Lineage-dependent partitioning of activities in chemoclines defines Woesearchaeota ecotypes in an extreme aquatic ecosystem

Lilian A. Cloarec[1], Thomas Bacchetta[1], Maxime Bruto[2], Christophe Leboulanger[3], Vincent Grossi[4,11], Céline Brochier-Armanet[5,6], Jean-Pierre Flandrois[5], Adrian Zurmely[5], Cécile Bernard[7], Marc Troussellier[3], Hélène Agogué[8], Magali Ader[9], Christine Oger-Desfeux[10], Philippe M. Oger[1], Adrien Vigneron[1] and Mylène Hugoni[1,6]*

Abstract

Background DPANN archaea, including Woesearchaeota, encompass a large fraction of the archaeal diversity, yet their genomic diversity, lifestyle, and role in natural microbiomes remain elusive. With an archaeal assemblage naturally enriched in Woesearchaeota and steep vertical geochemical gradients, Lake Dziani Dzaha (Mayotte) provides an ideal model to decipher their in-situ activity and ecology.

Results Using genome-resolved metagenomics and phylogenomics, we identified highly diversified Woesearchaeota populations and defined novel halophilic clades. Depth distribution of these populations in the water column showed an unusual double peak of abundance, located at two distinct chemoclines that are hotspots of microbial diversity in the water column. Genome-centric metatranscriptomics confirmed this vertical distribution and revealed a fermentative activity, with acetate and lactate as end products, and active cell-to-cell processes, supporting strong interactions with other community members at chemoclines. Our results also revealed distinct Woesearchaeota ecotypes, with different transcriptional patterns, contrasted lifestyles, and ecological strategies, depending on environmental/host conditions.

Conclusions This work provides novel insights into Woesearchaeota in situ activity and metabolism, revealing invariant, bimodal, and adaptative lifestyles among halophilic Woesearchaeota. This challenges our precepts of an invariable host-dependent metabolism for all the members of this taxa and revises our understanding of their contributions to ecosystem functioning and microbiome assemblage.

Keywords Archaeal genomic diversity, Woesearchaeales, Ecotypes, Metagenomics, Metatranscriptomics, Extreme lake

Background

In recent decades, cultivation-independent genomics has greatly increased our knowledge of natural microbial diversity [1–3]. This has led to the discovery of many new major lineages, including phyla and superphyla, such as the DPANN archaea, and to the redrawing of the Tree of Life [1, 4–7]. The DPANN superphylum, which includes the '*Ca.* Diapherotrites', '*Ca.* Parvarchaeota', '*Ca.* Aenigmarchaeota', Nanoarchaeota, '*Ca.* Nanohaloarchaeota', as well as the Woesearchaeota, Micrarchaeota, Pacearchaeota, Huberarchaeota, Mamarchaeota, and Undinarchaeota phyla [2, 8], represents a large radiation of the archaeal diversity and has been ubiquitously found in both oxic and anaerobic biomes ranging from animal microbiomes to fresh and marine waters, including acidic, alkaline, and hypersaline ecosystems [4, 9–13]. Pioneering studies using electronic microscopy on enrichment cultures depicted DPANN archaea as nanosized cells [6, 14–16], while genomic predictions indicated small genomes and limited catabolic and anabolic capabilities [16–19], suggesting a fermentative and pyruvate-centered metabolism [5, 20]. However, additional catabolic pathways such as the Embden-Meyerhof-Parnas pathway, an incomplete Entner-Doudoroff pathway, the beta-oxidation pathway, and a RubisCO-dependent nucleoside degradation pathway have been also reported [10, 21], illustrating the genomic diversity in this taxonomically large superphylum.

Many DPANN genomes are also characterized by the absence of genes encoding the enzymes involved in the synthesis of amino acids, purines, pyrimidines, lipids, and vitamins [3, 5], suggesting multiple dependencies on the environment or potential hosts through (epi)symbiotic or parasitic interactions [1, 16, 22, 23]. The rare cultivated DPANNs do have an epi-symbiotic lifestyle. For example, the DPANN archaeon *Nanoarchaeum equitans* is dependent for growth on its archaeal host, *Ignicoccus hospitalis* [16, 17], which fills its metabolic deficiencies. Other works reported the presence of Micrarchaeota in co-culture with Thermoplasmatales [24, 25], and Nanohaloarchaeota with *Halorubrum lacusprofundi* [26]. A physical association was also identified between the two DPANN archaea *Huberiarchaeum crystalense* and *Ca. Altiarchaeum hamiconexum* [27, 28], with *H. crystalense* potentially scavenging the cytoplasmic content of *Ca. Altiarchaeum* in a CRISPR-mediated interaction [29]. However, extensive metabolic capacities have been reported for several members of '*Ca.* Diapherotrites', '*Ca.* Micrarchaeota', and '*Ca.* Parvarchaeota' phyla, suggesting that some DPANN archaea may also be free-living [20, 30, 31].

Among the recently described DPANN archaea phyla, the Woesearchaeota phylum (referenced as the Woesearchaeales order within the Nanoarchaeia class and Nanoarchaeota phylum in the Genome Taxonomy Database, GTDB [32]) has yet no cultivated representatives, leaving the lifestyle and metabolism of the members of this phylum enigmatic. Meta-analysis of available genomes reported that some Woesearchaeota may be capable of anaerobic heterotrophy with fermentative metabolism relying on lactate dehydrogenase, [FeFe] hydrogenase, and flavin-based electron bifurcation mechanisms [13], while some Woesearchaeota genomes recovered from marine and hypersaline environments harbored sulfur-cycling genes [33, 34]. While these results support a host-independent energetic metabolism for some lineages, key metabolic pathways (i.e., glycolysis, TCA cycle, amino acids synthesis, lipid metabolism, etc.) are also absent or incomplete in most Woesearchaeota, suggesting contrasted capabilities and a strong dependence on a host or environmental for their supply [13]. However, it is noteworthy that a large fraction of the genes in Woesearchaeota genomes remains unassigned (e.g., hypothetical proteins) and, to our knowledge no transcriptomic data have yet confirmed the activity of identified genes so far, weakening metabolic and lifestyle inferences from genomic data, as observed for nanosized Patescibacteria [35]. Several studies speculating that Woesearchaeota could interact with various microbial and/or eukaryotic lineages have attempted to identify their potential hosts or partners through co-occurrence patterns [10, 34] or cell-sorting [36]. However, analysis of co-sorted cells isolated from subsurface environments concluded that they do not form symbiotic cell–cell associations and that their divergent genomic and cell characteristics may result from an ancestral trait [37]. Taken together these observations warrant for in-depth functional characterization of the Woesearchaeota phylum and a validation of their genomic potential through cultivation and activity-based approaches.

Lake Dziani Dzaha (Comoros Archipelago, Mayotte) is a poly-extreme environment, with both hypersaline (salinity > 50 psu) and hyperalkaline (pH > 9) conditions. This crater lake shows original geochemical features with periodically very high H_2S/HS^- concentrations, anoxic conditions below 1.5 m [38–40], and a high and stable phytoplanktonic biomass (524 to 875 chl$a$.μL^{-1}) [41]. The lake has steep physicochemical gradients that define different ecological niches [41]. In particular, the alternation of dry and rainy seasons leads to the establishment of a seasonal surface chemocline and a permanent deep chemocline at about 14 m, characterized by elevated H_2S/HS^- concentrations [42]. Previous surveys, based on 16S and 18S rRNA gene sequencing, have revealed that the biodiversity of the lake is limited to microorganisms, with no aquatic metazoans and no multicellular algae

identified to date [42–44]. The bacterial community is strongly dominated by the cyanobacterium *Limnospira fusiformis* (formerly *Arthrospira fusiformis*), while the microeukaryotic community is dominated by the chlorophyte *Picocystis salinarum*, and the archaeal assemblage by Woesearchaeota, representing up to 75% of the archaeal 16S rRNA gene sequences [42]. The dominance of Woesearchaeota among archaeal assemblage together with strong oxygen and sulfur gradients in the water column of Lake Dziani Dzaha provides thus an excellent model to study the diversity and metabolism of the members of this enigmatic phylum. Combining metagenomic and metatranscriptomic approaches, we first aimed to resolve Lake Dziani Dzaha Woesearchaeota phylogeny and to determine the origin and the diversity of these extremophiles (both halophilic and alkaliphilic). Secondly, we investigated the genomic repertoire of Woesearchaeota to decipher their ecological success across the strong oxygen and sulfur gradients of this model ecosystem. We then tested whether in-situ transcriptomic activities could support these genomic-based metabolic inferences and if their activities vary across the specific ecological niches of the system. Our results provide original insights into Woesearchaeota ecophysiology and demonstrate a complex and lineage-dependent partitioning of their transcriptomic activity, reflecting invariant, bimodal, and adaptative lifestyles among Lake Dziani Dzaha Woesearchaeota community in response to environmental/host conditions.

Materials and methods
Study site, sampling, and environmental parameters
Lake Dziani Dzaha (12°46′15.6″S; 45°17′19.2″E) is a volcanic crater lake (Fig. 1A), located in the Comoros archipelago (western Indian Ocean), on the Petite Terre Island of Mayotte (12°46′ 15.6″ S; 45° 17′ 19.2″ E).

Water samples were collected in November 2017 at the deepest point of the lake (maximum depth 17 m) at 7 discrete depths (0.25 m, 1 m, 2.5 m, 5 m, 11 m, 14 m, and 16 m) using a horizontal 1.2 L Niskin bottle. Water subsamples (45 mL) were first filtered through 3 μm pore-size polycarbonate Isopore™ membrane filters (NB Merck-Millipore) and then through 0.2 μm pore-size polycarbonate Isopore™ membrane filters (NB Merck-Millipore, pressure < 10 kPa). This sequential filtration was mandatory to reduce the representation of Cyanobacteria in the metagenomic and metatranscriptomic reads and access to other members of the microbial community. Although we could not exclude that some Woesearchaeota lineages were discarded by this approach, our previous study, based on 16S rRNA gene sequencing, did not reveal any Woesearchaeota OTU that were exclusively identified in 3 μm pore size filters [42], suggesting that the diversity of Woesearchaeota collected in 0.2 μm pore size filters was not altered by the prefiltration. The filters were stored at − 20 °C until nucleic acid extraction.

As reported by Sarazin et al. (2020), vertical profiles for pH, dissolved O_2, temperature, and conductivity were recorded using a WTW 3630 multiparameter system equipped with optical dissolved oxygen, conductivity, and pH probes. Salinity was calculated from conductivity and temperature. Soluble sulfide levels (ΣS(-II)), referred to as H_2S/HS^- were determined by colorimetry in the field laboratory using an Aqualytic SpectroDirect spectrophotometer and Merck reagent kits.

DNA and RNA extraction
DNA was extracted from one 0.2 μm pore-size filter per depth (seven DNA extractions), using the ZymoBIOMICS DNA Miniprep kit (Zymo Research), modified with a phenol/chloroform/isoamyl alcohol (25:24:1) step to further purify the DNA [45]. DNA quality was checked by 1% (w/v) agarose gel electrophoresis and quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. DNA extracts were stored at − 20 °C until library preparation. Metagenomic libraries for all samples were constructed and sequenced by the Fasteris company (Plan-les-Ouates, Switzerland). Sequencing (2 × 150 bp) was performed on an Illumina HiSeq 4000 system.

RNA was extracted from three 0.2 μm pore-size filters per depth using the Quick-RNA™ MiniPrep kit (Zymo Research), optimized for the hypersaline samples, and the detailed procedure is described in the "Supplementary methods" section. After the quality check, RNA extracts were pooled per sample to obtain enough RNA. The library was prepared after rRNA depletion by RiboZero Gold + Bacterial kit, using the TruSeq stranded mRNA kit (Illumina, CA, USA) and sequenced on HiSeq 4000 2 × 150 bp by the Fasteris company (Plan-les-Ouates, Switzerland).

Genomic reconstruction from metagenomes
Quality filtration of the reads was performed using BBDuk from the JGI's BBTools v.38.95 suite (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/), resulting in an average of 40,149,540 ± 9,158,963 reads per sample. Metagenomic reads passing quality control filters were then pooled and co-assembled using MetaSPAdes v.3.15.5 (-k 33, 55, 77) [46]. Only contigs > 1000 bp were kept and reads passing filters were then mapped to contigs using Bowtie2 v.2.5.2 [47] and Samtools 1.13 [48] to estimate contig coverage.

Reconstruction of MAGs (Metagenomes Assembled Genomes) was performed using MetaBAT 2 v.2.15−2 [49] with default parameters. Completeness
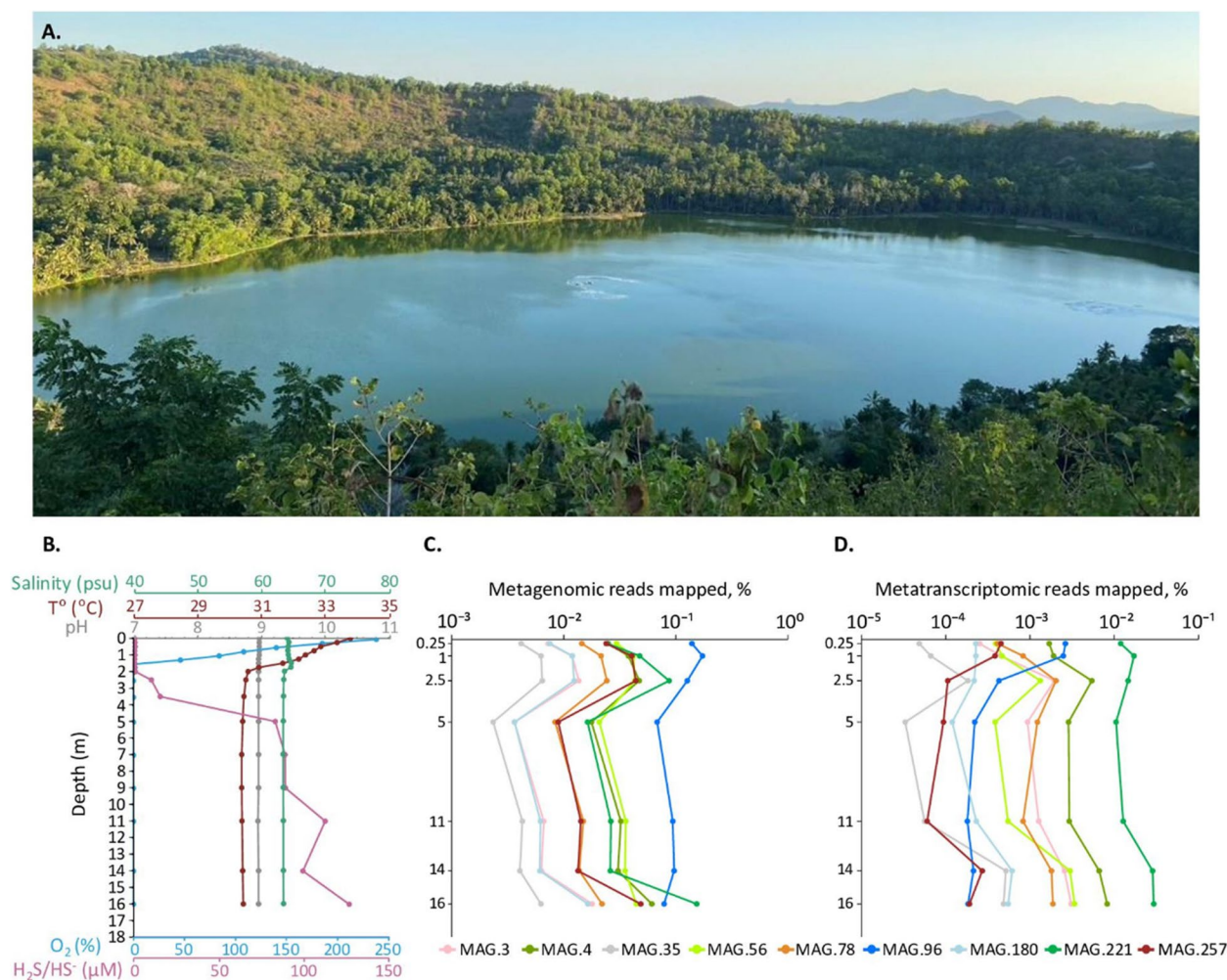
Fig. 1 Woesearchaeotal MAGs distribution across geochemical profiles of Lake Dziani Dzaha. **A** Picture of Lake Dziani Dzaha from the top of the crater (Credit: Mylène Hugoni, 2024). **B** Vertical profiles of environmental parameters recorded along the water column. Profiles of salinity (psu), temperature (°C), pH, O_2 saturation (%), and H_2S/HS^- concentration (µM) were recorded in November 2017 during the collection of microbial samples. Depth profiles of Woesearchaeota MAGs were estimated based on the percentage of mapped reads from **C** metagenomic and **D** metatranscriptomic data against woesearchaeotal MAGs recovered from this study. Scale for B and C is logarithmic

and contamination levels of the MAGs were assessed using CheckM2 1.0.1 [50]. Taxonomic assignment of MAGs was performed using GTDB-Tk v.2.1.1. [51] with the GTDB database v.207 [51]. MAGs affiliated to d__Archaea; p__Nanoarchaeota; c__Nanoarchaeia; o__Woesearchaeales by GTDB were selected from the dataset. MAG abundance was estimated by mapping reads from individual metagenomes to MAGs using Bowtie2 v.2.5.2 [47]. Average Amino Acid Identity (AAI) was calculated using the All-vs-All AAI matrix calculator tools developed at the Kostas lab [52] (Supplementary Table S1). The isoelectric point of Woeasearchaeota proteins were calculated using custom Python 3.10 and the Bio.SeqUtils package. Amino acid ratios of all predicted proteins

from Woesearchaeota MAGs were calculated using a custom Python 3.10 script available in Source data S5.

Functional annotation of MAGs was performed using default parameters of the NCBI Prokaryotic Genome Annotation Pipeline (PGAP, version 2023/10) [53], that includes more than 17 000 curated HMM profiles covering PFAM, TIGRFAM and CAZY categories and KofamScan using the KEGG database with e-value < 1.10^{-15} [54] (KO-HMM database and KO list from 2023/12/30) for a conservative approach and METABOLIC V.4 [55] for a more exploratory analysis. In addition, CAZYmes were also identified using dbCAN3 (e-value < 1.10^{-15}, coverage > 0.35) [56], and hydrogenases genes were affiliated using HydDB [57] (version June 2024) and phylogenic analysis was carried out as detailed in the "Supplementary

methods" section. ORFs labeled as pseudogenes were removed. Metabolic reconstruction of central carbon and amino acid pathways was performed using the KEGG PATHWAY database (https://www.genome.jp/kegg/pathway.html) and *MetaCyc* [58].

Identification of CRISPRs was performed using MinCED (https://github.com/ctSkennerton/minced) and default options. Spacer sequences were retrieved and searched in archaeal MAGs from Lake Dziani Dzaha with blastn [59] (options "blastn-short" and "qcov_hsp_perc 80", allowing to retrieve hits that align on at least 80% of the total length of the query sequence).

Phylogenetic analyses

A phylogenetic analysis of ribosomal proteins was performed to determine the phylogenetic position of the 15 woesearchaeotal MAGs from Lake Dziani Dzaha within the Woesearchaeota. All 636 genomes annotated as Woesearchaeota in NCBI (April 6, 2023) were downloaded (Supplementary Table S2). Genomes with < 70% completion, > 5% contamination level according to CheckM2 [50], > 150 contigs, < 10× sequencing coverage were removed from the dataset. The remaining genomes were then dereplicated using dRep v3.4.2 [60] with a 95% ANI threshold and a coverage threshold of 0.1. This allowed us to assemble a set of 297 non-redundant, high-quality, genomes of Woesearchaeota (Supplementary Table S2). We added 38 genomes from various DPANN lineages selected with the same quality criteria (Supplementary Table S3) as the outgroup for phylogenetic analysis. Altogether, the set of genomes used for phylogenetic analyses contained 350 genomes: 15 woesearchaeotal MAGs from this study, 297 woesearchaeotal genomes from public databases, and 38 additional DPANN lineages used as outgroups.

Genes coding for the 61 archaeal rprot families, were identified and extracted from the 350 genomes using the algorithm used to build the RiboDB database [61]. For each rprot family, single-copy coding genes present in each of the 350 genomes were retained for phylogenetic analysis, while multiple copies were omitted. Rprots sequences were aligned at the amino acid level using MAFFT v7.490 with the L-INSI option [62]. For each rprot family, the resulting multiple alignments were trimmed to suppress the most gapped amino acid positions and partial sequences, by removing columns (amino acid positions) and rows (genomes) containing more than 10% gaps. The multiple alignments were then combined into a supermatrix containing 7869 amino acid positions, with an overall proportion of gaps less than 12%.

ML phylogenetic trees were inferred with IQ-TREE v2.0.7. For the rprots supermatrix, we used the LG + C20 + F + R10 model, as suggested by ModelFinder (BIC criteria) [63].

Metatranscriptomic analysis

Metatranscriptomic reads were quality filtered as metagenomic sequences using BBDuk from the JGI's BBTools suite (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/), resulting in an average of 43,319, 000 ± 4,295,572 metatranscriptomic reads per sample. Retained metatranscriptomic reads were then mapped to open reading frames (ORFs) identified by PGAP in woesearchaeal MAGs using Bowtie2 v.2.5.2 [47] with a minimum end-to-end percentage of identity for qualifying reads of 98%. Mapping results were normalized by the length of each ORF and by the number of reads per metatranscriptome, then converted to transcripts per million (TPM) for comparisons.

Results

Environmental parameters

Consistent with the long-term monitoring of the lake [39], in-situ physicochemical measurements performed in November 2017 indicated that the water column has a steady high salinity (64 psu on average) and an alkaline pH ranging from 8.95 to 8.97 (Fig. 1B). Gradients of temperature, sulfides, and oxygen were also detected. A first transition zone including a thermocline, an oxycline, and a chemocline occurred between 0 and 2.5 m, shifting from 33.8 to 30.5 °C, 237.91 to 0%, and 0.63 to 10.38 µM for temperature, oxygen saturation and H_2S/HS^- concentration respectively. Below 2.5 m, temperature was stable with an average of 30.4 °C. An increase in H_2S/HS^- concentrations from 2.5 m to 14 m was reported, from 10.38 to 99.56 µM. Finally, the permanent deep chemocline previously reported in 2014 and 2015 [42] was also observed with an increase of H_2S/HS^- concentrations at 16 m (Fig. 1B). To investigate the genomic diversity of the Woesearchaeota populations, water samples were collected at seven discrete depths (0.25 m, 1 m, 2.5 m, 5 m, 11 m, 14 m, and 16 m), representing the distinct regions of geochemical redox gradients within the lake.

Genomic overview of Woesearchaeota from Lake Dziani Dzaha

After metagenomic sequencing, co-assembly of the reads from all the samples (with an average of 40×10^6 reads per sample) was performed. A total of 246 metagenome-assembled genomes (MAGs) were recovered, including 21 archaeal MAGs. Among them, 15 were assigned to the Woesearchaeales order by the GTDB-Tk pipeline, supporting the dominance of Woesearchaeota on the archaeal assemblage. These MAGs were estimated

to be 23.5 (MAG.72) to > 99% (MAG.56) complete based on CheckM2 pipeline, with an average completeness of 73.7 ± 23.4% and contamination of 0.6 ± 0.6% (Table 1). The average pairwise amino acid identity (AAI) between Woesearchaeota MAGs ranged from 37 ± 0.5 (MAG.40.165) to 43 ± 4% (MAG.221), suggesting a high degree of divergence (Supplementary Table S1). With the exception of MAG.250, R/K, and DE/IK ratios of the Woesearchaeota MAGs averaged at 0.49 ± 0.16 and 0.74 ± 0.09, respectively (Table 1). The isoelectric point distribution of Woesearchaeota proteins followed a bimodal distribution with both acidic and basic proteins with a median pI of Woesearchaeota MAGs averaged at 7.5 ± 1.0 (Supplementary Figure S1).

According to depth coverage, Woesearchaeota MAGs ranked between position 46th/246 (MAG.96) and 236th/246 (MAG.72), with estimated abundances similar to or below the median abundance of all the MAGs recovered from the dataset (21.9 reads per base) (Fig. 2A). MAG.96 had the highest coverage among the archaeal MAGs with an average of 95.2 reads per base in the contigs (Fig. 2B).

Woesearchaeota phylogeny

The phylogenetic position of Lake Dziani Dzaha MAGs within the Woesearchaeota phylum was confirmed for 14 MAGs by the analysis of the ribosomal protein set (rprots) (Fig. 3). However, MAG.40.165, which showed the lowest AAI percentages (< 0.38) with the other MAGs, was excluded from further analyses because it belongs to a lineage that branches outside of the Woesearchaeota phylum, between Pacearchaeota and Mamarchaeota (Fig. 3).

Regarding the Woesearchaeota phylum, our analyses, which doubled the number of analyzed genomes and included more stringent quality cut-offs compared to previous studies, defined well-resolved relationships with most ultrafast bootstrap values and aLRT supports > 0.95 (Fig. 3). We observed a deep split in the tree defining two major groups, referred hereafter as to group 1 and group 2, corresponding respectively to Group A and to a large cluster which encompasses groups B, C, D, E, F, G, H, I, and J (labeled on the leaves of the tree when available), from a previous meta-analysis by Huang et al. [13]. However, our analysis only partially supported the delineation of B, C, D, E, F, G, H, I, and J subgroups, indicating that further work is needed to fully resolve the phylogeny of Woesearchaeota.

Woesearchaeota populations from Lake Dziani Dzaha are highly diverse, as shown by the phylogeny of rprots, which showed that the 14 MAGs formed six distinct subgroups that branched at different locations within Group 2 (Fig. 3). Subgroups I and IV contained no

other representatives from public databases, suggesting that they may represent novel lineages specific to the Lake Dziani Dzaha ecosystem. Other MAGs clustered with sequences recovered from hypersaline meromictic lakes such as Cock Soda Lake (GCA_007117065, GCA_007128245, GCA_007116645), hypersaline soda lake brines (GCA_007117145) [33] or the hypersaline Lake El Tobar (GCA_021734865).

Metabolic potential of Woesearchaeota MAGs

To ensure the robustness of metabolic predictions, only MAGs with > 70% completeness (*n* = 9) were retained. As a first step, the functional annotation was performed with PGAP [53] and KofamScan [64]. This was followed by a more exploratory analysis using the METABOLIC pipeline [55] and dbCAN3 [56]. Combining these results, 52.7% to 59.9% of the ORFs were assigned to a potential function (Supplementary Figure S2).

Although we cannot exclude that some genes were not identified due to the incompleteness of the MAGs or because of high sequence divergence, genomic analyses indicated limited metabolic capabilities for Lake Dziani Dzaha Woesearchaeota (Fig. 4, Source data S1, S2, and S3).

No oxidative pathways or tricarboxylic acid cycle (TCA) genes were identified. All nine MAGs have the metabolic potential to degrade starch to glucose thanks to an alpha-amylase (*amy* – GH57) (Fig. 4). Genes encoding glycoside hydrolases (GH100, GH133, GH13, and GH15) were also detected in most (5/9) MAGs, indicating the potential for sucrose and glycogen degradation. However, the gene coding for the glucokinase (*glk*), which enables glucose phosphorylation, the limiting step of glycolysis, was only detected in two MAGs that were not in the same phylogenetic subgroup (MAG.78, MAG.96). Similarly, genes encoding the higher part of the glycolysis/gluconeogenesis pathway (i.e., 6-phosphofructokinase (*pfk*), class II fructose-1,6-bisphosphate aldolase (*fbaA*)) were not identified for most of the MAGs. However, the potential for conversion and phosphorylation of sugars to glyceraldehyde-3P via the non-oxidative pentose phosphate pathway was detected, linking hexose and pentose degradation to the lower part of the glycolysis and pyruvate metabolism.

Analysis of potential fermentation pathways suggested that MAG.4 and MAG.78 could ferment pyruvate to lactate using an ʟ-lactate dehydrogenase (LDH). In addition, MAG.4, MAG.78, and MAG.221 contained genes coding pyruvate ferredoxin oxidoreductase (*por*) and 2-oxoglutarate ferredoxin oxidoreductase (*kor*), which convert pyruvate to acetyl-CoA. MAG.4 and MAG.221 encoded the acetyl-CoA ligase (ADP-forming) (*acdAB*), which allows the production of acetate and ATP generation

Table 1 General genomic features of the 15 Woesearchaeota MAGs recovered from Lake Dziani Dzaha

MAG ID°	Compl. (%)	Cont. (%)	Nb. Contigs	N50	Ratio R/K	Ratio DE/IK	Depth coverage (X)	Genome size (bp)	Coding sequence (total)	GC (%)	Nb of tRNA	Presence of 16S rRNA gene	Functional annotation	NCBI ID
MAG.3	76.25	1.08	186	6400	0.51	0.80	7.0	1,033,299	1174	32	30	•	•	SAMN40749724
MAG.4	86.05	0.32	118	11,416	0.56	0.77	25.5	1,049,851	1118	34	27	•	•	SAMN40749725
MAG.35	70.94	0.62	116	5037	0.51	0.76	6.5	599,401	730	34	17		•	SAMN40749726
MAG.56	>99	0.98	70	24,409	0.61	0.77	25.5	1,115,666	1173	36	37	•	•	SAMN40749727
MAG.65	36.86	0.12	14	49,450	0.32	0.67	12.1	332,368	406	29	13			SAMN40749728
MAG.72	23.50	0.11	100	4032	0.39	0.67	5.6	416,575	478	32	13			SAMN40749729
MAG.74	52.36	2.67	184	4052	0.27	0.58	9.2	763,371	840	23	25	•		SAMN40749730
MAG.78	91.80	0.36	47	29,891	0.79	0.81	12.7	1,070,558	1132	39	39	•	•	SAMN40749731
MAG.96	99.18	0.07	14	103,126	0.80	0.98	95.2	960,688	1120	42	37	•	•	SAMN40749732
MAG.121	51.08	0.58	20	49,945	0.37	0.71	14.5	553,758	601	31	19	•		SAMN40749733
MAG.180	84.70	0.34	106	9891	0.58	0.72	8.3	847,139	990	34	33		•	SAMN40749734
MAG.221	98.10	0.36	12	232,784	0.42	0.66	25.1	1,557,840	1529	28	30	•	•	SAMN40749735
MAG.250	69.23	0.30	50	32,930	2.06	1.32	13.0	963,730	1115	57	39	•	•	SAMN40749736
MAG.257	76.54	0.25	7	167,132	0.39	0.74	25.5	817,310	940	30	29		•	SAMN40749737
MAG.40.165	90.58	0.80	56	33,820	0.43	0.74	59.1	1,092,002	1080	27	29	•		SAMN40749738

Completeness (Compl.) and contamination (Cont.) levels were estimated using CheckM2. Depth coverage was calculated by mapping the metagenomic reads against the MAGs with a similarity cutoff of 97%. Selection of the MAGs for "Functional annotation" was based on completeness and contamination levels R/K ratio arginine/lysine ratio, DE/IK ratio aspartic acid, DE/IK ratio aspartic acid *Glutamic acid/isoleucine * Lysine ratio
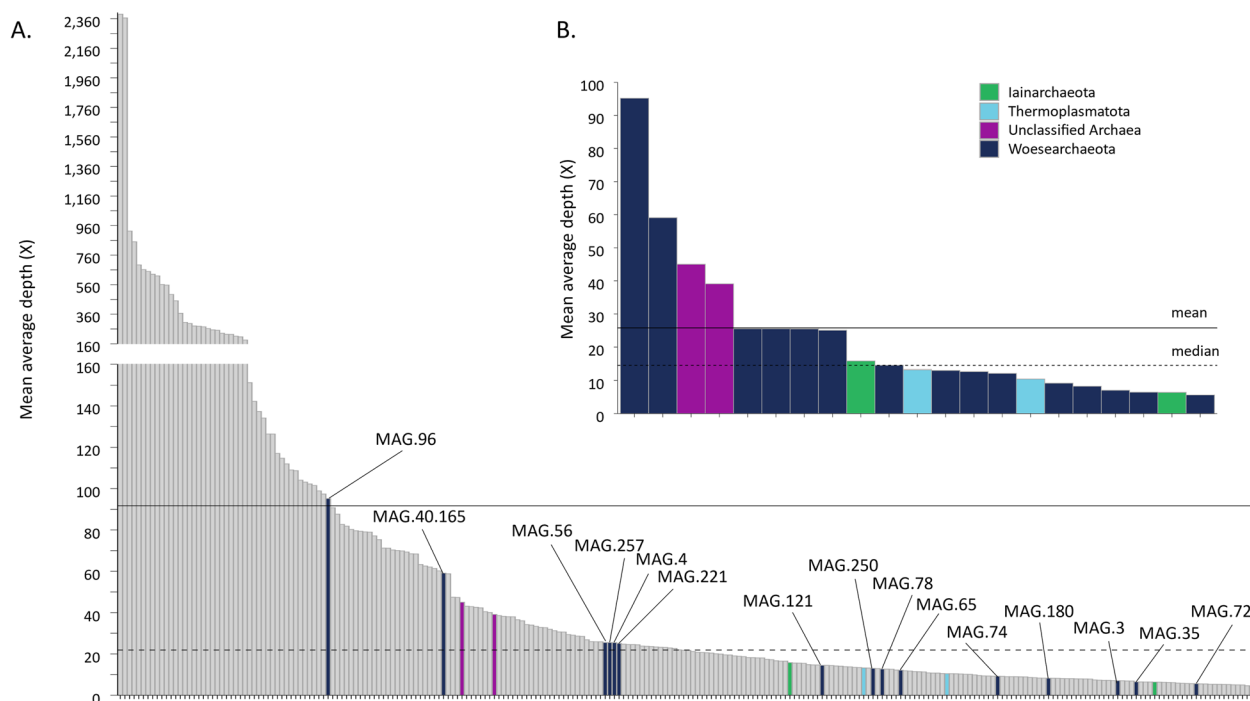
Fig. 2 Rank of Woesearchaeota MAGs. Average sequencing depth of **A** the 246 MAGs from the Lake Dziani Dzaha dataset and **B** 21 archaeal MAGs from Lake Dziani Dzaha (contamination < 5%). Solid and dashed lines represent the mean and the median of the average sequencing depth, respectively

from acetyl-CoA. This feature, usually found in fermentative archaea, may also be reversible and thus be involved in acetate utilization [65, 66]. Finally, genes encoding [FeFe] hydrogenases were detected for MAG.4, MAG.78, and MAG.221 (Fig. 4). Phylogenetic analysis of the corresponding proteins and the presence of *nuoF*-like and *nuoE*-like genes in the vicinity indicated that these hydrogenases are trimeric and belong to the electron bifurcating [FeFe] hydrogenase group A3, consistently with previous reports in fermentative microorganisms and other DPANN Archaea [67] (Supplementary Figure S3).

Pathways for the synthesis of essential amino acids (i.e., tryptophan, histidine, arginine, glutamine, cysteine, proline, alanine, valine, methionine, phenylalanine, isoleucine, tyrosine, aspartate, glutamate) were not detected. The nine MAGs encoded potential aspartate ammonialyase (*asnA*), which is used to synthesize asparagine. However, this pathway appeared to be blocked due to the absence of aspartate ammonia-lyase (*aspA*) and glutamine synthetase (*gnlA*). Similarly, five MAGs have an identified *glyA*, involved in serine synthesis, but none have the *serA/B/C* genes necessary for the downstream reactions (Fig. 4). Only MAG.96 showed a complete potential for the branched-chain amino acid transport pathway (*livK, livH, livM, livG, livF*). Genes involved in flagellar motility (*flaB, flaH, flaI, flaJ, flaK*) were detected

in most MAGs. The signal recognition particle receptor (*ftsY*) and the subunit SRP54 (*ffh*), as well as the gene involved in the secretion system (*secY*), were also identified in more than half of the Woesearchaeota MAGs. Na^+/Ca^{2+} antiporter genes were detected in MAG.3, MAG.4, MAG.35, MAG.180 and MAG.221. Gene coding a potassium transporter was also identified in MAG.221. In addition, most of the MAGs included multiple mechanosensitive channel genes (Fig. 4).

Abundance and distribution of Woesearchaeota MAGs

Mapping of the metagenomic reads against the nine retained MAGs showed that these Woesearchaeota represented a total of ~0.3% of the reads at all sampling depths. As expected from its coverage, MAG.96 was the most abundant MAG, representing an average of 0.11% of mapped reads across the water column, while other MAGs represented 0.0023 to 0.1535% of mapped reads (Fig. 1C). The nine MAGs showed similar depth distribution profiles with two peaks of increased abundance in the water column. The first peak was located near the surface at 1 m (MAG.96) or 2.5 m (the other eight MAGs) while the second peak was found at 14 m (MAG.96) or 16 m (the other eight MAGs) (Fig. 1C).

Two hypotheses could explain these increased abundances in the upper and the lower transition zones: (1)
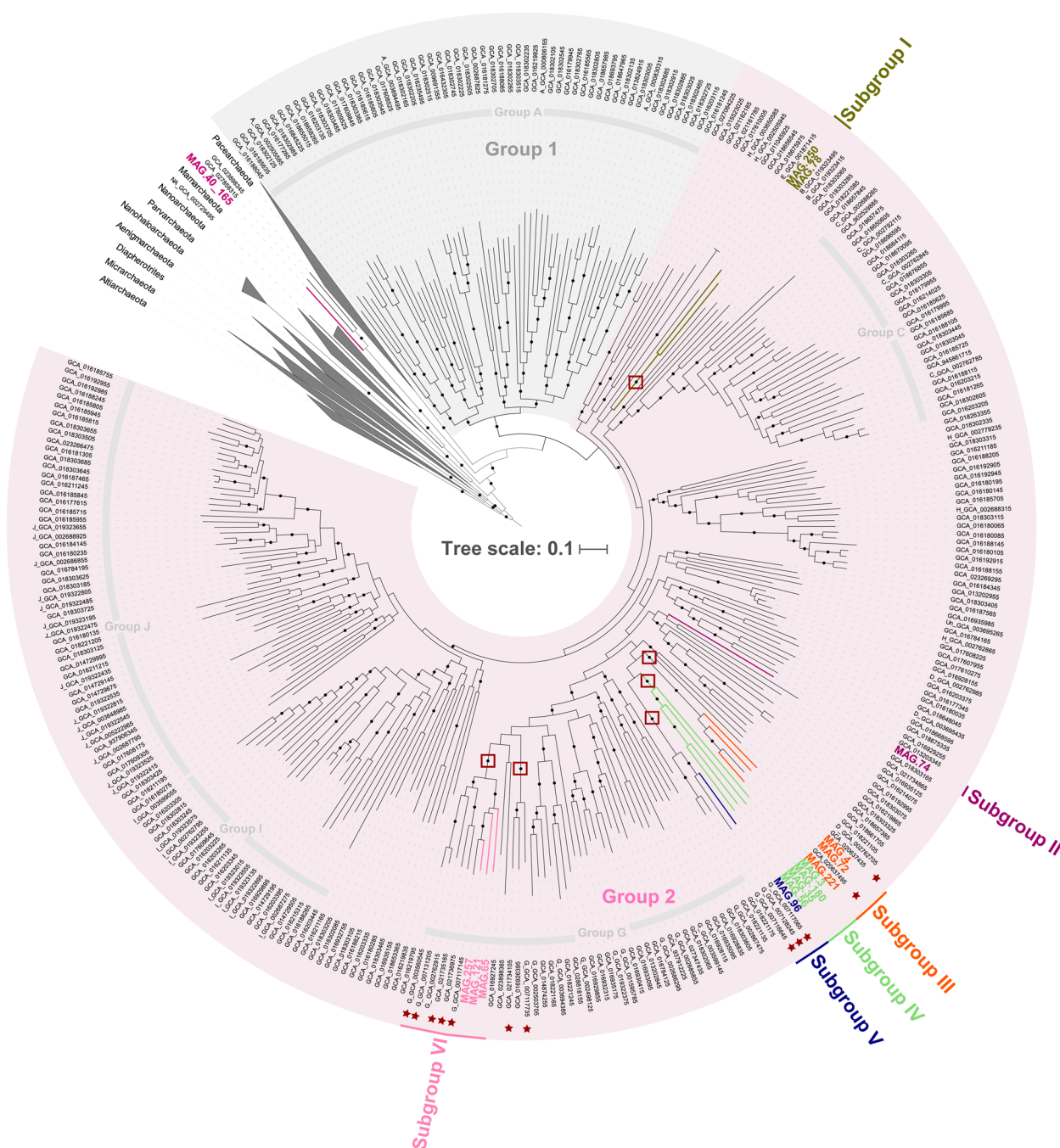
Fig. 3 Maximum likelihood phylogeny of the Woesearchaeota phylum. The tree was inferred using a large supermatrice gathering the sequences of 61 rprots (350 sequences, 7869 amino acid positions). The tree was computed with IQ-TREE with the LG + C20 + F + R10 model. The tree was rooted using representative genomes from other DPANN lineages. Branch robustness was evaluated using ultrafast bootstrap and aLRT supports computed with IQ-TREE. For clarity, only branches with both supports > 0.95 are highlighted (black circles). The scale bar corresponds to evolutionary distance (i.e., the average number of substitutions inferred per site). The clusters (A-J) defined by Huang et al. (2021) are labeled along the GCA code of the sequence and indicated as grey arcs in the tree. Outer colored lines and labels indicate subgroups delineated in this study. Red squares correspond to well-supported (bootstrap and aLRT > 0.95) clades of putative halophilic Woesearchaeota (red stars)

the 1 m/2.5 m peak could correspond to live/active Woesearchaeota, whereas the 14 m/16 m peak could be associated with decaying and dead/inactive cells, or (2)

Woesearchaeota could be active in these two contrasted ecological niches in the water column, possibly relying on different, niche-specific metabolisms.
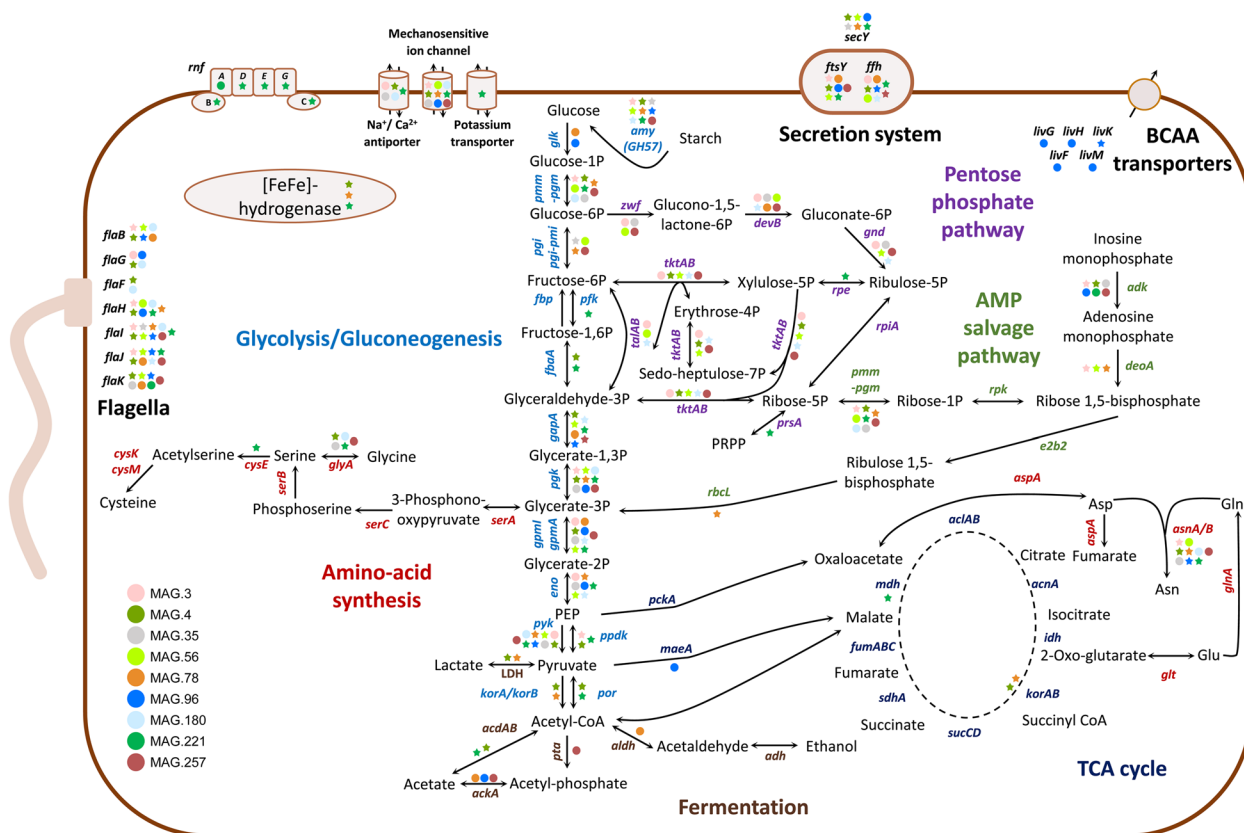
Fig. 4 Schematic overview of metabolic pathways identified in Woesearchaeota MAGs from Lake Dziani Dzaha. Presence of the genes in the nine retained MAGs (contamination < 5%, completion > 70%) is represented by colored circles. Stars indicate that corresponding transcripts for the gene were identified in at least one depth. Functional genes encoding the relevant proteins/enzymes are labeled for each metabolic step. Abbreviations: amy , amylase; glk , glucokinase; pmm-pgm , phosphomannomutase/phosphoglucomutase; pgi , glucose-6-phosphate isomerase; pgi-pmi , bifunctional phosphoglucose-phosphomannose isomerase; fbp , fructose-1,6-bisphosphatase I; pfk , 6-phosphofructokinase; fbaA , class II fructose-1,6-bisphosphate aldolase; gapA , type I glyceraldehyde-3-phosphate dehydrogenase; pgk , phosphoglycerate kinase; gpml , 2,3-bisphosphoglycerate-independent phosphoglycerate mutase; gpmA , 2,3-diphosphoglycerate-dependent phosphoglycerate mutase; eno , enolase; pyk , pyruvate kinase; ppdk , pyruvate, phosphate dikinase; korA , 2-oxoacid:acceptor oxidoreductase subunit alpha; korB , thiamine pyrophosphate-dependent enzyme; por , pyruvate:ferredoxin (flavodoxin) oxidoreductase; LDH , L-lactate dehydrogenase; acdAB , acetate–CoA ligase family protein; ackA , acetate kinase; pta , phosphate acetyltransferase; aldh , aldehyde dehydrogenase family protein; adh , alcohol dehydrogenase; serA , D-3-phosphoglycerate dehydrogenase/2-oxoglutarate reductase; serC , phosphoserine aminotransferase; serB , phosphoserine phosphatase; glyA , glycine hydroxymethyltransferase; cysE , serine O-acetyltransferase EpsC; cysK / cysM , cysteine synthase; talAB , transaldolase family protein; tktAB , transketolase; zwf , glucose-6-phosphate dehydrogenase; devB , 6-phosphogluconolactonase; rpe , ribulose-phosphate 3-epimerase; gnd , decarboxylating 6-phosphogluconate dehydrogenase; prsA , ribose-phosphate diphosphokinase; rpiA , ribose 5-phosphate isomerase A; pckA , phosphoenolpyruvate carboxykinase (ATP); maeA , NADP-dependent malic enzyme; mdh , malate dehydrogenase; fumABC , fumarate hydratase, class I; sdhA , succinate dehydrogenase flavoprotein subunit; sucC / sucD , succinyl-CoA synthetase; idh , isocitrate dehydrogenase; acnA , aconitate hydratase; aclA / aclB , ATP-citrate lyase; aspA , aspartate ammonia-lyase; asnA , aspartate–ammonia ligase; asnB , asparagine synthase; glnA , glutamine synthetase; glt , glutamate synthase (NADPH); flaB , archaeal flagellin FlaB; flaF , archaeal flagellar protein FlaF; flaG , archaeal flagellar protein FlaG; flaH , archaeal flagellin FlaH; flaI , archaeal flagellin FlaI; flaJ , archaeal flagellin FlaJ; flaK , archaeal preflagellin peptidase FlaK; ffh , signal recognition particle subunit SRP54; ftsY , fused signal recognition particle receptor; secY , preprotein translocase subunit SecY; livG / livF , branched-chain amino acid transport system ATP-binding protein; livM / livH , branched-chain amino acid transport system permease protein; livK , branched-chain amino acid transport system substrate-binding protein; rpk , Alpha-D-ribose-1-phosphate 5-kinase; e2b2 , eucaryal translation initiation factor 2B; deoA , AMP/thymidine phosphorylase; rbcL , type III ribulose-bisphosphate carboxylase; adk , adenylate kinase; rnf , rnfABCDE, electron transport complex Rnf

Overall gene expression profiles of Woesearchaeota MAGs
To test the two hypotheses associated with the double peak distribution, metatranscriptomic reads sequenced from the seven water depths were mapped against the ORFs of the nine Woesearchaeota MAGs (Fig. 1D). This approach revealed transcriptomic activity at all depths for all Woesearchaeota populations, indicating that they were likely alive and active throughout the entire water

column, or at least maintained a basal transcriptomic activity. The percentage of transcribed genes per depth varied between 12.98% and 71.74% depending on the considered MAGs. While MAG.96 was identified as the most abundant archaeal MAG in the metagenomic dataset, MAG.96 genes were not the most highly expressed. Expression levels of MAG.96 were higher at 0.25 m and 1 m, and then decreased with depth (Fig. 1D). In contrast, the transcription patterns of MAGs 3, 4, 35, 56, 78, 180, 221 as well as MAG.257 showed a double peak of activity at 1–2.5 and at 14–16 m, mirroring their depth distribution detected by metagenomics (Fig. 1C). Interestingly, the two most active MAGs, i.e., MAG.221 and MAG.4 belonged to the same phylogenetic subgroup III.

Analysis of the top 10 woesearchaeotal genes expressed at each depth in this study revealed several genes involved in archaellin type IV proteins (*flaB*), membrane formation (S-layer protein genes, YceI family protein), and cellular machinery (ATP synthase, TRAM domain-containing protein, ribosomal proteins, translation elongation factor EF-1, stress response translation inhibitor YciH and cold shock proteins), secretion system (*secY*) and cell division (protein FtsZ) (Supplementary Figure S4). Only a limited activity of amino acid synthesis was detected. For example, MAG.3, MAG.4, MAG.78, MAG.96, and MAG.221 expressed an asparagine synthase, while MAG.4 and MAG.221, expressed a serine/glycine hydroxymethyltransferase (Fig. 4). Transcripts of the non-oxidative pentose phosphate pathway and glycolysis genes were also identified in the dataset, as well as transcripts from the lactate dehydrogenase (LDH) of MAG.4 and MAG.78 and type III RubisCo from MAG.78 (Fig. 4). Transcripts of the genes involved in pyruvate fermentation to acetate were also identified for MAG.4

and MAG.221. MAG.4, MAG.78, and MAG.221 also expressed [FeFe] hydrogenase genes (Fig. 4).

Depth profiles of Woesearchaeota activity

To test whether the double peak of Woesearchaeota activity corresponds to a modulation of their metabolism to acclimatize to environmental conditions, gene expression patterns were compared between top (i.e., 0.25/2.5 m) and bottom (i.e., 14/16 m) samples. This comparison revealed contrasting patterns depending on the lineages (Fig. 5).

Transcript analysis identified five distinct transcriptomic patterns along the water column (Fig. 5). The first pattern concerns MAG.4, and MAG.221, both belonging to the phylogenetic subgroups III (Fig. 3). In these MAGs, the majority (48 and 58% for MAG.4, and MAG.221, respectively) of the genes, including [FeFe] hydrogenases, type V ATP synthases, glycoside hydrolases, cell division, and membrane protein genes were expressed at both top (1 m/2.5 m) and bottom (14/16 m) layers of the water column (Fig. 5). A complete rnfABCDGE electron transport complex was also expressed at both depths for MAG.221.

A relatively large overlap between gene expression patterns was also observed for MAG.3, 56, and 78, including archaellin and hydrogenases, but additional sets of genes were also expressed at specific depths. For instance, the lactate dehydrogenase gene of MAG.78 was expressed only at the bottom of the lake, while the gene encoding glycoside hydrolases (GH100 and GH57) and type III RuBisCo involved AMP recycling were only expressed at the top of the lake.

The third pattern was observed for MAG.180 and 35, which are phylogenetically close within subgroup IV. It is characterized by a very low number of genes expressed
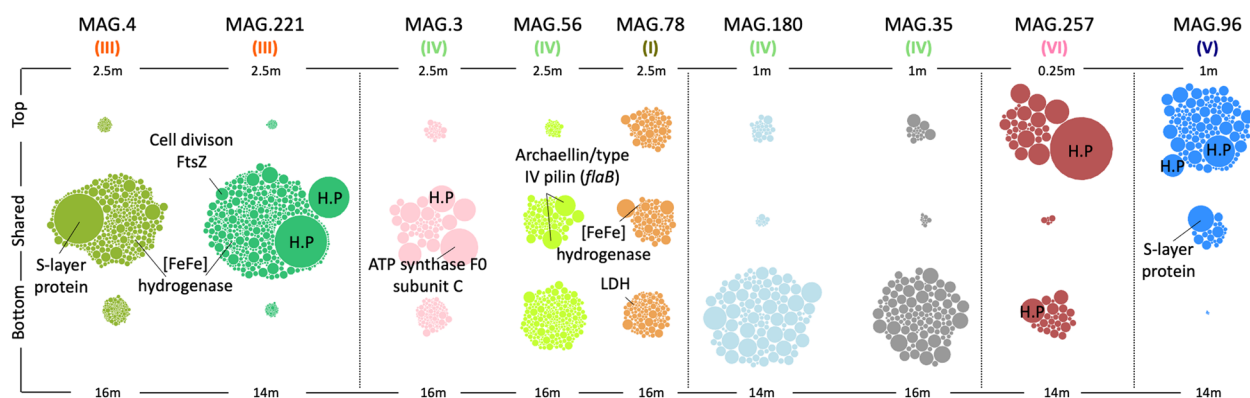

Fig. 5 Transcript partitioning of Woesearchaeota. Each dot represents a transcribed ORF that has been detected only at the top, bottom, or both (shared) water layers. The upper and lower depths corresponding to the peaks of activity are labeled for each MAG. For each MAG, the size of the dots was calculated based on the number of transcripts per ORF and then normalized by the number of transcripts per depth to compare top, shared, and bottom expressions. The number below each MAG indicates the subgroup to which it belongs. LDH, L-lactate dehydrogenase; H.P, Hypothetical protein

at both depths and a higher number of genes expressed at the bottom. This includes genes encoding GH15, phosphoenolpyruvate synthase, and glycogen synthase for both MAGs and chemotaxis proteins CheABCD for MAG.180.

A bimodal pattern with opposite expression profiles at the top and bottom peaks was detected for MAG.257, which was taxonomically related to subgroup VI (Fig. 5). However, the number of transcripts was too low and overly represented by hypothetical proteins to infer the metabolic activities. Finally, MAG.96 which was the most abundant MAG in the metagenomic dataset was mainly active at 1 m where it expressed archaellin, type V ATP synthase, transporters, and pyruvate metabolism genes.

Discussion

Woesearchaeota have been previously identified in numerous natural environments [11–13] and are taxonomically diverse, with at least 26 proposed subgroups based on 16S rRNA gene meta-analysis [10] and 10 subgroups based on a set of 38 orthologous proteins [13]. However, our larger-scale analysis of 61 ribosomal proteins, partially supported this previous taxonomic delineation of Woesearchaeota subgroups, highlighting that the known diversity of Woesearchaeota is far from being complete and that further work is needed to better understand their taxonomic arrangements and evolutionary history. Based on the taxonomic placements of MAGs recovered from Lake Dziani Dzaha and other high salinity environments (Laguna Del Tobar, Soda lakes), halophily appears to be paraphyletic within the Woesearchaeota phylum, with at least six cohesive clades of putative halophilic Woesearchaeota that emerged from three distinct evolutionary events. Although additional experiments are required to fully characterize the adaptative strategies of halophilic Woesearchaeota, no bias of amino acid composition or isoelectric point of proteins was identified, indicating that a salt-in strategy is unlikely [68]. By contrast, a strong expression level of the serine/glycine hydroxymethyltransferase was observed in the most active populations (MAG.221 and MAG.4), suggesting a salinity tolerance through osmoprotectant accumulation [69]. Transcripts of mechanosensitive channels were also detected (Fig. 4). These membrane proteins could potentially be involved in Woesearchaeota salinity resistance, as observed for other Archaea [70].

With six identified clusters, including two subgroups (I and IV) that could potentially be endemic, Lake Dziani Dzaha hosts an elevated genomic diversity for Woesearchaeota as observed in other stratified aquatic ecosystems [33, 34, 71]. This genomic diversity, possibly induced by their unique biology [72], may allow the

co-occurrence of multiple lineages in the same ecological niche. Consistently, mapping of the metagenomic data against Woesearchaeota MAGs revealed similar depth distributions for these phylogenetically distinct lineages, regardless of their relative abundance (dominant/rare) in the system. However, the unusual double peak of abundance that we detected at two different depths in Lake Dziani Dzaha challenged previous investigations of the vertical distribution of Woesearchaeota in aquatic ecosystems, which suggested a single niche preference along the vertical stratification of the water column [73]. Mapping of the MAGs against historical datasets from 2014 and 2015 showed similar patterns (Supplementary Figure S5) and metatranscriptomic data also confirmed the double peak of abundance, with an increased woesearchaeotal transcriptomic activity at both 1–2.5 m and 14–16 m depths, providing rare evidence of in-situ activity of Woesearchaeota.

Based on the geochemical profiles of the lake, these two depths correspond to the oxycline and the permanent deep chemocline that characterize the lake's geochemistry [42]. Transition zones are places of high chemical reactivity in lakes and hotspots of microbial diversity and activity due to overlapping distributions of electron donors and acceptors and intense interactions of different biotic components [74]. Since Woesearchaeota genomes do not appear to have the metabolic capacity to directly use the chemical species present in Lake Dziani Dzaha (i.e., oxygen and sulfur compounds, Fig. 4), we hypothesized that their double peak of abundance and activity could be linked to the increased biomass and diversity found in these layers [75].

At these depths, Woesearchaeota could benefit from multiple interactions with other microorganisms that would complement their metabolic gaps. Analysis of CRISPR-Cas signatures for the detection of a putative partner [29] was inconclusive, suggesting a non-CRISPR-Cas-mediated interaction. Woesearchaeota from Lake Dziani Dzaha expressed high levels of archaellin type IV (*fla*) proteins, which are also found in several DPANN members such as Nanoarchaeota, Woesearchaeota and Diapherotrites [3, 10, 76, 77]. In DPANN, archaellin genes may be involved in motility [78], but also in host attachment and interaction, as recently suggested for the co-cultured '*Ca. Nanoclepta minutus*' with its host, *Zestosphaera tikiterensis* [79]. These highly active genes are present in the different woesearchaeotal MAGs from the Lake Dziani Dzaha suggesting that they have a crucial function, potentially enhancing cell-to-cell contacts and interactions with other microorganisms in these transition zones.

In addition, metagenomic and metatranscriptomic data revealed starch, glycogen, and sugar degradation

and diverse lineage-dependent fermentations, with lactate and acetate as end products, providing transcriptomic evidence for a heterotrophic fermentative activity. Active hydrogen metabolism, via [FeFe] hydrogenase was also identified in some MAGs, supporting previous metabolic predictions of Woesearchaeota from other ecosystems [5, 10, 13]. Lactate, acetate, and hydrogen were found to be major intermediates in interactive metabolism and syntrophic relationships [80, 81], supporting microbial interactions between Woesearchaeota and other community members of the transition zones.

While presenting the same apparent double peak of abundance and activity in the water column, contrasted gene expression strategies were detected for Woesearchaeota populations. These include (i) identical gene expression at both surface and bottom water layers (MAGs 4, 221), (ii) constitutive expression of housekeeping genes but modulation of their activity related to carbon and energy metabolism (MAGs 3, 56, and 78), or (iii) extremely contrasted activities (MAGs 180, 35 and 257). These results show that beyond their observed metabolic versatility, Woesearchaeota populations could also be characterized by a complex and lineage-dependent tuning of their activity, challenging the extrapolation of their ecological role and lifestyle based on genomic data. The first pattern (i.e., identical expression) suggests that lineages clustered within subgroups III and IV are insensitive to local environmental conditions such as oxygen and sulfur concentrations and/or the potential state of their putative host, allowing them to avoid niche competition. In contrast, other patterns could indicate a gradient of transcriptomic responses from local acclimatation to changing conditions or different adapted lifestyles (i.e., associated vs free-living or active vs dormant) with differing gene expression requirements. Alternatively, this fine-tuning of activities could also reflect an adjustment to their potential host physiology that might differ between oxic and sulfidic zones as observed for instance in Cyanobacteria [82]. This reveals rapid modulation of their in-situ activities, different ecological and/or energy-saving tactics, and potential niche acclimatation for Woesearchaeota ecotypes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01956-0.

Supplementary Material 1.

Supplementary methods.

Supplementary references.

Supplementary Figure S1. Distribution of the Isoelectric point (pI) of the predicted proteins in Woesearchaeota MAGs from Lake Dziani Dzaha. To calculate pI distribution, pI values were rounded to one digit.

Supplementary Figure S2. Proportion of total ORFs assigned as hypothetical or annotated proteins for each of the nine MAGs of Woesearchaeota from Lake Dziani Dzaha (contamination <5%, completion >70%). Percentage of ORFs annotated as "annotated proteins" is labeled for each MAGs.

Supplementary Figure S3. Maximum likelihood phylogeny of the [FeFe] hydrogenase genes. The tree was inferred using 1384 [FeFe] hydrogenases sequences. The tree was computed with IQ-TREE, model LG+C20+R+F. The tree was rooted using [FeFe] hydrogenases affiliated to subgroups B and C. Branch robustness was evaluated using ultrafast bootstrap and aLRT supports computed with IQ-TREE. Only branches with both supports >0.95 are highlighted (black circles). The scale bar corresponds to evolutionary distance (i.e., the average number of substitutions inferred per site). Subgroups from A to G are colored within the corresponding branches in the tree. Uncolored branches identify sequences that were not previously analyzed. Phylum levels of identified Archaea sequences are colored at the corresponding labels in the tree. Source data are available in Source data 4.

Supplementary Figure S4. Transcriptomic activity of the nine woesearchaeal MAGs from Lake Dziani Dzaha for A. 0.25m, B. 1m, C. 2.5m, D. 5m, E. 11m, F. 14m and G. 16m. Each circle represents a transcript, whose size depends on the level of activity (based on the output of the jgi_summarize_bam_contig_depths from MetaBAT 2). The ten most active transcripts for each depth were numbered and the PGAP annotation filled in. (n = total number of active transcripts for the nine MAGs).

Supplementary Figure S5. Woesearchaeota distribution in historically stratified periods in 2014 and 2015. Depth profiles of Woesearchaeota populations were estimated based on the percentage of mapped reads from metagenomes produced from 2014 and 2015 samples. Scale is logarithmic. According to Hugoni et al. (2018) [9], the seasonal chemoclines are represented in blue, and the deep chemoclines in pink.

Supplementary Table S1. Average Amino Acid Identity (AAI) between the 15 Woesearchaeota MAGs from Lake Dziani Dzaha.

Supplementary Table S2. Genomic characteristics of the 636 Woesearchaeota MAGs retrieved from NCBI database. Woesearchaeota MAGs that reach our threshold (contamination <5%, completion >70%, numbers of contigs ≤150, sequencing depth >10X) were specified. In 136 cases the sequencing depth was not available. N.A. was then filled in. The last column indicated subgroups of MAGs used in Huang et al. (2021).

Supplementary Table S3. Genomic characteristics of the 38 DPANN MAGs retrieved from NCBI database (contamination <5%, completion >70%, numbers of contigs ≤150, sequencing depth >10X). In 3 cases the sequencing depth was not available, N.A. was then filled in.

Source data S1: Presence/absence of genes involved in metabolic pathways of Woesearchaeota, with their respective KO number and gene name according to KEGG and PGAP. Activity for each predicted ORF was calculated for each sampling depth and normalized by reads per kilo base per million mapped reads (TPM).

Source data S2: Transcriptomic activity for each predicted ORF in quality MAGs (n=9). Calculation was conducted for the different sampling depths and normalized by reads per kilo base per million mapped reads (TPM).

Source data S3: Summary of metabolic annotations of MAGs of Woesearchaeota from Lake Dziani Dzaha using PGAP, KOFAMscan, CAZy, METABOLIC and HydDB.

Source data S4: [FeFe]-hydrogenases sequences used in Supplementary Figure 2. Source data S5: Custom Python 3.10 script for inferring the ratio of amino acids in the Woesearchaeota MAGs from Lake Dziani Dzaha.

Authors' contributions
M.H., M.B., P.M.O., A.V., designed research. M.H., L.A.C., M.B., T.B., C.O.D. performed research. M.H., L.A.C., T.B., C.B.A., JP.F., A.Z., P.M.O. analyze data. M.H., L.A.C, T.B., C.B.A., JP.F. prepared the figures. L.A.C., A.V., T.B., C.B.A., JP.F., M.H, wrote the manuscript. M.B., C.L., V.G., C.B., M.T., H.A., M.A, C.O.D., P.M.O. revised the manuscript.

Data availability
Availability of data and materials Raw metagenomic and metatranscriptomic reads as well as sequences of the 15 Woesearchaeota MAGs have been deposited in the NCBI public database under Bioproject PRJNA1037317. Source data (annotation Tables) are available at: https://doi.org/10.6084/m9.figshare.26212661.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]UMR5240 Microbiologie Adaptation Et Pathogénie, Université, INSA Lyon, CNRS, Claude Bernard Lyon 1, Villeurbanne 69621, France. [2]Université de Lyon, UMR Mycoplasmoses Animales, VetAgro Sup, AnsesMarcy L'Etoile 69280, France. [3]MARBEC, Univ Montpellier, IRD, CNRS, Sète, Ifremer, France. [4]UMR 5276, Laboratoire de Géologie de Lyon: Terre, Univ Lyon, UCBL, CNRS, Environnement (LGL-TPE), PlanètesVilleurbanne 69622, France. [5]Laboratoire de Biométrie Et Biologie Évolutive, UMR5558, Université Claude Bernard Lyon 1, CNRS, VetAgro Sup, Villeurbanne, France. [6]Institut Universitaire de France (IUF), Paris, France. [7]UMR 7245 Molécules de Communication Et Adaptations Des Microorganismes (MCAM) MNHN-CNRS, Muséum National d'Histoire Naturelle, CP 39, 12 Rue Buffon, Paris F-75231, France. [8]UMR 7266, LIENSs, La Rochelle Université-CNRS, 2 Rue Olympe de Gouges, La Rochelle 17000, France. [9]Institut de Physique du Globe de Paris, Université de Paris, Paris, France. [10]Université, PRABI-ASMB, Claude Bernard Lyon 1, Villeurbanne R3728 BIOEENVIS, France. [11]Present address: Mediterranean Institute of Oceanography (MIO), Aix Marseille Univ-CNRS, Marseille, France.

References
1. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013;499:431–7.
2. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain bacteria. Nature. 2015;523:208–11.
3. Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol. 2015;25:690–701.
4. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.
5. Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell. 2018;172:1181–97.
6. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol Lett. 2019;366: fnz008.
7. Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, et al. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. Nat Commun. 2020;11:3939.
8. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J. 2017;11:2407–25.
9. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, 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.
10. Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, et al. Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. Microbiome. 2018;6:102.
11. Liu X, Wang Y, Gu J-D. Ecological distribution and potential roles of Woesearchaeota in anaerobic biogeochemical cycling unveiled by genomic analysis. Comput Struct Biotechnol J. 2021;19:794–800.
12. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data. 2018;5: 170203.
13. Huang W-C, Liu Y, Zhang X, Zhang C-J, Zou D, Zheng S, et al. Comparative genomic analysis reveals metabolic flexibility of Woesearchaeota. Nat Commun. 2021;12:5281.
14. Johnson MD, Shepherd DC, Sakai HD, Mudaliyar M, Pandurangan AP, Short FL, et al. Cell-to-cell interactions revealed by cryo-tomography of a DPANN co-culture system. Nat Commun. 2024;15:7066.
15. Probst AJ, Moissl-Eichinger C. "Altiarchaeales": uncultivated archaea from the subsurface. Life Basel Switz. 2015;5:1381–95.
16. Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature. 2002;417:63–7.
17. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, et al. The genome of Nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci. 2003;100:12984–8.
18. Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, et al. Enigmatic, ultrasmall, uncultivated Archaea. Proc Natl Acad Sci. 2010;107:8806–11.
19. Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, et al. Lineages of acidophilic Archaea revealed by community genomic analysis. Science. 2006;314:1933–5.
20. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 2018;16:629–45.
21. Chen L-X, Méndez-García C, Dombrowski N, Servín-Garciadueñas LE, Eloe-Fadrosh EA, Fang B-Z, et al. Metabolic versatility of small archaea micrarchaeota and parvarchaeota. ISME J. 2018;12:756–75.
22. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun. 2015;6:6372.

23. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, et al. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat Microbiol. 2021;6:354–65.

24. Golyshina OV, Toshchakov SV, Makarova KS, Gavrilov SN, Korzhenkov AA, La Cono V, et al. "ARMAN" archaea depend on association with euryarchaeal host in culture and in situ. Nat Commun. 2017;8:60.

25. Krause S, Bremges A, Münch PC, McHardy AC, Gescher J. Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. Sci Rep. 2017;7:3289.

26. Hamm JN, Erdmann S, Eloe-Fadrosh EA, Angeloni A, Zhong L, Brownlee C, et al. Unexpected host dependency of Antarctic Nanohaloarchaeota. Proc Natl Acad Sci. 2019;116:14661–70.

27. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol. 2018;3:328–36.

28. Schwank K, Bornemann TLV, Dombrowski N, Spang A, Banfield JF, Probst AJ. An archaeal symbiont-host association from the deep terrestrial subsurface. ISME J. 2019;13:2135–9.

29. Esser SP, Rahlff J, Zhao W, Predl M, Plewka J, Sures K, et al. A predicted CRISPR-mediated symbiosis between uncultivated archaea. Nat Microbiol. 2023;8:1619–33.

30. Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites.' ISME J. 2015;9:447–60.

31. Bird JT, Baker BJ, Probst AJ, Podar M, Lloyd KG. Culture independent genomic comparisons reveal environmental adaptations for Altiarchaeales. Front Microbiol. 2016;7:1221.

32. Rinke C, Chuvochina M, Mussig AJ, Chaumeil PA, Davín AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. A standardized archaeal taxonomy for the Genome Taxonomy Database. Nat Microbiol. 2021;6(7):946–59. https://doi.org/10.1038/s41564-021-00918-8.

33. Vavourakis CD, Mehrshad M, Balkema C, van Hall R, Andrei A-Ş, Ghai R, et al. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. BMC Biol. 2019;17:69.

34. Vigneron A, Cruaud P, Culley AI, Couture R-M, Lovejoy C, Vincent WF. Genomic evidence for sulfur intermediates as new biogeochemical hubs in a model aquatic microbial ecosystem. Microbiome. 2021;9:46.

35. Vigneron A, Cruaud P, Guyoneaud R, Goñi-Urriza M. Into the darkness of the microbial dark matter in situ activities through expression profiles of Patescibacteria populations. Front Microbiol. 2022;13:1073483.

36. Bruto M, Oger PM, Got P, Bernard C, Melayah D, Cloarec LA, et al. Phytoplanktonic species in the haloalkaline Lake Dziani Dzaha select their archaeal microbiome. Mol Ecol. 2023;32:6824–38.

37. Beam JP, Becraft ED, Brown JM, Schulz F, Jarett JK, Bezuidt O, et al. Ancestral absence of electron transport chains in patescibacteria and DPANN. Front Microbiol. 2020;11: 1848.

38. Cadeau P, Jézéquel D, Leboulanger C, Fouilland E, Le Floc'h E, Chaduteau C, et al. Carbon isotope evidence for large methane emissions to the Proterozoic atmosphere. Sci Rep. 2020;10:18186.

39. Sarazin G, Jézéquel D, Leboulanger C, Fouilland E, Floc'h EL, Bouvy M, et al. Geochemistry of an endorheic thalassohaline ecosystem: the Dziani Dzaha crater lake (Mayotte Archipelago, Indian Ocean). Comptes Rendus Géoscience. 2020;352:559–77.

40. Sala D, Grossi V, Agogué H, Leboulanger C, Jézéquel D, Sarazin G, et al. Influence of aphotic haloclines and euxinia on organic biomarkers and microbial communities in a thalassohaline and alkaline volcanic crater lake. Geobiology. 2022;20:292–309.

41. Leboulanger C, Agogué H, Bernard C, Bouvy M, Carré C, Cellamare M, et al. Microbial diversity and cyanobacterial production in Dziani Dzaha crater lake, a unique tropical thalassohaline environment. PLoS ONE. 2017;12: e0168879.

42. Hugoni M, Escalas A, Bernard C, Nicolas S, Jézéquel D, Vazzoler F, 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.

43. Cellamare M, Duval C, Drelin Y, Djediat C, Touibi N, Agogué H, Leboulanger C, Ader M, Bernard C. Characterization of phototrophic microorganisms and description of new cyanobacteria isolated from the salinealkaline crater-lake Dziani Dzaha (Mayotte, Indian Ocean). FEMS Microbiol Ecol. 2018;94(8). https://doi.org/10.1093/femsec/fiy108.

44. Escalas A, Troussellier M, Melayah D, Bruto M, Nicolas S, Bernard C, 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. 2022;97: fiab163.

45. Melayah D, Bontemps Z, Bruto M, Nguyen A, Oger P, Hugoni M. Metabarcoding of the three domains of life in aquatic saline ecosystems. Methods Mol Biol Clifton NJ. 2023;2605:17–35.

46. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

48. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10: giab008.

49. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7: e7359.

50. 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.

51. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315–6.

52. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Inc.; 2016.

53. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44:6614–24.

54. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

55. Zhou Z, Pq Tran, Am Breister, Liu Y, Kieft K, Es Cowley, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome. 2022;10:33.

56. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51:W115–21.

57. Søndergaard D, Pedersen CNS, Greening C. HydDB: A web tool for hydrogenase classification and analysis. Sci Rep. 2016;6: 34212.

58. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res. 2020;48:D445–53.

59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

60. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.

61. Jauffrit F, Penel S, Delmotte S, Rey C, de Vienne DM, Gouy M, et al. RiboDB database: a comprehensive resource for prokaryotic systematics. Mol Biol Evol. 2016;33:2170–2.

62. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160–6.

63. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587.

64. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.

65. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun. 2013;4:2120.

66. Barnhart EP, McClure MA, Johnson K, Cleveland S, Hunt KA, Fields MW. Potential role of acetyl-CoA synthetase (*acs*) and malate dehydrogenase (*mae*) in the evolution of the acetate switch in Bacteria and Archaea. Sci Rep. 2015;5: 12498.

67. Greening C, Cabotaje PR, Valentin Alvarado LE, Leung PM, Land H, Rodrigues-Oliveira T, et al. Minimal and hybrid hydrogenases are active from archaea. Cell. 2024;187:3357–3372.e19.

68. Oren A. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. Front Microbiol. 2013;4:4.

69. Waditee-Sirisattha R, Sittipol D, Tanaka Y, Takabe T. Overexpression of serine hydroxymethyltransferase from halotolerant cyanobacterium in *Escherichia coli* results in increased accumulation of choline precursors and enhanced salinity tolerance. FEMS Microbiol Lett. 2012;333:46–53.

70. Echeveste Medrano MJ, Leu AO, Pabst M, Lin Y, McIlroy SJ, Tyson GW, et al. Osmoregulation in freshwater anaerobic methane-oxidizing archaea under salt stress. ISME J. 2024;18:wrae137.

71. Patin NV, Dietrich ZA, Stancil A, Quinan M, Beckler JS, Hall ER, et al. Gulf of Mexico blue hole harbors high levels of novel microbial lineages. ISME J. 2021;15:2206–32.

72. Paul BG, Burstein D, Castelle CJ, Handa S, Arambula D, Czornyj E, et al. Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. Nat Microbiol. 2017;2:1–7.

73. Ortiz-Alvarez R, Casamayor EO. High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. Environ Microbiol Rep. 2016;8:210–7.

74. Ferro I, Morrone JJ. Biogeographical transition zones: a search for conceptual synthesis. Biol J Linn Soc. 2014;113:1–12.

75. Bernard C, Escalas A, Villeriot N, Agogué H, Hugoni M, Duval C, et al. Very low phytoplankton diversity in a propical saline-alkaline lake, with co-dominance of *Arthrospira fusiformis* (Cyanobacteria) and *Picocystis salinarum* (Chlorophyta). Microb Ecol. 2019;78:603–17.

76. Jarett JK, Nayfach S, Podar M, Inskeep W, Ivanova NN, Munson-McGee J, et al. Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. Microbiome. 2018;6:161.

77. St John E, Flores GE, Meneghin J, Reysenbach A-L. Deep-sea hydrothermal vent metagenome-assembled genomes provide insight into the phylum Nanoarchaeota. Environ Microbiol Rep. 2019;11:262–70.

78. Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, et al. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. Nat Commun. 2016;7: 12115.

79. St. John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, et al. A new symbiotic nanoarchaeote (Candidatus Nanoclepta minutus) and its host (Zestosphaera tikiterensis gen. nov., sp. nov.) from a New Zealand hot spring. Syst Appl Microbiol. 2019;42:94–106.

80. McDaniel EA, Scarborough M, Mulat DG, Lin X, Sampara PS, Olson HM, et al. Diverse electron carriers drive syntrophic interactions in an enriched anaerobic acetate-oxidizing consortium. ISME J. 2023;17:2326–39.

81. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. Mol Syst Biol. 2007;3:92.

82. Stal LJ. Physiological ecology of cyanobacteria in microbial mats and other communities. New Phytol. 1995;131:1–32.

Publisher's Note