Title:

Functional insights of novel Bathyarchaeia reveal metabolic versatility in their role in Peatlands of the Peruvian Amazon

Authors: Michael J. Pavia[1,2,3], Arkadiy I. Garber[1,3], Sarah Avalle[1], Franco Macedo-Tafur[4], Rodil Tello-Espinoza[4,5], Hinsby Cadillo-Quiroz[1,2,3*]

[1]School of Life Sciences, Arizona State University, Tempe, AZ, United States

[2]Swette Center for Environmental Biotechnology, Biodesign Institute, Arizona State University, Tempe, AZ, United States

[3]Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, AZ, United States

[4] Laboratory of Soil Research, Research Institute of Amazonia's Natural Resources, National University of the Peruvian Amazon, Iquitos, Loreto, Peru

[5] School of Forestry, National University of the Peruvian Amazon, Iquitos, Loreto, Peru

*Corresponding Author: Hinsby Cadillo-Quiroz, hinsby@asu.edu

Author Contributions:

MJP – data curation, data analyses, data visualization, and validation, writing original and editing draft; AIG – data analysis, review & editing draft; SC - data analysis, review & editing draft; FMT – field and molecular data collection, review & editing draft, RTE - field and molecular data collection, review & editing draft; HCQ – conceptualization, data collection and curation, funding acquisition, project supervision, writing, review & editing draft.

Funding:

This material is based upon work supported by the National Science Foundation under Grant No. CAREER-1749252 to HC-Q, and work (proposal: 10.46936/10.25585/60000849) conducted by

the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Abstract:

The decomposition of soil organic carbon within tropical peatlands is influenced by the functional composition of the microbial community. In this study, building upon our previous work, we recovered a total of 28 metagenome-assembled genomes (MAGs) classified as Bathyarchaeia from the tropical peatlands of the Pastaza-Marañón Foreland Basin (PMFB) in the Amazon. Using phylogenomic analyses, we identified nine genus-level clades to have representatives from the PMFB, with four forming a putative novel family ("*Candidatus* Paludivitaceae"*)* endemic to peatlands. We focus on the *Ca.* Paludivitaceae MAGs due to the novelty of this group and the limited understanding of their role within tropical peatlands. Functional analysis of these MAGs reveals that this putative family comprises facultative anaerobes, possessing the genetic potential for oxygen, sulfide, or nitrogen oxidation. This metabolic versatility can be coupled to the fermentation of acetoin, propanol, or proline. The other clades, outside *Ca.* Paludivitaceae are putatively capable of acetogenesis, *de novo* amino acid biosynthesis, and encode a high amount of Fe^{3+} transporters. Crucially, the *Ca.* Paludivitaceae are predicted to be carboxydotrophic, capable of utilizing CO for energy generation or biomass production. Through this metabolism, they could detoxify the environment from CO, a byproduct of methanogenesis, or produce methanogenic substrates like CO_2 and H_2. Overall, our results show the complex metabolism and various lineages of Bathyarchaeia within tropical peatlands pointing to the need to further evaluate their role in these ecosystems.

Importance**:**

With the expansion of the *Candidatus* Paludivitaceae family by the assembly of 28 new metagenome assembled genomes, this study provides novel insights into their metabolic

diversity and ecological significance in peatland ecosystems. From a comprehensive phylogenic and functional analysis, we have elucidated their putative unique facultative anaerobic capabilities and CO detoxification potential. This research highlights their crucial role in carbon cycling and greenhouse gas regulation. These findings are essential for resolving the microbial processes affecting peat soil stability, offering new perspectives on the ecological roles of previously underexplored and underrepresented archaeal populations.

58

59 **Introduction:**

60 Peatlands, a type of wetland, are characterized by their large carbon storage capacity, in
61 which the rates of primary production exceed decomposition rates (1). The slow process of
62 organic matter (OM) decomposition in peatlands is a consequence of several limiting factors
63 which include continuous water-logged soils leading to anoxic conditions (2), lower
64 temperatures compared to surrounding soils from low conductive properties (3), acidification
65 from biological metabolism (4), lower concentrations of nutrients from low input or depletion
66 (5), and accumulation of phenolic compounds due to inhibition of phenol oxidases in the absence
67 of oxygen (6).

68 The Pastaza-Marañón Foreland Basin (PMFB) is home to a large expanse of tropical
69 peatlands (~67,000 km^2) in the Amazon and is estimated to contain 3.1 Pg of stored soil carbon
70 (7, 8). In addition to functioning as a terrestrial carbon reservoir, the PMFB emits 3.16 to 41.1 Tg
71 of CH_4 per year (9). Understanding the roles of distinct microbial populations in the
72 transformation of stored organic carbon in these peatlands is essential to addressing the projected
73 shift of this region from a carbon sink to a carbon source as a result of climate change (10, 11).
74 Archaea constitute a significant proportion of the microbial community in the soils of the PMFB;
75 most notably the Bathyarchaeia, which range in relative abundance from 3 to 8% in the shallow
76 soils (≤ 20 cm below surface) and up to 20% of the community in the deeper soils (12, 13).
77 Previously classified as the 'Miscellaneous Crenarchaeotal Group', their presence has been
78 detected in multiple terrestrial and aquatic environments, making them a potentially important
79 contributor to sedimentary carbon cycling (14, 15). Despite their potential importance in the

carbon cycle, we currently lack axenic cultured representatives, hindering our ability to understand their physiology and functional role within the environment.

Metagenomic approaches have provided valuable information on the environmental distribution and metabolic diversity of Bathyarchaeia (16–20). Metabolic reconstructions predict a broad range of possible metabolisms within members of the Bathyarchaeia including, but not limited to, anaerobic methane oxidation (21), acetogenesis (22), phototrophy (16, 23), heterotrophy based on methylated compounds (19, 24), aromatic compounds (25), or detrital proteins (17, 26, 27). Furthermore, recovery of metagenome-assembled genomes (MAGs) classified as Bathyarchaeia from the Surat Basin (Eastern Australia) detected the potential for methylotrophic methanogenesis; yet, these MAGs contain few methanogenic marker genes and mostly 'Mcr-like' genes (24, 28). Cultivation-centered studies enriching for Bathyarchaeia have demonstrated their growth on various complex carbon substrates (29–32), as well as the production of acetate from guaiacol (33) so far only growing in mixed cultures. In addition to functioning in the cycling of carbon, the Bathyarchaeia are predicted to play a role in sulfur and nitrogen cycling, as well as vitamin B12 production (16).

While understanding the distribution and potential environmental drivers of the Bathyarchaeia is important, it remains unclear what metabolisms or functions they may carry out within tropical peatlands, such as the PMFB. Thus, to investigate the putative phylogenetic variation and metabolic potential of Bathyarchaeia in the PMFB region, we evaluated MAGs from shallow and deep peat soil layers from one nutrient-rich and three oligotrophic peatlands with distinct vegetation dominance respectively: mixed forest (Buena Vista), pole forest (San Jorge), palm swamp (Quistococha), and open (Maquía) peatland (34, 35). Using Bathyarchaeia MAGs from the PMFB plus nearly all publicly available high-quality MAGs, we conducted

phylogenomic, pangenomic, and functional analyses to understand (1) the phylogenetic and metagenomic-assembled gene content distribution of PMFB Bathyarchaeia within this diverse phylum and (2) predict the putative metabolisms unique to the PMFB Bathyarchaeia.

Methods:

Study sites, Metagenomes, and MAG assembly

We evaluated MAGs from metagenomes of four peatlands, previously described (12, 35, 36) within the PMFB: Buena Vista (BVA), San Jorge (SJO), Quistococha (QUI), and Maquía (MAQ). The general characteristics of each site are detailed in Table S1. Earlier sampling and metagenome sequencing for shallow soils in BVA, QUI, and SJO are described in our earlier report (37). Briefly, soils of 0-10 and 10-20 cm were collected between July-October 2015 and January-February 2016, and extracted DNA was sequenced with Illumina Hi Seq 2500 2 x 151 bp technologies at the Joint Genome Institute (JGI). To expand our studies, deep soil samples were collected at 60-100 cm deep in August 2016 for QUI, SJO, and August 2017 for MAQ, transported and stored frozen until DNA extraction using the previously reported method (37). Library preparation and Illumina sequencing were conducted under the same approach at JGI as above, all part of the JGI's Community Sequencing Program (proposal: doi:/10.46936/10.25585/60000849) and detailed in Supplementary File 1.

For MAG assemblies, metagenomic reads were decontaminated, trimmed, and quality score filtered following the Joint Genome Institute IMG protocol (38). Quality control metagenomes were assembled using MEGAHIT (v1.1.3) (39) using the default settings and quality was assessed with QUAST (v3) (40). MAQ metagenomes were co-assembled by transect, while SJO and QUI were a single assembly due to low DNA and low sequencing yield of

replicates. Contigs were quality-controlled as described elsewhere (37) and binned using MetaBAT2 (v2.12.1) (41) with a minimum cut-off length of 2000bp. The resulting bins were dereplicated and curated using Anvi'o (v6.1) (42). MAGs, >50% completeness and <10% redundant as determined by CheckM (43) were classified using GTDB-tk (44) against the Genome Taxonomy Database release 07-RS207. The relative abundance of MAGs was calculated by mapping metagenomes back to MAGs using bowtie2 (45), and the JGI script provided in MetaBAT2 (41) was used to calculate the total contig coverage; total coverage was then normalized by MAG size. MAG analysis code is available at https://github.com/Hinsby/BathyarchaeaMAGs2023.

Phylogenetic Inference and Average Amino Acid Identity (AAI) Comparisons

To determine the distribution of PMFB Bathyarchaeia within the class, we performed phylogenetic inference using an in-group of 238 Bathyarchaeia MAGs, which included both those from this study and publicly available ones (Supplementary File 2). Prior to phylogenetic analysis, Bathyarchaeia MAGs were filtered for bias based on genome size (>1Mb), redundancy (<10%), completeness (>50%), and AAI scores greater than 95% from MAGs sampled from the same environment. In instances where more than one MAG from the same environment had an AAI greater than 95%, we considered them replicates and selected the MAG with higher completeness and lower contamination. The outgroup consisted of 37 publicly available genomes representing both Thermoproteota and Euryarchaeota organisms. Anvi'o (v6.1) (46) was used to identify, concatenate, and align 54, single copy genes (Table S3) that were found in at least 82% (191) of MAGs (as determined from the mean completeness score for just Bathyarchaeia MAGs). The alignment was used for Maximum likelihood inference supported by bootstrap-resampling 300 times using RAxML (v8.2.12)(47) with the PROTCATBLOSUM62 model of

amino acid substitution. AAI scores, calculated using github.com/mooreryan/aai, were clustered based on ≥65% similarity, and clusters were assigned an operational taxonomic rank at the genus level, previously proposed by Konstantinidis et al. (48). In this study, we have denoted each putative genus-level cluster as a Bathyarchaeia clade (BC). A chi-squared test was used to detect if there was a significant association between BCs and ecosystem type. GC% and predicted genome size were assessed for significance to BC's with Kruskal-Wallis and Tukey's multiple comparison and were used on BCs that contained more than four MAGs. All differences were considered significant at a P-value <0.05 and visualized with ggstatsplot (49).

Meta-pangenome and Functional Enrichment Analysis of Bathyarchaeia

The Aniv'o pangenomic workflow (46) was used to identify gene clusters within BCs that contained MAGs from the PMFB (9 BCs in total). Briefly, this workflow calculates similarities across all open reading frames and removes weak hits with a chosen minbit heuristic score of 0.6, and uses the MCL algorithm for gene cluster identification, partial genes were excluded (50). Completion and contamination statistics were used for operational cutoffs to designate gene clusters as relaxed-core, shell, cloud, and singletons (Table S4). Gene clusters were annotated using eggNOG (v5) (51) and the nr database *(accessed August 2021)* (52) using a consensus sequence built from the alignments of each gene cluster. Briefly, the consensus sequence was built on the frequency of amino acids present at each position, and in positions where there was a tie, a residue was chosen at random.

Results and Discussion:

Microbial community composition from assemblies of new MAGs

New assemblies were completed for deeper soils' metagenomes from QUI and SJO, and shallow soil metagenome from MAQ, followed by genome binning resulting in a total of 122 high-quality (HQ) and medium-quality (MQ) MAGs (53) (Figure S1). Acidobacteriae and Nitrososphaeria were highly represented in MAGs recovered across all three sites. Moreover, we recovered fourteen novel MAGs belonging to the Bathyarchaeia and three HQ MAGs classified as Lokiarchaeia (Figure S1). Consistent with previous studies utilizing 16S rRNA gene amplicons (12, 13), Bathyarchaeia were found to be abundant in deep peat, accounting for ~15% of the community. This study extends these findings by recovering HQ and MQ MAGs, providing a more detailed genomic representation of the Bathyarchaeia populations within peatlands from QUI, SJO, and MAQ from the PMFB (12, 13).

Taxonomic placement of PMFB Bathyarchaeia

Bathyarchaeia MAGs from the PMFB (28 in total, seven from the new assemblies passing inclusion criteria) were evaluated in combination with 210 other MAGs from published studies with a median completeness of 81.9% and redundancy 3.3% (detailed in Supplementary File 2). Phylogenomic clades within the Bathyarchaeia were inferred from the robustness of tree topology based on a 54 single-copy gene (SCG) maximum likelihood (ML) phylogenomic tree in combination with AAI between MAGs (Figure 1 and Figure S2). Consensus between the phylogenomic tree and AAI suggests there are 60 distinct clades (operational taxonomic units at the genus level) within our Bathyarchaeia dataset. This is consistent with a recent report on the phylogenetic diversity of Bathyarchaeia (Hou et al., 2023). We note the clade equivalents with Hou et al., 2023 but expand the analysis by advancing predictions for novel Bathyarchaeia from Amazon peatlands. Among the Bathyarchaeia clades (BCs), nine contained MAGs from the

PMFB, with three comprised exclusively of PMFB MAGs (BC15, BC39, and BC42). Found within BC1 and BC3 (also noted as Baizomonadales), are Bathyarchaeia MAGs from both the PMFB and other environments such as hot springs, termite guts, and permafrost. BC36 (also noted as Houtuarculales), is primarily comprised of MAGs from grasslands, with only one from the PMFB. While the PMFB MAGs span the diversity of Bathyarchaeia several shallow branches (BC38 – BC42), within the noted Houtuarculales, were comprised primarily of our tropical peatland populations. BC15, also noted as Baizomonadales, and BC39 and BC42, located within the Houtuarculales group, have not been previously reported thus representing novel lineages of Bathyarchaeia inhabit the soils of the PMFB.

The distribution of MAGs recovered by environment and the sequence composition among BC's exhibited non-random distribution by source and clade-specific genomic signatures (Figure 2 and Figure S3). A Chi-squared test showed that among the 16 BCs with more than four representative MAGs, 13 BCs are significantly associated with one ecosystem type (Figure 2). The BCs that show the strongest association with an ecosystem were BC18 (also noted as Baizomonadales), BC5 (also noted as Baizomonadales), and BC47 (also noted as Houtuarculales) in hot springs; BC19 (also noted as Baizomonadales), in hydrothermal vents; BC29, BC8, and BC12 (all in the group also noted as Baizomonadales) in marine zones; BC36 (also noted as Houtuarculales) in grasslands; and BC38 (also noted as Houtuarculales) in peatlands. Other BCs (particularly BC1, BC11, and BC24, all in the group also noted Baizomonadales) are less significant and show a widespread distribution across multiple environments. Previous studies have suggested that Bathyarchaeia are non-randomly distributed along salinity and oxygen concentrations (54, 55), which may be applicable for BC8 and BC24 which were recovered from estuarine environments. Other studies have also indicated that total

organic carbon influences lineage distribution (56, 57). In line with these observations, BC11 is primarily comprised of MAGs recovered from estuarine sediment ecosystems including an enrichment culture on lignin.

Specialization in genomic characteristics, such as genome size and GC content, are consistent within archaeal species (58, 59). To assess whether the environment is a predictor of genomic signatures in Bathyarchaeia, we performed a Kruskal-Wallis test on clades containing more than four representative MAGS (Figure S3). Minimal significant differences were observed in both predicted genome size and GC content between MAGs recovered from different ecosystems, with many being indiscernible from each other. The weak relationship between archaeal genome size and phylogeny has been observed in other groups and may be influenced by the complexity of archaeal genomes (60). Alternatively, we found that GC content can be a strong indicator for BC. Many basal BCs have a GC content within 40-50%, whereas shallower branching clades (BC25, BC27, BC29, and BC36) have a significantly higher GC%, ranging from 50-60%. These four clades were also found to be significantly associated with specific ecosystems and may have undergone selective pressures linked to chronic energy stress (61). These clades are primarily recovered from marine sediment, littoral marine zones, and grasslands which are characteristic of being anoxic with low primary production and are energy-starved ecosystems. Commonly found in hot springs, BC47 has the lowest GC content, which to our knowledge differs from what is a common trait of thermophilic archaea (62, 63). Distribution of BCs by ecosystem and significant differences in GC content between BCs suggest divergent evolutionary trajectories across clades and potential for distinct functions in the Bathyarchaeia.

Metabolic comparison and putative differences between PMFB clades (and associated MAGs)

To investigate the relationship between phylogenetic distance and functional potential within BC clades containing PMFB MAGs, gene clusters from a meta-pangenomic analysis were annotated using a combination of eggNOG (v5) (51) and the nr database (52). We observed a weak correlation between phylogenetic distance and functional dissimilarities between the nine BCs tested (Mantel statistic r: 0.03198 P-value < 0.05). Below we detail common metabolic pathways and distinct functions between BC1, BC3, and BC38-41, which include multiple PMFB MAGs (singleton BCs, BC15, and BC42, were excluded from this step), and additionally recognize BC36 as a distinct cluster with one PMFB MAG but comprised primarily of other environments. We build upon and refine previous metabolic models of these groups (64) with specific emphasis on metabolism found within the PMFB.

Carbon Metabolism

Within the evaluated clades most MAGs possess a partial Embden-Meyerhof-Parnas (EMP) glycolysis pathway but lack hexokinase (*glk*) which is used for the initial phosphorylation of glucose (Supplementary File 3). Loh et al. (19) has suggested that due to the incomplete EMP pathway and the presence of phosphoenolpyruvate synthetase (*pps*) and fructose-1,6-bisphosphate aldolase (*fbaAB*) (FBP), these genes function exclusively in gluconeogenesis. Phylogenetic analysis shows that most MAGs within BC1, BC3, BC36, and BC38-41 have a class 2 FBP, while some from BC1, BC3, and one from BC40 have a class 1 FBP (Figure S4A). Class 1 FBP activity can be induced in *Escherichia coli* when grown on gluconeogenic carbon substrates, while class 2 is constitutively expressed and indicative of a primary use with glycolytic substrates (65). The presence of a class 1 FBP in only some populations of BC1, BC3, and BC40 suggests that a gluconeogenic function is less common in *Bathyarchaeia.* Sugar-

phosphate utilization may be possible in class 2 FBP-containing populations from BC1, BC3, BC36, and BC38-41, although sugar transporters could not be identified. While different isoforms of FBP are found within *Bathyarchaeia* experimental evidence is required to identify if sugar utilization is possible.

BC36 lacks phosphofructokinase (*pfk*) but encodes the full glyoxylate pathway; an alternative route for sugar-phosphate utilization in Bathyarchaeia (15). Findings in Haloarchaea (66, 67) suggest that the glyoxylate pathway and incomplete EMP function in neither anaplerotic nor gluconeogenic pathways, but rather used for growth on acetate. Carbohydrate utilization varies greatly across BCs, with BC38-41 putatively using EMP for gluconeogenic purposes.

Alternative to carbohydrate utilization many MAGs from BC38-41 possess the genetic potential for acetoin degradation (Figure 3). This is evidenced by the presence of genes *acoL,* *acoC,* and *acoAB*, which facilitate the conversion of acetoin into acetaldehyde and acetyl-CoA. Furthermore, in 65% of BC38-41 MAGs, the oxidation of propanol to acetaldehyde might be possible via *adhP,* an alcohol dehydrogenase with preference for propanol. The resulting acetaldehyde, formed either from acetoin or propanol, could undergo further conversion to acetate via an aldehyde dehydrogenase present in MAGs from BC38 and BC40. The likelihood of alcohol fermentation in BC38-41 is further supported by the presence of an acetyl-CoA synthetase (*acs*). We note that acetate formation from acetyl-CoA is a common feature in most MAGs from the seven BCs.

Carbon Fixation

Autotrophic lifestyles appear common among the Bathyarchaeia. The archaeal Wood-Ljungdahl (WL) pathway was found in most MAGs from BC1 and BC3 (Supplementary File 3).

286 However, previous reports have suggested that the absence of the *fmdE* subunit in the
287 Bathyarchaeia Formyl-MFR dehydrogenase complex is characteristic of this group (19), yet our
288 findings suggest this to be the case only for BC1 because in BC3 a homolog of *fmdE* was
289 detected. The lack of Methylene-H_4MPT reductase (*mer*) in 13% of BC1 or BC3 MAGs supports
290 proposals (19) that deeper branching Bathyarchaeia have lost the capacity to reduce CO_2 all the
291 way to the methyl level. The carbonyl branch of the WL pathway is carried out by the CO
292 dehydrogenase complex (*cdhABCE)* which condenses CO with a methyl group and CoA to form
293 acetyl-CoA. The *cdhABCE* complex was found in almost all MAGs from BC1 and BC3.

294 In comparison to the CO dehydrogenase complex present in BC1 and BC3, all but one
295 MAG from BC38-41 have a putative aerobic molybdenum containing CO dehydrogenase,
296 *coxMSL* capable of producing CO_2 and H_2 (Figure 3). A phylogenetic analysis of the large
297 subunit (*coxL)* indicates that BC38-40, clusters closely with other thermophilic CO-oxidizing
298 archaea (68, 69) (Figure S5). We note that we could not identify a *coxL* homolog in MAGs from
299 BC41, but they did have the other two subunits *coxM* and *coxS.* High concentrations of CO,
300 approximately 4mM, have been shown to inhibit methanogenesis (70). These concentrations are
301 not uncommon in natural environments (71) and suggest a potential for BC38-41 populations to
302 alleviate CO stress and provide metabolites to peatland methanogens. As a byproduct of
303 methanogenesis CO could be detoxified by BC38-41 and in turn produce compounds CO_2 and
304 H_2 that could be used to fuel methanogen metabolism.

305 Alternatively, CO oxidation can be coupled to CO_2 fixation (72) through the Calvin cycle
306 via ribulose-1,5-bisphosphate carboxylase (rbcL) which was present in clades BC38, BC39, and
307 BC40. However, only two MAGs have the genetic potential for the near full cycle, lacking only
308 phosphoribulokinase (*prkB*). *Archaeoglobus fulgidus* grows on CO with formate as an

intermediate (73), via a novel type of formate dehydrogenase (FDH); yet only two out of 13 *BC38-41* MAGs have FDH homologs. Instead, BC36 and BC38-41 have both methylene-H_4F reductase (*metF*) and methenyltetrahydrafrofolate cyclohydrogenasese (*folD*), involved in the bacterial WL pathway, which suggests the potential for formate utilization. While no formate transporters were found, it is plausible that *BC38-41* can convert formate to 5,10-methylene-THF and subsequently feed into the glycine cleavage system.

Oxygen, Nitrogen and Sulfur metabolism

BC38-41 are facultative anaerobes inhabiting water-logged peatland ecosystems. The low-affinity O_2 cytochrome aa3 oxidase was present in most of the MAGs in BC38-41, with four having all subunits (Figure 3). This is consistent with our findings that *BC38-41* can putatively fix CO aerobically. Bathyarchaeia has only been suggested to tolerate aerobic conditions (74), but our findings suggest that BC38-41 may prefer to occupy oxygen-rich niches, which is in agreement with the recent analysis from Hou et. al., 2023.

Existence in anaerobic soil by BC38-41 is also possible through alternative chemotrophic mechanisms. MAGs from BC36 and BC38 encode a flavocytochrome sulfide dehydrogenase (FCSD) and BC40 harbors a type 3 sulfide:quinone oxidoreductase (SQR) (Figure S4B). This suggests these two clades in BC38-41 can potentially utilize H_2S as an electron donor facilitating sulfide-dependent respiration (75). Bathyarchaeia have been found in sulfur rich environments (76, 77), and they may be playing a direct role in sulfur cycling in PMFB peatlands. N reduction, via *nosZ* or *nirK*, was identified in many MAGs from BC36 and BC38 but only in one MAG in BC39 and two in BC40. Direct fixation of N was only found in three MAGs from BC1 and BC41 and is likely a sparse function. Contrary to the report of Pan et. al. 2020, a significant contribution of NH_4 by Bathyarchaeia is not a common characteristic. The respiratory potential

to occupy either aerobic or anaerobic soil conditions is common in BC38-41 MAGs. These adaptations are well suited for changes in O_2 availability brought on by the seasonality of flood-driven tropical peatlands (2).

Amino acids and Transport

Amino acids have been suggested as a primary carbon source for the Bathyarchaeia (17). However, this appears unlikely for BC1 which has mostly complete *de novo* biosynthesis pathways for 11 amino acids (Figure 3). Additionally, BC1 has a low gene copy number of amino acid transporters relative to the other BCs (Figure 4). Peptide/Ni^+ and amino acid transporters were observed to be two-fold more abundant in *BC38-41* compared to the other clades (Figure 4, Supplementary File 3). MAGs from BC3, BC36, and BC38-41 have the genetic potential for histidine to glutamate conversion; however, in most BC38-41 we find gene copies for glutamate dehydrogenase allowing for the conversion of glutamate to oxaloacetate. The gene copy number for *aspB* (involved in the interconversion between aspartate and oxaloacetate) was found two-fold higher in BC38-41. In addition, BC38-41 also harbors *iaaA*, which catalyzes the conversion of asparagine to aspartate. BC38-41 encodes two cytosolic peptidases (*pip* and *pepP*) with the capacity to cleave proline from imported peptides. The liberated proline may be converted to glutamate, as evidenced by the presence of both *putB* and *rocA*. Collectively, this suggests that BC38-41 are adapted for growth on peptides, for either biosynthetic requirements and/or gluconeogenic purposes, with oxaloacetate serving as a key intermediate metabolite in their metabolism. The PMFB Bathyarchaeia BC38-41 harbor genomic evidence for the utilization of proline. Proline accumulation in plants is a common response to abiotic stressors such as extreme heat and drought (78). Projections of increased drought frequency and heat severity in the PMFB (11), suggest that local vegetation may accumulate more proline in

response to these conditions. When these stressed plants die and decompose, soil proline levels will increase providing a selective advantage and favor the growth of BC38-41 populations.

The high-affinity phosphate transport system (*ptsABCS*) was universally present in all BC's, indicating an adaptation for low phosphate environments. BC1 has many transports for various inorganic ions, with a high prevalence of Fe^{3+} transporters (Figure 4). In summary, a diverse array of nutrient acquisition and metabolic capabilities are exhibited among BCs, with BC38-41 potentially relying on the presence of amino acids within their niche.

Methanogenesis

We recovered many genes that are accessory to methanogenesis, such as *hrdABCD*, *acdAB*, *mchADG*, and *mtrH* in all clades excluding BC36 and BC39 (Supplementary File 3). However, no homologs for mcrA or 'MCR-like' genes were detected in any of the MAGs from this study. These findings suggest that a functional methanogenic pathway in the Bathyarchaeia is unlikely in tropical peatlands (19, 26, 79).

Metapangenomic analysis of PMFB Bathyarchaeia clades

To extend our understanding of the functional partitioning between tropical peatlands' Bathyarchaeia we conducted a metapangenomic analysis on the nine BCs containing MAGs from the PMFB. Anvi'o identified 50,379 gene clusters across all BCs, with 19,338 found in at least one MAG from PMFB (Figure 5). Due to the fragmented nature of MAGs and the evolutionary distance between BCs in analysis, no gene clusters were found in all PMFB MAGs. Only one gene cluster was found in 82 MAGs (88%) and another in 25 MAGS (89%) from the PMFB. Then under this status, the metapangenome of all MAGs from the PMFB represents an open pangenome whose analysis can mainly provide general trends (Figure S6).

The Bathyarchaeia metapangenome identified a distinct distribution of gene clusters within the relaxed-core of PMFB BCs (Figure 5, with Table S3 detailing each cluster membership and annotation). Gene clusters were highly conserved within BC1 and BC3 MAGs, while BC38-BC41 showed a comparable high prevalence of conserved genes within their cluster evidencing a noticeable gene pool separation of these clades. BC15 and BC42, which are represented by a single MAG from the PMFB share only a third of their gene clusters (not including singletons), 41% and 28% respectively, with at least one other PMFB MAG. More than half of the gene clusters in the relaxed-core and shell of BC36 were exclusive to this clade pointing to its uniqueness although deeper sampling can better test this observation. The relaxed-core gene clusters for all BCs accounted for 0.3%-21% of gene clusters found in each of the nine clades. The small percentage of gene clusters categorized as relaxed-core is analogous to previously reported thresholds at the class level (80). However, given our chosen completeness cut-off of >50%, we assume that many gene clusters categorized as shell are likely core genes misclassified because of the effects of assembly or binning errors. This assumption would bring the BC relaxed-core gene clusters to thresholds that are consistent with those reported for pan-genomes at the genus level (46, 81–83).

Proposal of BC38-41 as Candidatus Paludivitaceae

Multiple lines of evidence in this study show that clades BC38-41 represent a distinct and cohesive group of Bathyarchaeia with unique genomic and metabolic features that seem specifically adapted to peatland environments. Several characteristics set them apart from other Bathyarchaeia and justify the proposal of a putative novel family within the phylum abundant in Amazon peatlands (Bathyarchaeia has an average read mapping frequency of 4.6%, QUI:6.9%,

SJO: 7.1% in BVA, QUI and SJO metagenomes respectively). Genomic data for BC38-41 show that AAI scores between all 16 MAGs average 64.3%, which is below the cut-off for genus-level classification (48). Furthermore, these clades display minimal variation in GC content (46.7 - 51.8%) and coding density (84.3 - 90.4%). The MAGs are phylogenetically grouped based on 54 single-copy genes within a clade supported by a bootstrap value of 100. In addition, there are 1135 gene clusters shared only among the 21 MAGs highlighting their relatedness.

Metabolically, BC38-41 would use the EMP for gluconeogenic purposes with the potential to degrade acetoin or ferment propanol. Members are carboxydotrophic and putatively capable of carbon fixation through a proposed pathway of CO reduction to the Calvin cycle. In addition, BC38-41 are facultative anaerobes, having the respiratory potential to occupy both aerobic (*coxABCD*) and anaerobic (*nirK*, *nosZ*, and SQR) niches which may fluctuate seasonally in flooding tropical peatlands. Moreover, BC38-41 seem to be adapted for growth on peptides, particularly peptides with a terminal proline, given their high gene copy number of amino acid transporters and interconversions that lead to oxaloacetate.

This putative family of Bathyarchaeia has been recently proposed as in the order Houtuarculales [55], but given their persistent recovery from peatland environments and their potential unique metabolic adaptations for this type of environment we delineate and propose the BC38-41 as the novel putative family *Candidatus* Paludivitaceae. The standing taxonomic name of this BC or *Candidatus* family (or order) is to be established along with the future isolation of a culture representative. The *Ca.* Paludivitaceae's name is derived from the Latin words "palus" referring to marsh or swamp environments and "vita" meaning life. Together this implies the commonality of both ecological dwelling and metabolic traits suited for life found in peatland environments .

Conclusions

Bathyarchaeia are abundant within PMFB soils, yet our knowledge of their metabolic potential and ecological functions is limited. Here we detail the presence of nine clades within the Bathyarchaeia in PMFB. Genomic evidence points out that key groups (BC38-41 or *Ca.* Paludivitaceae) are facultative anaerobic carboxydotrophs capable of conserving energy through the aerobic oxidation of CO. Moreover, they are mixotrophic able to generate energy from organic compounds such as acetoin, propanol, or peptides with terminal prolines. The *Ca.* Paludivitaceae has also the genomic potential to use CO for both energy and biomass in aerobic environments. These various metabolic findings propose that *Ca.* Paludivitaceae can play a significant role through various interactions including those with methanogenesis in the carbon cycle of tropical peatland environments.

References:

1. Limpens J, Berendse F, Blodau C, Canadell JG, Freeman C, Holden J, Roulet N, Rydin H, Schaepman-Strub G. 2008. Peatlands and the carbon cycle: from local processes to global implications – a synthesis. Biogeosciences 5:1475–1491.
2. Page SE, Baird AJ. 2016. Peatlands and global change: response and resilience. Annu Rev Environ Resour 41:35–57.
3. Kelly TJ, Baird AJ, Roucoux KH, Baker TR, Honorio Coronado EN, Ríos M, Lawson IT. 2014. The high hydraulic conductivity of three wooded tropical peat swamps in northeast Peru: Measurements and implications for hydrological function. Hydrol Process 28:3373–3387.

4. Marwanto S, Watanabe T, Iskandar W, Sabiham S, Funakawa S. 2018. Effects of seasonal rainfall and water table movement on the soil solution composition of tropical peatland. Soil Sci Plant Nutr 64:386–395.

5. Lähteenoja O, Ruokolainen K, Schulman L, Alvarez J. 2009. Amazonian floodplains harbour minerotrophic and ombrotrophic peatlands. Catena 79:140–145.

6. Freeman C, Ostle NJ, Fenner N, Kang H. 2004. A regulatory role for phenol oxidase during decomposition in peatlands. Soil Biol Biochem 36:1663–1667.

7. Hergoualc'h K, Dezzeo N, Verchot L V., Martius C, van Lent J, del Aguila-Pasquel J, López Gonzales M. 2020. Spatial and temporal variability of soil N2O and CH4 fluxes along a degradation gradient in a palm swamp peat forest in the Peruvian Amazon. Glob Chang Biol 26:7198–7216.

8. Draper FC, Roucoux KH, Lawson IT, Mitchard ETA, Honorio Coronado EN, Lähteenoja O, Montenegro LT, Sandoval EV, Zaráte R, Baker TR. 2014. The distribution and amount of carbon in the largest peatland complex in Amazonia. Environ Res Lett 9.

9. Page SE, Rieley JO, Banks CJ. 2011. Global and regional importance of the tropical peatland carbon pool. Glob Chang Biol 17:798–818.

10. Davidson EA, De Araüjo AC, Artaxo P, Balch JK, Brown IF, Mercedes MM, Coe MT, Defries RS, Keller M, Longo M, Munger JW, Schroeder W, Soares-Filho BS, Souza CM, Wofsy SC. 2012. The Amazon basin in transition. Nature 481:321–328.

11. Wang S, Zhuang Q, Lähteenoja O, Draper FC, Cadillo-Quiroz H. 2018. Potential shift from a carbon sink to a source in Amazonian peatlands under a changing climate. Proc Natl Acad Sci U S A 115:12407–12412.

12. Finn DR, Ziv-El M, van Haren J, Park JG, del Aguila-Pasquel J, Urquiza–Muñoz JD,

Cadillo-Quiroz H. 2020. Methanogens and methanotrophs show nutrient-dependent community assemblage patterns across tropical peatlands of the Pastaza-Marañón Basin, Peruvian Amazonia. Front Microbiol 11:746.

13. Buessecker S, Zamora Z, Sarno AF, Finn DR, Hoyt AM, van Haren J, Urquiza Muñoz JD, Cadillo-Quiroz H. 2021. Microbial communities and interactions of nitrogen oxides with methanogenesis in diverse peatlands of the Amazon Basin. Front Microbiol 12:1564.

14. Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. 2003. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. Environ Microbiol 5:787–797.

15. Zhou Z, Pan J, Wang F, Gu JD, Li M. 2018. Bathyarchaeota: Globally distributed metabolic generalists in anoxic environments. FEMS Microbiol Rev 42:639–655.

16. Pan J, Zhou Z, Béjà O, Cai M, Yang Y, Liu Y, Gu JD, Li M. 2020. Genomic and transcriptomic evidence of light-sensing, porphyrin biosynthesis, Calvin-Benson-Bassham cycle, and urea production in Bathyarchaeota. Microbiome 8.

17. Yin X, Zhou G, Cai M, Zhu QZ, Richter-Heitmann T, Aromokeye DA, Liu Y, Nimzyk R, Zheng Q, Tang X, Elvert M, Li M, Friedrich MW. 2022. Catabolic protein degradation in marine sediments confined to distinct archaea. ISME J 2022 1–10.

18. Mohapatra M, Yadav R, Rajput V, Dharne MS, Rastogi G. 2021. Metagenomic analysis reveals genetic insights on biogeochemical cycling, xenobiotic degradation, and stress resistance in mudflat microbiome. J Environ Manage 292:112738.

19. Loh HQ, Hervé V, Brune A. 2021. Metabolic potential for reductive acetogenesis and a novel energy-converting [NiFe] hydrogenase in Bathyarchaeia from termite guts – A genome-centric analysis. Front Microbiol 11:3644.

20. Feng X, Wang Y, Zubin R, Wang F. 2019. Core metabolic features and hot origin of Bathyarchaeota. Engineering 5:498–504.
21. Harris RL, Lau MCY, Cadar A, Bartlett DH, Cason E, van Heerden E, Onstott TC. 2018. Draft genome sequence of " Candidatus Bathyarchaeota" archaeon BE326-BA-RLH, an uncultured denitrifier and putative anaerobic methanotroph from South Africa's deep continental biosphere. Microbiol Resour Announc 7.
22. He Y, Li M, Perumal V, Feng X, Fang J, Xie J, Sievert SM, Wang F. 2016. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. Nat Microbiol https://doi.org/10.1038/nmicrobiol.2016.35.
23. Wang F, Meng J, Wang F, Zheng Y, Peng X, Zhou H, Xiao X. 2009. An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. ISME J 3:106–116.
24. Borrel G, Adam PS, McKay LJ, Chen LX, Sierra-García IN, Sieber CMK, Letourneur Q, Ghozlane A, Andersen GL, Li WJ, Hallam SJ, Muyzer G, de Oliveira VM, Inskeep WP, Banfield JF, Gribaldo S. 2019. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. Nat Microbiol https://doi.org/10.1038/s41564-019-0363-3.
25. Meng J, Xu J, Qin D, He Y, Xiao X, Wang F. 2014. Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. ISME J 8:650–659.
26. Lazar CS, Baker BJ, Seitz K, Hyde AS, Dick GJ, Hinrichs KU, Teske AP. 2016. Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. Environ Microbiol 18:1200–1211.

27. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S, Schramm A, Jorgensen BB. 2013. Predominant archaea in marine sediments degrade detrital proteins. Nat 2013 4967444 496:215–218.
28. Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, Tyson GW. 2015. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science. 350:434–438.
29. Compte-Port S, Fillol M, Gich F, Borrego CM. 2020. Metabolic versatility of freshwater sedimentary archaea feeding on different organic carbon sources. PLoS One 15:e0231238.
30. Li Y, Zhao J, Achinas S, Zhang Z, Krooneman J, Euverink GJW. 2020. The biomethanation of cow manure in a continuous anaerobic digester can be boosted via a bioaugmentation culture containing Bathyarchaeota. Sci Total Environ 745:141042.
31. Suominen S, van Vliet DM, Sánchez-Andrea I, van der Meer MTJ, Sinninghe Damsté JS, Villanueva L. 2021. Organic matter type defines the composition of active microbial communities originating from anoxic Baltic Sea sediments. Front Microbiol 12.
32. Yu T, Wu W, Liang W, Lever MA, Hinrichs KU, Wang F. 2018. Growth of sedimentary Bathyarchaeota on lignin as an energy source. Proc Natl Acad Sci U S A 115:6022–6027.
33. Yu T, Hu H, Zeng X, Wang Y, Pan D, Deng L, Liang L, Hou J, Wang F. 2023. Widespread Bathyarchaeia encode a novel methyltransferase utilizing lignin-derived aromatics. mLife 2:272–282.
34. Lähteenoja O, Ruokolainen K, Schulman L, Alvarez J. 2009. Amazonian floodplains harbour minerotrophic and ombrotrophic peatlands. Catena 79:140–145.
35. Lahteenoja O, Page S. 2011. High diversity of tropical peatland ecosystem types in the Pastaza-Marañón basin, Peruvian Amazonia. J Geophys Res Biogeosciences 116:2025.

36. Lähteenoja O, Page S. 2011. High diversity of tropical peatland ecosystem types in the Pastaza-Marañón basin, Peruvian Amazonia. J Geophys Res Biogeosciences 116:2025.
37. Pavia MJ, Finn D, Macedo-Tafur F, Tello-Espinoza R, Penaccio C, Bouskill N, Cadillo-Quiroz H. 2023. Genes and genome-resolved metagenomics reveal the microbial functional make up of Amazon peatlands under geochemical gradients. Environ Microbiol 1–16.
38. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40.
39. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. Academic Press Inc. https://doi.org/10.1016/j.ymeth.2016.02.020.
40. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.
41. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 2015:e1165.
42. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-

led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6.

43. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

44. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36:1925–1927.

45. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

46. Delmont TO, Eren EM. 2018. Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. PeerJ 6:e4320.

47. Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758–771.

48. Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264.

49. Patil I. 2021. Visualizations with statistical details: The "ggstatsplot" approach. J Open Source Softw 6:3167.

50. Van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol Biol 804:281–295.

51. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.

52. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K,

Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20.

53. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 2017 358 35:725–731.

54. Fillol M, Auguet JC, Casamayor EO, Borrego CM. 2016. Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. ISME J 10:665–677.

55. Liu J, Yang H, Zhao M, Zhang XH. 2014. Spatial distribution patterns of benthic microbial communities along the Pearl Estuary, China. Syst Appl Microbiol 37:578–589.

56. Yu T, Liang Q, Niu M, Wang F. 2017. High occurrence of Bathyarchaeota (MCG) in the deep-sea sediments of South China Sea quantified using newly designed PCR primers. Environ Microbiol Rep 9:374–382.

57. Xiang X, Wang R, Wang H, Gong L, Man B, Xu Y. 2017. Distribution of Bathyarchaeota

Communities Across Different Terrestrial Settings and Their Potential Ecological Functions. Sci Rep 7:45028.

58. Campbell A, Mrázek J, Karlin S. 1999. Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. Proc Natl Acad Sci U S A 96:9184–9189.

59. Karlin S, Mrázek J, Campbell AM. 1997. Compositional biases of bacterial genomes and evolutionary implications. J Bacteriol 179:3899–3913.

60. Kellner S, Spang A, Offre P, Szöllosi GJ, Petitjean C, Williams TA. 2018. Genome size evolution in the Archaea. Emerg Top Life Sci 2:595.

61. Valentine DL. 2007. Adaptations to energy stress dictate the ecology and evolution of the Archaea. Nat Rev Microbiol 5:316–323.

62. Miralles F. 2010. Compositional properties and thermal adaptation of SRP-RNA in bacteria and archaea. J Mol Evol 70:181–189.

63. Kreil DP, Ouzounis CA. 2001. Identification of thermophilic species by the amino acid compositions deduced from their genomes. Nucleic Acids Res 29:1608–1615.

64. Hou J, Wang Y, Zhu P, Yang N, Liang L, Yu T, Niu M, Konhauser K, Woodcroft BJ, Wang F. 2023. Taxonomic and carbon metabolic diversification of Bathyarchaeia during its coevolution history with early Earth surface environment. Sci Adv 9:1–18.

65. Scamuffa MD, Caprioli RM. 1980. Comparison of the mechanisms of two distinct aldolases from Escherichia coli grown on gluconeogenic substrates. Biochim Biophys Acta 614:583–590.

66. Kuprat T, Johnsen U, Ortjohann M, Schönheit P. 2020. Acetate Metabolism in Archaea: Characterization of an Acetate Transporter and of Enzymes Involved in Acetate Activation and Gluconeogenesis in Haloferax volcanii. Front Microbiol 11:3102.

67. Borjian F, Han J, Hou J, Xiang H, Berg IA. 2016. The methylaspartate cycle in haloarchaea and its possible role in carbon metabolism. ISME J 10:546.
68. Nishimura H, Nomura Y, Iwata E, Sato N, Sako Y. 2010. Purification and characterization of carbon monoxide dehydrogenase from the aerobic hyperthermophilic archaeon Aeropyrum pernix. Fish Sci 76:999–1006.
69. Sorokin DY, Merkel AY, Messina E, Tugui C, Pabst M, Golyshin PN, Yakimov MM. 2022. Anaerobic carboxydotrophy in sulfur-respiring haloarchaea from hypersaline lakes. ISME J 16:1534–1546.
70. Esquivel-Elizondo S, Miceli J, Torres CI, Krajmalnik-Brown R. 2018. Impact of carbon monoxide partial pressures on methanogenesis and medium chain fatty acids production during ethanol fermentation. Biotechnol Bioeng 115:341–350.
71. Stegenta-Dąbrowska S, Drabczyński G, Sobieraj K, Koziel JA, Białowiec A. 2019. The biotic and abiotic carbon monoxide formation during aerobic co-digestion of dairy cattle manure with green waste and sawdust. Front Bioeng Biotechnol 7:283.
72. Weber CF, King GM. 2007. Physiological, ecological, and phylogenetic characterization of Stappia, a marine CO-oxidizing bacterial genus. Appl Environ Microbiol 73:1266–1276.
73. Henstra AM, Dijkema C, Stams AJM. 2007. Archaeoglobus fulgidus couples CO oxidation to sulfate reduction and acetogenesis with transient formate accumulation. Environ Microbiol 9:1836–1841.
74. Gagen EJ, Huber H, Meador T, Hinrichs KU, Thomm M. 2013. Novel cultivation-based approach to understanding the Miscellaneous Crenarchaeotic Group (MCG) archaea from sedimentary ecosystems. Appl Environ Microbiol 79:6400–6406.

654 75. Brito JA, Sousa FL, Stelter M, Bandeiras TM, Vonrhein C, Teixeira M, Pereira MM,
655 Archer M. 2009. Structural and functional insights into sulfide:quinone oxidoreductase.
656 Biochemistry 48:5613–5622.
657 76. Dahle H, Økland I, Thorseth IH, Pederesen RB, Steen IH. 2015. Energy landscapes shape
658 microbial communities in hydrothermal systems on the Arctic Mid-Ocean Ridge. ISME J
659 2015 97 9:1593–1606.
660 77. Kubo K, Lloyd KG, F Biddle J, Amann R, Teske A, Knittel K. 2012. Archaea of the
661 Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine
662 sediments. ISME J 2012 610 6:1949–1965.
663 78. Hayat S, Hayat Q, Alyemeni MN, Wani AS, Pichtel J, Ahmad A. 2012. Role of proline
664 under changing environments. Plant Signal Behav 7:1456–1466.
665 79. Maus I, Rumming M, Bergmann I, Heeg K, Pohl M, Nettmann E, Jaenicke S, Blom J,
666 Pühler A, Schlüter A, Sczyrba A, Klocke M. 2018. Characterization of Bathyarchaeota
667 genomes assembled from metagenomes of biofilms residing in mesophilic and
668 thermophilic biogas reactors. Biotechnol Biofuels Bioprod 11:167.
669 80. Zhang Y, Sievert SM. 2014. Pan-genome analyses identify lineage- and niche-specific
670 markers of evolution and adaptation in Epsilonproteobacteria. Front Microbiol 0:110.
671 81. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, Giovannoni S, Eren AM.
672 2019. Single-amino acid variants reveal evolutionary processes that shape the
673 biogeography of a global SAR11 subclade. Elife 8.
674 82. Bezuidt OK, Pierneef R, Gomri AM, Adesioye F, Makhalanyane TP, Kharroub K, Cowan
675 DA. 2016. The Geobacillus Pan-Genome: Implications for the Evolution of the Genus.
676 Front Microbiol 0:723.

83. Gaba S, Kumari A, Medema M, Kaushik R. 2020. Pan-genome analysis and ancestral state reconstruction of class halobacteria: probability of a new super-order. Sci Reports 2020 101 10:1–16.

84. Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics 33:2938–2940.

Figure 1. Maximum likelihood phylogenetic inference of 233 Bathyarchaeia MAGs constructed using 54 concatenated protein sequences. The robustness of MAG placement was assessed using 300 bootstrapping support. Bootstrap support of ≥70% and <70% (black) is represented by hollow and black circles, respectively. Clades with gray backgrounds and that are labeled represent potential genera (based on tree topology and AAI) with more than four MAGs. MAGs recovered from the PMFB are in red. Orders with representatives from the PMFB are labeled at the node. Corresponding heatmaps display the predicted genome size and GC%, for each along with the environmental source from which each MAG was recovered.

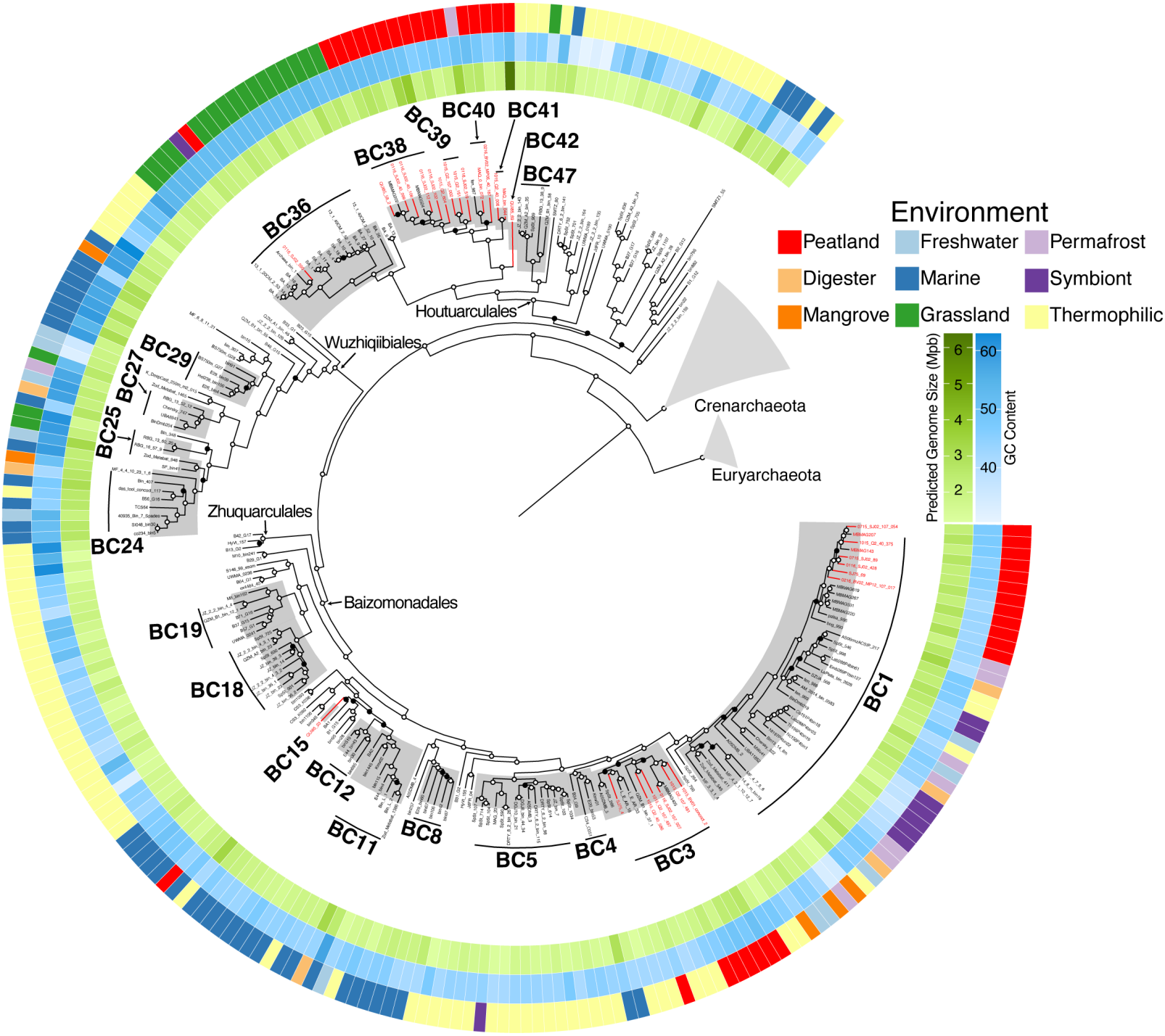
*Figure 2. Mosaic plot of the frequency of MAGs recovered by ecosystems in BCs with more than four representatives. BC labels in red represent clades with PMFB MAG representatives. Asterisks above each column represent the significance level of the relationship of an ecosystem to BC. Significance scale is as follows: *** - [0,0.001], ** - (0.002,0.01], * - (0.01,0.05], ns – not significant.*

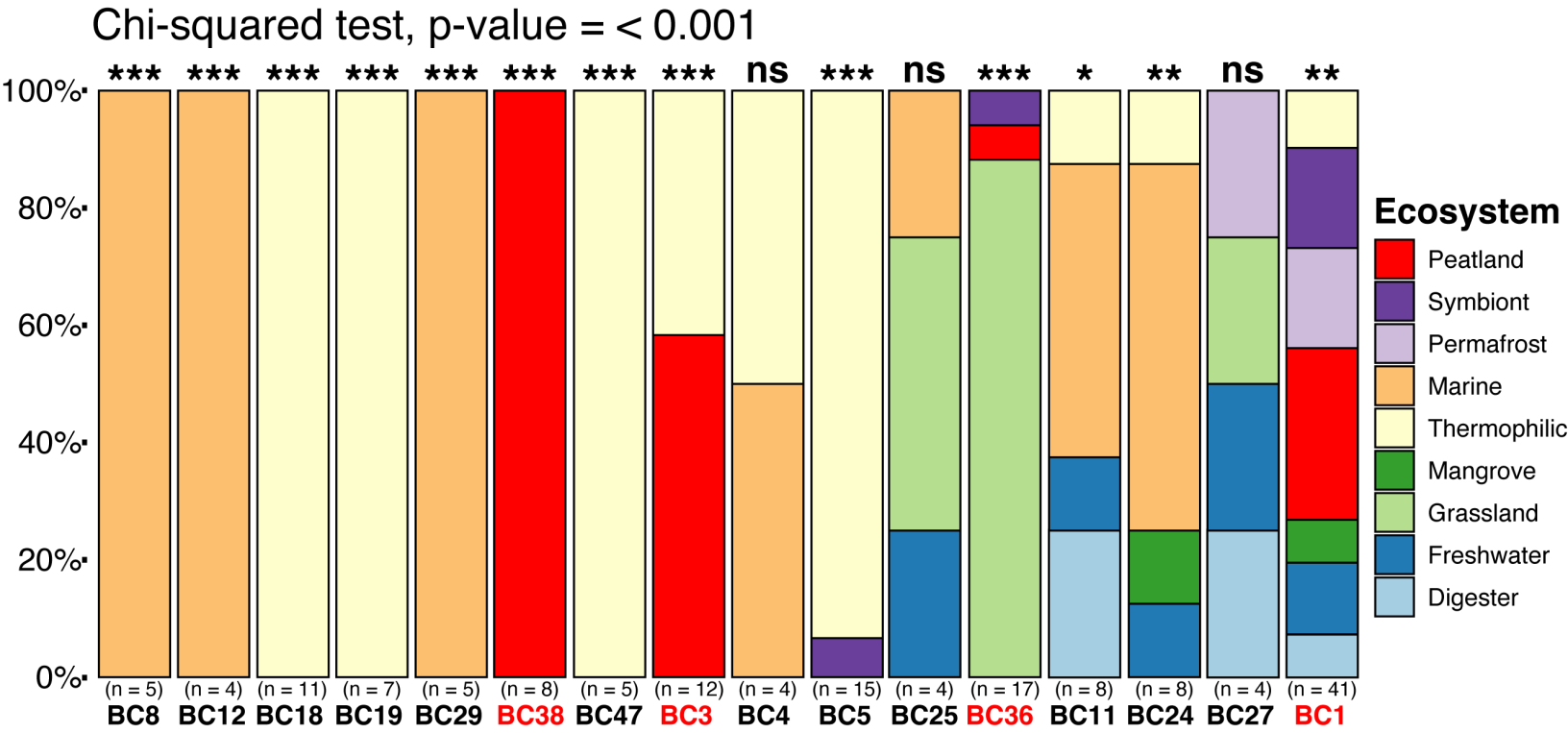
Figure 3. Proposed key (high frequency) and variable metabolic characteristics in Ca. Paludivitaceae*. Gene names are color-coded by the frequency of MAGs with that gene by BC. Key intermediates and pathway outputs are displayed in green. Question mark symbol represents an unknown function. A detailed list of corresponding KEGG numbers, gene names, and presence/absence in respective MAGs are listed in Supplementary File 3.*

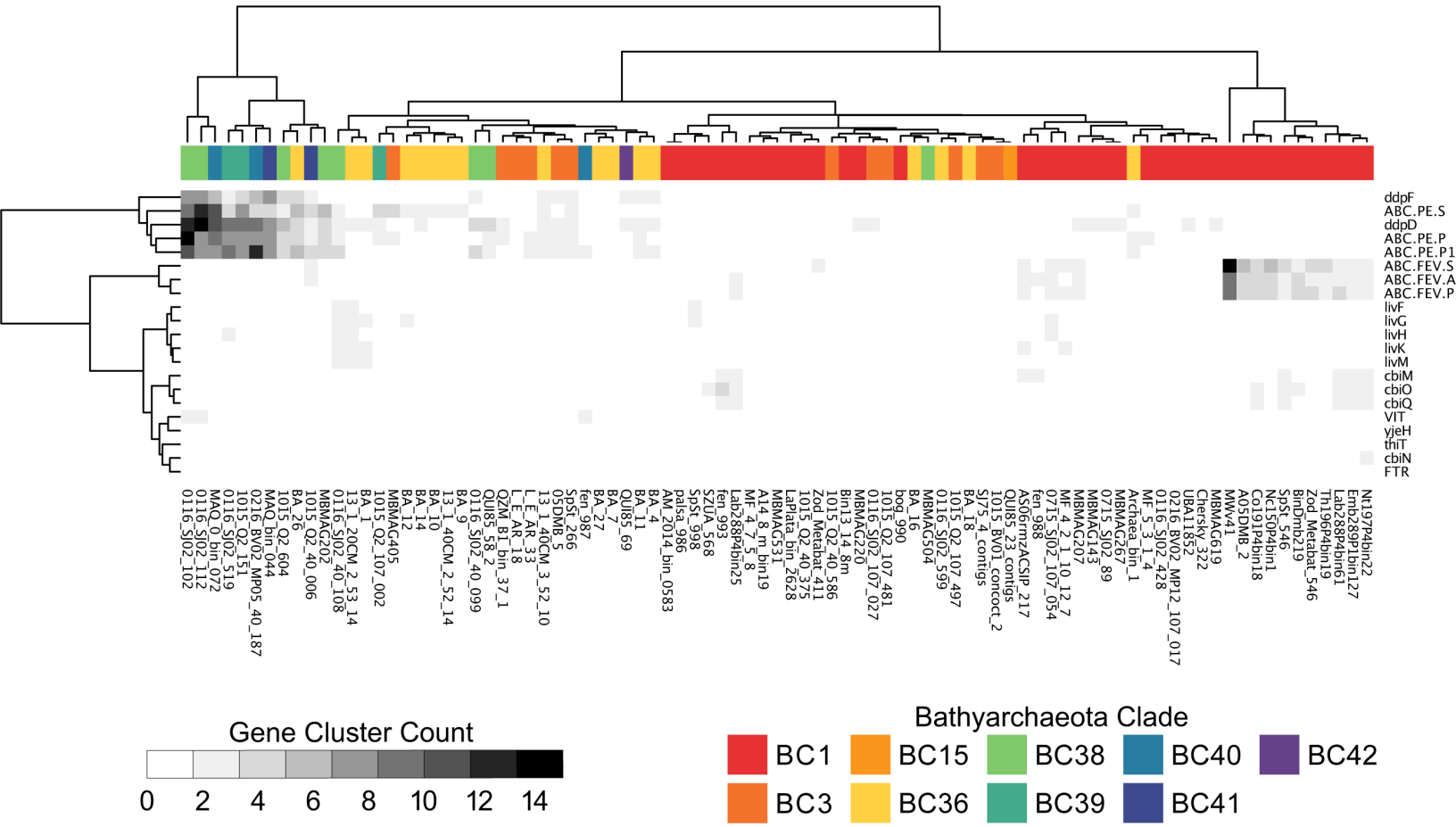
Figure 4. Analysis of abundance of cations and amino acid transport gene clusters. Clustering analysis was completed for gene cluster frequency (side dendrogram) and the phylogenetic

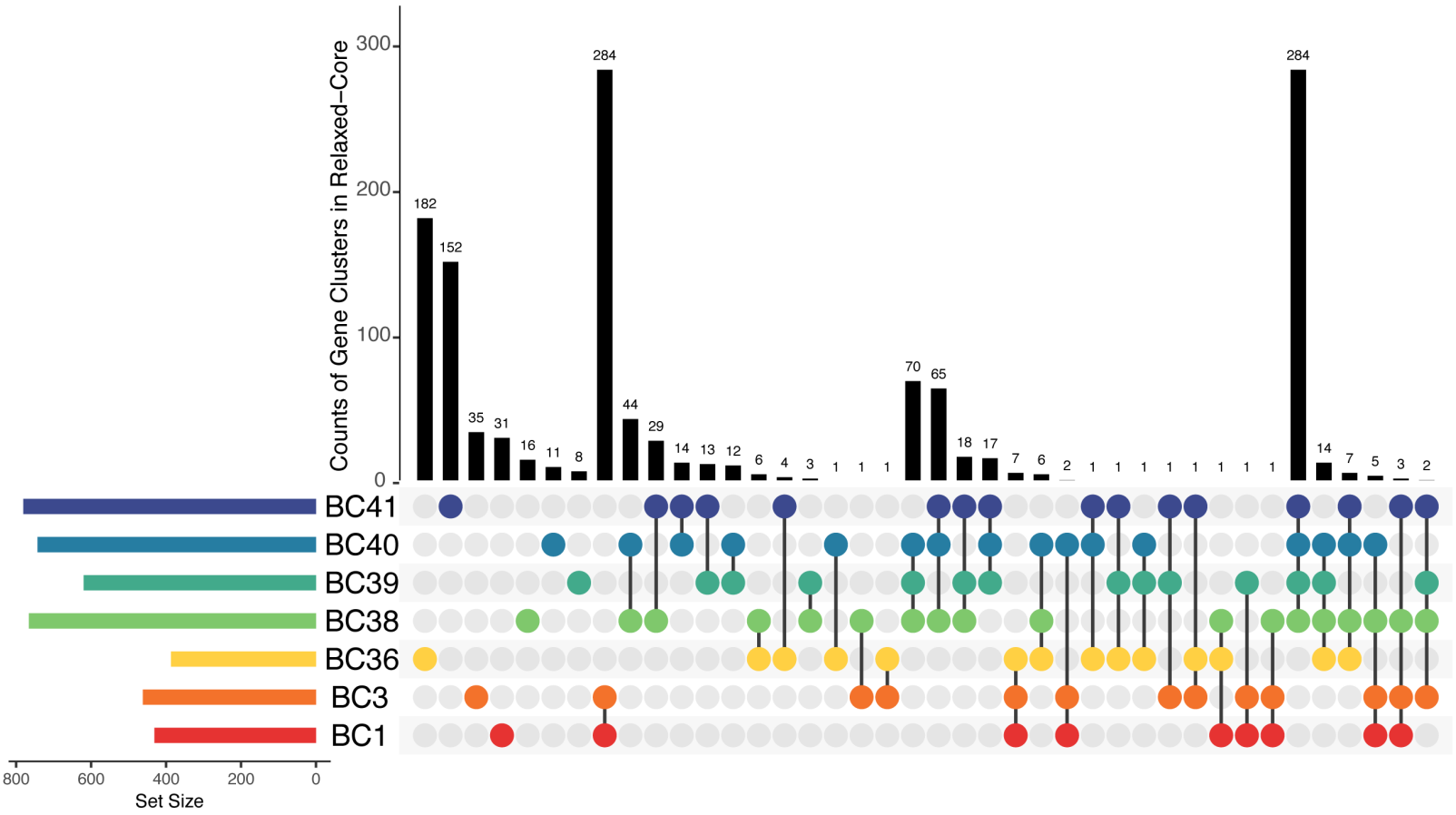
affiliation of MAGs is depicted in the top dendrogram. The colored bar on the top indicates the BC of MAG.

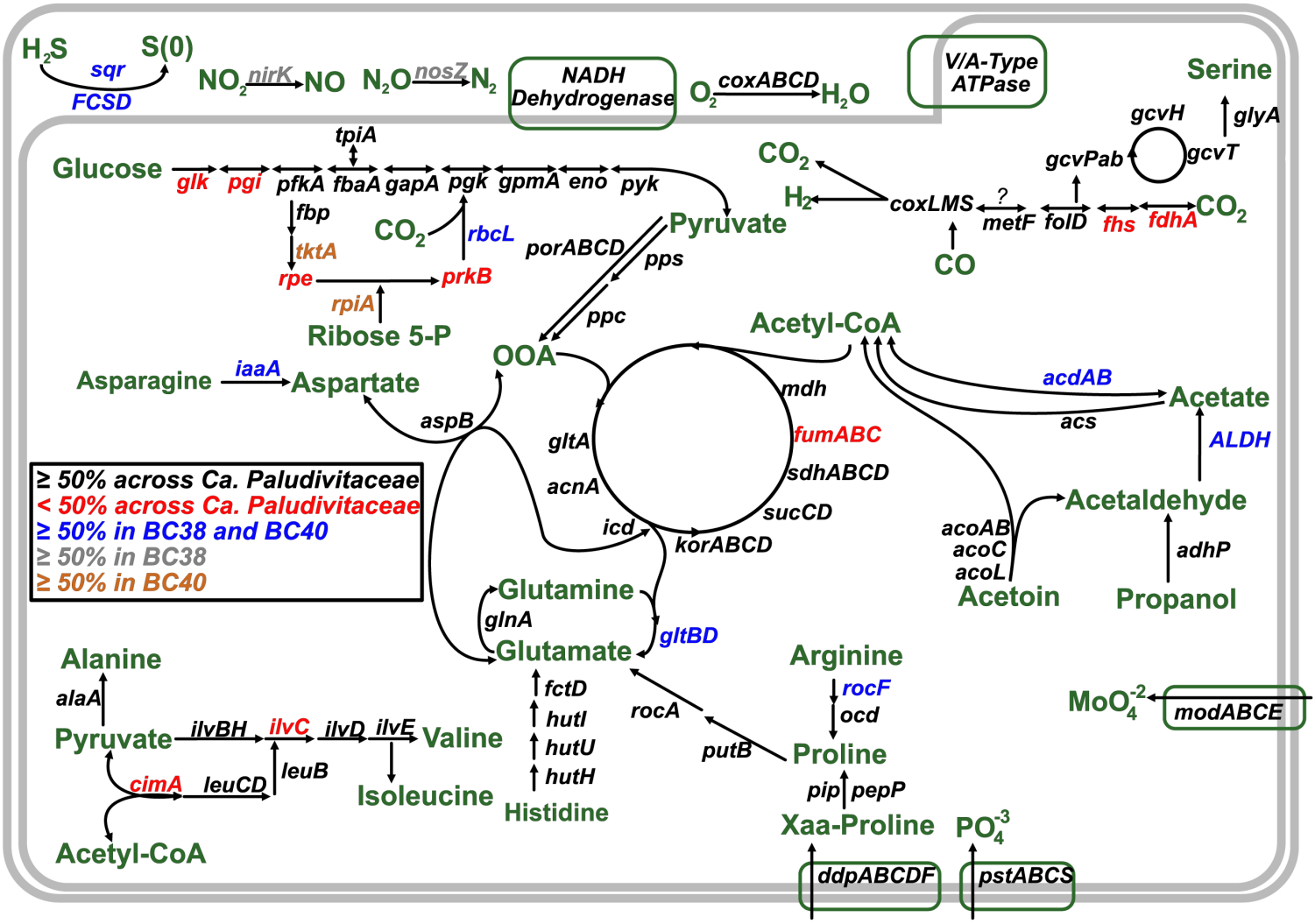
Figure 5. Meta-pangenome analysis of selected BC clades: BC1, BC3, BC36, BC38, BC39, BC40, and BC41. Upset plots (84)*, show the distribution (upper black columns) and overlap of gene clusters between BCs (connected circles) found within the relaxed-core of clades (indicating the size of the set of genes in lower left bars).*


BC40
BC41
BC39
BC38
BC42
BC47
BC36
Houtuarculales
Wuzhiqiibiales
BC29
BC27
BC25
BC24
Zhuquarculales
Baizomonadales
BC19
BC18
BC15
BC12
BC11
BC8
BC5
BC4
BC3
BC1
Crenarchaeota
Euryarchaeota
Environment
Peatland
Freshwater
Permafrost
Digester
Marine
Symbiont
Mangrove
Grassland
Thermophilic
Predicted Genome Size (Mpb)
6
5
4
3
2
GC Content
60
50
40


Chi-squared test, p-value = < 0.001
*** *** *** *** *** *** *** *** ns *** ns *** * ** ns **
100%
80%
60%
40%
20%
0%
(n = 5) (n = 4) (n = 11) (n = 7) (n = 5) (n = 8) (n = 5) (n = 12) (n = 4) (n = 15) (n = 4) (n = 17) (n = 8) (n = 8) (n = 4) (n = 41)
BC8 BC12 BC18 BC19 BC29 BC38 BC47 BC3 BC4 BC5 BC25 BC36 BC11 BC24 BC27 BC1
Ecosystem
Peatland
Symbiont
Permafrost
Marine
Thermophilic
Mangrove
Grassland
Freshwater
Digester


ddpF
ABC.PE.S
ddpD
ABC.PE.P
ABC.PE.P1
ABC.FEV.S
ABC.FEV.A
ABC.FEV.P
livF
livG
livH
livK
livM
cbiM
cbiO
cbiQ
VIT
yjeH
thiT
cbiN
FTR
0116_SJ02_102
0116_SJ02_112
MAQ_0_bin_072
0116_SJ02_519
1015_Q2_151
0216_BV02_MP05_40_187
MAQ_bin_044
1015_Q2_604
BA_26
1015_Q2_40_006
MBMAG202
0116_SJ02_40_108
13_1_20CM_2_53_14
BA_1
1015_Q2_107_002
MBMAG405
BA_12
BA_14
BA_10
13_1_40CM_2_52_14
BA_9
0116_SJ02_40_099
QUI85_58_2
QZM_BT_bin_37_1
L_E_AR_18
L_E_AR_33
13_1_40CM_3_52_10
05DMB_5
SpSt_266
fen_987
BA_27
BA_7
QUI85_69
BA_11
BA_4
AM_2014_bin_0583
palsa_986
SpSt_998
SZUA_568
fen_993
Lab28BP4bin25
MF_4_7_5_8
A14_8_m_bin19
MBMAG531
LaPlata_bin_2628
1015_Q2_40_375
Zod_Metabat_411
1015_Q2_40_586
Bin13_14_8m
MBMAG220
0116_SJ02_107_027
1015_Q2_107_481
boq_990
BA_16
MBMAG504
0116_SJ02_599
1015_Q2_107_497
BA_18
SJ75_4_contigs
1015_BV01_concoct_2
QUI85_23_contigs
AS06rmzACSIP_217
fen_988
MF_4_2_1_10_12_7
0715_SJ02_107_054
MBMAG207
MBMAG143
0715_SJ02_89
MBMAG267
Archaea_bin_1
MF_5_3_1_4
0116_SJ02_428
0216_BV02_MP12_107_017
UBA11852
Chersky_322
MBMAG619
MW.v41
A05DMB_2
Co191P4bin18
SpSt_546
Nc150P4bin1
BinDmb219
Zod_Metabat_546
Th196P4bin19
Lab28BP4bin61
Emb289P1bin127
Nt197P4bin22
Gene Cluster Count
0 2 4 6 8 10 12 14
Bathyarchaeota Clade
BC1
BC15
BC38
BC40
BC42
BC3
BC36
BC39
BC41


Counts of Gene Clusters in Relaxed−Core
300
200
100
0
182
152
35
31
16
11
8
284
44
29
14
13
12
6
4
3
1
1
1
70
65
18
17
7
6
2
1
1
1
1
1
1
1
1
284
14
7
5
3
2
BC41
BC40
BC39
BC38
BC36
BC3
BC1
800
600
400
200
0
Set Size


H2S
S(0)
sqr
FCSD
NO2
nirK
NO
N2O
nosZ
N2
NADH Dehydrogenase
O2
coxABCD
H2O
V/A-Type ATPase
Serine
gcvH
glyA
gcvT
gcvPab
Glucose
tpiA
glk
pgi
pfkA
fbaA
gapA
pgk
gpmA
eno
pyk
fbp
tktA
rpe
prkB
rpiA
Ribose 5-P
CO2
rbcL
porABCD
pps
ppc
Pyruvate
CO2
H2
coxLMS
CO
?
metF
folD
fhs
fdhA
CO2
Acetyl-CoA
OOA
Asparagine
iaaA
Aspartate
aspB
mdh
fumABC
sdhABCD
sucCD
korABCD
icd
acnA
gltA
acdAB
acs
Acetate
ALDH
Acetaldehyde
adhP
Propanol
acoAB
acoC
acoL
Acetoin
≥ 50% across Ca. Paludivitaceae
< 50% across Ca. Paludivitaceae
≥ 50% in BC38 and BC40
≥ 50% in BC38
≥ 50% in BC40
Glutamine
glnA
Glutamate
gltBD
Alanine
alaA
Pyruvate
ilvBH
ilvC
ilvD
ilvE
Valine
Isoleucine
leuB
leuCD
cimA
Acetyl-CoA
fctD
hutI
hutU
hutH
Histidine
rocA
putB
Arginine
rocF
ocd
Proline
pip
pepP
Xaa-Proline
MoO4-2
modABCE
PO4-3
ddpABCDF
pstABCS