Open access • Posted Content • DOI:10.1101/2020.03.09.984617

Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the Methanomassiliicoccales — Source link

Laura Zinke, Paul N. Evans, Alena L. Schroeder, Donovan H. Parks ...+4 more authors

Institutions: University of California, Berkeley, University of Queensland, University of New Hampshire, Ohio State University

Published on: 11 Mar 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Thermoplasmata

Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the *Methanomassiliicoccales*

Laura A. Zinke[1], Paul N. Evans[2], Alena L. Schroeder[1], Donovan H. Parks[2], Ruth K. Varner[3,4], Virginia I. Rich[5], Gene W. Tyson[2], Joanne B. Emerson[1]

[1]Department of Plant Pathology, University of California, Davis, California, USA

[2]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia

[3]Earth Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire, USA

[4]Department of Earth Sciences, University of New Hampshire, Durham, New Hampshire, USA

[5]Department of Microbiology, The Ohio State University, Columbus, Ohio, USA

Running title: Non-methanogens basal to the *Methanomassiliicoccales*

Abstract

Recent discoveries of *mcr* and *mcr*-like complexes in genomes from diverse archaeal lineages suggest that methane (and more broadly alkane) metabolism is an ancient pathway with complicated evolutionary histories. The conventional view is that methanogenesis is an ancestral metabolism of the archaeal class *Thermoplasmata*. Through comparative genomic analysis of 12 *Thermoplasmata* metagenome-assembled genomes (MAGs), we show that these microorganisms do not encode the genes required for methanogenesis, which suggests that this metabolism may have been laterally acquired by an ancestor of the order *Methanomassiliicoccales*. These MAGs include representatives from four orders basal to the *Methanomassiliicoccales*, including a high-quality MAG (95% complete) that likely represents a new order, *Ca.* Lunaplasma lacustris ord. nov. sp. nov. These MAGs are predicted to use diverse energy conservation pathways, such as heterotrophy, sulfur and hydrogen metabolism, denitrification, and fermentation. Two of these lineages are globally widespread among anoxic, sedimentary environments, with the exception of *Ca.* Lunaplasma lacustris, which has thus far only been detected in alpine caves and subarctic lake sediments. These findings advance our understanding of the metabolic potential, ecology, and global distribution of the *Thermoplasmata* and provide new insights into the evolutionary history of methanogenesis within the *Thermoplasmata*.

Introduction

High-throughput sequencing of environmental DNA, metagenomic assembly of DNA sequence reads into contigs, and binning of contigs into metagenome-assembled genomes (MAGs) has provided unprecedented insights into the metabolic potential and evolutionary history of many uncultivated lineages (Hug *et al.*, 2016; Brown *et al.*, 2015; Woodcroft *et al.*,

2018; Castelle *et al.*, 2013; Crits-Christoph *et al.*, 2018; Castelle *et al.*, 2015; Anantharaman *et al.*, 2016). In addition to revealing numerous new phyla, MAG analyses have identified new clades of microorganisms within longstanding phylogenetic groups, assigned biogeochemical roles to known but uncultivated lineages, and attributed new functions to diverse relatives of model prokaryotes (Graham *et al.*, 2018; Mondav *et al.*, 2014; Tully, 2019; Boyd *et al.*, 2019; Solden *et al.*, 2016; Singleton *et al.*, 2018; Martinez *et al.*, 2019). With these advances in sequencing technology, a more complex view of microbial evolution and metabolism has emerged. Increasingly, metabolic divergence between even closely related organisms is now recognized, with suggestions that some metabolism types, such as denitrification, are readily and repeatedly transferred to other microorganisms (Meyer and Kuever, 2007).

Within the last decade, our view of alkane/methane metabolism has evolved from being ascribed to specific *Euryarchaeota* orders to being found in multiple archaeal phyla (Martiny *et al.*, 2015; Evans *et al.*, 2019; Vanwonterghem *et al.*, 2016; Borrel *et al.*, 2019). In addition to the phylogenetic diversity of methanogens, more variations on the methane-cycling biochemical pathways have been discovered (as in Evans *et al.*, 2019; Hua *et al.*, 2019). Genes typically associated with methane metabolism have also been linked to alkane oxidation in multiple Archaea (Laso-Pérez *et al.*). These findings have led to vigorous investigation into the evolutionary history of methanogens and the Archaea (e.g. Wolfe and Fournier, 2018a; Roger and Susko, 2018; Wolfe and Fournier, 2018b; Berghuis *et al.*, 2019; Spang *et al.*, 2017; Borrel *et al.*, 2019). For example, it is hypothesized that the short H_2-dependent methylotrophic methanogenesis pathway might be transferred through horizontal gene transfer more easily than the longer H_2/CO_2 dependent pathway (Borrel *et al.*, 2016; Evans *et al.*, 2019), though a consensus has not been reached.

The current understanding of methanogenesis evolution is that ancestors of the *Euryarchaeota* phylum were methanogens (Evans *et al.*, 2019), or potentially methanotrophs or alkanotrophs (Spang *et al.*, 2017), with methanotrophy-driven acetogenesis also being proposed as an early metabolism type (Russell and Nitschke, 2017). It has been suggested that some euryarchaeal lineages, such as the *Thermoplasmatales*, lost the methanogenesis pathway, while others, such as the *Methanomassiliicoccales*, retained at least part of the pathway (Evans *et al.*, 2019). However, a recent taxonomic reclassification based on relative evolutionary distance (RED) has proposed a new phylum, *Ca.* Thermoplasmatota, which includes the *Thermoplasmatales*, the *Methanomassiliicoccales*, the *Aciduliprofundales*, the MGII/MGIII archaea, and other uncharacterized lineages (Rinke *et al.*, 2019). Within *Ca.* Thermoplasmatota, only the *Methanomassiliicoccales* are known methanogens (Borrel *et al.*, 2014; Dridi *et al.*, 2012), using a truncated H_2-dependent methylotrophic methanogenesis pathway (Lang *et al.*, 2015), while the other *Ca.* Thermoplasmatota lineages utilize a diverse set of metabolic strategies for energy conservation (Rinke *et al.*, 2019; Tully, 2019; Sapra *et al.*, 2003). This suggests that the basal ancestor to the *Ca.* Thermoplasmatota potentially was not a methanogen, which this study further supports.

Here, MAGs from several clades that are phylogenetically close to the methylotrophic methanogenic *Methanomassiliicoccales* are described. These lineages are consistently basal to the *Methanomassiliicoccales* in phylogenetic trees, although in some cases these MAGs were previously classified as *Methanomassiliicoccales* in public databases. Phylogenetic trees, MAG annotation, and functional profiling revealed key differences between these clades and *Methanomassiliicoccales*, most notably a lack of methanogenesis potential in the basal clades. Finally, the global distribution of three of these populations was determined and a new order

within the *Thermoplasmata*, sister to the *Methanomassiliicoccales* and the *Thermoplasmatales,* is proposed, named here as *Ca*. Lunaplasmatales ord. nov.

Materials and Methods

Data acquisition

Previous work suggested that a MAG, referred to here as subarctic lake (SAL) 16, assembled from methanogenic sediments of a subarctic lake in northern Sweden, lacked methanogenesis genes, despite being assigned as a *Methanomassiliicoccales* based on the genomic information and RDP assignment of the assembled and binned 16S rRNA gene (Seitz *et al.*, 2016). To further investigate whether these missing genes were the result of incomplete genome binning or were likely to be a true lack of methanogenic capabilities, MAGs were identified that were closely related to SAL16 in the Genome Taxonomy Database (GDTB; https://gtdb.ecogenomic.org/tree) release 86 archaeal tree, including those which were also in the *Ca.* Thermoplasmata_A class in GTDB (Parks *et al.*, 2018). Additionally, publicly available MAGs and genomes within the *Thermoplasmata* and *Ca.* Thermoplasmata_A classes, which together encompass the *Methanomassiliicoccales*, *Thermoplasmatales*, and *Aciduliprofundum* (Parks *et al.*, 2018), were included here for comparison. MAGs and genomes were downloaded from the National Center for Biotechnology Information (NCBI) database from the accession numbers, the GTDB, or from the publications listed in Supplemental Table 1. MAGs beginning with the "UBA" prefix have been binned and published previously as described in Parks *et al.*, 2017, while those beginning with "RBG" prefix are from Anantharaman *et al.*, 2016, SAL16 is from Emerson *et al*. 2020, and SG8-5 is from Lazar *et al.*, 2017. Genome statistics for all MAGs/genomes were determined using CheckM (Parks *et al.*, 2015; Supplemental Table 2). The completeness and redundancy of

the 12 MAGs of interest were further assessed using the MiGA webserver (http://enve-omics.ce.gatech.edu:3000/, accessed November, 2018; Rodriguez-R *et al.*, 2018; Supplemental Table 3).

MAG identification

MAGs were screened for 16S rRNA gene sequences using the MiGA webserver (Rodriguez-R *et al.*, 2018). Three MAGs contained 16S rRNA gene sequences, which were classified using the Ribosomal Database Project (RDP) classifier (Wang *et al.*, 2007). The 16S rRNA gene sequences were also uploaded to the SILVA Alignment, Classification and Tree (ACT) service (Quast *et al.*, 2012; Pruesse *et al.*, 2007). The sequences were aligned to the global SILVA SSU alignment and classified with a minimum identity of 0.95 and 10 neighbor sequences (Pruesse *et al.*, 2012). The 16S rRNA gene sequences were also compared to the NCBI nucleotide (nt) database using BLASTn for additional insight into taxonomy.

The 12 MAGs of interest were additionally compared to the NCBI Genome database (prokaryotes) using the MiGA server (Rodriguez-R *et al.*, 2018) (Supplemental Tables 5 and 6). Genome and MAG taxonomy was also compared to GTDB taxonomy, which uses RED values to determine taxonomic groupings (Parks *et al.*, 2018). The GTDB Toolkit (GTDB-Tk) v0.1.3 classification workflow (https://github.com/Ecogenomics/GtdbTk) was used to determine RED-based taxonomic placement of MAGs not already in the GTDB database (http://gtdb.ecogenomic.org/) (Supplemental Table 5). Pairwise average amino acid identities (AAIs) between all genomes and MAGs were computed using the envi-omics AAI calculator (http://enve-omics.ce.gatech.edu/aai/, accessed January 2019; Rodriguez-R and Konstantinidis, 2016) (Supplemental Table 7).

Phylogenetic trees

The phylogenetic tree based on 16 concatenated ribosomal proteins used in (Hug *et al.*, 2016) (Figure 1) was constructed using the methodology outlined in (Graham *et al.*, 2018). Briefly, open reading frames in MAGs and genomes were determined using Prodigal v2.6.3 (Hyatt *et al.*, 2010), and ribosomal proteins were identified from these ORFs with HMMER v3.1b2 using the command hmmsearch with an e-value cutoff of 1E-5 (Eddy, 2011). Individual proteins were aligned in Muscle v3.8.31 (Edgar, 2004), and alignments were trimmed using TrimAL v.1.2rev59 in automatic1 mode (Capella-Gutierrez *et al.*, 2009). Proteins were concatenated, and a maximum likelihood tree was calculated using FastTree v.2.1.10 with the parameters -lg -gamma and 1000 bootstraps (Price *et al.*, 2010; 2009). The resulting tree was visualized through the interactive Tree of Life (iTOL) webserver (Letunic and Bork, 2016).

The 16S rRNA gene tree (Supplemental Figure 1) was made using full-length or near full-length sequences from the 14 MAGs and genomes that had 16S rRNA sequences (*Thermoplasmatales* archaeon BRNA1, *Methanomassiliicoccus* sp. UBA386, *Methanomassiliicoccus* sp. UBA345, *Methanomassiliicoccus* sp. UBA6, *Methanomassiliicoccaceae* archaeon UBA593, *Ca.* Methanomethylophilus alvus Mx1201, methanogenic archaeon ISO4-H5, *Aciduliprofundum boonei* T469, *Aciduliprofundum* sp. MAR08-339, *Aciduliprofundum* sp. EPR07-39, *Thermoplasma acidophilum*, *Cuniculiplasma divulgatum* S5, *Cuniculiplamsa divulgatum* PM4, and *Acidiplasma* sp. MBA-1), as well as sequences from the NCBI database (accessed December, 2018). Sequences were aligned and trimmed using Muscle v3.8.45 implemented in Geneious v11.0.5 (100 maximum iterations)

(Edgar, 2004), and a tree was built with RAxML 8.2.12 with the parameters -m GTRCAT -f a -x 123 -p 456 and 1000 bootstraps (Stamatakis, 2014).

The RED-based phylogenetic tree (Supplemental Figure 2) was constructed using the GTDB Toolkit (GTDB-Tk) de_novo workflow (https://github.com/Ecogenomics/GtdbTk), including the GTDB classes Thermoplasmata and Thermoplasmata_A and using Crenarchaeota as the outgroup.

MtrH and MttB trees (Supplemental Figures 3 and 4 and) were made using sequences identified by Prokka v1.13.3, BlastKoala v2.1, and InterProScan v5.30-69.0 in the MAGs and genomes, and select sequences from the NCBI GenBank and UniProt databases (accessed January, 2019). MtrH was selected instead of MtrA for phylogenetic tree reconstruction since more of the *Methanomassiliicoccales* MAGs contained MtrH than MtrA. Sequences were aligned and trimmed using Muscle (Edgar, 2004) implemented in Geneious (100 maximum iterations), and a tree was built with RAxML with the parameters -f a -m PROTGAMMAAUTO -p 12345 -x 12345 and 100 bootstraps (Stamatakis, 2014).

Functional annotation

For all genomes, putative open reading frames (ORFs) were called, translated to amino acid sequences, and functionally annotated in Prokka with the kingdom set to Archaea (Seemann, 2014). InterProScan was used with default settings to compare the ORF amino acid sequence outputs from Prokka to the InterPro, TIGRFAM, and PFAM databases (Jones *et al.*, 2014). ORFs from the 12 MAGs of interest were also uploaded to the blastKOALA server and annotated with KEGG Orthologies using the Archaea taxonomy setting and the "family_eukaryotes + genus_prokaryotes" KEGG GENES database (Kanehisa *et al.*, 2016)

(Supplemental Tables 9-20). These ORFs were also screened to determine which contained export signals using psortb in Archaea mode (Yu *et al.*, 2010). Additional verification through BLASTp against the NCBI nr database was performed for manual curation of MAGs.

Recovering global 16S rRNA gene sequence distributions

The 16S rRNA gene sequences recovered from SG8-5, UBA147, and SAL16 (*Ca.* Lunaplasma lacustris) (16S rRNA gene sequences were recovered from three of the 12 MAGs) were compared to nucleotide sequences in the NCBI-nt database using methods similar to (Mondav *et al.*, 2014). Briefly, for each 16S rRNA gene sequence, standalone blast v2.2.31 command blastall was used with the settings -v 200000 -b 200000 -p blastn -m 8 against the nt database (downloaded in November, 2018) (Lipman *et al.*, 1997). The results were parsed using the bio-table command (https://github.com/pjotrp/bioruby-table) at 97% similarity and requiring the matched sequence to be at least 200 bp long, which removed spurious hits to conserved regions. Hits were manually searched by their GenBank identifiers. Those that could be associated with peer-reviewed publications and included latitude and longitude of the sample origin were used for mapping the distribution of these organisms (Supplemental Table 21). Additionally, the map included locations of the metagenomes from which the MAGs of interest were recovered (Table 1). The map was created in R using the 'maps' and 'ggplot2' packages (Brownrigg *et al.*; Wickham, 2011).

Results and Discussion

Phylogeny of MAGs within and basal to Methanomassiliicoccales

In our previous work, a MAG recovered from subarctic lake sediment metagenomes in Northern Sweden, referred to here as SAL16, was characterized taxonomically as a member of the archaeal order *Methanomassiliicoccales* (Seitz *et al.*, 2016). However, despite being 95% complete, the MAG lacked genetic evidence for methanogenesis, which was unusual, given that all previously identified *Methanomassiliicoccales* were thought to be methanogens (Borrel *et al.*, 2014; Söllinger *et al.*, 2015). Similarly, study of archaeal genomes recovered from aquifer water and sediments near the Colorado River in Rifle, Colorado suggested that *mcrA* gene sequences (indicative of methanogenesis) were not co-binned with the *Methanomassiliicoccales* or their close relatives, but these *Methanomassiliicoccales*-related MAGs were less than 50% complete, so the authors could not conclude whether or not these organisms were methanogens (Castelle *et al.*, 2015). In order to more fully assess the phylogeny and metabolic potential of the SAL16 MAG and its close relatives, we downloaded MAGs and genomes at least 70% complete from NCBI and the GTDB, targeting those classified as or grouped phylogenetically closely to the *Methanomassiliicoccales*, as well as sister clades *Thermoplasmatales* and *Aciduliprofundales* (GTDB classes Thermoplasmata and Thermoplasmata_A).

In total, 68 MAGs and genomes from these lineages were examined, with a focus on 12 publicly available MAGs (including SAL16) that we found to be closely related to, but not within, the *Methanomassiliicoccales* (Table 1; for MAG accession numbers, see Supplemental Table 1). It is important to note, however, that some of these MAGs were labeled as "Methanomassiliicoccales archaeon" in the NCBI database and categorized as "Methanomassiliicoccus" by the RDP classification tool, while being classified within a separate order or even class by other pipelines, such as GTDB-Tk (Parks *et al.*, 2018), MiGA (Rodriguez-R *et al.*, 2018), SILVA SINA (Pruesse *et al.*, 2007; Yilmaz *et al.*, 2013; Quast *et al.*, 2012), and

amino acid identity (AAI) (Rodriguez-R and Konstantinidis, 2016), as described in further detail below. The MAGs range in completeness from 70.8 to 95.2% with up to 6.3% redundancy (Table 1; Supplemental Table 2), as determined by CheckM (Parks *et al.*, 2015). SAL16 is considered a high-quality draft genome by the MIMAG standards (Bowers *et al.*), except for the lack of a binned 5S SSU rRNA gene. SAL16 was characterized as an 'excellent' quality genome through the MiGA webserver (Rodriguez-R *et al.*, 2018), with the other 11 MAGs generally categorized as 'high' (8 MAGs) or 'intermediate' (2 MAGs) quality, apart from one 'low' quality MAG (UBA10834) (Supplemental Table 3).

An assembled 16S rRNA gene sequence was recovered from each of three MAGs: SAL16 (1,465 bp), UBA147 (1,174 bp), and SG8-5 (1,459 bp), the last of which was previously determined to group phylogenetically within the Rice Cluster III lineage (Lazar *et al.*, 2017) (Supplemental Figure 1). The SAL16 16S rRNA gene sequence shares 87% nucleotide identity with both the UBA147 and SG8-5 sequences, indicating that these MAGs belong to different orders within the same class. SG8-5 and UBA147 share 95% similarity between their 16S rRNA gene sequences, revealing that they could represent the same genus, though the 59.7% AAI between them only supports that they represent the same family (Konstantinidis *et al.*, 2017). In all cases, the closest cultured relative was *Methanomassiliicoccales luminyensis* B10, at 86% rRNA gene identity with SAL16 and 88-89% with UBA147 and SG8-5, placing these MAGs in at least novel orders within the *Thermoplasmata* class (Supplemental Table 4) (Konstantinidis *et al.*, 2017). In contrast, the Ribosomal Database Project (RDP) classifier (https://rdp.cme.msu.edu/classifier/classifier.jsp) run with a 95% confidence threshold classified all three MAGs within the *Methanomassiliicoccus* genus (Cole *et al.*, 2009) (Supplemental Table 4). If these sequences had been recovered from 16S rRNA-based amplicon studies, they likely

would have been assigned as *Methanomassilicoccales*, which could lead to spurious assignments of metabolic capability (*i.e.*, an assumption that these sequences represent methanogens).

The RED calculated by the GTDB-Tk workflow (Parks *et al.*, 2018) supported the placement of SAL16 as a novel order within the Thermoplasmata class, which also contains the *Methanomassiliicoccales* (Supplemental Table 5). The 11 other MAGs were previously classified within orders containing no isolated representatives: the orders RBG-16-68-12 (3 MAGs), UBA10834 (5 MAGs), and SG8-5 (3 MAGs) (Supplemental Table 5; Parks *et al.*, 2018). Similarly, taxonomic classification and novelty calculations through the MiGA webserver also indicated that most of these MAGs represent novel orders, and potentially novel classes in some cases (Supplemental Table 5). The five MAGs in the order UBA10834 (RBG_COMBO_56_21; RBG_13_57_23; RBG_16_62_10, UBA10834, and UBA9653 (Figure 1; Supplemental Figure 2; Supplemental Table 5) were estimated to be within the *Thermoplasmata* class, but likely not within the *Methanomassiliicoccales* (Supplemental Table 5). The AAI values between these five MAGs and *Methanomassiliicoccus luminyensis* B10 were 46.9-47.7% AAI (Supplemental Tables 6 and 7), placing them at the lower end of the range suggested for being within the same family (Konstantinidis *et al.*, 2017). These results, taken together with the 16S rRNA gene sequence phylogenies, indicate that these MAGs are all at least in novel orders within the *Thermoplasmata,* and SAL16 potentially represents a novel class within the Euryarchaeota. Here, we will refer to these as novel orders within the Thermoplasmata_A class, in agreement with the RED based taxonomic classification (Parks *et al.*, 2018), or the *Thermoplasmata* as in NCBI.

Phylogenetic trees constructed from 16 concatenated ribosomal proteins (RPs) (Figure 1), 122 proteins used by GTDB (Supplemental Figure 2), and 16S rRNA genes (Supplemental

Figure 1) all show good agreement on the position of these orders as closely related to, but distinct from and basal to, the *Methanomassiliicoccales*. Some slight differences in topology are notable, however; both the RP and 122 protein-based trees show SAL16 (*Ca.* Lunaplasmata lacustris), the RBG-16 order, and the SG8-5 order MAGs as branching basal to the *Methanomassiliicoccales*, which is further supported by the 16S rRNA gene tree for the MAGs with recovered 16S rRNA genes. Placement of the order UBA10834 showed the most variation between the RP and RED-based trees. In the RP tree, UBA10834 order MAGs are grouped in a clade with the three orders listed above (though with low bootstrap support of 0.127), and this clade is basal to the *Methanomassiliicoccales* (Figure 1). However, the RED-based tree places the order UBA10834 as grouping separately from the other three orders, and more closely with the *Methanomassiliicoccales* (Supplemental Figure 2). This placement is supported by a slightly higher AAI between order UBA10834 and *Methanomassiliicoccus* MAGs than between UBA10834 MAGs and MAGs belonging to RBG-16, SG8-5, and SAL16 (*Ca.* Lunaplasma lacustris).

The highly complete (>95%) and minimally contaminated (<5%) SAL16 MAG meets the recently proposed MiMAG standards for naming organisms (Bowers *et al.*; Chuvochina *et al.*, 2019). Thus, we propose naming this microorganism *Ca.* Lunaplasma lacustris ord. nov. sp. nov. within the new order *Cand.* Lunaplasmatales ("Luna" referring to moon, "plasma" referring to being within the Ca. Thermoplasmatota, lacustris referring to lake; formal description below). This names reflects the previous enrichment of a representative of this species from carbonate cave deposits (referred to as moonmilk) in Austria (Reitschuler *et al.*, 2016; 2014), for which there is no published isolate or genomic characterization.

Comparison of the basal lineages to Methanomassiliicoccales

These MAGs are deeply branching members of the same phylogenetic clade that otherwise only includes the methanogenic *Methanomassiliicoccales* (Figure 1). The cultivated *Methanomassiliicoccales* produce methane through methylotrophic methanogenesis, with variations in terms of actual and predicted substrates along with enzyme complexes involved in the metabolism of these compounds compared to more well studied methanogens (Borrel *et al.*, 2014; Speth and Orphan, 2018). A common feature of these microorganisms is a truncated methanogenesis pathway, relative to the canonical hydrogenotrophic methanogenesis pathway observed in many euryarchaeal methanogens (Borrel *et al.*, 2014; Kaster *et al.*, 2011). In the *Methanomassiliicoccales*, the transfers of methyl groups from compounds such as methylamines, methylsulfides, or methanol to Coenzyme M (CoM) to generate methyl-CoM are predicted to be mediated by methyltransferase (MtbA, MtsA/MtsB, MtaA) and methyltransferase/coronoid proteins (MtmBC, MtbBC, MttBC, MtaBC) (Borrel et al. 2013; Lang et al. 2015). The methyl group from the methyl-CoM is then reduced to methane with electrons and H^+ supplied from H_2. These *Methanomassiliicoccales* organisms are H_2-dependent due to the absence of the Wood-Ljungdahl pathway that would otherwise provide reducing equivalents in a similar mechanism to conventional methylotrophic methanogens from the *Methanosarcinales* (Rother et al. 2005). Concomitantly during the reduction of methyl-CoM to methane, another enzyme called Coenzyme B (CoB) replaces the methyl group of methyl-CoM to form a heterodisulfide bond with CoM. The CoM is regenerated by the heterodisulfide/methylviologen reductase complex (HdrABC-MvhADG) with electrons generated from oxidized H_2 in a mechanism similar to typical methanogen electron bifurcating reactions (Lang et al. 2015). The reduced ferredoxin generated from this electron bifurcation is then predicted to oxidize a membrane-bound Fpo-like + hdrD complex to drive H^+ and/or Na^+ translocation across the inner membrane and create the

317 proton or sodium motive force used by ATPases to produce ATP (Borrel *et al.*, 2014; Kröninger
318 *et al.*, 2015; Lang *et al.*, 2015). While these *Methanomassiliicoccales* genomes/MAGs contain
319 many core methanogenesis proteins, notable exceptions include genes encoding for the N5-
320 methyltetrahydromethanopterin:CoM methyltransferase (Mtr) complex, Fmd/FwdABCD, and
321 methanogenesis marker genes 10 and 14 (Borrel *et al.*, 2014).

322 In contrast to the *Methanomassiliicoccales*, the 12 basal MAGs identified in this study
323 lack characterized genes that would confer the potential for H_2-dependent or other forms of
324 methanogenesis. No *mcr* or *mcr*-like open-reading frames (ORFs) were observed in any of these
325 MAGs (Figure 1, Supplemental Figure 5). We recognize that the lack of *mcr* gene sequences in
326 these MAGs does not definitively exclude the possibility that such sequences were in fact present
327 in the genomes, but were not binned within the MAGs. However, all 12 MAGs lack
328 methanogenesis associated genes. Furthermore, our inference that these populations are not
329 methanogenic is supported by the apparent absence of genes coding for the insertion of the
330 amino acid pyrrolysine, along with pyrrolysine biosynthesis genes (*pylBCDS*), and these MAGs
331 lack many of the core "methanogenesis marker genes" observed in *Methanomassiliicoccales*
332 (Figure 1, Supplemental Figure 5) (Borrel *et al.*, 2014; Kaster *et al.*, 2011). Two exceptions were
333 homologs of the methanogenesis marker protein 4, which was found in the three MAGs
334 (UBA9653, RBG-13-57-23, and RBG-19FT-COMBO-56-21), and a methanogenesis marker
335 protein 2 homolog found in RBG-13-57-23. These proteins do not have verified functions
336 (Kaster *et al.*, 2011), so we are unable to assess the role of these putative genes in these MAGs.

337 The three RBG-16 and five UBA10834 order MAGs contain ORFs annotated as
338 methyltransferases in the same protein families as found in the *Methanomassiliicoccales* (Figure
339 1). Most of the putative methyltransferase genes were annotated as coding for MttB (*i.e.*, within

the MttB superfamily) and some for the corresponding corrinoid proteins, MtbC or MttC; no putative methyltransferase alpha subunits, *e.g.* MtaA, were found to be encoded. In methanogens, MttB transfers methyl groups from trimethylamine to the corrinoid protein MttC, then MtaA or MtbA will transfer the methyl group to CoM. Many of the traditional MttB proteins in methanogens contain the amino acid pyrrolysine (Gaston *et al.*, 2011; Borrel *et al.*, 2014). However, non-pyrrolysine containing members of this protein family are widely found in non-methanogens, including MttB superfamily proteins which utilize glycine betaine (Ticak *et al.*, 2014). In the basal MAGs analyzed here, the ORFs annotated as *mttB* do not encode pyrrolysine, and they are phylogenetically distinct from *mttB* sequences of known methanogens (Supplemental figure 4), grouping instead with diverse lineages not associated with methanogens. Therefore, we infer that they encode non-methanogenic MttBs. While some of these putative *mttB* ORFs group more closely to glycine betaine reductases than the methanogenic MttBs, it is unclear what role these MttBs play in the non-methanogenic MAGs. Potentially, these MttBs may confer a methylotrophic metabolism not linked to methane formation.

Notably, ORFs annotated as *mtrAH* genes were observed in some of the environmental *Methanomassiliicoccales* MAGs, three of the SG8-5 order MAGs, but none of the other nine basal MAGs (Figure 1). In euryarchaeal methanogens a complete methyl-H_4MPT–coenzyme-M-methyltransferase (MtrABCDEFGH) complex translocates Na^+ ions across the cell membrane during hydrogenotrophic or acetoclastic methanogenesis, but it is not necessary for methylotrophic methanogenesis (Welander and Metcalf, 2005). Perhaps related to this, genes encoding Mtr subunits are not typically observed in the *Methanomassiliicoccales* (Borrel *et al.*, 2014). In the first observation of a *Methanomassiliicoccales* genome with *mtr* (Speth and

Orphan, 2018), the authors identified *mtrAH* homologs in an environmental *Methanomassiliicoccales* MAG referred to as MAssiliicoccales Lake Pavin (MALP.) Here, we find *mtrAH* homologs in many *Methanomassiliicoccales* MAGs closely related to MALP. Both the *Methanomassiliicoccales* and the SG8-5 *mtrH* sequences group separately from other known methanogens (Supplemental Figure 3), supporting the notion that either *Methanomassiliicoccales* methanogenic abilities might have been gained through horizontal gene transfer or that the *mtrAH* genes were horizontally transferred. Speth and Orphan (2018) proposed that *mtrAH* actually encode methyltetrahydrofolate:CoM methyltransferase, which could be used in the Wood-Ljungdahl pathways found in the MALP MAG (Speth and Orphan, 2018). However, the three SG8-5 MAGs do not contain genes for the Wood-Ljungdahl pathway (Figure 3), so the *mtrAH* homologs in this lineage have an unknown function, but it is likely involved in the transfer of methyl groups.

Evolutionary history of methanogenesis within the Thermoplasmata class

Taken together, these results support our conclusion that Ca. Lunaplasma lacustris, SG8-5, RBG-16, and UBA10834 order MAGs do not represent methanogenic Archaea, despite grouping phylogenetically adjacent and basal to the methanogenic *Methanomassiliicoccales*. These basal MAGs do however have some features in common with the *Methanomassiliicoccales,* such as retaining *hdrABCD* and, in some cases, a select few methanogenesis marker genes (Figure 1). Within the *Thermoplasmata*, only the *Methanomassiliicoccales* are known to be methanogenic, with orders including the *Thermoplasmatales* and the uncultured Candidatus Poseidoniales (previously known as Marine Group II or MGII) containing no known methanogens or methanotrophs (Tully, 2019; Rinke *et al.*, 2019).

386 It has previously been proposed that methanogenesis was lost in the *Thermoplasmatales*
387 and retained in the *Methanomassiliicoccales* (Evans *et al.*, 2019), but with the addition of extra
388 genomes here, it now appears that the *Methanomassiliicoccales*-related MAGs could have gained
389 genes for methane metabolism, further complicating the evolutionary history of metabolic
390 properties in the *Thermoplasmata*. Given the paraphyletic nature of groups basal to the
391 *Methanomassiliicoccales* within the *Thermoplasmata* (Figure 1), the most parsimonious
392 explanation for the lack of methanogenesis in these basal groups is that methanogenesis was in
393 fact not present in the last common ancestor of the *Thermoplasmata.* Instead, methanogenesis
394 was gained by an ancestor of the *Methanomassiliicoccales* through horizontal gene transfer, as
395 hypothesized for the *Ca.* Bathyarchaeota BA1 and BA2 and *Ca.* Verstraetearchaota lineages
396 (Evans *et al.*, 2015; Vanwonterghem *et al.*, 2016). Although the phylogenetic trees presented
397 here show topological differences (including two clades vs. one clade of these MAGs basal to the
398 *Methanomassiliicoccales*, as described above), the preservation of vertical transmission of
399 methanogenesis within the *Thermoplasmata* class to the *Methanomassiliicoccales* would still
400 require more than one loss of methanogenesis in the *Thermoplasmata*. However, as *mcr* or *mcr*-
401 like genes are found throughout the Archaeal tree, potentially mostly due to vertical descent
402 (Hua *et al.*, 2019), then the less parsimonious explanation (an inference of multiple losses of
403 methanogenesis in the *Thermoplasmatales* and other lineages basal to *Methanomassiliicoccales*)
404 is entirely possible.
405
406 *Metabolic potential of lineages basal to the* Methanomassiliicoccales
407 The *Ca.* Lunaplasma lacustris MAG was the highest quality MAG in our dataset, thus we
408 focused our metabolic analyses on this lineage, examining ORF annotations to attribute potential

409 metabolic functional to this MAG. For the less complete MAGs in the SG8-5, UBA10834, and
410 RBG-16 orders, we examined the MAGs as groups and compared these predicted functions to
411 those of *Ca*. Lunaplasma lacustris. In all cases, these MAGs were previously either metabolically
412 uncharacterized (MAGs beginning with the prefix "UBA" (Parks *et al.*, 2018) or characterized as
413 part of larger studies of many MAGs (SG8-5 as in (Lazar *et al.*, 2017) MAGs beginning with the
414 prefix "RBG" (Anantharaman *et al.*, 2016)).

415 Annotation suggests that *Ca.* Lunaplasma lacustris likely can conserve energy through
416 amino acid metabolism, similar to the metabolism proposed for the MBG-D Single Amplified
417 Genome (SAG) in (Lloyd *et al.*, 2013). In *Ca*. Lunaplasma lacustris, amino acids and short
418 peptides are predicted to be imported into the cell by branched chain and polar amino acid
419 transporters (Figure 2). Aminotransferases are predicted to convert the amino acids into 2-keto
420 acids, which could be oxidized through the action of the oxidoreductases Ior, Vor, or Kor.
421 Ferredoxin reduction is predicted to be coupled to amino acid oxidation (Sapra *et al.*, 2003;
422 Lloyd *et al.*, 2013). Ferredoxin (Fd_{red}) can then be oxidized by a membrane-bound Fpo-like +
423 HdrD and/or glcD complex and a membrane-bound hydrogenase, creating a proton motive force.
424 ATP can then be produced through a V/A-type ATPase.

425 Multiple ORFs with homology to hdrABC/mvhADG subunits were found in *Ca.*
426 Lunaplasma lacustris (Figure 2; Supplemental Table 9). In some methanogens, the
427 HdrABC/MvhADG complex reduces the heterodisulphide CoM-S-S-CoB and Fd_{red} (Buan and
428 Metcalf, 2010). However, Hdr subunits are also found in non-methanogenic sulfate-reducing
429 bacteria and archaea, where they have been proposed to function in electron bifurcation
430 associated with sulfur metabolism pathways (Ramos *et al.*, 2015). In *Ca.* Lunaplasma lacustris,
431 this complex could be coupling hydrogen oxidation to the reduction of ferredoxin and an

unknown heterodisulphide or other sulfur compound, as proposed for SG8-5 (Lazar *et al.*, 2017), *Archaeoglobus profundus* (Mander *et al.*, 2004), and an uncultured bacterial population (Castelle *et al.*, 2013).

A beta-oxidation pathway showing similarity, based on the presence of genes similar to *atoB, paaH, crt,* and *bcd/etf*, to those described in the methanogenic *Archaeoglobus* was also observed in *Ca.* Lunaplasma lacustris (Boyd *et al.*, 2019), so it is possible that these organisms degrade fatty acids for energy conservation (Figure 2; Supplemental Table 8). ORFs annotated as acetate-CoA ligase (*acd)* were also present, possibly conferring the ability to produce ATP through fermentation (Schäfer *et al.*, 1993) and providing another route for energy conservation in this apparently versatile population.

The three RBG16 order MAGs (RBG-19FT-COMBO-69-17, RBG-16-67-27, and UBA8695; Supplemental Table 10-12) are the most closely related to *Ca.* Lunaplasma lacustris, and they appear to have similar metabolic potential, *i.e.*, amino acid respiration, fermentation, and beta oxidation (Figure 3). These MAGs and *Ca.* Lunaplasma lacustris also contain ORFs annotated as a sulfide-quinone reductase (*sqr*). In some organisms, Sqr can oxidize H_2S to $S_{(n)}$, which can be coupled to oxygen reduction (Brito *et al.*, 2009; Lencina *et al.*, 2013). In addition to putative *sqr* genes, these MAGs encode rubredoxin, an electron carrier found in some sulfur respiring organisms (Ma *et al.*, 1993; Lumppio *et al.*, 2001), and sulfur dehydrogenases, which can act as bifunctional S^0 or S_n reductases or Fd:NADPH oxidoreductases (Ma and Adams, 1994). Based on the presence of these putative genes, it is likely that these MAGs can participate in sulfur cycling by consuming or producing sulfides.

A distinctive difference between the RBG-16 cluster MAGs and *Ca.* Lunaplasma lacustris is the presence of ORFs annotated as archaeal nitrate reductase complex (*narGH*)

(Yoshimatsu *et al.*, 2000) and nitrite reductase *nirK* in RBG-16-67-27 and RBG-19FT-COMBO-69-17 (Fig. 3). In the less complete UBA8695 (70.76% estimated completeness), *narGH* homologs were not found. UBA8695 did however contain a putative *narC*, which encodes cytochrome b-561 in the archaeon *Haloarcula marismortui* (Yoshimatsu *et al.*, 2007). This indicates that these MAGs likely represent nitrate- and nitrite-reducing archaea, which is consistent with their brief metabolic characterization in the source publication, which examined key functional genes in over 2,500 genomes (Anantharaman *et al.*, 2016). Two MAGs in the RBG-16 order contain cytochrome C oxidase subunits, pointing to the capacity for aerobic respiration (Capaldi *et al.*, 1983). Lending further support to this is the likely lack of rubrethythrin, a protein used to combat oxidative stresses in anaerobes, in the RBG-16 order (Weinberg *et al.*, 2004; Lencina *et al.*, 2013).

ORFs annotated as part of the *nrfD* family, which can function in nitrite oxidation and sulfur/polysulfide reduction (Hussain *et al.*, 1994), were also found in the RBG-16 order and two UBA10834 group MAGS (RBG-19FT-COMBO-56-21 and RBG-13-57-23) (Figure 3; Supplemental Tables 16 and 17). The putative *nrfD* subunits in these MAGs were all located next to a *dmsB* gene (Supplemental Figure 6), which can encode part of a dimethyl sulfoxide or trimethylamine N-oxide reductase complex (Müller and DasSarma, 2005). Only one MAG, RBG-19FT-COMBO-56-21, also contained *dmsA* and *dmsD* in the same operon as *dmsB* and a *nfrD* subunit, which was alternatively annotated as *dmsC*. In RBG-13-57-23, *nrfD* was also located next to a formate dehydrogenase subunit (*fdhD*), and in RBG-19FT-COMBO-69-17, one *nrfD* was between a molybdopterin oxidoreductase encoding gene annotated as encoding tetrothionate reductase subunit A (*ttrA)* and *dmsB*. Boyd et al. (2019) found *dmsB* and *nrfD* located on the same operon in an Archaeoglobus MAG recovered from the deep subseafloor and

478 posited that these genes encode proteins used in sulfur redox chemistry (Boyd *et al.*, 2019).
479 Similarly, sulfate-reducing organisms use NrfD-like proteins as electron shuttles during sulfate
480 reduction (Pereira *et al.*, 2011). These ORFs indicate that the reduction of sulfur compounds (S^0,
481 PS, and/or DMSO) may be a metabolic strategy in these organisms, but due to the various
482 functionalities of the predicted gene products, it remains inconclusive exactly which metabolisms
483 they confer.

484 SG8-5 order MAGs also appear to have a similar metabolic scheme to *Ca.* Lunaplasma
485 lacustris, with some notable deviations and variability among genomes (Figure 3; Supplemental
486 Tables 13-15). In Lazar *et al.*, 2017, SG8-5 was described as a peptide degrader, which supports
487 our findings that all three SG8-5 order organisms could be peptide-degrading. A difference
488 between the three MAGs, though, is that SG8-5 was hypothesized to be using
489 HdrABC/MvhADG as part of its machinery to reduce ferredoxin Lazar *et al.*, 2017. However,
490 UBA147, which is 89.6% complete compared to SG8-5's 72.24% completeness, does not
491 contain any predicted *hdr* genes, nor does UBA280 (77.8% completeness) (Figure 1). This
492 suggests that Hdr may not play a role in UBA147, and instead the membrane-bound
493 hydrogenases and potentially the P-type pyrophosphatases are presumed to be the main drivers
494 of the proton motive force.

495 The predicted metabolic schemes in the order UBA10834 are perhaps the most divergent
496 from the other MAGs basal to the *Methanomassiliicoccales*. Though these MAGs still contain
497 ORFs indicative of amino acid respiration and fermentation as likely metabolic strategies, these
498 MAGs lack evidence of beta-oxidation pathways (Figure 3). However, they contain ORFs
499 annotated as acetyl-CoA synthase/carbon monoxide dehydrogenase, which along with those for
500 the Wood-Ljungdahl pathway are required for carbon fixation (Ragsdale and Pierce, 2008).

Several MAGs in the UBA10834 order contained multiple ORFs annotated as sulfhydrogenase (*hydABDG)*, which can function as a hydrogenase and S^0 or S_n reductase (Ma *et al.*, 1993), though the function here is not confirmed. Thus the UBA10834 MAGs may represent autotrophic organisms that can couple sulfur reduction to hydrogen oxidation.

Global distribution of populations basal to the Methanomassiliicoccales

The NCBI non-redundant nucleotide (nt) database was searched to determine where 16S rRNA gene sequences from these *Methanomassiliicoccales*-related populations have been previously observed. 16S rRNA gene sequences recovered from three MAGs were used for this analysis (Supplemental Table 4). The 16S rRNA gene sequences from UBA147 and SG8-5 were found in globally distributed samples from environments such as sediments, coal beds, oil-sands tailing ponds, and groundwater (Fig. 4; Supplemental Table 21). These findings are consistent with the environmental origins of the MAGs recovered from these lineages and suggest that these lineages may be exclusively environmental. This is in contrast to some *Methanomassiliicoccales* lineages that, thus far, have been exclusively found in the human and animal gastrointestinal tracts (Figure 1) (Söllinger *et al.*, 2015).

The environments containing 16S rRNA gene sequences similar to UBA147 and SG8-5 tended to be anoxic, which is consistent with this lineage having no evidence of aerobic metabolism, unlike the RBG-16 order lineages discussed here. These anaerobic environments included methane-rich environments, such as the methanogenic coal beds in the Powder River Basin in the western United States and methane cold seeps in the northeast Pacific Ocean (Barnhart *et al.*, 2013; Marlow *et al.*, 2014). The majority of the UBA147-related sequences were from anaerobic zones of tailing ponds in Alberta, Canada. The SG8-5 related sequences were found in similar, and sometimes the same, environments as UBA147-related sequences

(Figure 4). However, apart from the coalbed-associated sequences, the SG8-5-related sequences were recovered from marine influenced areas, such as bays, continental margins, or deep-sea sediments, and the SG8-5 MAG itself was recovered from estuary sediments. It thus appears that within the order SG8-5, UBA147 represents a more widely distributed clade, whereas SG8-5-like organisms tend to be more marine-associated.

In contrast to the ubiquitous UBA147 and SG8-5 sequences, 16S rRNA sequences from approximately the same species as *Ca.* Lunaplasma lacustris (i.e. defined as > 97% similar) were found exclusively in calcite deposits ('moonmilk deposits') from alpine caves in the Austrian Alps (Reitschuler et al, 2016). Intriguingly, Reitschuler *et al.* were able to enrich these closely related 'Moonmilk Archaea' under anaerobic conditions at low temperatures (10°C), concluding that these organisms were unlikely to be methanogenic, methanotrophic, or nitrate- or iron-reducing. These conclusions support our findings that *Ca.* Lunaplasma lacustris is likely to be heterotrophic and non-methanogenic. As the *Ca*. Lunaplasma lacustris MAG was recovered from methanogenic sediments in post-glacial freshwater lakes that have seasonal freeze-thaw cycles in northern Sweden (Seitz *et al.*, 2016), it is possible that these organisms prefer cold, anoxic environments, consistent with the metabolic predictions for this lineage.

Description of Ca. Lunaplasma lacustris

Candidatus Lunaplasma (Lu.na.plas'ma. N.L. fem. n. *luna* (from Latin. fem. noun. *luna*) moon, Gr. neut. n. *plasma*, something formed or molded, a form). *Candidatus* Lunaplasma lacustris (la.cus.tris. M.L. fem. adj., lacustris inhabiting lake). "Lunaplasmatales" (Lu.na.plas'ma.tal.es N.L. n. "Lunaplasmatales" -entis, type genus of the family; suff. -ales, ending to denote an order; N.L. neut. pl. n. Lunaplasmatales, the order of the genus "Lunaplasma").

Conclusions

From our results we find that four orders of uncharacterized archaea closely related to *Methanomassiliicoccales* methanogens do not have pathways that would suggest a methanogenic lifestyle. This result is inconsistent with NCBI and RDP database classifications of these MAGs as belonging to the methanogenic order *Methanomassiliicoccales*. Instead, MAGs in these orders (basal to *Methanomassiliicoccales*) contain genes potentially conferring various combinations of sulfur, nitrogen, hydrogen, and aerobic metabolisms, with one order possibly autotrophic. Relationships of 16S rRNA gene sequences to those from published amplicon datasets indicate that species in one order, SG8-5, are widespread in both marine and terrestrial environments, while the *Ca*. Lunaplasmata lacustris MAG apparently represents a species with a constrained ecological distribution. In combination with phylogenetic analyses of the *Methanomassiliicoccales* and its basal lineages, our metabolic reconstructions suggest an alternative evolutionary history of methanogenesis in the *Thermoplasmata*, relative to a recent hypothesis that suggested that methanogenesis was present in the ancestor of the *Thermoplasmata* and vertically transmitted to the *Methanomassiliicoccales* but lost in the *Thermoplasmatales*. We suggest instead that the capacity for methanogenesis was absent in the ancestor of the *Thermoplasmata* and was gained in the ancestor of the *Methanomassiliicoccales* via horizontal gene transfer after divergence from the basal lineages analyzed here. Another possible explanation is that the ability to perform methanogenesis was present in a basal ancestor to the *Thermoplasmatales,* but was lost multiple times in the descending lineages. These findings highlight the potential roles of non-methanogenic *Thermoplasmata* in the environment and contribute to our rapidly developing understanding of the evolution of methanogenesis.

Acknowledgements: We thank Elaina Graham and Lily Momper for advice on phylogenetic tree construction, and members of the NSF funded Northern Ecosystems Research for Undergraduates REU Site (EAR#1063037) team Kaitlyn Steele, Martin Wik, and Nancy Freitas for sample collection. The support and resources from the High Performance Computing group in the Bioinformatics Core at the University of California, Davis are gratefully acknowledged. Our funding sources include new laboratory start-up to JBE from the UC Davis College of Agricultural and Environmental Sciences and the UC Davis Department of Plant Pathology, and SAL-16 MAG recovery via support from the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, grants DE-SC0010580 and DE-SC0016440 to VIR and GWT.

Competing interests: The authors declare no competing financial interests.

REFERENCES

Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, *et al.* (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Comms* **7**: 13219.

Barnhart EP, De León KB, Ramsay BD, Cunningham AB, Fields MW. (2013). Investigation of coal-associated bacterial and archaeal populations from a diffusive microbial sampler (DMS). *International Journal of Coal Geology* **115**: 64–70.

Berghuis BA, Yu FB, Schulz F, Blainey PC, Woyke T, Quake SR. (2019). Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *Proc Natl Acad Sci USA* **116**: 5037–5044.

Borrel G, Adam PS, Gribaldo S. (2016). Methanogenesis and the Wood–Ljungdahl Pathway: An Ancient, Versatile, and Fragile Association. *Genome Biology and Evolution* **8**: 1706–1711.

Borrel G, Adam PS, McKay LJ, Chen L-X, Sierra-García IN, Sieber CMK, *et al.* (2019). Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat Microbiol* **4**: 603–613.

Borrel G, Parisot N, Harris HMB, Peyretaillade E, Gaci N, Tottey W, *et al.* (2014). Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC genomics* **15**: 679–679.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **35**: 725–731.

Boyd JA, Jungbluth SP, Leu AO, Evans PN, Woodcroft BJ, Chadwick GL, *et al.* (2019). Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. *ISMEJ* **13**: 1269–1279.

Brito JA, Sousa FL, Stelter M, Bandeiras TM, Vonrhein C, Teixeira M, *et al.* (2009). Structural and Functional Insights into Sulfide:Quinone Oxidoreductase . *Biochemistry* **48**: 5613–5622.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, *et al.* (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208–211.

Brownrigg R, Becker RA, Wilks AR. maps: Draw Geographical Maps. R package version 2.3-9.

Buan NR, Metcalf WW. (2010). Methanogenesis by Methanosarcina acetivoransinvolves two structurally and functionally distinct classes of heterodisulfide reductase. *Molecular Microbiology* **75**: 843–853.

Capaldi RA, Malatesta F, Darley-Usmar VM. (1983). Structure of cytochrome c oxidase. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **726**: 135–148.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

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

Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, *et al.* (2015). Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. *Current Biology* **25**: 690–701.

630 Chuvochina M, Rinke C, Parks DH, Rappé MS, Tyson GW, Yilmaz P, *et al.* (2019). The
631 importance of designating type material for uncultured taxa. *Systematic and Applied*
632 *Microbiology* **42**: 15–21.

633 Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, *et al.* (2009). The Ribosomal Database
634 Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37**:
635 D141–D145.

636 Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. (2018). Novel soil
637 bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**: 440–444.

638 Dridi B, Fardeau M-L, Ollivier B, Raoult D, Drancourt M. (2012). Methanomassiliicoccus
639 luminyensis gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *IJSEM*
640 **62**: 1902–1907.

641 Eddy SR. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**: e1002195.

642 Edgar RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and
643 space complexity. *BMC Bioinformatics* **5**: 113.

644 Emerson JB, Varner RK, Wik M, Parks DH, Neumann R, Johnson JE, *et al.* (2020)
645 Spatiotemporal differences in subarctic lake methane emission patterns linked to sediment
646 microbial communities. *bioRxiv* 2020.02.08.934661
647
648 Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, *et al.* (2019). An
649 evolving view of methane metabolism in the Archaea. *Nature Reviews Microbiology* **17**: 219–
650 232.

651 Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, *et al.* (2015).
652 Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric
653 metagenomics. *Science* **350**: 434.

654 Gaston MA, Jiang R, Krzycki JA. (2011). Functional context, biosynthesis, and genetic encoding
655 of pyrrolysine. *Cell Regulation* **14**: 342–349.

656 Graham ED, Heidelberg JF, Tully BJ. (2018). Potential for primary productivity in a globally-
657 distributed bacterial phototroph. *ISMEJ* **12**: 1861–1866.

658 Hua Z-S, Wang Y-L, Evans PN, Qu Y-N, Goh KM, Rao Y-Z, *et al.* (2019). Insights into the
659 ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea.
660 *Nat Comms* **10**: 4574.

661 Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, *et al.* (2016). A new
662 view of the tree of life. *Nat Microbiol* **1**: 16048.

663 Hussain H, Grove J, Griffiths L, Busby S, Cole J. (1994). A seven-gene operon essential for
664 formate-dependent nitrite reduction to ammonia by enteric bacteria. *Molecular Microbiology* **12**:
665 153–163.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, *et al.* (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Kanehisa M, Sato Y, Morishima K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**: 726–731.

Kaster A-K, Goenrich M, Seedorf H, Liesegang H, Wollherr A, Gottschalk G, *et al.* (2011). More than 200 genes required for methane formation from H_2 and CO_2 and energy conservation are present in Methanothermobacter marburgensis and Methanothermobacter thermautotrophicus. *Archaea* **2011**: 973848–973848.

Konstantinidis KT, Rosselló-Móra R, Amann R. (2017). Uncultivated microbes in need of their own taxonomy. *ISMEJ* **11**: 2399–2406.

Kröninger L, Berger S, Welte C, Deppenmeier U. (2015). Evidence for the involvement of two heterodisulfide reductases in the energy-conserving system of Methanomassiliicoccus luminyensis. *FEBS J* **283**: 472–483.

Lang K, Schuldes J, Klingl A, Poehlein A, Daniel R, Brune A. (2015). New Mode of Energy Metabolism in the Seventh Order of Methanogens as Revealed by Comparative Genome Analysis of 'Candidatus Methanoplasma termitum'. *Appl Environ Microbiol* **81**: 1338.

Laso-Pérez R, Wegener G, Knittel K, Widdel F, Harding KJ, Krukenberg V, *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**: 396–401.

Lazar CS, Baker BJ, Seitz KW, Teske AP. (2017). Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISMEJ* **11**: 1118–1129.

Lencina AM, Ding Z, Schurig-Briccio LA, Gennis RB. (2013). Characterization of the Type III sulfide:quinone oxidoreductase from Caldivirga maquilingensis and its membrane binding. *Biochimica et biophysica acta* **1827**: 266–275.

Letunic I, Bork P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**: W242–W245.

Lipman DJ, Zhang J, Madden TL, Altschul SF, Schäffer AA, Miller W, *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, *et al.* (2013). Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**: 215–218.

Lumppio HL, Shenvi NV, Summers AO, Voordouw G, Kurtz DM. (2001). Rubrerythrin and Rubredoxin Oxidoreductase in Desulfovibrio vulgaris: a Novel Oxidative Stress Protection System. *J Bacteriol* **183**: 101.

Ma K, Adams MW. (1994). Sulfide dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus: a new multifunctional enzyme involved in the reduction of elemental sulfur. *J Bacteriol* **176**: 6509.

Ma K, Schicho RN, Kelly RM, Adams MW. (1993). Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci USA* **90**: 5341.

Mander GJ, Pierik AJ, Huber H, Hedderich R. (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon Archaeoglobus profundus. *European Journal of Biochemistry* **271**: 1106–1116.

Marlow JJ, Steele JA, Case DH, Connon SA, Levin LA, Orphan VJ. (2014). Microbial abundance and diversity patterns associated with sediments and carbonates from the methane seep environments of Hydrate Ridge, OR. *Front Mar Sci* **1**: 195.

Martinez MA, Woodcroft BJ, Ignacio Espinoza JC, Zayed AA, Singleton CM, Boyd JA, *et al.* (2019). Discovery and ecogenomic context of a global Caldiserica-related phylum active in thawing permafrost, Candidatus Cryosericota phylum nov., Ca. Cryosericia class nov., Ca. Cryosericales ord. nov., Ca. Cryosericaceae fam. nov., comprising the four species Cryosericum septentrionale gen. nov. sp. nov., Ca. C. hinesii sp. nov., Ca. C. odellii sp. nov., Ca. C. terrychapinii sp. nov. *Systematic and Applied Microbiology* **42**: 54–66.

Martiny JBH, Jones SE, Lennon JT, Martiny AC. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science* **350**: aac9323.

Meyer B, Kuever J. (2007). Molecular Analysis of the Diversity of Sulfate-Reducing and Sulfur-Oxidizing Prokaryotes in the Environment, Using aprA as Functional Marker Gene. *AEM* **73**: 7664–7679.

Mondav R, Woodcroft BJ, Kim E-H, McCalley CK, Hodgkins SB, Crill PM, *et al.* (2014). Discovery of a novel methanogen prevalent in thawing permafrost. *Nat Comms* **5**: 3212.

Müller JA, DasSarma S. (2005). Genomic Analysis of Anaerobic Respiration in the Archaeon Halobacterium sp. Strain NRC-1: Dimethyl Sulfoxide and Trimethylamine N-Oxide as Terminal Electron Acceptors. *J Bacteriol* **187**: 1659.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, *et al.* (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**: 996–1004.

738 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing
739 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
740 *Genome Research*.

741 Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, *et al.* (2017).
742 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life.
743 *Nat Microbiol* **2**: 1533–1542.

744 Pereira IAC, Ramos AR, Grein F, Marques MC, da Silva SM, Venceslau SS. (2011). A
745 Comparative Genomic Analysis of Energy Metabolism in Sulfate Reducing Bacteria and
746 Archaea. *Front Microbiol* **2**: 69.

747 Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 – Approximately Maximum-Likelihood
748 Trees for Large Alignments Poon AFY (ed). *PLOS ONE* **5**: e9490.

749 Price MN, Dehal PS, Arkin AP. (2009). FastTree: Computing Large Minimum Evolution Trees
750 with Profiles instead of a Distance Matrix. *mbe* **26**: 1641–1650.

751 Pruesse E, Peplies J, Glöckner FO. (2012). SINA: Accurate high-throughput multiple sequence
752 alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

753 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, *et al.* (2007). SILVA: a
754 comprehensive online resource for quality checked and aligned ribosomal RNA sequence data
755 compatible with ARB. *Nucleic Acids Research* **35**: 7188–7196.

756 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2012). The SILVA
757 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic*
758 *Acids Research* **41**: D590–D596.

759 Ragsdale SW, Pierce E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO2
760 fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1784**: 1873–1898.

761 Ramos AR, Grein F, Oliveira GP, Venceslau SS, Keller KL, Wall JD, *et al.* (2015). The
762 FlxABCD-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide
763 reductase widespread in anaerobic bacteria and involved in ethanol metabolism in Desulfovibrio
764 vulgaris Hildenborough. *Environ Microbiol* **17**: 2288–2305.

765 Reitschuler C, Lins P, Wagner AO, Illmer P. (2014). Cultivation of moonmilk-born non-
766 extremophilic Thaum and Euryarchaeota in mixed culture. *No oxygen-still vigorous: 8th*
767 *International Symposium on Anaerobic Microbiology (ISAM 8) in Innsbruck, Austria* **29**: 73–79.

768 Reitschuler C, Spötl C, Hofmann K, Wagner AO, Illmer P. (2016). Archaeal Distribution in
769 Moonmilk Deposits from Alpine Caves and Their Ecophysiological Potential. *Microbial Ecology*
770 **71**: 686–699.

771 Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, *et al.* (2019). A
772 phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca.
773 Poseidoniales ord. nov.). *ISMEJ* **13**: 663–675.

774 Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Móra R, Tiedje JM, Cole JR, *et al.* (2018).
775 The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of
776 Archaea and Bacteria at the whole genome level. *Nucleic Acids Research* **46**: W282–W288.

777 Rodriguez-R LM, Konstantinidis KT. (2016). The enveomics collection: a toolbox for
778 specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* **4**: e1900v1.

779 Roger AJ, Susko E. (2018). Molecular clocks provide little information to date methanogenic
780 Archaea. *Nature Ecology & Evolution* **2**: 1676–1677.

781 Russell MJ, Nitschke W. (2017). Methane: Fuel or Exhaust at the Emergence of Life?
782 *Astrobiology* **17**: 1053–1066.

783 Sapra R, Bagramyan K, Adams MWW. (2003). A simple energy-conserving system: Proton
784 reduction coupled to proton translocation. *Proc Natl Acad Sci USA* **100**: 7545.

785 Schäfer T, Selig M, Schönheit P. (1993). Acetyl-CoA synthetase (ADP forming) in archaea, a
786 novel enzyme involved in acetate formation and ATP synthesis. *Archives of Microbiology* **159**:
787 72–83.

788 Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–
789 2069.

790 Singleton CM, McCalley CK, Woodcroft BJ, Boyd JA, Evans PN, Hodgkins SB, *et al.* (2018).
791 Methanotrophy across a natural permafrost thaw environment. *ISMEJ*.

792 Solden LM, Hoyt DW, Collins WB, Plank JE, Daly RA, Hildebrand E, *et al.* (2016). New roles
793 in hemicellulosic sugar fermentation for the uncultivated Bacteroidetes family BS11. *ISMEJ* **11**:
794 691–703.

795 Söllinger A, Schleper C, Urich T, Schwab C, Loy A, Weinmaier T, *et al.* (2015). Phylogenetic
796 and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals
797 clade-specific habitat preferences. *FEMS Microbiology Ecology* **92**. e-pub ahead of print, doi:
798 10.1093/femsec/fiv149.

799 Spang A, Caceres EF, Ettema TJG. (2017). Genomic exploration of the diversity, ecology, and
800 evolution of the archaeal domain of life. *Science* **357**: eaaf3883.

801 Speth DR, Orphan VJ. (2018). Metabolic marker gene mining provides insight in global mcrA
802 diversity and, coupled with targeted genome reconstruction, sheds further light on metabolic
803 potential of the Methanomassiliicoccales. *PeerJ* **6**: e5614.

804 Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of
805 large phylogenies. *Bioinformatics* **30**: 1312–1313.

806 Ticak T, Kountz DJ, Girosky KE, Krzycki JA, Ferguson DJ. (2014). A nonpyrrolysine member
807 of the widely distributed trimethylamine methyltransferase family is a glycine betaine
808 methyltransferase. *Proc Natl Acad Sci USA* **111**: E4668.

809 Tully BJ. (2019). Metabolic diversity within the globally abundant Marine Group II Euryarchaea
810 offers insight into ecological patterns. *Nat Comms* **10**: 271.

811 Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, *et al.* (2016).
812 Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat*
813 *Microbiol* **1**: 16170.

814 Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naïve Bayesian Classifier for Rapid
815 Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol* **73**:
816 5261.

817 Weinberg MV, Jenney FE, Cui X, Adams MWW. (2004). Rubrerythrin from the
818 Hyperthermophilic Archaeon Pyrococcus furiosus Is a Rubredoxin-Dependent, Iron-Containing
819 Peroxidase. *J Bacteriol* **186**: 7888.

820 Welander PV, Metcalf WW. (2005). Loss of the mtr operon in Methanosarcina blocks growth on
821 methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. *Proc Natl*
822 *Acad Sci U S A* **102**: 10664.

823 Wickham H. (2011). ggplot2: Elegant Graphics for Data Analysis. *Biometrics* **67**: 678–679.

824 Wolfe JM, Fournier GP. (2018a). Horizontal gene transfer constrains the timing of methanogen
825 evolution. *Nature Ecology & Evolution* **2**: 897–903.

826 Wolfe JM, Fournier GP. (2018b). Reply to 'Molecular clocks provide little information to date
827 methanogenic Archaea'. *Nature Ecology & Evolution* **2**: 1678–1678.

828 Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, *et al.* (2018).
829 Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**: 49–54.

830 Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, *et al.* (2013). The SILVA and
831 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Research* **42**:
832 D643–D648.

833 Yoshimatsu K, Araya O, Fujiwara T. (2007). Haloarcula marismortui cytochrome b-561 is
834 encoded by the narC gene in the dissimilatory nitrate reductase operon. *Extremophiles* **11**: 41–
835 47.

836 Yoshimatsu K, Sakurai T, Fujiwara T. (2000). Purification and characterization of dissimilatory
837 nitrate reductase from a denitrifying halophilic archaeon, Haloarcula marismortui. *FEBS Letters*
838 **470**: 216–220.

839 Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, *et al.* (2010). PSORTb 3.0: improved
840 protein subcellular localization prediction with refined localization subcategories and predictive
841 capabilities for all prokaryotes. *Bioinformatics* **26**: 1608–1615.

842

Figures

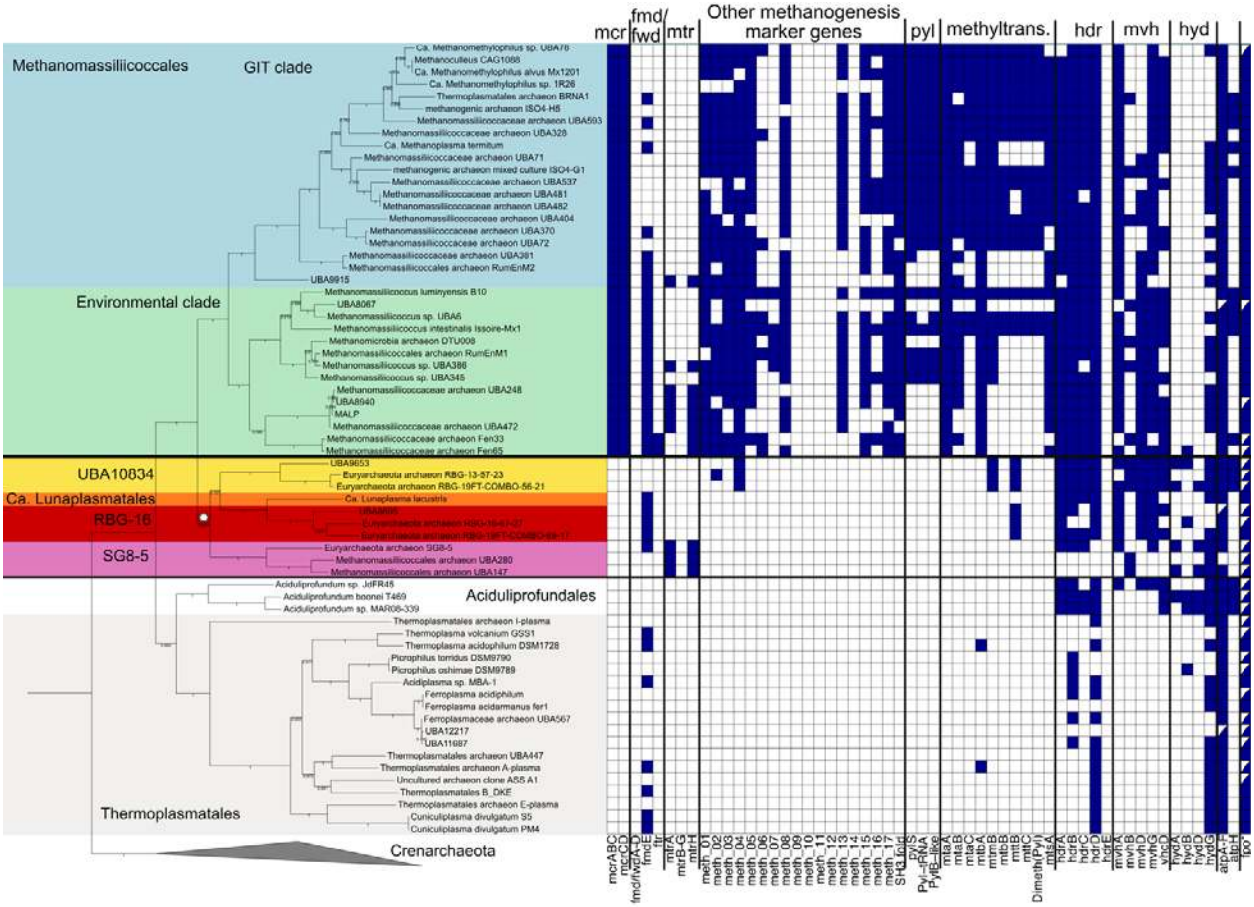


Figure 1. Ribosomal protein (RP) tree. Tree (based on the 16 proteins used in Hug *et al.*, 2016), showing the phylogenetic relationships among 10 of the non-methanogenic MAGs basal to the *Methanomassiliicoccales* (in gold, orange, red, and pink backgrounds) and closely related Archaea, including the *Methanomassiliicoccales* (gastrointestinal tract (GIT) and environmental clades in blue and green backgrounds, respectively, following clade designations from Söllinger *et al.* (2016)), *Aciduliprofundales* (white background), and *Thermoplasmatales* (grey background). The white star represents the proposed gain of methanogenesis in the *Methanomassiliicoccales*. Bootstrap values (of 1000 bootstraps) are shown as a filled black dot where support was higher than 80%, as an empty dot for more than 50% support, and as values

for branches with less than 50% support. The heatmap to the right demonstrates the genomic potential for methanogenesis in each of the genomes and MAGs (genes colored blue were detected, white were not, half-filled boxes indicate where some subunits were missing; see Supplemental Table 8 for further gene information). Methanogenesis marker genes are those listed in Borrel *et al*. 2014. Two MAGs in the UBA10834 group are not included in the tree because they lack at least 8 of the 16 RP required to be included in the tree. The methanogenesis complement for these MAGs is in Supplemental Figure 5. *Fpo includes subunits ABCDHIJKLMN.

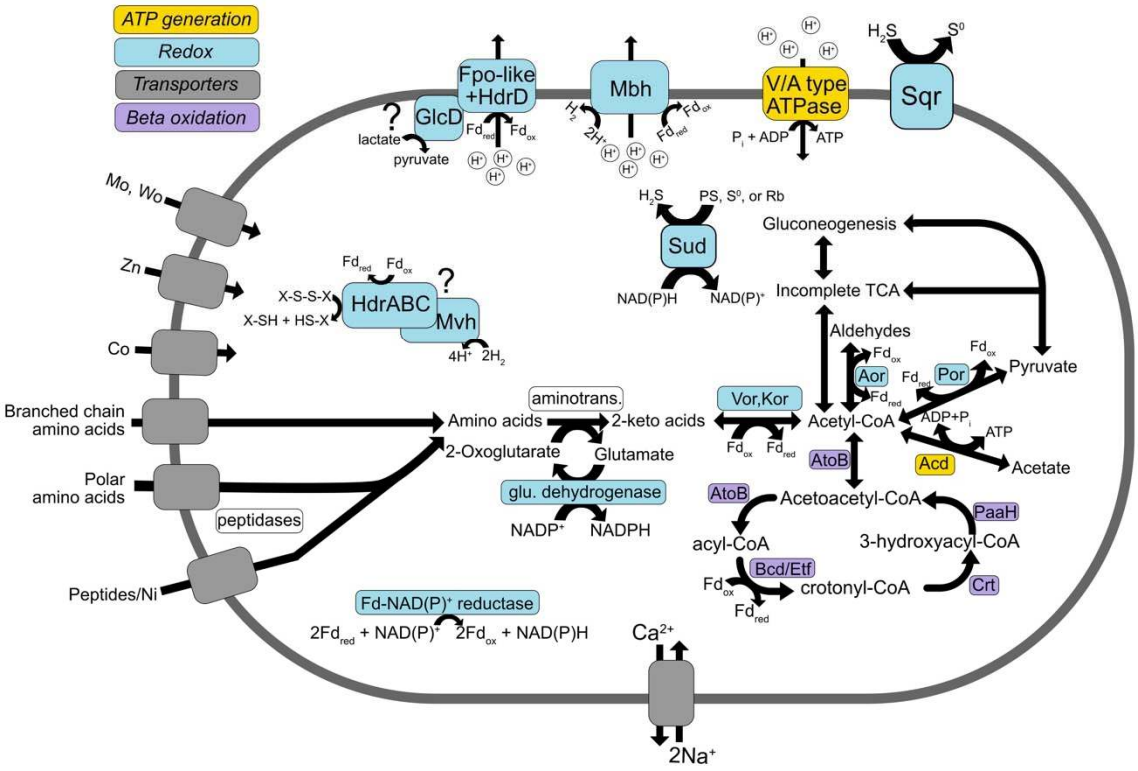


Figure 2. Proposed metabolic scheme for *Ca*. Lunaplasma lacustris based on genome reconstruction. *Ca*. L. lacustris putatively uptakes peptides or amino acids through transporters, then intracellular peptidases break down peptides. Aminotransferases and glutamate dehydrogenases convert amino acids into 2-keto acids, which can be oxidized to acetyl-CoA by various oxidoreductases (Ior, Vor, Kor), producing reduced ferredoxin (Fd_{red}). Oxidation of acetyl-CoA by Aor into aldehydes could serve as another source of Fd_{red}. Fpo-like subunits complexed with HdrD could drive a proton (or sodium) motive force using reducing power from Fd_{red}. A membrane-bound hydrogenase similarly could perform this reaction, coupling Fd oxidation to hydrogen production. A beta-oxidation pathway similar to that proposed in the *Ca*. Polytropus marinifundus, or substrate-level oxidation through Acd could be used for additional

875 reducing power or ATP synthesis. ORFs annotated as sulfide:quinone reductase (Sqr) and sulfide
876 dehydrogenase (SuDH) genes suggest sulfur compounds as a source of energy as well,
877 potentially including sulfide oxidation or S_0 or polysulfide (PS) oxidation.
878

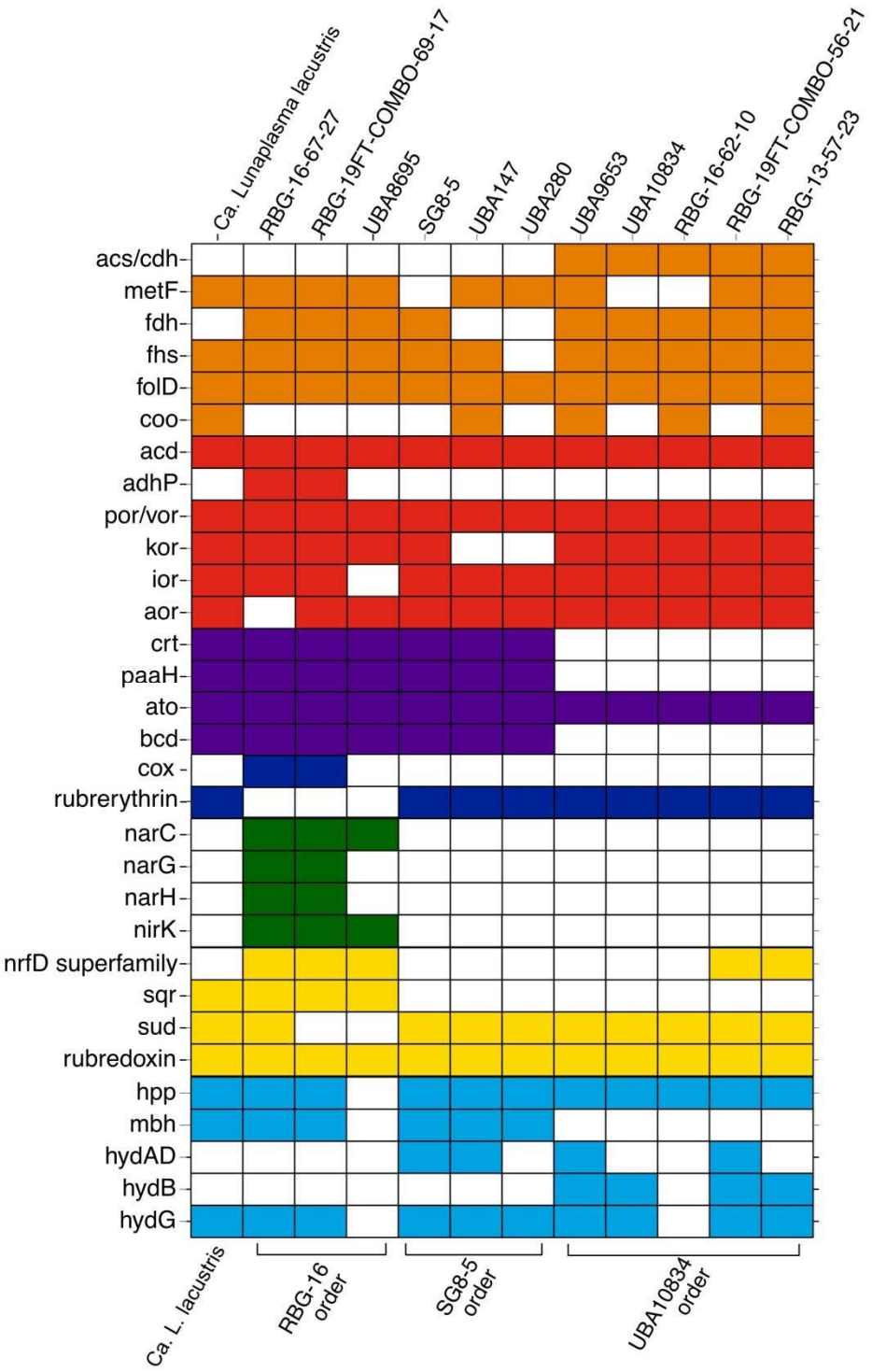
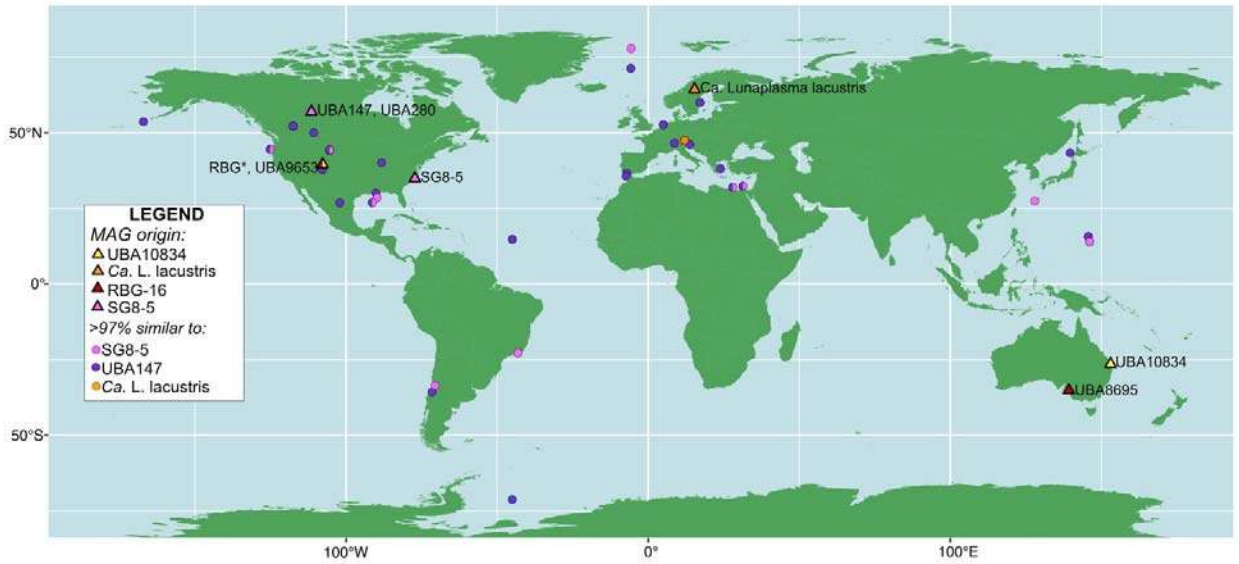


Figure 3. Key metabolic genes and their presence in the MAGs basal to the Methanomassiliicoccales. MAGs are grouped by their phylogenetic relationships, and putative gene assignments are grouped and colored by function or biogeochemical cycle (Orange: acetyl-

883 CoA pathway: red: fermentation and amino acid degradation; purple: beta oxidation; dark blue:
884 oxygen; green: nitrogen; yellow: sulfur; light blue: hydrogenases and pyrophosphatase). Gene
885 names and products are detailed in Supplemental table 8. For genes with multiple subunits,
886 presence was marked if at least half of the subunits were present.
887

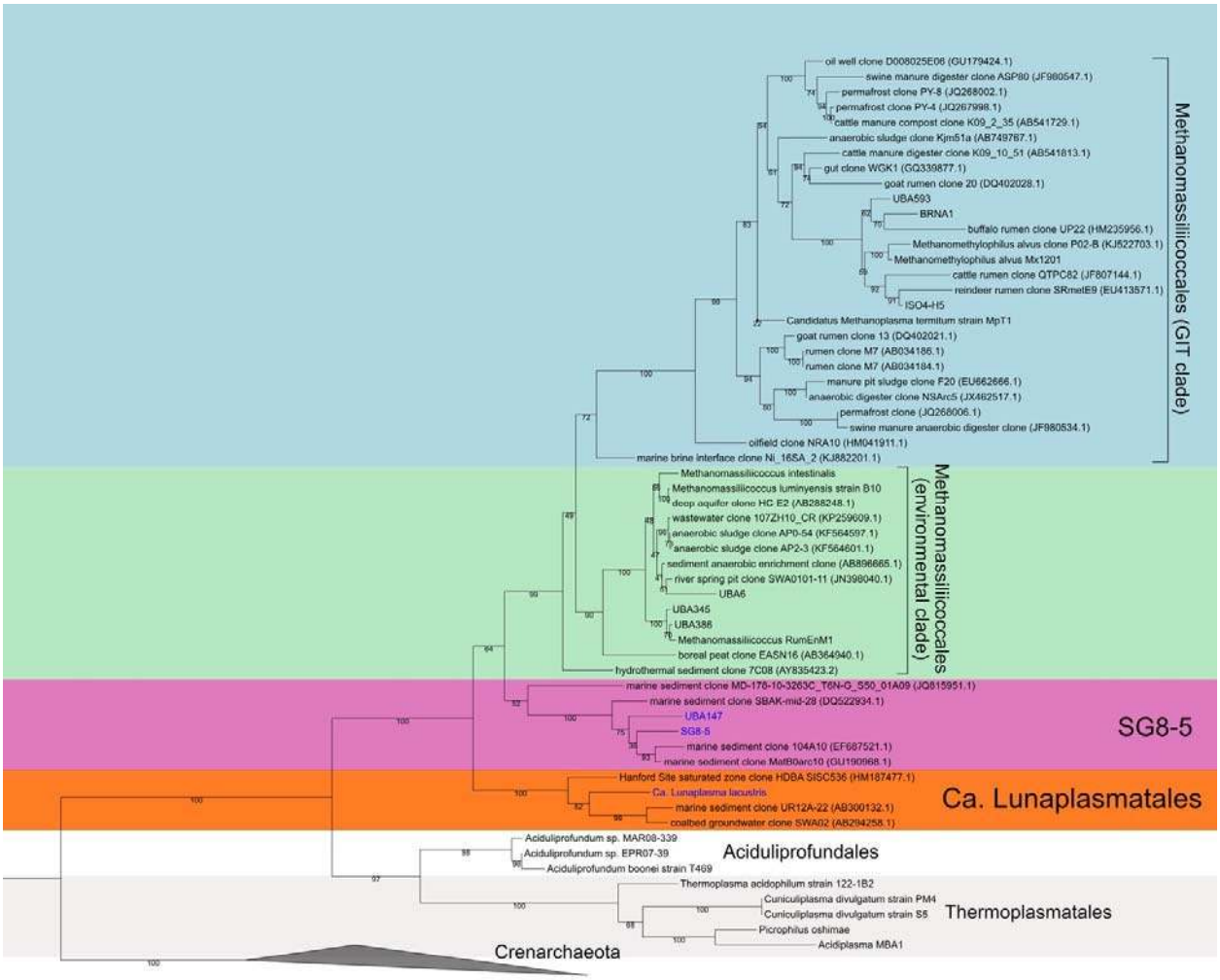
888



889

890 **Figure 4.** Map of MAGs (triangles) and 16S rRNA gene fragments (circles) in the NCBI
891 nucleotide (nt) database within 97% similarity to 16S rRNA genes recovered in three MAGs
892 from two orders: SG8-5 (SG8-5 and UBA147) and Lunaplasmatales (*Ca*. L. lacustris) (16S
893 rRNA gene sequences were not recovered from order RBG-16 or UBA10834 MAGs). All MAGs
894 and sequences were from environmental samples (that is, not host-associated). Shapes with
895 multiple colors indicate recovery of multiple orders from the same site. MAG IDs are next to
896 their locations. *The five MAGs beginning with "RBG" were binned from data generated from a
897 Rifle creek site in Colorado, United States.

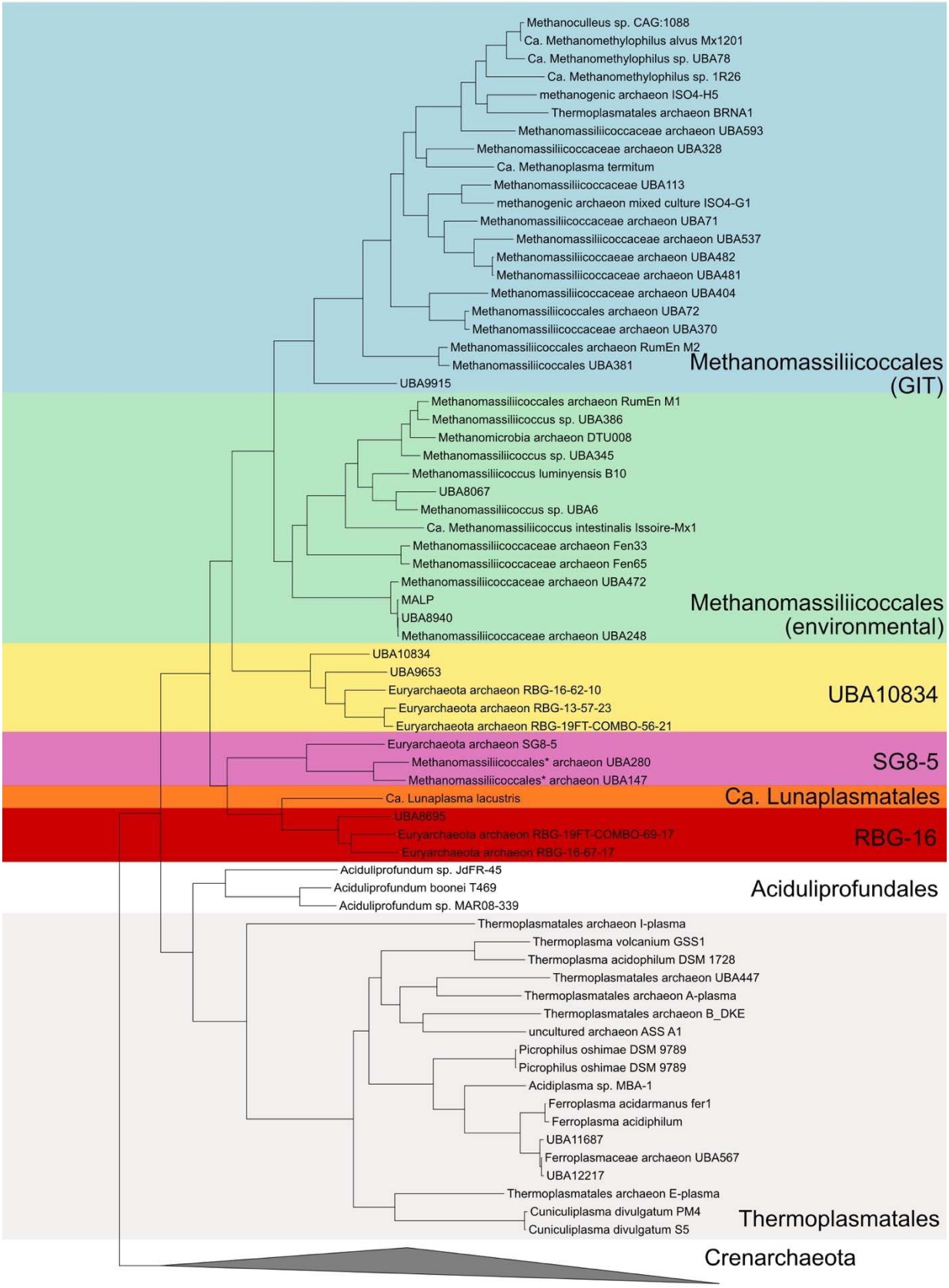
898



899

900 **Supplemental figure 1.** 16S rRNA gene phylogenetic tree, including the three 16S rRNA gene
901 sequences recovered from three MAGs in two orders basal to the *Methanomassiliicoccales* (blue
902 text): Lunaplasmatales (*Ca*. Lunaplasma lacustris) and SG8-5 (UBA147 and SG8-5). The
903 Crenarchaeota were used as an outgroup. Clades of *Methanomassiliicoccales* are labeled as in
904 Soellinger et al. (2016). Bootstrap values are calculated from 1000 bootstraps. Background
905 colors are as in Figure 1.

906



907

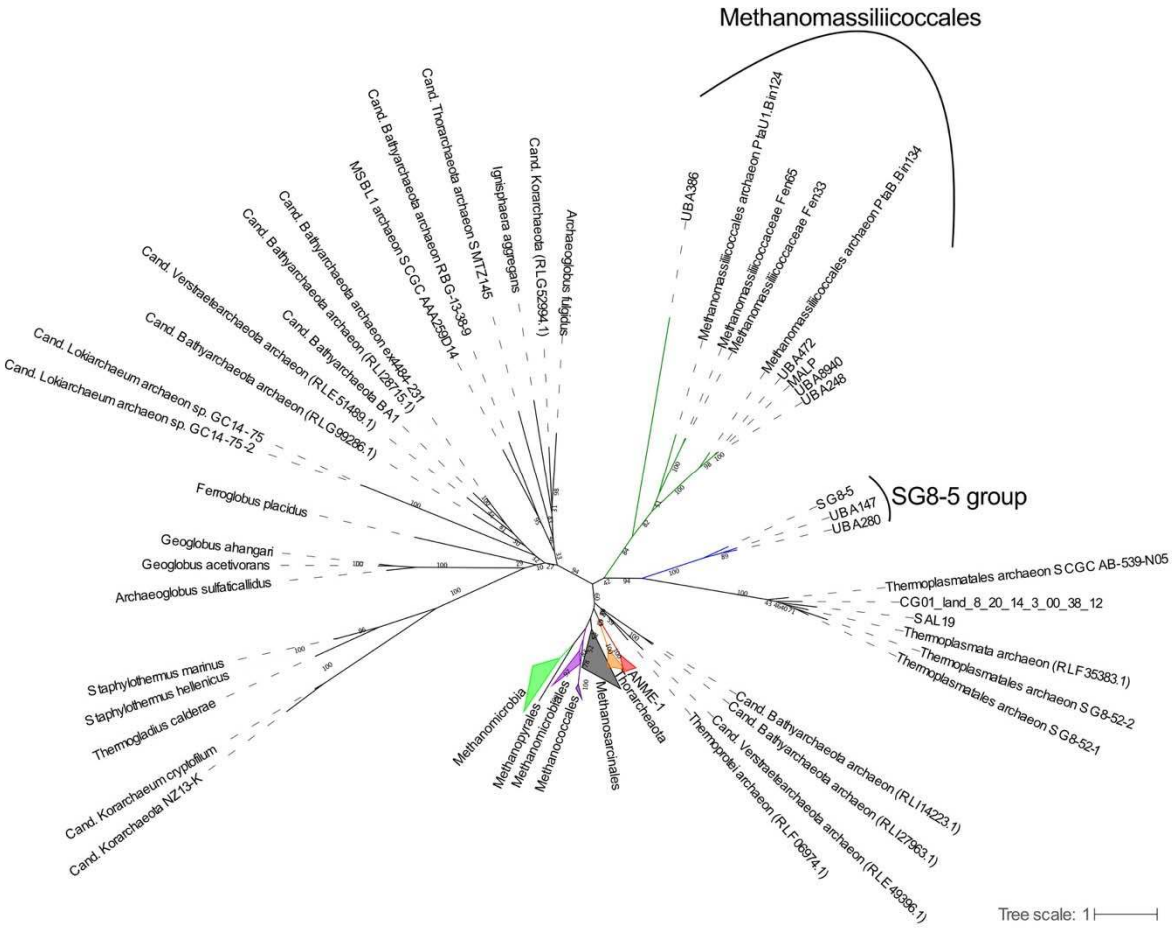
908

909 **Supplemental figure 2.** Phylogenetic tree of MAGs and genomes, based on relative

910 evolutionary divergence (RED), as calculated by the Genome Taxonomy Database toolkit

911 (GTDB-Tk). Crenarchaeota were used as the outgroup.

912



913

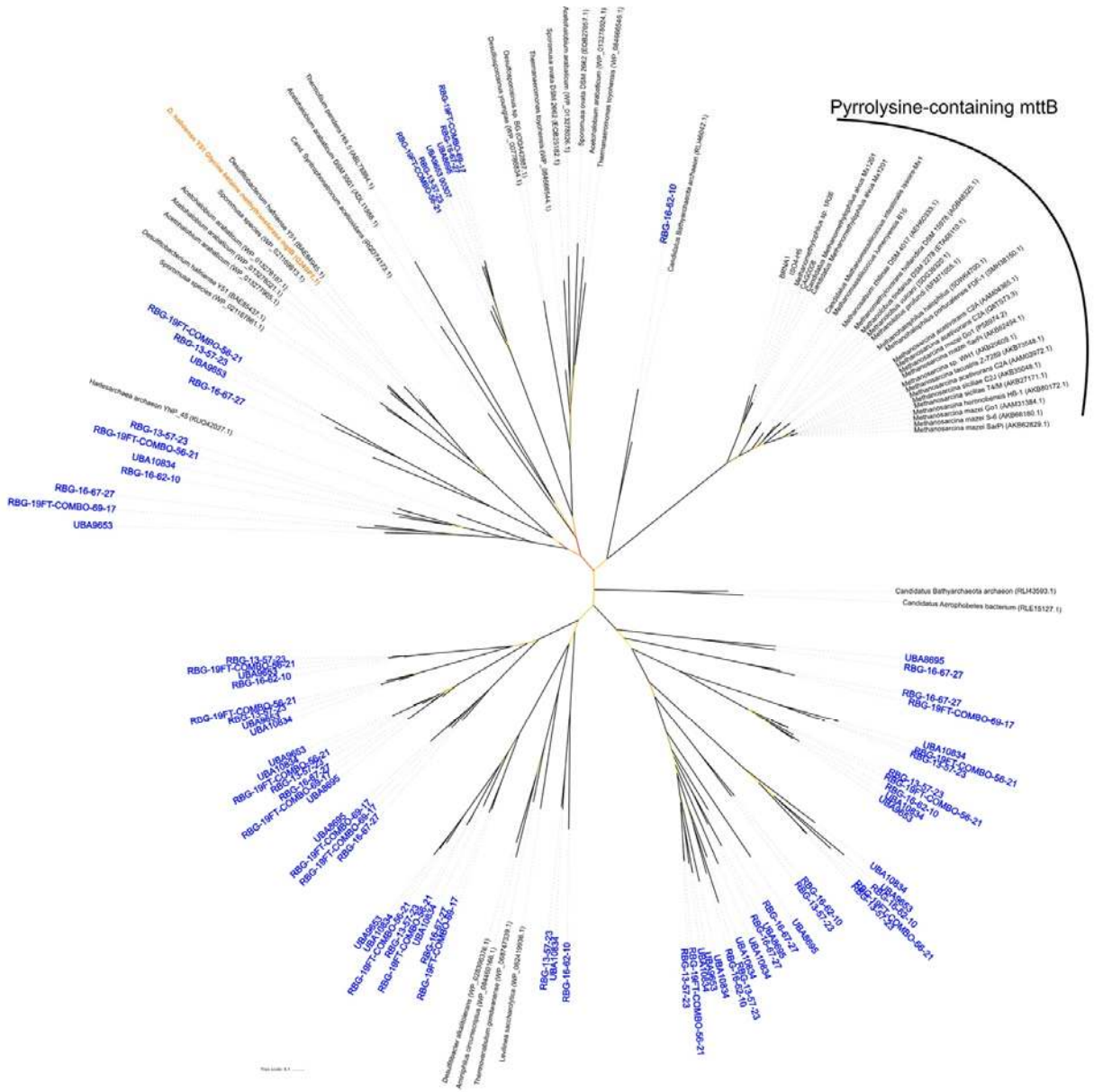
914

915 **Supplemental figure 3.** Tree of MtrH subunits, including those from *Methanomassiliicoccales*

916 and the SG8-5 order. Branches are colored to highlight various clades. Bootstrap support (out of

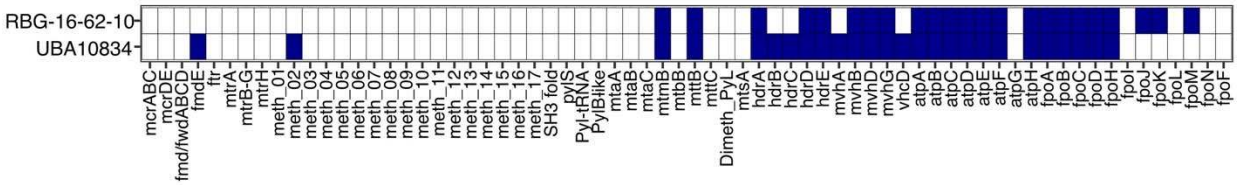
917 100 bootstraps) is shown along branches. Dashed lines extending branches do not represent

918 branch lengths, and are present to connect branch labels to branches for easier reading.



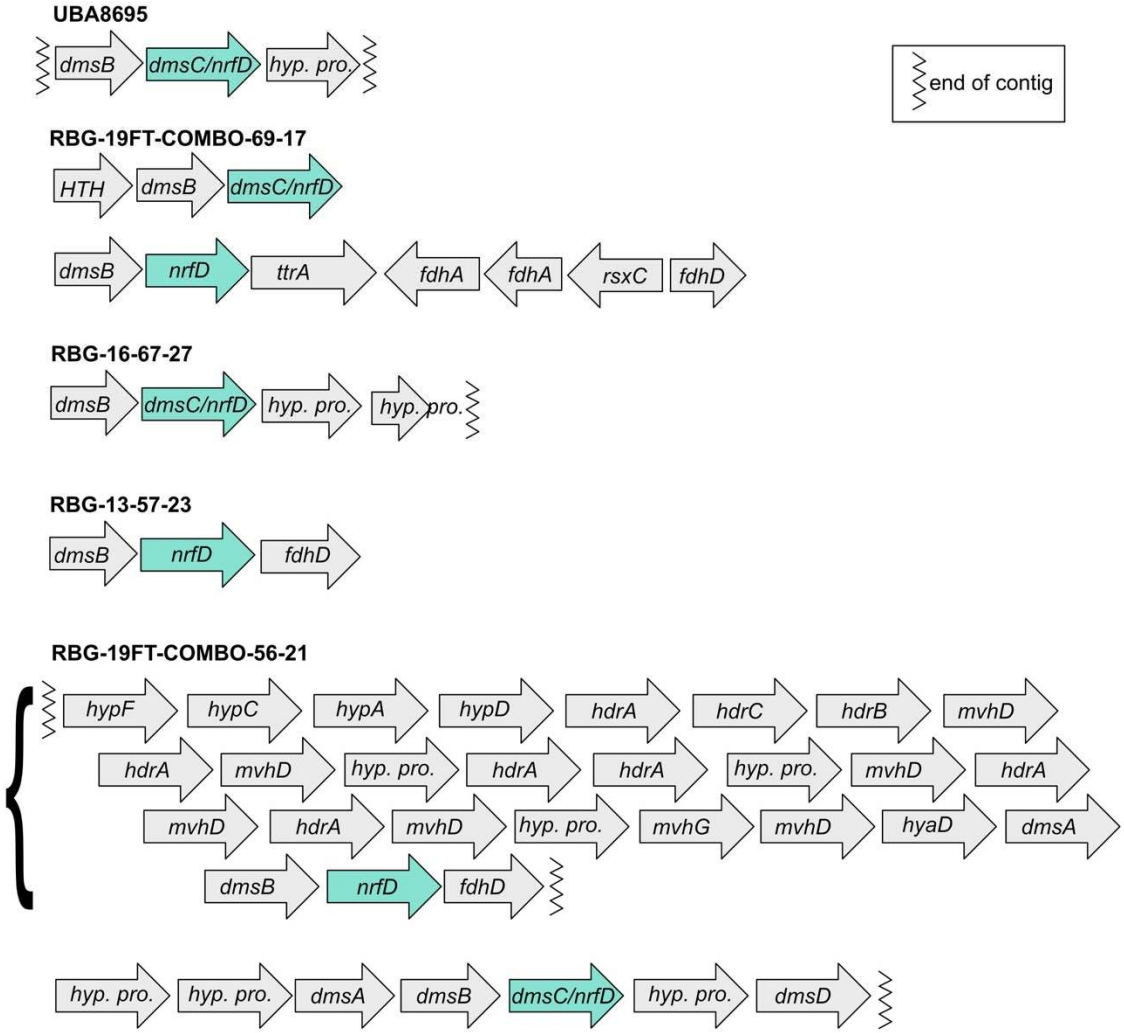
Supplemental figure 4. MttB superfamily tree including MttB sequences from euryarchaeal methanogens and *Methanomassiliicoccales*, glycine betaine transferases (part of the MttB superfamily but functionally distinct; in orange), and predicted MttB-like sequences recovered the MAGs basal to the *Methanomassiliicoccales* (all blue labels). Bootstrap support (out of 100 bootstraps) is indicated by branch color, displayed as a gradient, where black is the highest support (up to 100), red is the lowest support (5 or more), yellow is the midpoint (52). Dashed

926 lines extending branches do not represent branch lengths, and are present to connect branch
927 labels to branches for easier reading.
928
929
930



931
932

933 **Supplemental figure 5.** Presence of methanogenesis-related ORFs as in Figure 1 for two
934 UBA10834 order MAGs that did not have the minimum 8 of 16 RP required for inclusion in the
935 RP tree.



936

937

938 **Supplemental Figure 6.** Operons containing *nrfD* annotated genes from MAGs basal to the

939 *Methanomassiliicoccales*. Ends of contigs are marked with the zipper line. The bracket to the left

940 of the top RBG-19FT-COMBO-56-21 contig indicates that the four lines of arrows are putatively

941 one operon (*i.e.*, the four lines should be read as one continuous line).

942

943

944

Tables

Table 1. Genome statistics of MAGs from *Ca*. Lunaplasma lacustris and the orders RBG-16, UBA10834, and SG8-5. Completeness, contamination, contigs, and genome size (bp) were produced using CheckM. Sample type, metagenome origin, and reference were retrieved from NCBI or GTDB.

MAG name	Completeness	Contamination	Contigs	Genome Size	Sample type	Metagenome origin	Reference
Ca. Lunaplasma lacustris	95.2	4.8	91	2535897	sediment	Inre and Mellestra Harrsjon, Sweden	Emerson et al. *submitted* (on bioRxiv)
Euryarchaeota archaeon RBG_19FT_COMBO_69_17	96.67	6.34	449	2210690	sediment	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
Euryarchaeota archaeon RBG_19FT_COMBO_56_21	92.8	0.8	110	1868766	sediment	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
UBA10834	89.79	2.4	319	1705491	sediment	Noosa River estuary, SE Queensland, Australia	Parks *et al.* (2017) Nat. Micro.
Methanomassiliicoccales archaeon UBA147	89.6	3.2	80	1995084	waste water	Suncor tailings pond 6, Alberta, Canada	Parks *et al.* (2017) Nat. Micro.
UBA9653	88.67	0.8	500	1711095	ground water	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
Euryarchaeota archaeon RBG_13_57_23	83.8	0.8	108	1520809	sediment	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
Euryarchaeota archaeon RBG_16_67_27	77.47	0.4	144	1546722	sediment	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
Methanomassiliicoccales archaeon UBA280	76.8	2.4	46	1749312	waste water	Suncor tailling pond 5, Alberta, Canada	Parks *et al.* (2017) Nat. Micro.
Euryarchaeota archaeon RBG_16_62_10	74.45	1.6	971	1494907	sediment	Rifle, Colorado, USA	Anantharaman et al. (2016) Nat. Comms.
Euryarchaeota archaeon SG8-5	72.24	0.8	114	1247229	sediment	White Oak Estuary, North Carolina, USA	Lazar *et al.* (2017) ISME J.
UBA8695	70.76	0	317	1709476	soil	Murray Bridge, South Australia, Australia	Parks *et al.* (2017) Nat. Micro.

Supplemental table 1. Sources of genomes and MAGs used in this study. SRA source for MAGs retrieved from GTDB are from the source metagenomes of the MAGs.

Supplemental table 2. Detailed genome stats from CheckM for all genomes and MAGs.

Supplemental table 3. Genome statistics calculated by the MiGA webserver.

Supplemental table 4. Taxonomic classification of 16S rRNA gene sequences from UBA147, SG8-5, and *Ca*. Lunaplasma lacustris by RDP classifier, Silva SINA, and Blastn searches.

Supplemental table 5. Taxonomic classification and novelty of MAGs by MiGA and GTDBTk.

Supplemental table 6. MiGA calculated AAI comparison to genomes/MAGs in the NCBI prokaryotic database.

966

967 Supplemental table 7. AAI between MAGs and genomes here, computed by the envi-omics AAI
968 calculator.

969

970 Supplemental table 8. Genes used in Figures 1, 2, and 3.

971

972 Supplemental table 9. Comparisons of three 16S rRNA gene sequences recovered from
973 metagenome-assembled genomes (MAGs) in this study with 16S rRNA gene sequences
974 recovered from previous amplicon studies.

975

976 Supplemental table 10. Functional annotations from prokka, BlastKOALA, and InterProScan for
977 *Ca.* L. lacustris.

978

979 Supplemental table 11. Functional annotations from prokka, BlastKOALA, and InterProScan for
980 RBG-19FT-COMBO-69-17.

981

982 Supplemental table 12. Functional annotations from prokka, BlastKOALA, and InterProScan for
983 RBG-16-67-27.

984

985 Supplemental table 13. Functional annotations from prokka, BlastKOALA, and InterProScan for
986 UBA8695.

987

Supplemental table 14. Functional annotations from prokka, BlastKOALA, and InterProScan for SG8-5.

Supplemental table 15. Functional annotations from prokka, BlastKOALA, and InterProScan for UBA147.

Supplemental table 16. Functional annotations from prokka, BlastKOALA, and InterProScan for UBA280.

Supplemental table 17. Functional annotations from prokka, BlastKOALA, and InterProScan for RBG-19FT-COMBO-56-21.

Supplemental table 18. Functional annotations from prokka, BlastKOALA, and InterProScan for RBG-13-57-21.

Supplemental table 19. Functional annotations from prokka, BlastKOALA, and InterProScan for RBG-16-62-10.

Supplemental table 20. Functional annotations from prokka, BlastKOALA, and InterProScan for UBA9653.

Supplemental table 21. Functional annotations from prokka, BlastKOALA, and InterProScan for UBA10834.

1011