



OPEN ACCESS

EDITED BY

Magdalena R. Osburn,
Northwestern University, United States

REVIEWED BY

Mircea Podar,
Oak Ridge National Laboratory (DOE),
United States
Eric Daniel Becraft,
University of North Alabama,
United States

*CORRESPONDENCE

Adrien Vigneron
✉ avignero@gmail.com

SPECIALTY SECTION

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 18 October 2022

ACCEPTED 05 December 2022

PUBLISHED 09 January 2023

CITATION

Vigneron A, Cruaud P, Guyoneaud R
and Goñi-Urriza M (2023) Into the
darkness of the microbial dark matter
in situ activities through expression
profiles of *Patescibacteria* populations.
Front. Microbiol. 13:1073483.
doi: 10.3389/fmicb.2022.1073483

COPYRIGHT

© 2023 Vigneron, Cruaud, Guyoneaud
and Goñi-Urriza. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Into the darkness of the microbial dark matter *in situ* activities through expression profiles of *Patescibacteria* populations

Adrien Vigneron^{1*}, Perrine Cruaud², Rémy Guyoneaud¹ and Marisol Goñi-Urriza¹

¹IBEAS, Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, IPREM, Pau, France,

²Independent Researcher, Lourenties, France

Patescibacteria form a highly diverse and widespread superphylum of uncultured microorganisms representing a third of the global microbial diversity. Most of our knowledge on *Patescibacteria* putative physiology relies on metagenomic mining and metagenome-assembled genomes, but the *in situ* activities and the ecophysiology of these microorganisms have been rarely explored, leaving the role of *Patescibacteria* in ecosystems elusive. Using a genome-centric metatranscriptomic approach, we analyzed the diel and seasonal gene transcription profiles of 18 *Patescibacteria* populations in brackish microbial mats to test whether our understanding of *Patescibacteria* metabolism allows the extrapolation of their *in situ* activities. Although our results revealed a circadian cycle in *Patescibacteria* activities, a strong streamlined genetic expression characterized the *Patescibacteria* populations. This result has a major consequence for the extrapolation of their physiology and environmental function since most transcribed genes were uncharacterized, indicating that the ecophysiology of *Patescibacteria* cannot be yet reliably predicted from genomic data.

KEYWORDS

CPR bacteria, Parcubeacteria, MAGs, metatranscriptomic, microbial mats, genome-centric metatranscriptomic, ecophysiology

Introduction

Metagenomic sequencing has significantly expanded our knowledge of microbial diversity in various environments (Parks et al., 2017). This approach has notably revealed the “microbial dark matter,” gathering a large fraction of the global microbial diversity that escaped culture and PCR-based surveys (Rinke et al., 2013; Nobu et al., 2015). The candidate phyla radiation (CPR), also known as the *Patescibacteria* superphylum, has emerged from this microbial dark matter. Members of this superphylum have now been detected almost ubiquitously in natural environments, including in human-related samples (Castelle et al., 2018),

representing a significant proportion (7.5%) of the communities in most ecosystems (Beam et al., 2020). However, when the filtration strategy of aquatic samples included a size fraction lower than 0.22 μm , the relative proportion of *Patescibacteria* lineages frequently exceeded those of other bacteria in diversity surveys, suggesting that the abundance of ultra-small *Patescibacteria* is frequently underestimated (Luef et al., 2015; Anantharaman et al., 2016; Danczak et al., 2017; Vigneron et al., 2020; Chaudhari et al., 2021). Most of our knowledge about these microorganisms derives from population genomes recovered from metagenomic datasets (Brown et al., 2015; Hug et al., 2016; Castelle and Banfield, 2018; Tian et al., 2020). With more than 5000 draft genomes of *Patescibacteria* in public databases, the metabolic repertoire of the *Patescibacteria* members is rapidly expanding. Shared features of *Patescibacteria* include a small genome size with limited known biosynthetic and metabolic pathways (Castelle and Banfield, 2018; Tian et al., 2020). Therefore, a potential episymbiotic and fermentative lifestyle is frequently assumed for the *Patescibacteria* (Anantharaman et al., 2016; Castelle and Banfield, 2018; Castelle et al., 2018). While the majority of this superphylum remains uncultivated, some representatives of the *Saccharimonadia* group (previously known as TM7) were found to be episymbiotic of actinobacterial cells (He et al., 2015; Bor et al., 2020; Murugkar et al., 2020). Few members of the *Gracilibacteria* were also characterized as predatory bacteria of phototrophic *Gammaproteobacteria* (Moreira et al., 2021; Yakimov et al., 2021), and members of the *Yanofskybacteria* have been observed in the vicinity of methanogenic archaea (Kuroda et al., 2022), supporting the general assumption of a host-associated lifestyle for *Patescibacteria*. However, evidence for a host-dependent lifestyle is currently limited to the few characterized lineages, and the putative metabolism of most of the *Patescibacteria* lineages remains misunderstood. Free-living *Patescibacteria* were also observed, suggesting a broad range of potential life strategies in this phylum (Chiriac et al., 2022). However, these results might have pictured a “free-floating” stage in the life cycle of episymbiotic/parasitic *Patescibacteria*, as observed with *Saccharibacteria* (Bor et al., 2018). Since essential functions, such as genetic information processing, including nucleotide metabolic processes, were conserved, the reduced genome size observed in *Patescibacteria* could also be associated with an evolutionary streamlining to reduce functional redundancy and energy waste for the free-living bacteria (Tian et al., 2020). Consistently, a limited catabolic potential has been detected for different *Patescibacteria* lineages, suggesting a potential niche partitioning based on different carbohydrate utilization (Danczak et al., 2017; Vigneron et al., 2020). However, due to the lack of cultured representatives and ecophysiological studies, these metabolic hypotheses predicted from genomic data have not been tested, leaving the activity and ecological roles of these microorganisms elusive.

Taking benefit of an extensive metagenomic (370 Gb of reads) and metatranscriptomic (300 Gb of rRNA-depleted reads) dataset (SRP063590) covering sample heterogeneity as well as seasonal and diel variability of two microbial mats located in brackish lagoons (Salins du Lion and Etang de Berre, France) (Vigneron et al., 2021), we use a genome-centric metatranscriptomic approach to investigate *in situ* activities of *Patescibacteria* lineages. More specifically, we tested if our current understanding of the *Patescibacteria* genomic repertoire allowed the extrapolation of their putative activities.

Materials and methods

Metagenomic and metatranscriptomic samples

In this study, we analyzed metagenomic and metatranscriptomic sequences recovered from samples collected in two coastal microbial mats developing in neighboring brackish lagoons located at l'Etang de Berre (EDB) and Salins du Lion (SL). These microbial mats were strongly stratified with steep gradients of oxygen, nitrogen, and sulfur compounds throughout the thickness of the mats, leading to both aerobic and anoxic niches (Aubé et al., 2020; Vigneron et al., 2021). As previously described (Aubé et al., 2020), the sequenced DNA and RNA originating from 30 microbial mat samples were collected in triplicate in September 2011, April 2012, and September 2012 during both daytime (4 PM) and nighttime (4 AM). Nucleic acid extraction and library preparation, including ribosomal RNA depletion of the metatranscriptomes, were previously described (Aubé et al., 2020). Illumina HiSeq sequencing of the libraries led to a total of 670 Gb of sequencing reads (370 GB of metagenomes and 300 GB of rRNA-depleted metatranscriptomes). Datasets were quality filtered using the Trimmomatic v.0.39 tool keeping both R1 and R2 reads when reads overlapped (Bolger et al., 2014). The 16S rRNA reads were isolated from the metagenomic reads using Infernal v.1.1.4 (Nawrocki and Eddy, 2013), and their taxonomic assignments were performed with Mothur (Schloss et al., 2009) using BLAST against Silva database release 138 as reference (Pruesse et al., 2007). Since metagenomic 16S rRNA reads were only 100-bp long and spanned various regions of the 16S rRNA gene, taxonomic assignments were limited to the genus level and above.

Binning and functional characterization

For metagenome-assembled genome (MAG) reconstruction, all quality-filtered sequences were pooled and co-assembled using MEGAHIT v.1.2.9 (Li et al., 2015), as previously described (Vigneron et al., 2021). Read coverage

of the contigs against all metagenomic datasets was carried out using *bwa-mem* v.0.7.17 (<http://bio-bwa.sourceforge.net>), and the binning of the contigs longer than 2000 bp was carried out by MetaBAT-2 (Kang et al., 2015). Coding density, completeness, and contamination levels of the MAGs were then evaluated using CheckM v.1.1.2 (Parks et al., 2015) with *cpr_43_markers.hmm* files that provide a better estimation of CPR bacteria completeness (Parks et al., 2015). Taxonomic affiliation of the MAGs was performed using the GTDB_Tk classify workflow (GTDB-tk release 207) with genomes containing a minimum of 40% of the bacterial marker genes (Chaumeil et al., 2020). In addition, a phylogenomic tree was constructed based on the concatenated alignment of all ribosomal protein sequences detected in the MAGs (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24, rpS3, 6, 8, 10, 17, 19) using *metabolisHMM* (McDaniel et al., 2020). Ribosomal proteins were aligned using *mafft* v.7.471 (Katoh et al., 2002), and a phylogenetic tree of the concatenated alignment (62 sequences with 3729 amino acid positions) was produced using *FastTree2* with 1000 bootstraps (Price et al., 2010). Open reading frames (ORFs) were identified using *Prodigal* (Hyatt et al., 2010) with coding table 11, except for *Gracilibacteria* MAGs that required coding table 23 (Campbell et al., 2013). ORFs were then compared against COG (Galperin et al., 2021), TIGRFam (Haft et al., 2003), and KEGG (Kanehisa et al., 2016) databases on IMG/MER platform using IMG annotation pipeline v.5.0.19 (Markowitz et al., 2009), leading to 5.7×10^6 genes coding for proteins with the product name (58.96% of the genes). ORFs were also compared against the CAZY database using *dbcan2*, and glycoside hydrolase genes were analyzed to infer the catabolic potential (Zhang et al., 2018). The type of ribulose-1,5-bisphosphate carboxylase (Rubisco) protein was determined using publicly available HMM profiles of the different Rubisco categories (Anantharaman et al., 2016). Metatranscriptomic reads passing quality filtration ($4.12 \pm 0.7 \times 10^7$ reads passing quality filtration per sample for a total of 1.24×10^9 metatranscriptomic reads) were normalized and then mapped against all ORF identified in the MAGs using *bwa-mem* with a default value to determine the expression level of each gene under all tested conditions. To limit computing time, the mapping of the metatranscriptomic reads was performed only against samples where MAGs were recovered. For the comparison with other phyla, only MAGs with a coverage similar to *Patescibacteria* MAGs were analyzed, and only phylum with at least eight representative MAGs were included.

Results and discussion

The microbial community composition of the two microbial mats was characterized using 16S rRNA genes recovered from the 30 metagenomes. Overall, *Patescibacteria*

constituted a minor but diverse fraction of the microbial mats community, representing in total 2% of the 16S rRNA reads from the metagenomic dataset (Figure 1A). Nonetheless, spatial and seasonal variations in the *Patescibacteria* community composition were detected (Supplementary Figure 1), suggesting rapid responses to environmental conditions. After co-assembly of the metagenomic datasets and binning of the contigs longer than 2000 bp, 18 MAGs (average completeness: $90 \pm 10\%$) over the 407 MAGs recovered from the mats were affiliated to *Patescibacteria*, including representatives of the *Candidatus Pacebacteria*, *Dojkabacteria*, *Gracilibacteria*, *Moranbacteria*, *Campbellbacteria*, and *Kaiserbacteria* phyla (Figure 1B), that were identified as dominant *Patescibacteria* by 16S rRNA gene analysis. The coverage of the *Patescibacteria* MAGs averaged at 2.75 reads per base in contigs (max: 12.68, min: 0.85), which is below the average coverage of the 407 recovered MAGs (3.5) but higher than the median value (1.62). This result suggests that although *Patescibacteria* represented a small fraction of the community, each individual population has a similar relative abundance to most of the other populations found in the microbial mats and could, therefore, play an important role in these ecosystems.

Genomic composition of *Patescibacteria*

Consistently with publicly available genomes, the *Patescibacteria* MAGs were reduced in size, and an average of 940 ORFs per MAG were identified, indicating a lower genomic content for a similar completeness level compared with other phyla found in the microbial mats (Figure 2A and Supplementary Figure 2). Comparison against Kegg (Kanehisa et al., 2016), COG (Galperin et al., 2021), TIGRFAM (Haft et al., 2003), and CAZY (Cantarel et al., 2009) databases assigned together a predicted function for $65 \pm 4\%$ of the proteins encoded in ORFs (Supplementary Figure 3), which are lower than the average number of proteins with a predicted function for the entire microbial mat community (70%) but comparable to *Planctomycetes* or *Cyanobacteria* (Figures 2B, C). Although some key functional genes might not have been detected due to the incompleteness of some MAGs and sequence divergence, the overall *Patescibacteria* metabolic potential, inferred from the identified proteins, was consistent with previously published genomes (Castelle and Banfield, 2018).

Most of the MAGs included the pentose phosphate pathway and the full or terminal part of glycolysis/gluconeogenesis (Figure 3). Genes involved in the AMP salvage pathway (Wrighton et al., 2016), including gene coding for the type II/III ribulose-1,5-bisphosphate carboxylase (Rubisco), the AMP phosphorylase, and the ribose-1,5-bisphosphate isomerase were identified in *Ca. Dojkabacteria* MAGs, supporting the

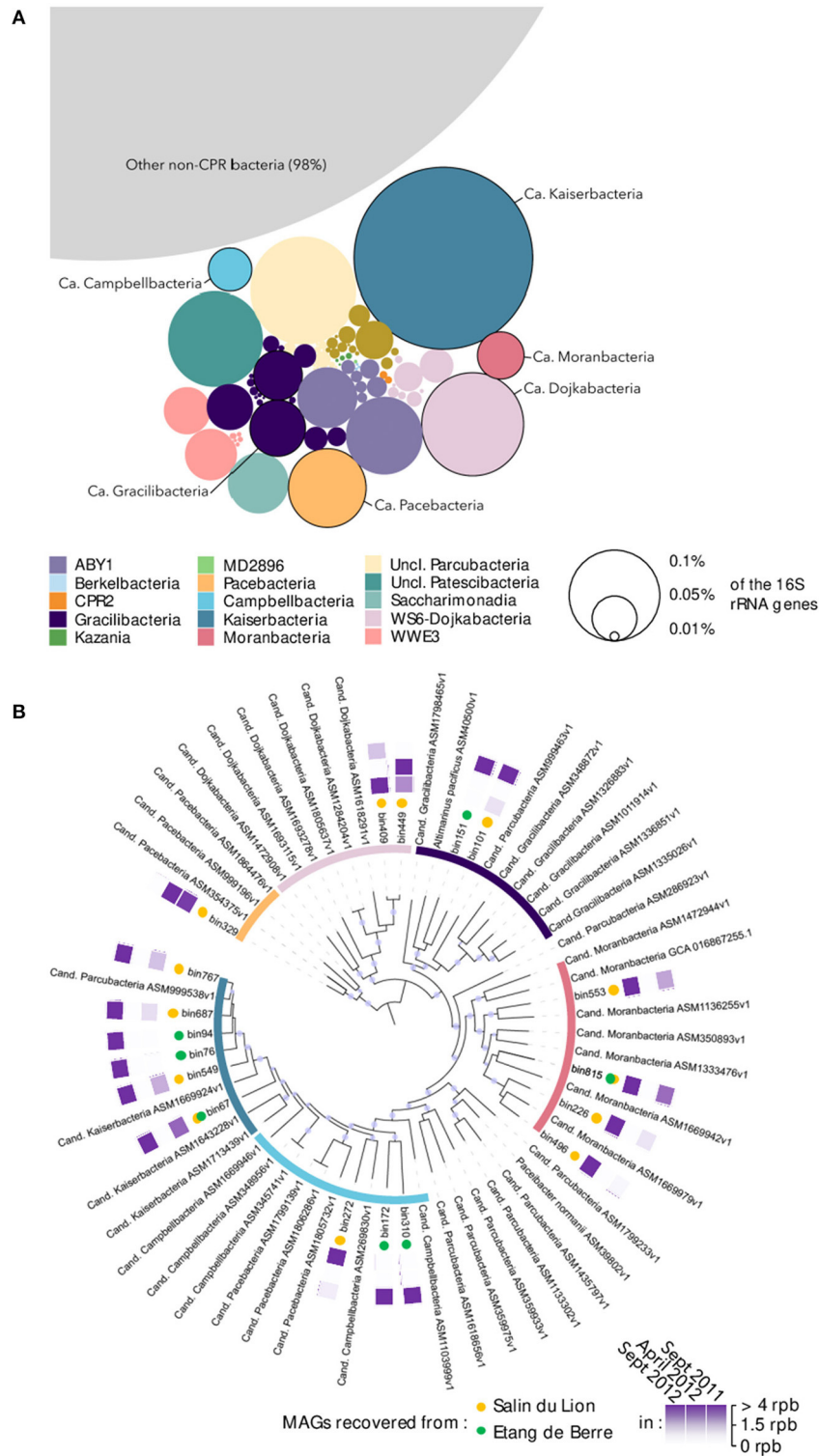
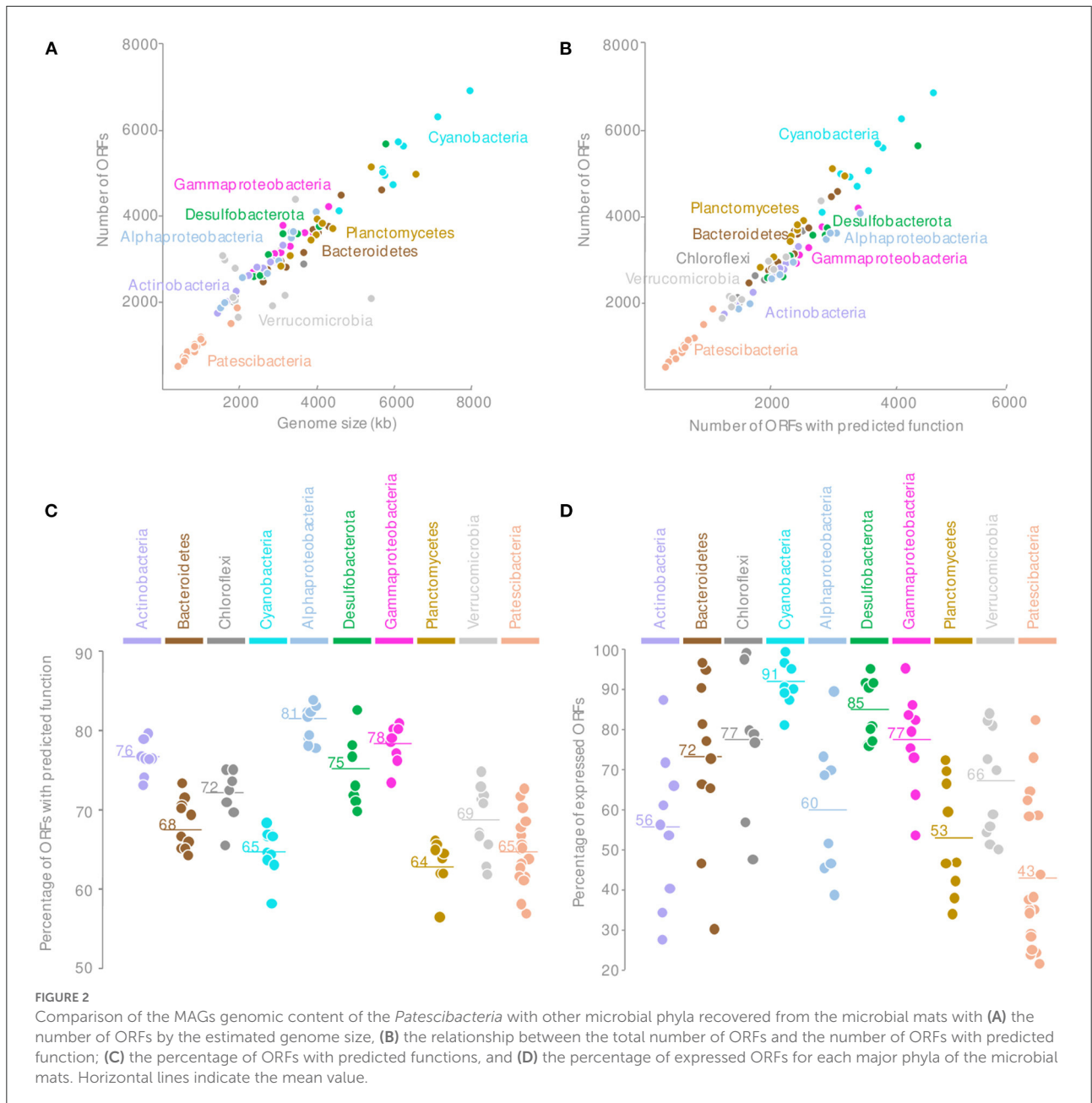


FIGURE 1

(A) Overall *Patescibacteria* community composition in the microbial mats based on 16S rRNA genes identified in the metagenomic dataset. The size of the bubbles represents the average relative proportion of each lineage across all samples. Circled bubbles indicate lineages with representative MAGs. (B) Phylogenomic tree of the *Patescibacteria* MAGs recovered in the microbial mats. The tree was constructed based on concatenated ribosomal protein genes alignment with the closest representative genomes available on NCBI. The purple points in the branches of the tree represent bootstrap values >0.8. The inner circle is color coded as in (A). Yellow/green dots indicate the origin of the MAG, respectively. Purple squares are heat maps indicating the season in which MAGs were detected based on the coverage of the MAGs (rpb, reads per bases). For example, bin767 was detected in September 2011, not detected in April 2012, and strongly detected again in September 2012.



potential for ribose degradation in some *Patescibacteria* phyla (Wrighton et al., 2016). While forms I and II of the Rubisco can fix atmospheric carbon dioxide during primary production via the Calvin–Benson–Bassham cycle, forms II/III and III, found primarily in archaea, enable light-independent CO₂ incorporation into sugars derived from nucleosides (Sato et al., 2007). Previously detected in various Patescibacteria lineages, including other *Dojkabacteria* (Castelle et al., 2018) and DPANN Archaea (Vigneron et al., 2022), enzymes coded by these genes catalyze the conversion of nucleoside

to glycerate-3P, which is then incorporated through the terminal part of glycolysis. Genes of proteins converting phosphoenolpyruvate generated by the glycolysis or the AMP salvage pathway into pyruvate were detected in most of the MAGs. Genes coding for Fe–Fe hydrogenase and lactate, malate, formate, and alcohol dehydrogenases were detected in the MAGs, providing various pathways for pyruvate fermentation (Figure 3) and supporting a basic metabolism centered on simple metabolites (Tian et al., 2020). A limited number of genes coding for transporters were detected in the MAGs,



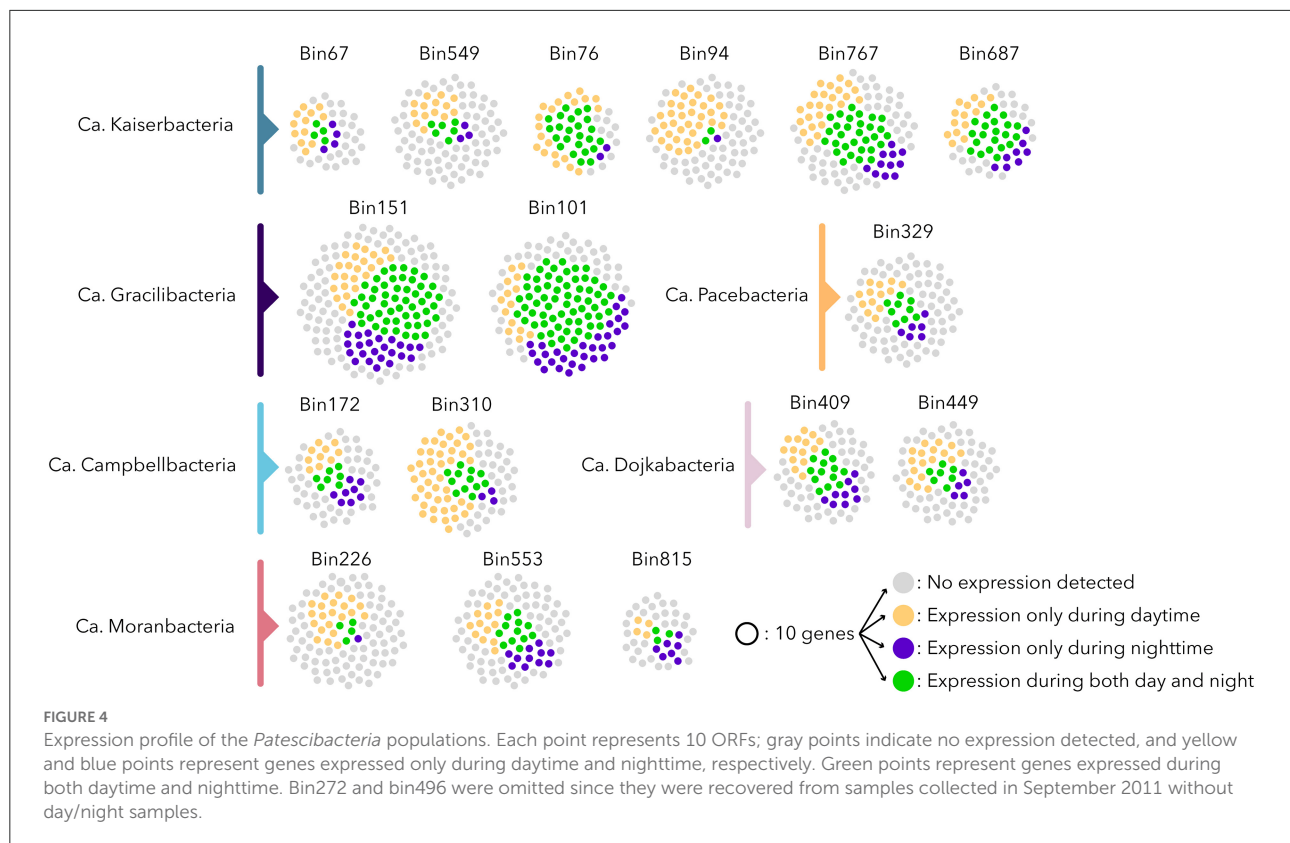
as previously reported (Tian et al., 2020; Vigneron et al., 2020). Transport systems for simple sugars, polysaccharides, amino acids, and volatile fatty acids were identified, notably in *Campbellbacteria*, suggesting a potential uptake of these compounds from the environment or a potential host. Genes encoding di-/tri-carboxylate transporters, allowing the uptake of intermediates of the tricarboxylic acid and Krebs cycles (e.g., citrate, fumarate, succinate, and malate), were also identified in *Gracilibacteria* MAGs (Figure 3). These MAGs also include oxaloacetate-decarboxylating malate dehydrogenase genes to convert these substrates into pyruvate, and even the TCA cycle for bin151, supporting the assimilation of di-/tricarboxylates produced by a potential host or excreted in the microbial mat matrix by other microorganisms (Krom et al., 2003).

Limited catabolic capabilities were identified among the different *Patescibacteria* lineages (max: nine glycoside hydrolase genes per MAG, Supplementary Figure 4). Consistently with previous investigations, potential substrates included cellulose and its components (GH3, GH5, GH39, and GH78), amylose (GH13, GH15, and GH57), chitin (GH18), polysaccharides (GH28, GH65, and GH100), and peptidoglycan (GH23 and GH108) (Danczak et al., 2017; Vigneron et al., 2019). Laccases and peroxidases genes were also identified in few

MAGs, suggesting that some *Patescibacteria* lineages could also degrade lignin.

Alternative energy-conserving pathways were scarce in *Patescibacteria* MAGs but included genes coding for bacteriorhodopsin, thiosulfate dehydrogenase, and nitrate and nitrite reductases, supporting previous results (León-Zayas et al., 2017; Jaffé et al., 2021). The microbial mats are characterized by steep gradients of light, oxygen, nitrogen, and sulfur compounds (Aubé et al., 2016), and sulfur and nitrogen cycles have been previously identified as major processes in these mats (Aubé et al., 2020), supporting such metabolism across the thickness of the mats.

In addition, the large subunit of the aerobic carbon monoxide dehydrogenase form II (CoxL) was identified in one *Moranbacteria* MAG (Figure 3), suggesting the potential for CO oxidation. Aerobic and anaerobic carbon monoxide dehydrogenase genes have not been reported in *Patescibacteria* so far (Adam et al., 2018); however, the catalytic site motif AYRGAGR, and HGRDH, HSHGQ motifs that occur consistently in characterized form II CoxL were detected in the sequence (King and Weber, 2007), supporting the identification of the gene. A comparison of the sequence against the NCBI database indicated that the sequence shared 74% similarity with the carbon monoxide dehydrogenase gene from



an uncultured marine organism and 67% similarity with the carbon monoxide dehydrogenase gene of *Rokubacteria* MAGs. Nonetheless, other subunits of the complex were not identified based on sequence similarity. However, it has been suggested that the lack of other subunits may indicate that this enzyme could act on substrates other than CO (King and Weber, 2007). Therefore, further work is needed to characterize the activity and distribution of putative aerobic carbon monoxide dehydrogenase in *Patescibacteria*.

In situ activities of *Patescibacteria*

The extended metatranscriptomic dataset ($4.12 \pm 0.7 \times 10^7$ reads passing quality filtration per sample for a total of 1.24×10^9 metatranscriptomic reads) that covered the sample heterogeneity, as well as the seasonal (April and September) and diel variations of the microbial mats, was mapped against *Patescibacteria* ORFs (Supplementary Table 2). An analysis of the expression level of ribosomal protein genes, DNA polymerases, and the number of expressed genes indicated differential activity levels for the *Patescibacteria* populations of the microbial mats (Figure 4). Only 32% of the expressed genes were transcribed under both day and night conditions, suggesting contrasted diel activities in *Patescibacteria* (Figure 4

and Supplementary Table 2). Apart from *Ca. Gracilbacteria* that showed higher nighttime activity, all other *Patescibacteria* populations expressed on average three times more genes during daytime than nighttime, with a maximum of 19 times for a *Ca. Kaiserbacteria* (bin94) population with a bacteriorhodopsin-like gene expressed only during daytime. Bacteriorhodopsin genes have been previously identified in freshwater *Patescibacteria* (Chiriach et al., 2022; Jaffe et al., 2022), and proton transport assays confirmed the light-induced proton translocation role of this rhodopsin (Jaffe et al., 2022), supporting the ability to produce energy from light for some *Patescibacteria* lineages. This result suggests that some *Patescibacteria* activity might be influenced directly by light. Water temperature, that fluctuated from 21.5°C during the day to 13°C at night, as well as dissolved oxygen (day: 10.9 mg/L, night: 4.6 mg/L) and redox potential (day: -200 Mv, night: -295 Mv) that are both linked to the activity of light harvesting microorganisms, also varied between daytime and nighttime and could also potentially shape *Patescibacteria* activities. In addition, activities of other diel microorganisms, such as photosynthetic *Cyanobacteria* or anoxygenic phototrophs that represent a large fraction of the microbial mat community (Vigneron et al., 2021), might also influence the gene expression pattern of *Patescibacteria* that rely on these lineages to fulfill their metabolic dependences.

Surprisingly, only $43.4 \pm 18\%$ of the *Patescibacteria* ORFs were expressed in at least one sample regardless of the estimated level of activity (Figure 2D). This proportion of transcribed ORFs is lower than for any other MAGs from other phyla recovered from the same samples and with similar coverage (Figure 2D). These results indicated that although *Patescibacteria* lineages harbored a reduced genomic content, a large fraction of their genomes were not expressed under the contrasting environmental conditions of our sampling campaigns. These results might indicate a rapid transcriptomic control, as illustrated by the dial shift, for a fine-tuning of their activities. This streamlined activity could represent an ecological benefit in the context of the Black Queen Hypothesis that characterizes the *Patescibacteria* genome evolution (Tian et al., 2020). In a complex and dense microbial ecosystem where by- and end-products of catabolic reactions are probably abundant, *Patescibacteria* populations could save their energy by streamlining their genetic expression and benefit from catabolic products available in the microbial mat matrix. This could also illustrate a rapid synchronization of the *Patescibacteria* to the activity of their potential host. Alternatively, this result might indicate that the genome reduction process is still ongoing. However, the coding density of the *Patescibacteria* MAGs was similar to the rest of the microbial community ($89 \pm 3\%$), providing little support for gene inactivation and erosion (Liu et al., 2004).

Among the expressed genes, an average of 66.7% (206 ± 132 ORFs per MAG) encoded proteins with a predicted function (Supplementary Figure 3 and Supplementary dataset). However, only 13% of the genes used for the metabolic characterization of the MAGs were expressed in more than 50% of the samples, and 52% of the samples were not expressed in any of the samples (Figure 3 and Supplementary Figure 4). Furthermore, $45 \pm 19\%$ of the most expressed genes in MAGs were not associated with a known function (Supplementary Figure 2), confirming that substantial activities remain uncharacterized. These results have major consequences for the extrapolation of the *Patescibacteria* metabolism and environmental function. For instance, transcripts of alcohol, lactate, or malate dehydrogenases were rarely detected, challenging the fermentative metabolism assumed from the detection of these genes in the genomic data. Although we cannot exclude that some transcripts were not sequenced, these results suggest that essential genes and pathways remain uncharacterized in the microbial dark matter or unidentified due to overly divergent sequences compared with existing databases. This indicates that *in situ* activities of *Patescibacteria* could hardly be inferred from metagenomic data alone and that activity-based and cultivation data are mandatory.

Data availability statement

The datasets presented in this study can be found in NCBI Genbank under the Bioproject accession number SRP063590.

Author contributions

AV and PC analyzed the data, designed the figures, and wrote the manuscript with contributions and corrections from RG and MG-U. AV, RG, and MG-U designed the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the French National Research Agency [ANR FUNHYMAT ANR11 BSV7 014 01 and ANR MICROMER ANR-21-CE34-0014] and by the Excellence Initiative of Université de Pau et des Pays de l'Adour – I-Site E2S UPPA [Hub MeSMic], a French Investissements d'Avenir program.

Acknowledgments

The authors thank Johanne Aubé for collaborating in the dataset production.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1073483/full#supplementary-material>

References

- Adam, P. S., Borrel, G., and Gribaldo, S. (2018). Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci.* 115, E1166–E1173. doi: 10.1073/pnas.1716667115
- 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, 13219. doi: 10.1038/ncomms13219
- Aubé, J., Senin, P., Bonin, P., Pringault, O., Jeziorski, C., Bouchez, O., et al. (2020). Meta-omics Provides Insights into the Impact of Hydrocarbon Contamination on Microbial Mat Functioning. *Microb. Ecol.* 80, 286–295. doi: 10.1007/s00248-020-01493-x
- Aubé, J., Senin, P., Pringault, O., Bonin, P., Deflandre, B., Bouchez, O., et al. (2016). The impact of long-term hydrocarbon exposure on the structure, activity, and biogeochemical functioning of microbial mats. *Mar. Pollut. Bull.* 111, 115–125. doi: 10.1016/j.marpolbul.2016.07.023
- Beam, J. P., Becraft, E. D., Brown, J. M., Schulz, F., Jarett, J. K., Bezuidt, O., et al. (2020). Ancestral absence of electron transport chains in patescibacteria and DPANN. *Front. Microbiol.* 11, 1848. doi: 10.3389/fmicb.2020.01848
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bor, B., Collins, A. J., Murugkar, P. P., Balasubramanian, S., To, T. T., Hendrickson, E. L., et al. (2020). Insights obtained by culturing saccharibacteria with their bacterial hosts. *J. Dent. Res.* 99, 685–694. doi: 10.1177/0022034520905792
- Bor, B., McLean, J. S., Foster, K. R., Cen, L., To, T. T., Serrato-Guillen, A., et al. (2018). Rapid evolution of decreased host susceptibility drives a stable relationship between ultrasmall parasite TM7x and its bacterial host. *Proc. Natl. Acad. Sci.* 115, 12277–12282. doi: 10.1073/pnas.1810625115
- Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208. doi: 10.1038/nature14486
- Campbell, J. H., O'Donoghue, P., Campbell, A. G., Schwientek, P., Sczyrba, A., Woyke, T., et al. (2013). COG is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl. Acad. Sci.* 110, 5540–5545. doi: 10.1073/pnas.1303901110
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663
- Castelle, C. J., and Banfield, J. F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* 172, 1181–1197. doi: 10.1016/j.cell.2018.02.016
- Castelle, C. J., Brown, C. T., Anantharaman, K., Probst, A. J., Huang, R. H., and Banfield, J. F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645. doi: 10.1038/s41579-018-0076-2
- Chaudhari, N. M., Overholt, W. A., Figueroa-Gonzalez, P. A., Taubert, M., Bornemann, T. L. V., Probst, A. J., et al. (2021). The economical lifestyle of CPR bacteria in groundwater allows little preference for environmental drivers. *Environ. Microbiome* 16, 24. doi: 10.1186/s40793-021-00395-w
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848
- Chiriac, M.-C., Bulzu, P.-A., Andrei, A.-S., Okazaki, Y., Nakano, S., Haber, M., et al. (2022). Ecogenomics sheds light on diverse lifestyle strategies in freshwater CPR. *Microbiome* 10, 84. doi: 10.1186/s40168-022-01274-3
- Danczak, R. E., Johnston, M. D., Kenah, C., Slattery, M., Wrighton, K. C., and Wilkins, M. J. (2017). Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome* 5, 112. doi: 10.1186/s40168-017-0331-1
- Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281. doi: 10.1093/nar/gkaa1018
- Haft, D. H., Selengut, J. D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. doi: 10.1093/nar/gkg128
- He, X., McLean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S.-Y., et al. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci.* 112, 244–249. doi: 10.1073/pnas.1419038112
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1, 16048. doi: 10.1038/nmicrobiol.2016.48
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi: 10.1186/1471-2105-11-119
- Jaffe, A. L., Konno, M., Kawasaki, Y., Kataoka, C., Béjà, O., Kandori, H., et al. (2022). Saccharibacteria harness light energy using type-1 rhodopsins that may rely on retinal sourced from microbial hosts. *ISME J.* 16, 2056–2059. doi: 10.1038/s41396-022-01231-w
- Jaffe, A. L., Thomas, A. D., He, C., Keren, R., Valentin-Alvarado, L. E., Munk, P., et al. (2021). Patterns of gene content and co-occurrence constrain the evolutionary path toward animal association in candidate phyla radiation bacteria. *mBio* 12, e00521–e00521. doi: 10.1128/mBio.00521-21
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165. doi: 10.7717/peerj.1165
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- King, G. M., and Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5, 107–118. doi: 10.1038/nrmicro1595
- Krom, B. P., Warner, J. B., Konings, W. N., and Lolkema, J. S. (2003). Transporters involved in uptake of di- and tricarboxylates in *Bacillus subtilis*. *Antonie Van Leeuwenhoek* 84, 69–80. doi: 10.1023/A:1024445131925
- Kuroda, K., Yamamoto, K., Nakai, R., Hirakata, Y., Kubota, K., Nobu Masaru, K., et al. (2022). Symbiosis between candidatus patescibacteria and archaea discovered in wastewater-treating bioreactors. *mBio* 13, e01711–e01722. doi: 10.1128/mbio.01711-22
- León-Zayas, R., Peoples, L., Biddle, J. F., Podell, S., Novotny, M., Cameron, J., et al. (2017). The metabolic potential of the single cell genomes obtained from the Challenger Deep, Mariana Trench within the candidate superphylum Parcubacteria (OD1). *Environ. Microbiol.* 19, 2769–2784. doi: 10.1111/1462-2920.13789
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 3, 33. doi: 10.1093/bioinformatics/btv033
- Liu, Y., Harrison, P. M., Kunin, V., and Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* 5, R64. doi: 10.1186/gb-2004-5-9-r64
- Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6, 6372. doi: 10.1038/ncomms7372
- Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I.-M. A., Chu, K., and Kyrpides, N. C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271–2278. doi: 10.1093/bioinformatics/btp393
- McDaniel, E. A., Anantharaman, K., and McMahon, K. D. (2020). metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. *bioRxiv* 2019.12.20.884627. doi: 10.1101/2019.12.20.884627
- Moreira, D., Zivanovic, Y., López-Archilla, A. I., Iniesto, M., and López-García, P. (2021). Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat. Commun.* 12, 2454. doi: 10.1038/s41467-021-22762-4
- Murugkar, P. P., Collins, A. J., Chen, T., and Dewhirst, F. E. (2020). Isolation and cultivation of candidate phyla radiation Saccharibacteria (TM7) bacteria in coculture with bacterial hosts. *J. Oral Microbiol.* 12, 1814666. doi: 10.1080/20002297.2020.1814666

- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Nobu, M. K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S. G., Woyke, T., et al. (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 9, 1710–1722. doi: 10.1038/ismej.2014.256
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- 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
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5. doi: 10.1371/journal.pone.0009490
- 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
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352
- Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315, 1003–1006. doi: 10.1126/science.1135999
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Tian, R., Ning, D., He, Z., Zhang, P., Spencer, S. J., Gao, S., et al. (2020). Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* 8, 51. doi: 10.1186/s40168-020-00825-w
- Vigeneron, A., Cruaud, P., Aubé, J., Guyoneaud, R., and Goñi-Urriza, M. (2021). Transcriptomic evidence for versatile metabolic activities of mercury cycling microorganisms in brackish microbial mats. *Npj Biofilms Microbiomes* 7, 83. doi: 10.1038/s41522-021-00255-y
- Vigeneron, A., Cruaud, P., Langlois, V., Lovejoy, C., Culley, A. I., and Vincent, W. F. (2020). Ultra-small and abundant: candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. *Limnol. Oceanogr. Lett.* 5, 212–220. doi: 10.1002/lol2.10132
- Vigeneron, A., Cruaud, P., Lovejoy, C., and Vincent, W. F. (2022). Genomic evidence of functional diversity in DPANN archaea, from oxic species to anoxic vampiristic consortia. *ISME Commun.* 2, 4. doi: 10.1038/s43705-022-00088-6
- Vigeneron, A., Lovejoy, C., Cruaud, P., Kalenitchenko, D., Culley, A., and Vincent, W. F. (2019). Contrasting Winter Versus Summer Microbial Communities and Metabolic Functions in a Permafrost Thaw Lake. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.01656
- Wrighton, K. C., Castelle, C. J., Varaljay, V. A., Satagopan, S., Brown, C. T., Wilkins, M. J., et al. (2016). RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J.* 10, 2702–2714. doi: 10.1038/ismej.2016.53
- Yakimov, M. M., Merkel, A. Y., Gaisin, V. A., Pilhofer, M., Messina, E., Hallsworth, J. E., et al. (2021). Cultivation of a vampire: 'Candidatus Absconditicoccus praedator.' *Environ. Microbiol.* 3, 15823. doi: 10.1111/1462-2920.15823
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418