Genomic characterisation of novel extremophile lineages from the thalassohaline lake Dziani Dzaha expands the metabolic repertoire of the PVC superphylum

Adrien Vigneron[1*], Lilian A. Cloarec[1], Céline Brochier-Armanet[2,6], Jean Pierre Flandrois[2], Marc Troussellier[3], Cécile Bernard[4], Hélène Agogué[5], Philippe M. Oger[1] and Mylène Hugoni[1,6*]

Abstract

Background Extreme environments are useful systems to investigate limits of life, microbial biogeography and ecology, and the adaptation and evolution of microbial lineages. Many novel microbial lineages have been discovered in extreme environments, especially from the Planctomycetota–Verrucomicrobiota–Chlamydiota (PVC) superphyla. However, their evolutionary history and roles in ecosystem functioning and microbiome assemblage are poorly understood.

Results Applying a genome-centric approach on an 8-year metagenomic timeseries produced from the hypersaline and hyperalkaline waters of Lake Dziani Dzaha (Mayotte), we recovered 5 novel PVC extremophilic candidate lineages from the biosphere of the lake. Sibling to Elusimicrobia and Omnitrophota, these lineages represented novel halophilic clades, with global distributions bounded to soda lakes and hypersaline hydrosystems. Genome mining of these newly defined clades revealed contrasted, but ecologically relevant, catabolic capabilities involved in the carbon, hydrogen and iron/electron cycles of the Dziani Dzaha ecosystem. This also includes extracellular electron transfer for two of them, suggesting metal reduction or potential electron exchanges with other members of the lake community. By contrast, a putative extracellular giant protein with multiple carbohydrate binding domains and toxin-like structures, as observed in virulence factors, was identified in the genome of another of these clades, suggesting predatory capabilities.

Conclusions Our results provided genomic evidences for original metabolism in novel extremophile lineages of the PVC superphyla, revealing unforeseen implications for members of this widespread and diverse bacterial radiation in aquatic saline ecosystems. Finally, monitoring the in-situ distribution of these lineages through the timeseries reveals the drastic effects of environmental perturbations on extreme ecosystem biodiversity.

Keywords Hypersaline, Extreme environments, Biodiversity, MAGs, Omnitrophota, Elusimicrobia

*Correspondence:
Adrien Vigneron
avignero@gmail.com
Mylène Hugoni
mylene.hugoni@univ-lyon1.fr
Full list of author information is available at the end of the article

Vigneron *et al. Environmental Microbiome*   (2025) 20:48

Page 2 of 11

Background

Extreme environments such as hydrothermal systems [1–3], polar lakes [4, 5], hypersaline basins [6, 7] and deep underground aquifers [8, 9] are hotspots of unexplored microbial biodiversity with potentially unusual metabolic and adaptive traits. The rapid development of DNA sequencing and computational algorithms has overcome the limitation of microorganism cultivation and now allows the reconstruction and analysis of genomes from uncultured microbial populations [10]. This approach has revealed many new, widespread and highly diversified lineages, collectively referred to as the "microbial dark matter", expanding our vision of the tree of life [10]. This provides insights into the limits of life, the ecological roles of uncultured microorganisms, and the adaptive mechanisms to extreme conditions [1, 7]. For instance, our knowledge of the taxonomic, genomic, and functional diversities of the enigmatic Patescibacteria (bacterial Candidate Phyla Radiation) or DPANN archaea, has been significantly expanded by genome-centric metagenomics, revealing their ubiquity, abundance and complex entanglement with other microbial lineages [11, 12]. In addition to these major superphyla, which have received much attention, a myriad of new lineages with limited distribution and/or representation have been identified in various biomes, including the most extreme environments [13]. These include Omnitrophota, Elusimicrobia, Auribacterota, Ratteibacteria, Desantisbacteria, Firestonebacteria and Goldbacteria from the superphylum Planctomycetota–Verrucomicrobiota–Chlamydiota (PVC) [9, 14–16]. Despite their relatively low abundances in ecosystems and a presumably small cell size [9, 14, 17], these poorly characterized microorganisms have been found to be hyperactive and to contribute significantly to ecosystem functioning [17], sulfur [18] and/or carbon cycles through complex organic matter degradation and fermentation [14, 19]. Furthermore, Omnitrophota and Elusimicrobia, the best documented lineages of this phylogenetic radiation, display atypical evolution and metabolic capacities [16, 17]. For instance, part of their representatives have been found to be host-associated microorganisms in parasitic or symbiotic interactions, while others are free-living, indicating divergent evolutionary trajectories and contrasted metabolic capabilities [16]. In addition, the genomes of these lineages have been predicted to encode several giant proteins of uncharacterized function that could be involved in interspecies recognition and cell wall degradation [20]. These results argue for an unanticipated implication of these enigmatic and potentially highly active lineages in the global biochemical cycles and in the microbial loop that deserve more attention.

To move in this direction, we present a spatially and temporally genome-resolved metagenomic study conducted on water samples collected between 2014 and 2022 at Lake Dziani Dzaha. This ecosystem is a volcanic maar lake that formed 7000–4000 years ago, trapping oceanic waters that have evolved over time into a warm (~ 30 °C) hypersaline (up to 71 psu) and hyperalkaline (pH 9.4–10) environment [21]. Previous 16S and 18S rRNA gene surveys that characterized its microbial community suggested that microbial activities were initially (in 2014–2015) likely focused on the degradation of the abundant phytoplankton biomass (i.e. the microeukaryote *Picocystis salinarum* and the cyanobacterium *Limnospira fusiformis*) that colonize the water column [22, 23]. However, the seismic crisis that has significantly altered the lake's biogeochemistry since 2018 led to an intense restructuration of its microbiome, inhuming parts of the initial community while awaking novel lineages from the seed bank [24]. These studies also revealed that the lake microbiome had low sequence similarity to known species highlighting its uniqueness and its strong potential for the detection of novel microbial lineages. Focusing on the PVC superphylum, we revealed the undiscovered microbial diversity associated with this poly extreme environment and their potential roles in ecosystem functioning and microbiome assembly. In addition, we disclosed their preferred ecological niches across the multiple geochemical contexts of the lake and the effects of the seismic crisis on these polyextremophile microorganisms.

Methods

Study site and samples collection

Lake Dziani Dzaha, located in Petite Terre (Mayotte, Comoros archipelago, latitude 12°46.237′ S, longitude 45°17.315′ E) is a shallow (~ 4 to 5 m deep) volcanic crater lake, with a narrow and 17 m deep pit originating from the chimney of the eruption at the lake's origin, that has been monitored since 2007 [21]. Until 2017, the geochemistry of the water column was characterized by a strong stability with an oxycline located at ~ 1.5 m, increasing sulfide (up to 6 mM), methane (up to 2 mM) and ammonia (up to 4.5 mM) concentrations with depth, and a salinity gradient with the position of the halocline depending on the precipitation and evaporation levels associated with dry and wet seasons [25, 26]. This geochemical stratification led to the separation of microbial processes between oxic and anoxic niches [27]. However, an active seismicity has recently altered the ecosystem. Underground magmatic movement that began in 2018 led to a strong bubbling of CO_2. This disrupted the stratifications, averaged the pH (~ 8.1) and introduced oxygen throughout the water column [25]. The mechanical

Vigneron *et al. Environmental Microbiome*   (2025) 20:48

Page 3 of 11

mixing of the water by the advective plume of magmatic gazes emitted from the western slope of pit [25] induces now a deeper and rapidly fluctuating oxygen penetration, creating (micro)-oxic conditions throughout the water column. Consequently, sulfides, methane and ammonia concentrations decreased in the water column, with significant effects on the abundance and composition of the microbial community [24, 25]. Water samples of Lake Dziani Dzaha were collected in the pit during five campaigns (Nov. 2014, Nov. 2015, Nov. 2017, Jun. 2022 and Nov. 2022) at seven (0.25 m, 1 m, 2.5 m, 5 m, 11 m, 15 m and 17 m) discrete depths, covering the contrasted historical and current ecological niches of the lake as previously described [24]. Subsampling and nucleic acid extraction protocols were detailed in a previous study characterizing the microbial community by 16S and 18S rRNA gene sequencing [22].

Library preparation, sequencing and analysis

Eight metagenomes were previously produced from Nov. 2017 samples (ncbi Bioproject PRJNA1037317), and 42 additional metagenomic libraries, generated from Nov. 2014, Nov. 2015, Jun. 2022 and Nov. 2022 samples, were sequenced using the Illumina NovaSeq 2*150 bp platform by Fasteris company (Plan-Les-Ouates, Swiss), resulting in an average of $1.24 \pm 0.17 \times 10^8$ reads per samples (Supplementary Table 1). First, sequences were quality filtered using Bbduk v.38.90 [28]. Then, for metagenome assembly, retained sequences were corrected for potential errors using SPAdes v.3.15.4 [29] (−only-error-correction), then pooled and co-assembled using MEGAHIT v.1.2.9 [30]. Read coverage of the contigs was determined using bwa-mem (http://bio-bwa.sourceforge.net). Contigs longer than 2000 bp were binned using MetaBAT-2 [31]. The completeness and contamination levels of the MAGs were assessed using both CheckM v.1.1.5 [32] and CheckM v.2 [33]. The relative abundance of the MAGs across samples was estimated from the contig depths calculated using bwa-mem and extracted using the jgi_summarize_bam_contig_depths script available in MetaBAT-2. Average coverages were calculated per MAGs, then relative abundances were normalized by the number of sequences passing quality filtration per metagenome and multiplied by 1,000,000 to expressed results in reads per million of sequences (RPM). This normalization allows comparison of the MAGs abundance across samples while providing user-friendly numbers. The taxonomic assignment of the MAGs was performed using GTDB-Tk (v.2.4.0, database R220) [34]. In addition, amino acid sequences of ribosomal proteins (rpL2, rpL3, rpL4, rpL5, rpL6, rpL14, rpL15, rpL16, rpL18, rpL22, rpL24, rpS3, rpS6, rpS8, rpS10,

rpS17, rpS19) were recovered from the MAGs and reference genomes downloaded from Genbank and the Genome Taxonomy database using HMM profiles. Reference genomes were selected following a three steps strategy. First, we included the closest representative genomes from the genome taxonomy database (GTDB) based on gtdb-tk backbone tree. Then, we compared our genomes on IMG/MER platform and added the closest representatives that were not already included in GTDB. Finally, we included additional genomes from Genbank that were affiliated to Ca. Auribacterota. Dataset was curated for completeness and contamination then only MAGs with at least 70% of the 17 ribosomal proteins were considered. These sequences were aligned with using Mafft v7.511 [35]. Resulting multiple alignments were concatenated in a large supermatrix combining 166 sequences and 6413 amino acid positions. A maximum likelihood phylogenetic tree was inferred using IQ-TREE v3 [36] using model LG + F + I + G4 as determined by IQ-TREE. The branch robustness was measured with the fast bootstraps procedure implemented in IQ-TREE (1000 replicates) and the SH-aLRT test (1000 replicates). The resulting tree was visualized using iTOL v.6 [37]. Pairwise average amino-acids identity (AAI) percentages between MAGs were calculated using the Genome-based distance matrix calculator [38].

Open reading frames (ORFs) of the MAGs were identified using Prodigal v.2.6.3 [39], then the functional annotation was carried out using KofamScan with the KEGG database v110.0 [40]. To consolidate annotation results, MAGs were also processed locally through NCBI's prokaryotic genome annotation pipeline (PGAP v6.6 2023–10-03.build7061) [41] and screened with custom HMM profiles collected from METABOLIC v.4 [42], FeGenie [43] and metabolisHMM [44]. The results were manually checked for the presence of specific pathways. Predicted protein sequences were also compared against the CAZY database using dbCAN3 [45], and glycoside hydrolase genes were analyzed to infer the catabolic potential of the MAGs. Hydrogenase sequences were classified using HydDB [46]. Individual phylogenies of key genes (hydrogenase, *ndh2*, *eetA*, *hxlA*) were constructed as described below to confirm their identity. The protein sequences of these genes were compared against the NCBI non-redundant protein database using BLASTP. Best hits of Blast (10 per query) as well as reference sequences (reviewed Uniprot entries for *hxlA*, *ndh2* and *eetA* genes, hydrogenase sequence database from HydDB) were downloaded. Amino-acid sequences were aligned using Mafft [35] and maximum-likelihood trees were calculated with IQTREE v.3 with 1000 bootstraps and 1000 SH-aLRT [36].

Vigneron *et al. Environmental Microbiome*　　(2025) 20:48

Page 4 of 11

Large protein analysis

Proteins larger than 5000 amino acids with no characterized function were recovered from the MAGs and reference genomes then analyzed in detail. Domains were identified using the NCBI RPS-BLAST against the Conserved Domain Database, InterProScan 5.61–93.0 with Interpro database v.93 [47] and Phyre2 [48]. Presence of signal peptide was determined by SignalP 6.0 [49] and putative localization of the proteins and the number of transmembrane helixes by DeepTMHMM v1.0.24 [50] and DeepLocPro v1 [51].

Results and discussion

Lake Dziani Dzaha hosts novel lineages of extremophiles from the PVC superphylum

After binning of the contigs reconstructed from the co-assembly, a total of 1185 unique MAGs was recovered, including 582 medium- to high-quality MAGs (>70% completeness and <5% contamination). Taxonomic assignment of the MAGs performed using the GTDB-Tk pipeline indicated that 16 of the 582 good-quality MAGs belonged to 11 uncharacterized phyla, revealing that the poly-extreme Lake Dziani Dzaha harbors an unexplored genomic diversity of microorganisms. Although 16S rRNA genes were not included in the MAGs, the

phylogenomic analysis of ribosomal proteins indicated that a third (n=5) of these uncharacterized MAGs belonged to the Planctomycetes-Verrucomicrobia-Chlamydia superphylum, as siblings of the Omnitrophota and Elusimicrobia clades (Fig. 1 and supplementary Table 2).

Two of them, which we named Ca. Menyafoubacter (bin.170) and Ca. Nundrabacter (bin.143) (Menyafou and Nundra meaning "destructive" and "deep", respectively, in Shimaore, one of the languages of the indigenous people of Mayotte) branched out into the poorly resolved Auribacterota phyla, close to Ca. Erginobacter and Ca. Euphemobacter that were recovered from the anoxic saline waters of Lake Ace (Antarctica) [4]. In contrast, Ca. Fungwabacter (bin.56) and Ca. Vouabacter (bin.1117) (Fungwa and Voua meaning "trapped" and "rainfall", respectively, in Shimaore) belonged to the uncharacterized candidate phyla CG03, and formed a distinct (pairwise amino acid identity < 50% with other phyla, Supplementary Table 3), monophyletic and cohesive group with genomes recovered from anoxic saline lakes, which we propose to name Shingomicrobia; "Shingo" meaning salty in Shimaore. Finally, bin.1047, which branched deeply in the phylogenomic tree and shared a low (43.91%) average amino acid identity with the closest reference genome (JAHJDO01 GCA018812485)
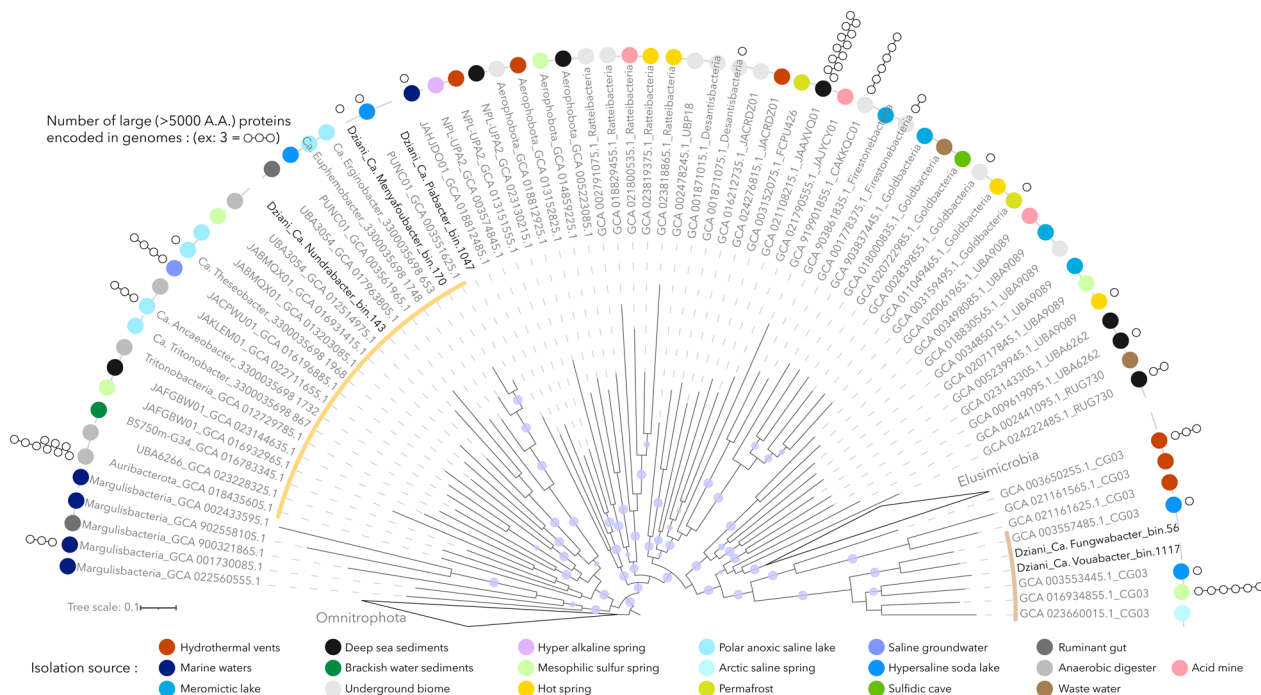
Fig. 1 Maximum likelihood tree of the PVC subgroups analysed in this study. The tree was calculated based on the concatenated alignment of 17 ribosomal proteins and rooted with the Omnitrophota phyla. Only bootstraps > 90% were represented by purple dots on the branches. Colored dots reflect the biome of origin where reference genomes were recovered. White circles at the end of the branches indicate the number of putative large (> 5000 amino-acids) proteins identified in genomes. Novel genomes (n = 5) are labeled in bold. Yellow and brown arcs indicate the Auribacterota and Shingomicrobia clades, respectively

may correspond to a novel class in an as yet undefined phylum, for which we propose the name of Ca. Piabacter ("Pia" meaning "novel" in Shimaore). Analysis of the environmental niches in which reference genomes were recovered indicated that, with the exception of Ca. Nundrabacter which forms a taxonomically distinct group with genomes recovered from anoxic saline waste digesters, all novel lineages cluster with genomes originating from hypersaline soda lakes and polar anoxic saline lakes [4, 6, 18], with environmental conditions very similar to Lake Dziani Dzaha. Furthermore, the depth and seasonal distribution of the MAGs in the water column of lake Dziani Dzaha indicated a niche preference for the highly sulfidic (4–6 mM of H_2S), anoxic environments and extreme salinity (60–70 psu) and alkalinity (pH 10) that characterized the bottom of the lake in 2014 and 2015 [22] and to a lesser extend in June 2022 (Fig. 2). For example, up to 2750 sequences per million of sequenced reads were detected in 2014 for Ca. Menyafoubacter. In contrast, the coverage of the MAGs decreased since 2017, reaching less than 1 sequences per million of sequenced reads in 2022, possibly corresponding to relict DNA. Altogether, these results suggest that these lineages form novel groups of extremophiles adapted to hypersaline and alkaline environments and illustrates that extreme environments are sources of an unexplored biodiversity.

Although, these bacteria were relatively rare in the Lake Dziani Dzaha microbial community, ranking from 115 to 470th in relative abundance among the 582 medium- to high-quality MAGs recovered from Lake Dziani Dzaha, minority lineages could have major ecological roles in ecosystems [52]. Genomes mining suggested a strictly anaerobic lifestyle without respiratory cytochromes (Fig. 3), which is consistent with the depth distribution of the lineages (Fig. 2) and genomes mining of the sibling groups Omnitrophota [17] and Elusimicrobia [16]. However, despite their taxonomic branching within the PVC superphylum, a large dissimilarity in gene content (Bray–Curtis similarity < 0.57) was detected between genomes, suggesting contrasted metabolic capabilities and ecological functions and illustrating the large genomic diversity within the PVC superphylum.

Extracellular electron transfer in Shingomicrobia (non-thermophilic CG03) lineages

Genes encoding group A3 FeFe bifurcating hydrogenases (Fig. 3 and Supplementary Fig. 1), RnfABCDEG and EtfAB/Bcd complexes were found in the genomes of
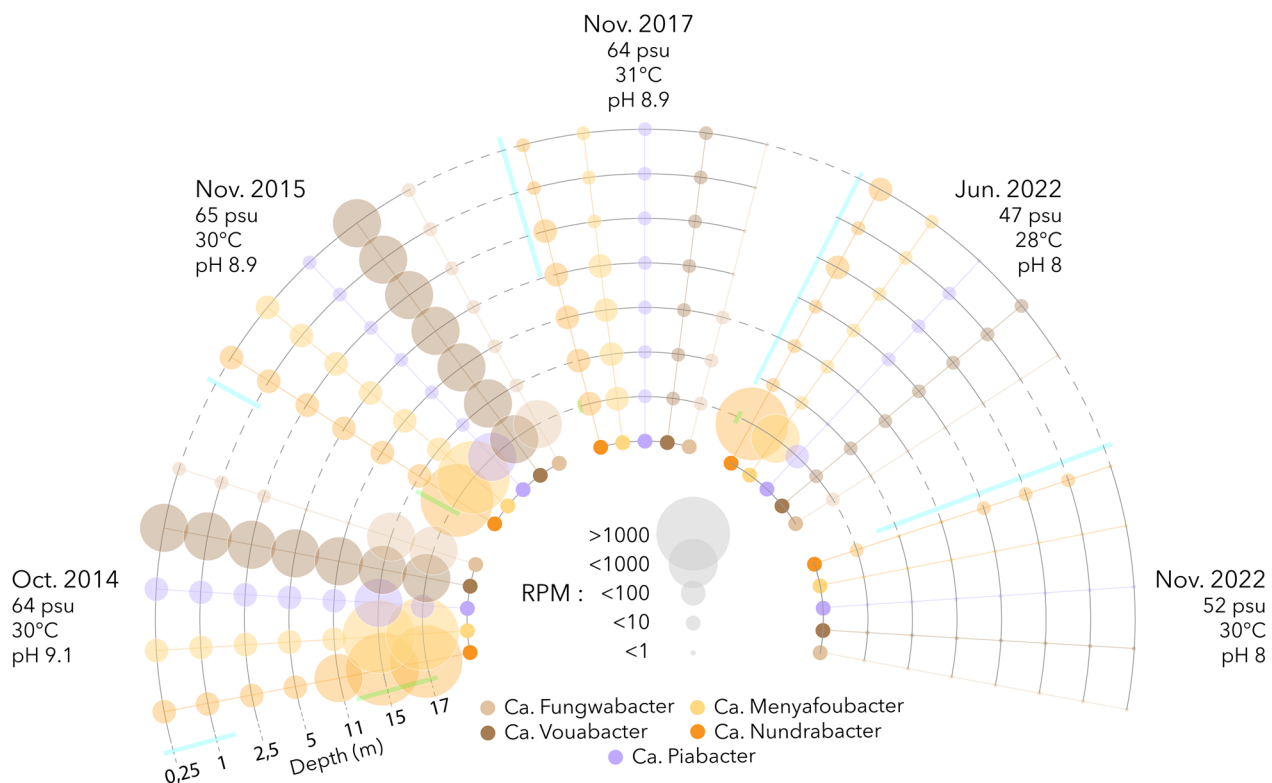

Fig. 2 Depth and seasonal distribution of the MAGs. Size of the dots represents the relative proportion (RPM: reads per million of sequences) of each MAG in the dataset. Blue and green bars localise the position of the oxic and sulfidic zones in the water column. Average salinity, temperature and alkalinity throughout the water column during the different sampling periods are indicated along the dates
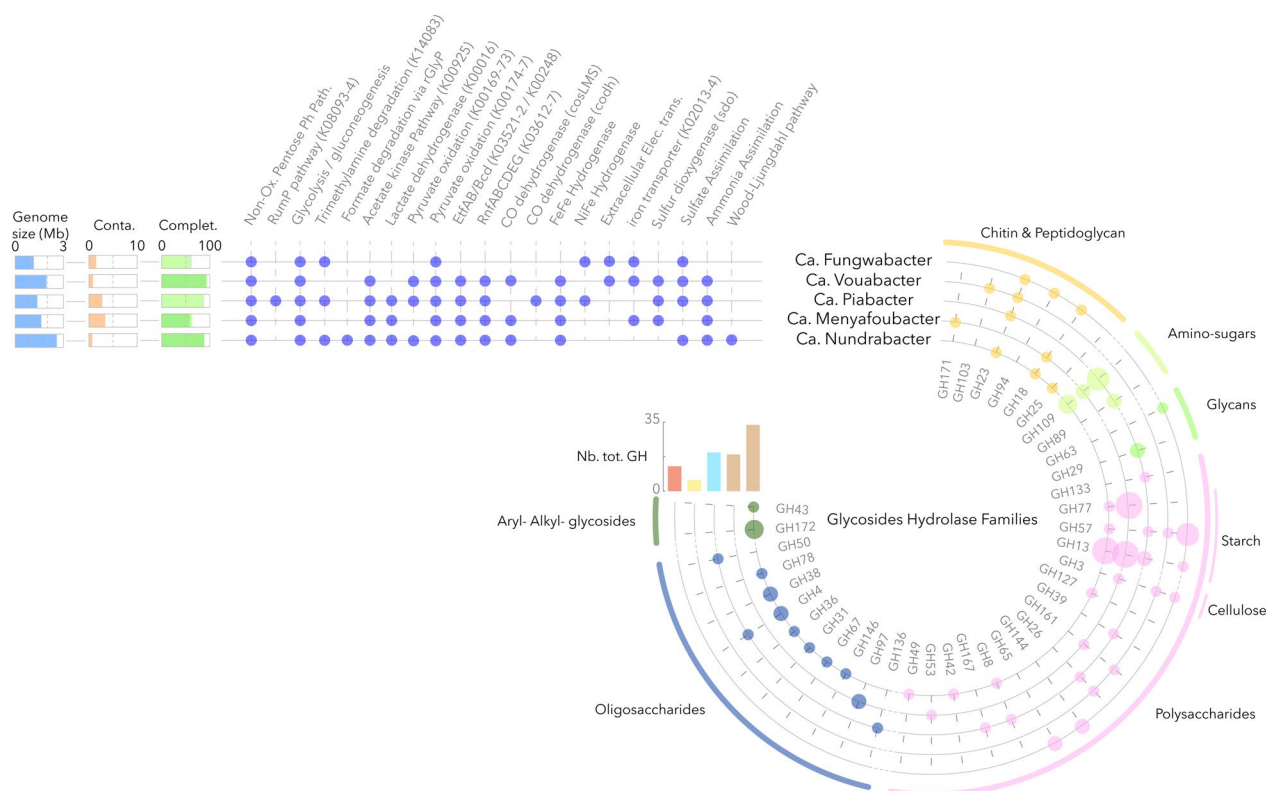
Fig. 3 Genome content and metabolic prediction of the MAGs. Genome size, contamination and completeness were determined using CheckM1 and CheckM2. The presence of metabolic pathways was assumed when genes were identified using both KofamScan and PGAP functional annotations. Extracellular electron transfer locus detection was performed using FeGenie and the sets of glycoside hydrolases were predicted by dbCAN3. Size of the dot in glycoside hydrolase circles indicates the number of copies of the gene (max=5), and colors indicate the potential substrate families

Ca. Vouabacter, Ca. Piabacter, Ca. Menyafoubacter and Ca. Nundrabacter (Fig. 3), indicating that flavin-based electron bifurcation systems are largely represented in the genomes of these lineages, as observed in the sister phyla Elusimicrobia [16]. These systems, often found in strictly anaerobic bacteria, provide these lineages with mechanisms for energy conservation, intracellular redox balance and hydrogen cycling, allowing exergonic reactions in the cells [53]. Interestingly, a flavin-based extracellular electron transfer (FLEET) locus [54] was also detected in the taxonomically close Ca. Vouabacter and Ca. Fungwabacter lineages (Fig. 4). These genes were also identified in the genomes of the non-thermophilic CG03, which clustered with these lineages within the Shingomicrobia clade, supporting the monophyly of this clade (Fig. 4), as well as in the recently described phyla of Candidatus Effluviviacota that also branched within the PVC superphyla [55]. Although further research is needed to explore the functional implications of these findings and the potential interactions between these genes and other metabolic pathways. The detection of the FLEET locus indicates that in addition to the Rnf complex and

hydrogenases, Ca. Vouabacter and Ca. Fungwabacter lineages could transfer electron extracellularly for energy production [54, 55]. Soluble iron (II) and ferrihydrite are supplied to the Lake Dziani Dzaha by the weathering of the crater [21]. These compounds have been found to accept electrons from FLEET [56], supporting an extracellular iron reduction metabolism for these microorganisms. This reaction could be particularly advantageous in Lake Dziani Dzaha, where elevated H_2S/HS^- concentrations, that were measured at the beginning of the survey (2014–2015) [24], form conductive iron sulfides that accelerate and extend the process through distant ecological niches [57]. Extracellular electron transfer has previously been associated with syntrophic relationships, such as in anaerobic methane oxidizing consortia, where sulfate-reducing bacteria act as electron acceptors [58]. Therefore, unidentified members of the community could also potentially be substitute electron acceptors for Shingomicrobia in an original syntrophic relationship. Alternatively, this pathway could enhance ecological fitness by contributing to the cellular $NADH/NAD^+$ and proton redox balance, or by enhancing iron bioavailability under
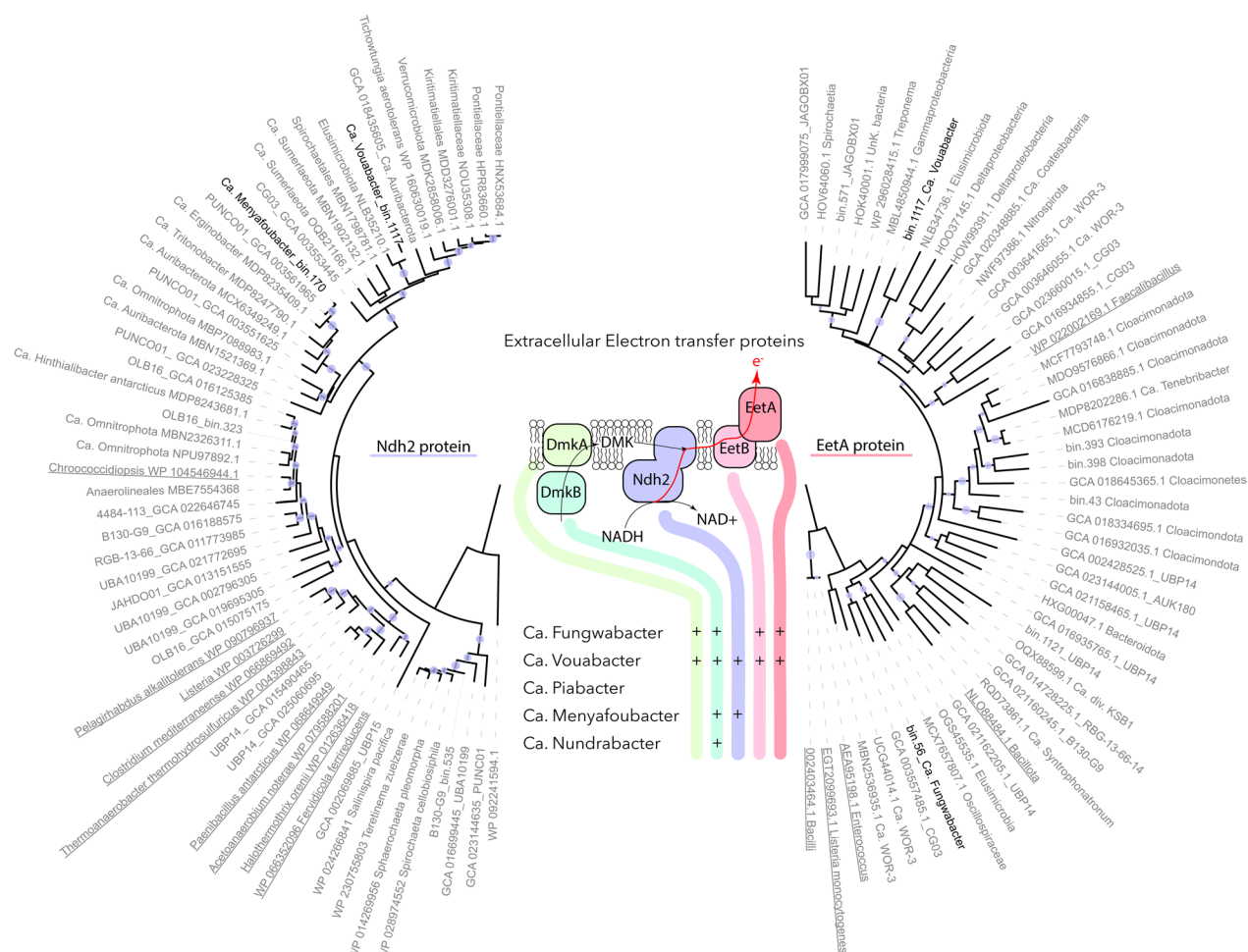
Fig. 4 Composition of the extracellular electron transport complex in the MAGs and phylogenetic trees of Ndh2 and EetA protein sequences. MAGs recovered from Lake Dziani Dzaha are shown in bold and lineages with probable functional extracellular electron transport complex are underlined. Only bootstraps > 90% are represented by purple dots on the branches

anoxic conditions [59], which would be assimilated via the iron complex transport system also identified in the genomes of these lineages (Fig. 3). Phylogenetic trees of the specialised type II NADH dehydrogenase Ndh2 (bin.1117_001767) and the EetA protein (bin.56_001019, bin.1117_001550) involved in the transmembrane apparatus revealed a taxonomic proximity with sequences related to Elusimicrobia, Auribacteria, Kiritimatiellales, Tichowtungiia and Cloacimonadota, supporting a broader but paraphyletic distribution of the FLEET locus within the PVC superphylum (Fig. 4).

Carbon monoxide, amines and formaldehyde assimilation by Ca. Piabacter

The study of Ca. Piabacter MAG revealed a specialization in the fermentation of small carbonated molecules, including (tri)methylamines and amino-sugars. A CO dehydrogenase (CODH) and acetyl-CoA synthase (ACS)

complex, coupled with [NiFe] hydrogenase group 4c was also detected, indicating a carbon monoxide metabolism. Although minimal, this complex provides an efficient primary strategy for energy production [60] and an escape from competition. In addition, the potential for formaldehyde assimilation through the *hxlAB* operon, encoding the 3-hexulose-6-phosphate synthase (HPS) and 6-phospho-3-hexuloisomerase (PHI), was also detected (bin.1047_000919 and 1047_000920). These genes, which are part of the ribulose monophosphate (RuMP) pathway, are usually detected as a one-carbon assimilation pathway in aerobic methylotrophic bacteria and as a detoxification process in non-methylotrophic aerobes [61]. Similar but taxonomically distant genes are also found in anaerobic archaea and function as an alternative pentose phosphate pathway [62]. However, the presence of this pathway in an anaerobic bacterium is questionable. BlastP of the *hxlAB* operon identified in Ca. Piabacter

Vigneron *et al. Environmental Microbiome* (2025) 20:48

Page 8 of 11

against the NCBI database and phylogenetic analysis of *hxlB* showed similarities with sequences recovered from sibling phyla of Ca. Omnitrophota, Ca. NPL-UPA2, Ca. Firestonebacteria, Desantisbacteria, Ca. Aerophobus, which formed a distant cluster of environmental sequences with some Desulfobacteraceae and Actinobacteria representatives (Fig. 5). Although experimental validation is required because neighboring sequences in the phylogenetic tree are uncharacterized, the presence of the *hxlAB* genes and the subsequent pathway for the degradation and assimilation of the fructose-6-phosphate supports the presence of an unforeseen formaldehyde assimilation/detoxification in anaerobic uncultivated lineages of the PVC superphylum.

Extended catabolic arsenal of Ca. Nundrabacter

Among the PVC MAGs from this study, Ca. Nundrabacter has the largest genome with up to 2.6 Mb and 2045 predicted coding genes. Consistently, an extended metabolic repertoire was detected, including a mixotrophic lifestyle with both CO_2 fixation potential via the Wood-Ljungdhal pathway and fermentations with lactate, acetate and potentially hydrogen as end products. Formate utilisation via the reductive glycine pathway was also identified, including genes encoding the formate tetrahydrofolate (THF) ligase (bin.143_002017), the methenyl-THF cyclohydrolase and dehydrogenase (bin.143_001550), and a complete glycine cleavage system (bin.143_001051-53) [63]. Trimethylamine degradation

potential (bin.143_001855) was also detected in Ca. Nundrabacter providing an additional source of carbon and nitrogen. A large catabolic arsenal was also predicted in Ca. Nundrabacter genome with up to 22 different (35 in total) subfamilies of glycoside hydrolases and 23 carbon-binding modules targeting various polysaccharides (rhamnose, sucrose, xylan, arabinan, glucan, mannan), reserve molecules (glycogen, glucosylglycerate, starch), chitin, peptidoglycan, and aryl- and alkyl-glycosides (Fig. 3). Similar extended degradation capabilities were also identified in the closest references genomes, named UBA3054 in the Genome Taxonomy database, that were recovered from ruminant gut and anaerobic digesters, suggesting that they form a distinct clade of anaerobic recyclers of complex organic matter within the phylum Auribacterota.

Novel extracellular large catabolic proteins in Menyafoubacteria

Numerous lineages in the PVC superphyla have been predicted to contain large to giant genes, suggesting the production of massive proteins (up to 85,800 amino acids) [20]. This is particularly relevant in the PVC clades of Omnitrophota [17], Elusimicrobia and CG03, where giant genes are common in the genomes [20]. While the post-transcriptional integrity of such large proteins remains unclear, the presence of multiple transmembrane helixes and carbohydrate degradation domains has been interpreted as massive surface weapons for
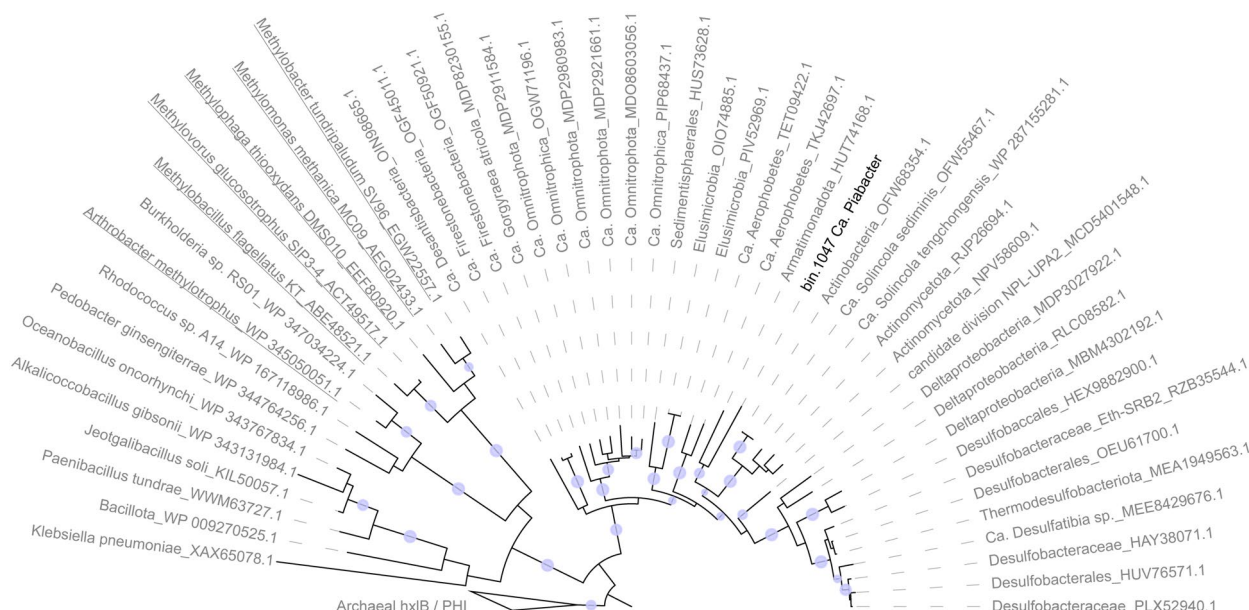
Fig. 5 Phylogenetic tree of the HxlB/PHI protein sequences. MAG of Ca. Piabacter recovered from Lake Dziani Dzaha is shown in bold and genomes from C_1 oxidizing lineages are underlined. Only bootstraps > 90% are represented by purple dots on the branches. Tree was rooted with archaeal Hxlb sequences

Vigneron *et al. Environmental Microbiome*   (2025) 20:48

Page 9 of 11

predatory bacteria [17, 20] or novel extracellular cellulosome-like structures [15]. Screening of the predicted proteomes revealed the presence of large proteins (> 5000 amino acids) (Fig. 1). However, the taxonomic distribution of large proteins was patchy, maybe due to methodological differences for the assembly and binning of genomic data. Among the MAGs recovered from Lake Dziani Dzaha, only Ca. Menyafoubacter harbored a predicted large (~ 6000 aa) protein (bin.170_000073) with a standard secretory signal peptide and no transmembrane helix suggesting an extracellular localization (Fig. 6).

The large protein of Ca. Menyafoubacter has several tandem carbohydrate-binding/sugar hydrolysis beta-helix (CASH) domains similar to pectin-lyase beta-helix and a large polysaccharide lyase 6 (e.g. alginate lyase) domain, suggesting a carbohydrate-degrading potential, focused on green algal cells. A disaggregase region, potentially inducing dissociation of cell aggregates [64] and various domains found in virulence factors such as in Rhs toxin that induces apoptosis of neighboring cells [65, 66], CglD-type adhesin and internalin domains were also identified in the protein, potentially favorizing its adhesion to other cells [67]. In addition, intrinsically disordered regions that allow structural flexibility [68] were also detected. Very similar proteins were also identified in taxonomically close MAGs recovered from a soda lake of the Kulunda steppe [6] (PUNC01 GCA_003551625) and anoxic saline waste digesters (PUNC01 GCA_035380285), supporting the distribution and structure of this large protein within this extremophile clade. Taken together, these results suggest that Ca. Menyafoubacter and related lineages could potentially catalyze the degradation of plant and/or green alga complex. The vegetation around Lake Dziani Dzaha and in the Kulunda steppe are highly contrasted. However, both ecosystems host a large population of the trilobed picoalgae *Picocystis salinarum* [69], which could potentially be the target of this large extracellular protein.

Roles in ecosystem functioning

Members of the PVC superphylum have been found to be involved in numerous biochemical transformations, such as in the nitrogen [70, 71], methane [72, 73] and sulfur [18, 74, 75] cycles. Our results indicate that the newly defined lineages are likely to be involved in the carbon, hydrogen and iron/electron cycles in hypersaline and alkaline ecosystems, expending our knowledge of the metabolism of the PVC members. In addition, these lineages may contribute to the recycling of the biomass as potential primary degraders or predators, extending previous observations from some members of the sister taxon Omnitrophota [17] and suggesting an unforeseen implication in the microbial loop of extreme environments. Although, these genomic investigations of the dark microbial matter lineages related to PVC superphylum require experimental and functional validations, they revealed a large metabolic diversification among the extremophilic lineages of this phylogenetic radiation, possibly related to niche differentiation and substrate availability and revealed contrasted interactive behaviors with other members of the community. Fluctuations in their distribution and abundance along the water column over time highlighted their high sensitivity to the "moderation" of the environmental conditions in the lake (e.g. oxygenation, decrease in pH and sulfide concentration). Although

Protein domains identified in excreted large proteins
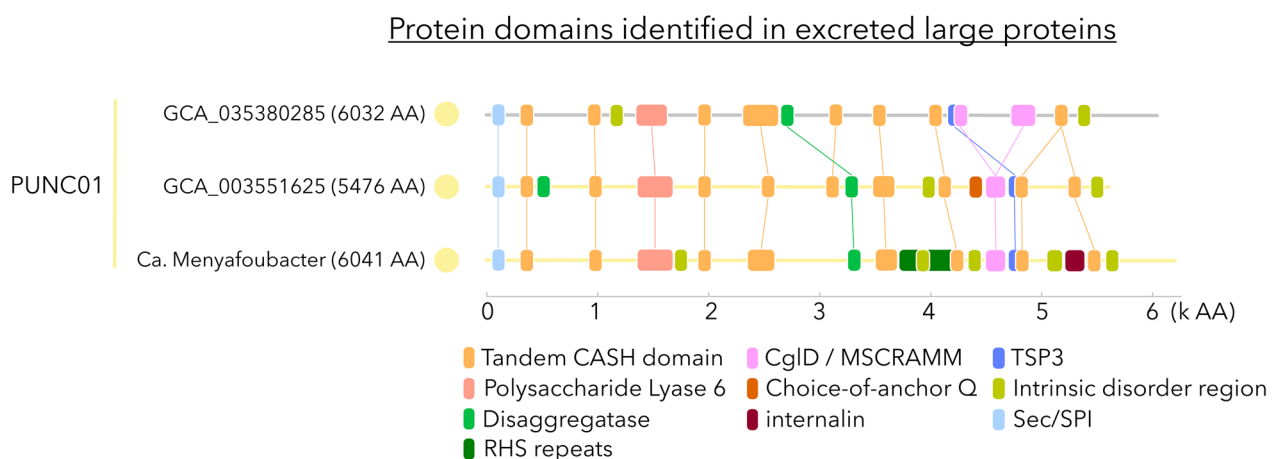
Fig. 6 Characterization of the putative large proteins identified in the MAG of Ca. Menyafoubacter and related reference genomes of the PUNC01 clade. Sec/SPI: secretory signal peptides transported by the Sec translocon and cleaved by signal peptidase I; CASH: carbohydrate-binding/sugar hydrolysis beta-helix; RHS: rearrangement hotspot; MSCRAMM: microbial surface components recognizing adhesive matrix molecules; TSP3: bacterial thrombospondin-3

Vigneron *et al. Environmental Microbiome*   (2025) 20:48

Page 10 of 11

all MAGs contained antioxidant enzymes (e.g. peroxiredoxin, super oxide dismutase or reductase, glutaredoxin), suggesting a relative tolerance to oxic stress, the changes in environmental conditions seemed to have been fatal for these lineages, confirming their strictly extremophilic lifestyle, and highlighting the major consequences of environmental perturbations for biodiversity in extreme environments.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40793-025-00699-1.

Additional file 1.

Additional file 2.

Additional file 3.

Author contributions

AV: analyzed data, interpreted results, wrote the manuscript and prepared Figs and Tables. LC: contributed to sampling, acquired data, revised the manuscript. CBA: analyzed data, interpreted results, revised the manuscript. JPF: analyzed data, interpreted results, revised the manuscript. MT: acquire funding, revised the manuscript. CB: contributed to sampling, acquire funding, revised the manuscript. HA: contributed to sampling, acquire funding, revised the manuscript. PO: revised the manuscript. MH: contributed to sampling, acquired data, analyzed data, interpreted results, acquire funding, revised the manuscript.

Data availability

Data availability Geochemical data has been previously published 25. The raw metagenomic reads produced from Lake Dziani Dzaha are available in NCBI SRA database under BioProjects PRJNA1037317 and PRJNA1222159 (description of the Bioprojects is available in Supplementary Table 1). Nucleotide and amino acid sequences of the MAGs as well as annotation files are freely available in figshare: https://doi.org/10.6084/m9.figshare.26347213.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹INSA Lyon, CNRS, UMR5240 Microbiologie Adaptation et Pathogénie, Universite Claude Bernard Lyon 1, 69621 Villeurbanne, France. ²CNRS, VetAgro Sup, Laboratoire de Biométrie et Biologie Évolutive, UMR5558, Universite Claude Bernard Lyon 1, Villeurbanne, France. ³MARBEC, IRD, CNRS, Ifremer, Univ. Montpellier, Sète, France. ⁴UMR 7245 Molécules de Communication et Adaptations des Microorganismes (MCAM) MNHN-CNRS, Muséum National d'Histoire Naturelle, CP 39, 12 Rue Buffon, 75231 Paris Cedex 05, France. ⁵LIENSs, UMR 7266, La Rochelle Université - CNRS, 2 Rue Olympe de Gouges, 17000 La Rochelle, France. ⁶Institut Universitaire de France (IUF), Paris, France.

References

1. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat Commun. 2018;9:4999.
2. Gong X, et al. New globally distributed bacterial phyla within the FCB superphylum. Nat Commun. 2022;13:7516.
3. De Anda V, et al. Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. Nat Commun. 2021;12:2404.
4. Williams TJ, Allen MA, Panwar P, Cavicchioli R. Into the darkness: the ecologies of novel 'microbial dark matter' phyla in an Antarctic lake. Environ Microbiol. 2022;24:2576–603.
5. Vigneron A, Vincent WF, Lovejoy C. Discovery of a novel bacterial class with the capacity to drive sulfur cycling and microbiome structure in a paleo-ocean analog. ISME Commun. 2023;3:82.
6. Vavourakis CD, et al. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome. 2018;6:168.
7. Belilla J, et al. Hyperdiverse archaea near life limits at the polyextreme geothermal Dallol area. Nat Ecol Evol. 2019;3:1552–61.
8. Probst AJ, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. Environ Microbiol. 2017;19:459–74.
9. Anantharaman K, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun. 2016;7:13219.
10. Parks DH, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.
11. Brown CT, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015;523:208–11.
12. Castelle CJ, et al. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 2018;16:629–45.
13. Baker BA, et al. Expanded phylogeny of extremely halophilic archaea shows multiple independent adaptations to hypersaline environments. Nat Microbiol. 2024;9:964–75.
14. Doud DFR, et al. Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. ISME J. 2020;14:659–75.
15. Williams TJ, Allen MA, Berengut JF, Cavicchioli R. Shedding light on microbial "dark matter": insights into novel cloacimonadota and omnitrophota from an antarctic lake. Front Microbiol. 2021. https://doi.org/10.3389/fmicb.2021.741077.
16. Méheust R, et al. Groundwater Elusimicrobia are metabolically diverse compared to gut microbiome Elusimicrobia and some have a novel nitrogenase paralog. ISME J. 2020;14:2907–22.
17. Seymour CO, et al. Hyperactive nanobacteria with host-dependent traits pervade Omnitrophota. Nat Microbiol. 2023;8:727–44.
18. Vigneron A, et al. Genomic evidence for sulfur intermediates as new biogeochemical hubs in a model aquatic microbial ecosystem. Microbiome. 2021;9:46.
19. López-Mondéjar R, Tláskal V, da Rocha UN, Baldrian P. Global distribution of carbohydrate utilization potential in the prokaryotic tree of life. mSystems. 2022;7:e00829-e922.
20. West-Roberts J. et al. Giant genes are rare but implicated in cell wall degradation by predatory bacteria. bioRxiv. 2023. https://doi.org/10.1101/2023.11.21.568195
21. Sarazin G, et al. Geochemistry of an endorheic thalassohaline ecosystem: the Dziani Dzaha crater lake (Mayotte Archipelago, Indian Ocean). Comptes Rendus Géoscience. 2020;352:559–77.
22. Hugoni M, et al. Spatiotemporal variations in microbial diversity across the three domains of life in a tropical thalassohaline lake (Dziani Dzaha, Mayotte Island). Mol Ecol. 2018;27:4775–86.

Vigneron *et al. Environmental Microbiome*   (2025) 20:48

Page 11 of 11

23. Bernard C, et al. Very low phytoplankton diversity in a tropical saline-alkaline lake, with co-dominance of arthrospira fusiformis (cyanobacteria) and *Picocystis salinarum* (Chlorophyta). Microb Ecol. 2019;78:603–17.
24. Vigneron A, et al. Seismic events as potential drivers of the microbial community structure and evolution in a paleo-ocean analog. Commun Earth Environ. 2024;5:504.
25. Cadeau P, Jézéquel D, Groleau A, Di Muro A, Ader M. Impact of the seismo-volcanic crisis offshore Mayotte on the Dziani Dzaha Lake. Comptes Rendus Géoscience. 2022;354:299–316.
26. Leboulanger C, et al. Microbial diversity and cyanobacterial production in Dziani Dzaha crater lake, a unique tropical thalassohaline environment. PLoS ONE. 2017;12:e0168879.
27. Escalas A, et al. Strong reorganization of multi-domain microbial networks associated with primary producers sedimentation from oxic to anoxic conditions in an hypersaline lake. FEMS Microbiol Ecol. 2021;97:fiab163.
28. Bushnell B. BBMap: a fast, accurate, splice-aware Aligner. (2014).
29. Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
30. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
31. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
33. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2023;20:1203–12.
34. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925–7.
35. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
36. Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
37. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
38. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr. 2016;4:e1900v1.
39. Hyatt D, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
40. Aramaki T, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.
41. Tatusova T, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24.
42. Zhou Z, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome. 2022;10:33.
43. Garber AI, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol. 2020. https://doi.org/10.3389/fmicb.2020.00037.
44. McDaniel EA, Anantharaman K, McMahon KD. metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. bioRxiv, 2020. https://doi.org/10.1101/2019.12.20.884627.
45. Zheng J, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51:W115–21.
46. Søndergaard D, Pedersen CNS, Greening C. HydDB: A web tool for hydrogenase classification and analysis. Sci Rep. 2016;6:34212.
47. Quevillon E, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33:W116–20.
48. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845–58.
49. Teufel F, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40:1023–5.
50. Hallgren J et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. https://doi.org/10.1101/2022.04.08.487609.
51. Moreno J, Nielsen H, Winther O, Teufel F. Predicting the subcellular location of prokaryotic proteins with DeepLocPro. bioRxiv, 2024. https://doi.org/10.1101/2024.01.04.574157.
52. Jousset A, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. 2017;11:853–62.
53. Buckel W, Thauer RK. Flavin-based electron bifurcation, a new mechanism of biological energy coupling. Chem Rev. 2018;118:3862–86.
54. Light SH, et al. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. Nature. 2018;562:140–4.
55. Su L, Marshall IPG, Teske AP, Yao H, Li J. Genomic characterization of the bacterial phylum Candidatus Effluviviacota, a cosmopolitan member of the global seep microbiome. MBio. 2024;15:e00992.
56. Stevens E, Marco ML. Bacterial extracellular electron transfer in plant and animal ecosystems. FEMS Microbiol Rev. 2023;47:fuad019.
57. Zhu F, et al. Biogenic iron sulfide functioning as electron-mediating interface to accelerate dissimilatory ferrihydrite reduction by Shewanella oneidensis MR-1. Chemosphere. 2022;288:132661.
58. Wegener G, Krukenberg V, Riedel D, Tegetmeyer HE, Boetius A. Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. Nature. 2015;526:587–90.
59. Jeuken LJC, Hards K, Nakatani Y. Extracellular electron transfer: respiratory or nutrient homeostasis? J Bacteriol. 2020. https://doi.org/10.1128/jb.00029-20.
60. Greening C, et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME J. 2016;10:761–77.
61. Kato N, Yurimoto H, Thauer RK. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. Biosci Biotechnol Biochem. 2006;70:10–21.
62. Orita I, et al. The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon thermococcus koda-karaensis. J Bacteriol. 2006;188:4698–704.
63. Sánchez-Andrea I, et al. The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. Nat Commun. 2020;11:5090.
64. Osumi N, et al. Identification of the gene for disaggregatase from Methanosarcina mazei. Archaea. 2008;2:949458.
65. González-Magaña A, et al. Structural and functional insights into the delivery of a bacterial Rhs pore-forming toxin to the membrane. Nat Commun. 2023;14:7808.
66. Koskiniemi S, et al. Rhs proteins from diverse bacteria mediate intercellular competition. Proc Natl Acad Sci. 2013;110:7032–7.
67. Pathak DT, Wall D. Identification of the cglC, cglD, cglE, and cglF genes and their role in cell contact-dependent gliding motility in Myxococcus xanthus. J Bacteriol. 2012;194:1940–9.
68. Trivedi R, Nagarajaram HA. Intrinsically disordered proteins: an overview. Int J Mol Sci. 2022;23:14050.
69. Singh J, Kaushik S, Maharana C, Jhingan GD, Dhar DW. Elevated inorganic carbon and salinity enhances photosynthesis and ATP synthesis in picoalga Picocystis salinarum as revealed by label free quantitative proteomics. Front Microbiol. 2023. https://doi.org/10.3389/fmicb.2023.1059199.
70. Strous M, et al. Missing lithotroph identified as new planctomycete. Nature. 1999;400:446–9.
71. Suarez C, et al. Metagenomic evidence of a novel family of anammox bacteria in a subsea environment. Environ Microbiol. 2022;24:2348–60.
72. Schmitz RA, et al. Verrucomicrobial methanotrophs: ecophysiology of metabolically versatile acidophiles. FEMS Microbiol Rev. 2021;45:fuab007.
73. Carere CR, et al. Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. ISME J. 2017;11:2599–610.
74. Wasmund K, Mußmann M, Loy A. The life sulfuric: microbial ecology of sulfur cycling in marine sediments. Environ Microbiol Rep. 2017;9:323–44.
75. Spring S, et al. Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. ISME J. 2016;10:2801–16.

Publisher's Note