like *Microcystis* sp., as their blooms can influence and alter bacterioplankton communities (45, 48). Likewise, this dependence is a plausible explanation for the reason why photoheterotrophic members of this phylum are proving to be difficult to culture.

The presented 45 metagenome-assembled genomes of *Gemmatimonadota* in this study represent the largest collection of freshwater metagenome-assembled genomes (MAGs) of this phylum so far available. The obtained MAGs further confirmed the limnic nature of *Gemmatimonadota*. With recovery of several MAGs that represent the same species based on AAI (with 99 or 100% similarity), we show that *Gemmatimonadota* MAGs are reassembled from metagenomes and recovered periodically from different years as well as different seasons. This confirms the strong resilience of their microdiversity in freshwater and cannot be taken as a random occurrence. Furthermore, 39 of the obtained MAGs are different from soil species (Fig. 1; also see Fig. S1 and S4 in the supplemental material) and form distinct phylogenetic groups (with AAI between 45 and 65%) with both photoheterotrophic and heterotrophic representatives (38). Six MAGs, gained from the hypolimnion of deep lakes, three from Lake Constance (CH-Jul18-bin44, CH-Jul18-bin76, and CH-Jul18-bin112), two from Lake Zurich (ZH-3nov15-207 and ZH-3nov15-212), and one from Lake Thun (TH-Jun18-bin75), (Fig. 1 and Fig. S1) are more closely related to genomes from soil and sediments. In the 16S rRNA phylogeny (Fig. S1), two previously mentioned MAGs (CH-Jul18-bin112 and ZH-3nov15-212) clustered within the BD2-11 terrestrial group based on SILVA SSU v138 database taxonomy (12). This is consistent with 16S rRNA gene abundance in Swiss lakes where we could episodically observe the occurrence of 16S rRNA sequences related with terrestrial groups, such as *Longimicrobia*, BD2-11, and S0134 that has, as yet, been connected only with soil environments (10). Environmental 16S rRNA gene sequences from freshwater have until now been associated only with the class *Gemmatimonadetes* (10–12), and all other groups were formed with soil and sediment representatives. While the six MAGs could represent new freshwater members of the soil-connected groups, due to the close phylogenomic similarity with soil representatives and not with other freshwater MAGs, it is more probable that they represent genomes recovered due to soil runoff. Nevertheless, the MAGs present in these different phylogenetic groups were assembled from metagenomes gained from different sampling times which shows *Gemmatimonadota* were recovered repeatedly from freshwater environments. All this evidence shows that at least 39 of the identified MAGs represent truly limnic and planktonic species that do not come from soils. Furthermore, with the notable genomic diversity of limnic *Gemmatimonadota*, we demonstrate the ecological relevance of this group, as different members are clearly able to persist in the water column, occupying different ecological niches as they occur both in the hypolimnion and in the epilimnion. Moreover, the distribution of *Gemmatimonadota* in contrasting freshwater lakes showed they are adapted to different types of aquatic environments. The highest diversity in terms of different genus level groups or species of *Gemmatimonadota*, especially photoheterotrophic representatives, was observed in Římov Reservoir. Despite this, we cannot associate the higher diversity to mesoeutrophic lakes, as the data set obtained from Římov Reservoir was substantially larger, allowing for an increased chance of recovering higher diversity. A higher trophic status of any lake is connected with higher phytoplankton productivity; therefore, lakes and reservoirs often show variations of microbial communities connected with phytoplankton productivity (49, 50). Specifically, the connection of bacterial communities with changes in phytoplankton has already been recorded in Římov Reservoir (51). Furthermore, anoxygenic phototrophic bacteria are often found in close association with algae (45, 52, 53), and since they seem to follow seasonal phytoplankton blooms in freshwater lakes (45, 48), it is suggested that they represent an important functional group in freshwater environments (54).

Phototrophic gene, phylogenomic, and AAI analyses have shown that phototrophy in the *Gemmatimonadota* spans multiple genera. The identified phototrophic *Gemmatimonadota* represent three different groups but share the same set of phototrophic genes with *G. phototrophica*. All identified PGCs share a very similar gene inventory (see Table S4 in the supplemental material) and organization (Fig. 3). It seems that the convergent orientation (→ ←) of superoperons *bchFNBHLM* and *crtFbchCXYZ-puf* is conserved among all the phototrophic *Gemmatimonadota*, in contrast to phototrophic *Proteobacteria*, where orientation of these superoperons may be divergent, convergent, or colinear (55). Moreover, the split of the PGC with inserted genes in the type strain *G. phototrophica* AP64 (14) is also present in several MAGs from the epilimnion. Previously, the PGC of *G. phototrophica* was compared with two PGCs reconstructed from the Odense wastewater metagenome (OdenseWW) and the Aalborg wastewater metagenome 2 (AalborgWW-2) which did not contain the insert of several hypothetical genes between two operons. The explanation proposed for the difference in having a split PGC or not was that it could represent different evolutionary history of phototrophic *Gemmatimonadota* originating from different environments (21). Apart from the split PGC present in some of the MAGs, we observed that members of phototrophic group 1 have coenzyme F420-reducing hydrogenase (*frhB* gene) instead of bacteriochlorophyll synthase 4.5-kDa chain (*bch2*), as part of the *bchP2G* operon. Coenzyme F420-reducing hydrogenase enzyme seems to have an important role in energy conservation and methanogenesis from CO_2 (56).

In support of the true photoheterotrophic nature of *Gemmatimonadota*, a previous study conducted in Římov Reservoir, found active expression of their *pufM* genes (a common molecular marker gene for aerobic anoxygenic phototrophs) (27). Expression of the photosynthetic apparatus of *Gemmatimonadota* showed that they are an active part of bacterial community and do not just passively contain the phototrophic genes (27).

Recently, several *Gemmatimonadota* MAGs that originate from sediments of a soda lake were reported to contain both phototrophic genes and genes related to the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase enzyme (57), suggesting that these soda lake MAGs represent the first photoautotrophic *Gemmatimonadota* (58). In contrast, some of our freshwater *Gemmatimonadota* contain phototrophic genes as well as type IV RuBisCO (Fig. S3), which is considered only a homologue of RuBisCO, since it does not have any carboxylation activity (59, 60). Type IV RuBisCO genes are present in many microorganisms, including both phototrophic and heterotrophic bacteria and *Archaea* and are thought to participate in some other metabolic pathways different from the Calvin-Benson cycle (61, 62). Therefore, these freshwater *Gemmatimonadota* MAGs are not photoautotrophs, rather have a photoheterotrophic metabolism, typical for aerobic anoxygenic photoheterotrophic bacteria. These bacteria do not fix inorganic carbon and need to rely on organic carbon produced by other organisms, so the ability to harvest light is used to supply energy for their mostly organoheterotrophic metabolism (40, 63, 64).

In conclusion, with MAGs from these contrasting freshwater lakes, we not only reveal the existence of several new phototrophic species that differ phylogenetically from the already cultured and characterized *G. phototrophica* but also show the considerable diversity of both photoheterotrophic and heterotrophic *Gemmatimonadota* in freshwater.

MATERIALS AND METHODS

Sampling for metagenomics analysis. Samples from five European freshwater lakes representing a large range of size, depth, and/or trophic status (Table 1) were used to obtain genomic information from *Gemmatimonadota*. Samples were taken from the surface mixed-sun exposed layer known as the epilimnion and the deeper, colder layer where typically there is reduced turbulence and a smaller amount of light is present. This deeper layer is called the hypolimnion. Římov Reservoir is a mesoeutrophic, canyon-shaped dimictic water body that was built during 1970s by damming a 13.5-km-long section of the River Malše (65). The sampling was performed between June 2015 and August 2017, above the deepest point of the reservoir by using a Friedinger sampler. A volume of 20 liters of water was collected from

both the epilimnion (0.5 m; $n = 10$) and hypolimnion (30 m; $n = 8$) and subjected to sequential peristaltic filtration through a series of 20-, 5-, and 0.2-μm-pore-size polycarbonate membrane filters (142 mm diameter) (Sterlitech Corporation, USA). Characteristics of the water column, depth, temperature, oxygen (GRYF XBQ4; Havlíčkův Brod, Czechia) and chlorophyll *a* (FluoroProbe TS-16-12; bbe Moldaenke, Kiel, Germany) were also measured. The sample collection and filtration steps were identical for the rest of the lakes unless otherwise stated. Jiřická pond is a dystrophic humic water body from which five epilimnion (0.5-m depth) water samples were collected between May 2016 and August 2017. Lake Zurich is an oligomesotrophic, perialpine monomictic water body, from which nine water samples were collected in a period between 2010 and 2018 from the epilimnion (5-m depth; $n = 5$) and hypolimnion (80/120-m depth; $n = 4$) of the lake and processed as described above. Lake Thun is an ultraoligotrophic, alpine water body. Two water samples were collected in June 2018 from 5-m and 180-m depth. Lake Constance is a large oligomesotrophic, perialpine lake. Four samples, which were collected in July and October 2018 from the epilimnion (5 m) and hypolimnion (200 m) were used for this study.

DNA extraction, sequencing, and assembly. DNA was extracted from the 0.22-μm filters (0.2- to 5-μm fraction) using the ZR Soil Microbe DNA MiniPrep kit (Zymo Research, Irvine, CA, USA) in accordance with the manufacturer's instructions. The total quantity of DNA was estimated using the Qubit dsDNA BR assay kit (Life Technologies, Foster City, CA, USA) on a Qubit 2.0 fluorometer (Life Technologies). DNA integrity was assessed by agarose gel (2%) electrophoresis and SYBR green I staining. Shotgun sequencing was performed using the Novaseq 6000 sequencing platform (2×150 bp) (Novogene, Hong Kong, China). Raw Illumina metagenomic reads were quality filtered in order to remove low-quality bases/reads and adapter sequences using bbmap package (66, 67). Briefly, the paired-end (PE) reads were interleaved by reformat.sh and quality trimmed by bbduk.sh (qtri = rltrimq = 18) (68). Subsequently, bbduk.sh (68) was used for adapter trimming and identification/removal of possible PhiX and p-Fosil2 contamination (k = 21 ref = vector file ordered cardinality). Additional checks (i.e., *de novo* adapter identification with bbmerge.sh [67]) were performed in order to ensure that the data sets met the quality threshold necessary for assembly. The obtained quality-filtered data sets were then assembled independently with MEGAHIT (v1.1.5) (69) using the k-mer sizes: 39, 49, 69, 89, 109, 129, 149, and default parameters.

16S rRNA abundance-based taxonomic classification. The obtained quality-filtered data sets were converted to FASTA format and randomly subsampled to 20 million reads by using reformat.sh (68). These subsets (containing 20 million sequences each) were queried against the SILVA SSU database, release 132 (70), in order to identify RNA-like sequences by using MMSeqs2 (71) and an E-value cutoff of 1e−3. The bona fide 16S rRNA gene sequences (as identified by SSU-ALIGN [72]) were further compared by blastn, in nucleotide space (using as cutoff the E value 1e−5), against the SILVA SSU database amended with 16S rRNA genes recovered from *Gemmatimonadota* MAGs (see below), and classified if the sequence identity was ≥80% and the alignment length was ≥90 bp (sequences failing these thresholds were not used for downstream analyses). The taxonomic affiliation of each identified 16S rRNA read was inferred based on its best blastn hit. The relative abundances of *Gemmatimonadota* taxonomic categories were calculated as a percentage of total 16S rRNA reads.

The statistical relationships between environmental data (oxygen, temperature, and chlorophyll *a*) (41) and *Gemmatimonadota* abundance in Římov Reservoir was analyzed by distance-based linear models and nonmetric multidimensional scaling (nMDS) in the PERMANOVA+ add-on package of the PRIMER7 software (Primer Ltd., Lutton, UK). Abundance data were square root transformed, and analysis was done using a stepwise selection procedure. The best model was selected based on statistical significance (9,999 permutations), and the value of the Akaike's information criterion (AICc) (73, 74). The same analysis was also done separately for the epilimnion and the hypolimnion.

Recovery of bacterial genomes. Quality-filtered metagenomics data sets were mapped using bbwrap.sh (75) (kfilter = 31 subfilter = 15 maxindel = 80) against the assembled contigs (longer than 3 kb) in a lake-dependent fashion. The resulting BAM files were used to generate contig abundance files with jgi_summarize_bam_contig_depths (–percentIdentity 97) (76). The contigs and their abundance files were used for binning with MetaBAT2 (default settings) (76). Bin completeness, contamination, and strain heterogeneity were estimated using CheckM (with default parameters) (77). Bins with estimated completeness above 40% and contamination below 5% were denominated as metagenome-assembled genomes (MAGs). MAGs were taxonomically classified with GTDB-Tk (2) using default settings.

Phylogenomics. MAGs belonging to *Gemmatimonadota* ($n = 45$), together with reference genomes ($n = 226$) recovered from public repositories (see Table S1 in the supplemental material) were annotated using the TIGRFAMs database (78). A total of 121 conserved marker proteins (Table S2) were extracted from the annotated *Gemmatimonadota* genomes. MAGs that had more than 40% markers present, together with reference genomes were used for phylogenetic reconstruction. Briefly, homologous proteins were independently aligned with PRANK (default settings) (79), trimmed with BMGE (-t AA -g 0.5 -b 3 -m BLOSUM30) (80), and concatenated. A maximum-likelihood phylogeny was constructed using IQ-TREE (81) with the LG+F+R10 substitution model (chosen as the best-fitting model by ModelFinder [82]) and 1000 ultrafast bootstrap replicates. Genomes appertaining to *Fibrobacterota* were used as an outgroup to root the tree.

The average amino acid identity (AAI) within coherent phylogenomic groups was determined by performing whole-genome pairwise coding DNA sequence (CDS) comparison, using BLAST, as previously described (83). Taxonomic categories for the MAGs were defined using the suggested standards (38).

The photosynthetic gene clusters within obtained phylogenomic groups were analyzed in Geneious Prime 2019.2.3. For each group, MAGs with the most complete photosynthetic gene cluster were chosen

as representatives. The alignment and identity matrix for nucleotide *pufM* and *acsF* gene sequences from photoheterotrophic MAGs and *G. photorophica* was done with ClustalW 2.0.10 (84).

Phylogenetics. The 16S rRNA sequences present in the recovered MAGs were identified by SSU-ALIGN (72). The ones with a length longer than 200 nucleotides (nt) were merged with a data set comprising *Gemmatimonadota* sequences recovered from SILVA SSU v138 database (reference NR 99; length, 1200 bp) (70). The sequences were aligned with PASTA v1.8.3 (85) and used to construct a maximum-likelihood phylogeny with IQ-TREE v1.6.10 (-m GTR+F+R10; chosen as the best fitting model by ModelFinder) (81, 82). Several sequences appertaining to *Cyanobacteria* and *Fibrobacterota* were used as outgroup to root the tree.

Eighteen proteins belonging to the large subunit of the ribulose 1,5-bisphosphate carboxylase enzyme (RuBisCO) were identified in the assembled MAGs ($n = 106$). These proteins, together with a data set comprised of 146 RuBisCO (types I to III) and RuBisCO-like (type IV) proteins (86) were treated with PREQUAL v1.02 (87) prior to alignment with PASTA v1.8.3 (85). The obtained alignment (854 aligned positions) was used to construct a maximum-likelihood phylogeny with IQ-TREE v.1.6.10 and the LG+F+R6 substitution model (chosen as the best-fitting model by ModelFinder) (81, 82).

Fragment recruitment. The obtained MAGs and the metagenomic data sets were used to compute genome coverage per gigabase using RazerS 3 (using cutoffs of >95% identity and alignment lengths of ≥50 bp) (88). All rRNA sequences (5S, 16S, and 23S) present in the MAGs were identified using rrna_hmm (89) and were masked prior to comparisons with quality-filtered metagenomic sequences. The quality-filtered data sets were used to compute abundance profiles for the *Gemmatimonadota* MAGs. Raw data showing coverage per gigabase is shown in Table S3, and a heatmap (Fig. 4) was created using http://heatmapper.ca.

Design of oligonucleotide probes. We designed a new oligonucleotide probe that targets a part of the class *Gemmatimonadetes* (Fig. S1). Originally, the probe was designed as a *Gemmatimonadetes*-specific broad-range (degenerate) PCR primer. *Gemmatimonadetes* and non-*Gemmatimonadetes* 16S rRNA gene sequences were retrieved from SSURef SILVA database (release 132 [70]), and aligned with the MUSCLE60 algorithm in MEGA7 software (90). The primers were then designed manually to target conserved regions specific for *Gemmatimonadetes*. The specificity of the primers was first tested *in silico* with the SILVA database (70) and then with PCR using DNA from *Gemmatimonas phototrophica* which was used as a positive control. Optimized PCR conditions were as follows: initial denaturation at 98°C for 3 min; 26 cycles with 1 cycle consisting of 98°C for 10 s, 60°C for 20 s, and 72°C for 20 s; and final elongation for 3 min. Primers were additionally tested with several other bacterial strains, which served as negative controls and showed no amplification. Finally, the primers were used on an environmental sample where their specificity was confirmed with Illumina amplicon sequencing. Briefly, three different primer sets were applied to an environmental sample collected at station DAM of Římov Reservoir. After Illumina sequencing, read quality was evaluated using FastQC v0.11.7 (Babraham Bioinformatics, Cambridge, UK). Primer sequences were trimmed using cutadapt v1.16 (91) and subsequently analyzed in the R/Bioconductor environment using the dada2 v1.6 package (37). Taxonomic assignment was performed using the SILVA 132 database (70). On the basis of these results, we opted to use the reverse primer (5′ TCG CTC CCC CAR RSA CCT AGT 3′) as a CARD-FISH-probe Gemma_801. As the probe-binding site was located within a hairpin loop and the accessibility of the binding site was low (92), we designed two 18-bp-long helpers to open the loop and facilitate hybridization (93): Gemma_801 H1 (5′-GCG CCG GCA YYC GAG GGG-3′) and Gemma_801 H2 (5′-GGG GDA CTT AAT GCG TTA-3′). The specificity of the probe was tested *in silico* using the ARB Probe_Match function and the SILVA online ProbeCheck tool against the ENA database (EMBL-EBI). The probe matches only six untargeted sequences and six more if one weighted mismatch is allowed (<0.0021% of all bacterial sequences), and all of them originate from soil or the rhizosphere. The coverage of the probe for class *Gemmatimonadetes* is 35%, so the usage of this probe was primarily justified with the intent to visualize the *Gemmatimonadetes* cells in the natural environment, and not to use it for absolute quantification. Hybridization conditions were optimized at 35°C and 46°C by a stepwise increase of formamide concentrations in the hybridization buffer (20% to 70% by 5%) using pure cultures of *G. phototrophica* AP64 and photoheterotrophic strain TET16 (36) as positive controls, and *Limnohabitans planktonicus* (*Gammaproteobacteria/Betaproteobacteriales*) and *Sphingomonas* sp. strain AAP5 (*Alphaproteobacteria*) as negative controls. The final protocol is described below.

Sampling for catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) analysis. We applied the newly designed probe on a sample set collected between 14 May and 24 August 2015 at Římov Reservoir. Six permanent stations (river, stations 2, 3, 4, and 7, and dam) along the longitudinal transect of the reservoir (94) (Fig. S2) were sampled with a Friedinger sampler at 3-week intervals. Samples were taken from 0.5-m depth at all stations and at additional depths at lacustrine stations: 5 m and 10 m at stations 4 and 7; and 5 m, 10 m, 20 m, 30 m, and 40 m at the dam site. Water samples were fixed on-site with formaldehyde (2% final concentration [vol/vol]) and transported to the laboratory within 2 h. For the enumeration of bacterial densities,1 to 2 ml was stained with 4′,6′-diamidino-2-phenylindole (DAPI) and filtered onto black 0.2-μm-pore-size polycarbonate membranes (25 mm diameter; SPI Supplies, USA), and bacterial abundance was determined via epifluorescence microscopy (95).

CARD-FISH analysis. Fixed formaldehyde water samples containing 1×10^6 to 2×10^6 bacteria were filtered onto polycarbonate membranes with a 0.2-μm pore size (47-mm diameter; Millipore, USA), and filters were stored at −20°C. The catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) analysis was conducted as in the previously described protocol (96) with slight modification in digestion procedure, optimized for samples from Římov Reservoir. In brief, the microorganisms were immobilized on the membranes with low-melting-point agarose and permeabilized with lysozyme (10 μg

ml^{-1}, 1 h) and achromopeptidase (60 U ml^{-1}, 20 to 25 min). Digestion was followed by neutralization of endogenous peroxidases and hybridization with the horseradish peroxidase-labeled probe Gemma_801 (biomers.net GmbH) and helpers in a hybridization buffer with 35% formamide concentration (vol/vol) at 35°C for 3 h. After the subsequent washing steps, amplification of the hybridization signal was performed with fluorescein-labeled tyramides (Invitrogen, Carlsbad, CA, USA) at 37°C for 30 min. Washed and dried filters were counterstained with DAPI (1 μg ml^{-1}) and analyzed with epifluorescence microscopy (Olympus BX-53F) using UNWU, U-WB, and U-WG optical filter sets. Proportions of CARD-FISH stained bacteria were determined by inspecting more than 1000 DAPI-stained cells per sample.

Data availability. Sequence data for all metagenomes generated in this work are archived at EBI European Nucleotide Archive and can be accessed under the BioProject accession number PRJEB35770. All metagenome-assembled genomes are also available at figshare (https://doi.org/10.6084/m9.figshare.12662132.v2).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.3 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 0.04 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.1 MB.
TABLE S1, XLSX file, 0.04 MB.
TABLE S2, XLSX file, 0.01 MB.
TABLE S3, XLSX file, 0.01 MB.
TABLE S4, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

This research was supported by the *PhotoGemm+* project (grant 19-28778X) from the Czech Science Foundation. R.G. was supported by the research grant 20-12496X (Grant Agency of the Czech Republic). A.-Ş.A. was supported by the research grants 17-04828S (Grant Agency of the Czech Republic) and MSM200961801 (Academy of Sciences of the Czech Republic). T.S. was supported by the research grant 20-23718Y (Grant Agency of the Czech Republic). M.M.S. was supported by the research grant 19-23469S (Grant Agency of the Czech Republic).

We are thankful to Yonghui Zeng and Thanassis Zervas (Aarhus University) for help in initial stages of this study. Petr Znachor and Pavel Rychtecký are acknowledged for providing samples from Římov Reservoir, and Petr Porcal for providing samples from Jiřická pond. We are also thankful to Thomas Posch and Eugen Loher (University of Zurich) for help with sampling of Lake Zurich, Laboratory for Water and Soil Protection of the Canton of Bern (GBL), Switzerland, for help with sampling of Lake Thun and the crew of the research vessel "Kormoran" (LUBW-ISF Langenargen) for help with sampling of Lake Constance. We are thankful to Alastair T. Gardiner for language corrections.

REFERENCES

1. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63. https://doi.org/10.1038/345060a0.
2. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.
3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.
4. Whitman WB, Oren A, Chuvochina M, da Costa MS, Garrity GM, Rainey FA, Rossello-Mora R, Schink B, Sutcliffe I, Trujillo ME, Ventura S. 2018. Proposal of the suffix –ota to denote phyla. Addendum to 'Proposal to include the rank of phylum in the International Code of Nomenclature of Prokaryotes.' Int J Syst Evol Microbiol 68:967–969. https://doi.org/10.1099/ijsem.0.002593.
5. Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL. 2001. Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria, with no known pure-culture representatives. Appl Environ Microbiol 67:411–419. https://doi.org/10.1128/AEM.67.1.411-419.2001.
6. Mummey DL, Stahl PD. 2003. Candidate division BD: phylogeny, distribution and abundance in soil ecosystems. Syst Appl Microbiol 26:228–235. https://doi.org/10.1078/072320203322346074.
7. Madrid VM, Aller JY, Aller RC, Chistoserdov AY. 2001. High prokaryote diversity and analysis of community structure in mobile mud deposits off French Guiana: identification of two new bacterial candidate divisions. FEMS Microbiol Ecol 37:197–209. https://doi.org/10.1111/j.1574-6941.2001.tb00867.x.
8. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y, Nakamura K. 2003. Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. Int J Syst Evol Microbiol 53:1155–1163. https://doi.org/10.1099/ijs.0.02520-0.

9. DeBruyn JM, Fawaz MN, Peacock AD, Dunlap JR, Nixon LT, Cooper KE, Radosevich M. 2013. Gemmatirosa kalamazoonesis gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum Gemmatimonadetes. J Gen Appl Microbiol 59:305–312. https://doi.org/10.2323/jgam.59.305.
10. Pascual J, García-López M, Bills GF, Genilloud O. 2016. Longimicrobium terrae gen. nov., sp. nov., an oligotrophic bacterium of the under-represented phylum Gemmatimonadetes isolated through a system of miniaturized diffusion chambers. Int J Syst Evol Microbiol 66:1976–1985. https://doi.org/10.1099/ijsem.0.000974.
11. Pascual J, Foesel BU, Geppert A, Huber KJ, Boedeker C, Luckner M, Wanner G, Overmann J. 2018. Roseisolibacter agri gen. nov., sp. nov., a novel slow-growing member of the under-represented phylum Gemmatimonadetes. Int J Syst Evol Microbiol 68:1028–1036. https://doi.org/10.1099/ijsem.0.002619.
12. Hanada S, Sekiguchi Y. 2014. The phylum Gemmatimonadetes, p 677–681. *In* Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The prokaryotes, 4th ed, vol 11. Springer, Berlin, Germany.
13. Park D, Kim H, Yoon S. 2017. Nitrous oxide reduction by an obligate aerobic bacterium, Gemmatimonas aurantiaca strain T-27. Appl Environ Microbiol 83:e00502-17. https://doi.org/10.1128/AEM.00502-17.
14. Zeng Y, Feng F, Medová H, Dean J, Koblížek M. 2014. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. Proc Natl Acad Sci U S A 111:7795–7800. https://doi.org/10.1073/pnas.1400295111.
15. Dachev M, Bína D, Sobotka R, Moravcová L, Gardian Z, Kaftan D, Šlouf V, Fuciman M, Polívka T, Koblížek M. 2017. Unique double concentric ring organization of light harvesting complexes in Gemmatimonas phototrophica. PLoS Biol 15:e2003943. https://doi.org/10.1371/journal.pbio.2003943.
16. Ward LM, Cardona T, Holland-Moritz H. 2019. Evolutionary implications of anoxygenic phototrophy in the bacterial phylum Candidatus Eremiobacterota (WPS-2). Front Microbiol 10:1658. https://doi.org/10.3389/fmicb.2019.01658.
17. Cardona T. 2019. Thinking twice about the evolution of photosynthesis. Open Biol 9:180246. https://doi.org/10.1098/rsob.180246.
18. Zeng Y, Selyanin V, Lukeš M, Dean J, Kaftan D, Feng F, Koblížek M. 2015. Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium Gemmatimonas phototrophica sp. nov., and emended descriptions of the genus Gemmatimonas and Gemmatimonas aurantiaca. Int J Syst Evol Microbiol 65:2410–2419. https://doi.org/10.1099/ijs.0.000272.
19. Cardona T. 2016. Origin of bacteriochlorophyll a and the early diversification of photosynthesis. PLoS One 11:e0151250. https://doi.org/10.1371/journal.pone.0151250.
20. DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. 2011. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. Appl Environ Microbiol 77:6295–6300. https://doi.org/10.1128/AEM.05005-11.
21. Zeng Y, Baumbach J, Barbosa EGV, Azevedo V, Zhang C, Koblížek M. 2016. Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. Environ Microbiol Rep 8:139–149. https://doi.org/10.1111/1758-2229.12363.
22. Huang Y, Zeng Y, Lu H, Feng H, Zeng Y, Koblížek M. 2016. Novel acsF gene primers revealed a diverse phototrophic bacterial population, including Gemmatimonadetes, in Lake Taihu (China). Appl Environ Microbiol 82:5587–5594. https://doi.org/10.1128/AEM.01063-16.
23. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325. https://doi.org/10.1126/science.aap9516.
24. Cary SC, McDonald IR, Barrett JE, Cowan DA. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nat Rev Microbiol 8:129–138. https://doi.org/10.1038/nrmicro2281.
25. Bakermans C, Skidmore ML, Douglas S, McKay CP. 2014. Molecular characterization of bacteria from permafrost of the Taylor Valley, Antarctica. FEMS Microbiol Ecol 89:331–346. https://doi.org/10.1111/1574-6941.12310.
26. Ren C, Chen J, Lu X, Doughty R, Zhao F, Zhong Z, Han X, Yang G, Feng Y, Ren G. 2018. Responses of soil total microbial biomass and community compositions to rainfall reductions. Soil Biol Biochem 116:4–10. https://doi.org/10.1016/j.soilbio.2017.09.028.
27. Fecskeová LK, Piwosz K, Hanusová M, Nedoma J, Znachor P, Koblížek M. 2019. Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. Sci Rep 9:18766. https://doi.org/10.1038/s41598-019-55210-x.
28. Song H, Li Z, Du B, Wang G, Ding Y. 2012. Bacterial communities in sediments of the shallow Lake Dongping in China. J Appl Microbiol 112:79–89. https://doi.org/10.1111/j.1365-2672.2011.05187.x.
29. Zhang J, Yang Y, Zhao L, Li Y, Xie S, Liu Y. 2015. Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. Appl Microbiol Biotechnol 99:3291–3302. https://doi.org/10.1007/s00253-014-6262-x.
30. Gugliandolo C, Michaud L, Lo Giudice A, Lentini V, Rochera C, Camacho A, Maugeri TL. 2016. Prokaryotic community in lacustrine sediments of Byers Peninsula (Livingston Island, Maritime Antarctica). Microb Ecol 71:387–400. https://doi.org/10.1007/s00248-015-0666-8.
31. Gołębiewski M, Całkiewicz J, Creer S, Piwosz K. 2017. Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria: the case study of the Vistula river estuary. Environ Microbiol Rep 9:129–143. https://doi.org/10.1111/1758-2229.12509.
32. Traving SJ, Rowe O, Jakobsen NM, Sørensen H, Dinasquet J, Stedmon CA, Andersson A, Riemann L. 2017. The effect of increased loads of dissolved organic matter on estuarine microbial community composition and function. Front Microbiol 8:351. https://doi.org/10.3389/fmicb.2017.00351.
33. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542. https://doi.org/10.1038/s41564-017-0012-7.
34. Holland-Moritz H, Stuart J, Lewis LR, Miller S, Mack MC, McDaniel SF, Fierer N. 2018. Novel bacterial lineages associated with boreal moss species. Environ Microbiol 20:2625–2638. https://doi.org/10.1111/1462-2920.14288.
35. Cabello-Yeves PJ, Zemskaya TI, Rosselli R, Coutinho FH, Zakharenko AS, Blinov VV, Rodriguez-Valera F. 2017. Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. Appl Environ Microbiol 84:e02132-17. https://doi.org/10.1128/AEM.02132-17.
36. Zeng Y, Nupur, Wu N, Madsen AM, Chen X, Gardiner AT, Koblížek M. 2021. Gemmatimonas groenlandica sp. nov. is an aerobic anoxygenic phototroph in the phylum Gemmatimonadetes. Front Microbiol 11:606612. https://doi.org/10.3389/fmicb.2020.606612.
37. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
38. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. ISME J 11:2399–2406. https://doi.org/10.1038/ismej.2017.113.
39. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
40. Koblížek M. 2015. Ecology of aerobic anoxygenic phototrophs in aquatic environments. FEMS Microbiol Rev 39:854–870. https://doi.org/10.1093/femsre/fuv032.
41. Kavagutti VS, Andrei AŞ, Mehrshad M, Salcher MM, Ghai R. 2019. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome 7:135. https://doi.org/10.1186/s40168-019-0752-0.
42. Ameryk A, Kownacka J, Zalewski M, Piwosz K. 2021. Typical freshwater and marine bacterial lineages dynamics at salinity between 0 and 4 in the Vistula Lagoon. Estuar Coast Shelf Sci 250:107100. https://doi.org/10.1016/j.ecss.2020.107100.
43. Morrison JM, Baker KD, Zamor RM, Nikolai S, Elshahed MS, Youssef NH. 2017. Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). PLoS One 12:e0177488. https://doi.org/10.1371/journal.pone.0177488.
44. Cai H, Jiang H, Krumholz LR, Yang Z. 2014. Bacterial community composition of size-fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater lake. PLoS One 9:e102879. https://doi.org/10.1371/journal.pone.0102879.
45. Shia L, Cai Y, Wang X, Li P, Yu Y, Kong F. 2010. Community structure of bacteria associated with Microcystis colonies from cyanobacterial blooms. J Freshw Ecol 25:193–203. https://doi.org/10.1080/02705060.2010.9665068.
46. Fauteux L, Cottrell MT, Kirchman DL, Borrego CM, Garcia-Chaves MC, Del Giorgio PA. 2015. Patterns in abundance, cell size and pigment content of aerobic anoxygenic phototrophic bacteria along environmental gradients in northern lakes. PLoS One 10:e0124035. https://doi.org/10.1371/journal.pone.0124035.
47. Hauruseu D, Koblízek M. 2012. Influence of light on carbon utilization in aerobic anoxygenic phototrophs. Appl Environ Microbiol 78:7414–7419. https://doi.org/10.1128/AEM.01747-12.

48. Xing P, Kong F, Cao H, Zhang M, Tan X. 2007. Variations of bacterioplankton community composition during Microcystis spp. blooms in a shallow eutrophic lake. J Freshw Ecol 22:61–67. https://doi.org/10.1080/02705060.2007.9664146.
49. Šimek K, Hornák K, Jezbera J, Mašín M, Nedoma J, Gasol JM, Schauer M. 2005. Influence of top-down and bottom-up manipulations on the R-BT065 subcluster of β-proteobacteria, an abundant group in bacterioplankton of a freshwater reservoir. Appl Environ Microbiol 71:2381–2390. https://doi.org/10.1128/AEM.71.5.2381-2390.2005.
50. Salcher MM, Pernthaler J, Posch T. 2011. Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria that rule the waves (LD12). ISME J 5:1242–1252. https://doi.org/10.1038/ismej.2011.8.
51. Simek K, Hornák K, Jezbera J, Nedoma J, Znachor P, Hejzlar J, Sed'a J. 2008. Spatio-temporal patterns of bacterioplankton production and community composition related to phytoplankton composition and protistan bacterivory in a dam reservoir. Aquat Microb Ecol 51:249–262. https://doi.org/10.3354/ame01193.
52. Yurkov VV, Beatty JT. 1998. Aerobic anoxygenic phototrophic bacteria. Microbiol Mol Biol Rev 62:695–724. https://doi.org/10.1128/MMBR.62.3.695-724.1998.
53. Goecke F, Thiel V, Wiese J, Labes A, Imhoff JF. 2013. Algae as an important environment for bacteria – phylogenetic relationships among new bacterial species isolated from algae. Phycologia 52:14–24. https://doi.org/10.2216/12-24.1.
54. Piwosz K, Vrdoljak A, Frenken T, González-Olalla JM, Šantić D, McKay RM, Spilling K, Guttman L, Znachor P, Mujakić I, Fecskeová LK, Zoccarato L, Hanusová M, Pessina A, Reich T, Grossart H-P, Koblížek M. 2020. Light and primary production shape bacterial activity and community composition of aerobic anoxygenic phototrophic bacteria in a microcosm experiment. mSphere 5:e00673-20. https://doi.org/10.1128/mSphere.00673-20.
55. Zheng Q, Zhang R, Koblížek M, Boldareva EN, Yurkov V, Yan S, Jiao N. 2011. Diverse arrangement of photosynthetic gene clusters in aerobic anoxygenic phototrophic bacteria. PLoS One 6:e25050. https://doi.org/10.1371/journal.pone.0025050.
56. Hendrickson EL, Leigh JA. 2008. Roles of coenzyme F420-reducing hydrogenases and hydrogen- and F420-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F420 and production of hydrogen during methanogenesis. J Bacteriol 190:4818–4821. https://doi.org/10.1128/JB.00255-08.
57. Vavourakis CD, Mehrshad M, Balkema C, Van Hall R, Andrei AŞ, Ghai R, Sorokin DY, Muyzer G. 2019. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. BMC Biol 17:69. https://doi.org/10.1186/s12915-019-0688-7.
58. Zorz JK, Sharp C, Kleiner M, Gordon PMK, Pon RT, Dong X, Strous M. 2019. A shared core microbiome in soda lakes separated by large distances. Nat Commun 10:4230. https://doi.org/10.1038/s41467-019-12195-5.
59. Hanson TE, Tabita FR. 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci U S A 98:4397–4402. https://doi.org/10.1073/pnas.081610398.
60. Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. Philos Trans R Soc Lond B Biol Sci 363:2629–2640. https://doi.org/10.1098/rstb.2008.0023.
61. Ashida H, Danchin A, Yokota A. 2005. Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? Res Microbiol 156:611–618. https://doi.org/10.1016/j.resmic.2005.01.014.
62. Ashida H, Saito Y, Kojima C, Kobayashi K, Ogasawara N, Yokota A. 2003. A functional link between RuBisCO-like protein of Bacillus and photosynthetic RuBisCO. Science 302:286–290. https://doi.org/10.1126/science.1086997.
63. Yurkov V, Hughes E. 2017. Aerobic anoxygenic phototrophs: four decades of mystery, p 193–214. *In* Modern topics in the phototrophic prokaryotes: environmental and applied aspects. Springer International Publishing, Cham, Switzerland.
64. Piwosz K, Kaftan D, Dean J, Šetlík J, Koblížek M. 2018. Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. Environ Microbiol 20:724–733. https://doi.org/10.1111/1462-2920.14003.
65. Znachor P, Hejzlar J, Vrba J, Nedoma J, Seda J, Simek K, Komarkova J, Kopacek J, Sorf M, Kubecka J, Matena J, Riha M, Peterka J, Cech M, Vasek M. 2016. Brief history of long-term ecological research into aquatic ecosystems and their catchments in the Czech Republic. Part I: manmade reservoirs. Biology Centre CAS, v.v.i., Institute of Hydrobiology, České Budějovice, Czechia.
66. Bushnell B. 2015. BBMap (version 35.14). https://sourceforge.net/projects/bbmap/.
67. Bushnell B, Rood J, Singer E. 2017. BBMerge – accurate paired shotgun read merging via overlap. PLoS One 12:e0185056. https://doi.org/10.1371/journal.pone.0185056.
68. Bushnell B. 2016. Reformat. https://github.com/BioInfoTools/BBMap/blob/master/sh/reformat.sh.
69. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
70. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196. https://doi.org/10.1093/nar/gkm864.
71. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35:1026–1028. https://doi.org/10.1038/nbt.3988.
72. Nawrocki EP. 2009. Structural RNA homology search and alignment using covariance models. PhD thesis. Washington University in St. Louis, St. Louis, MO.
73. Anderson MJ, Legendre P. 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. J Stat Comput Simul 62:271–303. https://doi.org/10.1080/00949659908811936.
74. Legendre P, Anderson MJ. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monogr 69:1–24. https://doi.org/10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2.
75. Bushnell B. 2016. BBMap short-read aligner, and other bioinformatic tools. https://sourceforge.net/projects/bbmap.
76. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
77. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
78. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373. https://doi.org/10.1093/nar/gkg128.
79. Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. Methods Mol Biol 1079:155–170. https://doi.org/10.1007/978-1-62703-646-7_10.
80. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210.
81. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
82. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.
83. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102:2567–2572. https://doi.org/10.1073/pnas.0409727102.
84. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404.
85. Mirarab S, Nguyen N, Guo S, Wang LS, Kim J, Warnow T. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol 22:377–386. https://doi.org/10.1089/cmb.2014.0156.
86. Bulzu PA, Andrei AŞ, Salcher MM, Mehrshad M, Inoue K, Kandori H, Beja O, Ghai R, Banciu HL. 2019. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. Nat Microbiol 4:1129–1137. https://doi.org/10.1038/s41564-019-0404-y.

87. Whelan S, Irisarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. Bioinformatics 34:3929–3930. https://doi.org/10.1093/bioinformatics/bty448.
88. Weese D, Holtgrewe M, Reinert K. 2012. RazerS 3: faster, fully sensitive read mapping. Bioinformatics 28:2592–2599. https://doi.org/10.1093/bioinformatics/bts505.
89. Huang Y, Gilna P, Li W. 2009. Identification of ribosomal RNA genes in metagenomic fragments. Bioinformatics 25:1338–1340. https://doi.org/10.1093/bioinformatics/btp161.
90. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
91. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17:10. https://doi.org/10.14806/ej.17.1.200.
92. Behrens S, Rühland C, Inácio J, Huber H, Fonseca A, Spencer-Martins I, Fuchs BM, Amann R. 2003. In situ accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. Appl Environ Microbiol 69:1748–1758. https://doi.org/10.1128/aem.69.3.1748-1758.2003.
93. Fuchs BM, Glockner FO, Wulf J, Amann R. 2000. Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes. Appl Environ Microbiol 66:3603–3607. https://doi.org/10.1128/aem.66.8.3603-3607.2000.
94. Rychtecký P, Znachor P. 2011. Spatial heterogeneity and seasonal succession of phytoplankton along the longitudinal gradient in a eutrophic reservoir. Hydrobiologia 663:175–186. https://doi.org/10.1007/s10750-010-0571-6.
95. Porter KG, Feig YS. 1980. The use of DAPI for identifying and counting aquatic microflora. Limnol Oceanogr 25:943–948. https://doi.org/10.4319/lo.1980.25.5.0943.
96. Sekar R, Pernthaler A, Pernthaler J, Warnecke F, Posch T, Amann R. 2003. An improved protocol for quantification of freshwater Actinobacteria by fluorescence in situ hybridization. Appl Environ Microbiol 69:2928–2935. https://doi.org/10.1128/aem.69.5.2928-2935.2003.

Paper II

Phylum Gemmatimonadota and its role in the environment

Mujakić, I., Piwosz, K., Koblížek, M. (2022)
Microorganisms, 10, 151

3.2. Summary results derived from Paper No. 2

Since the majority of information about Gemmatimonadota came from studies that focused on the whole microbial community, and thus was scattered in the literature, we decided to collect and summarize all available information in a short review.

The beginning describes the physiology and metabolism of six known cultures of Gemmatimonadota. The first isolate, strain *Gemmatimonas aurantiaca* T27^{T}, which gave the name to the phylum, originated from a wastewater treatment plant. Three more cultures, *Gemmatirosa kalamazoonesis*, *Longimicrobium terrae* and *Roseisolibacter agrii* came from different types of soil. These four species are chemoorganoheterotrophic, while two freshwater isolates, *Gemmatimonas phototrophica* and *Gemmatimonas groenlandica* are photoheterotrophs. All cultures can grow under aerobic or semi-aerobic conditions and are characterized by pigmentation varying between orange, pink or red caused by a variety of carotenoids they contain.

Gemmatimonadota is also one of seven bacterial phyla capable of anoxygenic photosynthesis. Both freshwater cultures are facultative photoheterotrophs that need a supply of organic carbon but can also harvest light energy using BChl-containing reaction centers. Their photosynthetic genes are organized in the PGC which, based on its close phylogenetic relationship and organization, was probably obtained through HGT from Proteobacteria.

We collected information about amplicon and metagenomics studies showing the wide distribution of Gemmatimonadota in various environments and their high uncultured diversity. Most of the studies focused on varying soil environments, where they are one of the top ten most abundant phyla. They are even found in the soils of Antarctica, which was once considered a sterile bacterial environment. In aquatic habitats they are found in freshwater lakes and sediments, estuaries and brackish waters, rivers and marine ecosystems, where they are mostly associated with deep hydrothermal vents, sediments or marine sponges. Information about other environments is limited, but they are also reported in wastewaters and as a part of biofilms and microbial mats. In addition, information about the diversity and distribution of photoheterotrophic Gemmatimonadota, specifically in freshwater environments, is reported in this review.

microorganisms

MDPI

Review

Phylum Gemmatimonadota and Its Role in the Environment

Izabela Mujakić [1,2], Kasia Piwosz [1,3] and Michal Koblížek [1,2,*]

[1] Centre Algatech, Institute of Microbiology, Czech Academy of Sciences, Novohradská 237, 379 81 Třeboň, Czech Republic; mujakic@alga.cz (I.M.); kpiwosz@mir.gdynia.pl (K.P.)
[2] Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 České Budějovice, Czech Republic
[3] National Marine Fisheries Research Institute, Kołłątaja 1, 81-332 Gdynia, Poland
* Correspondence: koblizek@alga.cz

Abstract: Bacteria are an important part of every ecosystem that they inhabit on Earth. Environmental microbiologists usually focus on a few dominant bacterial groups, neglecting less abundant ones, which collectively make up most of the microbial diversity. One of such less-studied phyla is Gemmatimonadota. Currently, the phylum contains only six cultured species. However, data from culture-independent studies indicate that members of Gemmatimonadota are common in diverse habitats. They are abundant in soils, where they seem to be frequently associated with plants and the rhizosphere. Moreover, Gemmatimonadota were found in aquatic environments, such as freshwaters, wastewater treatment plants, biofilms, and sediments. An important discovery was the identification of purple bacterial reaction centers and anoxygenic photosynthesis in this phylum, genes for which were likely acquired via horizontal gene transfer. So far, the capacity for anoxygenic photosynthesis has been described for two cultured species: *Gemmatimonas phototrophica* and *Gemmatimonas groenlandica*. Moreover, analyses of metagenome-assembled genomes indicate that it is also common in uncultured lineages of Gemmatimonadota. This review summarizes the current knowledge about this understudied bacterial phylum with an emphasis on its environmental distribution.

Keywords: Gemmatimonadota; *Gemmatimonadetes*; anoxygenic photosynthesis; photosynthetic gene cluster; MAGs

check for updates

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

Academic Editor: Remus T. Dame

Received: 1 December 2021
Accepted: 7 January 2022
Published: 12 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bacteria are an important component of all ecosystems, playing key roles in microbial food webs and the biogeochemical cycles. Initially, knowledge about microorganisms originated from work on species that could be cultivated and characterized based on their morphology, cell structure, chemical composition, and metabolic activities. This has changed since progress in molecular methods has enabled the study of microorganisms in their natural environment without the need for cultivation [1,2]. Gene surveys and analyses of 16S rRNA genes from environmental samples allow detection and identification of in situ microbial diversity, and due to shotgun metagenomic sequencing, our knowledge on microbial diversity is continuously growing [3].

One of the bacterial phyla established through molecular phylogenetic methods is Gemmatimonadota. It was first identified based on five environmental 16S rRNA gene sequences from deep-sea sediments, soils, and enhanced biological phosphorus removal (EBPR) reactor sludge, and it was named the BD-group [4–6]. Separately, based on the three 16S rRNA gene sequences from coastal sediment, it was called a candidate division KS-B [7]. It took only two years until a strain T27 belonging to the BD/KS-B group was isolated from a wastewater treatment plant in Japan. This organism was characterized and named *Gemmatimonas aurantiaca*, and the BD/KS-B group became a new phylum, Gemmatimonadota (initially named *Gemmatimonadetes* [8]). Phylogenetically, Gemmatimonadota are related with Fibrobacterota [9] and Chlorobi (Bacteroidota) [2,10,11]. Additionally, whole-genome

comparison of all available bacterial phyla in the Genome Taxonomy Database (GTDB), visualized by AnnoTree [12], relates them with several newly proposed phyla, such as '*Candidatus* Eisenbacteria', '*Candidatus* Krumholzibacteriota', and '*Candidatus* Edwardsbacteria' [13,14]. At present, there are six cultured species in the phylum, two of which are capable of anoxygenic photosynthesis [15,16]. Apart from Gemmatimonadota, anoxygenic photosynthesis is present in six other bacterial phyla and is scattered throughout the tree of life (Figure 1).

In contrast to the low number of described species, culture-independent methods indicate that members of Gemmatimonadota are ubiquitous and especially common in soils, limnic environments, and sediments [17]. Metagenomic analyses documented unexpected phylogenomic and metabolic diversity among uncultured Gemmatimonadota [18]. Moreover, several lineages of this phylum are capable of anoxygenic photosynthesis [15,16,19]. These results indicate that members of Gemmatimonadota likely play a specific role in the environment.

In this review, we summarize the current knowledge of culture and metagenomic studies on the ecology of Gemmatimonadota with a focus on their diversity and distribution in the environment.

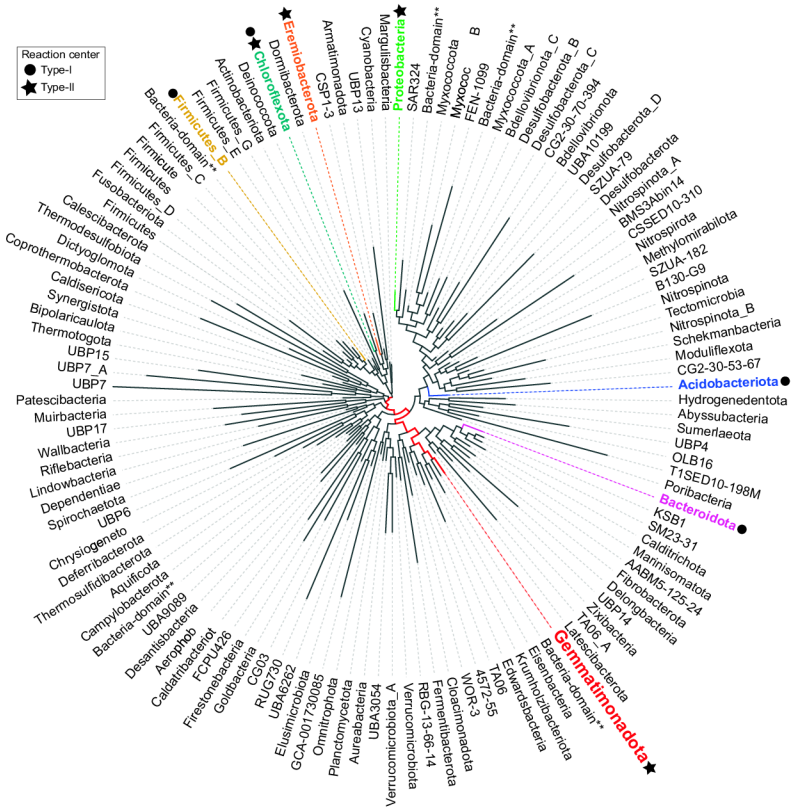

Figure 1. Phylogenomic tree of all bacterial phyla present in the Genome Taxonomy Database (GTDB) (Release RS95) [13,14] visualized using AnnoTree (version 1.2) [12], showing the position of Gemmatimonadota (marked in red). Additionally, along Gemmatimonadota, six other phyla that contain members capable of anoxygenic photosynthesis are marked in different colors, and the type of reaction center is indicated with a circle or a star (left legend). The double asterisk (**) means there is no corresponding taxonomy in GTDB for the genome so higher taxonomy level is used. The tree was edited in Inkscape (version 1.0).

2. Cultured Species

Gemmatimonas aurantiaca strain T-27^T was the first cultured Gemmatimonadota, which now represents the type species of class *Gemmatimonadetes*, genus *Gemmatimonas*. This aerobic chemoheterotrophic organism was isolated from an anaerobic–aerobic sequential batch reactor operated under enhanced EBPR conditions for wastewater treatment [8]. The other two members of the genus *Gemmatimonas* were isolated from freshwater habitats. A unique feature of these two organisms is the presence of bacteriochlorophyll-containing photosynthetic reaction centers [9,16]. Strain AP64^T, isolated from the shallow desert lake Tiān é hú (Swan Lake) in North China, shares 96.1% sequence identity of the 16S rRNA gene with the type species *G. aurantiaca* T-27^T. It is a facultative photoheterotroph, which requires organic substrates to grow, and the ability to harvest light provides an additional source of energy for its metabolism [20]. Based on its phototrophic lifestyle, strain AP64^T was named *Gemmatimonas phototrophica* sp. nov. [15]. Recently, the second photoheterotrophic strain TET16 was isolated from a stream in Northeast Greenland [16]. This strain represents a new species, as it has 95.7% sequence identity with the 16S rRNA gene of *G. phototrophica* and 95.9% identity with *G. aurantiaca*, and it was named *Gemmatimonas groenlandica* sp. nov. [16].

In contrast to the aquatic *Gemmatimonas* species, other cultured Gemmatimonadota originate from soils. A heterotrophic strain KBS708 was isolated from organically managed agricultural soil in Michigan USA. The strain with the tentative name '*Gemmatirosa kalamazoonensis*' gen. nov., sp. nov. shares only 89% 16S rRNA gene sequence identity with *G. aurantiaca* [21]. However, this organism still awaits its valid description. Another soil species belonging to class *Gemmatimonadetes* was *Roseisolibacter agri* gen. nov., sp. nov. strain AW1220^T, isolated from agricultural floodplain soil from Namibia [22]. Finally, *Longimicrobium terrae* strain CB-286315^T was isolated from a Mediterranean forest soil sample in Granada, Spain [23]. Since the new organism was relatively distant (16S rRNA identity 83–84%) from members of class *Gemmatimonadetes*, it established not only a new genus *Longimicrobium*, but also a novel class named *Longimicrobia* [23].

2.1. Physiology and Metabolism of Cultured Gemmatimonadota

All cultured Gemmatimonadota are chemoorganoheterotrophs, except for the facultative photoheterotrophic *G. phototrophica* and *G. groenlandica* [8,15,16,21–23]. They grow under aerobic or semiaerobic conditions, with temperature optima varying between 20 and 37 °C [16,22]. All species prefer low-salinity media—*G. aurantiaca* has the highest salinity tolerance of up to 0.8% NaCl [15]—as well as neutral or slightly alkaline pH. The only exception is '*Gemmatirosa kalamazoonensis*', with a pH optimum between 5.5 and 6.5 [21]. Interestingly, *G. aurantiaca* possesses a metabolic potential for reduction of N_2O, one of the strongest greenhouse gases [24,25], implying the possibly important ecological role of Gemmatimonadota in the environment.

Gemmatimonadota cells are typically short rods, but occasionally, they can form over 10-μm-long filaments (not reported for *G. aurantiaca*). They divide by binary fission [8,15,16,21,22] and, except for the *L. terrae*, show budding morphology [23]. A typical trait of all cultured Gemmatimonadota is an intense pigmentation varying between orange and pink or even a reddish color in the stationary phase (Figure 2a,b). The predominant respiratory quinones are menaquinone-9 [8,21,22] or menaquinone-8 [15,16,23]. They are also naturally resistant to some antibiotics: *G. aurantiaca*, *G. phototrophica*, and '*Gemmatirosa kalamazoonensis*' grow in the presence of ampicillin or penicillin [15,21], while *G. groenlandica* is resistant to bacitracin and chloramphenicol [16].

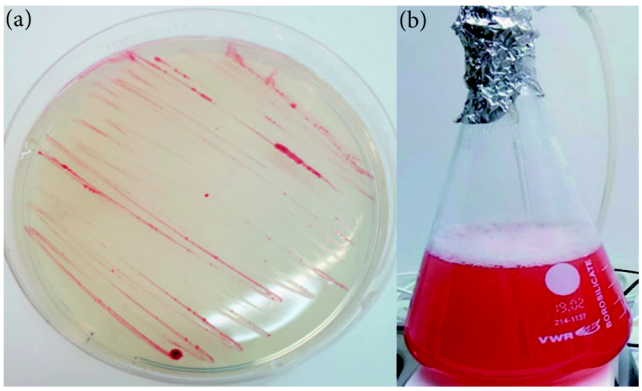

Figure 2. (**a**) Plates showing pure culture of *Gemmatimonas phototrophica* strain AP64 (**b**) and liquid culture of *Gemmatimonas groenlandica* strain TET16 (picture of TET16 is adapted from ref. [16]).

2.2. Anoxygenic Phototrophy in Gemmatimonadota

Gemmatimonadota, together with Firmicutes, Bacteroidota, Acidobacteriota, Proteobacteria, Chloroflexota, and *Candidatus* 'Eremiobacteriota', is one of few bacterial phyla capable of anoxygenic phototrophy [9–11] (Figure 1). Anoxygenic phototrophs can be distinguished based on the type of reaction center they have—either type I or type II, which differ with respect to the electron acceptors [26]. Both *G. phototrophica* and *G. groenlandica* contain type-II photosynthetic reaction centers [9,15,16]. They are facultative photoheterotrophs: They do not assimilate inorganic carbon and require a supply of organic substrates [9]. The ability of some members of Gemmatimonadota to harvest light provides them with additional energy, improves the efficiency of carbon utilization, and, consequently, increases their growth rate [20]. Light is used to generate ATP via photophosphorylation, which enables them to reduce respiration rate and, at the same time, increase assimilation rates of organic compounds, such as glucose and leucine (Figure 3a,b).

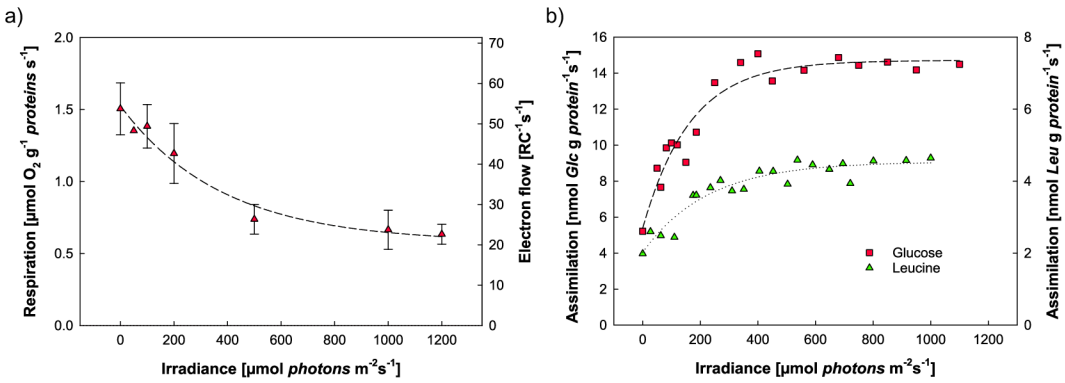

Figure 3. The figure shows (**a**) reduction of respiration in cells of *G. phototrophica* with increasing exposure to light and (**b**) increase in assimilation rates of ^{3}H-glucose and ^{3}H-leucine with exposure to light. The figure was adapted from ref. [20].

Phototrophic Gemmatimonadota organize their genes involved in bacteriochlorophyll biosynthesis, expression and assembly of the photosystem, carotenoid biosynthesis, and regulatory functions in a photosynthesis gene cluster (PGC) [9]. The PGCs in *G. phototrophica* and *G. groenlandica* share the same synteny with two big superoperons (Figure 4a) and high DNA sequence identities (70–100%) [16]. Based on this organization, which is also common

among phototrophic Proteobacteria [27], and the close phylogenetic relationship, it was suggested the source of the PGC for Gemmatimonadota was a horizontal gene transfer (HGT) event from Proteobacteria, specifically Gammaproteobacteria [9]. Intra-phylum HGT events are common in phototrophs and were reported for Cyanobacteria, Proteobacteria, and Chloroflexota [11,27–31]. The HGT of the complete PGC was documented for *Rhodobacteraceae* (Alphaproteobacteria) [32]. However, *G. phototrophica* represents the first known HGT event of a complete PGC between phototrophic and non-phototrophic representatives of different phyla [9,11,33].

The photosynthetic reaction centers of *G. phototrophica* are complexed with a unique double-concentric light-harvesting system, which manifests in two infra-red absorption bands [34] (Figure 4b). *G. phototrophica* synthesizes bacteriochlorophyll *a* constitutively. The photosynthetic apparatus was expressed even in cultures grown for one year under continuous darkness [15] and, in contrast to proteobacterial aerobic anoxygenic phototrophs [35], bacteriochlorophyll *a* is also produced under continuous light [20]. *G. phototrophica* contains over 10 different carotenoids [9], most still uncharacterized, which give the cells a specific red to pink color. Two identified carotenoids, also found in the heterotrophic member *G. aurantiaca*, are oscillol 2-rhamnoside and oscillol 2,2′-dirhamnoside [9,36]. They are believed to protect the cells from reactive oxygen species and excess light [9,37]. Recently, a unique carotenoid that is present only in the photosynthetic complex of *G. phototrophica* was described and named gemmatoxanthin [38]. Only limited information is available about the second phototrophic Gemmatimonadota *G. groenlandica*. It contains a slightly different set of carotenoids. Interestingly, despite the highly similar PGCs, *G. groenlandica* shows just one broad band in the absorption spectrum (Figure 4b), which may suggest a different organization of its light-harvesting antenna [16].

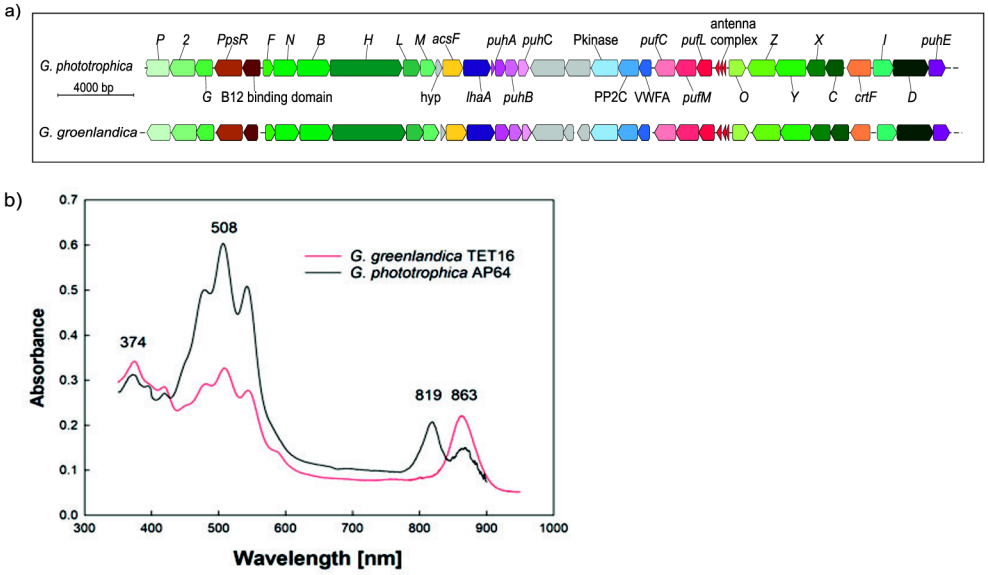

Figure 4. (**a**) Photosynthesis gene cluster of two cultured species of Gemmatimonadota, *G. phototrophica* and *G. groenlandica*. Different colors indicate genes involved in bacteriochlorophyll biosynthesis (green), carotenoid biosynthesis (orange), genes encoding the reaction center (*puf* operon-red), *puh* operon (pink/purple), other genes (brown and yellow), genes not involved in photosynthesis (blue), and hypothetical genes (gray). The figure was adapted from ref. [18]. (**b**) Absorption spectra of two cultures *G. phototrophica* (black) and *G. groenlandica* (red). In the near-infra-red range, *G. phototrophica* shows two peaks and *G. groenlandica* has only one. The figure was adapted from ref. [16].

3. Environmental Distribution

High-throughput sequencing of environmental 16S rRNA genes documented that the phylum Gemmatimonadota is cosmopolitan, with members distributed across a wide range of natural environments. They are found in soils [39–42], permafrost [43–46], rhizospheres [47–52], freshwater lakes and sediments [53–58], activated sludge [8,59,60], deep-sea sediments [61–64], marine sponge symbionts [65–68], and a brackish estuary [69].

The increasing number of environmental sequences has revealed high uncultured diversity within Gemmatimonadota. In 2014, Hanada and Sekiguchi proposed its phylogeny encompassing both cultured species and environmental sequences, dividing the phylum into five class-level lineages (Figure 5). Group 1, which corresponds to the class *Gemmatimonadetes*, was the most represented, with most sequences originating from soils and a high proportion from activated sludge in wastewater treatment plants and freshwaters. Group 2, also referred to as the BD2-11 terrestrial group, contains sequences from marine sediments and sponges in addition to those from soils [59]. Group 3 is an equivalent to class *Longimicrobia* and includes the species *Longimicrobium terrae* [23], as well as sequences from soils and other environments (oil field, gas hydrate). Group 4, also referred to as the PAUC43f marine benthic group [70], contains sequences from marine sediments and sponges [59]. Finally, Group 5, or the S0134 terrestrial group [71], contains environmental sequences with terrestrial origin, mostly from different types of soil.

A global picture of the distribution of Gemmatimonadota in various environments was obtained by a massive search in over 30,000 metagenomes that were publicly available at that time [17]. One thousand seven hundred and six metagenomes contained sequences related to Gemmatimonadota, which were most abundant in soils, with the largest proportion (0.3–1.8%) being found in agricultural soils (Figure 6). Gemmatimonadota were also abundant in wastewater treatment, biofilms, and plant-associated habitats, while smaller numbers were found in aquatic environments, such as lakes, rivers, estuaries, and springs.

3.1. Distribution in Soils

Gemmatimonadota are the eighth most abundant bacterial phylum in soils, accounting for about 1–2% of soil bacteria worldwide [41]. From the complete top-ten list (Proteobacteria, Actinobacteriota, Acidobacteriota, Planctomycetota, Chloroflexota, Verrucomicrobiota, Bacteroidota, Gemmatimonadota, Firmicutes, and Armatimonadota), Gemmatimonadota and Armatimonadota are plausibly the least-studied groups. A similar picture emerged from a study on the biogeographic distribution of Gemmatimonadota based on an analysis of sequences of 16S rRNA gene available in public databases [40]. They showed that, although they are present in sediments and other environments, the maximum number of sequences came from different types of soils, including grassland, agricultural, forest, or contaminated soils. The cosmopolitan distribution of Gemmatimonadota in various soils suggests that they are generalist species with a versatile metabolism that is able to adapt to a wide range of nutrients.

A unique environment where Gemmatimonadota have been found is in the glacier forefields of East Antarctica [76] and the McMurdo Dry Valleys, Antarctica. The soils of McMurdo Dry Valleys, originally thought to be sterile or to have low diversity of microorganisms [46], are dominated by Actinobacteriota, Acidobacteriota, Gemmatimonadota, and Bacteroidota [43,77]. Gemmatimonadota were suggested to be adapted to dry environments because they occur in high relative proportions in semiarid and arid soils and deserts [46,78–82]. Moreover, they were shown to be positively influenced by rainfall reduction and to be more abundant during drought [42]. Finally, even in soil aggregates, their relative abundance is higher in dry and semi-aerobic inner parts of microaggregates [83].

The pH is another factor that influences the abundance of Gemmatimonadota in soils, and it seems that they prefer neutral pH over acidic pH [40,84,85]. Interestingly, Gemmatimonadota dominated in alkaline [86] and highly saline soils and represented almost 17% of all bacterial reads [87].

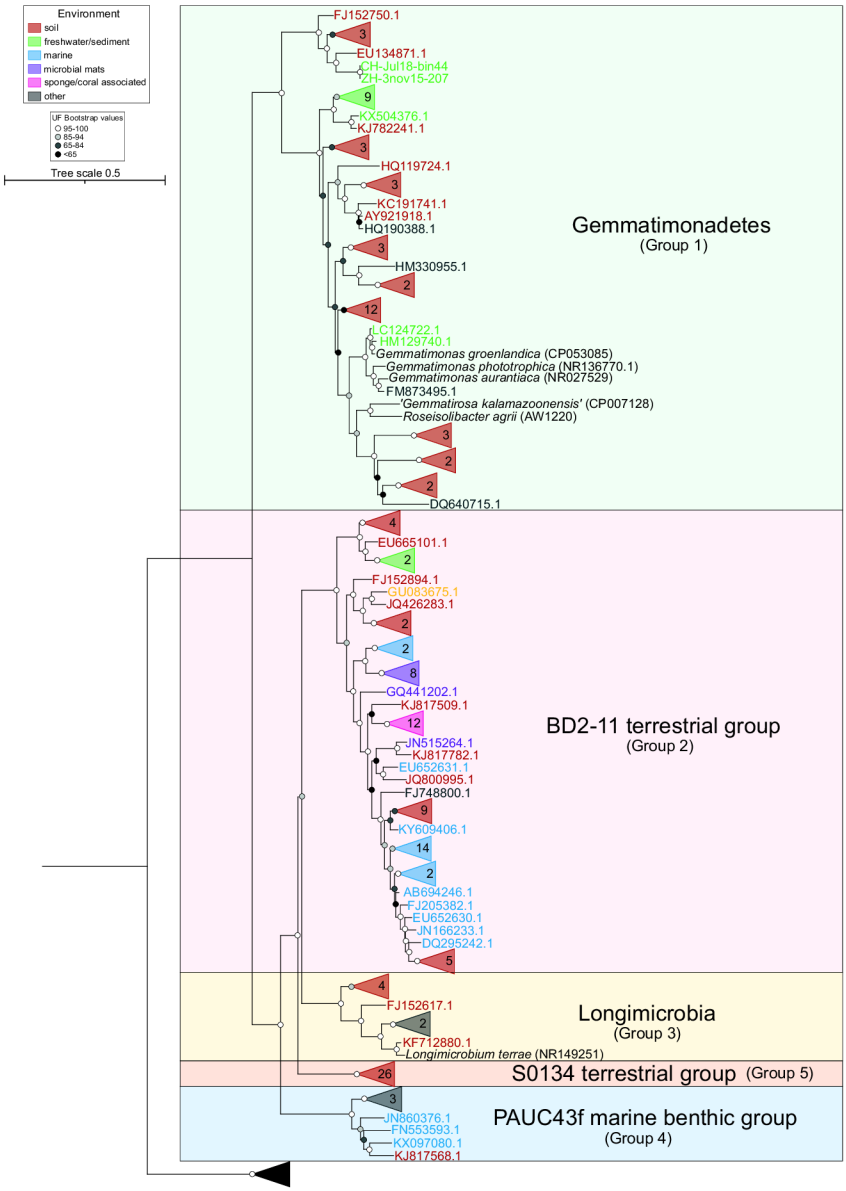

Figure 5. Maximum-likelihood phylogeny of 16S rRNA genes of Gemmatimonadota sequences recovered from the SILVA SSU v138 database (in total, 169 sequences, >1200 bp, sequence quality >94%, downloaded on 23 November 2021), 36 16S rRNA gene sequences whose accession numbers were taken from [22,59], and 12 sequences taken from [18]. All of the accession numbers are provided as Supplement Table S1. The phylogenetic tree was made with IQ-TREE [72,73], the TIM3 + F + I + G4 substitution model (chosen as the best-fitting model by ModelFinder [74]), and 1000 ultrafast bootstrap replicates. The sequences are colored based on the environment from which they originate (left legend). Numbers at collapsed branches indicate the number of sequences. All five class-level groups are indicated through vertical delimiters in the right part of the figure. The strength of support for internal nodes (assessed by ultrafast bootstrapping) is shown through gray-scale circles (left legend). Two sequences belonging to Fibrobacterota (*Fibrobacter succinogenes* and *Fibrobacter intestinalis*) were used as an outgroup. The phylogenetic tree was visualized using iTOL [75] and edited in Inkscape (version 1.0).

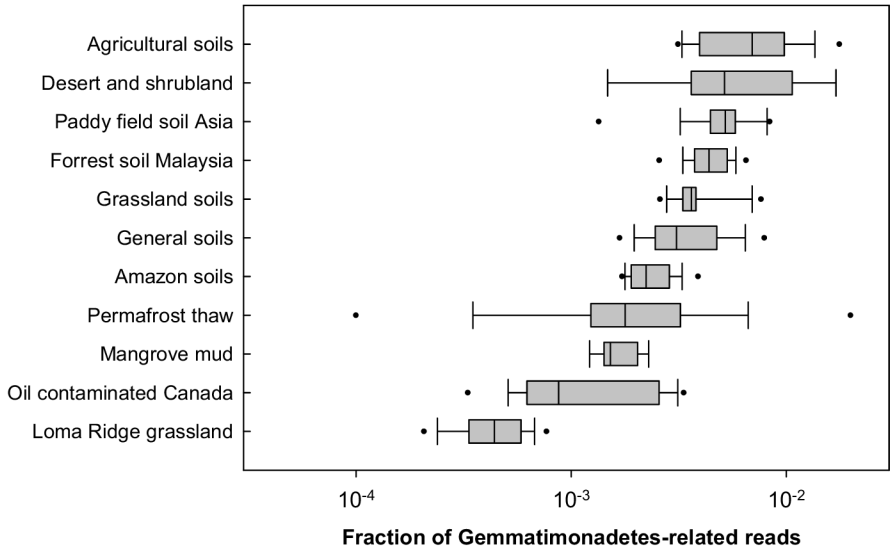

Figure 6. Gemmatimonadota-related reads recovered from different types of soils. The figure is based on metagenomic survey data published in [17].

Gemmatimonadota were also found to be one of the bacterial phyla that are positively correlated with vegetation restoration. As one of top-ten most abundant phyla that strongly increased with revegetation, their relative abundances were above 2% [88] and were positively correlated with plant richness and soil nutrients such as carbon. Moreover, they were one of seven more dominant bacterial phyla, with an abundance above 1%, which positively correlated with total carbon, nitrogen, and phosphorus in soil [89]. These studies showed the influence of high nutrient concentration on the abundance of Gemmatimonadota and their possible key role in soil ecosystems [88,89].

3.2. Distribution in Aquatic Habitats

Several reports noted the presence of Gemmatimonadota in freshwater lakes and sediments [19,53–56,90] or estuaries [69,91], and their sequences from public metagenomes were identified in lakes, rivers, estuaries, and marine ecosystems [17].

The analysis of the microbial community in the stratified, warm, monomictic, freshwater Grand Lake, OK, USA showed that Gemmatimonadota, mostly genus *Gemmatimonas*, had higher relative abundance in September in the hypolimnion of the lake, where the oxygen had lower concentrations due to the sedimentation of organic matter, while in March and June, they were rare [92]. Their average abundance was around 1%. In a metagenomic study of Lake Baikal, two novel metagenome-assembled genomes (MAGs) of Gemmatimonadota were reported, one closely related to *G. phototrophica,* which showed higher abundance at 20 m, and the other more similar to the soil representative '*Gemmatirosa kalamazoonensis*', with low abundances at both 5 and 20 m depth [93]. Until recently, it was not clear whether the Gemmatimonadota reported in freshwater environments were limnic species or they originated from surrounding soils. The limnic nature of this group was shown in a large study of several freshwater metagenomes from five freshwater lakes. The lakes differed in trophic status, and the presence of Gemmatimonadota was documented over the whole year at different depths in the lakes—both epilimnion and hypolimnion [18]. Their relative abundance based on the 16S rRNA gene ranged from 0.02 to 0.6% of total bacteria in epilimnion and up to 1% in hypolimnion. Genus *Gemmatimonas* occurred at both depths, but was more abundant in epilimnion. From these freshwater metagenomes, 45 MAGs of Gemmatimonadota were assembled, showing their great diversity in freshwaters [18]. Moreover, several novel genus-level clusters were proposed [18], including a newly defined photoheterotrophic cluster PG1 that is present in the hypolimnion (Figure 7).

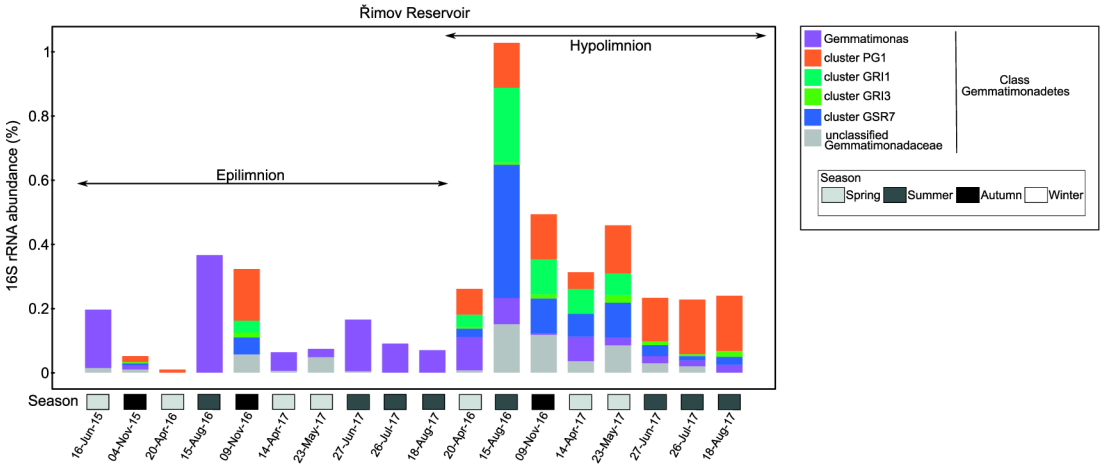

Figure 7. The relative abundances of 16S rRNA gene of Gemmatimonadota from metagenomes of a freshwater Římov Reservoir in the Czech Republic. Only the class *Gemmatimonadetes* is represented in Římov Reservoir, and newly defined clusters are shown (legend on the right). On the *x*-axis are shown sample dates and the season (gray-colored boxes), and on the *y*-axis is shown the percentage of Gemmatimonadota in the prokaryotic community. The figure was adapted from ref. [18].

Different cell morphologies, depending on the depth, were seen in samples from the meso-eutrophic Římov Reservoir in the Czech Republic by using catalyzed-reported deposition–fluorescence in situ hybridization (CARD-FISH). Small, free-living cells were present in the hypolimnion, whereas in the epilimnion, cells were larger and were found in association with diatoms (*Fragilaria* sp.) and cyanobacteria (*Microcystis* sp.) (Figure 8a,b) [18]. Gemmatimonadota could benefit in such a co-occurrence by obtaining organic carbon and, in return, providing inorganic nutrients [94]. Additionally, this dependence was suggested as one of the possible reasons for the low number of cultured members of this phylum. Further analyses of the metabolism of Gemmatimonadota are needed to confirm this relationship.

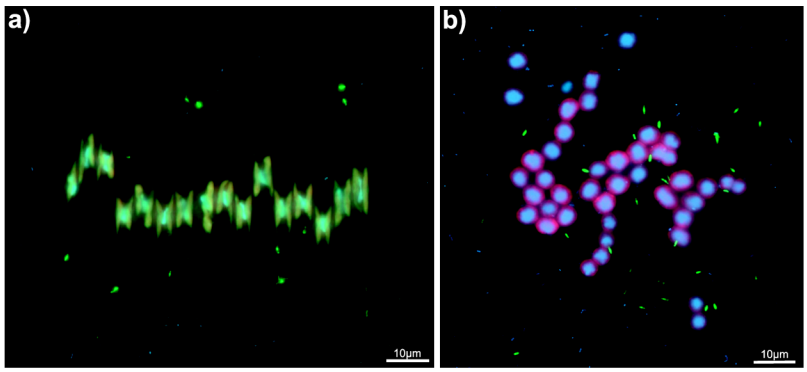

Figure 8. CARD-FISH images showing association of *Gemmatimonadetes* cells with colony of (**a**) *Fragilaria* sp. and (**b**) *Microcystis* sp. The signal of the probe is shown in green, DAPI staining in blue and autofluorescence in red. The photographs were taken by dr. Tanja Shabarova from Institute of Hydrobiology, Biology Centre of the Czech Academy of Sciences. Figure was adapted from ref. [18].

Sequences from BD2-11 terrestrial group, were reported in sediments of Siberian soda lakes [95]. In addition, MAGs containing phototrophic genes were recovered in soda lake sediments, including one that also contained genes encoding large subunit of

RuBisCo [19]. Furthermore, in sediment samples of shallow hypersaline Tuz lake in Turkey, Gemmatimonadota represented 2.7% of total 16S rRNA gene bacterial reads [96].

In marine environments, Gemmatimonadota seems to be mostly associated with sediments [97,98], deep-sea hydrothermal vents [64,99,100], and sponges [101–103]. In the deep-ocean sediment habitats of the Mariana and Massau trenches in the Pacific Ocean, the class *Gemmatimonadetes* was one of the dominant groups in the active bacterial community (in rRNA libraries), with an average relative abundance of OTUs of 13.30% and 9.93%, respectively [104]. *Gemmatimonadetes* were suggested to be a keystone group playing an important role in cycling of organic carbon due to their metabolic strategies [104]. In sediments of the South Eastern Arabian Sea, they were also abundant and represented 2.4% of the total bacterial 16S rRNA reads [105].

Finally, active Gemmatimonadota were found in brackish water in the Gulf of Gdańsk (Baltic Sea) at the estuary of the Vistula River [69]. The highest contribution of Gemmatimonadota to rRNA-based amplicon libraries was observed in summer in a river (>1.1%) and in the mixing zone at a salinity of 3.5‰ (Figure 9). The genus *Gemmatimonas* dominated in these habitats, especially in summer, when it made up over 90% of all Gemmatimonadota reads. Both the phylum Gemmatimonadota and genus *Gemmatimonas* were less active (<0.1% and <0.04% of reads, respectively) in the brackish waters of the Baltic Sea (salinity above 7‰). Interestingly, Gemmatimonadota were not reported at all from the DNA-based surveys of open Baltic Sea waters [106,107].

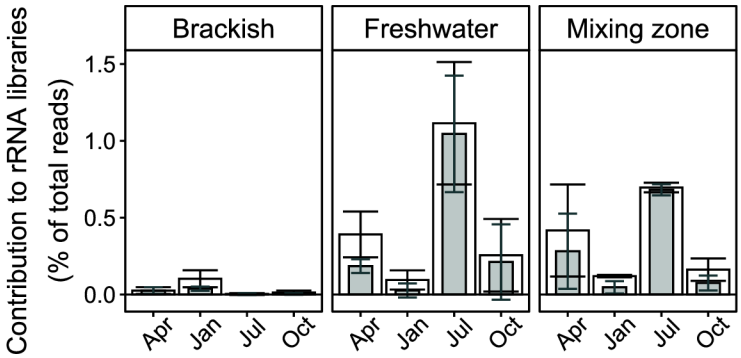

Figure 9. Gemmatimonadota rRNA reads recorded in the Gulf of Gdańsk (Baltic Sea) and the estuary of the Vistula River. The figure shows the percentage of reads associated with Gemmatimonadota (white), as well as the genus *Gemmatimonas* (gray). The salinity of brackish waters was ≥7‰, that of freshwater was <0.5‰, and that of the mixing zone was about 3.5‰. The data used to generate the figure were published in [69].

3.3. Other Environments

Studies of Gemmatimonadota in other environments are scarce. However, they seem to be very important members of bacterial communities in activated sludge and different wastewater treatments. The first cultured species, *G. aurantiaca,* was isolated from activated sludge [8], and higher proportions of both heterotrophic and photoheterotrophic Gemmatimonadota were detected in wastewater metagenomes [17]. In batch reactors used for pretreatment of urea wastewater, Gemmatimonadota became the dominant group and increased their relative abundance to over 50%, exceeding that of even Proteobacteria [108]. It seems that this group could be connected to intracellular urea hydrolysis [109], and urea could be used as an energy source and an important substrate [108]. Gemmatimonadota were also abundant in aquaculture wastewater and in soil irrigated with this water [110], an outlet of wastewater generated during nitrocellulose production [111], and wastewater treatment plants with high salinity [112].

Biofilms and microbial mats are yet another environment with a noticeable presence of Gemmatimonadota [17,113]. They were part of bacterial communities that formed the

base of biofilms attached to a substrate as opposed to streamer structures that floated in water [114] and part of communities of microplastic biofilms [115]. In hypersaline microbial mats under different tidal activity, Gemmatimonadota were the most active in autumn, yet showed high relative RNA proportions in all seasons in tidal mats characterized by dominance of diatoms and the influence of currents and waves [116].

3.4. Distribution of Phototrophic Gemmatimonadota

The distribution and diversity of anoxygenic phototrophs is frequently detected using specific photosynthetic genes serving as molecular markers. The most common molecular marker for phototrophic Proteobacteria, Chloroflexota, and Gemmatimonadota is the *pufM* gene encoding the M subunit of the bacterial type-II reaction center. A more universal marker is gene *bchY* encoding the chlorophyllide reductase subunit Y, since it targets all anoxygenic phototrophic species [117]. However, this marker is not suitable for phototrophic Gemmatimonadota because of its high similarity to the sequences of Proteobacteria. To avoid this, another marker, gene *acsF*, which encodes aerobic oxidative cyclase in many phototrophic organisms, was introduced in a study of the freshwater Lake Taihu [118]. The previous studies documented that the *acsF* gene marker can reliably differentiate between phototrophic Proteobacteria, Gemmatimonadota, and Cyanobacteria [9,118,119]. Amplification of the *acsF* genes from Lake Taihu samples showed that phototrophic Gemmatimonadota represented 17.3% of the *acsF* reads in deep samples and 10.5% in the shallow-sediment samples, while in the water column, they represented only 0.67% of the reads [118].

AcsF sequences from phototrophic Gemmatimonadota were found in 161 metagenomes from different environments, such as wastewater treatment plants, soils, lake water columns and sediments, estuarine waters, biofilms, and plant-associated habitats. However, no *acsF* sequences were found in marine waters [17]. This increased contribution in non-marine aquatic environments suggests that phototrophic Gemmatimonadota may prefer different habitats from those of non-phototrophic species. The diversity of uncultured phototrophic Gemmatimonadota seems to be comparable to the diversity of Proteobacteria [17]. Photoheterotrophic members were also shown to express the photosynthetic genes in freshwater environments. In a study of two lakes in the Czech Republic, the relative abundance of *pufM* from Gemmatimonadota in the libraries prepared from RNA exceeded that from DNA libraries, indicating that photoheterotrophic Gemmatimonadota were active members of planktonic communities of anoxygenic phototrophic bacteria [120].

Although only two cultured photoheterotrophic species are available so far, a recent metagenomic study reported a high diversity of photoheterotrophic Gemmatimonadota in freshwater lakes [18]. They recovered 19 MAGs belonging to different genera that contained PGCs with similar organizations of genes to those of *G. phototrophica* and *G. groenlandica* [16], which indicates that phototrophic genes are conserved in these phyla. The abundance of phototrophic MAGs varied over the seasons, and they were present in both the epilimnion and hypolimnion [18]. MAGs of phototrophic Gemmatimonadota have also been reported in deep layers of Lake Baikal [93] and sediments of a soda lakes [95,121]. Interestingly, while some photoheterotrophic MAGs recovered from freshwater lakes containedRuBisCO-like protein (RLP) [18], which is considered as only a homologue of RuBisCO without carboxylation activity (type IV RuBisCO) [122,123], MAGs recovered in a soda lake showed the presence of genes encoding a large subunit of the RuBisCO enzyme. Moreover, six MAGs from these soda lakes contained all genes involved in the Calvin cycle [121]. This suggests that they represent the first photoautotrophic Gemmatimonadota, expanding the list of roles that these bacteria play in the environment.

4. Summary and Perspective

Members of Gemmatimonadota are present in many different environments. At the moment, most of the information is available from soils [39–41], and recently also from freshwater lakes [18,92,120]. Future research should also focus on other environments,

such as sediments, plant-associated bacteria, or marine environments, where detailed information is currently missing.

Despite the fact that the information about Gemmatimonadota is accumulating, still, little is known about their metabolism and, thus, their environmental role. They usually form only a small fraction of the bacterial community, with relative abundances at around 1%. These low numbers in the environment could relate to their slower growth, which is often associated with the ability to withstand stressful conditions [124]. Members of this group are able to survive in extreme environments, such as saline soils [87], soils in Antarctica [43,46], hypersaline soda lakes [19], or deep-sea sediments [64,99]. This could signify that they are K-strategists with less active metabolisms and resistance against environmental stresses at the cost of lower growth rates. On the other hand, it cannot be excluded that the slow growth rates of cultured species could be the result of suboptimal media choices and/or the need for specific compounds, as specific growth rates of $>2d^{-1}$ of freshwater Gemmatimonadota were reported from a manipulation experiment [125]. The observed relationship with algae and cyanobacteria [18] could be another reason for their low numbers and the difficulties in culturing them. Such relationships could result in a patchy distribution and local predominance in specific microhabitats [83]. Further research should focus on their importance in food webs and biogeochemical cycles. Special attention should be paid to the key biogeochemical processes, such as photo(hetero)trophy, carbon assimilation, phosphorus acquisition, or nitrogen and sulfur metabolism. Potential roles in cycling of nutrients were already highlighted in discoveries of members capable of anoxygenic photosynthesis [15,16], phototrophic MAGs with a possible capacity for carbon fixation [19,121], and a member with potential N_2O reduction capabilities [24]. Additional laboratory experiments with cultured species are necessary in order to elucidate their metabolic properties and physiology. These experiments would be complementary to bioinformatic methods that enable the metabolic potential to be studied in many uncultured organisms by using the available metagenomes. The analysis of metagenome-assembled genomes offers an immense amount of information for studying this interesting but difficult-to-culture bacterial group.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/microorganisms10010151/s1, Table S1: Accession numbers of sequences used for the phylogenetic tree in Figure 5.

Author Contributions: Writing—original draft preparation, I.M., K.P. and M.K.; writing—review and editing, I.M., K.P. and M.K.; funding acquisition, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Science Foundation project PhotoGemm+ 19-28778X.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Alastair T. Gardiner for the language revision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Woese, C.R. Bacterial evolution. *Microbiol. Rev.* **1987**, *51*, 221–271. [CrossRef]
2. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K.; et al. A new view of the tree of life. *Nat. Microbiol.* **2016**, *1*, 16048. [CrossRef] [PubMed]
3. Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.A.; Woodcroft, B.J.; Evans, P.N.; Hugenholtz, P.; Tyson, G.W. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2017**, *2*, 1533–1542. [CrossRef]
4. Hugenholtz, P.; Tyson, G.W.; Webb, R.I.; Wagner, A.M.; Blackall, L.L. Investigation of candidate division tm7, a recently recognized major lineage of the domain bacteria with no known pure-culture representatives. *Appl. Environ. Microbiol.* **2001**, *67*, 411–419. [CrossRef] [PubMed]

5. Mummey, D.L.; Stahl, P.D. Candidate Division BD: Phylogeny, Distribution and abundance in soil ecosystems. *Syst. Appl. Microbiol.* **2003**, *26*, 228–235. [CrossRef] [PubMed]
6. Li, L.; Kato, C.; Horikoshi, K. Bacterial diversity in deep-sea sediments from different depths. *Biodivers. Conserv.* **1999**, *8*, 659–677. [CrossRef]
7. Madrid, V.M.; Aller, J.Y.; Aller, R.C.; Chistoserdov, A.Y. High prokaryote diversity and analysis of community structure in mobile mud deposits off French Guiana: Identification of two new bacterial candidate divisions. *FEMS Microbiol. Ecol.* **2001**, *37*, 197–209. [CrossRef]
8. Zhang, H.; Sekiguchi, Y.; Hanada, S.; Hugenholtz, P.; Kim, H.; Kamagata, Y.; Nakamura, K. *Gemmatimonas aurantiaca* gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 1155–1163. [CrossRef]
9. Zeng, Y.; Feng, F.; Medová, H.; Dean, J.; Koblížek, M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7795–7800. [CrossRef] [PubMed]
10. Ward, L.M.; Cardona, T.; Holland-Moritz, H. Evolutionary Implications of Anoxygenic Phototrophy in the Bacterial Phylum *Candidatus* Eremiobacterota (WPS-2). *Front. Microbiol.* **2019**, *10*, 1658. [CrossRef] [PubMed]
11. Cardona, T. Thinking twice about the evolution of photosynthesis. *Open Biol.* **2019**, *9*, 180246. [CrossRef]
12. Mendler, K.; Chen, H.; Parks, D.H.; Lobb, B.; Hug, L.A.; Doxey, A.C. AnnoTree: Visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **2019**, *47*, 4442–4448. [CrossRef] [PubMed]
13. Parks, D.H.; Chuvochina, M.; Rinke, C.; Mussig, A.J.; Chaumeil, P.-A.; Hugenholtz, P. GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **2021**, *50*, 785–794. [CrossRef]
14. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **2018**, *36*, 996–1004. [CrossRef]
15. Zeng, Y.; Selyanin, V.; Lukeš, M.; Dean, J.; Kaftan, D.; Feng, F.; Koblížek, M. Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 2410–2419. [CrossRef]
16. Zeng, Y.; Nupur; Wu, N.; Madsen, A.M.; Chen, X.; Gardiner, A.T.; Koblížek, M. *Gemmatimonas groenlandica* sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes. *Front. Microbiol.* **2021**, *11*, 606612. [CrossRef]
17. Zeng, Y.; Baumbach, J.; Barbosa, E.G.V.; Azevedo, V.; Zhang, C.; Koblížek, M. Metagenomic evidence for the presence of phototrophic *Gemmatimonadetes bacteria* in diverse environments. *Environ. Microbiol. Rep.* **2016**, *8*, 139–149. [CrossRef] [PubMed]
18. Mujakić, I.; Andrei, A.-Ş.; Shabarova, T.; Fecskeová, L.K.; Salcher, M.M.; Piwosz, K.; Ghai, R.; Koblížek, M. Common Presence of Phototrophic Gemmatimonadota in Temperate Freshwater Lakes. *mSystems* **2021**, *6*, e01241-20. [CrossRef] [PubMed]
19. Vavourakis, C.; Mehrshad, M.; Balkema, C.; Van Hall, R.; Andrei, A.-Ş.; Ghai, R.; Sorokin, D.Y.; Muyzer, G. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol.* **2019**, *17*, 69. [CrossRef] [PubMed]
20. Koblížek, M.; Dachev, M.; Bína, D.; Nupur; Piwosz, K.; Kaftan, D. Utilization of light energy in phototrophic Gemmatimonadetes. *J. Photochem. Photobiol. B Biol.* **2020**, *213*, 112085. [CrossRef]
21. DeBruyn, J.M.; Fawaz, M.N.; Peacock, A.D.; Dunlap, J.R.; Nixon, L.T.; Cooper, K.E.; Radosevich, M. *Gemmatirosa kalamazoonesis* gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum *Gemmatimonadetes*. *J. Gen. Appl. Microbiol.* **2013**, *59*, 305–312. [CrossRef] [PubMed]
22. Pascual, J.; Foesel, B.U.; Geppert, A.; Huber, K.J.; Boedeker, C.; Luckner, M.; Wanner, G.; Overmann, J. *Roseisolibacter agri* gen. nov., sp. nov., a novel slow-growing member of the under-represented phylum *Gemmatimonadetes*. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1028–1036. [CrossRef] [PubMed]
23. Pascual, J.; García-López, M.; Bills, G.F.; Genilloud, O. *Longimicrobium terrae* gen. nov., sp. nov., an oligotrophic bacterium of the under-represented phylum *Gemmatimonadetes* isolated through a system of miniaturized diffusion chambers. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1976–1985. [CrossRef] [PubMed]
24. Park, D.; Kim, H.; Yoon, S. Nitrous oxide reduction by an obligate. *Appl. Environ. Microbiol.* **2017**, *83*, 1–12. [CrossRef]
25. Chee-Sanford, J.; Tian, D.; Sanford, R. Consumption of N_2O and other N-cycle intermediates by *Gemmatimonas aurantiaca* strain T-27. *Microbiology* **2019**, *165*, 1345–1354. [CrossRef]
26. Hohmann-Marriott, M.F.; Blankenship, R.E. Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* **2011**, *62*, 515–548. [CrossRef]
27. Nagashima, S.; Nagashima, K.V.P. Comparison of Photosynthesis Gene Clusters Retrieved from Total Genome Sequences of Purple Bacteria. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 66, pp. 151–178; ISBN 9780123979230.
28. Zhaxybayeva, O.; Gogarten, J.P.; Charlebois, R.L.; Doolittle, W.F.; Papke, R.T. Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res.* **2006**, *16*, 1099–1108. [CrossRef]
29. Sousa, F.L.; Shavit-Grievink, L.; Allen, J.F.; Martin, W.F. Chlorophyll Biosynthesis Gene Evolution Indicates Photosystem Gene Duplication, Not Photosystem Merger, at the Origin of Oxygenic Photosynthesis. *Genome Biol. Evol.* **2013**, *5*, 200–216. [CrossRef]
30. Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.; Nagashima, K.V.P. Horizontal Transfer of the Photosynthesis Gene Cluster and Operon Rearrangement in Purple Bacteria. *J. Mol. Evol.* **2001**, *52*, 333–341. [CrossRef]

31. Ward, L.M.; Hemp, J.; Shih, P.M.; McGlynn, S.E.; Fischer, W.W. Evolution of Phototrophy in the Chloroflexi Phylum Driven by Horizontal Gene Transfer. *Front. Microbiol.* **2018**, *9*, 260. [CrossRef]
32. Brinkmann, H.; Göker, M.; Koblížek, M.; Wagner-Döbler, I.; Petersen, J. Horizontal operon transfer, plasmids, and the evolution of photosynthesis in Rhodobacteraceae. *ISME J.* **2018**, *12*, 1994–2010. [CrossRef]
33. Cardona, T. Origin of Bacteriochlorophyll a and the Early Diversification of Photosynthesis. *PLoS ONE* **2016**, *11*, e0151250. [CrossRef] [PubMed]
34. Dachev, M.; Bína, D.; Sobotka, R.; Moravcová, L.; Gardian, Z.; Kaftan, D.; Šlouf, V.; Fuciman, M.; Polívka, T.; Koblížek, M. Unique double concentric ring organization of light harvesting complexes in *Gemmatimonas phototrophica*. *PLoS Biol.* **2017**, *15*, 1–16. [CrossRef]
35. Yurkov, V.V.; Beatty, J.T. Aerobic Anoxygenic Phototrophic Bacteria. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 695–724. [CrossRef]
36. Takaichi, S.; Maoka, T.; Takasaki, K.; Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (*Gemmatimonadetes*): Identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2′-dirhamnoside. *Microbiology* **2010**, *156*, 757–763. [CrossRef]
37. Yurkov, V.; Csotonyi, J.T. New Light on Aerobic Anoxygenic Phototrophs. In *Advances in Photosynthesis and Respiration*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 31–55.
38. Nupur; Kuzma, M.; Hájek, J.; Hrouzek, P.; Gardiner, A.T.; Lukeš, M.; Moos, M.; Šimek, P.; Koblížek, M. Structure elucidation of the novel carotenoid gemmatoxanthin from the photosynthetic complex of *Gemmatimonas phototrophica* AP64. *Sci. Rep.* **2021**, *11*, 15964. [CrossRef] [PubMed]
39. Janssen, P.H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* **2006**, *72*, 1719–1728. [CrossRef]
40. DeBruyn, J.M.; Nixon, L.T.; Fawaz, M.N.; Johnson, A.M.; Radosevich, M. Global Biogeography and Quantitative Seasonal Dynamics of *Gemmatimonadetes* in Soil. *Appl. Environ. Microbiol.* **2011**, *77*, 6295–6300. [CrossRef] [PubMed]
41. Delgado-Baquerizo, M.; Oliverio, A.M.; Brewer, T.E.; Benavent-González, A.; Eldridge, D.J.; Bardgett, R.D.; Maestre, F.T.; Singh, B.K.; Fierer, N. A global atlas of the dominant bacteria found in soil. *Science* **2018**, *359*, 320–325. [CrossRef]
42. Ren, C.; Chen, J.; Lu, X.; Doughty, R.; Zhao, F.; Zhong, Z.; Han, X.; Yang, G.; Feng, Y.; Ren, G. Responses of soil total microbial biomass and community compositions to rainfall reductions. *Soil Biol. Biochem.* **2018**, *116*, 4–10. [CrossRef]
43. Bakermans, C.; Skidmore, M.L.; Douglas, S.; McKay, C.P. Molecular characterization of bacteria from permafrost of the Taylor Valley, Antarctica. *FEMS Microbiol. Ecol.* **2014**, *89*, 331–346. [CrossRef]
44. Frey, B.; Rime, T.; Phillips, M.; Stierli, B.; Hajdas, I.; Widmer, F.; Hartmann, M. Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw018. [CrossRef] [PubMed]
45. Tuorto, S.J.; Darias, P.; McGuinness, L.R.; Panikov, N.; Zhang, T.; Häggblom, M.M.; Kerkhof, L.J. Bacterial genome replication at subzero temperatures in permafrost. *ISME J.* **2014**, *8*, 139–149. [CrossRef] [PubMed]
46. Cary, S.C.; McDonald, I.R.; Barrett, J.E.; Cowan, D.A. On the rocks: The microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.* **2010**, *8*, 129–138. [CrossRef] [PubMed]
47. Breidenbach, B.; Pump, J.; Dumont, M.G. Microbial Community Structure in the Rhizosphere of Rice Plants. *Front. Microbiol.* **2016**, *6*, 1537. [CrossRef] [PubMed]
48. Yang, Y.; Wang, N.; Guo, X.; Zhang, Y.; Ye, B. Comparative analysis of bacterial community structure in the rhizosphere of maize by high-throughput pyrosequencing. *PLoS ONE* **2017**, *12*, e0178425. [CrossRef]
49. Gkarmiri, K.; Mahmood, S.; Ekblad, A.; Alström, S.; Högberg, N.; Finlay, R. Identifying the Active Microbiome Associated with Roots and Rhizosphere Soil of Oilseed Rape. *Appl. Environ. Microbiol.* **2017**, *83*, e01938-17. [CrossRef]
50. Ren, N.; Wang, Y.; Ye, Y.; Zhao, Y.; Huang, Y.; Fu, W.; Chu, X. Effects of Continuous Nitrogen Fertilizer Application on the Diversity and Composition of Rhizosphere Soil Bacteria. *Front. Microbiol.* **2020**, *11*, 1948. [CrossRef] [PubMed]
51. Rodriguez, M.T.F.; Valverde, N.B.; Lagurara, P.; Revale, S.; De Souza, J.A.M.; Vilaro, M.D.R. Soil and Rhizosphere Bacterial Diversity in Maize Agro- Ecosystem. *Sustain. Agric. Res.* **2017**, *6*, 35. [CrossRef]
52. Na, X.; Xu, T.; Li, M.; Zhou, Z.; Ma, S.; Wang, J.; He, J.; Jiao, B.; Ma, F. Variations of Bacterial Community Diversity Within the Rhizosphere of Three Phylogenetically Related Perennial Shrub Plant Species Across Environmental Gradients. *Front. Microbiol.* **2018**, *9*, 709. [CrossRef] [PubMed]
53. Gugliandolo, C.; Michaud, L.; Giudice, A.L.; Lentini, V.; Rochera, C.; Camacho, A.; Maugeri, T.L. Prokaryotic Community in Lacustrine Sediments of Byers Peninsula (Livingston Island, Maritime Antarctica). *Microb. Ecol.* **2016**, *71*, 387–400. [CrossRef]
54. Sheng, P.; Yu, Y.; Zhang, G.; Huang, J.; He, L.; Ding, J. Bacterial diversity and distribution in seven different estuarine sediments of Poyang Lake, China. *Environ. Earth Sci.* **2016**, *75*, 479. [CrossRef]
55. Röske, K.; Sachse, R.; Scheerer, C.; Röske, I. Microbial diversity and composition of the sediment in the drinking water reservoir Saidenbach (Saxonia, Germany). *Syst. Appl. Microbiol.* **2012**, *35*, 35–44. [CrossRef] [PubMed]
56. Song, H.; Li, Z.; Du, B.; Wang, G.; Ding, Y. Bacterial communities in sediments of the shallow Lake Dongping in China. *J. Appl. Microbiol.* **2012**, *112*, 79–89. [CrossRef] [PubMed]
57. Liu, Y.; Zhang, J.; Zhao, L.; Zhang, X.; Xie, S. Spatial distribution of bacterial communities in high-altitude freshwater wetland sediment. *Limnology* **2014**, *15*, 249–256. [CrossRef]
58. Zhang, J.; Yang, Y.; Yuzhao, L.; Li, Y.; Xie, S.; Liu, Y. Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 3291–3302. [CrossRef]

59. Hanada, S.; Sekiguchi, Y. The phylum Gemmatimonadetes. In *The Prokaryotes*, 4th ed.; Springer: Berlin/Heidelberg, Germany, 2014; Volume 11, pp. 677–681.
60. Qin, H.; Ji, B.; Zhang, S.; Kong, Z. Study on the bacterial and archaeal community structure and diversity of activated sludge from three wastewater treatment plants. *Mar. Pollut. Bull.* **2018**, *135*, 801–807. [CrossRef]
61. Durbin, A.M.; Teske, A. Microbial diversity and stratification of South Pacific abyssal marine sediments. *Environ. Microbiol.* **2011**, *13*, 3219–3234. [CrossRef] [PubMed]
62. Cerqueira, T.; Pinho, D.; Egas, C.; Froufe, H.; Altermark, B.; Candeias, C.; Santos, R.S.; Bettencourt, R. Microbial diversity in deep-sea sediments from the Menez Gwen hydrothermal vent system of the Mid-Atlantic Ridge. *Mar. Genom.* **2015**, *24*, 343–355. [CrossRef]
63. Zhang, J.; Sun, Q.-L.; Zeng, Z.-G.; Chen, S.; Sun, L. Microbial diversity in the deep-sea sediments of Iheya North and Iheya Ridge, Okinawa Trough. *Microbiol. Res.* **2015**, *177*, 43–52. [CrossRef] [PubMed]
64. Zhang, L.; Kang, M.; Xu, J.; Xu, J.; Shuai, Y.; Zhou, X.; Yang, Z.; Ma, K. Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. *Sci. Rep.* **2016**, *6*, 25982. [CrossRef]
65. Kamke, J.; Taylor, M.W.; Schmitt, S. Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME J.* **2010**, *4*, 498–508. [CrossRef] [PubMed]
66. Montalvo, N.F.; Hill, R.T. Sponge-Associated Bacteria Are Strictly Maintained in Two Closely Related but Geographically Distant Sponge Hosts. *Appl. Environ. Microbiol.* **2011**, *77*, 7207–7216. [CrossRef] [PubMed]
67. Souza, D.T.; Genuário, D.B.; Silva, F.S.P.; Pansa, C.C.; Kavamura, V.N.; Moraes, F.C.; Taketani, R.G.; Melo, I.S. Analysis of bacterial composition in marine sponges reveals the influence of host phylogeny and environment. *FEMS Microbiol. Ecol.* **2017**, *93*, fiw204. [CrossRef] [PubMed]
68. Kennedy, J.; Flemer, B.; Jackson, S.A.; Morrissey, J.P.; O'Gara, F.; Dobson, A.D.W. Evidence of a Putative Deep Sea Specific Microbiome in Marine Sponges. *PLoS ONE* **2014**, *9*, e91092. [CrossRef] [PubMed]
69. Gołębiewski, M.; Całkiewicz, J.; Creer, S.; Piwosz, K. Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria: The case study of the Vistula river estuary. *Environ. Microbiol. Rep.* **2017**, *9*, 129–143. [CrossRef] [PubMed]
70. Hentschel, U.; Hopke, J.; Horn, M.; Friedrich, A.B.; Wagner, M.; Hacker, J.; Moore, B.S. Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* **2002**, *68*, 4431–4440. [CrossRef]
71. Dunbar, J.; Barns, S.M.; Ticknor, L.O.; Kuske, C.R. Empirical and theoretical bacterial diversity in four arizona soils. *Appl. Environ. Microbiol.* **2002**, *68*, 3035–3045. [CrossRef]
72. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [CrossRef]
73. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
74. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Von Haeseler, A.; Jermiin, L.S. Model Finder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
75. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef] [PubMed]
76. Bajerski, F.; Wagner, D. Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. *FEMS Microbiol. Ecol.* **2013**, *85*, 128–142. [CrossRef] [PubMed]
77. Niederberger, T.D.; McDonald, I.R.; Hacker, A.L.; Soo, R.M.; Barrett, J.E.; Wall, D.H.; Cary, S.C. Microbial community composition in soils of Northern Victoria Land, Antarctica. *Environ. Microbiol.* **2008**, *10*, 1713–1724. [CrossRef] [PubMed]
78. Acosta-Martínez, V.; Dowd, S.; Sun, Y.; Allen, V. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biol. Biochem.* **2008**, *40*, 2762–2770. [CrossRef]
79. Kim, J.-S.; Dungan, R.S.; Crowley, D. Microarray analysis of bacterial diversity and distribution in aggregates from a desert agricultural soil. *Biol. Fertil. Soils* **2008**, *44*, 1003–1011. [CrossRef]
80. Ahmed, V.; Verma, M.K.; Gupta, S.; Mandhan, V.; Chauhan, N.S. Metagenomic Profiling of Soil Microbes to Mine Salt Stress Tolerance Genes. *Front. Microbiol.* **2018**, *9*, 159. [CrossRef] [PubMed]
81. Zhao, S.; Liu, J.J.; Banerjee, S.; Zhou, N.; Zhao, Z.Y.; Zhang, K.; Hu, M.F.; Tian, C.Y. Biogeographical distribution of bacterial communities in saline agricultural soil. *Geoderma* **2020**, *361*, 114095. [CrossRef]
82. Neilson, J.W.; Califf, K.; Cardona, C.; Copeland, A.; van Treuren, W.; Josephson, K.L.; Knight, R.; Gilbert, J.A.; Quade, J.; Caporaso, J.G.; et al. Significant Impacts of Increasing Aridity on the Arid Soil Microbiome. *mSystems* **2017**, *2*, 1–15. [CrossRef]
83. Mummey, D.; Holben, W.; Six, J.; Stahl, P. Spatial Stratification of Soil Bacterial Populations in Aggregates of Diverse Soils. *Microb. Ecol.* **2006**, *51*, 404–411. [CrossRef]
84. Lauber, C.L.; Strickland, M.S.; Bradford, M.A.; Fierer, N. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* **2008**, *40*, 2407–2415. [CrossRef]
85. Mendez, M.O.; Neilson, J.W.; Maier, R.M. Characterization of a Bacterial Community in an Abandoned Semiarid Lead-Zinc Mine Tailing Site. *Appl. Environ. Microbiol.* **2008**, *74*, 3899–3907. [CrossRef] [PubMed]
86. Malard, L.A.; Anwar, M.Z.; Jacobsen, C.S.; Pearce, D.A. Biogeographical patterns in soil bacterial communities across the Arctic region. *FEMS Microbiol. Ecol.* **2019**, *95*, fiz128. [CrossRef] [PubMed]

87. Guan, Y.; Jiang, N.; Wu, Y.; Yang, Z.; Bello, A.; Yang, W. Disentangling the role of salinity-sodicity in shaping soil microbiome along a natural saline-sodic gradient. *Sci. Total Environ.* **2021**, *765*, 142738. [CrossRef] [PubMed]
88. Liu, M.; Li, X.; Zhu, R.; Chen, N.; Ding, L.; Chen, C. Vegetation richness, species identity and soil nutrients drive the shifts in soil bacterial communities during restoration process. *Environ. Microbiol. Rep.* **2021**, *13*, 1758–2229. [CrossRef]
89. Deng, J.; Bai, X.; Zhou, Y.; Zhu, W.; Yin, Y. Variations of soil microbial communities accompanied by different vegetation restoration in an open-cut iron mining area. *Sci. Total Environ.* **2020**, *704*, 135243. [CrossRef]
90. Ye, W.; Liu, X.; Lin, S.; Tan, J.; Pan, J.; Li, D.; Yang, H. The vertical distribution of bacterial and archaeal communities in the water and sediment of Lake Taihu. *FEMS Microbiol. Ecol.* **2009**, *70*, 263–276. [CrossRef]
91. Traving, S.J.; Rowe, O.; Jakobsen, N.M.; Sørensen, H.; Dinasquet, J.; Stedmon, C.A.; Andersson, A.; Riemann, L. The Effect of Increased Loads of Dissolved Organic Matter on Estuarine Microbial Community Composition and Function. *Front. Microbiol.* **2017**, *8*, 1–15. [CrossRef]
92. Morrison, J.M.; Baker, K.D.; Zamor, R.M.; Nikolai, S.; Elshahed, M.S.; Youssef, N.H. Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLoS ONE* **2017**, *12*, e0177488. [CrossRef]
93. Cabello-Yeves, P.J.; Zemskaya, T.; Rosselli, R.; Coutinho, F.H.; Zakharenko, A.S.; Blinov, V.V.; Rodriguez-Valera, F. Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Appl. Environ. Microbiol.* **2018**, *84*, e02132-17. [CrossRef]
94. Shia, L.; Cai, Y.; Wang, X.; Li, P.; Yu, Y.; Kong, F. Community Structure of Bacteria Associated withMicrocystisColonies from Cyanobacterial Blooms. *J. Freshw. Ecol.* **2010**, *25*, 193–203. [CrossRef]
95. Vavourakis, C.D.; Andrei, A.S.; Mehrshad, M.; Ghai, R.; Sorokin, D.Y.; Muyzer, G. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments 06 Biological Sciences 0605 Microbiology 06 Biological Sciences 0604 Genetics. *Microbiome* **2018**, *6*, 168. [CrossRef] [PubMed]
96. Çınar, S.; Mutlu, M.B. Prokaryotic community compositions of the hypersaline sediments of tuz lake demonstrated by cloning and high-throughput sequencing. *Microbiology* **2020**, *89*, 756–768. [CrossRef]
97. Cui, G.; Li, J.; Gao, Z.; Wang, Y. Spatial variations of microbial communities in abyssal and hadal sediments across the Challenger Deep. *PeerJ* **2019**, *7*, e6961. [CrossRef] [PubMed]
98. Peoples, L.M.; Grammatopoulou, E.; Pombrol, M.; Xu, X.; Osuntokun, O.; Blanton, J.C.; Allen, E.E.; Nunnally, C.C.; Drazen, J.; Mayor, D.J.; et al. Microbial Community Diversity Within Sediments from Two Geographically Separated Hadal Trenches. *Front. Microbiol.* **2019**, *10*, 347. [CrossRef]
99. Kato, S.; Nakawake, M.; Kita, J.; Yamanaka, T.; Utsumi, M.; Okamura, K.; Ishibashi, J.-I.; Ohkuma, M.; Yamagishi, A. Characteristics of Microbial Communities in Crustal Fluids in a Deep-Sea Hydrothermal Field of the Suiyo Seamount. *Front. Microbiol.* **2013**, *4*, 85. [CrossRef]
100. Nunoura, T.; Nishizawa, M.; Hirai, M.; Shimamura, S.; Harnvoravongchai, P.; Koide, O.; Morono, Y.; Fukui, T.; Inagaki, F.; Miyazaki, J.; et al. Microbial Diversity in Sediments from the Bottom of the Challenger Deep, the Mariana Trench. *Microbes Environ.* **2018**, *33*, 186–194. [CrossRef]
101. Thiel, V.; Neulinger, S.C.; Staufenberger, T.; Schmaljohann, R.; Imhoff, J.F. Spatial distribution of sponge-associated bacteria in the Mediterranean sponge *Tethya aurantium*. *FEMS Microbiol. Ecol.* **2007**, *59*, 47–63. [CrossRef]
102. Slaby, B.M.; Hackl, T.; Horn, H.; Bayer, K.; Hentschel, U. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J.* **2017**, *11*, 2465–2478. [CrossRef]
103. Alvarez-Yela, A.C.A.; Mosquera-Rendón, J.; Noreña-P, A.; Cristancho, M.; López-Alvarez, D. Microbial Diversity Exploration of Marine Hosts at Serrana Bank, a Coral Atoll of the Seaflower Biosphere Reserve. *Front. Mar. Sci.* **2019**, *6*, 338. [CrossRef]
104. Liu, R.; Wang, Z.; Wang, L.; Li, Z.; Fang, J.; Wei, X.; Wei, W.; Cao, J.; Wei, Y.; Xie, Z. Bulk and Active Sediment Prokaryotic Communities in the Mariana and Mussau Trenches. *Front. Microbiol.* **2020**, *11*, 1521. [CrossRef]
105. Vipindas, P.V.; Mujeeb, R.K.M.; Jabir, T.; Thasneem, T.R.; Mohamed Hatha, A.A. Diversity of sediment bacterial communities in the South Eastern Arabian Sea. *Reg. Stud. Mar. Sci.* **2020**, *35*, 101153. [CrossRef]
106. Herlemann, D.P.R.; Labrenz, M.; Jürgens, K.; Bertilsson, S.; Waniek, J.J.; Andersson, A.F. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **2011**, *5*, 1571–1579. [CrossRef] [PubMed]
107. Hu, Y.O.O.; Karlson, B.; Charvet, S.; Andersson, A.F. Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea. *Front. Microbiol.* **2016**, *7*, 679. [CrossRef] [PubMed]
108. Chen, Y.; Chen, H.; Chen, Z.; Hu, H.; Deng, C.; Wang, X. The benefits of autotrophic nitrogen removal from high concentration of urea wastewater through a process of urea hydrolysis and partial nitritation in sequencing batch reactor. *J. Environ. Manag.* **2021**, *292*, 112762. [CrossRef]
109. Hu, H.; Deng, C.; Wang, X.; Chen, Z.; Zhong, Z.; Wang, R. Performance and mechanism of urea hydrolysis in partial nitritation system based on SBR. *Chemosphere* **2020**, *258*, 127228. [CrossRef]
110. Chen, L.; Feng, Q.; Li, C.; Wei, Y.; Zhao, Y.; Feng, Y.; Zheng, H.; Li, F.; Li, H. Impacts of aquaculture wastewater irrigation on soil microbial functional diversity and community structure in arid regions. *Sci. Rep.* **2017**, *7*, 11193. [CrossRef] [PubMed]
111. Ziganshina, E.E.; Ibragimov, E.M.; Ilinskaya, O.N.; Ziganshin, A.M. Bacterial communities inhabiting toxic industrial wastewater generated during nitrocellulose production. *Biologia* **2016**, *71*, 70–78. [CrossRef]

112. Correa-Galeote, D.; Roibás-Rozas, A.; Mosquera-Corral, A.; Juárez-Jiménez, B.; González-López, J.; Rodelas, B. Revealing the dissimilar structure of microbial communities in different WWTPs that treat fish-canning wastewater with different NaCl content. *J. Water Process. Eng.* **2021**, *44*, 102328. [CrossRef]
113. Rampadarath, S.; Bandhoa, K.; Puchooa, D.; Jeewon, R.; Bal, S. Early bacterial biofilm colonizers in the coastal waters of Mauritius. *Electron. J. Biotechnol.* **2017**, *29*, 13–21. [CrossRef]
114. Besemer, K.; Hödl, I.; Singer, G.; Battin, T.J. Architectural differentiation reflects bacterial community structure in stream biofilms. *ISME J.* **2009**, *3*, 1318–1324. [CrossRef] [PubMed]
115. Wu, X.; Pan, J.; Li, M.; Li, Y.; Bartlam, M.; Wang, Y. Selective enrichment of bacterial pathogens by microplastic biofilm. *Water Res.* **2019**, *165*, 114979. [CrossRef] [PubMed]
116. Cardoso, D.C.; Cretoiu, M.S.; Stal, L.J.; Bolhuis, H. Seasonal development of a coastal microbial mat. *Sci. Rep.* **2019**, *9*, 9035. [CrossRef] [PubMed]
117. Yutin, N.; Suzuki, M.T.; Rosenberg, M.; Rotem, D.; Madigan, M.T.; Süling, J.; Imhoff, J.F.; Béjà, O. BchY-Based Degenerate Primers Target All Types of Anoxygenic Photosynthetic Bacteria in a Single PCR. *Appl. Environ. Microbiol.* **2009**, *75*, 7556–7559. [CrossRef] [PubMed]
118. Huang, Y.; Zeng, Y.; Lu, H.; Feng, H.; Zeng, Y.; Koblížek, M. Novel acsF Gene Primers Revealed a Diverse Phototrophic Bacterial Population, Including *Gemmatimonadetes*, in Lake Taihu (China). *Appl. Environ. Microbiol.* **2016**, *82*, 5587–5594. [CrossRef]
119. Boldareva-Nuianzina, E.N.; Bláhová, Z.; Sobotka, R.; Koblížek, M. Distribution and Origin of Oxygen-Dependent and Oxygen-Independent Forms of Mg-Protoporphyrin Monomethylester Cyclase among Phototrophic Proteobacteria. *Appl. Environ. Microbiol.* **2013**, *79*, 2596–2604. [CrossRef] [PubMed]
120. Fecskeová, L.K.; Piwosz, K.; Hanusová, M.; Nedoma, J.; Znachor, P.; Koblížek, M. Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. *Sci. Rep.* **2019**, *9*, 18766. [CrossRef]
121. Zorz, J.K.; Sharp, C.; Kleiner, M.; Gordon, P.M.K.; Pon, R.T.; Dong, X.; Strous, M. A shared core microbiome in soda lakes separated by large distances. *Nat. Commun.* **2019**, *10*, 4230. [CrossRef]
122. Ashida, H.; Saito, Y.; Kojima, C.; Kobayashi, K.; Ogasawara, N.; Yokota, A. A Functional Link Between RuBisCO-like Protein of Bacillus and Photosynthetic RuBisCO. *Science* **2003**, *302*, 286–290. [CrossRef] [PubMed]
123. Ashida, H.; Danchin, A.; Yokota, A. Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism? *Res. Microbiol.* **2005**, *156*, 611–618. [CrossRef]
124. Gudelj, I.; Weitz, J.S.; Ferenci, T.; Horner-Devine, M.; Marx, C.J.; Meyer, J.R.; Forde, S.E. An integrative approach to understanding microbial diversity: From intracellular mechanisms to community structure. *Ecol. Lett.* **2010**, *13*, 1073–1084. [CrossRef] [PubMed]
125. Piwosz, K.; Shabarova, T.; Tomasch, J.; Šimek, K.; Kopejtka, K.; Kahl, S.; Pieper, D.H.; Koblížek, M. Determining lineage-specific bacterial growth curves with a novel approach based on amplicon reads normalization using internal standard (ARNIS). *ISME J.* **2018**, *12*, 2640–2654. [CrossRef] [PubMed]

Paper III

Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake

Villena-Alemany, C., Mujakić, I., Porcal, P., Koblížek, M. and Piwosz, K.
(2023)
Environmental Microbiology Reports, 15 (1), 60-71

3.3. Summary results derived from Paper No. 3

In this paper, we studied the temporal changes of the total bacterial and aerobic anoxygenic phototrophic (AAP) bacterial community in the epilimnion and the hypolimnion of a shallow freshwater Cep lake in the Czech Republic. For this, we used amplicon sequencing of the 16S rRNA gene and the *pufM* gene, a common marker for AAP bacteria.

We observed a similar abundance pattern for all bacteria in both the epilimnion and hypolimnion, with maximum values in spring or summer, respectively, while AAP bacteria showed the highest abundance for both depths in spring and with a higher contribution to total bacterial abundance in the epilimnion than in the hypolimnion.

Amplicons of the 16S rRNA gene showed that total bacterial alpha diversity was highest in summer in the hypolimnion, and the largest differences in the community composition between the two layers occurred during the stratification time in summer. As previously seen in other freshwater lakes, the relative abundance of Gemmatimonadota varied in two layers but was generally below 1%. They were represented by three ASVs belonging to the genus *Gemmatimonas*. Regarding other AAP bacteria, Chloroflexota represented less than 4% of total 16S rRNA gene sequences, and Proteobacteria were the most abundant, with Gammaproteobacteria having much higher numbers (51–77%) than Alphaproteobacteria (23–49%).

The number of *pufM* ASVs was also higher in the hypolimnion and, similarly to the total bacterial community, the AAP community started to differ between layers mostly in summer. From different AAP phyla, in this dataset, only Gemmatimonadota, Chloroflexota and Proteobacteria were detected. Gemmatimonadota formed less than 2% of the AAP community, except in September when their contribution increased to 9%. The Gemmatimonadota community consisted of three *pufM* ASVs, which showed the same temporal dynamic as ASVs identified in 16S rRNA analysis. The highest contribution to the AAP community had Proteobacteria (>90% of *pufM* reads), especially Gammaproteobacteria (>50%), while Chloroflexota usually formed up to 5 % of the *pufM* sequences.

Furthermore, we tested the influence of several environmental factors on total bacterial and AAP communities using distance-based linear models (DistLM) and distance-based redundancy analysis (dbRDA). Factors best explaining the variability of both communities were oxygen, temperature and dissolved organic carbon (DOC). However, their explanatory power differed, and the AAP community was mostly influenced by DOC and temperature, while oxygen was the best explanatory variable for the total bacterial community.

Received: 18 March 2022 | Accepted: 19 September 2022

DOI: 10.1111/1758-2229.13131

BRIEF REPORT

environmental microbiology reports Applied Microbiology International

Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake

Cristian Villena-Alemany[1,2] | **Izabela Mujakić**[1,2] | **Petr Porcal**[2,3] | **Michal Koblížek**[1,2] | **Kasia Piwosz**[4]

[1]Laboratory of Anoxygenic Phototrophs, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czechia

[2]Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czechia

[3]Department of Hydrochemistry and Ecosystem Modelling, Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czechia

[4]Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, Gdynia, Poland

Correspondence
Kasia Piwosz, Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, ul. Kołłątaja 1, 81-332 Gdynia, Poland.
Email: kpiwosz@mir.gdynia.pl

Abstract

Aerobic anoxygenic photoheterotrophic (AAP) bacteria represent a functional group of prokaryotic organisms that harvests light energy using bacteriochlorophyll-containing photosynthetic reaction centers. They represent an active and rapidly growing component of freshwater bacterioplankton, with the highest numbers observed usually in summer. Species diversity of freshwater AAP bacteria has been studied before in lakes, but its seasonal dynamics remain unknown. In this report, we analysed temporal changes in the composition of the phototrophic community in an oligo-mesotrophic freshwater lake using amplicon sequencing of the *puf*M marker gene. The AAP community was dominated by phototrophic Gammaproteobacteria and Alphaproteobacteria, with smaller contribution of phototrophic Chloroflexota and Gemmatimonadota. Phototrophic Eremiobacteriota or members of Myxococcota were not detected. Interestingly, some AAP taxa, such as *Limnohabitans*, *Rhodoferax,* Rhodobacterales or Rhizobiales, were permanently present over the sampling period, while others, such as Sphingomonadales, Rhodospirillales or Caulobacterales appeared only transiently. The environmental factors that best explain the seasonal changes in AAP community were temperature, concentrations of oxygen and dissolved organic matter.

INTRODUCTION

Photoheterotrophic bacteria represent an important component of freshwater bacterioplankton. These organisms harvest light energy but, as they do not fix carbon, require organic carbon to build biomass. Two main groups of aquatic photoheterotrophic bacteria are rhodopsin-containing bacteria and aerobic anoxygenic phototrophic (AAP) bacteria. Rhodopsins are membrane-bound proteins able to use light energy to translocate protons across the membrane. The proton gradient is used for ATP production (Lanyi, 2004; Spudich & Luecke, 2002). In contrast, AAP bacteria harvest light energy using photosynthetic complexes containing bacteriochlorophyll-*a* (BChl-*a*) (Koblížek, 2015; Yurkov & Csotonyi, 2009). Due to the captured energy, AAPs reduce utilization of organic carbon for respiration and increase their biomass yield up to 100% (Hauruseu & Koblížek, 2012; Piwosz et al., 2018). When compared with rhodopsin-containing bacteria, AAP bacteria are more effective in producing energy from light (Kirchman & Hanson, 2013).

AAP bacteria contribute 1–22% of all prokaryote abundance in the euphotic layer of freshwater lakes (Čuperová et al., 2013; Fauteux et al., 2015; Mašín et al., 2008, 2012; Ruiz-González et al., 2013; Ruiz-González et al., 2020). They have, on average, larger cell size and exhibit faster growth and mortality rates than heterotrophic bacteria (Cepáková et al., 2016; Garcia-Chaves et al., 2016). Recently, it has been shown that upon infra-red illumination AAP bacteria reduce total microbial respiration by 15% and increase

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Environmental Microbiology Reports* published by Applied Microbiology International and John Wiley & Sons Ltd.

microbial production by 6% (Piwosz et al., 2022). All these findings indicate the importance of AAPs in microbial food webs and the freshwater carbon cycle (Fauteux et al., 2015; Koblížek et al., 2005).

The most common representatives of AAP bacteria in freshwaters are members of the order Burkholderiales (Gammaproteobacteria), such as *Limnohabitans*, *Rhodoferax* and *Polynucleobacter* and members of orders Sphingomonadales and Rhodobacterales (Alphaproteobacteria), with a minor contribution of Chloroflexota and Gemmatimonadota (Caliz & Casamayor, 2014; Fecskeová et al., 2019; Kasalický et al., 2018; Martinez-Garcia et al., 2012; Mujakić et al., 2021; Salka et al., 2011). Unfortunately, most of these studies provided only snap-shot information on AAP community composition from a single or infrequent sampling of multiple lakes, even though seasonal time series have shown AAP bacteria undergo large changes in abundance, activity and growth rates (Cepáková et al., 2016; Čuperová et al., 2013; Kolářová et al., 2019). This has likely also a large impact on AAP community composition, which remain unexplored. Therefore, we decided to investigate how the diversity changes during the year.

Here, we applied amplicon sequencing of the 16S rRNA gene and *puf*M marker gene for AAP bacteria to study temporal changes in total bacterial and AAP bacteria community composition in the epi- and the hypolimnion in a shallow, meso-oligotrophic freshwater lake Cep. We hypothesized that the diversity of AAP bacteria would show distinct temporal patterns in the epi- and the hypolimnion, as had been observed for all bacteria (Zemskaya et al., 2020). We also hypothesized that the AAP community would be affected by different environmental factors than the overall bacterial community.

EXPERIMENTAL PROCEDURES

Sampling

Samples were collected from April to September 2016 from oligotrophic Cep lake (48°92′49.24″ N, 14°88′68.11″ E) located in the Třeboň Basin Protected Landscape Area, Czechia. The lake was created in the second half of the 20th century as a result of sand mining and was filled with groundwater seeping from the nearby river Lužnice. Water samples were collected from five depths: 0.5, 2, 5, 7 and 9 m using a Ruttner Water Sampler (model 11.003KC Denmark AS). Water was transported to the laboratory in closed plastic containers, which were prerinsed three times with the sampled water and stored in a cooled box. Temperature and oxygen profiles were taken with an EXO1 multi-parameter probe (YSI Inc., Yellow Springs, OH).

Bacteria and AAP bacteria microscopy counts

Samples of 50 ml were fixed with buffered, sterile-filtered paraformaldehyde (Penta, Prague, Czechia) to a final concentration of 1%, and 0.5 ml was filtered onto white polycarbonate filters (pore size 0.2 μm, Nucleopore, Whatman, Maidstone, UK). Cells were stained with 4′,6-diamidino-2-phenylindole (DAPI) at concentration of 1 mg L^{-1} (Coleman, 1980). Total and AAP bacterial abundances were determined using an epifluorescence Zeiss Axio Imager.D2 microscope equipped with Collibri LED module illumination system (Carl Zeiss, Jena, Germany). Ten microphotographs were taken for every sample under 325–370 nm excitation and 420–470 nm emission wavelengths for DAPI fluorescence (total bacteria), 450–490 nm excitation and 600–660 nm emission wavelengths for autofluorescence from Chl-*a* (algae and cyanobacteria), and combined 325–370 nm, 450–490 nm, 545–565 nm and 615–635 nm excitation and 645–850 emission wavelengths for autofluorescence from BChl-*a* (AAP bacteria). As some part of Chl-*a* autofluorescence is also visible in the infrared spectrum, only the IR-positive cells that did not show any autofluorescence from Chl-*a* were counted as AAP bacteria (Cottrell et al., 2006).

Nutrients and chlorophyll-*a*

Samples were filtered through glass fibre filters with 0.4 μm nominal porosity (GF-5, Macherey-Nagel, Düren, Germany). Concentrations of soluble reactive phosphorus (SRP) were determined spectrophotometrically (Kopáček & Hejzlar, 1993; Murphy & Riley, 1962). Concentrations of nitrate and ammonium were measured according to Procházková (Procházková, 1959) and Kopáček and Procházková (Kopáček & Procházková, 1993). DOC and dissolved nitrogen (DN) were determined by catalytic thermal combustion at 720°C in combination with chemiluminescence detection by Shimadzu TOC-L equipped with TNM-L Total Nitrogen module (Shimadzu, Kyoto, Japan).

For Chl-*a* measurements, phytoplankton was collected by filtration onto GF-5 glass fibre filters (Macherey-Nagel). The filters were dried of excess water by gently pressing in a paper towel, and flush frozen in liquid nitrogen. Pigments were extracted in acetone-methanol (7:2, v:v) mixture and analysed by HPLC as described in (Piwosz et al., 2020).

DNA isolation

Water (between 250 and 700 ml) was filtered through sterile 0.2 μm Nucleopore Track-Etch Membrane filter units (Whatman, Maidstone, United Kingdom). Filters

were put inside sterile cryogenic vials, flash-frozen in liquid nitrogen and stored at −80°C. DNA extraction was done using PowerWater DNA Isolation Kit (MO BIO Laboratories Inc., Carlsbad, CA).

DNA samples were pooled in equimolar concentrations for the subsequent analysis. The epilimnion samples included water collected from 0.5 and 2 m, and the hypolimnion ones from the deeper layers (5–9 m). Such division was based on the temperature profiles.

Bacterial community analysis

Amplicons of 16S rRNA gene were prepared using the primers set 341F-785R (Klindworth et al., 2013). PCR was performed in 20 μl reaction, using Phusion High-Fidelity PCR MasterMix (Thermo Scientific, USA). Reaction conditions were as follow: 98°C for 3 min, 25 cycles at 98°C for 10 s, 60°C for 20 s, 72°C for 20 s and a final extension at 72°C for 3 min. The reactions for each sample were done in triplicates, pooled and purified from the gel using the Wizzard SV Gel and PCR clean system (Promega), and quantified with Qubit dsDNA HS assay. Amplicons were sequenced on Illumina MiSeq (2 × 250 bp) platform at Genomic Service of the Universitat Pompeu Fabra (Barcelona, Spain).

Obtained reads were quality checked using FastQC v0.11.7 (Babraham Bioinformatics, Cambridge, UK). The primer sequences were trimmed using Cutadapt v1.16 (Martin, 2011) and further analysis was done in the R/Bioconductor environment using the DADA2 package (version 1.12.1) (Callahan et al., 2016). Low quality reads were filtered out and cut (dada2::filterAndTrim, truncLen = c(225, 225), maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = TRUE). Sequences were merged (dada2::mergePairs()), and using removeBimeraDenovo function (method = "pooled"), chimeras were removed and singletons and doubletons were eliminated using Phyloseq (McMurdie & Holmes, 2013) (phyloseq::filter_taxa (ps, function(x) sum(x > 3) > (0.2*length(x)), TRUE)) resulting in 685 amplicon sequence variants (ASV) (Table S1). The final ASV table contained from 18,551 to 71,613 reads per sample (average ± SD: 42,945 ± 14,271). Taxonomic assignment was done in DADA2 (dada2:: assignTaxonomy()) using the SILVA r138.1 database released on 27 August 2020. Graphs were done using phyloseq (McMurdie & Holmes, 2013) and ggplot2 (Wickham, 2009) packages.

AAP bacteria community analysis

The composition of AAP bacteria community was analysed using the *puf*M gene. Amplicons for *puf*M gene were prepared using *puf*M UniF and *puf*M UniR primers (Yutin et al., 2005). The hypolimnion sample from 15/06 could not be amplified and was excluded in further analysis. The PCR was done in a triplicate of 20 μl reaction using Phusion High-Fidelity PCR MasterMix (Thermo Scientific, USA). The conditions were as follows: initial denaturation for 3 min at 98°C, 30 cycles of 98°C for 10 s, 52°C for 30 s, 72°C for 30 s, final elongation at 72°C for 5 min. The obtained triplicate reactions were pooled and amplicons were purified from the gel using the Wizzard SV Gel and PCR clean system (Promega) and quantified with Qubit dsDNA HS assay. Amplicons sequencing was performed on Illumina MiSeq (2 × 250bp) platform by Genomic Service of the Universitat Pompeu Fabra (Barcelona, Spain). Obtained *puf*M gene reads were processed in the same manner as 16S rRNA gene amplicons with different filterAndTrim function values (truncLen = c(130, 130), maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = TRUE, multithread = FALSE). Final ASV table contained 468 pufM_ASVs and from 15,746 to 84,506 reads per sample (41,524 ± 12,440; Table S2). Taxonomic assignment was done in DADA2 (function: assignTaxonomy[]) using manually curated in-house database. It contained 1475 unique *puf*M sequences, downloaded from the Fungene repository on May 16, 2019 (http://fungene.cme.msu.edu), from metagenomes from the Římov reservoir (Andrei et al., 2019; Mehrshad et al., 2018) and from the Genome Taxonomy database accessed on 16 September 2020 (Parks et al., 2018).

The sequences of 16S amplicons were deposited in the NCBI database under Biosamples SAMN26677261–SAMN26677286 and of *puf*M amplicons under Biosamples SAMN26677246–SAMN26677260 as a part of BioProject PRJNA816466.

Phylogenetic analysis of *puf*M gene

The limitation of the taxonomic assignation of the *pufM* gene amplicons was evident from the initial classification using the default DADA2 algorithm, which failed to assign all Chloroflexota reads and 20% of Proteobacteria reads at class level, and 65% of Alphaproteobacteria reads at order level. Over 50% of reads remain unclassified at the order level after the DADA2 taxonomic assignment. Phylogenetic analysis reduced the contribution of these unclassified Alphaproteobacteria up to 95% in June. Moreover, even though only 5% of Gammaproteobacteria reads remained unclassified at order level, over a quarter of the Burkholderiales reads could not be assigned to the genus level. The *puf*M gene is more variable than 16S rRNA gene and the lack of reference sequences closely related to these unclassified pufM_ASVs in our database hampered a more precise taxonomic assignment.

To partially overcome this limitation and reveal the hidden diversity of freshwater AAP bacteria in the studied lake, we performed a phylogenetic analysis of these unclassified pufM_ASVs.

Amino acid sequences of pufM genes were obtained from the Genome Taxonomy database through AnnoTree tool (Mendler et al., 2019) on 13/09/2021. The sequences with percentage of identity ≤40% to the reference pufM sequence in GTDB were removed as they originated from the *pufL* gene. 14 *puf*M sequences from previous works (Fecskeová et al., 2019; Mujakić et al., 2021; Piwosz et al., 2020; Zeng et al., 2021), which are not shown in GTDB database but they represent a reference point in freshwater environments, were included in the analysis. The unclassified pufM_ASVs were translated to amino acids using the second forward open reading frame. Both the reference *puf*M sequences and the pufM_ASVs were aligned in Geneious v2019.2.3 using ClustalW v2.1 (Larkin et al., 2007). The phylogenetic analysis was done with IQTREE (Trifinopoulos et al., 2016). LG + F + I + G4 was selected as a best-fit nucleotide substitution model by ModelFinder (Kalyaanamoorthy et al., 2017) according to Bayesian information criterion, and the Maximum Likelihood tree was calculated using 1000 ultrafast bootstrap replicates and default settings. The trees were prepared for the publication using iTOL (Letunic & Bork, 2021) and Inkscape v1.01.

Statistical analysis

The statistical relationship between environmental data (Table T3), abundance of all bacteria and abundance of AAP bacteria was analysed by distance based linear models (DistML), distance based Redundancy Analysis (Anderson & Legendre, 1999; Legendre & Andersson, 1999) and non-metric multi-dimensional scaling (nMDS) in the PERMANOVA+ add-on package of the PRIMER7 software (Anderson et al., 2008) (Primer Ltd., Lutton, UK). Environmental variables for each layer were averaged from the corresponding depths (0.5 and 2 m for the epilimnion and 5 to 9 m for the hypolimnion) and normalized. Correlations between different environmental variables were checked using Draftsman plots, and in the case of the strong correlation (absolute value of correlation coefficient >0.7) one of the variables was excluded from the analysis (Table T3). Abundance data for all and AAP bacteria (from microscopic counts), and the amplicon data (after removing singletons and doubletons) were log $(X + 1)$ transformed. The best model was selected using a stepwise selection procedure based on statistical significance (9999 permutations) and the value of the Akaike's Information Criterion (AICc) (Anderson & Legendre, 1999; Legendre & Andersson, 1999).

RESULTS AND DISCUSSION

Environmental variables

Samples were collected from a freshwater lake Cep, from April to September. The environmental conditions were typical for oligo-mesotrophic lakes (Figure S1) indicating that our results are representative for Northern hemisphere temperate lakes (Dodds & Whiles, 2020; Verpoorter et al., 2014). Thermal stratification of the water column was formed in May and present until the end of the sampling. Nevertheless, hypoxic or anoxic conditions were not observed in the hypolimnion (minimal oxygen concentration was 5.69 mg O_2 L^{-1}; Figure S1). Concentrations of nutrients and chlorophyll-*a* (Chl-*a*) were higher in the hypolimnion.

Abundances of total and AAP bacteria

The abundance and dynamics of all bacteria were similar in the epi- and the hypolimnion, with maximum values in July for the epilimnion (4.62×10^6 cells ml^{-1}) and in May for the hypolimnion (5.19×10^6 cells ml^{-1}) (Figure 1A). The abundance of AAP bacteria ranged from 7.11×10^4 to 2.39×10^5 cells ml^{-1} in the epilimnion and from 5.57×10^4 to 2.56×10^5 cells ml^{-1} in

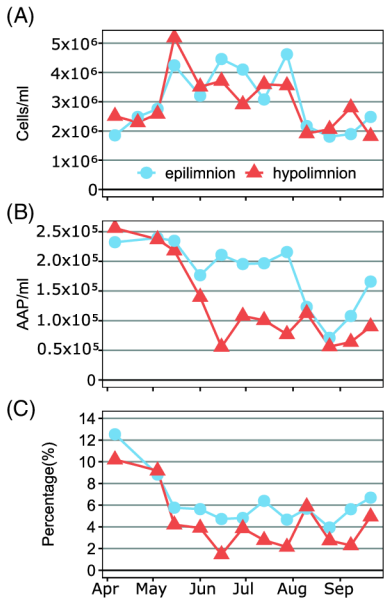

FIGURE 1 Weekly abundances of all prokaryotes (A), AAP bacteria (B) and relative abundance of AAP bacteria in the total bacterial community (C) in the epilimnion (blue-circles and line) and the hypolimnion (red-triangles and line) of Cep lake.

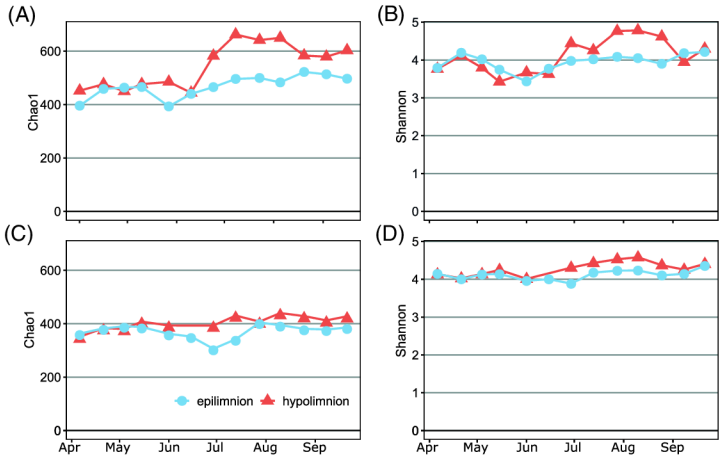

FIGURE 2 Weekly estimates of bacterial diversity in the epilimnion (blue-circles and line) and the hypolimnion (red-triangles and line) of Cep lake. (A) Chao1 index and (B) Shannon index estimated for total bacterial community based on 16S rRNA gene amplicon sequences; (C) Chao1 index and (D) Shannon index estimated for AAP bacterial community based on *puf*M gene amplicon sequences.

the hypolimnion (Figure 1B), with the highest values in April and May in both layers. AAP bacterial abundance decreased in the hypolimnion in June, while in the epilimnion only in August (Figure 1B). The percentage contribution of AAP bacteria to the total bacterial abundance was generally higher in the epilimnion (3.93–12.53%) than in the hypolimnion (1.45–10.20%, Figure 1C). Such temporal patterns of AAP bacteria abundance between in the epi- and the hypolimnion are common in temperate freshwater lakes (Čuperová et al., 2013; Fauteux et al., 2015).

Bacterial community composition

The bacterial alpha diversity was higher in the hypolimnion, especially in summer, based on 16S rRNA gene amplicons (16S_ASVs; Figure 2). The most abundant phyla were, in descending order, Actinobacteriota, Cyanobacteria, Proteobacteria, Bacteroidota, Verrucomicrobiota and Planctomycetota (Figure S2A). Non-metric multidimensional scaling (nMDS) analysis showed that the largest differences in community composition between the two layers were during stratification period from June to September (Figure S2). Since the composition and temporal changes of bacterial community were typical for the oligo-mesotrophic lakes (Cabello-Yeves et al., 2017; Morrison et al., 2017), below we have just focused on phyla known to contain AAP bacteria: Gemmatimonadota, Chloroflexota and Proteobacteria.

The relative abundance of Gemmatimonadota was <1%, which seems to be typical for freshwater lakes (Mujakić et al., 2021). Gemmatimonadota were represented by three 16S_ASVs (Figure 3A), classified as genus *Gemmatimonas* that is common in aquatic habitats (Gołębiewski et al., 2017; Mujakić et al., 2021). Their relative abundance varied substantially in both layers, with the changes in the epilimnion preceding those in the hypolimnion, suggesting a downward propagation of temporal changes (Cram et al., 2015).

Phylum Chloroflexota represented <4% of total 16S rRNA gene sequences (Figure S3A). Genera from class Chloroflexia known to contain *puf*M genes, such as *Roseiflexus*, *Kouleothrix* and UBA965 dominated during spring (Figure 3B). They were followed by the SL56 cluster (Ca. Limnocylindrus, Mehrshad et al., 2018) and they dominated in the epilimnion until the end of the sampling campaign. In the hypolimnion, a noticeable increase of the class Anaerolineae was observed in August and September (Figure 3B), when oxygen concentration was at its lowest. These bacteria are widely spread in different environments (Yamada & Sekiguchi, 2009) and were originally described to be anaerobic (Nakahara et al., 2019; Yamada et al., 2006) but recently an aerobic member containing type-II reaction centre (RC) was assembled from a metagenome (Martinez et al., 2020).

Proteobacteria were the most abundant phylum with known AAP species (Figure S3A). Alphaproteobacteria and Gammaproteobacteria were the dominant classes, with a contribution to the Proteobacteria reads ranging from 23–49% and from 51–77%, respectively.

Temporal patterns of Alphaproteobacteria were quite distinct between the epi- and the hypolimnion (Figure 3C). Caulobacterales, Rhizobiales, Rhodobacterales and Sphingomonadales, orders known to contain AAP species (Imhoff et al., 2019; Kopejtka

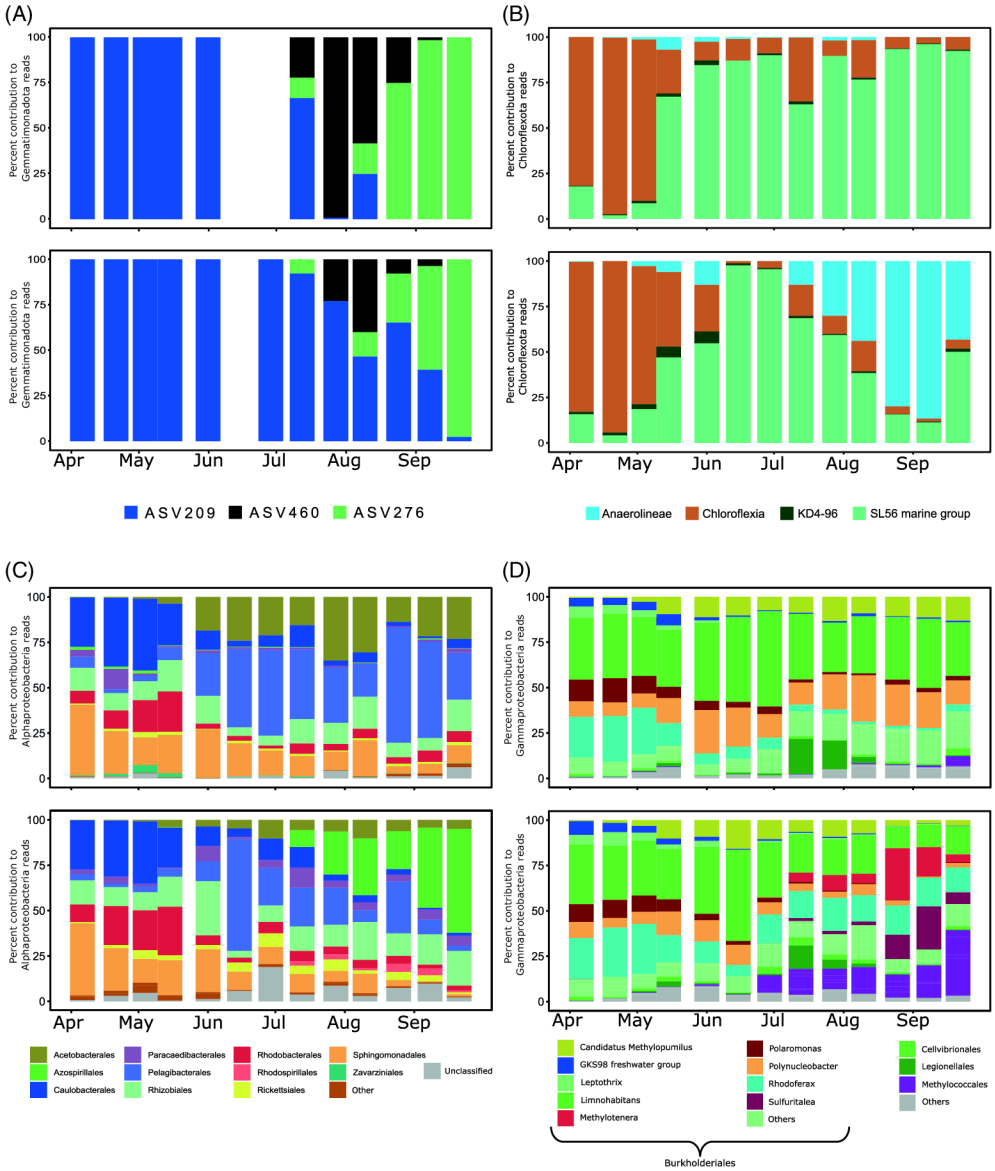

FIGURE 3 Bacterial community composition based on 16S rRNA gene amplicon sequencing in the epilimnion (upper plots in each panel) and the hypolimnion (bottom plots) in Cep lake. Only phyla and classes known to contain AAP species are shown. (A) Percent contribution of ASVs affiliated with Gemmatimonadota to the number of reads coming from Gemmatimonadota; (B) percent contribution of Chloroflexota of classes to the number of reads coming from Chloroflexota, (C) percent contribution of alphaproteobacterial orders to the number of reads coming from Alphaproteobacteria and (D) percent contribution of gammaproteobacterial orders to the number of reads coming from Gammaproteobacteria.

et al., 2017, 2021), dominated in spring, coinciding with the maxim values of AAP bacterial abundance. Pelagibacteriales dominated in the epilimnion and Azospirillales in the hypolimnion from June until the last sampling in September. They have already been reported as a part of microbial communities in a wide variety of freshwater environments (Galachyants et al., 2021; Tsementzi et al., 2019).

Gammaproteobacteria were dominated by the order Burkholderiales (Figure 3D), specifically by genera such as *Limnohabitans*, *Rhodoferax* and *Polynucleobacter*, important members of microbial food webs also known to contain AAP species (Hahn et al., 2012; Jezberová et al., 2017; Kasalický et al., 2018; Salka et al., 2011; Šimek et al., 2005). From the onset of stratification, the Gammaproteobacteria community started to develop differently in the depth layers. The relative abundance of Ca. Methylopumilus and *Polynucleobacter* increased in the epilimnion, while *Methylotenera*, *Sulfuritalea* and order Methylococcales increased in the hypolimnion.

AAP bacteria community composition

The AAP community composition was assessed with amplicon sequencing of the *puf*M gene, which is a routinely used marker for anoxygenic phototrophs containing type-II RC (Koblížek, 2015). The number of *puf*M ASVs (pufM_ASVs) estimated using the Chao1 index and the Shannon diversity index were higher in the hypolimnion than in the epilimnion (Figure 2C, D). Higher bacterial diversity in the hypolimnion seems to be a common feature of freshwater lakes (Shade et al., 2012), and here we demonstrated that this may be also a trend for AAP bacteria, at least in shallow lakes that are oxygenated and illuminated to the bottom. Moreover, AAP community started to diverge between the epi- and the hypolimnion in summer (Figure S2B), as observed also for all bacteria. Nevertheless, the difference in alpha and beta diversity between two layers was less conspicuous for AAP bacteria than for the total bacterial community (Figure 2 and Figure S1).

Gemmatimonadota, Chloroflexota and Proteobacteria were the only phyla detected in our samples, while none of the phototrophic members of Eremiobacteriota and Myxococcota were present in the *puf*M gene libraries (Figure S3B).

The contribution of Gemmatimonadota to the AAP bacteria community was in general below 2%, except in September, when it increased to 9% (Figure S3B). Gemmatimonadota community consisted of three pufM_ASVs (Figure 4A). pufM_ASV69, that clustered in the phylogenetic tree together with the environmental cluster Pg2 (Mujakić et al., 2021) (Figure S4), dominated in spring in the epilimnion and in spring and summer in the hypolimnion (Figure 4A). It was almost replaced by pufM_ASV30 and pufM_ASV209 (Figure S4) that, despite being closely related, showed a distinct seasonal pattern (Figure 4A). Such temporal separation of closely related phylotypes may have resulted from the differences in their physiology, as observed for *G. phototrophica* and *G. groenlandica* (Zeng et al., 2021). Interestingly, the dynamics of Gemmatimonadota by pufM_ASVs resembles that of 16S_ASVs (Figure 3A and Figure 4A). This suggests that pufM_ASV209, pufM_ASV30 and pufM_ASV69 from *puf*M amplicons might correspond to 16S_ASV460, 16S_ASV276 and 16S_ASV209 from 16S rRNA gene amplicons, respectively.

Chloroflexota contributed up to 5% to the *puf*M gene sequences and were more abundant in April and July–August (Figure S3B). They consisted of 5 pufM_ASVs from an uncultured freshwater clade of the Roseiflexaceae family (Mehrshad et al., 2018) that differed only by 1 or 2 amino acids (Figure S5). Despite this high similarity, these Chlorofleoxta ASVs showed distinct dynamics (Figure 4B). Interestingly, the most abundant pufM_ASV18 and pufM_ASV87 were shown to be highly active in August 2016 in the same lake, based on the comparison of DNA and RNA amplicon libraries (Fecskeová et al., 2019). Our results indicate that these Chloroflexota are core members of freshwater AAP bacteria communities over the whole sampling season.

Proteobacteria represented over 90% of all *puf*M reads (Figure S3B). This comprised between 2% and 14% of Alphaproteobacteria, while Gammaproteobacteria were always more than 50%, with a maximum >90% in April (Figure S2B).

Alphaproteobacteria community was similar in both layers in spring, being dominated by Rhodobacterales, Rhizobiales and Caulobacterales (Figure 4C), as observed also for 16S amplicons (Figure 4D). From the onset stratification, Rhodospirillales increased their contribution in the epilimnion, while Sphingomonadales and Rhizobiales showed their maximum contribution in the hypolimnion. Interestingly, the contribution of unclassified reads increased from July, and they reached up to 75% by late September (Figure 4C).

Phototrophic Gammaproteobacteria were more stable compared with other phototrophic phyla, and there was little difference between the epi- and the hypolimnion (Figure 4D). Burkholderiales made up >90% of reads in all samples. *Limnohabitans*, *Polynucleobacter* and *Rhodoferax* were the dominant genera, but 20–25% of the Burkholderiales remained unclassified at genus level. *Limnohabitans* and *Rhodoferax* showed the highest relative abundances in April until June, whereas *Polynucleobacter* showed this in August and September, reaching almost 30% of the Gammaproteobacteria reads in the epilimnion. Phototrophic activity of Burkholderiales may have been relatively low, as they were substantially underrepresented in the active AAP community investigated in August 2016 in the same lake (Fecskeová et al., 2019).

Most of the groups recovered in our time series were also reported in snapshots studies of freshwater lakes (Cepáková et al., 2016; Čuperová et al., 2013; Fecskeová et al., 2019). Here, we show that the composition of AAP community substantially varies in both epi- and hypolimnion, with many groups showing transient occurrence (Figure 4). This clearly indicated that the snapshot studies cannot sufficiently describe their diversity and the season

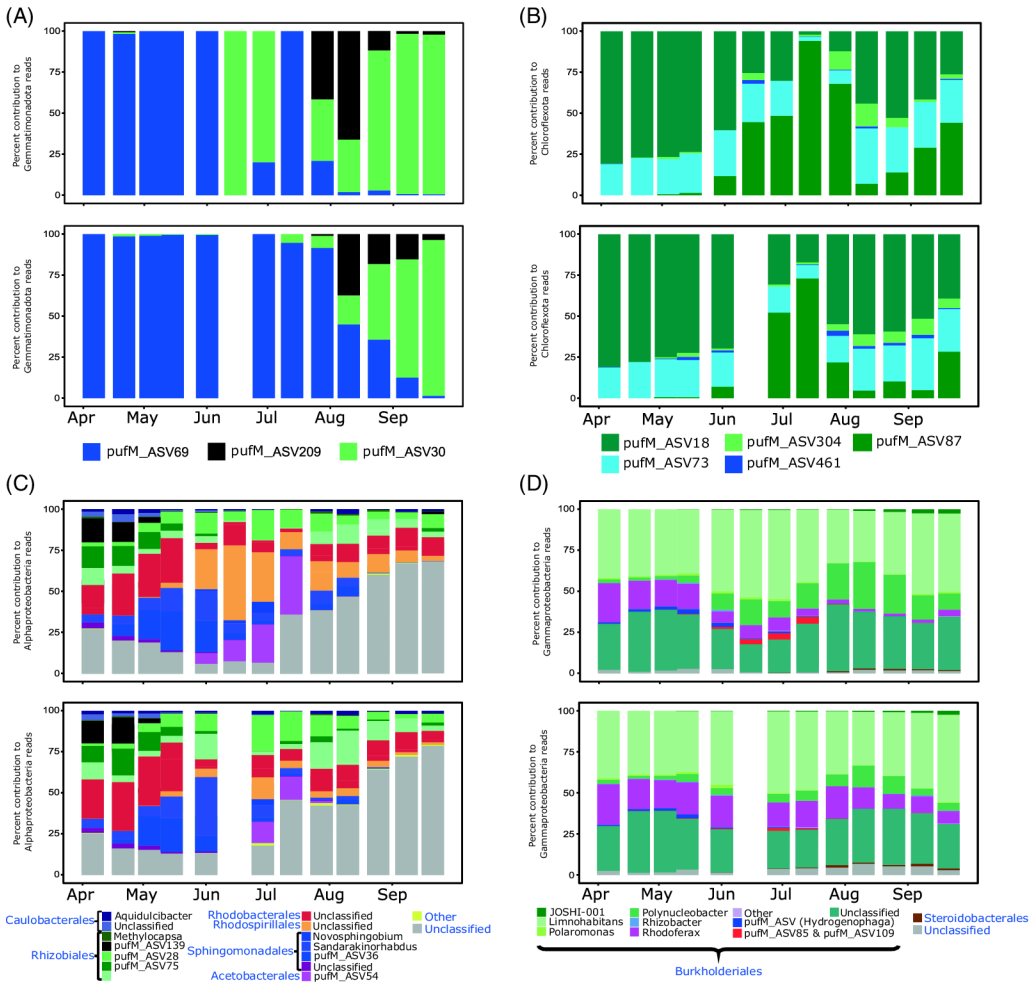

FIGURE 4 AAP bacteria community composition based on *puf*M gene amplicon sequencing in the epilimnion (upper plots in each panel) and the hypolimnion (bottom plots) in Cep lake. (A) Percent contribution of ASVs affiliated with Gemmatimonadota to the number of reads coming from Gemmatimonadota; (B) percent contribution of ASVs affiliated with Chloroflexota to the number of reads coming from Chloroflexota, (C) percent contribution of alphaproteobacterial genera and ASVs to the number of reads coming from Alphaproteobacteria and (D) percent contribution of gammaproteobacterial genera and ASVs to the number of reads coming from Gammaproteobacteria.

of sampling affects the results. As confirmation, a substantial proportion of the reads and pufM_ASVs in our study remained unassigned at genus or order level (Figure 4). This was not just the case for the understudied phyla, such as Chloroflexota and Gemmatimonadota, but also for Proteobacteria (Figure 4C, D), for which diversity in freshwater environments is well described (Ferrera et al., 2017). Furthermore, some pufM_ASVs could not be classified even to the class level. For instance, a group of 20 pufM_ASVs formed a branch between Rhodobacterales (Alphaproteobacteria) and Ectothiorhodospirales (Gammaproteobacteria) (Figure S6). The lack of reference sequences related to the unclassified pufM_ASVs hampered the taxonomic assignation of many of them. These results indicate the high potential of undescribed diversity of AAP bacteria in freshwaters that can be investigated for instance using metagenomics (Mehrshad et al., 2018; Mujakić et al., 2021; Ward et al., 2019).

Relationship with environmental variables

Temporal dynamics of AAP bacteria abundance has been shown to respond to changes in the

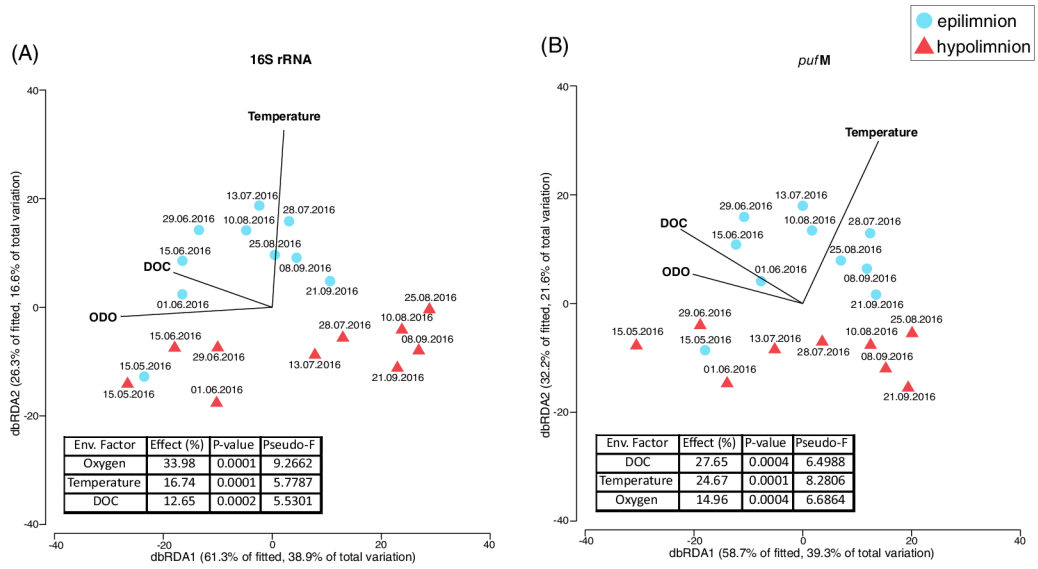

FIGURE 5 Distance-based redundancy analysis biplots relating the observed variability in the composition of communities for (A) all bacteria and for (B) AAP bacteria to the explanatory variables (black lines) in the epilimnion (blue-circles) and hypolimnion (red-triangles). Tables embedded in the graphs show the percentage of variability explained (effect [%]) by individual environmental variables (Env. Factor) and their statistical significance (*p* value and pseudo-F). DOC, dissolved organic carbon; ODO, optical dissolved oxygen.

environmental conditions (Kolářová et al., 2019; Lew et al., 2015; Mašín et al., 2008). However, how they influence the AAP community composition remains mostly unknown. Such information is important as, for example, high photoheterotrophic activity by AAP bacteria was linked to elevated relative abundance of few alphaproteobacterial orders: Caulobacteriales and Sphingomonadales (Piwosz et al., 2022). Thus, we conducted statistical analysis to reveal environmental factors that may have influenced the dynamics of AAP communities.

Distance-based linear models (DistLM) and distance-based redundancy analysis (dbRDA) (Anderson & Legendre, 1999; Legendre & Andersson, 1999) showed that the environmental factors that best explained the variability of the total bacterial community were oxygen, temperature and dissolved organic carbon (DOC) (Figure 5). The variability in the AAP community composition was explained by the same environmental variables but the explanatory power of particular variables differed. Namely, the best explanatory variable was DOC, followed by temperature, whereas oxygen, which explained most of the variability in the total bacterial community was less important for AAP bacteria. These environmental factors have been shown to be determinants in general for shaping bacterial and AAP communities (Ferrera et al., 2017; Mašín et al., 2008; Niño-García et al., 2016). Nevertheless, the observation that this functional group correlates differently with those variables than total bacteria indicates that their response to changing environmental conditions may differ as well.

SUMMARY AND CONCLUSIONS

In this report, we investigated temporal changes in the diversity of bacteria in a freshwater lake, with the focus on AAP bacteria. Our results showed that the AAP community followed different temporal patterns in the epilimnion and the hypolimnion when the lake was stratified. The high number of unclassified reads in the *puf*M libraries indicates that there is still hidden diversity of AAP bacteria within the known bacterial phyla. This was the case not only for less known Gemmatimonadota and Chloroflexota but, surprisingly, also for Alphaproteobacteria, the class to which the first described AAP species belongs (Harashima et al., 1978). Progress in linking this unexpected *puf*M diversity with taxonomic affiliation may be expected with increasing use of metagenomics, targeted single-cell sequencing and culturing of AAP bacteria (Parks et al., 2017; Woyke et al., 2017).

Environmental conditions shape bacterial communities, resulting in temporal changes in their diversity and composition, driven by adaptation that causes physiologic differences between distinct phylotypes. In

contrast to our expectations, the same set of environmental factors explained variability of all and AAP communities but they differed in their importance for total versus functional group. This indicates that AAP bacteria may respond differently to the changing environment, for example to temperature rise and deoxygenation of lakes due to global warming. More information on metabolic capacity of this photoheterotrophic bacteria is needed to fully understand their dynamics.

AUTHOR CONTRIBUTIONS

Cristian Villena-Alemany: analysis of the sequencing data, prepared figures and writing the manuscript; Izabela Mujakić: statistical analysis, prepared the amplicons, prepared a figure and commenting on the manuscript; Petr Porcal: measurement of environmental chemicals and commenting on the manuscript; Michal Koblížek: editing and commenting on the manuscript; Kasia Piwosz: participated in sampling, extracted DNA, supervised analysis of data, editing and commenting on the manuscript.

ACKNOWLEDGEMENTS

Authors thank Jason Dean and Martina Hanusová for help during the sampling campaign, and Alastair T. Gardiner for commenting on the manuscript. This study was supported by project PhotoGemm+ n° 19-28778X from the Czech Science Foundation awarded to Michal Koblížek.

ORCID

Cristian Villena-Alemany https://orcid.org/0000-0002-6158-1879
Izabela Mujakić https://orcid.org/0000-0001-5602-7331
Michal Koblížek https://orcid.org/0000-0001-6938-2340
Kasia Piwosz https://orcid.org/0000-0002-3248-3364

REFERENCES

Anderson, M.J., Gorley, R.N. & Clarke, K.R. (2008) PERMANOVA+ for PRIMER: guide to software and statistical methods. In: *Plymouth*. UK, PRIMER-E. pp. 1–214.

Anderson, M.J. & Legendre, P. (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62, 271–303.

Andrei, A.-Ş., Salcher, M.M., Mehrshad, M., Rychtecký, P., Znachor, P. & Ghai, R. (2019) Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. *The ISME Journal*, 13, 1056–1071.

Cabello-Yeves, P.J., Haro-Moreno, J.M., Martin-Cuadrado, A.-B., Ghai, R., Picazo, A., Camacho, A. et al. (2017) Novel Synechococcus genomes reconstructed from freshwater reservoirs. *Frontiers in Microbiology*, 8, 1151.

Caliz, J. & Casamayor, E.O. (2014) Environmental controls and composition of anoxygenic photoheterotrophs in ultraoligotrophic high-altitude lakes (Central Pyrenees). *Environmental Microbiology Reports*, 6, 145–151.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A. J.A. & Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583.

Cepáková, Z., Hrouzek, P., Žišková, E., Nuyanzina-Boldareva, E., Šorf, M., Kozlíková-Zapomělová, E. et al. (2016) High turnover rates of aerobic anoxygenic phototrophs in European freshwater lakes. *Environmental Microbiology*, 18, 5063–5071.

Coleman, A.W. (1980) Enhanced detection of bacteria in natural environments by fluorochrome staining of DNA1. *Limnology and Oceanography*, 25, 948–951.

Cottrell, M.T., Mannino, A. & Kirchman, D.L. (2006) Aerobic Anoxygenic phototrophic bacteria in the mid-Atlantic bight and the North Pacific gyre. *Applied and Environmental Microbiology*, 72, 557–564.

Cram, J.A., Xia, L.C., Needham, D.M., Sachdeva, R., Sun, F. & Fuhrman, J.A. (2015) Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. *The ISME Journal*, 9, 2573–2586.

Čuperová, Z., Holzer, E., Salka, I., Sommarug, R. & Koblížek, M. (2013) Temporal changes and altitudinal distribution of aerobic anoxygenic phototrophs in mountain lakes. *Applied and Environmental Microbiology*, 79, 6439–6446.

Dodds, W.K. & Whiles, M.R. (2020) Freshwater ecology: concepts and environmental applications of limnology. In: Chapter 9 - *Microbes and plants*, Third edition. Academic Press, Elsevier, London, UK, pp. 211–249.

Fauteux, L., Cottrell, M.T., Kirchman, D.L., Borrego, C.M., Garcia-Chaves, M.C. & del Giorgio, P.A. (2015) Patterns in abundance, cell size and pigment content of aerobic Anoxygenic phototrophic bacteria along environmental gradients in Northern Lakes. *PLoS One*, 10, 1–17.

Fecskeová, L., Piwosz, K., Hanusová, M., Nedoma, J., Znachor, P. & Koblížek, M. (2019) Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. *Scientific Reports*, 9, 18766.

Ferrera, I., Sarmento, H., Priscu, J.C., Chiuchiolo, A., González, J. M. & Grossart, H.-P. (2017) Diversity and distribution of freshwater aerobic Anoxygenic phototrophic bacteria across a wide latitudinal gradient. *Frontiers in Microbiology*, 8, 175.

Galachyants, A.D., Krasnopeev, A.Y., Podlesnaya, G.V., Potapov, S. A., Sukhanova, E.V., Tikhonova, I.V. et al. (2021) Diversity of aerobic Anoxygenic phototrophs and rhodopsin-containing bacteria in the surface microlayer, water column and Epilithic biofilms of Lake Baikal. *Microorganisms*, 9, 842.

Garcia-Chaves, M.C., Cottrell, M.T., Kirchman, D.L., Ruiz-González, C. & Del Giorgio, P.A. (2016) Single-cell activity of freshwater aerobic anoxygenic phototrophic bacteria and their contribution to biomass production. *The ISME Journal*, 10, 1579–1588.

Gołębiewski, M., Całkiewicz, J., Creer, S. & Piwosz, K. (2017) Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria: the case study of the Vistula river estuary. *Environmental Microbiology Reports*, 9, 129–143.

Hahn, M.W., Scheuerl, T., Jezberová, J., Koll, U., Jezbera, J., Šimek, K. et al. (2012) The passive yet successful way of planktonic life: genomic and experimental analysis of the ecology of a free-living Polynucleobacter population. *PLoS One*, 7, e32772.

Harashima, K., Shiba, T., Totsuka, T., Simidu, U. & Taga, N. (1978) Occurrence of bacteriochlorophyll a in a strain of an aerobic heterotrophic bacterium. *Agricultural and Biological Chemistry*, 42, 1627–1628.

Hauruseu, D. & Koblížek, M. (2012) Influence of light on carbon utilization in aerobic Anoxygenic phototrophs. *Applied and Environmental Microbiology*, 78, 7414–7419.

Imhoff, J.F., Rahn, T., Künzel, S. & Neulinger, S.C. (2019) Phylogeny of Anoxygenic photosynthesis based on sequences of photosynthetic reaction center proteins and a key enzyme in

bacteriochlorophyll biosynthesis, the Chlorophyllide reductase. *Microorganisms*, 7, 576.

Jezberová, J., Jezbera, J., Znachor, P., Nedoma, J., Kasalický, V. & Šimek, K. (2017) The Limnohabitans genus harbors Generalistic and opportunistic subtypes: evidence from spatiotemporal succession in a canyon-shaped reservoir. *Applied and Environmental Microbiology*, 83, e01530–e01517.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Kasalický, V., Zeng, Y., Piwosz, K., Šimek, K., Kratochvilová, H. & Koblížek, M. (2018) Aerobic anoxygenic photosynthesis is commonly present within the genus Limnohabitans. *Applied and Environmental Microbiology*, 84, 6–17.

Kirchman, D.L. & Hanson, T.E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environmental Microbiology Reports*, 5, 188–199.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. et al. (2013) Evaluation of general 16 S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41, e1.

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews*, 39, 854–870.

Koblížek, M., Stoń-Egiert, J., Sagan, S. & Kolber, Z.S. (2005) Diel changes in bacteriochlorophyll a concentration suggest rapid bacterioplankton cycling in the Baltic Sea. *FEMS Microbiology Ecology*, 51, 353–361.

Kolářová, E., Medová, H., Piwosz, K. & Koblížek, M. (2019) Seasonal dynamics of aerobic anoxygenic phototrophs in freshwater lake Vlkov. *Folia Microbiologia (Praha)*, 64, 705–710.

Kopáček, J. & Hejzlar, J. (1993) Semi-micro determination of Total phosphorus in fresh waters with Perchloric acid digestion. *International Journal of Environmental Analytical Chemistry*, 53, 173–183.

Kopáček, J. & Procházková, L. (1993) Semi-micro determination of ammonia in water by the rubazoic acid method. *International Journal of Environmental Analytical Chemistry*, 53, 243–248.

Kopejtka, K., Tomasch, J., Zeng, Y., Tichý, M., Sorokin, D.Y. & Koblížek, M. (2017) Genomic analysis of the evolution of Phototrophy among Haloalkaliphilic Rhodobacterales. *Genome Biology and Evolution*, 9, 1950–1962.

Kopejtka, K., Zeng, Y., Kaftan, D., Selyanin, V., Gardian, Z., Tomasch, J. et al. (2021) Characterization of the aerobic Anoxygenic phototrophic bacterium Sphingomonas sp. AAP5. *Microorganisms*, 9, 768.

Lanyi, J.K. (2004) Bacteriorhodopsin. *Annual Review of Physiology*, 66, 665–688.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947–2948.

Legendre, P. & Andersson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69, 512–524.

Letunic, I. & Bork, P. (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49, W293–W296.

Lew, S., Koblížek, M., Lew, M., Medová, H., Glińska-Lewczuk, K. & Owsianny, P.M. (2015) Seasonal changes of microbial communities in two shallow peat bog lakes. *Folia Microbiologia (Praha)*, 60, 165–175.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 10.

Martinez, J.N., Kawai, S., Saini, M.K., Tank, M., Hanada, S. & Thiel, V. (2020) Draft genome sequence of a filamentous Anoxygenic phototrophic bacterium, "Candidatus Roseilinea sp. strain NK_OTU-006," recovered from metagenomic data of a hot spring microbial mat. *Microbiology Resource Announcements*, 9, e01104–e01120.

Martinez-Garcia, M., Swan, B.K., Poulton, N.J., Gomez, M.L., Masland, D., Sieracki, M.E. et al. (2012) High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *The ISME Journal*, 6, 113–123.

Mašín, M., Čuperová, Z., Hojerová, E., Salka, I., Grossart, H. & Koblížek, M. (2012) Distribution of aerobic anoxygenic phototrophic bacteria in glacial lakes of northern Europe. *Aquatic Microbial Ecology*, 66, 77–86.

Mašín, M., Nedoma, J., Pechar, L. & Koblížek, M. (2008) Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environmental Microbiology*, 10, 1988–1996.

McMurdie, P.J. & Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8, e61217.

Mehrshad, M., Salcher, M.M., Okazaki, Y., Nakano, S.I., Šimek, K., Andrei, A.S. et al. (2018) Hidden in plain sight - highly abundant and diverse planktonic freshwater Chloroflexi 06 biological sciences 0602 ecology 05 environmental sciences 0502 environmental science and management. *Microbiome*, 6, 176.

Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A. & Doxey, A.C. (2019) AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Research*, 47, 4442–4448.

Morrison, J.M., Baker, K.D., Zamor, R.M., Nikolai, S., Elshahed, M. S. & Youssef, N.H. (2017) Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater Lake (grand lake, OK, USA). *PLoS One*, 12, e0177488.

Mujakić, I., Andrei, A.-Ş., Shabarova, T., Fecskeová, L.K., Salcher, M. M., Piwosz, K. et al. (2021) Common presence of phototrophic Gemmatimonadota in temperate Freshwater Lakes. *mSystems*, 6, 1241–1220.

Murphy, J. & Riley, J.P. (1962) A modified single solution method for the determination of phosphate in natural waters. *Analytica Chimica Acta*, 27, 31–36.

Nakahara, N., Nobu, M.K., Takaki, Y., Miyazaki, M., Tasumi, E., Sakai, S. et al. (2019) Aggregatilinea lenta gen. Nov., sp. nov., a slow-growing, facultatively anaerobic bacterium isolated from subseafloor sediment, and proposal of the new order aggregatilineales Ord. Nov. within the class anaerolineae of the phylum chloroflexi. *International Journal of Systematic and Evolutionary Microbiology*, 69, 1185–1194.

Niño-García, J.P., Ruiz-González, C. & Del Giorgio, P.A. (2016) Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *The ISME Journal*, 10, 1755–1766.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A. et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, 36, 996–1004.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N. et al. (2017) Recovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2, 1533–1542.

Piwosz, K., Kaftan, D., Dean, J., Šetlík, J. & Koblížek, M. (2018) Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. *Environmental Microbiology*, 20, 724–733.

Piwosz, K., Villena-Alemany, C. & Mujakić, I. (2022) Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *The ISME Journal*, 16, 1046–1054.

Piwosz, K., Vrdoljak Tomaš, A., Thijs, F., Gonzalez Olalla, J.M., Šantić, D., McKay, L. et al. (2020) Light and primary production shape bacterial activity and community composition of aerobic

Anoxygenic phototrophic bacteria in a microcosm experiment. *mSphere*, 5, 1–17.
Procházková, L. (1959) Bestimmung der Nitrate im Wasser. *Fresenius' Zeitschrift für Analytische Chemie*, 167, 254–260.
Ruiz-González, C., Garcia-Chaves, M.C., Ferrera, I., Niño-García, J. P. & Giorgio, P.A. (2020) Taxonomic differences shape the responses of freshwater aerobic anoxygenic phototrophic bacterial communities to light and predation. *Molecular Ecology*, 29, 1267–1283.
Ruiz-González, C., Proia, L., Ferrera, I., Gasol, J.M. & Sabater, S. (2013) Effects of large river dam regulation on bacterioplankton community structure. *FEMS Microbiology Ecology*, 84, 316–331.
Salka, I., Čuperová, Z., Mašín, M., Koblížek, M. & Grossart, H.-P. (2011) Rhodoferax-related pufM gene cluster dominates the aerobic anoxygenic phototrophic communities in German freshwater lakes. *Environmental Microbiology*, 13, 2865–2875.
Shade, A., Read, J.S., Youngblut, N.D., Fierer, N., Knight, R., Kratz, T.K. et al. (2012) Lake microbial communities are resilient after a whole-ecosystem disturbance. *The ISME Journal*, 6, 2153–2167.
Šimek, K., Horňák, K., Jezbera, J., MaŠín, M., Nedoma, J., Gasol, J. M. et al. (2005) Influence of top-down and bottom-up manipulations on the R-BT065 subcluster of β-Proteobacteria, an abundant Group in Bacterioplankton of a freshwater reservoir. *Applied and Environmental Microbiology*, 71, 2381–2390.
Spudich, J.L. & Luecke, H. (2002) Sensory rhodopsin II: functional insights from structure. *Current Opinion in Structural Biology*, 12, 540–546.
Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B.Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44, W232–W235.
Tsementzi, D., Rodriguez-R, L.M., Ruiz-Perez, C.A., Meziti, A., Hatt, J.K. & Konstantinidis, K.T. (2019) Ecogenomic characterization of widespread, closely-related SAR11 clades of the freshwater genus "Candidatus Fonsibacter" and proposal of Ca. Fonsibacter lacus sp. nov. *Systematic and Applied Microbiology*, 42, 495–505.
Verpoorter, C., Kutser, T., Seekell, D.A. & Tranvik, L.J. (2014) A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41, 6396–6402.
Ward, L.M., Cardona, T. & Holland-Moritz, H. (2019) Evolutionary implications of anoxygenic phototrophy in the bacterial phylum Candidatus Eremiobacterota (WPS-2). *Frontiers in Microbiology*, 10, 1–12.
Wickham, H. (2009) ggplot2: elegant graphics for data analysis.
Woyke, T., Doud, D.F.R. & Schulz, F. (2017) The trajectory of microbial single-cell sequencing. *Nature Methods*, 14, 1045–1054.
Yamada, T. & Sekiguchi, Y. (2009) Cultivation of uncultured Chloroflexi subphyla: significance and ecophysiology of formerly uncultured Chloroflexi "subphylum i" with natural and biotechnological relevance. *Microbes and Environments*, 24, 205–216.
Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H. et al. (2006) Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. Nov., sp. nov. and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the. *International Journal of Systematic and Evolutionary Microbiology*, 56, 1331–1340.
Yurkov, V. and Csotonyi, J.T. (2009) New light on aerobic Anoxygenic phototrophs. In *Advances in photosynthesis and respiration*. Daldal F, Thurnauer MC, Beatty J. (ed). Springer, Dordrecht, The Netherlands, pp. 31–55.
Yutin, N., Suzuki, M.T. & Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Applied and Environmental Microbiology*, 71, 8958–8962.
Zemskaya, T.I., Cabello-Yeves, P.J., Pavlova, O.N. & Rodriguez-Valera, F. (2020) Microorganisms of Lake Baikal—the deepest and most ancient lake on earth. *Applied Microbiology and Biotechnology*, 104, 6079–6090.
Zeng, Y., Nupur, W., N., Madsen, A.M., Chen, X., Gardiner, A.T. et al. (2021) Gemmatimonas groenlandica sp. nov. is an aerobic Anoxygenic phototroph in the phylum Gemmatimonadetes. *Frontiers in Microbiology*, 11, 1–18.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Villena-Alemany, C., Mujakić, I., Porcal, P., Koblížek, M. & Piwosz, K. (2022) Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environmental Microbiology Reports*, 1–12. Available from: https://doi.org/10.1111/1758-2229.13131

Paper IV

Multi-environment ecogenomics analysis of the cosmopolitan phylum Gemmatimonadota

Mujakić, I., Cabello-Yeves, P.J., Villena-Alemany, C., Piwosz, K., Rodriguez-Valera, F., Picazo, A., Camacho, A., Koblížek, M.

(2023)
MS under revision in the Microbiology Spectrum

3.4. Summary results derived from Paper No. 4

In the fourth chapter, we analyzed more than 400 metagenome-assembled genomes (MAGs) and 5 cultured species from various environments in order to obtain information about genomic differences and metabolic potential of Gemmatimonadota across multiple environments. Additionally, we assembled 16 Gemmatimonadota MAGs from four freshwater reservoirs in Spain to further expand the collection of freshwater Gemmatimonadota.

We classified Gemmatimonadota MAGs based on their environmental origin into 12 different categories. Genomes from different environments varied in their sizes and number of coding sequences (CDS). The smallest genomes with the smallest number of genes were those from marine waters, while genomes from potentially more nutrient-rich environments, such as soil, soda lake sediments or wastewater, had larger sizes and higher numbers of CDS. The GC content of Gemmatimonadota genomes also varied and was again the lowest in marine genomes and some genomes from hydrothermal vents, while genomes from soil, permafrost, or wastewater had higher GC content.

Principal coordinate analysis based on the presence/absence of genes showed grouping of the Gemmatimonadota genomes based on their environmental origin. Similar patterns were also observed in phylogenomic analysis, indicating their adaptation to different environments. Further, we analyzed the main metabolic commonalities and differences associated with each environment, with a focus on some of the key metabolic pathways like carbon, nitrogen, and sulfur cycles, and phototrophy. Gemmatimonadota from all environments contained genes for central carbohydrate metabolism and were generally organoheterotrophic. However, some metabolic pathways or genes were present in Gemmatimonadota from specific environments. For example, genes for anoxygenic phototrophy were found in freshwater, wastewater and soda lake sediment genomes, and were not present in soil or marine genomes. Gemmatimonadota from several environments contained the gene encoding the large subunit of RuBisCO. Furthermore, Gemmatimonadota from soil, host-associated (marine sponges and corals), wastewater, hydrothermal vent, and groundwater could potentially use CO as an alternative energy source since they contained large subunit of carbon monoxide dehydrogenase (*coxL*). Gemmatimonadota did not have complete pathways for denitrification or nitrogen fixation. However, in all environments except marine water and host-associated they had *nosZ* gene indicating the potential for reduction of N_2O, which is considered one of the strongest greenhouse-gases. Furthermore, genes for flagella assembly were present mostly in wastewater and freshwater genomes, while the presence of phosphate transport system and different genes important during phosphorus limitations varied in different environments but were in some form found in all Gemmatimonadota.

Multi-environment ecogenomics analysis of the cosmopolitan phylum Gemmatimonadota

Izabela Mujakić[1,2]#, Pedro J. Cabello-Yeves[3,4,5], Cristian Villena-Alemany[1,2], Kasia Piwosz[6], Francisco Rodriguez-Valera[4], Antonio Picazo[3], Antonio Camacho[3], Michal Koblížek[1,2]#

[1]Lab. of Anoxygenic Phototrophs, Institute of Microbiology of the Czech Academy of Sciences, 37981 Třeboň, Czechia

[2]Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czechia

[3]Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, E- 46980, Paterna, Valencia, Spain

[4]Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante, Spain

[5]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

[6]Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute, ul. Kołłątaja 1, 81-332 Gdynia, Poland

Address correspondence to: Michal Koblížek: koblizek@alga.cz; Izabela Mujakić: mujakic@alga.cz

Running title: Potential metabolic roles of Gemmatimonadota from different environments

Abstract

Gemmatimonadota is a diverse bacterial phylum commonly found in environments such as soils, rhizospheres, fresh waters, and sediments. So far, the phylum contains just six cultured species (5 of them sequenced), which limits our understanding of their diversity and metabolism. Therefore, we analyzed over 400 metagenome-assembled genomes (MAGs) and 5 culture-derived genomes representing Gemmatimonadota from various aquatic environments, hydrothermal vents, sediments, soils, and host-associated (with marine sponges and coral) species. The principal coordinate analysis based on the presence/absence of genes in Gemmatimonadota genomes and phylogenomic analysis documented that marine and host-associated Gemmatimonadota were the most distant from freshwater and wastewater species. A smaller genome size and CDS number reduction were observed in marine MAGs, pointing to an oligotrophic environment adaptation. Several metabolic pathways were restricted to specific environments. For example, genes for anoxygenic phototrophy were found only in freshwater, wastewater, and soda lake sediments genomes. There were several genomes from soda lake sediments and wastewater containing type IC/ID RuBisCO. Various genomes from wastewater harbored bacterial type II RuBisCO, whereas RuBisCO-like protein was found in genomes from fresh waters, soil, host-associated and marine sediments. Gemmatimonadota does not contain complete nitrogen fixation genes, however, the *nosZ* gene, involved in the reduction of N_2O, was present in genomes from most environments, missing only in marine waters and host-associated Gemmatimonadota. The presented data suggest that Gemmatimonadota evolved as an organotrophic species relying on aerobic respiration, and then remodeled its genome inventory when adapting to particular environments.

Importance

Gemmatimonadota is a rarely studied bacterial phylum consisting of a handful of cultured species. Recent culture-independent studies documented that these organisms are distributed in many environments, including soil, marine, fresh and waste waters. However, due to the lack of cultured species, information about their metabolic potential and environmental role is scarce. Therefore, we collected available Gemmatimonadota metagenome-assembled genomes (MAGs) from different habitats and performed a systematic analysis of their genomic characteristics and metabolic potential. Our results show how Gemmatimonadota have adapted their genomes to the different environments.

Keywords: *Gemmatimonadota, Gemmatimonadetes, anoxygenic phototrophs, MAGs, metagenome, RuBisCO*

1. Introduction

The bacterial phylum Gemmatimonadota was established in 2003 when the type species, *Gemmatimonas aurantiaca*, was isolated from a wastewater treatment plant (1). Since then, only five more species have been described. '*Gemmatirosa kalamazoonesis*', *Roseisolibacter agri*, and *Longimicrobium terrae* were isolated from various soils (2–4), while *Gemmatimonas phototrophica* and *Gemmatimonas groenlandica* originated from fresh waters (5, 6). Due to the low number of cultured species, our understanding of the metabolic properties of Gemmatimonadota is very limited. All isolates grow on liquid organic carbon media under aerobic or semi-aerobic conditions (7, 8). In addition, the cultured freshwater species are facultative photoheterotrophs. They perform anoxygenic phototrophy and can supplement their metabolism with light energy harvested using bacteriochlorophyll (BChl)-*a* containing photosystems, however, they require a supply of organic substrate for growth (5, 9, 10). Photoheterotrophic Gemmatimonadota, similar to Proteobacteria, have their photosynthesis genes organized in the photosynthesis gene cluster (PGC) containing *bch* and *crt* genes encoding enzymes of bacteriochlorophyll and carotenoid synthesis, *puf* and *puh* operons encoding the subunits of reaction center and light-harvesting complexes and various regulatory genes (6, 9, 11, 12).

Metagenomic analyses have documented that Gemmatimonadota is present in a wide range of environments (13, 14). They are one of the most abundant phyla in soils, representing on average 2% of 16S rRNA gene sequences (13, 15, 16), and are relatively common in fresh waters, where they typically constitute 1 % of bacteria but may contribute even up to 9% of the bacterial community (7, 12, 17, 18). Gemmatimonadota were found in soda lake sediments, where they represented ≥1% of 16S rRNA gene sequences (19). Only minimum numbers have been registered in the marine water column (20), and typically in marine environments they are found associated with sponges (21, 22), deep-sea hydrothermal vents (23, 24) or sediments (25, 26), where they represented up to 2.4 % of the total bacterial 16S rRNA reads (27).

Previously, we documented a high diversity of photoheterotrophic Gemmatimonadota in freshwater lakes (12). Interestingly, MAGs containing genes both for anoxygenic photosynthesis and carbon fixation were identified in soda lakes sediments (28, 29). However, there is only limited information about Gemmatimonadota inhabiting other environments such as soils or marine waters. Therefore, we analyzed all publicly available MAGs (up until the 3rd of May 2021) affiliated with Gemmatimonadota to get a global picture of their metabolic functions, patterns, and genomic differences across multiple environments. In addition, we assembled 16 MAGs from four Spanish freshwater reservoirs (Table S1, S2). We focused on key metabolic pathways, such as carbon assimilation, nitrogen and sulfur cycles, and photoheterotrophic capability, to define the

potential roles of Gemmatimonadota in nutrient cycling, and to decipher the specific differences in their physiology based on the environment from which they originate.

2. Results and Discussion

2.1 Basic characteristics of the Gemmatimonadota genomes

Gemmatimonadota MAGs were classified based on their environmental origin in twelve different categories (Figure 1A, 1B). The numbers of dereplicated genomes within each category were as follows: fresh waters 91, soil 90, wastewater 49, soda lake sediments 46, marine water 42, host-associated (i.e associated with marine sponges and coral) 25, permafrost 22, marine sediments 12, hydrothermal vent 18, groundwater 21, other sediments 13. The final category "Other" consisted of 13 genomes from varying environments and was not included in most analyses (unless stated otherwise). Genomes from all environments varied largely in size, with an average of 3.59 Mbp (1.68 - 7.77 Mbp) and 2744 CDS (coding sequences) (Figure S1A, S1B). The smallest genomes (1.68 - 3.01 Mbp) with the lowest number of genes and higher homogeneity were those from marine waters. Genomes from potentially more nutrient-rich environments, like soil, soda lake sediments, marine sediments, and wastewater, had larger sizes as well as a higher number of CDS (Figure S2A). This is consistent with previous studies documenting that nutrient limitation affects genome size, GC content or coding density (30–32). The average coding density was 92.9 % (84 - 97%) and, despite the high variability, it was on average higher in MAGs from fresh waters and soils than those from marine waters, marine sediments, or wastewaters (Figure S1C). The average median intergenic distance was 35.45 bp. Even though marine genomes are in general smaller, they have on average longer intergenic spacers than freshwater, soil, or wastewater genomes (Figure S1D, Figure S2B). Genomes from soda lake sediments and marine sediments have both larger sizes and longer median intergenic spacers. Lengths of intergenic spacers vary substantially among bacteria (32) and often contain regulatory elements with key functions (33).

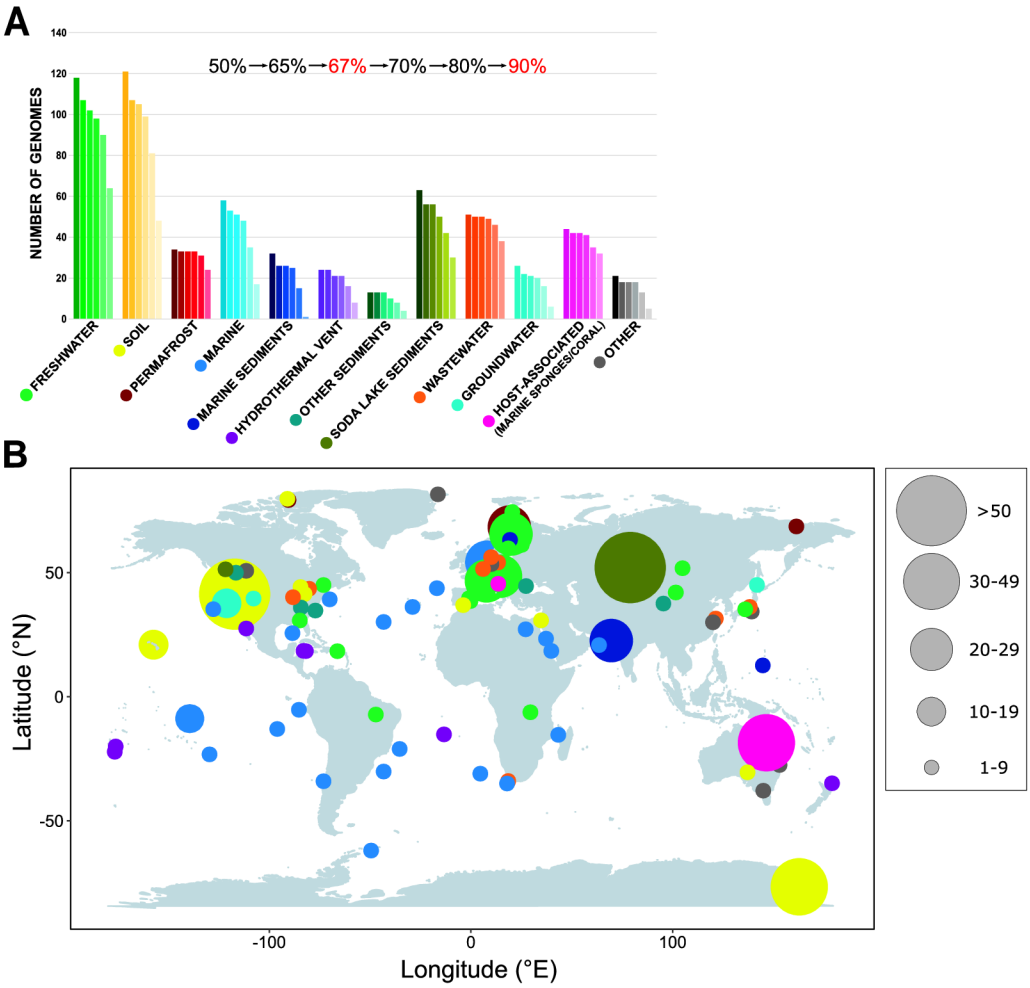

Figure 1. Distribution of Gemmatimonadota in different environments. **(A)** Bar plot showing numbers of Gemmatimonadota genomes present in NCBI database, including newly assembled freshwater MAGs, divided based on the environment of origin and completeness. Each bar represents the number of genomes with different completeness's, from MAGs with more than >50% up to >90% of completeness. The completeness levels used in subsequent analyses are marked red. **(B)** Map showing where the MAGs used in our analyses originated from. Environments are color coded, and size of the circle represents number of MAGs obtained from the specific location.

The GC content in the studied genomes ranged from 44.4 – 74.4 %, with an average of 65.6 % (Figure 2A, 2B). Marine genomes and several others from hydrothermal vents had the lowest GC content (range 44.4 – 62.7 %, and 45.5 – 69.8 %, respectively). GC content of bacterial communities is known to be influenced by the environment (34) and low GC content among marine bacteria is a common phenomenon (35) interpreted as an adaptation to low nitrogen (36) or a result of evolutionary history (37). It must be noted that the distribution of the GC content in marine genomes was trimodal (44 – 45 %, 49 –

54 % and 59 – 62.7 %), indicating an additional sub-environmental division of genomes from the same origin, possibly depending on parameters such as the water depth or water nutrient concentration. However, the associated metadata in the NCBI did not contain enough details to fully explain this pattern. The influence of environment on the GC content of bacterial communities can be observed even in closely related species which in different environments show significant differences in GC content (34). Freshwater genomes were also smaller but had higher GC content than marine genomes (Figure 2A). Gemmatimonadota genomes from other environments like soil, permafrost or wastewater varied in genome sizes and had on average a higher GC content than marine MAGs, a trait common for bacteria living in more nutrient-rich environments (38). This, combined with their larger genomes and higher number of CDS indicates their higher metabolic potential and advantages under different environmental conditions.

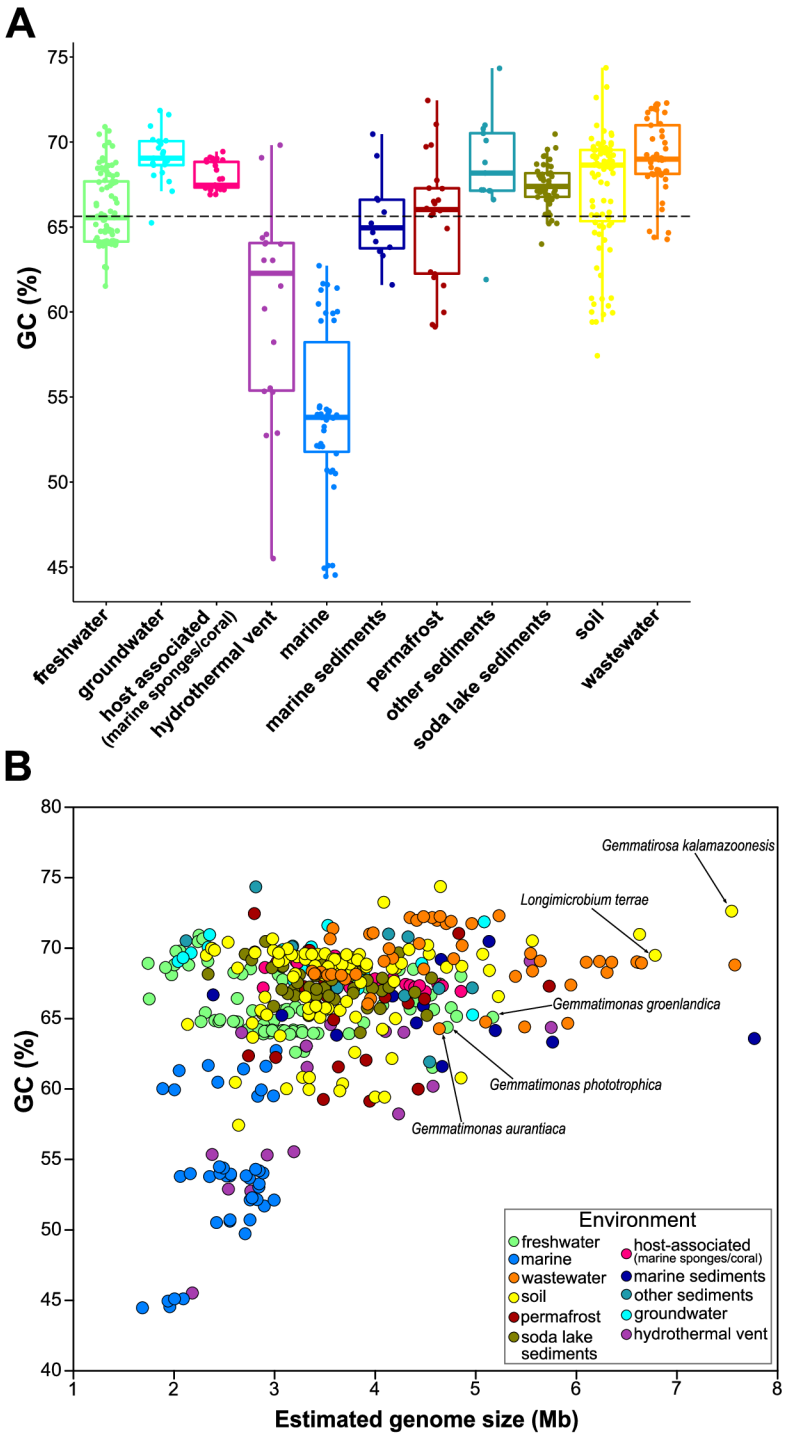

Figure 2. (A) Distribution of GC content (%) of Gemmatimonadota genomes based on their environmental origin. **(B)** A comparison of estimated genome size and GC content (%) of Gemmatimonadota genomes from different environments. Genomes are color coded based on their environment. Labels depict the cultured Gemmatimonadota species.

2.2 Gemmatimonadota habitat-related core and accessory genes analysis

We explored the main shared (core) and flexible (accessory) genome among Gemmatimonadota MAGs with >90% completeness and <10% of contamination across multiple origins. Generally, the size of the habitat-dependent core and flexible genome of Gemmatimonadota differed between environments and ranged from a lower average of 2677 genes in the marine environment (10 genomes), 2868 in the freshwater environment (29 genomes), to the highest average of 4659 genes in wastewater environment (13 genomes) (Table S3), indicating how contrasting environments differentially shape their gene inventories. Larger genomes found in wastewaters may encode a wider variety of enzymes for utilization in an environment often enriched with nutrients (38, 39). The size of the shared genes (strict core and soft core) also varied (Figure S3), while accessory genes formed by the shell (40) and cloud (41) categories represented more than 50% of the flexible/accessory genome in all environments except for marine, showing high variability in the gene inventories among the members of the phylum.

2.3 Multi-environment principal coordinate and phylogenomic analyses

The similarity among genomes was studied using a principal coordinate analysis (PCoA) based on the presence or absence of genes (Table S4). The genomes clustered based on their environmental origin (Figure 3), indicating their differential adaptation to specific environments. Permutational multivariate dispersion analysis (PERMDISP) documented a significant difference in heterogeneity levels (PERMDISP, $P < 0.05$) between various environments (Table S5), and significant differences in gene presence/absence were detected for Gemmatimonadota from all environments (PERMANOVA, $P < 0.001$) except marine sediments and other sediments (PERMANOVA, $P=0.019$). SIMPER analysis based on Bray-Curtis similarity showed host-associated (70.6 %) and marine water (69.9 %) MAGs to be the most similar among them, while those from other sediments (56.2%) marine sediments (59.2%) and groundwater (59.7%) were least similar (Table S5). In the comparison between different environments, marine MAGs were more like other marine-related environments such as hydrothermal vents, or host-associated (with marine sponges and coral) (average dissimilarity of 37.79 % and 40.21 %, respectively), while freshwater MAGs were more similar to wastewater, groundwater and permafrost MAGs (<41 % of average dissimilarity). The highest dissimilarities (>47 %) were seen between soil vs marine and host-associated (with marine sponges and coral) MAGs and between marine vs wastewater MAGs.

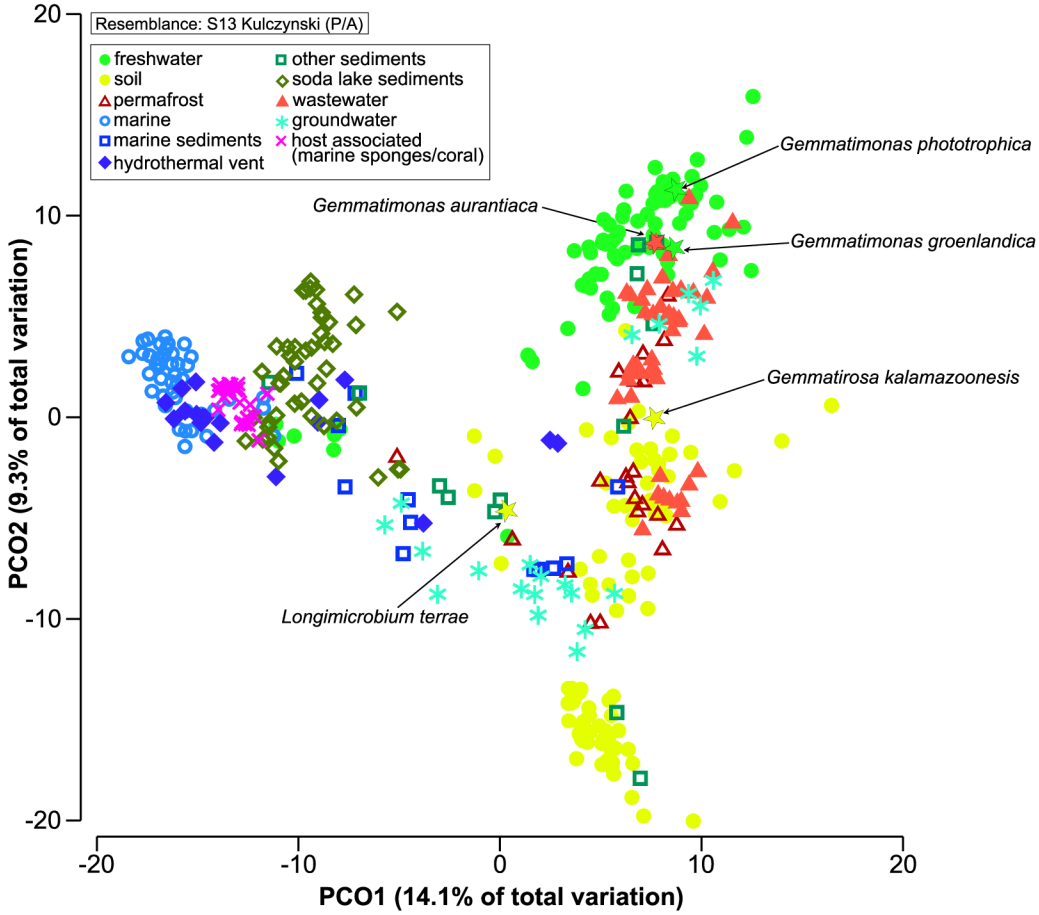

Figure 3. Principal coordinate analysis (PCoA) using Kulczynski resemblance matrix based on SEED presence/absence of the genes in Gemmatimonadota genomes showing grouping of genomes based on their environment. The legend in the upper left corner shows environments are color coded and a different symbol is assigned to each environment. Cultured Gemmatimonadota species are labeled and shown with star symbol.

Similar patterns were found in the phylogenomic analysis (Figure 4), albeit these genomes did not cluster exclusively according to the environment of origin. Most of the obtained MAGs belonged to two families inside the order Gemmatimonadales. The family *Gemmatimonadaceae* encompassed most of the MAGs from fresh-, waste- and groundwater, along with genomes from permafrost and soil (Figure S4), and cultured species *G. phototrophica, G. groenlandica, G. aurantiaca and G. kalamzoonesis*. This family also contained all MAGs from Spanish reservoirs reconstructed in this study. Eleven of them formed a clade related to MAGs from the hypolimnion of several Swiss lakes and Římov Reservoir (Czech Republic). The remaining four clustered together with a previously assigned group Pg2 (12), which consists of freshwater phototrophic Gemmatimonadota from the epilimnion of Lake Zurich (Switzerland) and Římov

Reservoir. The second family GWC2-71-9 mostly contained genomes from the soil, wastewater, permafrost, groundwater, and other sediments, and only a small number of genomes from freshwater lakes or marine sediments (Figure S4).

The second largest group was formed by MAGs belonging to the order Longimicrobiales, which was established based on the soil bacterium *Longimicrobium terrae* (4). This order mostly contains marine water, marine sediments, hydrothermal vents, soda lake sediments and host-associated (with marine sponges and coral) genomes together with several genomes from the hypolimnion of deep freshwater lakes. This is in line with our previous observations that Gemmatimonadota from deep freshwater lakes are related to those from marine environments or environments like soil and sediments (12). Host-associated MAGs (marine sponges and coral) were part of two different families (Longimicrobiales and a not assigned family) and while they are closely related to marine water genomes, the differences in the gene repertoire between these two environments were significant (PERMANOVA, $P < 0.0001$) and they represent real symbionts of marine sponges and corals (42, 43).

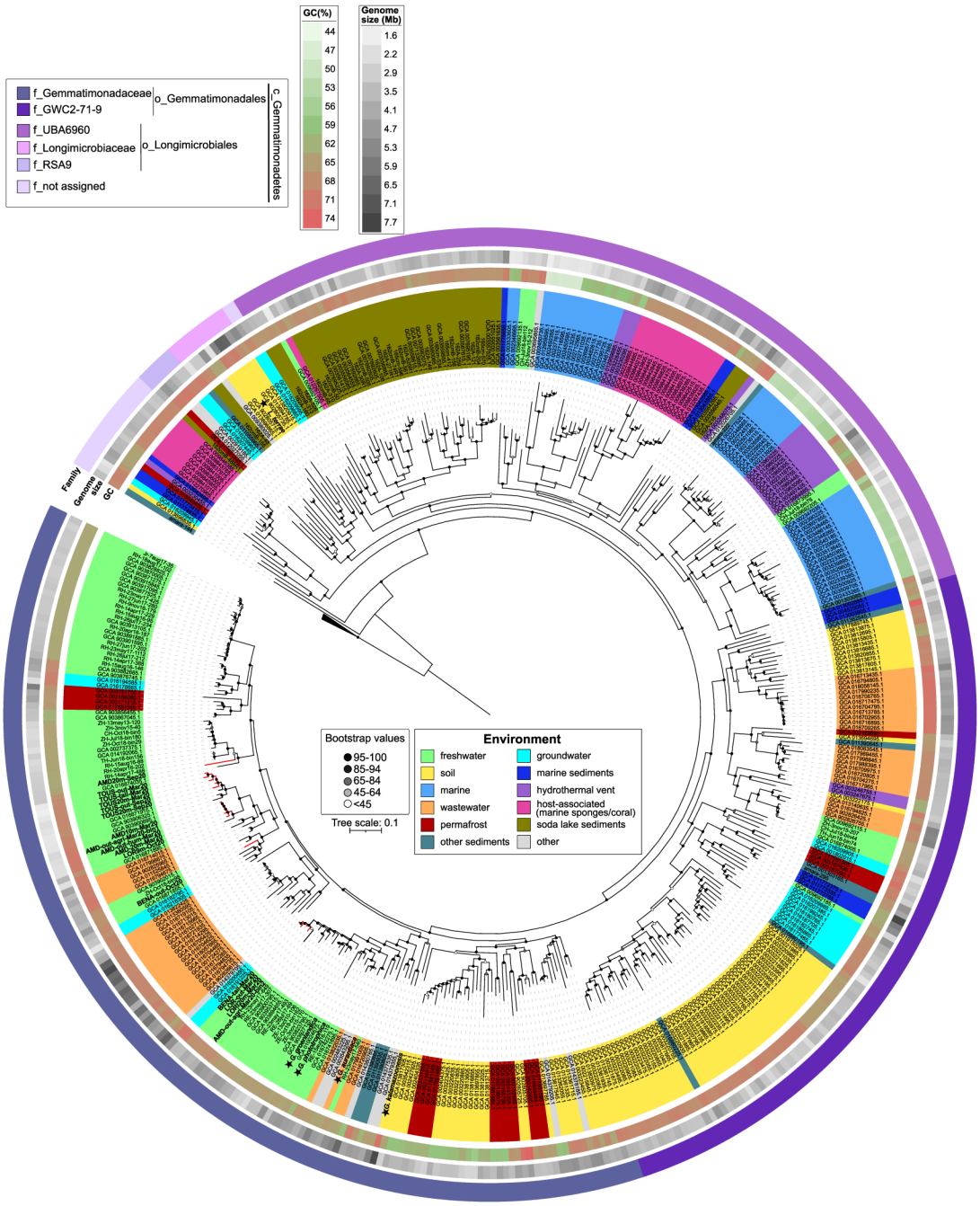

Figure 4. Phylogenomic tree of Gemmatimonadota genomes based on 400 universally conserved and most ubiquitous proteins present in PhyloPhlAn database (44, 45). The collapsed branch represents an outgroup consisting of three genomes from bacterial phylum Fibrobacterota (GCA_900142455.1 *Hallerella intestinalis*, GCA_900217845.1 *Fibrobacter elongates*, GCA_000146505.1 *F. succinogenes*). The strength of support for internal nodes is shown through grey-scale-colored circles (center legend). All genomes are color coded based on their environmental origin (center legend). Following

annotations starting from innermost to outermost indicate GC content (%), estimated genome size (Mb) and family level classification. The legend for each outer circle is represented in upper left corner. Details on all genomes can be found in Table S1 and Table S2 in the Supplemental material.

2.4 Main metabolic pathways across the Gemmatimonadota phylum

The core metabolism for the Gemmatimonadota phylum was reported recently (13). Still, both PCoA and phylogenomic analysis showed that their gene inventories vary depending on their origin, presumably due to an adaptation to the specific conditions and selection pressure in any particular habitat. To study this further, we looked for the main metabolic commonalities and uniqueness associated with each environment. To do so, we individually inspected genomes from all environments to reconstruct a metabolic model of the Gemmatimonadota phylum with pathways presence/absence for each environmental specialist (Figure 5, Figure S5, Figure 6, Table S6).

2.4.1 Basic energy metabolism

Gemmatimonadota from all environments contained basic genes for respiratory metabolism such as NADH:quinone oxidoreductase, cytochrome *c* oxidase, F-type ATPase, and subunits of succinate dehydrogenase involved in oxidative electron transfer chains, or enzymes of heme biosynthesis (Figure 6). Cytochrome *bd* ubiquinol oxidase (encoded by *cydAB* genes) with high affinity for oxygen (46) was present in MAGs from most environments, except for marine and host-associated ones. The host-associated genomes also lacked succinate dehydrogenase cytochrome b subunit (*sdhC*), while in marine water MAGs it was present only in one genome. Genes encoding fumarate reductase, a key enzyme in anaerobic respiration that catalyzes the reduction of fumarate to succinate, were found in host-associated (60%), soda lake sediments (36.9%), hydrothermal vent (33.3%), marine sediments (8.3%), groundwater (14.3%) and marine water (7.14%) MAGs.

Gemmatimonadota also contained genes necessary for central carbohydrate metabolism, including glycolysis (Embden-Meyerhof pathway), gluconeogenesis, tricarboxylic acid cycle, coenzyme A biosynthesis, aerobic route of oxidation of pyruvate to acetyl-CoA via pyruvate dehydrogenase (PDH) and the biosynthesis of phosphoribosyl diphosphate (PRPP), which is needed to produce purines and pyrimidines (Figure S5). The ED (Entner–Doudoroff) pathway to obtain pyruvate without glycolysis was reported as less common in Gemmatimonadota (13) and, similarly, we saw that the key enzyme for this pathway, 2-dehydro-3-deoxyphosphogluconate aldolase (*eda*), was present in MAGs from soil and permafrost (52.2% (47 genomes) and 27.3% (6 genomes) respectively), and almost absent in MAGs from marine, fresh, and wastewaters, other sediments, and marine

sediments, where it was found only in 3 genomes at most. Moreover, phosphogluconate dehydratase (*edd*), which catalyzes another key step in this pathway, was absent in MAGs from all environments. Alternatively, the ED pathway could be supplied through the degradation of D-Galacturonate (13), which can be an important carbon source for microorganisms. The pathway for degradation of D-Galacturnate was present in soil Gemmatimonadota (key enzymes were present from 18.8 to 54.4 % of MAGs, and complete pathway in 16.6 % MAGs) (Figure S5).

Gemmatimonadota from all environments encoded representative genes for the non-oxidative phase of the pentose phosphate pathway, while both key enzymes for the oxidative phase (*zwf*, PGD) were predominately found in permafrost (72.7%), soil (47.7%), and marine sediment genomes (41.6%). A common feature was also the presence of genes encoding the biosynthesis of dTDP-L-rhamnose, an important cell wall component, except for the host-associated MAGs, which lacked two key enzymes (*rfbC* and *rfbD)*. As these MAGs live in a symbiotic association, it is likely they do not require these enzymes, as the same pathway seems to be also missing in Alphaproteobacteria associated with marine sponges (47).

Several metabolic pathways were only common in Gemmatimonadota from specific environments. For example, key enzymes of the glyoxylate cycle (isocitrate lyase (*aceA*) and malate synthase (*aceB*)) were found in soil genomes (52.2%), other sediments (30.7%), groundwater (28.5%), permafrost (22.7%), wastewater (18.4%), marine sediments (16.6%), marine water (7.1%), soda lake sediments (4.3%) and fresh waters (3.3%). Bacteria harboring this pathway can assimilate acetate in the absence of complex substrates (48, 49). Moreover, while some of the genes encoding the pathway for conversion of propionyl-CoA to succinyl-CoA occurred in all environments, all key genes (Figure S5) were present only MAGs from soda lake sediments (21.7%).

Furthermore, MAGs from all environments showed the potential to degrade polysaccharides. Gene encoding endoglucanase (cellulase) was common among MAGs from all environments, while xylanase (endo-1,4-beta-xylanase*, xynA*) was present in soil (22.2%), permafrost (9.1%), and sporadically in freshwater (7.7%) and wastewater MAGs (2.04%). Chitinase (c*hiC*) was present in MAGs from wastewater (28.6%), permafrost (27.3%), marine sediments (25%), marine water (23.8%), soil (16.6%), as well as in several genomes of other sediments (7.7%) and fresh waters (8.8%). Additionally, MAGs from permafrost (59.1%), soil (58.8%), groundwater (47.6%), fresh waters (43.9%), hydrothermal vent (16.6%), and wastewater (6.1%) had chitin disaccharide deacetylase (*chbG*), which is suggested to catalyze the deacetylation of chitin making it an easily degradable substrate (50).

One of the main storage molecules, which helps bacteria to survive periods when nutrients or energy sources are scarce, is glycogen (51–54). The enzymes for its biosynthesis (1,4-

alpha-glucan branching enzyme, glucose-1-phosphate adenylyltransferase and glycogen synthase) were present in freshwater (21.9%), wastewater (18.3%), soil (11.1%), 3 groundwater and soda lake sediments genomes. Similarly, the complete pathway for glycogen degradation was mostly found in wastewater (46.9%), other sediments (30.7%), permafrost (18.2%), freshwater (16.5%), and soil (8.8%) genomes. The ability of some bacteria to accumulate glycogen as an energy reserve (51) allows them to quickly activate their metabolism when nutrients availability increases, providing a competitive advantage in nutrient fluctuating environments.

2.4.2 Carbon fixation strategies

Gemmatimonadota from several different environments contained genes encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO, *rbcL*), sometimes in two or three copies. There are three forms of RuBisCO (type I, II, and III) that catalyze the carboxylation and oxygenation of ribulose 1,5-bisphosphate (55). The most widespread Type I was reported earlier in six Gemmatimonadota MAGs reconstructed from soda lake sediments (28, 29), which were included in this analysis. Genes encoding type I RuBisCO, phylogenetically identified as type IC/ID (Figure S6) were present in soda lake sediments (28.3%), wastewater (36.7%), groundwater (4.7%), soil (3.3%), freshwater (1.1%), and one MAG from glacier ("Other"). Most of these MAGs also contained genes encoding the small subunit of RuBisCO (*rbcS*) and phosphoribulokinase (Figure 5). None of the MAGs contained the proteobacterial, α-cyanobacterial (IA) or the β-cyanobacterial form (IB) of RuBisCO (56).

Bacterial type II RuBisCO, which is less efficient in discriminating between CO_2 and O_2 and adapted to environments with low oxygen concentrations (55, 57), was present in wastewater MAGs (36.7%). Type II is commonly found in Proteobacteria (58) and organisms which also have type I (55), as is the case for some of the wastewater MAGs (12.2%), which had both types. In addition, 72 MAGs contained the so-called type IV *rbcL* gene which is probably not involved in carbon fixation (55, 59, 60).

Key genes encoding the phosphate acetyltransferase-acetate kinase pathway for carbon fixation, in which acetate produced from acetyl-CoA can be used as carbon source or electron donor, were found in soda lake sediments (80.4%), permafrost (59.1%), other sediments (38.4%), wastewater (24.5%), freshwater (19.8%), and soil MAGs (11.1%). In the freshwater environment, the presence of the two key genes was observed in two photoheterotrophic species, as well as in 4 MAGs from Spanish reservoirs.

Carbon monoxide (CO), atmospheric trace gas, can be an alternative energy source for some organo-heterotrophic bacteria during organic carbon starvation, enhancing their survival (61–63). The gene encoding the large subunit of carbon monoxide dehydrogenase (*coxL*) is highly abundant in soils where it can facilitate atmospheric CO removal, and

was reported as present in soil Gemmatimonadota (61, 64). In this study we found the c*oxL* gene, as well as the genes for small (*coxS*) and medium (*coxM*) subunits of carbon monoxide dehydrogenase in host-associated (80%), wastewater (57.1%), hydrothermal vent (50%), groundwater (42.8%), soil (30%), other sediments (23.1%), marine sediments (16.6%), permafrost (9.1%) and freshwater (5.5%) MAGs.

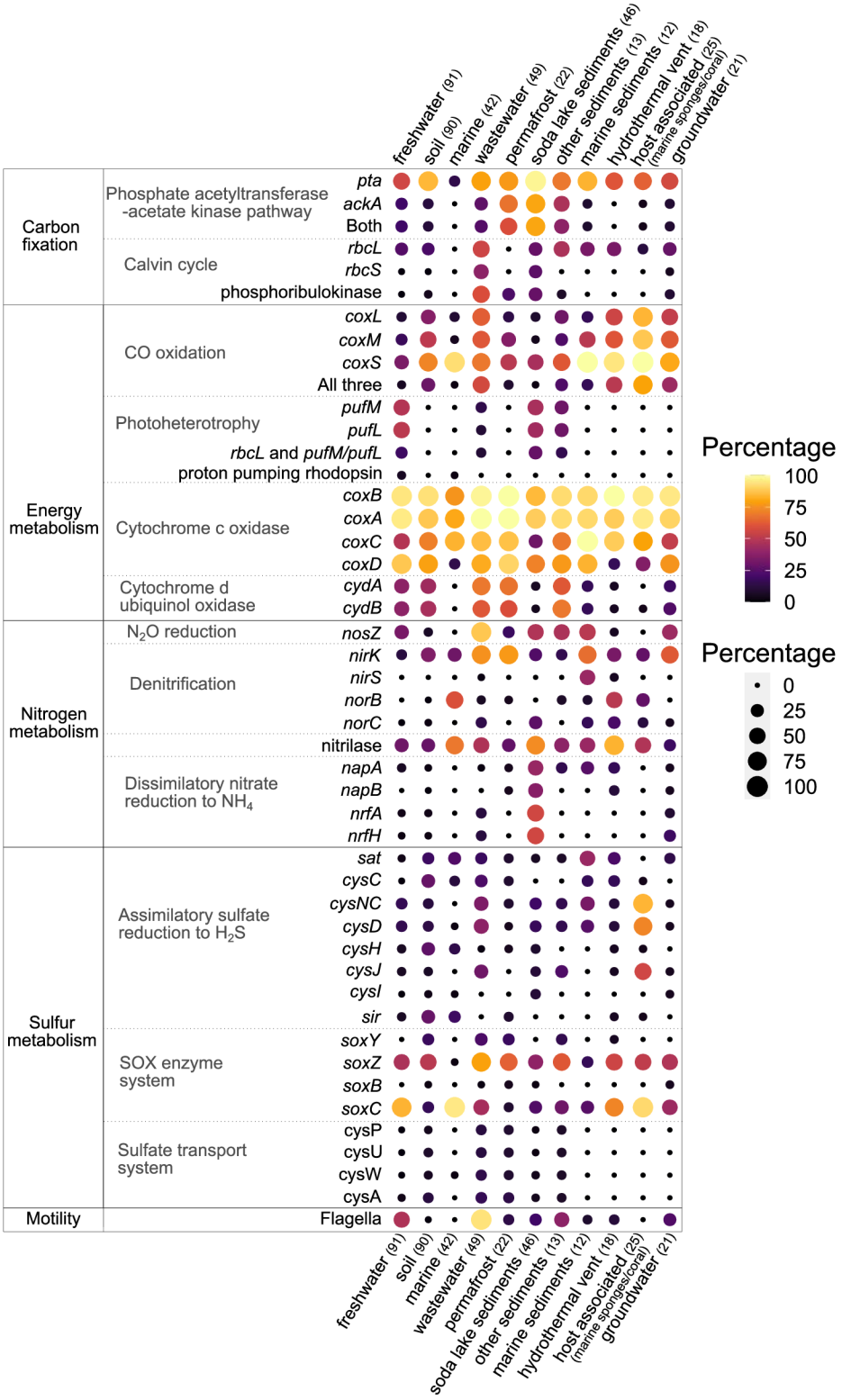
freshwater (91)
soil (90)
marine (42)
wastewater (49)
permafrost (22)
soda lake sediments (46)
other sediments (13)
marine sediments (12)
hydrothermal vent (18)
host associated (25) (marine sponges/coral)
groundwater (21)
Carbon fixation
Phosphate acetyltransferase -acetate kinase pathway
pta
ackA
Both
Calvin cycle
rbcL
rbcS
phosphoribulokinase
Energy metabolism
CO oxidation
coxL
coxM
coxS
All three
Photoheterotrophy
pufM
pufL
rbcL and pufM/pufL
proton pumping rhodopsin
Cytochrome c oxidase
coxB
coxA
coxC
coxD
Cytochrome d ubiquinol oxidase
cydA
cydB
Nitrogen metabolism
N2O reduction
nosZ
Denitrification
nirK
nirS
norB
norC
nitrilase
Dissimilatory nitrate reduction to NH4
napA
napB
nrfA
nrfH
Sulfur metabolism
Assimilatory sulfate reduction to H2S
sat
cysC
cysNC
cysD
cysH
cysJ
cysI
sir
SOX enzyme system
soxY
soxZ
soxB
soxC
Sulfate transport system
cysP
cysU
cysW
cysA
Motility
Flagella
Percentage
100
75
50
25
0
Percentage
0
25
50
75
100

Figure 5. Bubble plot showing the percentages of key genes involved in specific pathways present in Gemmatimonadota genomes from different environments. Dot color and size indicate the percentage of each gene in any given environment, with the darkest color and smallest size of the dot marking the absence of said gene in that environment. Number of MAGs of each environment are labelled in parenthesis. Details about genes presence/absence can be found in Table S6.

2.4.3 Phototrophy

Many freshwater Gemmatimonadota are aerobic anoxygenic phototrophic (AAP) species (6, 11, 12). A common marker gene for AAP bacteria is the *pufM* gene that encodes the M subunit of the bacterial photosynthetic reaction center (65). Here, we identified genes encoding type-2 photosynthetic reaction centers (*pufM* and/or *pufL*) in 51.6% of all freshwater MAGs (Figure 5), as well as in MAGs from soda lake (47.8%) and other sediments (30.7%), wastewater (16.3%), and group "Other" (glacier (3 MAGs), biofilm (1 MAG)), while they were absent from all other environments. Interestingly, 28.3% of phototrophic MAGs from soda lake sediments and 8.2% from wastewater also contained type I *rbcL*, indicating that these species may have the potential for photoautotrophic growth (28, 29).

Many aquatic microorganisms harvest light energy using proton-pumping rhodopsins (66, 67). However, among Gemmatimonadota this system is very rare. We found genes encoding green- or blue-light absorbing proteorhodopsins only in five genomes which originated from deeper layers of freshwater lakes Baikal, Constance, Zurich and Biwa. A xanthorhodopsin gene was identified in one marine and two glacier MAGs (category Other). One of the MAGs from the glacier also contained genes for BChl-*a* based photoheterotrophy indicating the potential for dual phototrophy (68).

2.4.4 Nitrogen cycle

Nitrogen metabolism in Gemmatimonadota is relatively simple. No nitrogen fixation genes were found in any of the analyzed MAGs, which means that Gemmatimonadota must rely on combined nitrogen sources, such as ammonium or amino acids. The gene encoding high-affinity ammonium transporter (Amt), a preferred nitrogen source for microbial growth, was present in MAGs from all environments, as well as the gene encoding nitrilase that hydrolyzes nitriles to ammonia. Branched-chain amino acid transporters were a common feature for host-associated, marine, hydrothermal vent and wastewater genomes, while spermidine/putrescine or nitrate-nitrite/taurine transporter was more common in wastewater, freshwater, and soil genomes (Figure 6).

The complete denitrification pathway was not identified in any analyzed MAGs. However, the nitrous oxide reductase (*nosZ)* gene was found in genomes from all environments except for marine waters and host-associated (Figure 5). This enzyme

catalyzes the final step of denitrification (69–72) but is also considered an independent respiratory reaction since it is often found in organisms lacking other genes for denitrification, such as *nirK*, *nirS,* and *nor* (73). In Gemmatimonadota MAGs, *nirK* gene (NO^- forming nitrite reductase was found in all environments, however, the *nirS* gene was only present in marine sediments and single genomes from hydrothermal vents and wastewater. Gemmatimonadota *nosZ* genes seem to be one of the most abundant in soil environments (74–76), and their high presence in other environments points to their potentially important role in reducing the N_2O. Both *G. aurantiaca* and *G. kalamazoonensis* have been suggested to use N_2O as a substitute for O_2 to survive temporary anoxia during transitions between oxic and anoxic states, which can be common in soil or wastewater environments (71, 77). Furthermore, 59.5% marine, 44.4% hydrothermal vent, and 28% host-related MAGs did not have *nosZ* but contained *norB* (nitric oxide reductase subunit B), which converts nitric oxide (NO) to N_2O and could point to their genetic potential to produce N_2O. The presence of this gene in host-associated Gemmatimonadota suggests their potential role in nitrogen cycling as part of the marine sponge microbiome (42). Finally, genes for the dissimilatory nitrate reduction to ammonia (*napAB* and *nrfAH*) were common in soda lake sediments (28), probably due to the anaerobic conditions that can occur in these habitats.

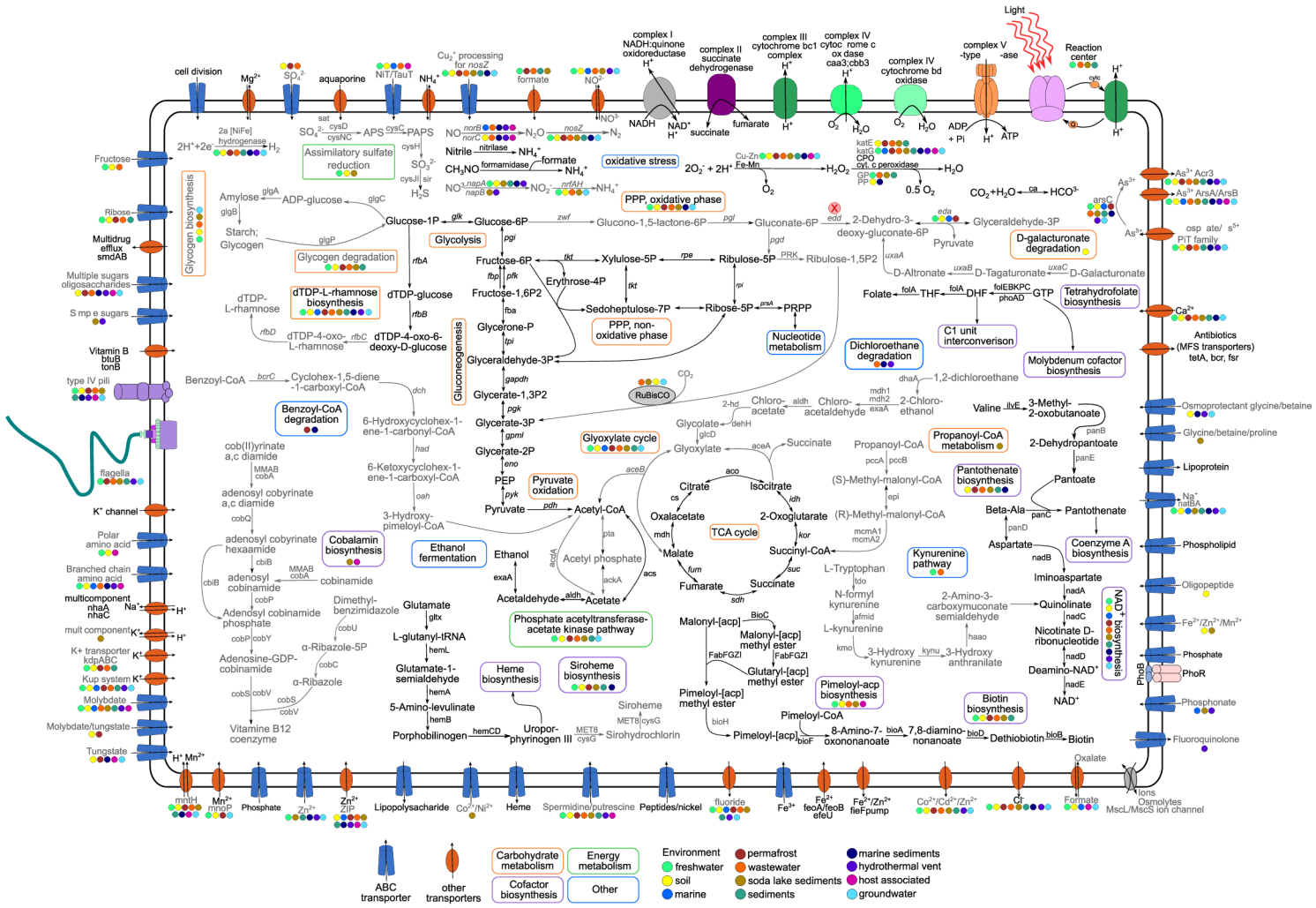

Figure 6. Metabolic reconstruction of Gemmatimonadota showing some of the key pathways. Four different colored rectangles depict names of pathways and metabolic processes. Pathways labelled with black are present in all Gemmatimonadota genomes,

while the ones in grey only in Gemmatimonadota genomes from certain environments. Color coded circles representing different environments of origin indicate the said gene/pathway/transporter was present in that environment (shown if at least two genomes showed presence). Details of genomes can be found in Table S6. Abbreviations for compounds: PEP- phosphoenolpyruvate; PPP- pentose phosphate pathway; PRPP- 5-Phosphoribosyl 1-pyrophosphate; CPO- chloroperoxidase; GP- glutathione peroxidase; PP- porphyrinogen peroxidase; THF- tetrahydrofolate; DHF- dihydrofolate; GTP-guanosine 5'-triphosphate; APS- adenylyl sulfate; PAPS- 3'-phosphoadenylyl sulfate; NAD^+- nicotinamide adenine dinucleotide.

2.4.5 Sulfur cycle

Regarding the sulfur cycle, the distribution of genes encoding enzymes involved in assimilatory sulfate reduction to H_2S (*sat*, *cysC, cysNC*, *cysD*, *cysH, cysJ, cysI* and sulfite reductase) was patchy (Figure 5). In this pathway, sulfate is reduced to H_2S, which is then incorporated into cysteine which can be subsequently used for the synthesis of other sulfur-containing molecules (78, 79). The complete pathway was found in the highest numbers in soil (18.9%) and soda lake sediments (10.8%). In fresh waters, hydrothermal vents, permafrost and host-associated it was only present in up to three genomes. Genes encoding for the sulfate transport system, which enables sulfate-sulfur assimilation, were mostly found in wastewater MAGs, with a lower occurrence in soil, and permafrost MAGs. Furthermore, the complete sox enzyme system, involved in thiosulfate oxidation to SO_4^{2-} was not found in any Gemmatimonadota genomes, although they contained some genes (*soxZ*, *soxY*, *soxC* or *soxB*), depending on the environment.

2.4.6 Phosphate

Phosphate is one of the main biogenic elements required for the biosynthesis of nucleic acids and lipids. Due to its low natural availability, it represents the limiting nutrient in many natural environments. The main route for its uptake in Gemmatimonadota from all environments was the high-affinity phosphate transport system (*pstSCAB*), and they could regulate its acquisition through the PhoR-PhoB two-component system. Additionally, marine (19%) and two MAGs from soda lake sediments, and hydrothermal vents had an uptake system for phosphonate (*phnCDE*), a good source of phosphorus under phosphate starvation (80). During phosphorous starvation, many bacteria can produce alkaline phosphatases (*phoA, phoX, phoD*), which facilitate the use of phosphoesters as a phosphorous source (81, 82). *PhoA* was found in Gemmatimonadota from all environments, with the lowest numbers in host-associated (8%) and highest in permafrost MAGs (68.2%). In contrast, *phoX* was present in up to two MAGs in marine sediment and fresh waters and was generally found in lower numbers in all environments except for host-associated (92%), marine waters (57.1%) and wastewater (55.1%). Finally, *phoD* was highly present in freshwater (92.3%), wastewater (69.4%) and hydrothermal vent

(66.6%) MAGs. The Polyphosphate kinase (PPK) gene used for the accumulation of polyphosphate, was present in all Gemmatimonadota MAGs except marine. The presence of all these genes, which are crucial during phosphorus limitation, as well as the high-affinity phosphate transport system indicates that Gemmatimonadota has different strategies to cope with phosphorous limitations.

2.4.7 Protection against oxidative stress

Gemmatimonadota is composed of mostly aerobic organisms depending on aerobic respiration. Therefore, their genomes encode many proteins involved in the protection from oxidative damage and stress that is associated with an aerobic lifestyle (Figure 6). [Fe-Mn] and [Cu-Zn] families of superoxide dismutases were present in MAGs from all environments, except for marine and hydrothermal vents, where the [Cu-Zn] family was not found. Cytochrome c peroxidase also occurred in all environments, while glutathione peroxidase was found in high numbers in wastewater (53.1%) and fresh waters (38.4%), although in other environments only in few representatives were found or not at all. Catalase peroxidase *katG* was present in all MAGs except soil, while catalase *katE* was found in soda lake sediments (39.1%), other sediments (23%), wastewater (22.4%), permafrost (18.2%), soil (12.2%), and only one freshwater bacterium, *G. groenlandica* (6). From other types of peroxidases, chloroperoxidase occurred in all environments, while porphyrinogen peroxidase was only present in two MAGs from soil and one from soda lake and marine sediments.

All so far cultured Gemmatimonadota contain large amounts of carotenoids. These pigments protect cells from excess light as well as against reactive oxygen species, and in AAPs can act as additional light harvesting pigments (11, 83). Gemmatimonadota MAGs from wastewater (85.7%, 67.3%), fresh water (58.2%, 63.7%), other sediments (69.2%, 76.9%), soda lake sediments (39.13%, 91.3%) and soil (24.4%, 21.11%) contained genes encoding the initial part for carotenoid biosynthesis, phytoene synthase (*crtB*) and phytoene dehydrogenase (*crtI*), respectively. They were almost absent in marine, permafrost, marine sediments, hydrothermal vents and groundwater genomes where they were found in up to two genomes. Host-associated genomes contained *crtI* gene (24%) but *crtB* was present only in one genome. Other carotenoid biosynthesis genes found were ß-carotene ketolase (*crtO*) present in all environments, lycopene beta-cyclase (*crtY*) found in a small number of soil and freshwater MAGs, and up to three genomes of wastewater, permafrost, marine, marine sediments and soda lake sediments and ß-carotene 3-hydroxylase (*crtZ*) found only in several freshwater and wastewater genomes.

2.4.8 Cofactors and vitamins

Gemmatimonadota cultures require a mixture of vitamins like biotin (Vitamin B_7), folic acid (Vitamin B_9), nicotinic acid (Vitamin B_3), pantothenic acid (Vitamin B_5) and cobalamin (Vitamin B_{12}) for growth (1–5). All analyzed MAGs contained the complete pathway for molybdenum cofactor synthesis. Molybdenum is a cofactor in numerous enzymes in living organisms (84). Folate biosynthesis could be inferred as complete in all environments, given that marine, host-associated, wastewater and other sediments MAGs that lack one of the key enzymes (*folA*) had the gene encoding *thyX*, suggested to function as *folA* (85). Furthermore, MAGs from most environments encoded genes involved in pantothenate biosynthesis, a precursor of coenzyme A, an essential molecule in metabolism. The exceptions were marine, freshwater, and host-associated MAGs, which lacked one of the key enzymes (*panD*). Genes encoding biosynthesis of biotin, an essential cofactor of enzymes involved in fatty acid synthesis or amino acid metabolism (86), were present mostly in freshwater, wastewater, permafrost, other sediments, and several soil MAGs. In other environments, several genes of biotin biosynthesis were missing. Genes for NAD^+ biosynthesis, an important metabolite and cofactor involved in nucleotide synthesis, were only sporadically present in host-associated MAGs, probably due to their incompleteness. Another biosynthetic pathway, kynurenine pathway that leads to quinolinate, a precursor of NAD (87), was present in freshwater and wastewater MAGs. Soda lake sediments and host-associated MAGs had several genes involved in late steps of cobalamin biosynthesis from cobyrinate a,c-diamide, however, genes involved in both aerobic and anaerobic cobalamin pathways were not found. This suggests that these genes may be used in the salvage pathway as a more effective way for obtaining cobalamin, since the *btuB* transporter and *tonB* protein, which function together in cobalamin transport (88), were present in Gemmatimonadota.

2.4.9 Other genes

Genes for flagella assembly were present in almost all wastewater (93.8%) and nearly half of freshwater (47.3%) MAGs (Figure 5). This included five of the Gemmatimonadota MAGs from Spanish reservoirs and many limnic and planktonic MAGs from Římov Reservoir (12). Additionally, the presence of flagella was already shown for both freshwater cultures *G. phototrophica* and *G. groenlandica* (8). Smaller numbers were found in other sediments (38.4%), groundwater (23.8%), soda lake sediments (19.6%), and permafrost (13.6%). Host-associated, and marine MAGs did not contain flagellar genes, while in soil, hydrothermal vents, and marine sediments they only occurred in 1 or 2 genomes, respectively. Additionally, genomes from all environments but marine had genes encoding type IV pili (Figure 6).

Some Gemmatimonadota also displays different enzymes for degrading alkanes, which they may potentially use as a source of carbon and energy (89). Alkanes are naturally found in environments from sources like decaying microorganisms, algae or plants but are also present in high content in crude oil, which can be a contaminant for the environment (89). The gene encoding alkanesulfonate monooxygenase (*alkB*) which degrades short alkanes (90) was found in the highest numbers in soil (30 %), wastewater (18.4%), freshwater (14.3%) and marine MAGs (14.3%). Alkane-1-monooxygenase (*alkM*), used in the degradation of longer alkanes (90) was less common and was missing from most environments except marine (45.2%), host-associated (36%), and freshwater (8.8%) MAGs. The presence of these genes could suggest a potentially ecologically relevant role in the biodegradation of hydrocarbons in Gemmatimonadota. This potential for biodegradation is also evident in Gemmatimonadota MAGs from permafrost and marine sediments, which had genes involved in the degradation of benzoyl-CoA, a central intermediate of synthetic aromatic compounds (91). Moreover, Gemmatimonadota from wastewater, fresh water, hydrothermal vents and one genome from marine sediments seems to be able to degrade 1,2- dichloroethane (13), an industrially produced pollutant in aquatic environments (92). Furthermore, Gemmatimonadota seems to utilize glycolate converting it to glyoxylate as they have genes encoding glycolate oxidase, a protein complex which consists of three subunits D, E, and F (*glcDEF*). This could explain the previous observation of a close association of limnic Gemmatimonadota with phytoplankton in freshwater environments, (12) since glycolate is one of the most common cyanobacterial and algal exudates that can be utilized by bacteria (93–95). Additionally, the *glc* operon contains malate synthase G (*glcB*) that further converts glyoxylate to malate, which is then used for energy production in the TCA cycle (96).

Several different antimicrobial compounds and multidrug transport systems were present in all Gemmatimonadota, while importers and efflux systems of ions like Fe^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+} and other heavy metals were present in different environments (Figure 6). The Na^+/H^+ antiporter system to remove Na^+ from cells, as well as Kch voltage-gated K^+ channels, important in all prokaryotes for maintaining cellular homeostasis (31, 97), were present in all Gemmatimonadota. Trk-type, fast but low-affinity, K^+ transporter (97) predominated in marine water, marine sediment, and hydrothermal vent genomes, while K^+ channels and the K^+/H^+ antiporter system in MAGs from soda lake sediments, freshwater, soil, permafrost, and wastewater. To deal with hypo-osmotic stress Gemmatimonadota from all environments had aquaporins, water channels which ease the water stress by enabling fast water efflux (98), and two types of mechanosensitive channels, MscL and/or MscS, which also help cells return to normal, isotonic size (99).

3. Conclusions

We have explored a large dataset of Gemmatimonadota MAGs to characterize their metabolic potential in different environments. Phylogenomics and gene content analyses indicated that Gemmatimonadota have diverse and flexible metabolisms and the ability to adapt to different conditions. A common feature of all MAGs was aerobic organoheterotrophy, but many pathways were specific to some environments. For instance, photoheterotrophy and motility (flagella) were more prevalent in freshwaters, soda lakes and wastewaters, whereas CO oxidation was more common in soils, marine sediments, hydrothermal vents, host-associated and groundwater. Differences between environments could also be observed in their genomes' sizes and GC content. The size and GC content of marine MAGs was the lowest, which is a common adaptation of marine bacteria to oligotrophic conditions. Moreover, Gemmatimonadota exhibit different strategies for survival under phosphorous limitation, some present in all genomes, and some like uptake system for phosphonate were more common in marine and hydrothermal vent genomes, or different alkaline phosphatases like *phoD* or *phoX* more common in host-associated, freshwater or wastewater genomes. Gemmatimonadota are unable to fix nitrogen, however a potential environmental role in the reduction of N_2O is highlighted by the presence of *nosZ* genes in all Gemmatimonadota except for marine, host-associated and hydrothermal vents, in which the presence of *norB* could suggest that they may rather produce N_2O. Finally, pathways for the degradation of synthetic solvents and aromatic compounds found in some Gemmatimonadota point to their potential role in biodegradation in the environment, and the ability to utilize glycolate indicates to a potential symbiotic relationship with phytoplankton.

4. Methods

Sampling, sequencing, and assembly

Samples collected from four Spanish freshwater reservoirs (Amadorio, Tous, Benageber and Loriguilla) (56) were reused in this analysis. All four monomictic reservoirs are located in the semi-arid eastern region of Spain, close to the Mediterranean Sea. Briefly, for each reservoir samples were taken in two campaigns (March - winter mixing period and September/October - summer stratification period in 2020) at three locations: dam (epilimnion, deep chlorophyll maximum-DCM and hypolimnion in summer; and epilimnion and hypolimnion in winter), outlet of the reservoir and tailwaters (0.5 m). In all cases, water samples were filtered through a series of 20, 5 and 0.22 µm filters and DNA was extracted as described in Cabello-Yeves et al., 2022. DNA extracted from 0.22 µm filters was sequenced with Illumina NovaSeq.

Metagenomes were assembled with IDBA-UD (100) which resulted in ca. 5000 contigs > 5 kb per metagenome which were used for further binning. Binning was conducted using METABAT2 (101) and we obtained a total of 16 metagenome assembled genomes (MAGs) ascribed to the Gemmatimonadota phylum (Table S1). A quality check of 16 MAGs was made using the CheckM v1.1.3 package (102). All MAGs had < 5% contamination, while completeness ranged from lowest 67.79% to 100%.

Analyzed dataset

The obtained MAGs from the Spanish reservoirs were expanded with five cultured representatives and all publicly available MAGs of Gemmatimonadota from NCBI (downloaded on 3rd of May 2021) together with previously published freshwater Gemmatimonadota (12). All genomes (731) (Table S2) were checked for completeness and contamination using CheckM package (102), and 68 MAGs with completeness below 50% and/or contamination above 10% were removed from further analysis. Moreover, basic metadata such as the environmental origin, assembly size, estimated genome size, GC content, median intergenic spacer, and coding density were obtained for each genome. Environmental origin and assembly size were collected from NCBI, GC content, median intergenic spacer and coding density were calculated using the in-house pipeline, and estimated genome size was calculated based on the formula: (total sequence length / completeness) × (100 - contamination). All of these were taxonomically classified with Genome taxonomy database (GTDB-Tk) with default settings (103). This identified 57 ambiguous genomes, which were re-classified as Latescibacterota and several other closely related bacterial phyla (e.g., Eisenbacteria, Krumholzibacteriota). These genomes were also removed from all subsequent analysis. Finally, in order to avoid bias and reduce redundancy, the remaining genomes were dereplicated using dRep v 2.3.3 (104), with parameters: -comp 50 –pa 0.99 -sa 0.995. The final data set consisted of 442 MAGs (completeness >67%, a value chosen a it represents 2/3 of the genome) that were divided based on their environmental origin into 12 different categories: freshwater, soil, marine water, marine sediment, hydrothermal vent, permafrost, soda lake sediments, other sediments, host associated (with marine sponges and coral), wastewater, groundwater and "Other". The latter category ("Other") included MAGs from glacier (6), biofilm (1), bioreactor (2), fossil (1), compost (1), hot spring (2), and an unknown metagenome (1) (Table S2). Since the environments in this category were too diverse to be considered together, the MAGs were excluded for all of the analysis, except for phylogenomics of Gemmatimonadota genomes, RuBisCO tree and phototrophy in metabolic analysis. Coding density plots showing comparison of estimated genome size with % GC, number of CDS and median intergenic spacers (bp) were performed for all genomes that had > 67% completeness, excluding the category "Other". The graphs were plotted using

SigmaPlot v.14.0 and Rstudio v.3.6.1 (packages ggplot2 and maps) and edited in Inkscape v.1.0.

Analysis of core and accessory genes of Gemmatimonadota from different habitats

The analysis of core and accessory genes from Gemmatimonadota genomes was done using the GET_HOMOLOGUES package based on diamond blastp and OMCL algorithms with default parameters (105). Only environments where MAGs/cultured genomes had completeness above 90% were analyzed. These included the following categories: freshwater (29), soil (34), permafrost (9), soda lake sediments (20), wastewater (13), marine (10), and host-associated (14). To avoid bias due to reduntant genomes and variability in completeness of similar MAGs, genomes with ≥ 98% of average nucleotide identity (ANI) in the same environment were excluded. This analysis must be treated with caution due to the varying level of completeness of the analyzed MAGs regardless of the environment where they had originated, and differing numbers of MAGs used for each environment. Average number of core, soft core, shell and cloud genes was calculated for each of this environmental groups (Table S3). Core genes were defined as present in all considered genomes of the analyzed environment, and soft core as genes present in 95% of them. The shell category comprises moderately conserved genes present in <90% of compared genomes. Finally, cloud genes are rare genes present in only one or two genomes (105).

PCoA/clustering plots and Phylogenies

A Principal coordinate ordination analysis with SEED (106) gene presence/absence (Table S4) was conducted for all genomes (excluding the category "Other") with completeness higher than 67%. Briefly, a Kulczynski resemblance matrix based on SEED presence/absence gene values was obtained and the derived triangular matrix was used to obtain a clustering and PCoA analysis where all genomes were distributed accordingly. Additionally, similarity percentage analysis (SIMPER) was done with the same SEED presence/absence gene values (Table S4) using Bray-Curtis (Table S5). Differences in dispersion of genes was tested by performing an analysis of permutational multivariate dispersion (PERMDISP) (107) which includes pairwise comparisons of environments. To test for significant differences between environments PERMANOVA was performed using 9999 permutations. All the calculations were conducted with PRIMER7 software (Primer Ltd., Lutton, UK), and the obtained graph was further edited in Inkscape v.1.0.
Phylogenomic analysis of Gemmatimonadota genomes was done with PhyloPhlAn 3.0 tool (44, 108). Three genomes from bacterial phylum Fibrobacterota were used as an outgroup (GCA_900142455.1 *Hallerella intestinalis*, GCA_900217845.1 *Fibrobacter elongates*, GCA_000146505.1 *F. succinogenes*). PhyloPhlAN uses USEARCH (109) to

screen for the presence of 400 universally conserved and most ubiquitous proteins (found in PhyloPhlAn database). The alignments of proteins against built-in database were done using MUSCLE (110), concatenated and used to generate maximum-likelihood tree with RAxML (111). The tree was visualized in iTOL (112) and edited using Inkscape v.1.0.
A RuBisCO tree was constructed with representative sequences of the large subunit (*rbcL/cbbL* genes) from various types including type IA, IB, IC, ID, II, intermediary II/III, III, IV and archaeal types. This RuBisCO dataset was aligned in Geneious Prime (version 2022.2.2) using MAFFT alignment (113, 114) (n=508). Sequences that were not obtained from Gemmatimonadota MAGs were downloaded from UniProt (115) or obtained from previous studies (19, 56, 116, 117). A maximum-likelihood phylogenetic tree was calculated using IQ-TREE (118) with LG+F+I+G4 substitution model chosen as the best-fitting model by ModelFinder according to Bayesian Information Criterion (BIC) (119) and 1000 ultrafast bootstrap replicates. The tree was visualized in iTOL (112) and edited using Inkscape v.1.0.

Metabolic analysis

This analysis was conducted in genomes with > 67% completeness. Gene predictions were performed with PROKKA (120) and diamond (v0.9.14.115) blastp was used to search versus KEGG/SEED databases (Table S4). Metabolic features of MAGS were also analyzed with RAST annotation pipeline database (106) and through BlastKOALA (121), which allowed us to obtain KO identifiers (K numbers) to orthologous genes present in all MAGs (Table S6). Metabolic pathways were then inferred from KEGG (121) and SEED (106) and manually examined for completeness. The percentages of the presence of key genes and pathways were calculated for each environment (Table S6). Plots showing the percentage of presence of specific metabolic pathways were done in Rstudio (package bubbleplot) and edited in Inkscape 1.0. In the Figure S5 the genomes which had present majority of the genes (>60%) related with flagella were considered as having it present. The figure of metabolic reconstruction was done in Inkscape 1.0, following Table S6.

Supplement Figures

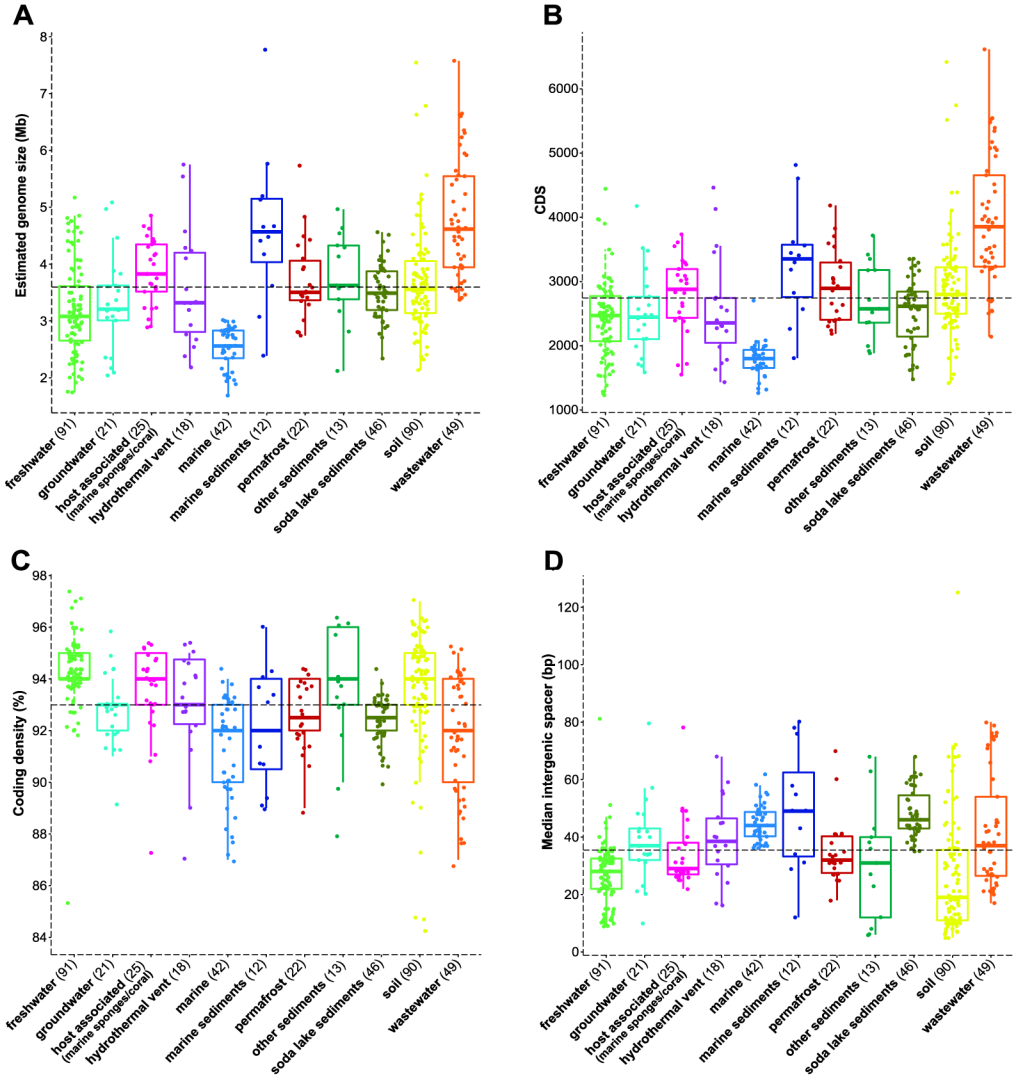

Figure S1. Basic characteristics of Gemmatimonadota genomes. Plots show **(A)** range of estimated genome sizes, **(B)** number of CDS, **(C)** ranges of coding density (%) and **(D)** ranges of median intergenic spacer (bp) of Gemmatimonadota genomes from different environments. Environments are color coded, and the number of genomes used in each environment is labeled.

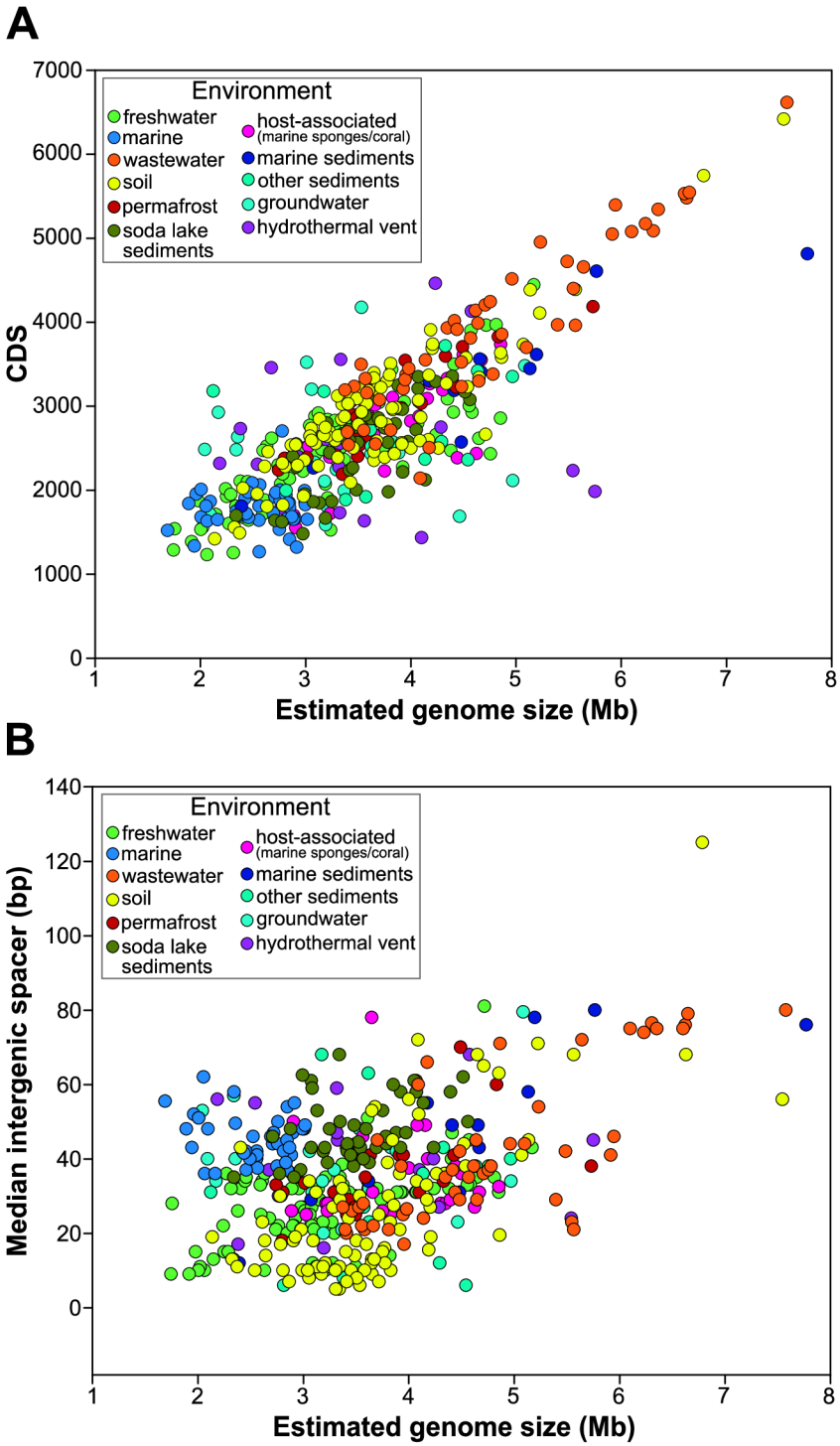

Figure S2. Coding density plots showing comparison of **(A)** CDS and **(B)** median intergenic spacer (bp) with the estimated genome size of Gemmatimonadota genomes from different environments color coded.

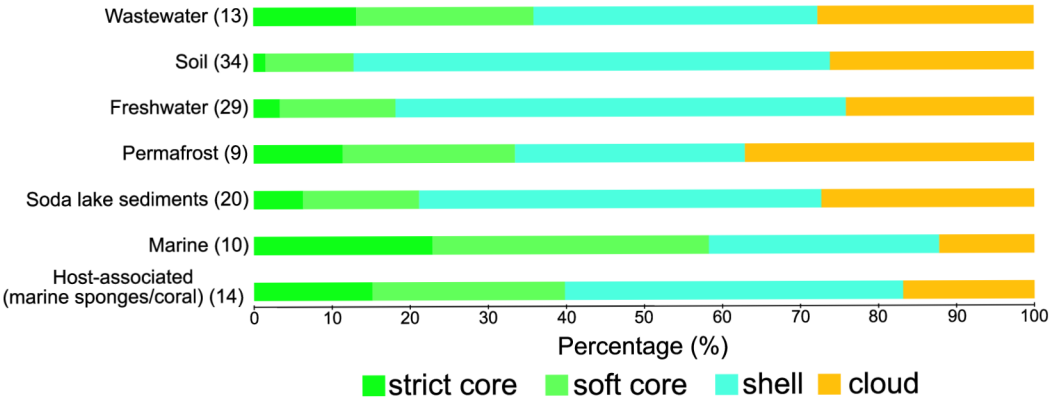

Figure S3. Habitat related core and accessory genes of Gemmatimonadota genomes from 7 environments showing percentage of genes forming strict core, soft core, cloud and shell part of the genomes.

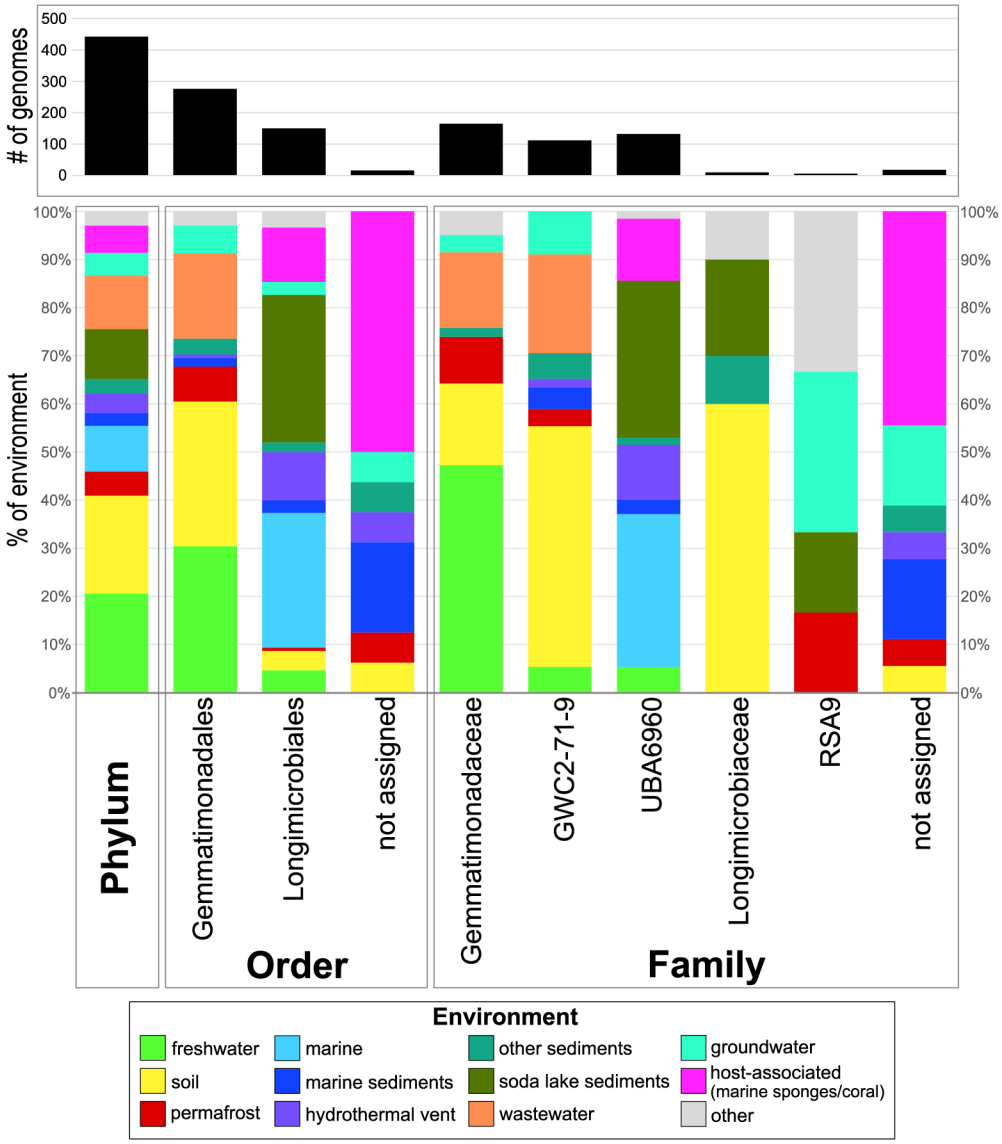

Figure S4. Environment related grouping of Gemmatimonadota on different taxonomic levels (phylum, family, order) based on taxonomic assignation by GTDB (103). Bottom legend shows color-coded environments from which the Gemmatimonadota genomes originate.

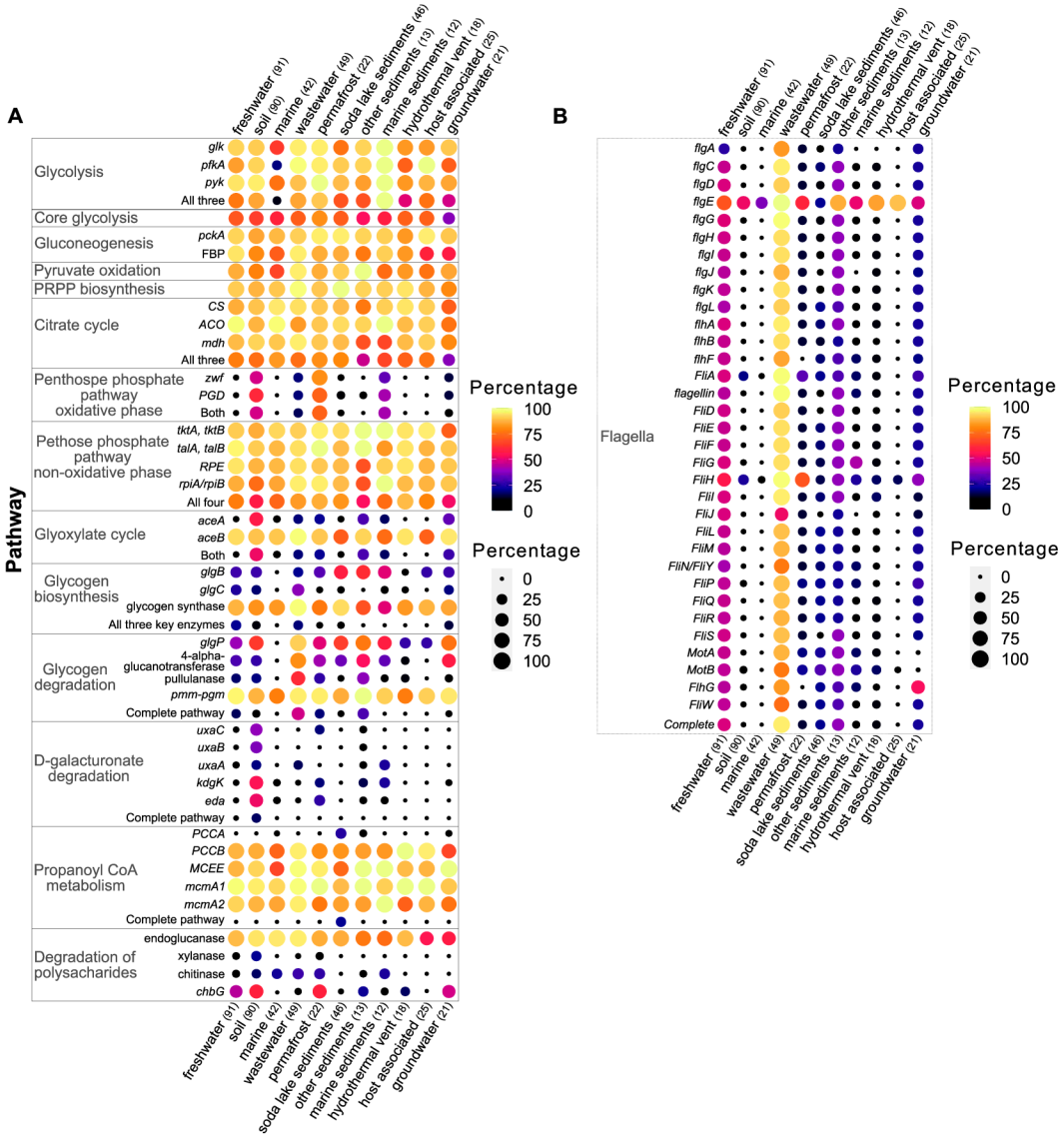

Figure S5. (A) Bubble plot showing the percentages of key genes involved in central and other carbohydrate metabolism present in Gemmatimonadota genomes from different environments. **(B)** Bubble plot showing flagellar assembly genes. Dot color and size indicate the percentage of each gene in certain environment, with darkest color and smallest size of the dot marking the absence of said gene in that environment. The number of MAGs from each environment are labelled in parentheses. Details about genes presence/absence can be found in Table S6.

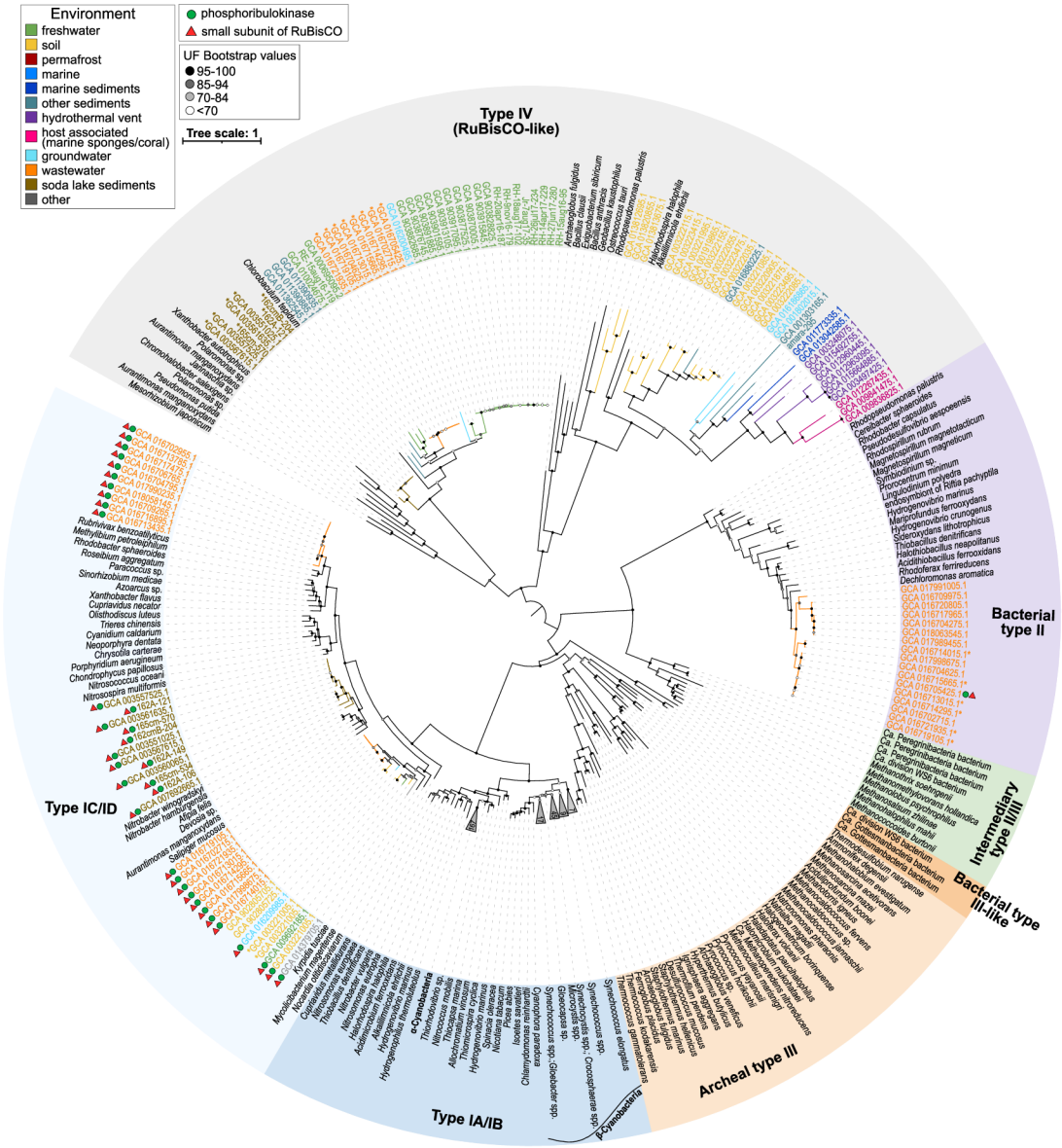

Figure S6. Maximum-likelihood phylogenetic tree (LG+F+I+G4 substitution model chosen as the best-fitting model by ModelFinder according to Bayesian Information Criterion (BIC) (119) and 1000 ultrafast bootstrap replicates) of the large subunit of RuBisCO (types I-III) and RuBisCO like (type IV) proteins (*rbcL*) showing position and classification of Gemmatimonadota RuBisCO sequences. Sequences are color coded based on environment of origin. The presence of small RuBisCO subunit (*rbcS*) and phosphoribulokinase, is labeled with red triangle and green circle, respectively. The numbers shown at collapsed branches (i.e., 7 and 46) indicate the numbers of genomes (not shown) comprising the respective taxonomic categories.

Table S1. Basic characteristics, taxonomic classification and accession numbers of 16 assembled Gemmatimonadota MAGs from four Spanish reservoirs.

Table S2. Basic characteristics and taxonomic classification of all Gemmatimonadota MAGs downloaded from NCBI (3rd of May, 2021), and previously published freshwater MAGs (12) (in total 731 MAGs) (Sheet 1). The second sheet shows only final set of MAGs chosen after removing MAGs/SAGs which taxonomically belonged to closely related bacterial phyla, and removing all MAGs/SAGs with completeness lower than 67% and contamination higher than 10%. The third sheet shows list of dereplicated MAGs used for analyses.

Table S3. List of genomes from 7 different environments used for reconstruction of the habitat-related core and accessory genes and average number and percentages of strict core, soft core, shell and cloud genes.

Table S4. List of presence or absence (value 1 means present, and value 0 means absent) of genes based on SEED (106) for Gemmatimonadota MAGs.

Table S5. SIMPER analysis of similarity and dissimilarity (using Bray-Curtis similarity) between Gemmatimonadota genomes from different environments, based on Table S3 of SEED presence/absence of the genes (Sheet 2 and 3). Sheet 4 shows result of Permutational multivariate dispersion analysis (PERMDISP) - pairwise comparison of environments (using Kulczynski resemblance) based on Table S3, and Sheet 5 is result of PERMANOVA analysis between environments also based on Table S3 (using Kulczynski resemblance).

Table S6. KO identifiers (K numbers) of genes present in Gemmatimonadota MAGs. Presence/absence (value 1 means present, and value 0 means absent) of genes from different pathways based on KEGG. Additionally, the percentage of key genes and pathways is shown.

Acknowledgements

This work was supported by project PhotoGemm+, no. 19-28778X from the Czech Science Foundation awarded to MK. IM was additionally supported by an Erasmus+ traineeship grant from the University of South Bohemia, Česke Budejovice (NO. 16/2020-2021/PRF). PJC-Y was supported by a APOSTD/2019/009 Post-Doctoral Fellowship from Generalitat Valenciana and is currently supported by a Post-Doctoral Fellowship from the Fundación Alfonso Martín Escudero, Spain. This work was supported by grants "VIREVO" CGL2016-76273-P [MCI/AEI/FEDER, EU] (cofounded with FEDER funds) to FRV, and CLIMAWET-CONS (PID2019-104742RB-I00) to AC, both from the Spanish Ministerio de Ciencia e Innovación and Agencia Estatal de Investigación, as well as "HIDRAS3" PROMETEO/2019/009 from Generalitat Valenciana granted to both researchers. FRV was also the beneficiary of the 5top100-program of the Ministry for Science and Education of Russia. The authors thank Alastair T. Gardiner for language corrections.

Author contributions

PJC-Y, AC, IM and MK conceived the study. PJC-Y, AP and AC performed the sampling campaigns from which metagenomes were derived. IM, PJC-Y, CVA, KP analyzed the sequence data. IM, CVA prepared the figures, FRV, AC and MK provided the funding. IM, MK, PJC-Y wrote the manuscript. All authors read, provided comments, and approved the manuscript.

Data availability

All data derived from this work are publicly available in NCBI-Genbank databases. All 16 MAGs assembled in this study have been deposited in the NCBI-Genbank database under Bioproject number PRJNA721863, biosample numbers SAMN32886101-SAMN32886116 and Genbank accession numbers JARIER000000000-JARIFG000000000. All these genomes were derived from Spanish lakes and reservoirs metagenomic datasets that were previously deposited under Bioproject number PRJNA721863 and SRA numbers SRR15198238- SRR15198275.

Competing interests

The authors declare there are no competing interests.

Consent for publication

All authors have read and commented on the manuscript, and have given consent for publication.

5. References

1. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y, Nakamura K. 2003. Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. Int J Syst Evol Microbiol 53:1155–1163.
2. DeBruyn JM, Fawaz MN, Peacock AD, Dunlap JR, Nixon LT, Cooper KE, Radosevich M. 2013. Gemmatirosa kalamazoonesis gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum Gemmatimonadetes. J Gen Appl Microbiol 59:305–312.
3. Pascual J, Foesel BU, Geppert A, Huber KJ, Boedeker C, Luckner M, Wanner G, Overmann J. 2018. Roseisolibacter agri gen. Nov., sp. nov., a novel slow-growing member of the under-represented phylum gemmatimonadetes. Int J Syst Evol Microbiol 68:1028–1036.
4. Pascual J, García-López M, Bills GF, Genilloud O. 2016. Longimicrobium terrae gen. Nov., sp. Nov., an oligotrophic bacterium of the under-represented phylum

gemmatimonadetes isolated through a system of miniaturized diffusion chambers. Int J Syst Evol Microbiol 66:1976–1985.

5. Zeng Y, Selyanin V, Lukeŝ M, Dean J, Kaftan D, Feng F, Koblížek M. 2015. Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium Gemmatimonas phototrophica sp. Nov., and emended descriptions of the genus Gemmatimonas and Gemmatimonas aurantiaca. Int J Syst Evol Microbiol 65:2410–2419.
6. Zeng Y, Nupur, Wu N, Madsen AM, Chen X, Gardiner AT, Koblížek M. 2021. Gemmatimonas groenlandica sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes. Front Microbiol 11:1–18.
7. Mujakić I, Piwosz K, Koblížek M. 2022. Phylum Gemmatimonadota and Its Role in the Environment. Microorganisms 10:151.
8. Shivaramu S, Tomasch J, Kopejtka K, Nupur N, Saini MK, Bokhari SNH, Küpper H, Koblížek M. 2022. The Influence of Calcium on the Growth, Morphology and Gene Regulation in Gemmatimonas phototrophica. Microorganisms 11:27.
9. Koblížek M, Dachev M, Bína D, Nupur, Piwosz K, Kaftan D. 2020. Utilization of light energy in phototrophic Gemmatimonadetes. J Photochem Photobiol B Biol 213:112085.
10. Qian P, Gardiner AT, Šímová I, Naydenova K, Croll TI, Jackson PJ, Kloz M, Čubáková P, Kuzma M, Zeng Y, Castro-hartmann P, Knippenberg B Van, Goldie KN, Kaftan D, Hrouzek P, Hájek J, Agirre J, Siebert CA, Bína D, Sader K, Stahlberg H, Sobotka R, Russo CJ, Polívka T, Hunter CN, Koblížek M. 2022. Phototrophica Photosystem 3139:1–11.
11. Zeng Y, Feng F, Medová H, Dean J, Koblížek M, Medova H, Dean J, Kobli ek M. 2014. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. Proc Natl Acad Sci U S A 111:7795–7800.
12. Mujakić I, Andrei A-Ş, Shabarova T, Fecskeová LK, Salcher MM, Piwosz K, Ghai R, Koblížek M. 2021. Common Presence of Phototrophic Gemmatimonadota in Temperate Freshwater Lakes. mSystems 6:e01241-20.
13. Zheng X, Dai X, Zhu Y, Yang J, Jiang H, Dong H, Huang L. 2022. (Meta)Genomic Analysis Reveals Diverse Energy Conservation Strategies Employed by Globally Distributed Gemmatimonadota. mSystems 7.
14. Zeng Y, Baumbach J, Barbosa EGV, Azevedo V, Zhang C, Koblížek M. 2016. Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. Environ Microbiol Rep 8:139–149.
15. Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol 72:1719–28.
16. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325.
17. Morrison JM, Baker KD, Zamor RM, Nikolai S, Elshahed MS, Youssef NH. 2017. Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). PLoS One

12:e0177488.

18. Villena-Alemany C, Mujakić I, Porcal P, Koblížek M, Piwosz K. 2022. Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. Environ Microbiol Rep 1–12.
19. Vavourakis CD, Andrei A-S, Mehrshad M, Ghai R, Sorokin DY, Muyzer G. 2018. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome 6:168.
20. Nunoura T, Hirai M, Yoshida-Takashima Y, Nishizawa M, Kawagucci S, Yokokawa T, Miyazaki J, Koide O, Makita H, Takaki Y, Sunamura M, Takai K. 2016. Distribution and niche separation of planktonic microbial communities in the water columns from the surface to the hadal waters of the Japan Trench under the Eutrophic Ocean. Front Microbiol 7:1–12.
21. Thiel V, Neulinger SC, Staufenberger T, Schmaljohann R, Imhoff JF. 2007. Spatial distribution of sponge-associated bacteria in the Mediterranean sponge Tethya aurantium. FEMS Microbiol Ecol 59:47–63.
22. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. ISME J 11:2465–2478.
23. Kato S, Nakawake M, Kita J, Yamanaka T, Utsumi M, Okamura K, Ishibashi JI, Ohkuma M, Yamagishi A. 2013. Characteristics of microbial communities in crustal fluids in a deep-sea hydrothermal field of the Suiyo Seamount. Front Microbiol 4:1–11.
24. Nunoura T, Nishizawa M, Hirai M, Shimamura S, Harnvoravongchai P, Koide O, Morono Y, Fukui T, Inagaki F, Miyazaki J, Takaki Y, Takai K. 2018. Microbial diversity in sediments from the bottom of the challenger deep, the mariana trench. Microbes Environ 33:186–194.
25. Cui G, Li J, Gao Z, Wang Y. 2019. Spatial variations of microbial communities in abyssal and hadal sediments across the Challenger Deep. PeerJ 7:1–17.
26. Peoples LM, Grammatopoulou E, Pombrol M, Xu X, Osuntokun O, Blanton J, Allen EE, Nunnally CC, Drazen JC, Mayor DJ, Bartlett DH. 2019. Microbial Community Diversity Within Sediments from Two Geographically Separated Hadal Trenches. Front Microbiol 10:347.
27. Vipindas P V., Mujeeb RKM, Jabir T, Thasneem TR, Mohamed Hatha AA. 2020. Diversity of sediment bacterial communities in the South Eastern Arabian Sea. Reg Stud Mar Sci 35:101153.
28. Vavourakis CD, Mehrshad M, Balkema C, Van Hall R, Andrei AŞ, Ghai R, Sorokin DY, Muyzer G. 2019. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. BMC Biol 17:1–20.
29. Zorz JK, Sharp C, Kleiner M, Gordon PMK, Pon RT, Dong X, Strous M. 2019. A shared core microbiome in soda lakes separated by large distances. Nat Commun 10:1–10.
30. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565.

31. Salcher MM, Schaefle D, Kaspar M, Neuenschwander SM, Ghai R. 2019. Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in Methylophilaceae. ISME J 13:2764–2777.
32. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245.
33. Thorpe HA, Bayliss SC, Hurst LD, Feil EJ. 2017. Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. Genetics 206:363–376.
34. Foerstner KU, von Mering C, Hooper SD, Bork P. 2005. Environments shape the nucleotide composition of genomes. EMBO Rep 6:1208–1213.
35. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, Gonźalez JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A 110:11463–11468.
36. McEwan CEA, Gatherer D, McEwan NR. 1998. Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. Hereditas 128.
37. Mann S, Chen YPP. 2010. Bacterial genomic G + C composition-eliciting environmental adaptation. Genomics 95:7–15.
38. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A 101:3160–3165.
39. Numberger D, Ganzert L, Zoccarato L, Mühldorfer K, Sauer S, Grossart HP, Greenwood AD. 2019. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. Sci Rep 9.
40. Kaas RS, Friis C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics 13.
41. Wolf YI, Makarova KS, Yutin N, Koonin E V. 2012. Updated clusters of orthologous genes for Archaea: A complex ancestor of the Archaea and the byways of horizontal gene transfer. Biol Direct 7:1–15.
42. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. 2020. Characterization of a sponge microbiome using an integrative genome-centric approach. ISME J 14:1100–1110.
43. Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA, Miller DJ, Forêt S, Ball E, Beeden R, Berumen M, Aranda M, Ravasi T, Bongaerts P, Hoegh-Guldberg O, Cooke I, Leggat B, Sprungala S, Fitzgerald A, Shang C, Lundgren P, Fyffe T, Rubino F, van Oppen M, Weynberg K, Foret S, Voolstra CR, Bourne DG, Tyson GW. 2019. A genomic view

of the reef-building coral Porites lutea and its microbial symbionts. Nat Microbiol 4:2090–2100.

44. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4.
45. Rashighi M, Harris JE. 2017. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Physiol Behav 176:139–148.
46. Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. 2011. The cytochrome bd respiratory oxygen reductases. Biochim Biophys Acta - Bioenerg 1807:1398–1413.
47. Engelberts JP, Robbins SJ, Herbold CW, Moeller FU, Jehmlich N, Laffy PW, Wagner M, Webster NS. 2023. Metabolic reconstruction of the near complete microbiome of the model sponge Ianthella basta. Environ Microbiol 25:646–660.
48. Kornberg HL. 1966. The role and control of the glyoxylate cycle in Escherichia coli. Biochem J 99:1–11.
49. Tang KH, Tang YJ, Blankenship RE. 2011. Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. Front Microbiol 2:1–23.
50. Zhao Y, Park RD, Muzzarelli RAA. 2010. Chitin deacetylases: Properties and applications. Mar Drugs 8:24–46.
51. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J. 2010. Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev.
52. He S, Stevens SLR, Chan L-K, Bertilsson S, Glavina del Rio T, Tringe SG, Malmstrom RR, Mcmahon KD. 2017. Ecophysiology of Freshwater Verrucomicrobia Inferred from. mSphere 2:e00277-17.
53. Wang L, Wise MJ. 2011. Glycogen with short average chain length enhances bacterial durability. Naturwissenschaften 98:719–729.
54. Wang L, Liu Q, Hu J, Asenso J, Wise MJ, Wu X, Ma C, Chen X, Yang J, Tang D. 2019. Structure and evolution of glycogen branching enzyme N-termini from bacteria. Front Microbiol 10:1–12.
55. Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. Philos Trans R Soc B Biol Sci 363:2629–2640.
56. Cabello-Yeves PJ, Scanlan DJ, Callieri C, Picazo A, Schallenberg L, Huber P, Roda-Garcia JJ, Bartosiewicz M, Belykh OI, Tikhonova I V., Torcello-Requena A, De Prado PM, Millard AD, Camacho A, Rodriguez-Valera F, Puxty RJ. 2022. α-cyanobacteria possessing form IA RuBisCO globally dominate aquatic habitats. ISME J https://doi.org/10.1038/s41396-022-01282-z.
57. Alfreider A, Tartarotti B. 2019. Spatiotemporal dynamics of different CO2 fixation strategies used by prokaryotes in a dimictic lake. Sci Rep 9:1–10.
58. Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about

Rubisco evolution and structure/function relationships. J Exp Bot 59:1515–1524.

59. Hanson TE, Tabita FR. 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci U S A 98:4397–4402.
60. Ashida H, Saito Y, Kojima C, Kobayashi K, Ogasawara N, Yokota A. 2003. A functional link between RuBisCO-like protein of Bacillus and photosynthetic RuBisCO. Science 302:286–290.
61. Cordero PRF, Bayly K, Man Leung P, Huang C, Islam ZF, Schittenhelm RB, King GM, Greening C. 2019. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. ISME J 13:2868–2881.
62. Islam ZF, Cordero PRF, Feng J, Chen YJ, Bay SK, Jirapanjawat T, Gleadow RM, Carere CR, Stott MB, Chiri E, Greening C. 2019. Two Chloroflexi classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide. ISME J 13.
63. King GM, Weber CF. 2007. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat Rev Microbiol.
64. Meier D V., Imminger S, Gillor O, Woebken D. 2021. Distribution of Mixotrophy and Desiccation Survival Mechanisms across Microbial Genomes in an Arid Biological Soil Crust Community. mSystems 6.
65. Koblížek M. 2015. Ecology of aerobic anoxygenic phototrophs in aquatic environments. FEMS Microbiol Rev 39:854–870.
66. Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, Delong EF. 2018. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. ISME J 12.
67. DeLong EF, Béjà O. 2010. The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. PLoS Biol 8.
68. Zeng Y, Chen X, Madsen AM, Zervas A, Nielsen K, Andrei AS, Chresten L, Hansen L, Liu Y, Hansen LH. 2020. Potential Rhodopsin and Bacteriochlorophyll - Based Dual Phototrophy in a High Arctic Glacier.
69. Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C, Rodríguez G, Massol-Deyá A, Krishnani KK, Ritalahti KM, Nissen S, Konstantinidis KT, Löffler FE. 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proc Natl Acad Sci U S A 109:19709–19714.
70. Yoon S, Nissen S, Park D, Sanford RA, Löffler E. 2016. Clade I NosZ from Those Harboring Clade II NosZ. Appl Envioronmental Microbiol 82:3793–3800.
71. Chee-Sanford J, Tian D, Sanford R. 2019. Consumption of N2O and other N-cycle intermediates by Gemmatimonas aurantiaca strain T-27. Microbiology 165:1345–1354.
72. Zumft WG. 1997. Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev 61:533–616.
73. Graf DRH, Jones CM, Hallin S. 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of

community structure for N2O emissions. PLoS One 9:1–20.

74. Orellana LH, Rodriguez-R LM, Higgins S, Chee-Sanford JC, Sanford RA, Ritalahti KM, Löffler FE, Konstantinidis KT. 2014. Detecting nitrous oxide reductase (nosZ) genes in soil metagenomes: Method development and implications for the nitrogen cycle. MBio 5.
75. Jones CM, Graf DRH, Bru D, Philippot L, Hallin S. 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: A potential nitrous oxide sink. ISME J 7:417–426.
76. Jones CM, Spor A, Brennan FP, Breuil MC, Bru D, Lemanceau P, Griffiths B, Hallin S, Philippot L. 2014. Recently identified microbial guild mediates soil N2O sink capacity. Nat Clim Chang 4.
77. Park D, Kim H, Yoon S. 2017. Nitrous Oxide Reduction by an Obligate. Appl Environ Microbiol 83:1–12.
78. Leustek T, Martin MN, Bick JA, Davies JP. 2000. Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. Annu Rev Plant Biol 51:141–165.
79. Grein F, Ramos AR, Venceslau SS, Pereira IAC. 2013. Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. Biochim Biophys Acta - Bioenerg 1827:145–160.
80. Villarreal-Chiu JF, Quinn JP, McGrath JW. 2012. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. Front Microbiol 3.
81. Santos-Beneit F. 2015. The Pho regulon: A huge regulatory network in bacteria. Front Microbiol 6:1–13.
82. Luo H, Benner R, Long RA, Hu J. 2009. Subcellular localization of marine bacterial alkaline phosphatases. Proc Natl Acad Sci U S A 106:21219–21223.
83. Yurkov V, Csotonyi JT. 2009. New Light on Aerobic Anoxygenic Phototrophs. Hunt CN, Daldal F, Thurnauer MC, Beatty JT Purple Phototrophic Bact Adv Photosynth Respir vol 28 Springer, Dordrecht 31–55.
84. Mendel RR. 2013. The molybdenum cofactor. J Biol Chem.
85. Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U. 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297:105–107.
86. Tong L. 2013. Structure and function of biotin-dependent carboxylases. Cell Mol Life Sci.
87. Kurnasov O, Goral V, Colabroy K, Gerdes S, Anantha S, Osterman A, Begley TP. 2003. NAD Biosynthesis: Identification of the Tryptophan to Quinolinate Pathway in Bacteria. Chem Biol 10.
88. Cadieux N, Kadner RJ. 1999. Site-directed disulfide bonding reveals an interaction site between energy-coupling protein TonB and BtuB, the outer membrane cobalamin transporter. Proc Natl Acad Sci U S A 96:10673–10678.
89. Rojo F. 2009. Degradation of alkanes by bacteria. Environ Microbiol 11:2477–

2490.

90. Appolinario LR, Tschoeke D, Paixão RVS, Venas T, Calegario G, Leomil L, Silva BS, Thompson CC, Thompson FL. 2019. Metagenomics sheds light on the metabolic repertoire of oil-biodegrading microbes of the South Atlantic Ocean. Environ Pollut 249:295–304.
91. Harwood CS, Burchhardt G, Herrmann H, Fuchs G. 1998. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. FEMS Microbiol Rev 22:439–458.
92. Dinglasan-Panlilio MJ, Dworatzek S, Mabury S, Edwards E. 2006. Microbial oxidation of 1,2-dichloroethane under anoxic conditions with nitrate as electron acceptor in mixed and pure cultures. FEMS Microbiol Ecol 56.
93. Paver SF, Kent AD. 2010. Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate with Phytoplankton Population Dynamics in Humic Lakes. Microb Ecol 60:406–418.
94. Smith DJ, Kharbush JJ, Kersten RD, Dick GJ. 2022. Uptake of Phytoplankton-Derived Carbon and Cobalamins by Novel Acidobacteria Genera in Microcystis Blooms Inferred from Metagenomic and Metatranscriptomic Evidence. Appl Environ Microbiol 88.
95. Yao S, Lyu S, An Y, Lu J, Gjermansen C, Schramm A. 2019. Microalgae–bacteria symbiosis in microalgal growth and biofuel production: a review. J Appl Microbiol 126:359–368.
96. Pellicer MT, Badía J, Aguilar J, Baldomà L. 1996. glc Locus of Escherichia coli: Characterization of genes encoding the subunits of glycolate oxidase and the glc regulator protein. J Bacteriol 178:2051–2059.
97. Chiriac MC, Haber M, Salcher MM. 2022. Adaptive genetic traits in pelagic freshwater microbes. Environ Microbiol 1–36.
98. Sleator RD, Hill C. 2002. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. FEMS Microbiol Rev 26.
99. Pivetti CD, Yen M-R, Miller S, Busch W, Tseng Y-H, Booth IR, Saier, MH. 2003. Two Families of Mechanosensitive Channel Proteins. Microbiol Mol Biol Rev 67.
100. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428.
101. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019:1–13.
102. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.
103. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927.
104. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: A tool for fast and

accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868.

105. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696–7701.

106. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:206–214.

107. Anderson MJ. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62:245–253.

108. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:1–10.

109. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

110. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

111. Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

112. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

113. Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066.

114. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780.

115. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, da Silva A, Denny P, Dogan T, Ebenezer TG, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuche B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C,

Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J, Ruch P, Teodoro D. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489.

116. Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. 2007. Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. Microbiol Mol Biol Rev 71:576–599.

117. Wrighton KC, Castelle CJ, Varaljay VA, Satagopan S, Brown CT, Wilkins MJ, Thomas BC, Sharon I, Williams KH, Tabita FR, Banfield JF. 2016. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. ISME J 10:2702–2714.

118. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–W235.

119. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

120. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

121. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–731.

4. Conclusion and future prospects

While data about the presence of Gemmatimonadota in different environments has been continuously growing, information about their ecology, distribution and diversity in freshwaters remained limited. This left a question whether Gemmatimonadota sequences previously identified in lakes truly originate from limnic species, or they are passively transferred to lakes with runoff waters from surrounding soil or re-suspended from sediments.

In this work, we used metagenomic data to confirm Gemmatimonadota persistence in freshwater lakes. We also showed their ubiquitous nature and ability to adapt to aquatic environments with varying trophic statuses. Furthermore, we documented their abundances in several lakes, as well as temporal changes during one year in a meso-oligotrophic lake. The 45 assembled MAGs represented, at that time, the largest collection of freshwater MAGs of this group. With it, we showed high genomic diversity of both heterotrophic and photoheterotrophic members of this phylum. Finally, by using CARD-FISH, we have for the first time visualized the cells of Gemmatimonadota in their natural environment, and even observed their association with other organisms.

As the data related to Gemmatimonadota, especially knowledge of their environmental distribution, has increased significantly, we summarized all available information in a short review. This emphasized that still their metabolism and environmental roles remain unexplored and led us to analyze their metabolic capabilities and genomic characteristics in dependence on the environment they inhabit. This study showed Gemmatimonadota have flexible metabolism and the ability to adapt to varying environmental conditions, and provided metabolic characterization of Gemmatimonadota that was previously missing.

I believe that all of the information obtained during my PhD study has expanded the knowledge about this interesting bacterial phylum and provided the basis for future studies and objectives. They could include, for instance, an in-depth analysis of the evolution of photosynthesis in Gemmatimonadota and HGT, culturing of new members from environments like soda lakes, which may have the capability for carbon fixation, or a bigger focus on environments where Gemmatimonadota remain unexplored, such as marine waters, biofilms or host-associated.

5. References

Amann, R., and Fuchs, B. M. (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* 6, 339–348. doi:10.1038/nrmicro1888.

Amann, R., Fuchs, B. M., and Behrens, S. (2001). The identification of microorganisms by fluorescence in situ hybridisation. *Curr. Opin. Biotechnol.* 12. doi:10.1016/S0958-1669(00)00204-4.

Amann, R. I., Krumholz, L., and Stahl, D. A. (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* 172. doi:10.1128/jb.172.2.762-770.1990.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169. doi:10.1128/mmbr.59.1.143-169.1995.

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7, 1–11. doi:10.1038/ncomms13219.

Arora-Williams, K., Olesen, S. W., Scandella, B. P., Delwiche, K., Spencer, S. J., Myers, E. M., et al. (2018). Dynamics of microbial populations mediating biogeochemical cycling in a freshwater lake. *Microbiome* 6, 1–16. doi:10.1186/s40168-018-0556-7.

Austin, B. (2017). The value of cultures to modern microbiology. *Antonie Van Leeuwenhoek* 110, 1247–1256. doi:10.1007/s10482-017-0840-8.

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi:10.3354/meps010257.

Bar-On, Y. M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. *Proc. Natl. Acad. Sci.* 115, 6506–6511. doi:10.1073/pnas.1711842115.

Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., and Tranvik, L. J. (2009). The boundless carbon cycle. *Nat. Geosci.* 2, 598–600. doi:10.1038/ngeo618.

Bauer, C. E., Setterdahl, A., Wu, J., and Robinson, B. R. (2009). "Regulation of Gene Expression in Response to Oxygen Tension," in *Springer*, 707–725. doi:10.1007/978-1-4020-8815-5_35.

Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., Stein, H. J., Hannah, J. L., et al. (2004). Dating the rise of atmospheric oxygen. *Nature* 427. doi:10.1038/nature02260.

Beman, J. M., Popp, B. N., and Francis, C. A. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *ISME J.* 2, 429–441. doi:10.1038/ismej.2007.118.

Berggren, M., Lapierre, J.-F., and del Giorgio, P. A. (2012). Magnitude and regulation

of bacterioplankton respiratory quotient across freshwater environmental gradients. *ISME J.* 6, 984–993. doi:10.1038/ismej.2011.157.

Bertaux, J., Gloger, U., Schmid, M., Hartmann, A., and Scheu, S. (2007). Routine fluorescence in situ hybridization in soil. *J. Microbiol. Methods* 69. doi:10.1016/j.mimet.2007.02.012.

Brinkmann, H., Göker, M., Koblížek, M., Wagner-Döbler, I., and Petersen, J. (2018). Horizontal operon transfer, plasmids, and the evolution of photosynthesis in Rhodobacteraceae. *ISME J.* 12, 1994–2010. doi:10.1038/s41396-018-0150-9.

Bristow, L. A., Mohr, W., Ahmerkamp, S., and Kuypers, M. M. M. (2017). Nutrients that limit growth in the ocean. *Curr. Biol.* 27, R474–R478. doi:10.1016/j.cub.2017.03.030.

Bryant, D. A., and Liu, Z. (2013). *Green Bacteria. Insights into Green Bacterial Evolution through Genomic Analyses.* Elsevier doi:10.1016/B978-0-12-397923-0.00004-7.

Cabello-Yeves, P. J., Ghai, R., Mehrshad, M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2017). Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. *Front. Microbiol.* 8. doi:10.3389/fmicb.2017.02131.

Cabello-Yeves, P. J., Zemskaya, T. I., Rosselli, R., Coutinho, F. H., Zakharenko, A. S., Blinov, V. V., et al. (2018). Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Appl. Environ. Microbiol.* 84, e02132-17. doi:10.1128/AEM.02132-17.

Cabello-Yeves, P. J., Zemskaya, T. I., Zakharenko, A. S., Sakirko, M. V., Ivanov, V. G., Ghai, R., et al. (2020). Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnol. Oceanogr.* 65, 1471–1488. doi:10.1002/lno.11401.

Cardona, T. (2015). A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth. Res.* 126, 111–134. doi:10.1007/s11120-014-0065-x.

Cardona, T. (2016). Origin of bacteriochlorophyll a and the early diversification of photosynthesis. *PLoS One* 11, 1–9. doi:10.1371/journal.pone.0151250.

Cepáková, Z., Hrouzek, P., Žišková, E., Nuyanzina-Boldareva, E., Šorf, M., Kozlíková-Zapomělová, E., et al. (2016). High turnover rates of aerobic anoxygenic phototrophs in European freshwater lakes. *Environ. Microbiol.* 18, 5063–5071. doi:10.1111/1462-2920.13475.

Chen, G., Bai, R., Zhang, Y., Zhao, B., and Xiao, Y. (2022). Application of metagenomics to biological wastewater treatment. *Sci. Total Environ.* 807, 150737. doi:10.1016/j.scitotenv.2021.150737.

Chen, P., Zhou, H., Huang, Y., Xie, Z., Zhang, M., Wei, Y., et al. (2021). Revealing the full biosphere structure and versatile metabolic functions in the deepest ocean sediment of the Challenger Deep. *Genome Biol.* 22, 1–28. doi:10.1186/s13059-021-02408-w.

Christensen, H., Hansen, M., and Sørensen, J. (1999). Counting and size classification

of active soil bacteria by fluorescence in situ hybridization with an rRNA oligonucleotide probe. *Appl. Environ. Microbiol.* 65, 1753–1761. doi:10.1128/aem.65.4.1753-1761.1999.

Chróst, R. J., Tomasz, A., Kalinowska, K., and Skowrońska, A. (2009). Abundance and structure of microbial loop components (bacteria and protists) in lakes of different trophic status. *J. Microbiol. Biotechnol.* 19, 858–868. doi:10.4014/jmb.0812.651.

Cole, J. J. (1982). Interactions between bacteria and algae in aquatic ecosystems. *Ann Rev. Ecol. Syst.* 13, 291–314.

Cole, J. J., Likens, G. E., and Strayer, D. L. (1982). Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. *Limnol. Oceanogr.* 27, 1080–1090. doi:10.4319/lo.1982.27.6.1080.

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., et al. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10, 171–184. doi:10.1007/s10021-006-9013-8.

Cottrell, M. T., and Kirchman, D. L. (2000). Community Composition of Marine Bacterioplankton Determined by 16S rRNA Gene Clone Libraries and Fluorescence In Situ Hybridization. 66, 5116–5122.

Coyne, K. J., Parker, A. E., Lee, C. K., Sohm, J. A., Kalmbach, A., Gunderson, T., et al. (2020). The distribution and relative ecological roles of autotrophic and heterotrophic diazotrophs in the McMurdo Dry Valleys, Antarctica. *FEMS Microbiol. Ecol.* 96, 1–16. doi:10.1093/femsec/fiaa010.

Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C., and Banfield, J. F. (2018). Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558, 440–444. doi:10.1038/s41586-018-0207-y.

Čuperová, Z., Holzer, E., Salka, I., Sommarug, R., and Koblížek, M. (2013). Temporal changes and altitudinal distribution of aerobic anoxygenic phototrophs in mountain lakes. *Appl. Environ. Microbiol.* 79. doi:10.1128/AEM.01526-13.

Dachev, M., Bína, D., Sobotka, R., Moravcová, L., Gardian, Z., Kaftan, D., et al. (2017). Unique double concentric ring organization of light harvesting complexes in Gemmatimonas phototrophica. *PLoS Biol.* 15, 1–16. doi:10.1371/journal.pbio.2003943.

DeBruyn, J. M., Fawaz, M. N., Peacock, A. D., Dunlap, J. R., Nixon, L. T., Cooper, K. E., et al. (2013). Gemmatirosa kalamazoonesis gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum Gemmatimonadetes. *J. Gen. Appl. Microbiol.* 59, 305–312. doi:10.2323/jgam.59.305.

DeBruyn, J. M., Nixon, L. T., Fawaz, M. N., Johnson, A. M., and Radosevich, M. (2011). Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl. Environ. Microbiol.* 77, 6295–6300. doi:10.1128/AEM.05005-11.

Dekas, A. E., Poretsky, R. S., and Orphan, V. J. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* 326, 422–426. doi:10.1126/science.1178223.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., et al. (2018). A global atlas of the dominant bacteria found in soil. *Science* 359, 320–325. doi:10.1126/science.aap9516.

Delgado-Baquerizo, M., Reich, P. B., Khachane, A. N., Campbell, C. D., Thomas, N., Freitag, T. E., et al. (2017). It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environ. Microbiol.* 19, 1176–1188. doi:10.1111/1462-2920.13642.

Dodds, W. K., and Whiles, M. R. (2020). "Freshwater Ecosystems," in *Freshwater Ecology* (Elsevier), 723–764. doi:10.1016/B978-0-12-813255-5.00024-7.

Dunlap, P. V. (2001). Microbial Diversity. *Encycl. Biodivers. Second Ed.*, 280–291. doi:10.1016/B978-0-12-384719-5.00435-4.

Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W., and Bock, E. (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, Nitrospira moscoviensis sp. nov. and its phylogenetic relationship. *Arch. Microbiol.* 164, 16–23. doi:10.1007/s002030050230.

Eickhorst, T., and Tippkötter, R. (2008). Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biol. Biochem.* 40, 1883–1891. doi:10.1016/j.soilbio.2008.03.024.

Escobar-Zepeda, A., De León, A. V. P., and Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front. Genet.* 6, 1–15. doi:10.3389/fgene.2015.00348.

Falkowski, P., Barber, R., and Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* 281, 200–207. doi:10.1126/science.281.5374.200.

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* 320, 1034–1039. doi:10.1126/science.1153213.

Fauteux, L., Cottrell, M. T., Kirchman, D. L., Borrego, C. M., Garcia-Chaves, M. C., and del Giorgio, P. A. (2015). Patterns in abundance, cell size and pigment content of aerobic anoxygenic phototrophic bacteria along environmental gradients in northern lakes. *PLoS One* 10, 1–17. doi:10.1371/journal.pone.0124035.

Fecskeová, L. K., Piwosz, K., Hanusová, M., Nedoma, J., Znachor, P., and Koblížek, M. (2019). Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-019-55210-x.

Ferrera, I., Gasol, J. M., Sebastián, M., Hojerová, E., and Kobížek, M. (2011). Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal mediterranean waters. *Appl. Environ. Microbiol.* 77, 7451–7458. doi:10.1128/AEM.00208-11.

Ferrera, I., Sarmento, H., Priscu, J. C., Chiuchiolo, A., González, J. M., and Grossart, H.-P. (2017). Diversity and Distribution of Freshwater Aerobic Anoxygenic Phototrophic Bacteria across a Wide Latitudinal Gradient. *Front. Microbiol.* 8.

doi:10.3389/fmicb.2017.00175.

Frank, J. A., Pan, Y., Tooming-Klunderud, A., Eijsink, V. G. H., McHardy, A. C., Nederbragt, A. J., et al. (2016). Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Sci. Rep.* 6, 1–10. doi:10.1038/srep25373.

French, E., Kozlowski, J. A., Mukherjee, M., Bullerjahn, G., and Bollmann, A. (2012). Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. *Appl. Environ. Microbiol.* 78, 5773–5780. doi:10.1128/AEM.00432-12.

Garcia-Chaves, M. C., Cottrell, M. T., Kirchman, D. L., Ruiz-González, C., and del Giorgio, P. A. (2016). Single-cell activity of freshwater aerobic anoxygenic phototrophic bacteria and their contribution to biomass production. *ISME J.* 10, 1579–1588. doi:10.1038/ismej.2015.242.

Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2013). Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* 3. doi:10.1038/srep02471.

Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2014). Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol. Ecol.* 23, 6073–6090. doi:10.1111/mec.12985.

Gieseke, A., Bjerrum, L., Wagner, M., and Amann, R. (2003). Structure and activity of multiple nitrifying bacterial populations co-existing in a biofilm. *Environ. Microbiol.* 5. doi:10.1046/j.1462-2920.2003.00423.x.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345, 60–63. doi:10.1038/345060a0.

Glöckner, F. O., Fuchs, B. M., and Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: A first comparison based on fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 65, 3721–3726. doi:10.1128/aem.65.8.3721-3726.1999.

Haro-Moreno, J. M., López-Pérez, M., de la Torre, J. R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6, 1–19. doi:10.1186/s40168-018-0513-5.

Hauruseu, D., and Koblížek, M. (2012). Influence of Light on Carbon Utilization in Aerobic Anoxygenic Phototrophs. *Appl. Environ. Microbiol.* 78, 7414–7419. doi:10.1128/AEM.01747-12.

He, S., Stevens, S. L. R., Chan, L.-K., Bertilsson, S., Glavina del Rio, T., Tringe, S. G., et al. (2017). Ecophysiology of Freshwater Verrucomicrobia Inferred from. *mSphere* 2, e00277-17. doi:https://doi.org/10.1128/mSphere.00277-17.

Herlemann, D. P. R., Woelk, J., Labrenz, M., and Jürgens, K. (2014). Diversity and abundance of "Pelagibacterales" (SAR11) in the Baltic Sea salinity gradient. *Syst. Appl. Microbiol.* 37. doi:10.1016/j.syapm.2014.09.002.

Hobbie, J. E., and Hobbie, E. A. (2013). Microbes in nature are limited by carbon and energy: The starving-survival lifestyle in soil and consequences for estimating microbial rates. *Front. Microbiol.* 4, 1–11. doi:10.3389/fmicb.2013.00324.

Hohmann-Marriott, M. F., and Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* 62, 515–548. doi:10.1146/annurev-arplant-042110-103811.

Hojerová, E., Mašín, M., Brunet, C., Ferrera, I., Gasol, J. M., and Koblížek, M. (2011). Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ. Microbiol.* 13, 2717–2725. doi:10.1111/j.1462-2920.2011.02540.x.

Holland-Moritz, H., Stuart, J., Lewis, L. R., Miller, S., Mack, M. C., McDaniel, S. F., et al. (2018). Novel bacterial lineages associated with boreal moss species. *Environ. Microbiol.* 20, 2625–2638. doi:10.1111/1462-2920.14288.

Holland, H. D. (2006). The oxygenation of the atmosphere and oceans. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 903–915. doi:10.1098/rstb.2006.1838.

Hooper, S. D., Raes, J., Foerstner, K. U., Harrington, E. D., Dalevi, D., and Bork, P. (2008). A molecular study of microbe transfer between distant environments. *PLoS One* 3, 1–6. doi:10.1371/journal.pone.0002607.

Howard, J. B., and Rees, D. C. (1996). Structural basis of biological nitrHoward, James B., and Douglas C. Rees. 1996. "Structural Basis of Biological Nitrogen Fixation." Chemical Reviews 96(7): 2965–82.ogen fixation. *Chem. Rev.* 96, 2965–2982.

Huang, Y., Zeng, Y., Lu, H., Feng, H., Zeng, Y., and Koblížek, M. (2016). Novel acsF Gene Primers Revealed a Diverse Phototrophic Bacterial Population, Including Gemmatimonadetes, in Lake Taihu (China). *Appl. Environ. Microbiol.* 82, 5587–5594. doi:10.1128/AEM.01063-16.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016a). A new view of the tree of life. *Nat. Microbiol.* 1, 1–6. doi:10.1038/nmicrobiol.2016.48.

Hug, L. A., Castelle, C. J., Wrighton, K. C., Thomas, B. C., Sharon, I., Frischkorn, K. R., et al. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 1, 1–17. doi:10.1186/2049-2618-1-22.

Hug, L. A., Thomas, B. C., Sharon, I., Brown, C. T., Sharma, R., Hettich, R. L., et al. (2016b). Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ. Microbiol.* 18, 159–173. doi:10.1111/1462-2920.12930.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3, 1–8. doi:10.1186/gb-2002-3-2-reviews0003.

Hugenholtz, P., Goebel, B. M., and Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180, 6793.

Hugenholtz, P., Tyson, G. W., Webb, R. I., Wagner, A. M., and Blackall, L. L. (2001).

Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria, with no known pure-culture representatives. *Appl. Environ. Microbiol.* 67, 411–419. doi:10.1128/AEM.67.1.411-419.2001.

Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., and Nagashima, K. V. P. (2001). Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.* 52, 333–341. doi:10.1007/s002390010163.

Ishii, K., Mußmann, M., MacGregor, B. J., and Amann, R. (2004). An improved fluorescence in situ hybridization protocol for the identification of bacteria and archaea in marine sediments. *FEMS Microbiol. Ecol.* 50, 203–213. doi:10.1016/j.femsec.2004.06.015.

Jannasch, H. W., and Jones, G. E. (1959). Bacterial Populations in Sea Water as Determined by Different Methods of Enumeration1. *Limnol. Oceanogr.* 4, 128–139. doi:10.4319/lo.1959.4.2.0128.

Jugnia, L. B., Tadonléké, R. D., Sime-Ngando, T., Foto, S. M., and Kemka, N. (1998). Short-term variations in the abundance and cell volume of bacterioplankton in an artificial tropical lake. *Hydrobiologia* 385, 113–119. doi:10.1023/a:1003453709384.

Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* 39, 854–870. doi:10.1093/femsre/fuv032.

Koblížek, M., Dachev, M., Bína, D., Nupur, Piwosz, K., and Kaftan, D. (2020). Utilization of light energy in phototrophic Gemmatimonadetes. *J. Photochem. Photobiol. B Biol.* 213, 112085. doi:10.1016/j.jphotobiol.2020.112085.

Koblížek, M., Mašín, M., Ras, J., Poulton, A. J., and Prášil, O. (2007). Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ. Microbiol.* 9, 2401–2406. doi:10.1111/j.1462-2920.2007.01354.x.

Koblížek, M., Stoń-Egiert, J., Sagan, S., and Kolber, Z. S. (2005). Diel changes in bacteriochlorophyll a concentration suggest rapid bacterioplankton cycling in the Baltic Sea. *FEMS Microbiol. Ecol.* 51, 353–361. doi:10.1016/j.femsec.2004.09.016.

Koblížek, M., Zeng, Y., Horák, A., and Oborník, M. (2013). *Regressive Evolution of Photosynthesis in the Roseobacter Clade*. Elsevier doi:10.1016/B978-0-12-397923-0.00013-8.

Kolářová, E., Medová, H., Piwosz, K., and Koblížek, M. (2019). Seasonal dynamics of aerobic anoxygenic phototrophs in freshwater lake Vlkov. *Folia Microbiol. (Praha).* 64, 705–710. doi:10.1007/s12223-019-00735-x.

Kolber, Z. S., Plumley, F. G., Lang, A. S., Beatty, J. T., Blankenship, R. E., VanDover, C. L., et al. (2001). Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* 292, 2492–2495. doi:10.1126/science.1059707.

Könneke, M., Bernhard, A. E., De La Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546. doi:10.1038/nature03911.

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi:10.1038/ismej.2017.113.

Kopejtka, K., Tomasch, J., Zeng, Y., Tichý, M., Sorokin, D. Y., and Koblížek, M. (2017). Genomic analysis of the evolution of phototrophy among haloalkaliphilic rhodobacterales. *Genome Biol. Evol.* 9, 1950–1962. doi:10.1093/gbe/evx141.

Kowalchuk, G. A., and Stephen, J. R. (2001). Ammonia-oxidizing bacteria: A model for molecular microbial ecology. *Annu. Rev. Microbiol.* 55, 485–529. doi:10.1146/annurev.micro.55.1.485.

Kubota, K. (2013). CARD-FISH for environmental microorganisms: Technical advancement and future applications. *Microbes Environ.* 28, 3–12. doi:10.1264/jsme2.ME12107.

Kuypers, M. M. M., Lavik, G., Woebken, D., Schmid, M., Fuchs, B. M., Amann, R., et al. (2005). Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 6478–6483. doi:10.1073/pnas.0502088102.

Kysela, D. T., Randich, A. M., Caccamo, P. D., and Brun, Y. V. (2016). Diversity Takes Shape: Understanding the Mechanistic and Adaptive Basis of Bacterial Morphology. *PLoS Biol.* 14, 1–15. doi:10.1371/journal.pbio.1002565.

Lawson, C. E., Wu, S., Bhattacharjee, A. S., Hamilton, J. J., McMahon, K. D., Goel, R., et al. (2017). Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Commun.* 8, 1–12. doi:10.1038/ncomms15416.

LeBauer, D. S., and Treseder, K. K. (2008). Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* 89, 371–379. doi:10.1890/06-2057.1.

Lee, J. Z., Craig Everroad, R., Karaoz, U., Detweiler, A. M., Pett-Ridge, J., Weber, P. K., et al. (2018). Metagenomics reveals niche partitioning within the phototrophic zone of a microbial mat. *PLoS One* 13, 1–19. doi:10.1371/journal.pone.0202792.

Lehtovirta-Morley, L. E. (2018). Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiol. Lett.* 365, 1–9. doi:10.1093/femsle/fny058.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809. doi:10.1038/nature04983.

Li, L., Kato, C., and Horikoshi, K. (1999). Bacterial diversity in deep-sea sediments from different depths. *Biodivers. Conserv.* 8, 659–677. doi:10.1023/A:1008848203739.

Linz, A. M., He, S., Stevens, S. L. R., Anantharaman, K., Rohwer, R. R., Malmstrom, R. R., et al. (2018). Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* 2018, 1–24. doi:10.7717/peerj.6075.

Llobet-Brossa, E., Rabus, R., Böttcher, M. E., Könneke, M., Finke, N., Schramm, A.,

et al. (2002). Community structure and activity of sulfate-reducing bacteria in an intertidal surface sediment: A multi-method approach. *Aquat. Microb. Ecol.* 29. doi:10.3354/ame029211.

Llobet-Brossa, E., Rosselló-Mora, R., and Amann, R. (1998). Microbial community composition of wadden sea sediments as revealed by fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 64. doi:10.1128/aem.64.7.2691-2696.1998.

Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi:10.1038/s41576-020-0236-x.

Lombard, N., Prestat, E., van Elsas, J. D., and Simonet, P. (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol.* 78, 31–49. doi:10.1111/j.1574-6941.2011.01140.x.

Ludwig, W., Bauer, S. H., Bauer, M., Held, I., Kirchhof, G., Schulze, R., et al. (1997). Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol. Lett.* 153, 181–190. doi:10.1016/S0378-1097(97)00256-5.

Lyons, T. W., Reinhard, C. T., and Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506, 307–315. doi:10.1038/nature13068.

Maberly, S. C., Barker, P. A., Stott, A. W., and De Ville, M. M. (2013). Catchment productivity controls CO2 emissions from lakes. *Nat. Clim. Chang.* 3, 391–394. doi:10.1038/nclimate1748.

Madigan, T., Martinko, J., Stahl, D., and Clark, D. (2010). *Brock Biology of Microorganisms (13th Edition)*. 13th editi. Pearson Education, Inc., publishing as Benjamin Cummings.

Madrid, V. M., Aller, J. Y., Aller, R. C., and Chistoserdov, A. Y. (2001). High prokaryote diversity and analysis of community structure in mobile mud deposits off French Guiana: Identification of two new bacterial candidate divisions. *FEMS Microbiol. Ecol.* 37, 197–209. doi:10.1016/S0168-6496(01)00165-9.

Maixner, F., Noguera, D. R., Anneser, B., Stoecker, K., Wegl, G., Wagner, M., et al. (2006). Nitrite concentration influences the population structure of Nitrospira-like bacteria. *Environ. Microbiol.* 8. doi:10.1111/j.1462-2920.2006.01033.x.

Man, M., Nedoma, J., Pechar, L., and Koblek, M. (2008). Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environ. Microbiol.* 10, 1988–1996. doi:10.1111/j.1462-2920.2008.01615.x.

Marx, V. (2023). Method of the year: long-read sequencing. *Nat. Methods* 20, 6–11. doi:10.1038/s41592-022-01730-w.

Mašín, M., Čuperová, Z., Hojerová, E., Salka, I., Grossart, H., and Koblížek, M. (2012). Distribution of aerobic anoxygenic phototrophic bacteria in glacial lakes of northern Europe. *Aquat. Microb. Ecol.* 66, 77–86. doi:10.3354/ame01558.

Mehrshad, M., Salcher, M. M., Okazaki, Y., Nakano, S., Šimek, K., Andrei, A.-S., et al. (2018). Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* 6, 176. doi:10.1186/s40168-018-0563-8.

Meier, D. V., Pjevac, P., Bach, W., Hourdez, S., Girguis, P. R., Vidoudez, C., et al. (2017). Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J.* 11, 1545–1558. doi:10.1038/ismej.2017.37.

Mendes Monteiro, J., Vogwill, R., Bischoff, K., and Gleeson, D. B. (2019). Comparative metagenomics of microbial mats from hypersaline lakes at Rottnest Island (WA, Australia), advancing our understanding of the effect of mat community and functional genes on microbialite accretion. *Limnol. Oceanogr.*, 1–17. doi:10.1002/lno.11323.

Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., and Doxey, A. C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* 47, 4442–4448. doi:10.1093/nar/gkz246.

Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., and Capone, D. G. (2004). High rates of N2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* 430, 1027–1031. doi:10.1038/nature02824.

Morrison, J. M., Baker, K. D., Zamor, R. M., Nikolai, S., Elshahed, M. S., and Youssef, N. H. (2017). Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLoS One* 12, e0177488. doi:10.1371/journal.pone.0177488.

Moter, A., and Göbel, U. B. (2000). Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *J. Microbiol. Methods* 41, 85–112. doi:10.1016/S0167-7012(00)00152-4.

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

Mummey, D. L., and Stahl, P. D. (2003). Candidate division BD: Phylogeny, distribution and abundance in soil ecosystems. *Syst. Appl. Microbiol.* 26, 228–235. doi:10.1078/072320203322346074.

Musat, N., Werner, U., Knittel, K., Kolb, S., Dodenhof, T., van Beusekom, J. E. E., et al. (2006). Microbial community structure of sandy intertidal sediments in the North Sea, Sylt-Rømø Basin, Wadden Sea. *Syst. Appl. Microbiol.* 29, 333–348. doi:10.1016/j.syapm.2005.12.006.

Nagashima, S., and Nagashima, K. V. P. (2013). "Comparison of Photosynthesis Gene Clusters Retrieved from Total Genome Sequences of Purple Bacteria," in *Advances in Botanical Research* (Elsevier), 151–178. doi:10.1016/B978-0-12-397923-0.00005-9.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., et al. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiol. Mol. Biol. Rev.* 77, 342–356. doi:10.1128/mmbr.00051-12.

Neuenschwander, S. M., Ghai, R., Pernthaler, J., and Salcher, M. M. (2018). Microdiversification in genome-streamlined ubiquitous freshwater

Actinobacteria. *ISME J.* 12, 185–198. doi:10.1038/ismej.2017.156.

Neuenschwander, S. M., Salcher, M. M., and Pernthaler, J. (2015). Fluorescence in situ hybridization and sequential catalyzed reporter deposition (2C-FISH) for the flow cytometric sorting of freshwater ultramicrobacteria. *Front. Microbiol.* 6, 1–8. doi:10.3389/fmicb.2015.00247.

Nisbet, E. G., and Fowler, C. M. R. (2003). "The Early History of Life," in *Treatise on Geochemistry* doi:10.1016/B0-08-043751-6/08166-4.

Nishimura, K., Shimada, H., Ohta, H., Masuda, T., Shioi, Y., and Takamiya, K. I. (1996). Expression of the puf operon in an aerobic photosynthetic bacterium, Roseobacter denitrificans. *Plant Cell Physiol.* 37, 153–159. doi:10.1093/oxfordjournals.pcp.a028926.

Olson, J. M. (2006). Photosynthesis in the Archean era. *Photosynth. Res.* 88. doi:10.1007/s11120-006-9040-5.

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., et al. (2015). Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform. Biol. Insights* 9, 75–88. doi:10.4137/BBI.S12462.

Overmann, J., and Garcia-Pichel, F. (2013). "The Phototrophic Way of Life," in *The Prokaryotes: Prokaryotic Communities and Ecophysiology*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 203–257. doi:10.1007/978-3-642-30123-0.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi:10.1126/science.276.5313.734.

Pace, N. R., Stahl, D. A., Lane, D. J., and Olsen, G. J. (1986). "The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences," in *ASM American Society for Microbiology News*, 1–55. doi:10.1007/978-1-4757-0611-6_1.

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., and Hugenholtz, P. (2021). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, 785–794. doi:10.1093/nar/gkab776.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi:10.1038/nbt.4229.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. doi:10.1038/s41564-017-0012-7.

Pascual, J., Foesel, B. U., Geppert, A., Huber, K. J., Boedeker, C., Luckner, M., et al. (2018). Roseisolibacter agri gen. Nov., sp. nov., a novel slow-growing member of the under-represented phylum gemmatimonadetes. *Int. J. Syst. Evol. Microbiol.* 68, 1028–1036. doi:10.1099/ijsem.0.002619.

Pascual, J., García-López, M., Bills, G. F., and Genilloud, O. (2016). Longimicrobium terrae gen. Nov., sp. Nov., an oligotrophic bacterium of the under-represented phylum gemmatimonadetes isolated through a system of miniaturized diffusion chambers. *Int. J. Syst. Evol. Microbiol.* 66, 1976–1985. doi:10.1099/ijsem.0.000974.

Pedrós-Alió, C., and Brock, T. D. (1982). Assessing Biomass and Production of Bacteria in Eutrophic Lake Mendota, Wisconsin. *Appl. Environ. Microbiol.* 44, 203–218. doi:10.1128/aem.44.1.203-218.1982.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* 68, 3094–3101. doi:10.1128/AEM.68.6.3094-3101.2002.

Pessi, I. S., Viitamäki, S., Virkkala, A.-M., Eronen-Rasimus, E., Delmont, T. O., Marushchak, M. E., et al. (2022). In-depth characterization of denitrifier communities across different soil ecosystems in the tundra. *Environ. Microbiome* 17, 30. doi:10.1186/s40793-022-00424-2.

Pierella Karlusich, J. J., Pelletier, E., Lombard, F., Carsique, M., Dvorak, E., Colin, S., et al. (2021). Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods. *Nat. Commun.* 12, 1–18. doi:10.1038/s41467-021-24299-y.

Piwosz, K., Kaftan, D., Dean, J., Šetlík, J., and Koblížek, M. (2018). Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. *Environ. Microbiol.* 20, 724–733. doi:10.1111/1462-2920.14003.

Piwosz, K., Villena-Alemany, C., and Mujakić, I. (2022). Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME J.* 16, 1046–1054. doi:10.1038/s41396-021-01142-2.

Piwosz, K., Vrdoljak Tomaš, A., Thijs, F., Gonzalez Olalla, J. M., Šantić, D., L McKay, R. M., et al. (2020). ic bacteria in a microcosm eLight and primary production shape bacterial activity and community composition of aerobic anoxygenic phototrophxperiment. *mSphere* 00354–20.

Porter, J., Morris, S. A., and Pickup, R. W. (2004). Effect of Trophic Status on the Culturability and Activity of Bacteria from a Range of Lakes in the English Lake District. *Appl. Environ. Microbiol.* 70, 2072–2078. doi:10.1128/AEM.70.4.2072-2078.2004.

Porter, K. G., and Feig, Y. S. (1980). The Use of DAPI for Identifying and Counting Aquatic Microflora. *Limnol. Oceanogr.* 25, 943–948. doi:10.4319/lo.1980.25.5.0943.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi:10.1093/nar/gkm864.

Qian, P., Gardiner, A. T., Šímová, I., Naydenova, K., Croll, T. I., Jackson, P. J., et al. (2022). Phototrophica Photosystem. 3139, 1–11.

Rappé, M. S., Connon, S. A., Vergin, K. L., and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418. doi:10.1038/nature00917.

Rappé, M. S., and Giovannoni, S. J. (2003). The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* 57, 369–394. doi:10.1146/annurev.micro.57.030502.090759.

Raymond, J., and Segrè, D. (2006). The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 311, 1764–7. doi:10.1126/science.1118439.

Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y., and Blankenship, R. E. (2002). Whole-genome analysis of photosynthetic prokaryotes. *Science* 298, 1616–1620. doi:10.1126/science.1075558.

Raynaud, X., and Nunan, N. (2014). Spatial ecology of bacteria at the microscale in soil. *PLoS One* 9. doi:10.1371/journal.pone.0087217.

Risse, J., Thomson, M., Patrick, S., Blakely, G., Koutsovoulos, G., Blaxter, M., et al. (2015). A single chromosome assembly of Bacteroides fragilis strain BE1 from Illumina and MinION nanopore sequencing data. *Gigascience* 4, 1–7. doi:10.1186/s13742-015-0101-6.

Ruiz-González, C., Proia, L., Ferrera, I., Gasol, J. M., and Sabater, S. (2013). Effects of large river dam regulation on bacterioplankton community structure. *FEMS Microbiol. Ecol.* 84, 316–331. doi:10.1111/1574-6941.12063.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5, 0398–0431. doi:10.1371/journal.pbio.0050077.

Salcher, M. M., Neuenschwander, S. M., Posch, T., and Pernthaler, J. (2015). The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J.* 9, 2442–2453. doi:10.1038/ismej.2015.55.

Salcher, M. M., Pernthaler, J., and Posch, T. (2011). Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria that rule the waves (LD12). *ISME J.* 5, 1242–1252. doi:10.1038/ismej.2011.8.

Salcher, M. M., Posch, T., and Pernthaler, J. (2013). In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J.* 7, 896–907. doi:10.1038/ismej.2012.162.

Schönhuber, W., Fuchs, B., Juretschko, S., and Amann, R. (1997). Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Appl. Environ. Microbiol.* 63. doi:10.1128/aem.63.8.3268-3273.1997.

Schulz, H. N., Brinkhoff, T., Ferdelman, T. G., Mariné, M. H., Teske, A., and Jorgensen, B. B. (1999). Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* 284, 493–495. doi:10.1126/science.284.5413.493.

Scott, A. C. (1984). The early history of life on land. *J. Biol. Educ.* 18, 207–219.

doi:10.1080/00219266.1984.9654637.

Sekar, R., Pernthaler, A., Pernthaler, J., Warnecke, F., Posch, T., and Amann, R. (2003). An improved protocol for quantification of freshwater Actinobacteria by fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 69. doi:10.1128/AEM.69.5.2928-2935.2003.

Sener, M., Strumpfer, J., Singharoy, A., Hunter, C. N., and Schulten, K. (2016). Overall energy conversion efficiency of a photosynthetic vesicle. *Elife* 5, 1–30. doi:10.7554/eLife.09541.

Shah, N., Tang, H., Doak, T. G., and Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pacific Symp. Biocomput. 2011, PSB 2011*, 165–176. doi:10.1142/9789814335058_0018.

Smith, D. J., Kharbush, J. J., Kersten, R. D., and Dick, G. J. (2022). Uptake of Phytoplankton-Derived Carbon and Cobalamins by Novel Acidobacteria Genera in Microcystis Blooms Inferred from Metagenomic and Metatranscriptomic Evidence. *Appl. Environ. Microbiol.* 88. doi:10.1128/aem.01803-21.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120. doi:10.1073/pnas.0605127103.

Song, H., Li, Z., Du, B., Wang, G., and Ding, Y. (2012). Bacterial communities in sediments of the shallow Lake Dongping in China. *J. Appl. Microbiol.* 112, 79–89. doi:10.1111/j.1365-2672.2011.05187.x.

Sorokin, D. Y., Lücker, S., Vejmelkova, D., Kostrikina, N. A., Kleerebezem, R., Rijpstra, W. I. C., et al. (2012). Nitrification expanded: Discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J.* 6, 2245–2256. doi:10.1038/ismej.2012.70.

Speth, D. R., In'T Zandt, M. H., Guerrero-Cruz, S., Dutilh, B. E., and Jetten, M. S. M. (2016). Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat. Commun.* 7. doi:10.1038/ncomms11172.

Staley, J. (1985). Measurement of In Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. *Annu. Rev. Microbiol.* 39, 321–346. doi:10.1146/annurev.micro.39.1.321.

Strous, M., Jhon, A. F., Evelien, H. M. K., Susanne, L., Gerard, M., Katinka, T. van de P.-S., et al. (1999). Missing lithotroph identified as new planctomycete. *Nature* 400, 446–449.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348, 1261359. doi:10.1126/science.1261359.

Swingley, W. D., Blankenship, R. E., and Raymond, J. (2009). "Evolutionary Relationships Among Purple Photosynthetic Bacteria and the Origin of Proteobacterial Photosynthetic Systems," in *The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration*, ed. J. T. Hunter, C.N., Daldal, F., Thurnauer, M.C., Beatty (Springer, Dordrecht). doi:10.1007/978-1-4020-8815-

5_2.

Teske, A., Alm, E., Regan, J. M., Toze, S., Rittmann, B. E., and Stahl, D. A. (1994). Evolutionary Relationships among Ammonia- and Nitrite-Oxidizing Bacteria. 176, 6623–6630.

Torsvik, V., Goksoyr, J., and Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56, 782–787. doi:10.1128/aem.56.3.782-787.1990.

Tran, P. Q., Bachand, S. C., McIntyre, P. B., Kraemer, B. M., Vadeboncoeur, Y., Kimirei, I. A., et al. (2021). Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. *ISME J.* 15, 1971–1986. doi:10.1038/s41396-021-00898-x.

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., et al. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* 54, 2298–2314. doi:10.4319/lo.2009.54.6_part_2.2298.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative metagenomics of microbial communities. *Science* 308, 554–557. doi:10.1126/science.1107851.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428. doi:10.1038/nature02340.

Vavourakis, C. D., Andrei, A.-S., Mehrshad, M., Ghai, R., Sorokin, D. Y., and Muyzer, G. (2018). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* 6, 168. doi:10.1186/s40168-018-0548-7.

Vavourakis, C. D., Mehrshad, M., Balkema, C., Van Hall, R., Andrei, A. Ş., Ghai, R., et al. (2019). Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol.* 17, 1–20. doi:10.1186/s12915-019-0688-7.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi:10.1126/science.1093857.

Villena-Alemany, C., Mujakić, I., Porcal, P., Koblížek, M., and Piwosz, K. (2022). Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environ. Microbiol. Rep.*, 1–12. doi:10.1111/1758-2229.13131.

Vitousek, P. M., Cassman, K., Cleveland, C., Crews, T., Field, C. B., Grimm, N. B., et al. (2002). Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* 57–58, 1–45. doi:10.1023/A:1015798428743.

Volland, J.-M., Gonzalez-Rizzo, S., Gros, O., Tyml, T., Ivanova, N., Schulz, F., et al. (2022). A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. *Science* 376, 1453–1458. doi:10.1126/science.abb3634.

Ward, L. M., Cardona, T., and Holland-Moritz, H. (2019). Evolutionary Implications

of Anoxygenic Phototrophy in the Bacterial Phylum Candidatus Eremiobacterota (WPS-2). *Front. Microbiol.* 10, 1–12. doi:10.3389/fmicb.2019.01658.

Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E., and Fischer, W. W. (2018). Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer. *Front. Microbiol.* 9, 1–16. doi:10.3389/fmicb.2018.00260.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578–83. doi:10.1073/pnas.95.12.6578.

Wick, R. R., Judd, L. M., and Holt, K. E. (2023). Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput. Biol.* 19, e1010905. doi:10.1371/journal.pcbi.1010905.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271. doi:10.1128/mmbr.51.2.221-271.1987.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms (archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny). *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090.

Wooley, J. C., Te, Y., and Ye, Y. (2010). Metagenomics: Facts and artifacts, and computational challenges. *J Comput Sci Technol* 25, 71–81. doi:10.1007/s11390-010-9306-4.

Wu, Q. L., and Hahn, M. W. (2006). High predictability of the seasonal dynamics of a species-like Polynucleobacter population in a freshwater lake. *Environ. Microbiol.* 8, 1660–1666. doi:10.1111/j.1462-2920.2006.01049.x.

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., Van Bleijswijk, J., Timmers, P., et al. (2006). Archaeal nitrification in the ocean. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12317–12322. doi:10.1073/pnas.0600756103.

Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S. M., Meng, J., et al. (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J.* 5, 414–426. doi:10.1038/ismej.2010.144.

Yool, A., Martin, A. P., Fernández, C., and Clark, D. R. (2007). The significance of nitrification for oceanic new production. *Nature* 447, 999–1002. doi:10.1038/nature05885.

Yurkov, V., and Csotonyi, J. T. (2009). New Light on Aerobic Anoxygenic Phototrophs. *Hunt. C.N., Daldal F., Thurnauer M.C., Beatty J.T. Purple Phototrophic Bact. Adv. Photosynth. Respir. vol 28. Springer, Dordrecht.*, 31–55. doi:10.1007/978-1-4020-8815-5_3.

Yurkov, V. V., and Beatty, J. T. (1998). Aerobic Anoxygenic Phototrophic Bacteria. *Microbiol. Mol. Biol. Rev.* 62, 695–724. doi:10.1128/mmbr.62.3.695-724.1998.

Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: A cross-system comparison. *Environ. Microbiol.* 5, 539–554. doi:10.1046/j.1462-2920.2003.00451.x.

Zeng, Y., Baumbach, J., Barbosa, E. G. V., Azevedo, V., Zhang, C., and Koblížek, M.

(2016). Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environ. Microbiol. Rep.* 8, 139–149. doi:10.1111/1758-2229.12363.

Zeng, Y., Feng, F., Medová, H., Dean, J., Koblížek, M., Medova, H., et al. (2014). Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7795–7800. doi:10.1073/pnas.1400295111.

Zeng, Y., Nupur, Wu, N., Madsen, A. M., Chen, X., Gardiner, A. T., et al. (2021). Gemmatimonas groenlandica sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes. *Front. Microbiol.* 11, 1–18. doi:10.3389/fmicb.2020.606612.

Zeng, Y., Selyanin, V., Lukeŝ, M., Dean, J., Kaftan, D., Feng, F., et al. (2015). Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium Gemmatimonas phototrophica sp. Nov., and emended descriptions of the genus Gemmatimonas and Gemmatimonas aurantiaca. *Int. J. Syst. Evol. Microbiol.* 65, 2410–2419. doi:10.1099/ijs.0.000272.

Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., et al. (2003). Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *Int. J. Syst. Evol. Microbiol.* 53, 1155–1163. doi:10.1099/ijs.0.02520-0.

Zhang, J., Yang, Y., Zhao, L., Li, Y., Xie, S., and Liu, Y. (2015). Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Appl. Microbiol. Biotechnol.* 99, 3291–3302. doi:10.1007/s00253-014-6262-x.

Zorz, J. K., Sharp, C., Kleiner, M., Gordon, P. M. K., Pon, R. T., Dong, X., et al. (2019). A shared core microbiome in soda lakes separated by large distances. *Nat. Commun.* 10, 1–10. doi:10.1038/s41467-019-12195-5.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of history. *J. Theor. Biol.* 8, 357–366.

Curriculum vitae

Izabela Mujakić

izabelamujakic@yahoo.com
Date of Birth: 31.08.1988
Place of birth: Zagreb, Croatia

Google Scholar: https://scholar.google.com/citations?hl=cs&user=DTCGVGcAAAAJ
Research gate: https://www.researchgate.net/profile/Izabela-Mujakic-2
OrcID https://orcid.org/0000-0001-5602-7331

EDUCATION

October 2017 – Current
PhD student – University of South Bohemia, Faculty of Science, Department of Ecosystem Biology, Hydrobiology
PhD student – Laboratory of Anoxygenic phototrophs, Algatech Center, Institute of Microbiology, Czech Academy of Sciences
Thesis: *Ecology of Gemmatimonadota*

October 2012 – November 2014
Master in Environmental Sciences- University of Zagreb, Faculty of Science, Department of Biology
Thesis work done at Ruđer Bošković Institute, Division for marine and environmental research, Laboratory of aquaculture and pathology of aquatic organisms
Thesis: *Molecular characterization of intestinal parasites (Acanthocephala) in gut of freshwater fishes*

October 2008 – September 2012
Bachelor in Environmental Sciences - University of Zagreb, Faculty of Science, Department of Biology

WORK EXPERIENCE

February 2017 – Current
Laboratory of Anoxygenic phototrophs, Algatech Center, Institute of Microbiology, Czech Academy of Sciences

September 2016 – January 2017
Associate at Ruđer Bošković Institute, Division for marine and environmental research, Laboratory of aquaculture and pathology of aquatic organisms

November 2015 – September 2016
Internship at State office for Radiological and Nuclear safety, Division for Environment and Radioactive waste

PRESENTATION AT CONFERENCES

ISME18 (14th–19th of August 2022) Lausanne, Switzerland

poster presentation Multi-environment ecogenomics of the cosmopolitan phylum Gemmatimonadota

SAME16 (1st–16th of September 2019) Potsdam, Germany
poster presentation Analysis of metagenome assembled genomes (MAGs) of freshwater Gemmatimonadetes

6th Freshblood for Fresh Water Conference (23rd–27th of April 2019) Tihany, Hungary
poster presentation Analysis of metagenome assembled genomes (MAGs) of freshwater Gemmatimonadetes

WORKSHOPS AND OTHER

Part of team at Algatech Center, Institute of Microbiology involved in SARS-CoV-2 testing (30th March–31st May 2020)

Galaxy workshop- Introduction to Galaxy and NGS analysis, RNA-seq analysis with Galaxy and R, Genome Annotation (15th–19th February 2021)

Workshop for basic statistics and the usage of R at Algatech Center, Institute of Microbiology, Czech Republic (April–June 2018)

SAME15- Helping the organizing team during the conference (Zagreb, Croatia 3rd–8th September 2017)

Part of a team during 10th International GAP meeting 19th–30th August 2017, Třeboň, Czech Republic

MicroCokit Project training school- Methods for detecting and quantifying aquatic microbial communities (Rome, Italy 4th–6th April 2017)

SCHOLARSHIPS AND STUDY STAYS

Scholarship for PhD students from University of South Bohemia, Faculty of Science

Erasmus+ Traineeship (1st May–01 July) University of Valencia, Cavanilles Institute for Biodiversity and Evolutionary Biology, Group of Limnology and Department of Microbiology and Ecology, School of Biology

Study stay (1st–30th November 2018) at Aarhus University, Department of Environmental Science, Roskilde, Denmark

LANGUAGE SKILLS

Croatian (mother tongue), English (C1), Spanish (C1)

ACADEMIC PUBLICATIONS

Villena-Alemany, C., **Mujakić, I**., Porcal, P., Koblížek, M. & Piwosz, K. (2023) Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environmental Microbiology Reports*, 15(1), 60–71.

Gardiner, A.T., **Mujakić, I.**, Bína, D., Gardian, Z., Kopejtka, K., Nupur, Quian, P., Koblížek, M. (2023) Characterisation of the photosynthetic complexes from the marine gammaproteobacterium Congregibacter litoralis KT71. *Biochimica et Biophysica Acta*

(BBA) - Bioenergetics, Volume 1864, Issue 2

Mujakić, I., Piwosz, K., Koblížek, M. (2022) Phylum Gemmatimonadota and Its Role in the environment. Microorganisms, 10, 151.

Piwosz, K., Villena-Alemany, C. & **Mujakić, I.** (2022) Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME J* **16**, 1046–1054.

Kačírová, J., Maďari, A., Mucha, R., Fecskeová, L.K., **Mujakić, I.**, Koblížek, M., Nemcová, R., Maďar, M. (2021) Study of microbiocenosis of canine dental biofilms. *Sci Rep* **11**, 19776.

Mujakić, I., Andrei, A-Ş., Shabarova, T., Fecskeová. L.K., Salcher, M.M., Piwosz, K., Ghai, R., Koblížek, M. (2021) Common presence of phototrophic Gemmatimonadota in temperate freshwater lakes. mSystems 6:e01241-20.

Kolda, A., **Mujakić, I.**, Perić, L., Vardić-Smrzlić, I., Kapetanović, D. (2020) Microbiological Quality Assessment of Water and Fish from Karst Rivers of the Southeast Black Sea Basin (Croatia), and Antimicrobial Susceptibility of *Aeromonas* Isolates. *Curr Microbiol* **77**, 2322–2332.

Piwosz, K., Vrdoljak, A., Frenken, T., González-Olalla, J.M., Šantić, D., McKay, R.M., Spilling, K., Guttman, L., Znachor, P., **Mujakić, I.**, Fecskeová, L.K., Zoccarato, L., Hanusová, M., Pessina, A., Reich, T., Grossart, H-P., Koblížek, M. (2020) Light and primary production shape bacterial activity and community composition of aerobic anoxygenic phototrophic bacteria in a microcosm experiment. mSphere 5:e00673-20.

Pjevac, P., Dyksma, S., Goldhammer, T., **Mujakić, I.**, Koblížek, M., Mußmann, M., Amann, R. and Orlić, S. (2019) *In situ* abundance and carbon fixation activity of distinct anoxygenic phototrophs in the stratified seawater lake Rogoznica. Environ Microbiol, 21: 3896-3908

© for non-published parts Izabela Mujakić
izabelamujakic@yahoo.com

Ecology of Gemmatimonadota
Ph.D. Thesis Series, 2023, No. 12

All rights reserved
For non-commercial use only
Printed in the Czech Republic by Typodesign
Edition of 15 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz