AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Occurrence, Diversity, and Genomes of "*Candidatus* Patescibacteria" along the Early Diagenesis of Marine Sediments

Rui Zhao,[a,c] Ibrahim F. Farag,[a] Steffen L. Jørgensen,[b] Jennifer F. Biddle[a]

[a]School of Marine Science and Policy, University of Delaware, Lewes, Delaware, USA

[b]Centre for Deep Sea Research, Department of Earth Science, University of Bergen, Bergen, Norway

[c]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

ABSTRACT The phylum "*Candidatus* Patescibacteria" (or Candidate Phyla Radiation [CPR]) accounts for roughly one-quarter of microbial diversity on Earth, but the presence and diversity of these bacteria in marine sediments have been rarely charted. Here, we investigate the abundance, diversity, and metabolic capacities of CPR bacteria in three sediment sites (Mohns Ridge, North Pond, and Costa Rica Margin) with samples covering a wide range of redox zones formed during the early diagenesis of organic matter. Through metagenome sequencing, we found that all investigated sediment horizons contain "*Ca.* Patescibacteria" (0.4 to 28% of the total communities), which are affiliated with the classes "*Ca.* Paceibacteria," "*Ca.* Gracilibacteria," "*Ca.* Microgenomatia," "*Ca.* Saccharimonadia," "*Ca.* ABY1," and "*Ca.* WWE3." However, only a subset of the diversity of marine sediment "*Ca.* Patescibacteria," especially the classes "*Ca.* Paceibacteria" and "*Ca.* Gracilibacteria," can be captured by 16S rRNA gene amplicon sequencing with commonly used universal primers. We recovered 11 metagenome-assembled genomes (MAGs) of CPR from these sediments, most of which are novel at the family or genus level in the "*Ca.* Paceibacteria" class and are missed by the amplicon sequencing. While individual MAGs are confined to specific anoxic niches, the lack of capacities to utilize the prevailing terminal electron acceptors indicates that they may not be directly selected by the local redox conditions. These CPR bacteria lack essential biosynthesis pathways and may use a truncated glycolysis pathway to conserve energy as fermentative organotrophs. Our findings suggest that marine sediments harbor some novel yet widespread CPR bacteria during the early diagenesis of organic matter, which needs to be considered in population dynamics assessments in this vast environment.

IMPORTANCE Ultrasmall-celled "*Ca.* Patescibacteria" have been estimated to account for one-quarter of the total microbial diversity on Earth, the parasitic lifestyle of which may exert a profound control on the overall microbial population size of the local ecosystems. However, their diversity and metabolic functions in marine sediments, one of the largest yet understudied ecosystems on Earth, remain virtually uncharacterized. By applying cultivation-independent approaches to a range of sediment redox zones, we reveal that "*Ca.* Patescibacteria" members are rare but widespread regardless of the prevailing geochemical conditions. These bacteria are affiliated with novel branches of "*Ca.* Patescibacteria" and have been largely missed in marker gene-based surveys. They do not have respiration capacity but may conserve energy by fermenting organic compounds from their episymbiotic hosts. Our findings suggest that these novel "*Ca.* Patescibacteria" are among the previously overlooked microbes in diverse marine sediments.

KEYWORDS marine sediments, early diagenesis, Patescibacteria, redox zones, Candidate Phyla Radiation, biogeochemistry, metagenome

Editor Jennifer B. Glass, Georgia Institute of Technology

Address correspondence to Rui Zhao, zhaorui087@gmail.com, or Jennifer F. Biddle, jfbiddle@udel.edu.

The authors declare no conflict of interest.

Received 17 August 2022
Accepted 14 November 2022
Published 5 December 2022

Marine sediments are known to harbor at least half of the microbial cells in the marine realm (1), which exert an essential control over the global biogeochemical cycling and climate (e.g., atmospheric oxygenation, CO_2 concentration) on Earth (2–4). The full-spectrum identities and functions of these organisms are crucial for us to fully understand the mechanisms of how they regulate the biogeochemical cycles and to better predict future environmental scenarios on the planet. With the advances in cultivation-independent methods, microbial diversity in a wide range of marine sediments has been studied via molecular methods (for examples, see references 5 to 8), revealing that organic carbon content and the presence/absence of oxygen are the critical controlling factors of the microbial community composition in marine sediments (7). However, due to the well-known biases associated with marker gene surveys (for examples, see references 9 and 10), it remains unclear whether we have captured all the microbial diversity in global marine sediments.

The Candidate Phyla Radiation (CPR, also called "*Ca.* Patescibacteria" [11]) represents a large group of bacterial lineages defined through single-cell genomes (11) and metagenome-assembled genomes (MAGs) (10, 12) and have been suggested to harbor one of the top 10 genera found in marine sediments (13). First detected by marker gene surveys in the environments (14), CPR is now predicted to contain >73 individual phyla and comprise roughly one-quarter of bacterial diversity on Earth (15–17). Although CPR was initially placed as the most basal branch of the tree of bacteria (12), this placement was recently challenged (18) and the evolutionary significance of CPR is still under debate. CPR genomes have been revealed to be widespread in groundwater environments (10, 19–24) and have also been documented in many other environments, such as soda lake sediments (25), a thermokarst lake ecosystem (26), wastewater treatment plants (27, 28), terrestrial hot springs (29–31), and diverse terrestrial subsurface environments (32). CPR bacteria in these diverse environments appear to share unusual characteristics, including small cell sizes, reduced genomic repertoires, and restricted metabolic and biosynthetic capacities (15, 33). They may conserve energy by fermenting organic matters from their hosts (33). Through the predation of their host cells (for examples, see references 27 and 34), CPR bacteria may have a profound ecological impact on controlling the population size and turnover of their prokaryotic host cells, which so far have largely been attributed to the overall energy availability (35–37) and viral lysis (37, 38). To date, despite that the presence and genomes of CPR have been reported in several marine sediment locations (for examples, see references 39 to 42), only a single study has been dedicated to CPR in sediments beneath the Mariana Trench (43), leaving the distribution, metabolic functions, and ecological impacts of CPR in the vast majority of the global marine sediments mostly uncharacterized.

Microbial surveys using the 16S rRNA gene as a phylogenetic marker theoretically can capture only a subset of the total CPR population, if they are present, in marine sediments. Although in typical microbiome research sedimentary microbes are not size selected before DNA extraction and thus should not be heavily masked from bulk DNA extraction, their high divergences and frequent insertions of 16S rRNA gene sequences prevent many specific CPR phyla from being detected by typical PCR surveys with the universal bacterial primer set 515F/806R (9, 10, 17). Assembled metagenomic surveys may recover more CPR genomic information. Also, because very few CPR bacteria in natural and engineering environments have recently been cultured (34, 44), obtaining CPR genomic information from natural habitats is valuable for deciphering their metabolic strategies and potential interactions with other microbes in the communities. As such, we utilized the 16S rRNA gene amplicon sequencing and metagenome sequencing data sets from three marine sediment sites with contrasting environmental contexts to study the occurrence, diversity, and metabolic potentials of CPR in marine sediments. Our results suggest that CPR bacteria of novel lineages are rare yet widespread across a wide range of sediment redox zones.

RESULTS AND DISCUSSION

Environmental context of the three sediment locations. We investigated CPR in marine sediment cores from three locations: Mohns Ridge (MR) (which is part of the

Version of record at: https://doi.org/10.1128/aem.01409-22

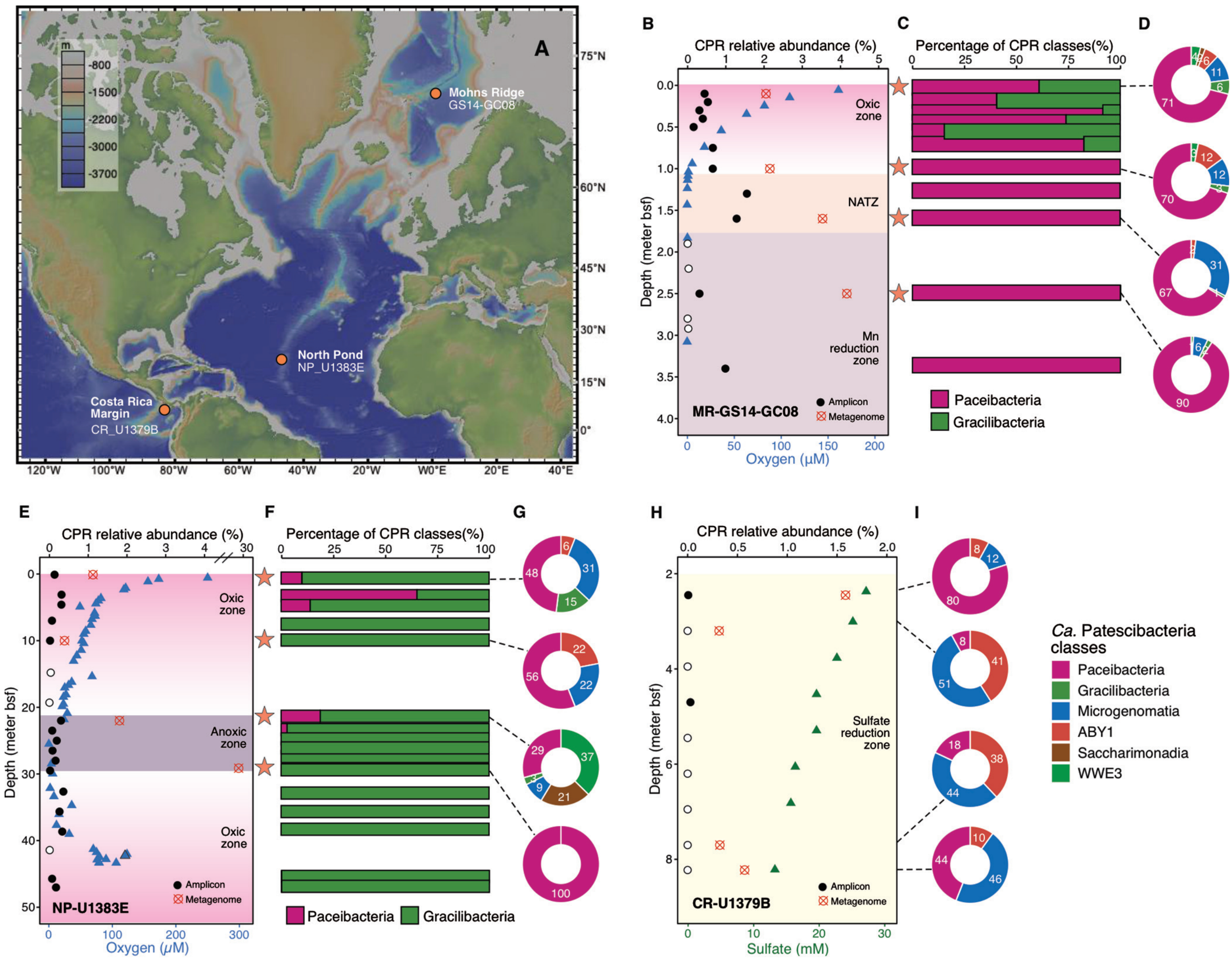

FIG 1 Occurrence of "*Ca.* Patescibacteria" (or CPR) in the three sediment sites investigated in this study. (A) Bathymetric map showing the locations of the three study sites. The map was made with GeoMapApp (www.geomapapp.org). (B to I) Relative abundance and community structure of CPR in cores MR-GS14-GC08 (B, C, and D), NP-U1383E (E, F, and G), and CR-U1379B (H and I) assessed by 16S rRNA genes recovered from metagenome sequencing and amplicon sequencing. In panels B, E, and H, CPR relative abundances assessed by metagenome sequencing are represented by red circles with crosses and those assessed by amplicon sequencing are represented by solid black circles. Sediment horizons with <10 reads of CPR in MR-GS14-GC08 and NP-U1383E are represented by open circles, while those in CR-U1379B horizons without any CPR reads are represented by open circles. Sediment horizons selected for metagenome sequencing in each core are highlighted with stars. The class-level classifications of "*Ca.* Patescibacteria" in the amplicon sequencing data are shown with bars (C and F), while for the metagenomes, they are shown using pie charts (D, G, and I). Redox zonation in MR-GS14-GC08, NP-U1383E, and CR-U1379B was determined based on the geochemical data reported in references 45, 46, and 47, respectively. In panels C, D, F, G, and I, CPR sequences were classified against the SILVA 138.1 release, and the community structure at the class level is reported.

Arctic Mid-Ocean Ridge) core MR-GS14-GC08, North Pond (NP) core NP-U1383E, and organic-rich sediments off the Costa Rica Margin (CR) at site U1379 (core CR- U1379B) (Fig. 1A), where concomitant metagenome and 16S rRNA gene amplicon sequencing data were published (45–47) and are available for detailed analyses of CPR. The sea-floor depths of these sites are in the range of 127 to 4,435 m. Sediments at these sites are characterized by a nearly constant rate of sediment deposition but without under-lying energy inputs like hydrothermal vents or episodic sediment changes like mass wasting events in trenches. Based on the existing geochemical profiles (see Fig. S1 in the supplemental material) and definitions described previously (45–47), the investi-gated sediments represent a variety of redox zones ranging from the oxic zone down to the sulfate reduction zone (Fig. 1), although they do not represent a coherent time series of sediment samples formed at a single location. The four sediment horizons from the MR core span the upper oxic zone (10 cm), the oxic-anoxic transition zone

(OATZ; 100 cm), the nitrate-ammonium transition zone (NATZ; 160 cm), and the Mn reduction zone (250 cm) (Fig. 1B). Additionally, a total of four oligotrophic sediment horizons from the NP core were included, two of which are located in the upper oxic zone (100 cm and 1,000 cm), and the remaining two represent the transition zones between the oxic and anoxic sediments (i.e., the oxic-anoxic transition zone [at ~2,200 cm] and the anoxic-oxic transition zone [AOTZ, at 2,950 cm below seafloor {bsf}]) (Fig. 1E). Finally, for the CR core (47), four sediment horizons in the interval of 2 to 9 m below the seafloor were included (Fig. 1H). Sulfate should be the dominant electron acceptor in sediments of this interval, due to the combination of the following three observations. (i) Sulfate in the porewater decreased with depth (Fig. 1H), although its depletion depth (~40 m below seafloor [mbsf]) was not covered by the investigated sediment layers of this study. (ii) The sulfur isotope of sulfate ($\delta^{34}SO_4$) increased with depth (48), indicating biological sulfate reduction. (iii) Dissolved Mn in the porewater, the product of Mn oxide reduction, showed a decreasing rather than increasing trend with depth in the investigated sediment interval (Fig. S1C), indicating that Mn oxides were no longer an important terminal electron acceptor. Therefore, this extensive data set allowed us to capture CPR diversity and genomes from a wide range of redox conditions formed in the sediment's early diagenesis.

Only part of the CPR diversity can be captured by 16S rRNA gene amplicon sequencing. We first examined the presence and diversity of CPR in the metagenome sequencing data. By searching and classifying the putative 16S rRNA gene reads using phyloFlash (49), we found that CPR bacteria account for 2.0 to 4.2% of the total microbial communities in MR core GC08 (Fig. 1B), 0.4 to 28% in NP (Fig. 1E), and 0.3 to 1.6% in CR (Fig. 1H). We noted that the variations in the relative abundance of CPR in different sediment layers, especially between those layers in the NP core, were not artifacts caused by different metagenome sequencing depths between samples, because the total 16S rRNA gene reads detected in the metagenome sequencing data sets were comparable (Table S1). In contrast, based on the concomitant 16S rRNA gene amplicon sequencing data, there were only 26 (1.1% of the total recovered operational taxonomic units [OTUs]), 27 (0.4% of the total), and 3 (0.3% of the total) OTUs affiliated with CPR at NP, MR, and CP, respectively. These CPR OTUs accounted for <2% of the total microbial communities of the three sediment sites (Fig. 1B, E, and H), lower than those estimated by metagenome sequencing. The discrepancy between these two methods may be related to the fact that the amplicon sequencing underestimates the relative abundance of CPR because most of the divergent and intron-containing 16S rRNA genes of CPR can evade the detection of universal primers (10). Nevertheless, these results suggested that members of CPR are usually part of the rare community (<5% of the total prokaryotic communities, irrespective of the detection method) yet are widespread (frequently encountered) in marine sediments of various redox zones, which was also supported by reports from marine sediments of other locations such as hadal trench sediments in the Pacific Ocean (41, 42).

Metagenome sequencing also revealed a higher diversity of CPR than did the 16S rRNA gene amplicon sequencing. Based on the taxonomical classification of the 16S rRNA gene reads in the metagenome sequencing data, CPR of the classes "*Candidatus* Paceibacteria," "*Ca*. Gracilibacteria," "*Ca*. Microgenomatia," "*Ca*. Saccharimonadia," "*Ca*. ABY1," and "*Ca*. WWE3" were present in the three sediment sites. Among the total 12 sediment layers where metagenome sequencing data are available, "*Ca*. Paceibacteria" was generally the most abundant CPR class, although the occasional dominances of "*Ca*. Gracilibacteria" at 22 mbsf of NP (Fig. 1G) and "*Ca*. Microgenomatia" and "*Ca*. ABY1" in the three deep layers of CR were also detected (Fig. 1I). In contrast, only members of the classes "*Ca*. Paceibacteria" and "*Ca*. Gracilibacteria" (see Fig. S2 for the phylogenetic placements of individual OTUs) were detected by 16S rRNA gene amplicon sequencing (Fig. 1C and F): "*Ca*. Paceibacteria" was the dominant CPR class (~80% of the total CPR community) in most of the sediment layers of MR core GC08 (Fig. 1C), and members of "*Ca*. Gracilibacteria" dominated in most of the examined sediment layers of NP (Fig. 1F). The discrepancy in obtained CPR community structure from the two methods again demonstrated that the amplicon sequencing method can probably capture only a subset of the total CPR population in complex communities.

Novelty of CPR MAGs recovered from marine sediments. To further characterize CPR in these sediments, we recovered CPR genomes from the existing metagenome sequencing data of these sites. We obtained four CPR MAGs (MR_Bin143, MR_Bin147, MR_Bin1662, and MR_Bin1762) from Mohns Ridge, two (NP_Bin194 and NP_Bin050) from North Pond, and five from sediments of the Costa Rica site (CR_Bin034, CR_Bin039, CR_Bin047, CR_Bin053, and CR_Bin021) (Table 1). These 11 CPR MAGs are considerably small in genome size (0.39 to 1.10 Mbp) and contain 12 to 153 scaffolds (Table 1). These genomes have 462 to 1,264 coding sequences, with the coding density varying in the range of 86.2 to 92.0% (Table 1). Based on the universal 43 single-copy genes of CPR bacteria (10), all except two (MR_Bin1762 and CR_Bin021) were estimated to be of >97.7% completeness with <2.3% redundancy (Table 1). CPR bacteria are known to lack some of the single-copy genes found in non-CPR bacteria (10), so based on the single-copy genes automatically determined for bacteria by CheckM (50), completion estimates were lower (61.0 to 80.3% complete) (Table 1). All of these CPR MAGs except CR_Bin039 have a reconstructed 16S rRNA gene and thus can be regarded as high-quality MAGs.

The novelty of these CPR MAGs was evident from phylogenetic analyses based on (i) the 14 concatenated ribosomal proteins (Fig. 2A) and (ii) the 16S rRNA gene (Fig. 2B). To ensure consistency with the current literature, we adopted the updated genomic distance-based classification scheme of GTDB (51) to name CPR lineages. While six CPR classes were present in the three sediment sites, as revealed by the bulk metagenome sequencing data (Fig. 1), all 11 recovered CPR MAGs were classified as members of novel genera, families, or even orders within the class "*Ca.* Paceibacteria." The phylogenetic novelty of these genomes was also supported by the GTDB classification, in which all the CPR MAGs showed <80% average nucleotide identities (ANI) with their most similar genomes included in the GTDB (07-RS207) database, and relative evolutionary divergence (RED) values of 0.74 to 0.83 were calculated. In particular, NP_Bin050 represented a new order (named o__JAHCSD01 in GTDB 07-RS207), with the closest phylogenetic relationship to members of the order *Portnoybacterales* (Fig. 2A). MR_Bin147 was a member of a new genus in the family GWB1-50-10 in the order UBA6257, an order mainly constituted by MAGs that previously were classified as *Jorgensenbacteria* recovered from the groundwater environment (20, 52) (Fig. 2A). The other CPR MAGs, belonging to the order "*Ca.* Paceibacterales," showed phylogenetic novelties at suborder levels. MR_Bin1662 and MR_Bin1762 formed a new family (we provisionally named it "*Candidatus* Bathypaceibacteraceae") (Fig. 2A). Three MAGs from the CR core (CR_Bin039, CR_Bin053, and CR_Bin034) also formed a new family (provisionally named "*Candidatus* Sedimentipaceibacteraceae") (Fig. 2A). NP_Bin194 represented a new genus within the RBG-13-36-15 family. In addition, MR_Bin143 represented a new genus in the UBA10102 family, which is composed mainly of genomes previously classified as *Wildermuthbacteria*. Finally, CR_Bin047 represented a new genus within the family GWA2-38-27 (Fig. 2A).

The phylogenetic novelty of these CPR MAGs was supported by the phylogenetic analysis of the 16S rRNA gene, which was broadly congruent with the concatenated ribosomal protein tree (Fig. 2B). It also showed that close relatives of these CPR MAGs were previously detected by clone libraries in anoxic marine sediments of various locations, including the Gulf of Mexico (53), Shimokita subseafloor sediments (54), Angola Basin sediments (55), and South Pacific Gyre ferromanganese nodules (56), although quantitative information about their abundance at these sites is lacking. The phylogenetic relatedness between CPR MAGs in marine sediments of different oceanographic regions may reflect the habitat or ecological niche preferences or the availability of the hosts of the CPR groups.

After integration of the 16S rRNA gene amplicon sequences into the 16S rRNA gene phylogenetic tree, it was clear that three of the four CPR MAGs recovered from MR were also captured by the 16S rRNA gene amplicon sequencing: MR_Bin1662 showed

TABLE 1 Summary of CPR MAGs recovered from marine sediments examined in this study

Parameter	Value for:										
	Mohns Ridge				North Pond		Costa Rica Margin				
	MR_Bin143	MR_Bin147	MR_Bin1662	MR_Bin1762	NP_Bin050	NP_Bin194	CR_Bin034	CR_Bin039	CR_Bin047	CR_Bin053	CR_Bin021
Genome size (Mbp)	1.10	0.79	0.64	1.00	0.80	0.84	0.80	0.48	0.82	0.61	0.39
No. of scaffolds	47	12	15	58	153	40	26	14	23	25	45
Matching OTU	OTU_152	OTU_525	—[c]	OTU_556	—	—	—	—	—	—	—
% GC	37.7	42.8	32.5	33.1	44.6	38.8	34.9	34.4	35.8	35.5	36.8
% Completion[a]	66.1	77.9	71.5	71.8	71.5	66.5	61.0	77.0	59.6	80.3	72.6
% Redundancy[a]	0.99	0	0	1.7	1.2	2.8	0	0	0	1.7	4.6
% Strain heterogeneity[a]	0	0	0	0	0	33	0	0	0	50	0
% Completion[b]	97.7	100	97.7	86.1	90.1	97.7	97.7	97.7	100	97.7	90.1
% Redundancy[b]	2.33	0	2.33	4.65	0	2.33	2.33	0	2.33	4.65	32.6
% Strain heterogeneity[b]	0	0	0	0	0	0	0	0	0	50	0
N_{50} of contigs	568,579	621,264	84,628	27,524	6,688	42,357	72,099	52,442	110,521	43,180	10,236
No. of coding sequences	1,264	799	671	1,077	833	908	831	492	813	623	462
% Coding density	91.4	90.2	91.6	88.7	86.2	88.0	86.9	91.6	88.3	91.2	92.0
rRNAs	4 (5S, 2 × 16S, 23S)	2 (16S, 23S)	3 (5S, 16S, 23S)	3 (5S, 16S, 23S)	3 (5S, 2 × 16S)	3 (5S, 16S, 23S)	3 (5S, 16S, 23S)	0	3 (5S, 16S, 23S)	2 (2 × 16S)	1 (16S)
tRNAs	69	46	40	46	38	48	48	42	48	42	37
NCBI accession no.	JANBVJ000000000	JANBVK000000000	JANBVL000000000	JANBVM000000000	JAHCSD000000000	JAHCSE000000000	JAHCRE000000000	JAHCRF000000000	JAHCRG000000000	JAHCRH000000000	JAHCRD000000000

[a]Genome quality assessed by CheckM with single-copy genes automatically determined for bacteria.

[b]Genome quality assessed by CheckM using a custom-defined workflow based on the 43 single-copy genes proposed by Brown et al. (10).

[c]—, unknown.

Version of record at: https://doi.org/10.1128/aem.01409-22

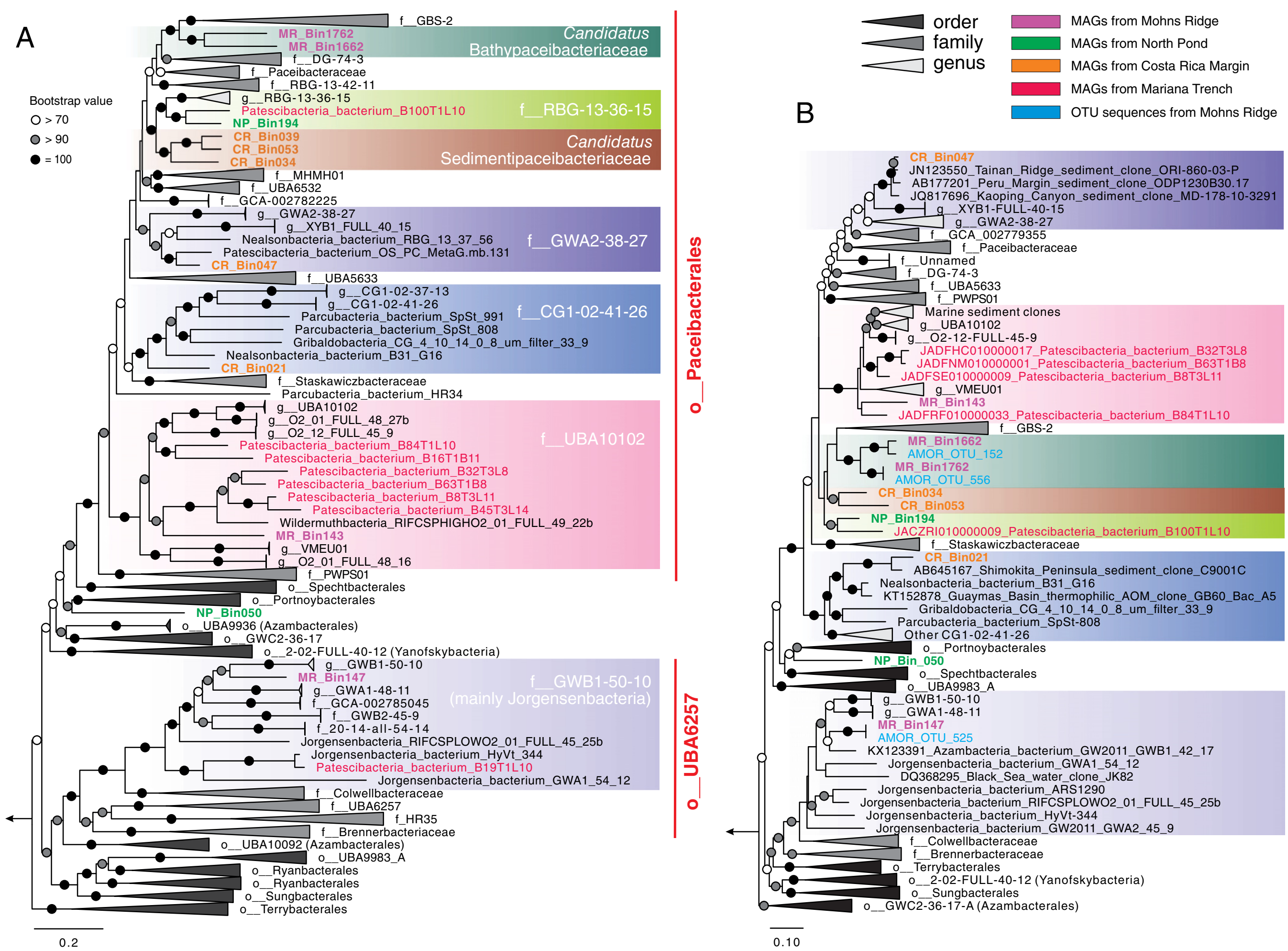

FIG 2 Maximum-likelihood phylogenetic trees of CPR genomes based on the concatenated ribosomal proteins (A) and the 16S rRNA gene (B). Both trees were inferred using IQ-TREE with the best-fit evolutional models and 1,000 bootstrap replicates. Both phylogenomic trees are rooted to four "*Ca.* Doudnabacteria" MAGs. In both trees, lineages harboring the MAGs recovered from the marine sediments investigated in this study are highlighted with boxes of the same color. For readability, lineages of reference genomes were collapsed in dark gray branches for orders and in light gray branches for families. CPR MAGs recovered from MR sediments are shown in purple, those from NP in green, and those from CR in orange. In addition, MAGs from the Mariana Trench are highlighted in red. Ultrafast bootstrap values are shown with circles of different colors according to the figure key. The scale bars indicate estimated substitutions per residue.

a 100% match with OTU_152, MR_Bin1762 matched OTU_556, and MR_Bin147 corresponded to OTU_525. It was also clear that the recovered CPR genomes represent only a subset of the CPR diversity revealed by the 16S rRNA gene amplicon sequencing. Despite that most CPR MAGs from NP and CR contained a 16S rRNA gene sequence, none of them was a good match with the amplicon-recovered OTUs from these sediments. The misrepresentation of these novel CPR MAGs in the 16S rRNA gene amplicon sequencing data may be due to the divergence of their 16S rRNA genes because they have one or more mismatches with the forward PCR primer used than those recovered from the MR sediments (see Fig. S3 in the supplemental material). Another possibility is that these CPR MAGs were rare taxa such that stochasticity and local heterogeneity might have led to their evasion in the PCR amplification. "*Ca.* ABY1" and "*Ca.* Microgenomatia," two important CPR classes that prevail in the three investigated sediment cores, especially in the CR core (Fig. 1I), were absent in our genome inventory but had been recovered from Mariana Trench hadal sediments (40). This indicated that the CPR MAGs detected by bulk metagenome sequencing are indeed part of the microbiome of the vast marine sedimentary environment. The mismatch of 16S rRNA gene sequences between the MAGs and the amplicon sequencing also suggested that some yet unknown CPR bacteria remained to be revealed in global marine sediments

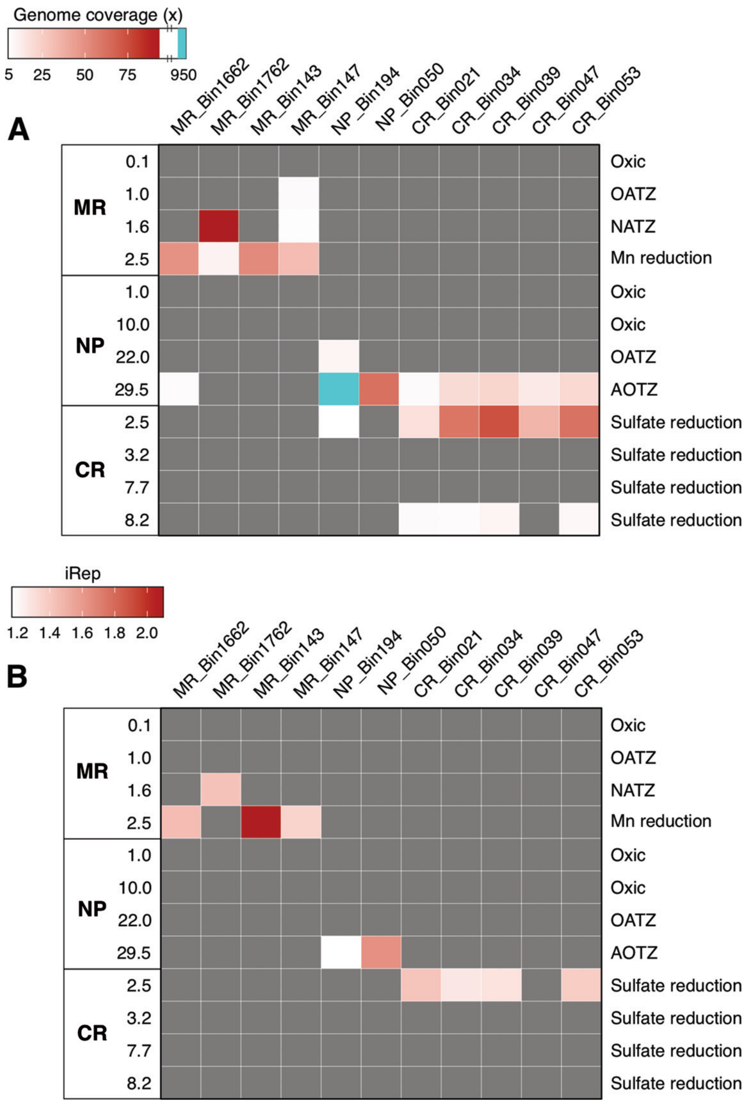

FIG 3 Confined distribution of CPR MAGs in marine sediment cores. (A) Genome coverage is shown as a proxy of the relative abundance of genomes in complex communities in different sediment layers of the three cores. The numbers on the *y* axis denote sediment depth in the unit of meters below seafloor (mbsf). (B) Index of replication (iRep) of genomes in complex communities, in different sediment layers of the three examined cores. In both panels A and B, gray squares indicate the absence of the examined items (either <2 genome coverages [A] or uncalculatable iRep [B]).

(for an example, see reference 57) and highlights the necessity of genome recovery for the discovery of novel CPR bacteria.

Varied preferred niches of marine sediment CPR MAGs. The presence of the 11 CPR MAGs was confined to anoxic sediment layers at their source locations. Based on the genome coverage calculation, CPR bacteria represented by these MAGs prefer anoxic sediment layers. The four MAGs from MR were present mainly in anoxic sediments in the NATZ or the Mn reduction zone below the oxygen depletion depth (Fig. 3A), the two CPR MAGs from NP mainly existed in the AOTZ (29.50 mbsf), and the five MAGs from CR were detected in the uppermost sequenced sediment horizon (2.0 mbsf) and to a lesser extent in the deepest horizon (~9.0 mbsf) (Fig. 3A). Read mapping across the three sediment sites suggested that the individual CPR MAGs were generally specific to the sediment core of origin (Fig. 3A), except that the five CPR MAGs recovered from CR were also detectable in the AOTZ of core NP-U1383E (Fig. 3A). Similar site-specific diversity of CPR genomes was also recently reported in the groundwater and freshwater lake environment by large-scale metagenomic surveys (23, 58).

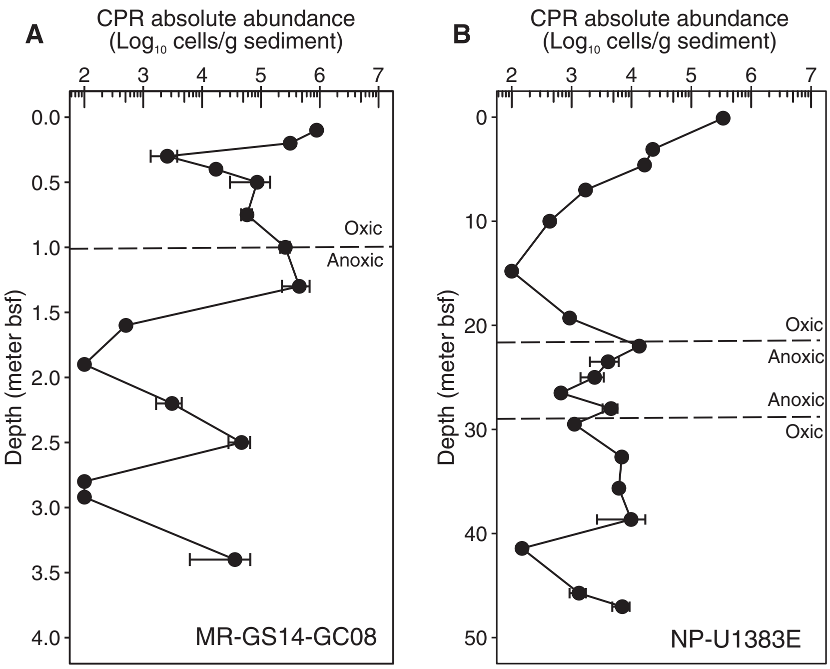

FIG 4 Estimated abundances of CPR in the MR (A) and NP (B) cores. The transitions between oxic and anoxic zones are marked with dashed lines. The error bars are derived from the standard deviation of the triplicate quantitative PCR (qPCR) quantification of the total cell numbers. For some data points, the error bar is not visible because it is smaller than the symbol.

As with where they were detected, the growth of the 11 CPR MAGs, inferred from the index of replication (iRep; values of >1 indicate proliferation) (59), occurred only in their primary niches (i.e., the sediment layer where the highest genome coverage was detected for a certain genome). The highest iRep of the four individual MAGs from MR occurred in either the NATZ or the Mn reduction zone, the two from NP in the AOTZ, and the five from CR in the uppermost anoxic zone (Fig. 3B). The calculated iRep (in the range of 1.2 to 2.2) suggested that these CPR MAGs were actively replicating in their primary niches at the time of sampling.

CPR are present in a wide range of redox niches but may not be directly selected by prevailing electron acceptors. Through amplicon sequencing and genome binning, we detected CPR in a variety of redox niches in marine sediments, with the dominant electron acceptors shifting from oxygen to nitrate, Mn oxides, and sulfate. Although none of the CPR MAGs was recovered from oxic sediment metagenomes, CPR 16S rRNA genes were detected in the investigated oxic sediments, especially in cores MR-GS14-GC08 and NP-1383E, which contain extensive oxic zones (Fig. 1). We estimated the absolute abundances of CPR in these two cores as the product of the total cell numbers (i.e., the sum of archaeal and bacterial 16S rRNA gene abundances) (46) and the relative abundances of CPR shown in Fig. 1. In cores of both MR and NP, the estimated absolute abundance of CPR can reach over 10^5 cells g^{-1} in the oxic zones (Fig. 4). Among the CPR genomes detected in MR sediments, MR_Bin147 was present in the oxic zone of GC08 (Fig. 3), despite that this genome lacks genes encoding the cytochrome *o* ubiquinol oxidase (complex IV) involved in oxygen reduction (see results below). Most known CPR organisms were previously detected in oxygen-limited or anoxic environments (33, 60), whereas some CPR organisms were also occasionally reported in oxic groundwater (24, 60–62), freshwater lakes (58), and soils (63). Our survey suggests that the overall CPR population was present in marine sediments throughout the early diagenesis processes even in bulk oxygenated sediments, but individual genomes showed preferences for different anoxic layers in different cores. Whether they live in the presence of oxygen or inhabit microenvironments in sediment particles without oxygen remains unknown. Considering that (i) the recovered CPR genomes seem not to

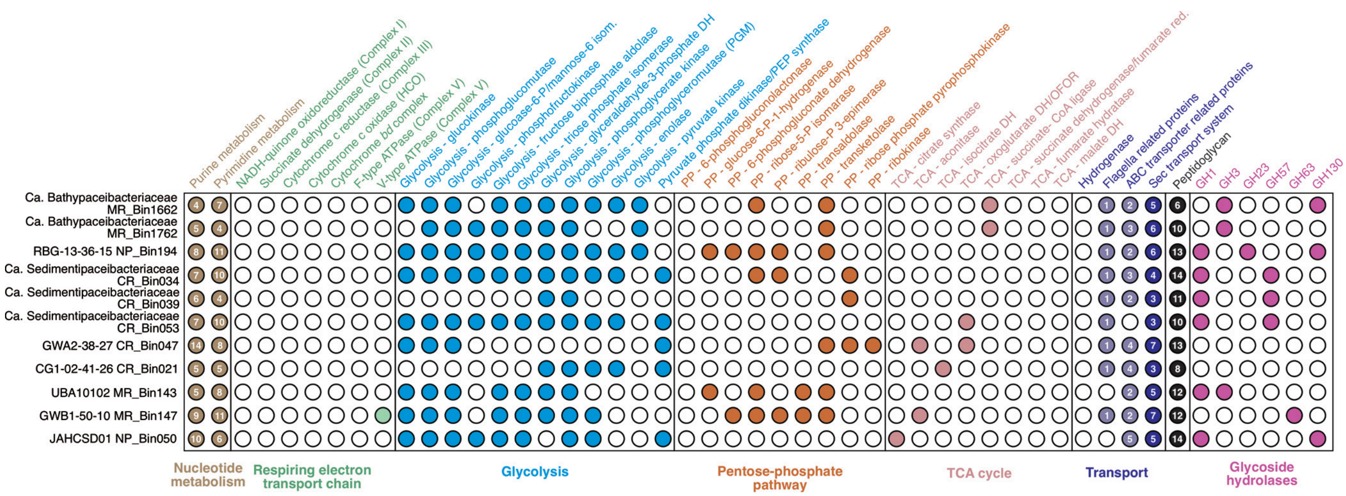

FIG 5 Metabolic potential of CPR genomes recovered from marine sediments. Specific proteins/pathways are shown on the top, while the metabolic pathways are indicated at the bottom. The presence of specific proteins/pathways is indicated by filled circles, while their complete absence is indicated by open circles. For some pathways, the numbers of the encoding genes of the key enzymes are also indicated.

have metabolic pathways sensitive to the external redox condition changes (the switching of the dominant terminal electron acceptors) and (ii) they have their narrow niches (indicated by the restricted distributions of their presence and active replication [i.e., iRep values]), the distribution of CPR bacteria in marine sediments may not be directly selected by the prevailing electron acceptors. The lack of direct environmental dependences of CPR has also been observed in a recent large-scale genome survey in the groundwater environment (24).

Limited energy metabolisms and biosynthesis capacities in marine sediment CPR. The CPR organisms recovered from marine sediments are likely fermentative organotrophs with simplified energy metabolism pathways. All recovered MAGs lack hydrogenases, especially the cytoplasmic bidirectional group 3 [NiFe] hydrogenase that has been proposed to be involved in pumping protons to build the proton motive force and help ATP generation in CPR genomes (33). All marine sediment CPR MAGs, except MR_Bin147, lack ATPases (Fig. 5), indicating that they are not capable of conserving energy by proton motive force. Instead, they may synthesize ATP through partial glycolysis via fermentation and substrate-level phosphorylation (33), similar to the recently characterized CPR bacterium *Vampirococcus lugosii* (64). This was confirmed by the presence of the glycolysis pathway-related genes in all sediment CPR MAGs recovered in this study, which facilitates the degradation of glucose to produce pyruvate or further to acetate, although some genes that regulate few intermediate steps were missing likely due to the incomplete nature of the CPR genomes (Fig. 5). The important intermediate compound of glycolysis, fructose 6-phosphate, can be provided by the pentose phosphate pathway, which is encoded in most of the recovered CPR MAGs (Fig. 5). We also note the lack of a complete tricarboxylic acid cycle (TCA), despite some CPR MAGs having a small subset of the enzymes in this cycle, likely for biosynthetic purposes. Like their close relatives from other environments (33), the marine sediment CPR MAGs lack a respiratory electron transport chain, as evidenced by the absence of NADH dehydrogenase (complex I) and complexes II to IV of the oxidative phosphorylation pathway (Fig. 5), suggesting that they are nonrespiring.

Although various glycoside hydrolases (GHs) have been detected in some CPR MAGs in groundwater (22, 65) and their expression in some CPR has been detected in subseafloor sediments (66), the GHs detected in our CPR MAGs are very limited. Six MAGs have a GH1 (for hydrolyzing carbohydrate moiety), while GH3, GH57 (starch), GH63, and GH130 (mannose) were also detected in fewer than three CPR MAGs (Fig. 5), suggesting that marine CPR MAGs have very limited saccharolytic capacities. However, the majority of the detected carbohydrate-active enzymes (CAZymes) belong to families 2 and 4, which associated with glycolipid synthesis, similar to CPR genomes detected in groundwater (21). Marine sediment CPR MAGs also lack the genes encoding nitrate reductase that were previously reported in genomes recovered from hadal sediments (43). Although the copper-containing nitrite reductase (NirK) has been previously noticed in some CPR genomes (22, 33), none of the marine sediment CPR genomes contain this gene. Therefore, whether CPR MAGs have an impact on the nitrogen cycle in marine sediments remains unclear.

Most of the sediment CPR genomes lack complete biosynthesis pathways for amino acids, lipids, cofactors, and nucleotides. Among the 20 known amino acids, only the genes for the biosynthesis of lysine and arginine are present (Fig. 5), indicating that they may need to obtain the rest of the amino acids from the external environment or their presumed host cells. Similarly, for nucleotide synthesis, they have only 4 to 14 genes for the metabolisms of purine and pyrimidine (Fig. 5). The CPR genomes have no recognized genes responsible for the synthesis of lipids or cofactors. Also lacking in these sediment CPR genomes are genes for flagellar biosynthesis (only 1 of the 46 required genes is encoded) (Fig. 4), indicating that they are probably nonmotile in the sediment environment. These CPR genomes also lack ABC type transporters and have very limited genes ($<$5 genes) for the Sec secretion system (Fig. 5). However, like other CPR genomes (33), all sediment CPR MAGs have extensive genes for the biosynthesis of peptidoglycan, suggesting that they may have intact cell walls.

Given the multiple auxotrophies detected in the recovered genomes and small genome sizes, we anticipate an episymbiotic lifestyle for the sediment CPR MAGs similar to that of their nonmarine relatives (34, 64, 67, 68). The lack of ABC transporters in the CPR MAGs may force them to rely on the host cells to obtain the necessary substrates for their metabolism. In MR_Bin147, the genome carries genes for hemolysin synthesis and hemolysin transporter, which could export hemolysin to the surface of the host cells and contribute to the host cell wall and membrane disruption and cell content release (Fig. 6). Similar to CPR genomes in other environments (27, 34, 69, 70), marine sediment CPR genomes harbor genes for type IV pilus synthesis (Fig. 6), which may provide access for the membrane-bound translocation complex to environmental double-stranded DNA (71). In addition, it also has a competence-related integral membrane protein, ComEC, which plays a role in the uptake of host DNA (72) that can be degraded to various restriction endonucleases to provide the nucleotides necessary for growth (Fig. 6).

Potential hosts of CPR in the MR sediments. We attempted to explore the potential prokaryotic hosts of CPR in sediments of the MR core, where CPR organisms were the most abundant among the three sites, by performing a cooccurrence network analysis based on the existing 16S rRNA gene amplicon sequencing data. Our results indicated that MR_Bin147 (represented by OTU_525) (Fig. S4) has a tight association with OTU_59 only, a member of the order *Gemmatimonadales* in the phylum *Gemmatimonadota*. Therefore, this analysis suggested that *Gemmatimonadales* bacteria could be hosts of CPR in MR sediments. It is worth noting that the (relative) abundances of OTU_59 (1.20% to 2.50% of the total communities) were almost 4-fold higher than those of of OTU_525 (01.4% to 0.74%), suggesting that only a fraction of *Gemmatimonadota* cells may harbor episymbionts in the MR sediments. This is consistent with the observation that "*Candidatus* Nanosynbacter lyticus TM7x" (*Saccharibacterium*) cells establish a long-term parasitic association with host cells (*Actinomyces odontolyticus* XH001) by infecting only a subset of the population (73). Future cultivation efforts are necessary to confirm such prediction and to also explore the hosts of the other CPR bacteria identified in these marine sediments.

These CPR genomes are also present in other sediments. Although the 11 CPR MAGs are phylogenetically novel and were confined within their ideal redox niches, they are not geographically exclusive to the three sediment cores investigated in this study. Recently, Zhou et al. (40) reported $>$500 microbial MAGs from the Challenger Deep sediments of the Mariana Trench, which included 24 "*Ca.* Patescibacteria" MAGs. Based on the GTDB-tk classification, "*Ca.* Paceibacteria" was the most abundant class (n = 14), while members of the classes "*Ca.* ABY1" (n = 4), "*Ca.* Andersenbacteria"

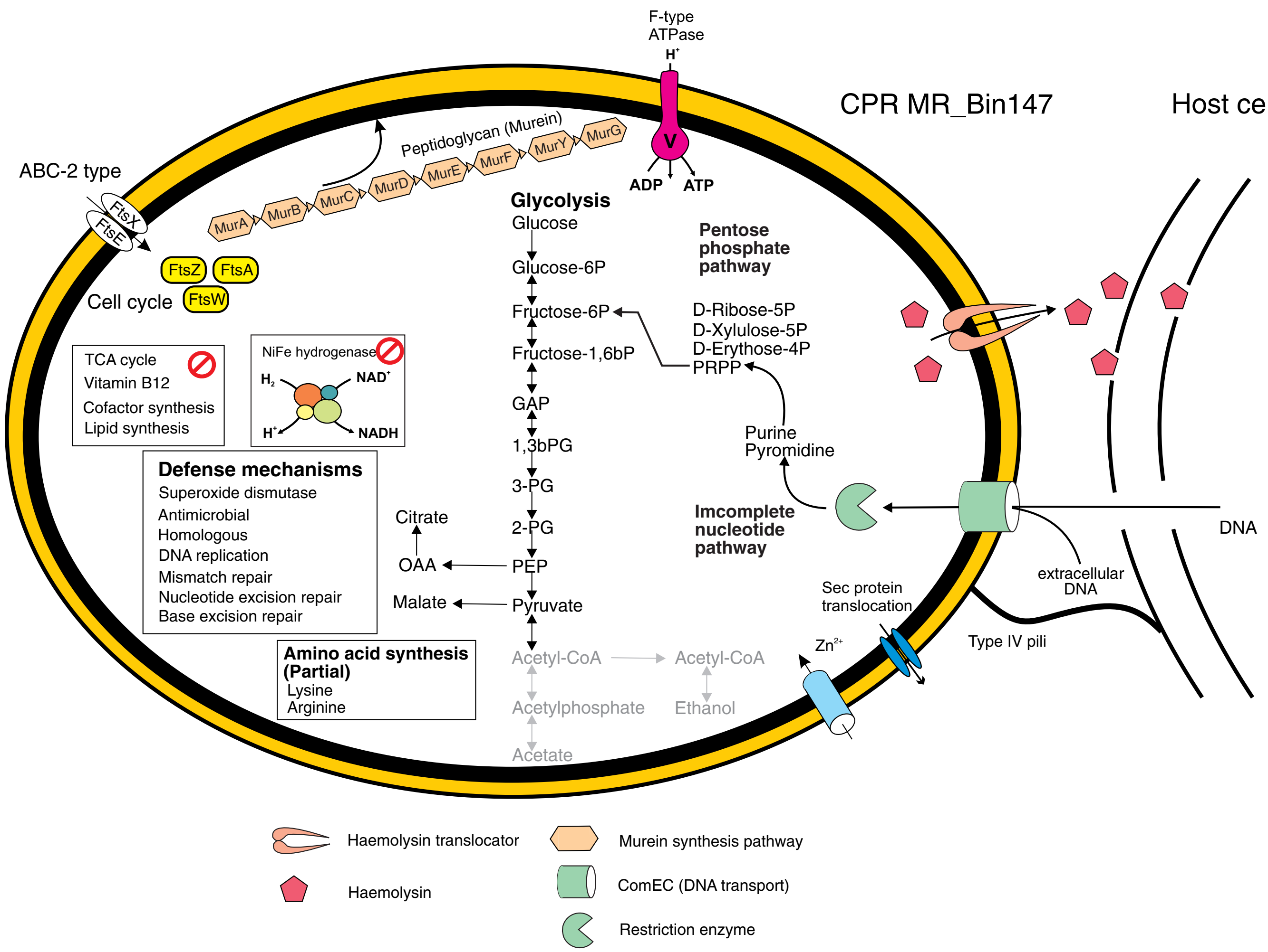

FIG 6 Metabolic capacities of CPR bacterium MR_Bin147. Pathways not detected in the CPR bacterial genome are shown in gray or highlighted in boxes. Glucose-6-P, glucose-6-phosphate; fructose-6-P, fructose-6-phosphate; GAP, glyceraldehyde 3-phosphate; 2-PG, 2-phosphoglyceric acid; PEP, phosphoenolpyruvate.

(n = 1), and "*Ca.* Microgenomatia" (n = 5) were also detected. By adding the "*Ca.* Paceibacteria" MAGs into our phylogenetic trees (Fig. 2A and B), we found that eight of these "*Ca.* Paceibacteria" MAGs were affiliated with the novel branches together with our MAGs. In particular, six fell into the family UBA10102 (containing MR_Bin143), one into RBG-13-36-15 (containing NP_Bin194), and the final into the order UBA6257 (containing MR_Bin174) (Fig. 2A). This comparison indicated that the CPR MAGs of these branches may have wider distribution ranges.

Conclusion and outlook. We report the abundance, diversity, and metabolic capacities of CPR bacteria in three sediment sites with contrasting environmental contexts, in which the examined sediments span most of the redox zones developed during the early diagenesis of organic matter. In this vast environment, CPR bacteria are generally rare (accounting for $<$5% in all but one sample) in the total microbial communities and are mainly in anoxic sediments, although $\sim 10^5$ cells g^{-1} of CPR cells are also present in the oxic zone. 16S rRNA gene amplicon sequencing tends to severely underestimate the diversity and abundance of CPR bacteria in the examined marine sediments, probably due to primer biases. The recovered CPR MAGs are novel at the family or genus level in the "*Ca.* Paceibacteria" class. All recovered CPR MAGs should be fermentative organotrophs relying on the truncated glycolysis for energy generation but lack complete biosynthesis pathways for amino acids, lipids, cofactors, and nucleotides and therefore are either symbionts or closely dependent on other community members for these key building blocks. The novel CPR genomes recovered in this study lay a foundation for further phylogenomic analysis of CPR in marine sediments. Despite some recent progress from untargeted metagenome sequencing surveys (for examples, see references 39 and 40), the total number of CPR MAGs from marine sediments is still much smaller than that from other natural habitats such as the groundwater (10, 23, 24), freshwater lakes (58), and soil (63). More targeted approaches (with some preselection before sequencing efforts) should be employed in future investigations to fully capture the overall diversity of CPR and other ultrasmall-celled microbes in global marine sediments. Cultivating diverse prokaryotes from native sediments and using cultured hosts to cue the presumed episymbionts in the laboratory (for examples, see references 27 and 44) could be among the promising and scalable ways to coculture CPR and to shed more light on these microbes in the vast habitat of marine sediments. Our study indicates that in global marine sediments, there are still some microbes that are yet uncaptured by single-gene-based surveys.

MATERIALS AND METHODS

Characteristics of study sites. We analyzed MAGs affiliated with Candidate Phyla Radiation (CPR) in marine sediments recovered from three different locations, the Arctic Mid-Ocean Ridge (MR-GS14-GC08) and North Pond (NP-U1383E) and Pacific coastal sediments (core CR-U1379B) off the Coast Rica Margin (47). The details about sample collection, DNA extraction, 16S rRNA gene amplicon sequencing and analysis, metagenome sequencing, assembly, binning, and refinement were described in previous reports (46, 47). We highlighted the contrasting redox conditions between the three sediment sites based on previously reported geochemical profiles (45–47). For those measurements, sediment porewater samples at discrete depths were extracted using either Rhizon samplers (for Arctic and North Pond cores [74]) or a titanium squeezer (for the Costa Rica core [75]) without air contacts to minimize alterations caused by sampling, handling, and measurement processes. Briefly, oxygen concentrations in sediment cores from the Arctic and North Pond cores were measured directly when the cores were still in the core liner by using a needle-type fiber optic oxygen microsensor (PreSens, Regensburg, Germany). Other porewater constituents, including nitrate, nitrite, ammonium, and dissolved inorganic carbon, were measured by a QuAAtro continuous flow analyzer (SEAL Analytical Ltd., Southampton, UK), using colorimetrical methods specific for each solute (46). Sulfate concentrations in the porewater of CR-U1379B were measured using a Dionex ICS-3000 ion chromatograph.

Analysis of CPR distribution in MR and NP sediments. We leveraged the previously generated shotgun metagenome sequencing data and 16S rRNA gene amplicon sequencing data of these three sites to examine the diversity and distribution of CPR bacteria. For the metagenome sequencing data, 16S rRNA gene reads in the quality-controlled sequences were identified and taxonomically classified using phyloFlash v3.2 beta1 (49), which uses BBMap (76) to identify and then classify putative 16S rRNA gene reads using the SILVA 138 release (77) as the reference database. The fractions of CPR classes in the total prokaryotic communities in the sediment samples were calculated and are shown as pie charts.

All the amplicon sequencing data sets were generated using "universal primers" (Uni519F/806R) (45) that target the variable region 4 (V4) of archaeal and bacterial 16S rRNA genes. Among the nine variable regions, the V4 region contains the second lowest frequency of insertions after the V8 region (10) and therefore was selected as the targeting region in our primer-dependent approach to capture CPR in complex communities, although primer-free methods (for an example, see reference 78) in theory should be better alternatives to characterize novel microbial populations. OTUs (220 bp, clustered at the cutoff of 97% nucleotide similarity) of putative CPR MAGs classified by CREST (79) against the SILVA 138.1 release were extracted from the OTU table. OTUs with $<$10 total reads across all sediment samples were excluded. Sediment samples with $<$5 CPR reads (i.e., 0.025% of the total community in GC08 and 0.05% of the total community in NP-U1383E) were excluded and are indicated as "not detected." The relative abundance of each CPR OTU in each sample was calculated using normalization scaling. OTUs were aggregated at the CPR class level and visualized using bar charts made using the R package ggplot2 (80).

CPR MAGs with identifiable 16S rRNA gene fragments were matched with the short amplicon OTUs by blast alignment using the MAG 16S rRNA gene as the query and the OTU sequences from the corresponding sediment cores as the reference database. For example, if the 16S rRNA gene of MAG_A showed a $>$99% match with OTU_001 in core NP-U1383E, they were considered matches and the distribution of MAG_A in NP-U1383E was considered to be that of OTU_001. This method is helpful to describe the distribution of MAGs in all the sampled sediment horizons rather than just those of the metagenome sequenced.

Metagenome binning and genome refinement. The eight metagenome sequencing data sets from MR (n = 4) and NP (n = 4) were coassembled, while the four from CR were also coassembled using MEGAHIT v1.2.9 (81), using k-mers from 21 to 141 (–k-min 21 –k-max 141 –k-step 10 –presets meta-large). Contigs longer than 1,000 bp were binned using MaxBin2 v2.2.5 (82) and MetaBAT v2.15.3 (83), and the highest-quality ones were selected using DAS_Tool (84) with the default parameters. The resulting MAGs were quality assessed using CheckM v1.1.3 (50) and taxonomically classified using GTDB-Tk v2.0.0 (parameters: classify_wf defaults) (85). As a proxy of relative abundance, the coverage of each genome in each sample was determined by read mapping using BBMap v38.76 (76), with a 99% identity threshold (minid = 99, idfilter = 99). To refine and improve the quality of each MAG, reads in the highest-coverage samples were recruited onto the scaffolds of the MAG and reassembled using SPAdes v.3.12.0 (86), with the k-mers of 21, 33, 55, and 77. The resulting assembly was refined using gbtools (87), mainly by plotting coverages of contigs in the samples in which the two highest coverages were determined and outlier contigs were removed. We used CheckM v1.1.3 (50) to assess the quality of the CPR MAGs based on two sets of genes: (i) the single-copy genes that are automatically determined by CheckM for bacteria and (ii) the 43 single-copy genes proposed by Brown et al. (10) that were implemented in CheckM as a specific custom-defined workflow (for detailed steps, see https://github.com/Ecogenomics/CheckM/wiki/Workflows#using-cpr-marker-set).

Genome annotation. Genes in the CPR genomes were predicted using Prodigal v2.6.3 (88). Genome annotation was conducted using Prokka v.1.13 (89), eggNOG v2.1.4 (90), and BlastKoala (91) using the KEGG database. The functional assignments of genes of interest were also confirmed using BLASTp (E value $<$ 1e$-$10, sequence identity $>$ 50%, and shortest alignment rate $>$ 60%) against the NCBI RefSeq database. The metabolic pathways were reconstructed using KEGG Mapper (92). Carbohydrate-active enzymes (CAZYmes) were searched using dbCAN hmm profiles (downloaded April 2019 from http://bcb.unl.edu/dbCAN2/download/Databases/dbCAN-old@UGA/), using an E value of 1e$-$18 and a coverage of $>$0.35, per the instructions of the developers.

Phylogenetic analysis. Phylogenomic analysis was performed separately for the 11 MAGs of CPR recovered. In addition, the 14 MAGs of the "*Ca.* Paceibacteria" class from the Mariana Trench sediments reported previously (40) were also included. This phylogenomic analysis was based on marker genes consisting of 14 concatenated ribosomal protein genes (rpL2, -3, -4, -5, -6, -14, -16, -18, and -22 and rpS3, -8, -10, -17, and -19). These genes in the genomes were identified in Anvi'o v6.2 (93) by hidden Markov model (HMM) profiles. Sequences of four "*Candidatus* Doudnabacteria" genomes were used as the outgroup. Sequences of each marker gene were aligned individually using MUSCLE (94), and alignment gaps were removed using trimAl (95), with the mode "automated." Individually trimmed alignments were concatenated. The maximum-likelihood phylogenetic tree was reconstructed using IQ-TREE v1.6.10 (96) with LG+F+R6 as the best-fit evolutionary model selected by ModelFinder (97) and 1,000 ultrafast bootstraps by UFBoot2 (98).

Likewise, phylogenetic analyses of the 16S rRNA gene were also performed for the CPR MAGs. 16S rRNA gene sequences in the MAGs were annotated using barrnap (99) implemented in the Prokka package. In some cases where more than one copy of 16S rRNA gene fragments was found in a genome, these fragments were assembled by aligning them with other high-quality intron-free CPR 16S rRNA gene sequences in Unipro UGENE (100) and removing the 100% overlap region. These sequences of MAGs were used to identify their close relatives by a BLASTn search in the NCBI database with a similarity threshold of 90%. All sequences were aligned using MAFFT-LINSi (101). Given the known large insertions/introns in CPR 16S rRNA gene sequences (10), the alignment was manually inspected using Unipro UGENE (100) and gaps were manually removed. A maximum-likelihood phylogenetic tree was constructed using IQ-TREE v1.6.10 (96) with SYM+R6 as the best-fit evolution model selected by ModelFinder (97), and 100 ultrafast bootstraps were performed by UFBoot2 (98). Relevant amplicon OTU sequences (220 bp) were added to the phylogenetic tree without changing the overall topology to identify the clades of the OTU sequences.

iRep calculation. Index of replication (iRep) values of the recovered CPR MAGs in all the metagenomic sequenced depths were calculated in accordance with the procedure outlined in reference 59. Before the calculations, contigs of $<$5 kb in each genome bin were removed. The quality (completeness and contamination levels) of the bins assessed by CheckM (50) was not affected by this omission and fulfilled the prerequisite of genome quality of such calculation.

Cooccurring network analysis. Network analysis was performed for the amplicon sequencing data sets of the MR core GC08. Given that the CPR MAGs were recovered mainly from the anoxic sediments of GC08 (Fig. 3A), the sediment horizons collected from the oxic zone were removed to minimize the effect of niche filtration on network analysis outcomes. We also retained only those OTUs representing $>$100 reads totally across all samples in this analysis to reduce the effect of spurious correlations caused by low-abundance taxon distribution. The relative abundance data were converted to ratio data using the centered log ratio (Clr) method. The network was inferred using SPIEC-EASI (102) using the "glasso" method and characterized using the igraph package (103). The nodes in the reconstructed network represent OTUs at 97% identity, and their sizes are proportional to their degrees (number of connections), while the edges (i.e., connections) correspond to strong and significant (positive or negative) correlations between nodes. Only positive correlations were plotted, whereas four weak negative correlations existed in the data set but were not shown.

Data availability. All sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive under project numbers PRJNA529480 (MR), PRJNA489438 (NP), and PRJNA599172 (CR). Raw metagenomic sequencing data are available in NCBI under BioSample numbers SAMN11268098 (GC08_10cm), SAMN11268104 (GC08_100cm), SAMN11268106 (GC08_160cm), and SAMN11268109 (GC08_250cm). The MAGs described in this study are available from NCBI under the accession numbers provided in Table 1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.8 MB.

ACKNOWLEDGMENTS

We thank the scientists and crew members of R/V *JOIDES Resolution* and R/V *G. O. Sars* for their efforts in collecting the sediment samples used in this study. We are also grateful to Sophie Abby, Christa Schleper, and Thomas Pollak for metagenome sequencing data generation for the MR and NP samples and to the University of Delaware DNA Sequencing & Genotyping Center staff for the CR samples.

Computational support from the University of Delaware Center for Bioinformatics and Computational Biology Core Facility and the use of the BIOMIX compute cluster were made possible through funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute. This work was funded by the K. G. Jebsen Foundation and the Trond Mohns Science Foundation (to S.L.J.). R.Z. and J.F.B. were funded in part by the W. M. Keck Foundation.

R.Z. and S.L.J. collected the samples and metagenome sequencing data of the MR and NP sediments. J.F.B. and R.Z. were involved in the metagenome sequencing data collection of the Costa Rica Margin sediments. R.Z. binned and refined the MAGs. R.Z. and I.F.F. analyzed the genomic data. R.Z. drafted the manuscript, and all authors contributed substantially to the editing of this paper.

We declare that we have no conflicts of interest.

REFERENCES

1. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. Proc Natl Acad Sci U S A 109:16213–16216. https://doi.org/10.1073/pnas.1203849109.
2. D'Hondt S, Pockalny R, Fulfer VM, Spivack AJ. 2019. Subseafloor life and its biogeochemical impacts. Nat Commun 10:3519. https://doi.org/10.1038/s41467-019-11450-z.
3. Berner RA, Canfield DE. 1989. A new model for atmospheric oxygen over Phanerozoic time. Am J Sci 289:333–361. https://doi.org/10.2475/ajs.289.4.333.
4. Ridgwell A, Hargreaves JC. 2007. Regulation of atmospheric CO2 by deep-sea sediments in an Earth system model. Global Biogeochem Cycles 21:GB2008. https://doi.org/10.1029/2006GB002764.
5. Teske A, Sørensen KB. 2008. Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J 2:3–18. https://doi.org/10.1038/ismej.2007.90.
6. Inagaki F, Nunoura T, Nakagawa S, Teske A, Lever M, Lauer A, Suzuki M, Takai K, Delwiche M, Colwell FS, Nealson KH, Horikoshi K, D'Hondt S, Jørgensen BB. 2006. Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments, on the Pacific Ocean Margin. Proc Natl Acad Sci U S A 103:2815–2820. https://doi.org/10.1073/pnas.0511033103.
7. Hoshino T, Doi H, Uramoto G-I, Wörmer L, Adhikari RR, Xiao N, Morono Y, D'Hondt S, Hinrichs K-U, Inagaki F. 2020. Global diversity of microbial communities in marine sediment. Proc Natl Acad Sci U S A 117:27587–27597. https://doi.org/10.1073/pnas.1919139117.
8. Parkes RJ, Cragg B, Roussel E, Webster G, Weightman A, Sass H. 2014. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. Marine Geol 352:409–425. https://doi.org/10.1016/j.margeo.2014.02.009.
9. Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. 2016. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol 1:15032. https://doi.org/10.1038/nmicrobiol.2015.32.
10. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523:208–211. https://doi.org/10.1038/nature14486.
11. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352.
12. 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.
13. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. mSystems 3:e00055-18. https://doi.org/10.1128/mSystems.00055-18.
14. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. 1998. Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180:366–376. https://doi.org/10.1128/JB.180.2.366-376.1998.
15. Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172:1181–1197. https://doi.org/10.1016/j.cell.2018.02.016.
16. 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.
17. Schulz F, Eloe-Fadrosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, Kyrpides NC, Woyke T. 2017. Towards a balanced view of the bacterial tree of life. Microbiome 5:140. https://doi.org/10.1186/s40168-017-0360-9.
18. Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz P, Szöllősi GJ, Williams TA. 2021. A rooted phylogeny resolves early bacterial evolution. Science 372:eabe0511. https://doi.org/10.1126/science.abe0511.
19. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7:13219. https://doi.org/10.1038/ncomms13219.
20. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF. 2018. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol 3:328–336. https://doi.org/10.1038/s41564-017-0098-y.
21. Geesink P, Wegner CE, Probst AJ, Herrmann M, Dam HT, Kaster AK, Küsel K. 2020. Genome-inferred spatio-temporal resolution of an uncultivated Roizmanbacterium reveals its ecological preferences in groundwater. Environ Microbiol 22:726–737. https://doi.org/10.1111/1462-2920.14865.
22. Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ. 2017. Members of the candidate phyla radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. Microbiome 5:112. https://doi.org/10.1186/s40168-017-0331-1.
23. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, Banfield JF. 2021. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat Microbiol 6:354–365. https://doi.org/10.1038/s41564-020-00840-5.
24. Chaudhari NM, Overholt WA, Figueroa-Gonzalez PA, Taubert M, Bornemann TLV, Probst AJ, Hölzer M, Marz M, Küsel K. 2021. The economical lifestyle of

CPR bacteria in groundwater allows little preference for environmental drivers. Environ Microbiome 16:24. https://doi.org/10.1186/s40793-021-00395-w.
25. 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. https://doi.org/10.1186/s40168-018-0548-7.
26. Vigneron A, Cruaud P, Langlois V, Lovejoy C, Culley AI, Vincent WF. 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. https://doi.org/10.1002/lol2.10132.
27. Batinovic S, Rose JJA, Ratcliffe J, Seviour RJ, Petrovski S. 2021. Cocultivation of an ultrasmall environmental parasitic bacterium with lytic ability against bacteria associated with wastewater foams. Nat Microbiol 6: 703–711. https://doi.org/10.1038/s41564-021-00892-1.
28. Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH, Kondrotaite Z, Karst SM, Dueholm MS, Nielsen PH, Albertsen M. 2021. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nat Commun 12:2009. https://doi.org/10.1038/s41467-021-22203-2.
29. Merino N, Kawai M, Boyd ES, Colman DR, McGlynn SE, Nealson KH, Kurokawa K, Hongoh Y. 2020. Single-cell genomics of novel Actinobacteria with the Wood-Ljungdahl pathway discovered in a serpentinizing system. Front Microbiol 11:1031. https://doi.org/10.3389/fmicb.2020.01031.
30. Chen L-X, Al-Shayeb B, Méheust R, Li W-J, Doudna JA, Banfield JF. 2019. Candidate Phyla Radiation Roizmanbacteria from hot springs have novel and unexpectedly abundant CRISPR-Cas systems. Front Microbiol 10: 928. https://doi.org/10.3389/fmicb.2019.00928.
31. Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A, Morrill PL, Inagaki F, Kuenen JG, Nealson KH. 2017. Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars. ISME J 11:2584–2598. https://doi.org/10.1038/ismej.2017.111.
32. Beam JP, Becraft ED, Brown JM, Schulz F, Jarett JK, Bezuidt O, Poulton NJ, Clark K, Dunfield PF, Ravin NV, Spear JR, Hedlund BP, Kormas KA, Sievert SM, Elshahed MS, Barton HA, Stott MB, Eisen JA, Moser DP, Onstott TC, Woyke T, Stepanauskas R. 2020. Ancestral absence of electron transport chains in Patescibacteria and DPANN. Front Microbiol 11:1848. https://doi.org/10.3389/fmicb.2020.01848.
33. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol 16:629–645. https://doi.org/10.1038/s41579-018-0076-2.
34. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, Hallsworth JE, Klyukina AA, Tikhonova EN, Gorlenko VM. 2022. Cultivation of a vampire: 'Candidatus Absconditicoccus praedator'. Environ Microbiol 24:30–49. https://doi.org/10.1111/1462-2920.15823.
35. LaRowe DE, Amend JP. 2015. Catabolic rates, population sizes and doubling/replacement times of microorganisms in natural settings. Am J Sci 315:167–203. https://doi.org/10.2475/03.2015.01.
36. Zhao R, Mogollón JM, Roerdink DL, Thorseth IH, Økland I, Jørgensen SL. 2021. Ammonia-oxidizing archaea have similar power requirements in diverse marine oxic sediments. ISME J 15:3657–3667. https://doi.org/10.1038/s41396-021-01041-6.
37. Jørgensen BB, Marshall IPG. 2016. Slow microbial life in the seabed. Annu Rev Mar Sci 8:311–332. https://doi.org/10.1146/annurev-marine-010814-015535.
38. Engelhardt T, Orsi WD, Jørgensen BB. 2015. Viral activities and life cycles in deep subseafloor sediments. Environ Microbiol Rep 7:868–873. https://doi.org/10.1111/1758-2229.12316.
39. Dong X, Zhang C, Li W, Weng S, Song W, Li J, Wang Y. 2021. Functional diversity of microbial communities in inactive seafloor sulfide deposits. FEMS Microbiol Ecol 97:fiab108. https://doi.org/10.1093/femsec/fiab108.
40. Zhou Y-L, Mara P, Cui G-J, Edgcomb VP, Wang Y. 2022. Microbiomes in the Challenger Deep slope and bottom-axis sediments. Nat Commun 13: 1515. https://doi.org/10.1038/s41467-022-29144-4.
41. Hiraoka S, Hirai M, Matsui Y, Makabe A, Minegishi H, Tsuda M, Juliarni, Rastelli E, Danovaro R, Corinaldesi C, Kitahashi T, Tasumi E, Nishizawa M, Takai K, Nomaki H, Nunoura T. 2020. Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. ISME J 14:740–756. https://doi.org/10.1038/s41396-019-0564-z.
42. Schauberger C, Glud RN, Hausmann B, Trouche B, Maignien L, Poulain J, Wincker P, Arnaud-Haond S, Wenzhöfer F, Thamdrup B. 2021. Microbial community structure in hadal sediments: high similarity along trench axes and strong changes along redox gradients. ISME J 15:3455–3467. https://doi.org/10.1038/s41396-021-01021-w.
43. Leon-Zayas R, Peoples L, Biddle JF, Podell S, Novotny M, Cameron J, Lasken RS, Bartlett DH. 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. https://doi.org/10.1111/1462-2920.13789.
44. Kuroda K, Yamamoto K, Nakai R, Hirakata Y, Kubota K, Nobu MK, Narihiro T. 2022. Symbiosis between Candidatus Patescibacteria and archaea discovered in wastewater-treating bioreactors. mBio 0:e01711-22. https://doi.org/10.1128/mbio.01711-22.
45. Zhao R, Hannisdal B, Mogollon JM, Jørgensen SL. 2019. Nitrifier abundance and diversity peak at deep redox transition zones. Sci Rep 9:8633. https://doi.org/10.1038/s41598-019-44585-6.
46. Zhao R, Mogollón JM, Abby SS, Schleper C, Biddle JF, Roerdink DL, Thorseth IH, Jørgensen SL. 2020. Geochemical transition zone powering microbial growth in subsurface sediments. Proc Natl Acad Sci U S A 117: 32617–32626. https://doi.org/10.1073/pnas.2005917117.
47. Zhao R, Biddle JF. 2021. Helarchaeota and co-occurring sulfate-reducing bacteria in subseafloor sediments from the Costa Rica Margin. ISME Commun 1:25. https://doi.org/10.1038/s43705-021-00027-x.
48. Riedinger N, Torres ME, Screaton E, Solomon EA, Kutterolf S, Schindlbeck-Belo J, Formolo MJ, Lyons TW, Vannucchi P. 2019. Interplay of subduction tectonics, sedimentation, and carbon cycling. Geochem Geophys Geosys 20:4939–4955. https://doi.org/10.1029/2019GC008613.
49. Gruber-Vodicka HR, Seah BKB, Pruesse E. 2020. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. mSystems 5:e00920-20. https://doi.org/10.1128/mSystems.00920-20.
50. 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.
51. 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.
52. Anantharaman K, Brown CT, Burstein D, Castelle CJ, Probst AJ, Thomas BC, Williams KH, Banfield JF. 2016. Analysis of five complete genome sequences for members of the class Peribacteria in the recently recognized Peregrinibacteria bacterial phylum. PeerJ 4:e1607. https://doi.org/10.7717/peerj.1607.
53. Yang T, Speare K, McKay L, MacGregor BJ, Joye SB, Teske A. 2016. Distinct bacterial communities in surficial seafloor sediments following the 2010 Deepwater Horizon Blowout. Front Microbiol 7:1384. https://doi.org/10.3389/fmicb.2016.01384.
54. Inagaki F, Hinrichs KU, Kubo Y, Bowles MW, Heuer VB, Hong WL, Hoshino T, Ijiri A, Imachi H, Ito M, Kaneko M, Lever MA, Lin YS, Methe BA, Morita S, Morono Y, Tanikawa W, Bihan M, Bowden SA, Elvert M, Glombitza C, Gross D, Harrington GJ, Hori T, Li K, Limmer D, Liu CH, Murayama M, Ohkouchi N, Ono S, Park YS, Phillips SC, Prieto-Mollar X, Purkey M, Riedinger N, Sanada Y, Sauvage J, Snyder G, Susilawati R, Takano Y, Tasumi E, Terada T, Tomaru H, Trembath-Reichert E, Wang DT, Yamada Y. 2015. Exploring deep microbial life in coal-bearing sediment down to similar to 2. Science 349:420–424. https://doi.org/10.1126/science.aaa6882.
55. Schauer R, Røy H, Augustin N, Gennerich H-H, Peters M, Wenzhoefer F, Amann R, Meyerdierks A. 2011. Bacterial sulfur cycling shapes microbial communities in surface sediments of an ultramafic hydrothermal vent field. Environ Microbiol 13:2633–2648. https://doi.org/10.1111/j.1462-2920.2011.02530.x.
56. Shiraishi F, Mitsunobu S, Suzuki K, Hoshino T, Morono Y, Inagaki F. 2016. Dense microbial community on a ferromanganese nodule from the ultra-oligotrophic South Pacific Gyre: implications for biogeochemical cycles. Earth Planet Sci Lett 447:10–20. https://doi.org/10.1016/j.epsl.2016.04.021.
57. Gong X, Rio A, Xu L, Langwig M, Su L, Sun M, Huerta-Cepas J, De Anda V, Baker B. 2022. New globally distributed bacteria with high proportions of novel protein families involved in sulfur and nitrogen cycling. Res Square https://doi.org/10.21203/rs.3.rs-1620321/v1.
58. Chiriac M-C, Bulzu P-A, Andrei A-S, Okazaki Y, Nakano S-i, Haber M, Kavagutti VS, Layoun P, Ghai R, Salcher MM. 2022. Ecogenomics sheds

light on diverse lifestyle strategies in freshwater CPR. Microbiome 10:84. https://doi.org/10.1186/s40168-022-01274-3.

59. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016. Measurement of bacterial replication rates in microbial communities. Nat Biotechnol 34:1256–1263. https://doi.org/10.1038/nbt.3704.
60. Herrmann M, Wegner C-E, Taubert M, Geesink P, Lehmann K, Yan L, Lehmann R, Totsche KU, Küsel K. 2019. Predominance of Cand. Patescibacteria in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. Front Microbiol 10:1407. https://doi.org/10.3389/fmicb.2019.01407.
61. Nelson WC, Stegen JC. 2015. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol 6:713. https://doi.org/10.3389/fmicb.2015.00713.
62. Castelle CJ, Brown CT, Thomas BC, Williams KH, Banfield JF. 2017. Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. Sci Rep 7:40101. https://doi.org/10.1038/srep40101.
63. Nicolas AM, Jaffe AL, Nuccio EE, Taga ME, Firestone MK, Banfield JF. 2021. Soil candidate phyla radiation bacteria encode components of aerobic metabolism and co-occur with Nanoarchaea in the rare biosphere of rhizosphere grassland communities. mSystems 6:e01205-20. https://doi.org/10.1128/mSystems.01205-20.
64. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. 2021. Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nat Commun 12:2454. https://doi.org/10.1038/s41467-021-22762-4.
65. Solden L, Lloyd K, Wrighton K. 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. Curr Opin Microbiol 31:217–226. https://doi.org/10.1016/j.mib.2016.04.020.
66. Orsi WD, Richards TA, Francis WR. 2018. Predicted microbial secretomes and their target substrates in marine sediment. Nat Microbiol 3:32–37. https://doi.org/10.1038/s41564-017-0047-9.
67. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun 6:6372. https://doi.org/10.1038/ncomms7372.
68. He XS, McLean JS, Edlund A, Yooseph S, Hall AP, Liu SY, Dorrestein PC, Esquenazi E, Hunter RC, Cheng GH, Nelson KE, Lux R, Shi WY. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci U S A 112:244–249. https://doi.org/10.1073/pnas.1419038112.
69. Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, Yousef M, Watson AR, Lolans K, Esen ÖC, Lee STM, Downey N, Morrison HG, Dewhirst FE, Mark Welch JL, Eren AM. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biol 21:292. https://doi.org/10.1186/s13059-020-02195-w.
70. Xie B, Wang J, Nie Y, Chen D, Hu B, Wu X, Du W. 2021. EpicPCR-directed cultivation of a Candidatus Saccharibacteria symbiont reveals a type IV pili-dependent epibiotic lifestyle. bioRxiv. https://doi.org/10.1101/2021.07.08.451036.
71. Chen I, Dubnau D. 2004. DNA uptake during bacterial transformation. Nat Rev Microbiol 2:241–249. https://doi.org/10.1038/nrmicro844.
72. Hahn J, Inamine G, Kozlov Y, Dubnau D. 1993. Characterization of comE, a late competence operon of Bacillus subtilis required for the binding and uptake of transforming DNA. Mol Microbiol 10:99–111. https://doi.org/10.1111/j.1365-2958.1993.tb00907.x.
73. Bor B, McLean JS, Foster KR, Cen L, To TT, Serrato-Guillen A, Dewhirst FE, Shi W, He X. 2018. Rapid evolution of decreased host susceptibility drives a stable relationship between ultrasmall parasite TM7x and its bacterial host. Proc Natl Acad Sci U S A 115:12277–12282. https://doi.org/10.1073/pnas.1810625115.
74. Seeberg-Elverfeldt J, Schlüter M, Feseker T, Kölling M. 2005. Rhizon sampling of pore waters near the sediment/water interface of aquatic systems. Limnol Oceanogr Methods 3:361–371. https://doi.org/10.4319/lom.2005.3.361.
75. Torres ME, Muratli JM, Solomon EA. 2014. Data report: minor element concentrations in pore fluids from the CRISP-A transect drilled during Expedition 334, p 2. *In* Vannucchi P, Ujile K, Stroncik N, Malinverno A, Expedition 334 Scientists (ed), Proceedings of the Integrated Ocean Drilling Program, Tokyo, Japan.
76. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory, Berkeley, CA.
77. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.
78. Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M. 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat Biotechnol 36:190–195. https://doi.org/10.1038/nbt.4045.
79. Lanzen A, Jørgensen SL, Huson DH, Gorfer M, Grindhaug SH, Jonassen I, Ovreas L, Urich T. 2012. CREST—classification resources for environmental sequence tags. PLoS One 7:e49334. https://doi.org/10.1371/journal.pone.0049334.
80. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York, NY.
81. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
82. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
83. 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 7:e7359. https://doi.org/10.7717/peerj.7359.
84. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.
85. 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.
86. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
87. Seah BK, Gruber-Vodicka HR. 2015. gbtools: interactive visualization of metagenome bins in R. Front Microbiol 6:1451. https://doi.org/10.3389/fmicb.2015.01451.
88. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
89. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
90. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.
91. 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. https://doi.org/10.1016/j.jmb.2015.11.006.
92. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109–D114. https://doi.org/10.1093/nar/gkr988.
93. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.
94. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
95. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
96. 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.
97. 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.

98. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
99. Seemann T. 2015. Barrnap. https://github.com/tseemann/barrnap.
100. Okonechnikov K, Golosova O, Fursov M, Team U, UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167. https://doi.org/10.1093/bioinformatics/bts091.
101. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
102. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 11:e1004226. https://doi.org/10.1371/journal.pcbi.1004226.
103. Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal Complex Systems 1695.