LETTER

Ultra-small and abundant: Candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem

Adrien Vigneron [1,2,3]* Perrine Cruaud [3,4] Valérie Langlois [2,3,4] Connie Lovejoy [1,2,3,5] Alexander I. Culley [2,3,4] Warwick F. Vincent [1,2,3]

[1]Département de Biologie, Université Laval, Québec, Québec, Canada; [2]Centre d'études nordiques (CEN), Takuvik Joint International Laboratory, Université Laval, Québec, Québec, Canada; [3]Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, Québec, Canada; [4]Département de Biochimie, de Microbiologie et de Bio-informatique, Université Laval, Québec, Québec, Canada; [5]Québec Océan, Université Laval, Québec, Québec, Canada

Scientific Significance Statement

Metagenomic analyses have recently uncovered a hidden radiation of the bacterial tree of life that may be widespread in freshwater ecosystems. However, these uncultured candidate phyla radiation (CPR) bacteria are poorly resolved with standard sampling methods and molecular protocols in aquatic microbial ecology. Combining a water filtration protocol for ultra-small particles with high-throughput metagenomics, we investigated the contribution of these microorganisms to the functioning of a permafrost thaw lake, a major type of aquatic ecosystem in northern landscapes and an intense source of greenhouse gas emissions to the atmosphere. Our results suggest that previously overlooked CPR bacteria are abundant in both summer and winter, and may play a key biogeochemical role in transforming complex organic carbon in aquatic environments.

Abstract

The candidate phyla radiation (CPR) is a diverse group of uncultured bacterial lineages with poorly understood metabolic functions. CPR bacteria can represent a large proportion of the total planktonic microbial community in subarctic thermokarst lakes, but their functional roles remain unexplored. We applied sequential water filtration and metagenomic shotgun sequencing to a peatland permafrost thaw lake, and found high proportions of CPR bacteria in both summer and winter (> 40% of 16S rRNA reads in the 0.02–0.22 μm pore-size fraction). The metagenome-assembled genomes of CPR bacteria representatives showed capacities to degrade and ferment permafrost- and peatland-derived organic matter. Potential products of their metabolic activities included acetate, CO_2, and hydrogen, implying a syntrophic relationship

*Correspondence: avignero@gmail.com

Associate editor: Emily Stanley

Additional Supporting Information may be found in the online version of this article.

with other community members, including methanogens and methanotrophs. The results indicate biogeochemical interdependencies in organic matter utilization within thermokarst microbial communities, with CPR members playing a key intermediate role in carbon and methane cycling.

The candidate phyla radiation (CPR) is a large and diverse radiation of bacteria that to date consists exclusively of uncultured microorganisms. CPR bacteria tend to be absent or underestimated in 16S rRNA gene amplicon surveys, likely because of introns within their 16S rRNA gene that interfere with polymerase chain reaction (PCR) amplification, and mismatches with standard "universal" primers (Brown et al. 2015). However, genome resolved metagenomics and single cell sequencing of environmental samples have revealed this large and previously hidden bacterial radiation. With more than 74 phyla clustered within two major superphyla, the *Parcubacteria* and the *Microgenomates*, CPR taxa may contribute up to half of the global bacterial biodiversity (Rinke et al. 2013; Brown et al. 2015; Hug et al. 2016; Castelle et al. 2018).

High proportions of CPR bacteria have been recently reported in subarctic thermokarst lakes and ponds (Wurzbacher et al. 2017; Vigneron et al. 2019). These small waterbodies form by the thawing and collapse of ice-rich permafrost, and are abundant throughout the subarctic region. In these ecosystems, the permafrost-derived organic matter is mobilized and transformed by microbial activities, leading to high rates of greenhouse gas emission to the atmosphere (Matveev et al. 2016). Geochemical conditions such as oxygen, methane, and sulfide concentrations contrast markedly between the ice-free and ice-cover periods (Matveev et al. 2019), and this seasonal change is accompanied by pronounced shifts in microbial community structure and metabolic capacities (Vigneron et al. 2019). In the oxygenated surface waters in summer, the microbial communities are predominantly aerobic bacteria (*Betaproteobacteria* and *Actinobacteria*, Fig. 1), with many having the capacity for phototrophic metabolism, and the enzymatic potential to degrade labile organic matter or to oxidize one-carbon compounds (Vigneron et al. 2019). In contrast, the microbiome of the fully anoxic water column in winter is composed of anaerobic bacteria (including *Planctomycetes*, *Deltaproteobacteria*, *Chloroflexi*, and methanogens, Fig. 1) that may be involved in syntrophic and methanogenic degradation of complex organic matter (Vigneron et al. 2019). Using metagenomic data sets, CPR bacteria, including *Parcubacteria* and *Microgenomates*, were found to represent a large proportion (15–25%) of the microbial community in a subarctic thermokarst lake in both seasonal contexts (Vigneron et al. 2019), raising questions about their potential roles and lifestyles in such a poorly explored environment.

CPR bacteria were initially recovered from groundwaters and aquifers, where they have been found to account for up to 20% of the microbial community (Kantor et al. 2013; Luef et al. 2015). They have also been detected as minority members of the community in other environments, including surface freshwaters (Linz et al. 2017), terrestrial systems (Parks et al. 2017) and marine habitats (Dombrowski et al. 2017; Tully et al. 2018), as well as in animal microbiomes (Camanocha and Dewhirst 2014). Consistent with their enrichment in groundwater samples that have been prefiltered through 0.22 μm filters (Luef et al. 2015; Anantharaman et al. 2016; Danczak et al. 2017; Castelle et al. 2018), transmission electron microscopy indicates that some of the *Parcubacteria* members are extremely small (0.1 μm in diameter; Luef et al. 2015). This "ultra-small" cell size could also explain their relatively low proportions in aquatic microbial samples collected on 0.22 μm filters, since they would be poorly retained on such filters.

The functional capacities of CPR phyla in natural ecosystems remain unclear as their apparently reduced genomes (~ 1 Mb; Brown et al. 2015) are mainly composed of uncharacterized genes. In addition, some of the core metabolic genes thought to be essential in other bacterial phyla are consistently absent in the metagenomic assembled genomes (Castelle et al. 2018). Extensive genome data mining of *Parcubacteria* and *Microgenomates* has indicated a lack of known respiratory pathways for the large majority of phyla, suggesting syntrophic/symbiotic and fermentative lifestyles for these bacteria, which could also help to explain the absence of cultured representatives (Wrighton et al. 2012; Castelle et al. 2018). However, genomes of some *Parcubacteria* species indicate a potential involvement in nitrogen cycling (Castelle et al. 2017; Danczak et al. 2017; León-Zayas et al. 2017). Analysis of glycoside hydrolase (GH) families suggests that the metabolic capacity for amylose and cellulose degradation is widespread within the radiation, indicating a role in carbon cycling and complex organic matter degradation. Other catabolic capabilities have also been reported implying differences in ecological niches and carbon substrate utilization (Danczak et al. 2017).

In the present study, we combined sequential filtration of thermokarst waters on 0.22 and 0.02 μm pore-size filters and PCR-free metagenomic sequencing to determine whether the distribution of CPR members in thermokarst lakes might have been previously underestimated due to the small cell-size of some representatives. We then used metagenomic binning to investigate the metabolic capacities of these microbial lineages and to address the question of their role in permafrost organic carbon cycling in thermokarst lake ecosystems.

Methods

Water samples were collected from a 2.8 m deep thermokarst lake in the Sasapimakwananisikw (SAS) River Valley located in the sporadic permafrost zone of subarctic
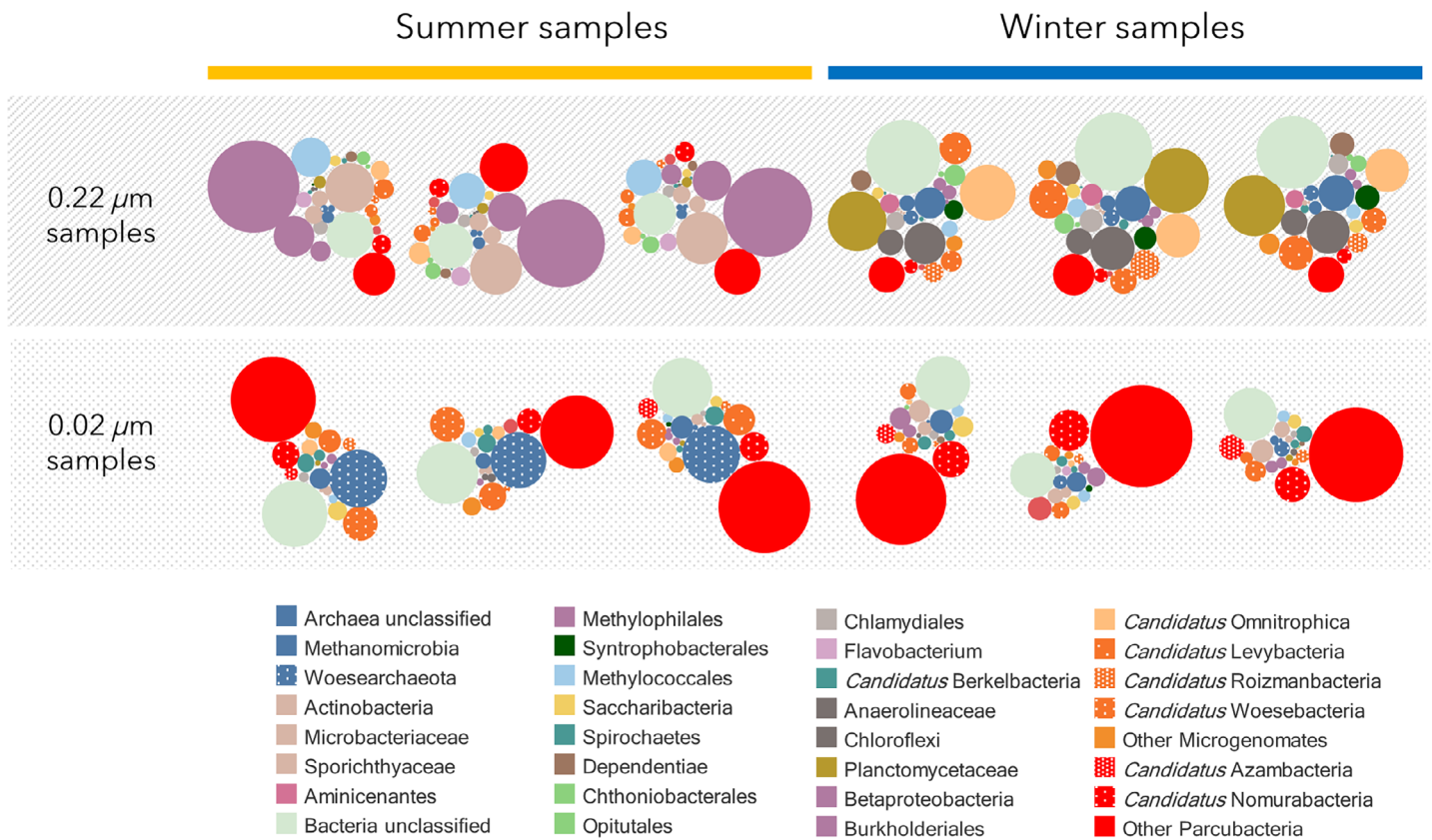
Fig. 1. Microbial community composition of SAS2A thermokarst lake samples based on 16S rRNA genes identified in the metagenomic data sets. The size of the circles represents the relative abundance of genes for each lineage. Only lineages with relative proportions > 1% of the reads in at least one sample are shown.

Quebec. Water samples were collected from the anoxic water column under ice during winter 2015 and at the oxycline (0.5 m depth) during summer 2016 as previously described in Vigneron et al. (2019). Further site details are given in the Supporting Information. Three independently sampled replicates were analyzed per season and all water samples were collected by serial filtration onto 0.22 Sterivex™ filter units (EMD Millipore, MA, U.S.A.) then 0.02 μm Anotop™ filters (Whatman, GE Healthcare, UK) using three independent lines of a peristaltic pump system. The filters were immediately frozen and stored at −70°C until DNA extraction.

For the six 0.22 μm Sterivex samples, nucleic acids were extracted using Qiagen Allprep DNA/RNA Mini Kit with modifications (Cruaud et al. 2017). For the six 0.02 μm Anotop filters, nucleic acids were extracted using Master Pure Complete DNA purification kit as in Mueller et al. (2014). Due to the nature of the filters (membrane or glass microfibers) a single DNA extraction procedure was not possible, nonetheless both DNA extraction protocols were composed of similar enzymatic (Lysozyme and Proteinase K) and chemical lysis. Metagenomic libraries were prepared using Illumina Nextera XT kit (for Sterivex samples) and SWIFT Accel NGS® 1S kit (allowing lower DNA input for Anotop samples) following the manufacturer's

recommendations. DNA was sequenced on Illumina Nextseq for 0.22 μm filters (at the CGEB—Integrated Microbiome Resource, Dalhousie University) and on an Illumina HiSeq for the 0.02 μm filters (Génome Québec sequencing facility, University McGill), both producing 150 bp reads. The bioinformatics methods for microbial community composition analysis and metagenomic assembly, binning and analysis are given in the Supporting Information. Bacterial relative abundance was determined by quantitative PCR using the 1369f/1492r primer set as detailed in Supporting Information.

Results and discussion

Ultra-small bacteria in thermokarst lake waters

Specific bacterial lineages that passed through the 0.22 μm pore-size filters but were retained by 0.02 μm pore-size filters were identified by 16S rRNA genes extracted from metagenomes of winter and summer samples (Fig. 1). The sequential filtration enriched for small-sized cells, but spores, bacteria with flexible membrane, viruses, and environmental DNA adsorbed onto small particles would also have been recovered in this fraction. CPR bacteria represented on average (± SD), 55.7% ± 7.8% and 43.0% ± 7.7% of the 16S rRNA gene reads on the 0.02 μm filters

in winter and summer samples, respectively (Fig. 1). For comparison, CPR bacteria represented on average $16.2\% \pm 3.6\%$ and $13.4\% \pm 2.3\%$ of the community on the 0.22 μm filters in winter and summer samples, respectively. In particular, *Parcubacteria* lineages (including *Candidatus* Azambacteria and *Candidatus* Nomurabacteria) were significantly enriched (12 times greater) in metagenomes from the 0.02 μm filters compared to the 0.22 μm filters (*t*-test, $p = 0.003$), regardless of season (Fig. 1). These results extend previous observations carried out on groundwater samples to permafrost pond environments and are consistent with the small cell size previously inferred or observed for these lineages (Brown et al. 2015; Luef et al. 2015; Danczak et al. 2017). Nonetheless, other *Parcubacteria* lineages, including Ca. Moranbacteria and Ca. Falkowbacteria and *Microgenomates* lineages, including Ca. Daviesbacteria and Ca. Levybacteria were detected in similar proportions in both size filters (Fig. 1).

The abundance of bacterial 16S rRNA genes on the 0.02 and 0.22 μm filters was estimated by quantitative PCR (Fig. 2). Our results for the 0.02–0.22 μm fraction indicated an average of 9.9×10^3 16S rRNA genes mL^{-1} and 4.16×10^5 16S rRNA genes mL^{-1} in winter and summer, respectively, while the bacterial abundance quantified on the 0.22 μm filters averaged 4.03×10^5 and 6.74×10^5 16S rRNA genes mL^{-1} for the same seasons. Assuming the same gene copy numbers per cell in both fractions, this would imply that bacteria passing through 0.22 μm filters represent 2.4% of the total bacterial community (0.22 + 0.02 μm filters together) in winter and 38.2% in summer (Fig. 2). However, all complete CPR genomes reported to date have only one copy of the 16S rRNA gene and some CPR bacteria would also likely escape quantification due to primer mismatches (Brown et al. 2015). These abundance
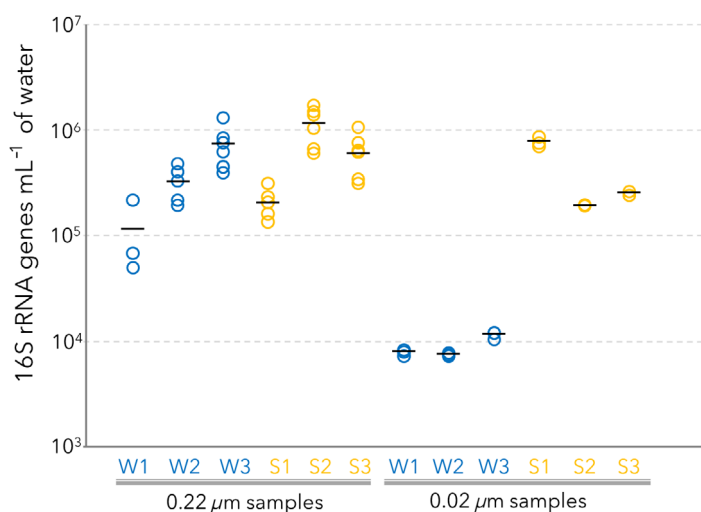
estimates therefore represent lower bounds for total bacteria and relative abundance compared to lineages with multiple 16S rRNA gene copies.

The metagenome analyses indicated that *Parcubacteria* were predominant members of the 0.02–0.22 μm community in both summer (31% of the 16S rRNA reads) and winter (51% of the 16S rRNA reads), despite differences in the microbial community composition on the 0.22 μm filter samples (Fig. 1) and geochemical changes between the summer and winter, including in oxygen and sulfide concentrations (Vigneron et al. 2019). This result is also supported by previous RNA-based amplicon sequencing carried out on the same samples that identified *Parcubacteria* members in both summer and winter samples (Vigneron et al. 2019). Although this result only concerns the *Parcubacteria* lineage that can be amplified by standard primer sets, it suggested that at least the PCR-detectable part of the *Parcubacteria* are probably active in summer and winter seasons. Although the lower abundance observed by quantitative PCR (qPCR) in winter samples compared to summer samples should be taken into consideration, these results imply that *Parcubacteria* members may be more resistant than other lineages to oxygen concentrations and other variable factors observed in permafrost ponds, such as light availability, microbial grazers and viruses, and inhibitory compounds including hydrogen sulfide. This resistance might explain their success in thermokarst lake waters where strong seasonal environmental variations restructure the rest of the microbial community (Vigneron et al. 2019).

Based on their limited metabolic repertoire and absence of biosynthetic genes, previous studies have proposed an epi-symbiotic relationship for *Parcubacteria* whereby they are attached to the surface of other cells (Nelson and Stegen 2015; Castelle et al. 2018). However, the microbial communities were markedly different between summer and winter, and no other microbial lineage was found to covary with *Parcubacteria*. This weakens any hypothesis of a stable symbiotic host for *Parcubacteria* at least in this ecosystem type. Nonetheless, we cannot exclude interdependencies or syntrophy among *Parcubacteria* lineages.

In contrast to *Parcubacteria*, the relative proportion of *Microgenomates* lineages and *Candidatus* Berkelbacteria were enriched in the 0.02 μm filter samples only in summer, possibly contributing to the log difference in relative abundance detected by qPCR (9.9×10^3 16S rRNA genes mL^{-1} vs. 4.16×10^5 16S rRNA genes mL^{-1} in winter and summer, respectively; Fig. 2). These results suggest a different response of these lineages to the thermokarst aquatic environment than by the predominant *Parcubacteria* members, with the possibility of symbiotic relationships with taxa in the summer microbial community.

Carbon metabolism by thermokarst lake CPR bacteria

After binning of the coassembled metagenomes, we analyzed seven genomic bins affiliated to *Parcubacteria*, and one bin affiliated to *Microgenomates* (Supporting Information

Fig. 2. Relative abundance of the bacterial community of SAS2A thermokarst lake samples based on 16S rRNA gene quantification (qPCR). Average values are represented by black bars, for winter (W) and summer (S).

Table S1). The completeness of these bins ranged from 47% to 73% based on a set of 101 bacterial markers that includes genes consistently not detected in CPR bins (Nelson and Stegen 2015), and with a contamination level below 7% (Fig. 3). Based on 16S rRNA genes and other ribosomal protein (rps) genes analysis, two of the *Parcubacteria* bins were related to *Candidatus* Yanofskybacteria (bin13 and bin20), three to *Candidatus* Moranbacteria (bin9, 23, and 91) and a single genomic bin was related to *Candidatus* Kaiserbacteria (bin137). In addition, bin120 was also reconstructed but 16S rRNA gene was missing and rps genes analysis was unable assign this bin to a specific taxon within the *Parcubacteria* superphylum. Finally, the single bin affiliated to *Microgenomates* superphylum was related to *Candidatus* Daviesbacteria (bin78).

Using IMG/MR and KAAS servers for genome annotation and pathway reconstruction (Moriya et al. 2007; Markowitz et al. 2009), characterized genes were identified on approximately a third of the contigs (> 2000 bp), representing 249 ± 57 different KEGG orthologies (KO) per bin. Their genomic composition mirrored taxonomic affiliation, suggesting similar gene composition within members of the same phylum (Fig. 3). Although the incompleteness of the bins will have led to nondetection of certain genes, all *Parcubacteria* and *Microgenomates* genomic bins recovered from our samples lacked a respiratory chain and tricarboxylic acid cycle, consistent with fermentative metabolism that would be adapted to the winter anoxic thermokarst lake environment. Nonetheless, catalase (K03781), peroxiredoxin (K11188), and superoxide dismutase (K04565) genes were identified in the genomic bins, suggesting enzymatic resistance to oxygen for thermokarst lake *Parcubacteria*

lineages and potentially explaining their high relative proportion in summer oxic waters (Supporting Information Table S2). All genomic bins encode for the lower part of the glycolytic pathway (phosphofructokinase genes were not detected, excluding the possibility for neoglucogenesis by the usual enzymatic pathway), with the upper glycolysis being completed for three bins (bins 20, 78, and 137) by the nonoxidative pentose phosphate pathway (Fig. 3). Genes involved in downstream transformation of pyruvate to acetyl-CoA and acetate were detected in *Candidatus* Moranbacteria and *Candidatus* Yanofskybacteria, suggesting that these lineages might ferment carbon compounds to acetate (Fig. 3). The capacity to generate H_2 via NiFe hydrogenases was also found in *Candidatus* Yankofskybacteria bin13. Lactate dehydrogenase gene was detected in *Candidatus* Kaiserbacteria, suggesting that lactate may be an end product of its fermentation. In addition, the capacity for CO_2 producing (decarboxylative) malate fermentation was identified in *Candidatus* Moranbacteria and *Candidatus* Kaiserbacteria (Fig. 3). Although these results described only a limited part of the CPR bacteria community identified in our samples, these metabolic capabilities align with metabolic functions identified in *Parcubacteria* bins recovered from other freshwater environments (Nelson and Stegen 2015; Castelle et al. 2018), and extend our knowledge of CPR bacterial metabolism to thermokarst lakes and ponds, one of the most abundant ecosystem types in the circumpolar North.

Degradation of permafrost-derived organic matter by CPR bacteria

An average of 29 ± 12 carbohydrate-active enzyme (CAZy) genes per CPR genomic bin was identified (Fig. 4), indicating
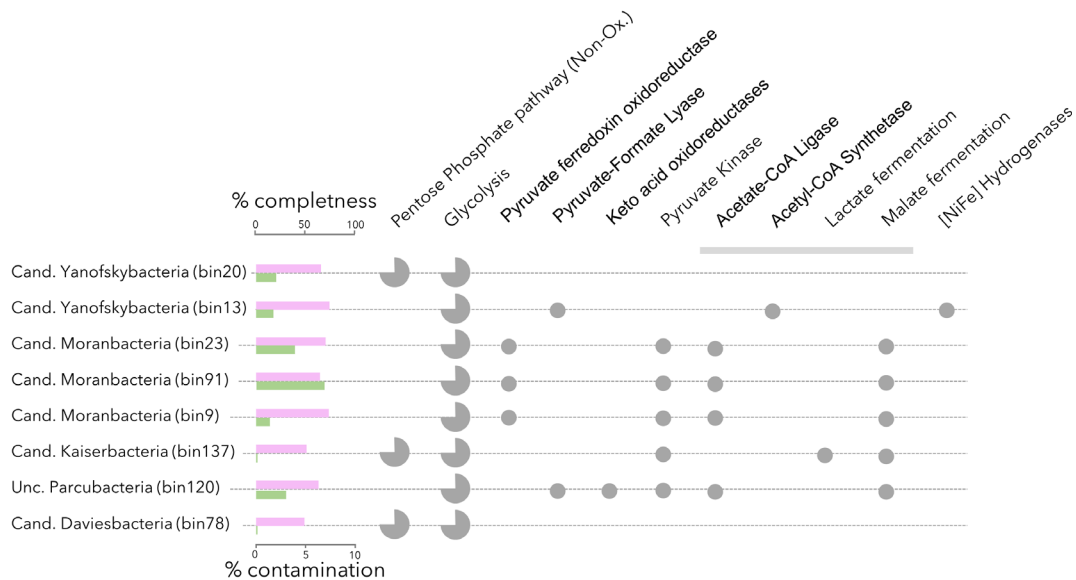
Fig. 3. Central carbon metabolism identified in the CPR bacteria bins. Bins are sorted according to their genome similarity, based on identified KOs and Bray-Curtis similarity index. Taxonomic affiliation of the bins was determined using 16S rRNA genes and other ribosomal protein genes. Pink bars: completeness (min: 47%, max: 73%). Green bars: contamination (min: 0.37%, max: 7%). Gray bar: potential for fermentations.

Fig. 4. Genes for carbohydrate active enzymes (CAZy) identified in the CPR bacteria bins. Bins were sorted according to their genome similarity, based on identified KOs and Bray-Curtis similarity index. Taxonomic affiliation of the bins was determined using 16S rRNA genes and other ribosomal protein genes. The size of the filled circles represents the number of detected genes (min: 1, max: 16). Green: plant or algal detritus degradation; Orange: chitin degradation; Blue: cell envelope/necromass degradation; Yellow: all classes of degradation. Percentages of CAZy genes were calculated based on the number of total coding regions per bin.

the potential for degradation of complex carbon substrates (Cantarel et al, 2009). The number of CAZy genes was lower than in other complex organic carbon degrading bacterial lineages from the thermokarst waters (Vigneron et al. 2019), but is consistent with previous analysis of CPR genomes (Danczak et al. 2017). Genes coding for amylases which usually occur in *Parcubacteria* genomes from marine and freshwater samples were detected in most of the bins along with the capacities for degrading cellulose (GH1) and mannose (GH130), supporting previous results from aquifer samples (Danczak et al. 2017). Cellulose and mannose are the most abundant and stable constituents of Sphagnum peats (Theander et al. 1954), which constitute most of the available or permafrost-stored organic carbon in our peatland permafrost study site (Arlen-Pouliot and Bhiry 2005; Matveev et al. 2016). Our results thereby might indicate the potential for permafrost and peatland-derived organic carbon degradation and fermentation by CPR bacteria in these permafrost thaw lakes. A capacity for chitin degradation (GH18) was also detected in *Candidatus* Moranbacteria, suggesting that some *Parcubacteria* members might also degrade remains of fungi and invertebrates. In addition, various genes coding glycosyltransferases (GT) were identified, suggesting that conversion of (poly)saccharides, (poly)peptides, and glycerolipids might be an important mechanism in CPR bacteria, which frequently lack de novo synthetic pathways for

these compounds. This metabolic potential would benefit these small bacteria since such conversions could aid cellular membrane synthesis and cell to cell interactions (León-Zayas et al. 2017), supporting a close relationship with other members of the community.

Who benefits from *Parcubacteria* metabolism?

Our results indicate that acetate would be one of the major end products of *Parcubacteria* metabolism. We therefore searched for potential acetate-utilizing microbial lineages based on the presence of acetyl-CoA synthetase (K01895), acetate kinase (K00925), or phosphate acetyltransferase (K00625) genes, conferring the ability for assimilatory or dissimilatory use of acetate (Wolfe 2005), in the genomic bins for the same samples (Vigneron et al. 2019). Various microbial lineages were identified as potential acetate utilizers and together represented up to $21\% \pm 1\%$ and $55\% \pm 1\%$ of the community on 0.22 μm pore-size filters in winter and summer, respectively (Fig. 5). Although acetyl-CoA synthetase is bidirectional and might be involved in acetate production (Wolfe 2005), acetate-utilizing microorganisms include members of the *Actinobacteria*, *Betaproteobacteria*, *Planctomycetes*, and *Chloroflexi* phyla as well as methanogens and methanotrophs. In addition, genes involved in acetate transport were also identified in most of these lineages, supporting this potential (Fig. 5). These results suggest that in addition to their CO_2 production by fermentation,
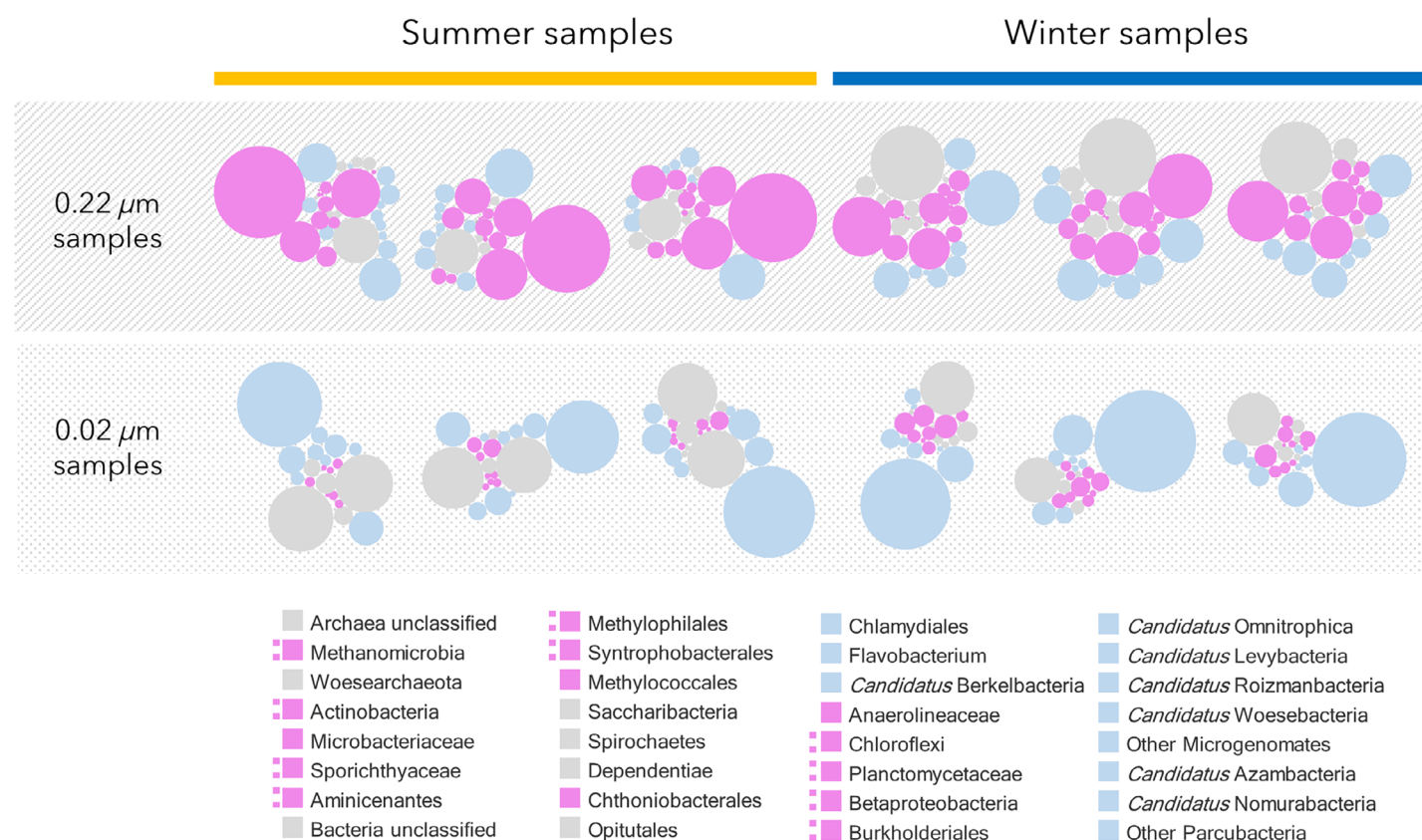
Fig. 5. Members of the thermokarst lake microbial communities with capacities for acetate utilization (in purple) based on the presence of acetyl-CoA synthetase (K01895), acetate kinase (K00925), or phosphate acetyltransferase (K00625) genes in related genomic bins. Double purple squares in the legend indicate the presence in genomic bins of genes coding for cation/acetate symporter (K14193), monocarboxylic acid transporter (K08179-90), and GPR1/FUN34/YaaH family proteins (PF01184) involved in acetate transport and uptake. Gray lineages: no genomic bins recovered. Blue lineages: genes not detected in related bins. The size of the circles represents the relative abundance of genes for each lineage.

certain end products of *Parcubacteria* metabolism might potentially fuel greenhouse-gas producing members of the microbial community.

Conclusions

Our results show that CPR bacteria are a major component of the thermokarst lake microbial community, accounting for more than 20% of the microbial reads detected on 0.22 μm filters and up to 56% of the microorganisms in 0.02 μm filters. These high percentages were maintained in both aerobic surface waters in summer and in the anoxic under-ice conditions in winter. These ultra-small microorganisms were also potentially abundant, representing quantitatively up to a third of the total bacterial community quantified in this study. Reconstruction of *Parcubacteria* genomes was consistent with previous reports and suggested that they may degrade and ferment permafrost and peatland-derived organic matter, thereby providing volatile fatty acids and hydrogen to methanogenic and other thermokarst pond microorganisms. In addition, they may also contribute to thermokarst pond CO_2 emissions through their

fermentative metabolism. Although our approach remains to be applied to a range of permafrost lakes and ponds before extrapolating these conclusions more widely, the results imply that *Parcubacteria* could play a major role in the carbon cycling network of these ecosystems and may act as catalysts for carbon transformation in all seasons. Given their poor resolution by standard water filtration methods and molecular protocols in aquatic microbial ecology, and their role suggested here in generating labile organic carbon substrates, the importance of the CPR group is likely to have been greatly underestimated or overlooked in lake ecosystems in general.

References

Anantharaman, K., and others. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. **7**: 13219. doi: 10.1038/ncomms13219

Arlen-Pouliot, Y., and N. Bhiry. 2005. Palaeoecology of a palsa and a filled thermokarst pond in a permafrost peatland, subarctic Québec, Canada. Holocene **15**: 408–419. doi:10. 1191/0959683605hl818rp

Brown, C. T., and others. 2015. Unusual biology across a group comprising more than 15% of domain bacteria. Nature **523**: 208. doi:10.1038/nature14486

Camanocha, A., and F. E. Dewhirst. 2014. Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions. J. Oral Microbiol. **6**. doi:10.3402/jom.v6.25468

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glyco-genomics. Nucleic Acids Res. **37**: D233–D238. doi:10.1093/nar/gkn663

Castelle, C. J., C. T. Brown, B. C. Thomas, K. H. Williams, and J. F. Banfield. 2017. Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. Sci. Rep. **7**: 40101. doi:10.1038/srep40101

Castelle, C. J., C. T. Brown, K. Anantharaman, A. J. Probst, R. H. Huang, and J. F. Banfield. 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

Cruaud, P., A. Vigneron, M.-S. Fradette, S. J. Charette, M. J. Rodriguez, C. C. Dorea, and A. I. Culley. 2017. Open the Sterivex™ casing: An easy and effective way to improve DNA extraction yields. Limnol. Oceanogr.: Methods **15**: 1015–1020. doi:10.1002/lom3.10221

Danczak, R. E., M. D. Johnston, C. Kenah, M. Slattery, K. C. Wrighton, and M. J. Wilkins. 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

Dombrowski, N., K. W. Seitz, A. P. Teske, and B. J. Baker. 2017. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiome **5**: 106. doi:10.1186/s40168-017-0322-2

Hug, L. A., and others. 2016. A new view of the tree of life. Nat. Microbiol. **1**: 16048. doi:10.1038/nmicrobiol.2016.48

Kantor, R. S., K. C. Wrighton, K. M. Handley, I. Sharon, L. A. Hug, C. J. Castelle, B. C. Thomas, and J. F. Banfield. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio **4**: e00708–e00713. doi:10.1128/mBio.00708-13

León-Zayas, R., L. Peoples, J. F. Biddle, S. Podell, M. Novotny, J. Cameron, R. S. Lasken, and D. H. Bartlett. 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

Linz, A. M., B. C. Crary, A. Shade, S. Owens, J. A. Gilbert, R. Knight, and K. D. McMahon. 2017. Bacterial community composition and dynamics spanning five years in freshwater bog lakes. mSphere **2**: e00169–e00117. doi:10.1128/mSphere.00169-17

Luef, B., and others. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat. Commun. **6**: 6372. doi:10.1038/ncomms7372

Markowitz, V. M., K. Mavromatis, N. N. Ivanova, I.-M. A. Chen, K. Chu, and N. C. Kyrpides. 2009. IMG ER: A system for microbial genome annotation expert review and curation. Bioinformatics **25**: 2271–2278. doi:10.1093/bioinformatics/btp393

Matveev, A., I. Laurion, B. N. Deshpande, N. Bhiry, and W. F. Vincent. 2016. High methane emissions from thermokarst lakes in subarctic peatlands. Limnol. Oceanogr. **61**: S150–S164. doi:10.1002/lno.10311

Matveev, A., I. Laurion, B. N. Deshpande, and W. F. Vincent. 2019. Concentrations of dissolved methane, carbon dioxide and oxygen in thermokarst lakes and ponds in palsa peatlands, northern Quebec, Canada. Nordicana **D48**. doi:10.5885/45588CE-5A12A84DFAAF4D36

Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. **35**: W182–W185. doi:10.1093/nar/gkm321

Mueller, J. A., A. I. Culley, and G. F. Steward. 2014. Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. Appl. Environ. Microbiol. **80**: 3930–3942. doi:10.1128/AEM.00245-14

Nelson, W., and J. Stegen. 2015. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front. Microbiol. **6**: 713. doi:10.3389/fmicb.2015.00713

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

Rinke, C., and others. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature **499**: 431–437. doi:10.1038/nature12352

Theander, O., H. Soling, E. J. Matikkala, and N. A. Sörensen. 1954. Studies on *Sphagnum* peat. III. A quantitative study on the carbohydrate constituents of *Sphagnum* mosses and *Sphagnum* peat. Acta Chem. Scand. **8**: 989–1000. doi:10.3891/acta.chem.scand.08-0989

Tully, B. J., E. D. Graham, and J. F. Heidelberg. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci. Data **5**: 170203. doi:10.1038/sdata.2017.203

Vigneron, A., C. Lovejoy, P. Cruaud, D. Kalenitchenko, A. Culley, and W. F. Vincent. 2019. Contrasting winter versus summer microbial communities and metabolic functions in a permafrost thaw lake. Front. Microbiol. **10**: 1656. doi:10.3389/fmicb.2019.01656

Wolfe, A. J. 2005. The acetate switch. Microbiol. Mol. Biol. Rev. **69**: 12. doi:10.1128/MMBR.69.1.12-50.2005

Wrighton, K. C., and others. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science **337**: 1661–1665. doi:10.1126/science.1224041

Wurzbacher, C., R. H. Nilsson, M. Rautio, and S. Peura. 2017. Poorly known microbial taxa dominate the microbiome of permafrost thaw ponds. ISME J. **11**: 1938. doi:10.1038/ismej.2017.54