ORIGINAL ARTICLE

Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches

CS Lazar[1,2,3,5], BJ Baker[4,5], KW Seitz[4] and AP Teske[1]

[1]Department of Marine Sciences, University of North Carolina Chapel Hill, Chapel Hill, NC, USA; [2]Department of Geosciences, Organic Geochemistry Group, MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany; [3]Department of Aquatic Geomicrobiology, Institute of Ecology, Friedrich Schiller University Jena, Jena, Germany and [4]Department of Marine Science, University of Texas Austin, Marine Science Institute, Port Aransas, TX, USA

Genomic bins belonging to multiple archaeal lineages were recovered from distinct redox regimes in sediments of the White Oak River estuary. The reconstructed archaeal genomes were identified as belonging to the rice cluster subgroups III and V (RC-III, RC-V), the Marine Benthic Group D (MBG-D), and a newly described archaeal class, the Theionarchaea. The metabolic capabilities of these uncultured archaea were inferred and indicated a common capability for extracellular protein degradation, supplemented by other pathways. The multiple genomic bins within the MBG-D archaea shared a nearly complete reductive acetyl-CoA pathway suggesting acetogenic capabilities. In contrast, the RC-III metabolism appeared centered on the degradation of detrital proteins and production of H_2, whereas the RC-V archaea lacked capabilities for protein degradation and uptake, and appeared to be specialized on carbohydrate fermentation. The Theionarchaea appeared as complex metabolic hybrids; encoding a complete tricarboxylic acid cycle permitting carbon (acetyl-CoA) oxidation, together with a complete reductive acetyl-CoA pathway and sulfur reduction by a sulfhydrogenase. The differentiated inferred capabilities of these uncultured archaeal lineages indicated lineage-specific linkages with the nitrogen, carbon and sulfur cycles. The predicted metabolisms of these archaea suggest preferences for distinct geochemical niches within the estuarine sedimentary environment.
The ISME Journal (2017) **11,** 1118–1129; doi:10.1038/ismej.2016.189; published online 13 January 2017

Introduction

Particulate organic matter in the ocean is continuously sinking to the seafloor and accumulates in subsurface sediments, where it is assimilated, transformed and oxidized mainly by resident bacteria and archaea that constitute the global marine sedimentary biosphere (Kallmeyer *et al.*, 2012). Although Bacteria are generally thought to dominate prokaryotic communities in subsurface sediments, Archaea were shown to contribute significantly to marine sediment microbial communities and their biogeochemical cycles (Biddle *et al.*, 2006; Lipp *et al.*, 2008; Teske and Sorensen, 2008). On the basis of 16S rRNA gene surveys, benthic archaeal communities consist of distinct lineages with few to no cultured

representatives (Teske and Sorensen, 2008). Hence, the physiological capabilities and specific biogeochemical roles of uncultured clades of Archaea in marine sediment ecosystems remain poorly understood and represent a major research challenge.

Coastal estuaries, where sea water and riverine fresh water meet and mix, are particularly rich in nutrients in both the water and sediment columns. In estuarine environments, reactive organic matter accumulating in sediments fuels intense microbial activity (Poremba *et al.*, 1999). Microbial processes in estuarine sediments have been well-studied in the White Oak River (WOR) estuary, a drowned river valley that drains a swamp forest-dominated coastal plain of North Carolina, USA (Land and Paull, 2001). The WOR sediments have been thoroughly characterized with regard to sulfate reduction, methanogenesis and anaerobic methane oxidation (Martens and Goldhaber, 1978; Kelley *et al.*, 1990; Kelley *et al.*, 1995; Lloyd *et al.*, 2011). Recently, bacterial genomes—including representatives of two novel bacterial phyla—were reconstructed from different sulfate-reducing and methane-producing sediment

layers of the WOR estuary (Baker *et al.*, 2015). The sulfate-reducing zone (SRZ) hosted Deltaproteobacteria and uncultured Gammaproteobacteria with a potential for sulfur oxidation and nitrate/nitrite reduction; whereas the bacteria of the sulfate–methane transition zone (SMTZ) and the methane-producing zone (MPZ) were predominantly involved in fermentative organic carbon degradation.

The WOR estuary also represents a preferred habitat for uncultured benthic Archaea, as demonstrated by archaeal 16S rRNA gene and metagenomics surveys. In particular, the WOR estuary harbors a natural enrichment of the Miscellaneous Crenarchaeotal Group, now renamed Bathyarchaeota (Meng *et al.*, 2014), these archaea dominated the 16S rRNA gene pool of all analyzed sediment depths (Lazar *et al.*, 2015). Analysis of genomes belonging to four different subgroups of the Bathyarchaeota showed that they shared common pathways for protein degradation and acetogenesis, but suggested distinct ecological niches for the different subgroups (Lazar *et al.*, 2016). Genomes of two other uncultured WOR archaeal groups, the former South African Gold Mine Euryarchaeotal Group, now renamed Hadesarchaea (Baker *et al.*, 2016), and a novel archaeal phylum called Thorarchaeota (Seitz *et al.*, 2016), contained the genomic blueprint of CO and H_2 oxidation, and of elemental sulfur and thiosulfate reduction, respectively.

In this study, we used *de novo* assembly of random shotgun genomic libraries to reconstruct and identify archaeal genomic bins recovered from moderately to strongly reduced sediment layers, and their distinct geochemical niches in the WOR sediments. We also analyzed the reconstructed genomes for insights into the metabolic capabilities of these uncultured archaea. These reconstructed archaeal genomes were identified as belonging to the rice cluster subgroups III and V (RC-III, RC-V), the Marine Benthic Group D (MBG-D), and one newly described archaeal phylum-level group previously identified as Z7ME43, and renamed Theionarchaea.
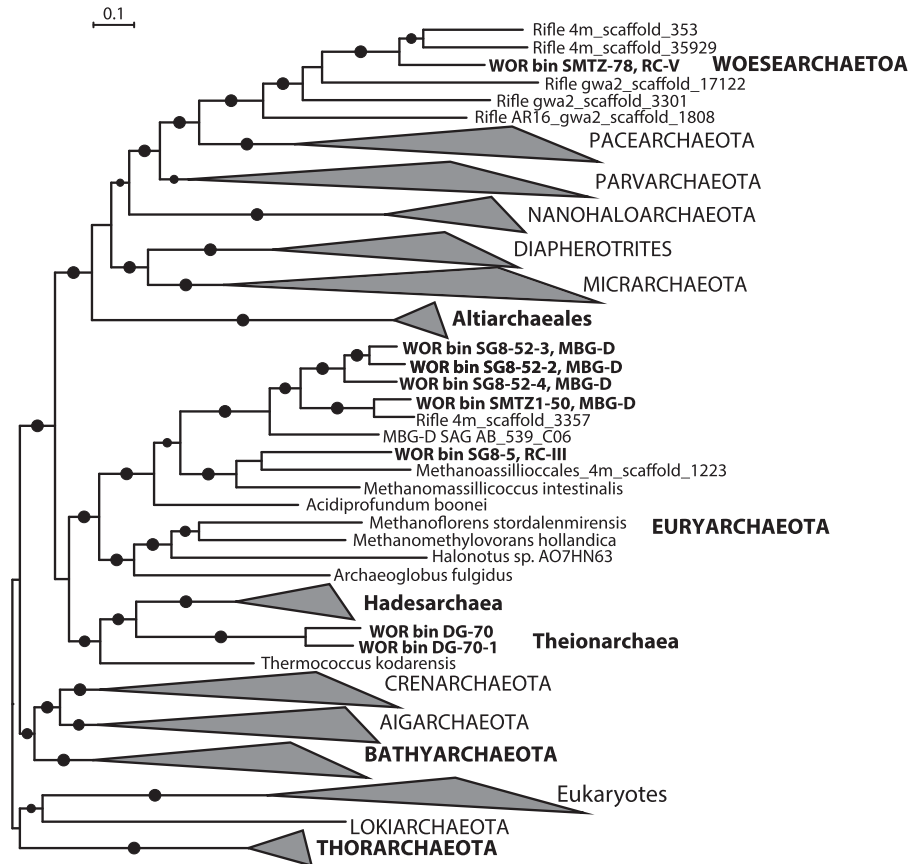
Materials and methods

Sampling, DNA extraction and sequencing
Six 1 m sediment plunger cores, each containing *ca.* 60 cm sediment, were collected at specific site coordinates in three mid-estuary locations of the WOR (North Carolina) in October 2010 (Lazar *et al.*, 2015). These sediment cores were previously used to characterize the archaeal 16S rRNA gene diversity for all three sites at selected depths, representing the SRZ, SMTZ and MPZ, based on pore water concentrations of methane, sulfate and sulfide (Lazar *et al.*, 2015). Subsequently, a wide range of archaeal and bacterial genomes was obtained by metagenomic analyses of these estuarine sediments (Baker *et al.*, 2015; Lazar *et al.*, 2016; Seitz *et al.*, 2016). Sediment cores were processed and DNA was extracted

and prepared as detailed previously (Lazar *et al.*, 2015). Briefly, a focused-ultrasonicator (Covaris, Woburn, MA, USA) was used to shear 100 ng of DNA to 270 bp. The sheared DNA fragments were size selected for a preferred size of 270 bp using SPRI beads (Beckman Coulter, Brea, CA, USA). The selected fragments were then end-repaired, A-tailed and ligated with Illumina compatible adapters (IDT, Inc., San Jose, CA, USA) using KAPA-Illumina library creation kit (KAPA Biosystems, Wilmington, USA).

Genomic assembly, binning and annotation
Illumina (HiSeq 2500 PE 2x150) shotgun genomic reads were screened against Illumina artifacts (adapters, DNA spike-ins) using a sliding window with a kmer size of 28 and a step size of 1 using cutadapt (v1.1, https://pypi.python.org/pypi/cutadapt). Reads with three or more N's or with average quality score of less than Q20 and a length < 50 bps were removed. Screened reads were trimmed from both ends using a minimum quality cutoff of 5 using Sickle (v1.33 https://github.com/najoshi/sickle). Trimmed, screened, paired-end Illumina reads were assembled using IDBA-UD (version 1.0.9; Peng *et al.*, 2012) with the following parameters (pre_correction –mink 55 –maxk 95 –step 10 –seed_kmer 55). To maximize assembly reads from different sites were co-assembled. The SRZ assembly was a combination of high-quality reads (474 179 948 with an average read length 148 bp) from sites 2 (8–12 cm), and 3 (8–10 cm). The SMTZ assembly was generated from a combination of high-quality reads (698 574 240, average read length 143 bp and average insert 274 bp) from sites 2 (30–32 cm) and 3 (24–28 cm). The MPZ assembly was generated from high-quality reads (378 027 948, average read length 124 bp and average insert 284 bp) of site 1 (52–54 cm). The co-assembled samples have been grouped previously by LINKTREE analysis, a form of constrained cluster analysis involving a divisive partition of the sediment samples with their archaeal communities into increasingly specific groups (Lazar *et al.*, 2015). As we were not able to co-assemble all three of the samples from the SMTZ due to computational limits (computer memory maxed), an additional assembly was generated from the third sample (site 1 26–30 cm) from high-quality 345 710 832 reads (average length of 129 bp and average insert 281 bp). The contigs from this SMTZ sample (site 1) were co-binned with the combined assembly of the other two, mutually closely related SMTZ samples of sites 2 and 3 (Lazar *et al.*, 2015). This sequence of binning the most closely related sample pair before adding the more distinct third sample, resulted in differentiating some closely related but distinct bins, SM1-50 from site 1 and SM23-78 from the site 2/3 co-assembly. Contigs with genes of particular interest were checked for chimeras by looking for dips in coverage within read mappings.

Figure 1 Archaeal phylogenetic tree inferred from 16 syntenous ribosomal protein genes present within genomic bins from the sediment metagenomic assemblies. Ribosomal protein genes list can be found in Supplementary Table S5. Each individual sequence in bold is from a single genomic bin. Phyla names in bold represent other lineages recovered from the WOR and are included in the overview shown in Figure 3. The taxonomic identity of these lineages was confirmed by 16S rRNA gene phylogeny (Supplementary Figure S1). Large circles show >90 percent Bayesian posterior probabilities, and smaller circles show >70 percent. The scale bar indicates 10% estimated phylogenetic divergence.

Initial binning of the assembled fragments was done using tetra-nucleotide frequencies signatures using 5 kb fragments of the contigs (Dick *et al.*, 2009). The emergent self-organizing maps (Ultsch and Moerchen, 2005) were manually delineated based on distinct clusters that were separated by visible valleys within the emergent self-organizing maps, see Baker *et al.*, 2015 for the maps used in this study. The completeness and contamination of the genomic bins was then estimated by counting universal ribosomal proteins and single-copy genes using CheckM (v1.0.5; Parks *et al.*, 2014). Genomic bins that were found to contain contaminants were re-examined with a second round of binning, using differential coverage. Candidate contaminants were identified by plotting the coverage of all the contigs in the bin between different sites; the coverage outliers that emerged in these plots were then removed from the bins. Duplications of single-copy genes identified with checkM were also considered during the cleaning of the genomic bins. Single-copy genes with extremely low or high coverage were also considered likely contaminant contigs, and removed. Coverage was determined by recruiting reads (from each individual library/sample) to scaffolds by

BLASTN (bitscore >75), which was then normalized to the number of reads from each library in order to effectively compare sample coverages. Comparison of the genomic bins, each of the 5 kb sub-portions of the contigs, was used to assess the accuracy of the binning. If the contig was >15 kb then the contig was assigned to the bin where the majority of the 5 kb sub-portions were assigned. Genes were called and putative function was assigned using the JGI IMG/MER system (Markowitz *et al.*, 2012). The predicted gene and protein sequence data supporting the results of this article are being made available in NCBI Genbank under the accession numbers LSSB00000000 to LSSJ00000000.

Phylogenetic analyses
The concatenated ribosomal protein tree was generated using 16 genes that have been shown to have limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19; Sorek *et al.*, 2007). The reference data sets were derived from the Phylosift database (Darling *et al.*, 2014), which additional sets from the Joint Genomic Institute IMG database (Castelle *et al.*, 2013). Scaffolds, from bins

Table 1 General bin characteristics and CheckM completeness analysis of the genomic bins

Bin ID	SM1-50	SG8-52-1	SG8-52-2	SG8-52-3	SG8-5	SG8-52-4	DG-70	DG-70-1	SM23-78
Completeness	94.78	71.62	61.51	85.32	72.59	77.38	72.49	82.44	82.41
Contamination	3.24	7.20	2.38	2.78	0.79	8.33	7.47	7.55	0.93
GC (%)	39	31	31	31	56	31	41	41	37
Total length of bin (bp)	2 091 705	1 914 472	1 355 901	1 909 404	1 247 229	2 246 471	3 559 548	4 794 391	1 480 827
Longest scaffold (bp)	111 154	93 713	68 198	131 306	53 836	104 598	66 889	144 352	96 136
Number of scaffolds	94	79	68	84	111	91	177	194	47
Lineage	MBG-D	MBG-D	MBG-D	MBG-D	RC-III	MBG-D	Z7ME43	Z7ME43	RC-V

CheckM completeness was determined using multiple archaeal gene markers. High contamination is representative of multiple marker hits within a genomic bin.

in this study, containing <50% of the 16 selected ribosomal proteins were not included in the analyses. We searched NCBI to include reference amino acid sequences for phylogenetic analyses. The trees shown in amino acid alignments of the concatenated ribosomal proteins (Figure 1) were generated using MAFFT (v1.3.3; Katoh *et al.*, 2002) and trimmed to 1904 residues using BMGE (v2.0; Criscuolo and Gribaldo, 2010; with the following settings: -m BLOSUM30 –g 0.5). The curated alignments were then concatenated for phylogenetic analyses, and then phylogeny was generated by running the RAxML (maximum likelihood) tool (version 2.2.1) with PROTGAMMA rate distribution model and CAT substitution model. Bayesian posterior probabilities were generated using MRBAYES (v2.2.2; Huelsenbeck and Ronquist, 2001) for 120 000 generations.

The phylogenetic archaeal 16S rRNA tree was calculated with ARB (Ludwig *et al.*, 2004); using the neighbor-joining method based on Jukes–Cantor distances. The 16S rRNA gene sequences were aligned using the SINA webaligner, available online at http://www.arb-silva.de/ (Pruesse *et al.*, 2012). Extremely short 16S rRNA sequences (<130 bp) that distorted the phylogeny by introducing predominantly unknown positions into the alignment were assigned phylogenetically by closest match (>99% similarity) obtained by BLASTn on the NCBI database. The phylogenetic trees of the functional genes (hydA, hydB, hydD, hydG and cdhA) were calculated in MEGA4 (Tamura *et al.*, 2007) using neighbor-joining method, and based on amino acid sequence alignments. The amino acid sequences were aligned using ClustalW (v1.4).

Results and discussion

Genome reconstruction of WOR archaea
De novo genomic assembly and binning of three sediment cores from three sites situated mid-estuary of the WOR, resulted in the reconstruction of nine draft archaeal genomes (Table 1) from the distinct redox regimes of the SRZ, SMTZ and MPZ (Lazar *et al.*, 2015). Phylogenies based on concatenated ribosomal proteins (Figure 1) and 16S rRNA genes (Supplementary Figure S1) showed that these genomic bins belong to the RC-III, RC-V MBG-D and Z7ME43 archaeal clades. Based on the presence of single-copy genes, these genomes range from partially to near complete (61–94% complete, Table 1). Some of the bins contain fragments from duplication of single-copy genes, which may represent strain-level variants that could not be distinguished using differential coverage. Therefore, these contigs were retained in the genomic bins. Binning was manually curated by generating phylogenetic trees from markers on contigs and by accounting for differences in coverage, to avoid contamination between the subgroups. Four of the reconstructed genomes (SG8-52-1, -2, -3 and -4) were initially binned together based on tetra-emergent self-organizing maps clustering, due to their high similarity, and then were separated based on differential coverage. All the genomic bins show considerable size variation, with RC-III being 1.25 Mb, RC-V being 1.48 Mb, MBG-D averaging 1.9 Mb and Z7ME43 averaging 4.2 Mb (Table 1).

Archaea in the SRZ
The shallow SRZ layers of the WOR sediments were characterized by high sulfate concentrations (*ca*. 12–16 mM) coinciding with barely detectable sulfide and methane concentrations (Lazar *et al.*, 2015). Here, bins belonging to two *Thermoplasmatales* clades within the *Euryarchaeota* were identified. The RC-III archaea, which were first identified in anoxic incubations of flooded rice roots (Grosskopf *et al.*, 1998), are represented by the 72% complete genomic bin SG8-5 (Table 1). The MBG-D archaea, which were initially detected in shallow sediments of the Atlantic continental slope (Vetriani *et al.*, 1999), are represented by four closely related genomic bins SG8-52-1, -2, -3 and -4 that are 73, 61, 85 and 77% complete, respectively (Table 1).

This metagenomic analysis suggests that the metabolisms of the RC-III and MBG-D archaea are centered on the degradation of extracellular detrital proteins (Supplementary Figures 2 and 3), consistent with previous studies (Kemnitz *et al.*, 2005; Lloyd *et al.*, 2013). The MBG-D bins have genes encoding secreted extracellular proteases such as clostripain (Merops family C11) and gingipain (Merops family C25), interpain (Merops family C10) and legumain

(Merops family C13) cysteine peptidases. No genes encoding extracellular proteases were detected in the RC-III bin, with the caveat that these genes could be located in the missing parts of this incomplete genome. Both RC-III and MBG-D bins contain genes encoding di- and oligo-peptides membrane transporters, intracellular aminopeptidases and proteases involved in the breakdown of amino acids, as well as pyruvate:ferredoxin and indolpyruvate:ferredoxin oxidoreductases converting 2-oxo acids to acyl-CoA and CO_2 using ferredoxin as electron carrier (Supplementary Figures 2 and 3; Supplementary Tables S1 and S2). In addition, the MBG-D bins encode an oligopeptide transporter system, a possible adaptation for utilizing small amino acids, depending on environmental conditions. Indeed this gene was shown to be regulated and enhanced by the presence of small peptides in sulfur-depleted environments (Wiles *et al.*, 2006). The MBG-D genomes also contain genes encoding enzymes involved in assimilatory sulfate reduction to sulfite (Supplementary Figure 3), the first steps of cysteine and methionine biosynthesis. In contrast to frequently found peptide transporters, genes involved in carbohydrate uptake and assimilation were not detected in the MBG-D bins, suggesting that peptides are the sole carbon sources for MBG-D.

Both the RC-III and MBG-D bins contain genes encoding subunits of a Ni, Fe hydrogenase type 3 and genes encoding an ATP synthase (Supplementary Figures 2 and 3). Thus, the RC-III and MBG-D archaea might possess a simple energy-conserving respiratory system similar to the one described in *Pyrococcus furiosus* (Sapra *et al.*, 2003). In this system, reduced ferredoxin is used as an electron donor and protons are used as terminal electron acceptors, and production of H_2 is directly coupled to ATP synthesis via proton translocation. Consistent with this scenario, a combined qPCR and culturing study (Kemnitz *et al.*, 2005) demonstrated no growth stimulation in an RC-III archaeal enrichment after addition of various electron acceptors (NO^{3-}, SO_4^{2-}, Fe(III), S^0, $S_2O_3^{2-}$ and fumarate). In addition, the RC-III bin contains 8 of the 11 subunits of a NADH dehydrogenase complex I (without a NADH-binding module), all subunits of a heterodisulfide reductase HdrABC as well as subunits of a F420-non-reducing hydrogenase MvhADG (Supplementary Figure 2). It has been suggested that the 11 subunit NADH dehydrogenase complex could recycle ferredoxin by accepting electrons from reduced ferredoxin (Battchikova *et al.*, 2011), possibly provided by the HdrABC/MvhADG hydrogenase complex (Castelle *et al.*, 2013). In non-methanogenic archaea, the HdrABC/MvhADG complex could couple the reduction of ferredoxin with H_2 to the reduction of an unknown disulfide (Mander *et al.*, 2004; Castelle *et al.*, 2013), hence yielding reduced ferredoxin.

The RC-III bin contains a pyruvate-formate lyase as well as a pyruvate-formate lyase activating enzyme (Supplementary Figure 2). This enzyme catalyzes the conversion of pyruvate, obtained by protein degradation, to acetyl-CoA and formate, and its expression is typically induced and enhanced under anaerobic conditions (Sawers and Bock, 1988). The acetyl-CoA produced during protein degradation can be converted to acetate and ATP via an ATP-producing acetyl-CoA synthetase. As no genes involved in fermentation were detected (for example, alcohol or lactate dehydrogenases), the originally suggested fermentative metabolism for the RC-III archaea (Kemnitz *et al.*, 2005) could not be confirmed. However, the MBG-D archaea could ferment proteins to ethanol as their bins contain genes encoding an ATP-producing acetyl-CoA synthetase, commonly found in peptide-degrading archaea (Sapra *et al.*, 2003), an aldehyde:ferredoxin oxidoreductase and an alcohol dehydrogenase.

Typically, organisms growing on amino acids require an Embden–Meyerhoff–Parnas gluconeogenesis pathway (EMP) to provide carbon sources for biosynthesis of cellular building blocks (Ng *et al.*, 2000). Both the RC-III and MBG-D bins include genes encoding a partial EMP. Specifically, the *fbp* gene encodes the unidirectional gluconeogenetic fructose 1,6 bisphosphatase; which produces fructose 6-phosphate for further use in anabolic pathways (Supplementary Figures 2 and 3; Supplementary Tables S1 and S2).

An additional near-complete pathway providing anabolic carbon sources in the MBG-D archaea is the reductive acetyl-CoA (Wood–Ljungdahl) pathway using tetrahydrofolate as a cofactor (Supplementary Figure S3 and Supplementary Table S2). The gene *cdhA,* encoding the alpha subunit of the CO dehydrogenase/acetyl-CoA synthetase, was detected in the MBG-D bins (Supplementary Table S2). A phylogenetic tree based on amino acid sequences placed the MBG-D CdhA sequences within a cluster containing the heterotrophic, acidophilic and sulfur-reducing archaeon *Aciduliprofundum* (Supplementary Figure S4), a sister lineage of the MBG-D archaea within the *Thermoplasmatales*. The *Aciduliprofundum* sequences are annotated as [Fe4-S4] metal clusters which were shown to be present in the active sites of the CO dehydrogenase/acetyl-CoA synthetase of anaerobic bacteria (Dobbek *et al.*, 2001). Thus, the low sequence similarity of the *cdhA* MBG-D genes to other versions of the *cdhA* gene suggests a modified CdhA protein in the MBG-D archaea.

So far, the existence of Wood–Ljungdahl genes in MBG-D does not provide conclusive evidence for autotrophy. As the tricarboxylic acid (TCA) cycle is complete in the MBG-D bins, it would produce citrate, oxoglutarate or oxaloacetate for biosynthesis of amino and/or fatty acids; the byproduct CO_2 could be recycled in the Wood–Ljungdahl pathway (Schuchmann and Müller, 2014). To date no cultured anaerobic representative of the *Thermoplasmatales* has been shown to use the reductive acetyl-CoA pathway for carbon fixation. However, *Ferroplasma*, a cultured aerobic *Thermoplasmatales*, has been predicted to use a

modified Wood–Ljungdahl pathway with an aerobic CO dehydrogenase (Cárdenas *et al.*, 2009).

The MBG-D bins harbor numerous genes encoding one- and two-component systems (sensory protein and response regulators) that are involved in sensing and responding to a variety of changes in the surrounding environment; for example, osmolarity, oxidative stress or nutrient availability (Supplementary Figure S3 and Supplementary Table S2). In addition to regulatory systems for chemotaxis, the MBG-D genomes possess genes involved in archaeal flagellin proteins and flagella assembly (Patenge *et al.*, 2001; Chaban *et al.*, 2007; Supplementary Table S2). Hence, the MBG-D archaea are predicted to be motile, and capable of chemotactic response towards different environmental factors. These genomes also contain a gene encoding a CRISPR exonuclease which might confer resistance to phages (Barrangou *et al.*, 2007), and the *uspA* gene encoding a small cytoplasmic protein that is induced during starvation periods when substrate limitation inhibits the growth rate (Nystrom and Neidhardt, 1994).
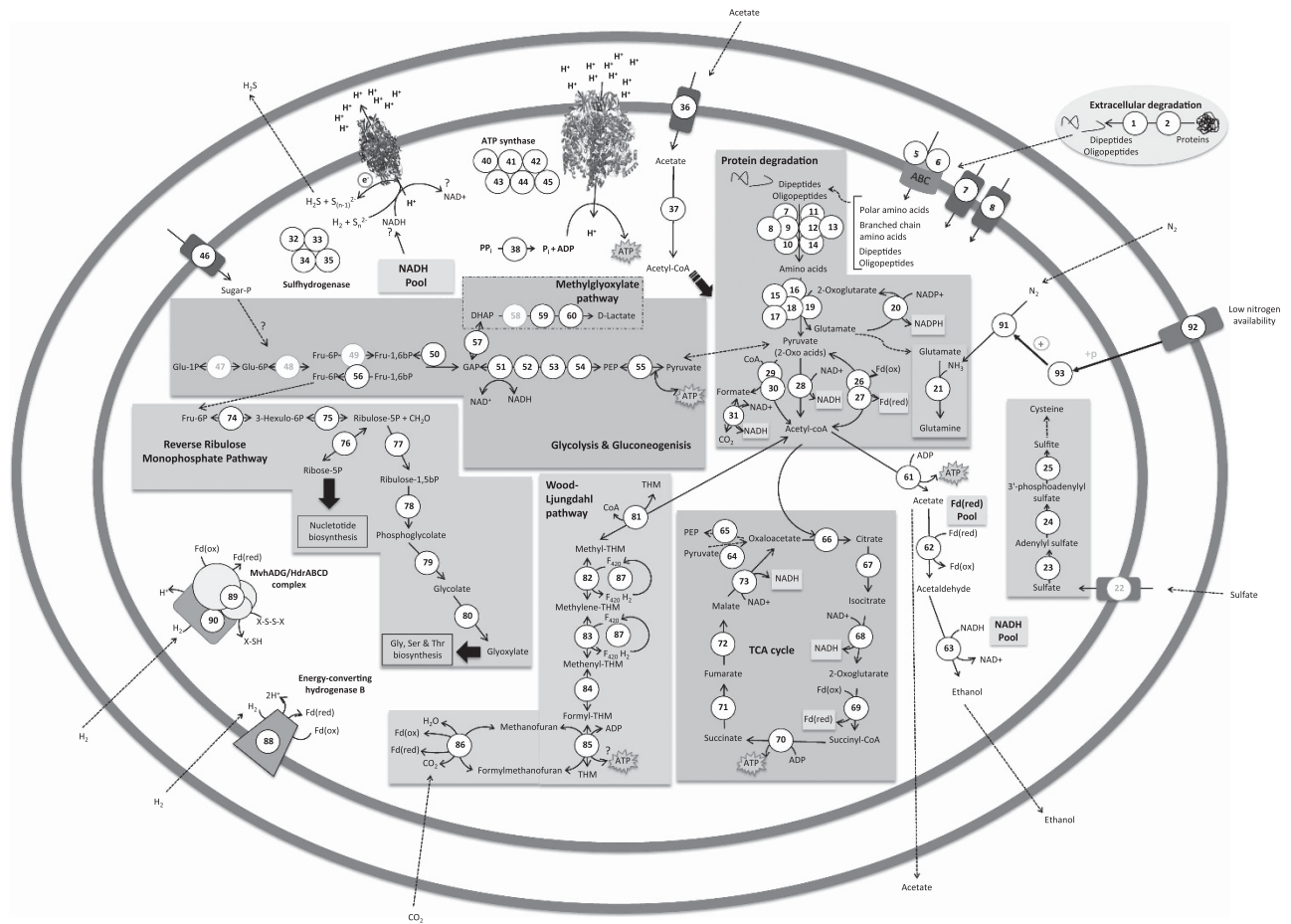
Archaea in the SMTZ

Bins belonging to two archaeal clades—the RC-V (bin SM23-78, 82.4% complete; Table 1) and an additional MBG-D (bin SM1-50, 94.8% complete; Table 1) distinct from those recovered from the SRZ —were identified from the SMTZ layers of the WOR sediments, where overlapping pore water sulfate and methane gradients coincided with a sulfide peak (Lazar *et al.*, 2015). Phylogenetic analyses based on concatenated ribosomal proteins (Figure 1) and 16S rRNA genes (Supplementary Figure S1) place the WOR RC-V genomic bin within the Woesearchaeota phylum (Castelle *et al.*, 2015). The core metabolism of the WOR RC-V archaea appears to be centered on carbohydrate fermentation, whereas the SMTZ MBG-D core metabolism appears to be centered on the degradation of extracellular detrital proteins. In comparison with the MBG-D bins from the SRZ, genes encoding interpain and legumain cysteine peptidases were not detected in the SMTZ bin; and both the gluconeogenetic EMP and the TCA cycle appear to be incomplete (Supplementary Table S2), with the caveat that the pathway gaps could be artifacts resulting from incomplete genomic bins.

The RC-V bin contains two genes encoding potential extracellular enzymes involved in hydrolyzing plant-derived polymeric carbohydrates to smaller oligosaccharides and glucose, that is, an exo-beta-1,3-glucanase and a glucan-1,4-alpha-glucosidase (Supplementary Figure S5 and Table S3). The exo-beta-1,3-glucanase has been shown to use laminarin, storage polysaccharides found in brown algae, as specific substrate (Bara *et al.*, 2003). Furthermore, the RC-V archaea have the potential to transport these extracellular sugars into the cell as the RC-V bin contains genes encoding an ABC-type sugar transport system and a sugar permease. Once

transported into the cytoplasm, intracellular polysaccharides may undergo further breakdown to glucose-1P or fructose-6P via the galactose and/or mannose degradation pathway (Supplementary Figure S5). The glucose-1P and/or fructose-6P can then enter the glycolysis pathway, producing ATP and NADH. The RC-V bin contains all genes constituting the EMP pathway (Supplementary Figure S5) including the *pfk* and *pyk* genes encoding a phosphofructokinase and a pyruvate kinase which operate exclusively in the catabolic direction (Bräsen *et al.*, 2014). The end product of glycolysis, pyruvate, can then be converted to acetyl-CoA via a pyruvate: ferredoxin oxidoreductase using ferredoxin as electron carrier (Supplementary Figure S5 and Supplementary Table S3). Genes encoding proteins involved in cellular respiration and in the TCA cycle were not detected; hence the WOR RC-V archaea might be strict fermenters relying on carbohydrates as sole source of energy and carbon. The RC-V bin also contains the *pta* and *ackA* genes encoding a phosphate acetyltransferase and an acetate kinase, catalysing the conversion of acetyl-CoA to acetate via acetyl-P, and producing ATP. This pathway potentially used by the RC-V archaea leading to acetate production, was thought to be unusual for Archaea as the acetate-forming archaea are suggested to mainly use the acetyl-CoA synthetase pathway (Schäfer *et al.*, 1993). However, an acetate kinase was detected in the methanogen *Methanosarcina barkeri*, in which cell-free enzyme extracts showed activity in the direction of acetate formation (Smith and Lequerica, 1985). Recently, the ack gene was detected in Bathyarchaeota, and the expressed and purified protein showed reversible catalysing activity; suggesting that the Bathyarchaeota can consume and/or produce acetate through the pta-ack pathway (He *et al.*, 2016). Recycling of oxidized ferredoxin and NAD+ could occur via ethanol fermentation (Köpke *et al.*, 2010), as the RC-V bin has genes encoding an aldehyde:ferredoxin oxidoreductase and an alcohol dehydrogenase (Supplementary Figure S5).

Theionarchaea in the MPZ

Two bins (DG-70 and DG-70-1, 74.5 and 83% complete; Table 1) belonging to an as-yet undescribed deeply branched class of the Euryarchaeota (Figure 1), were recovered from the MPZ layers (52–54 cmbsf) of the WOR sediments, where methane concentrations near 0.5 mM were measured, no sulfate was detected and sulfide concentrations were low (Lazar *et al.*, 2015). This euryarchaeal group was previously identified as Z7ME43 (Sahl *et al.*, unpublished; NCBI GeneBank reference FJ902807); we propose to call this new WOR archaeal class Theionarchaea, as their genomes indicate strong capabilities for sulfur cycling (Figure 2). This new archaeal lineage appears to comprise widespread archaea, as phylogenetic analysis based on 16S rRNA

Figure 2 Reconstructed metabolic pathways of genes detected in Theionarchaeota, based on the DG-70 and DG-70-1 bins. Predicted proteins indicated by numbers in the figure can be found in the Supplementary Table S4. A list of predicted signal peptides can be found in the Supplementary Table S6. DHAP, dihydroxyacetone phosphate; Fd, ferredoxin; Fru-6P/-1,6bP, Fructose-6-phophshate/-1,6-bisphosphate; GAP, glyceraldehyde 3-phosphate; Glu-1P/-6P, Glucose-1-phosphate/-6-phosphate; NAD, nicotinamide adenine dinucleotide; PEP, phosphoenolpyruvate; THM, tetrahydromethanopterin.

genes show that sequences belonging to this new phylum were recovered from coastal marine sediments, estuary sediments, lagoon carbonaceous sediments, a subsurface limestone sinkhole biomat, lake sediment, deep subsurface marine sediments from the Peru Margin and a terrestrial mud volcano (Supplementary Figure S1).

As their most distinctive feature, bins DG-70 and DG-70-1 contain genes encoding all four subunits of a sulfhydrogenase (or sulfur reductase; Figure 2; Supplementary Figure S6). This enzyme has been described in hyperthermophilic archaea such as *P. furiosus*, in which sulfur reduction was initially assumed to function as sink for fermentatively generated H_2 (Fiala and Stetter, 1986). However, sulfur reduction catalyzed by the sulfhydrogenase was later suggested to be an energy-conserving mechanism allowing disposal of excess reductant using polysulfides or S^o (or H^+) as electron acceptors and H_2 or organic compounds as electron donor (Ma *et al.*, 1993; Schicho *et al.*, 1993). Two cytoplasmic (sulf)hydrogenases (H-I and H-II) have been identified and purified in *P. furiosus*, and whereas both

enzymes have been shown to preferentially reduce S^o rather than protons, H-II was shown to have a higher affinity for S^o and polysulfides than H-I (Ma *et al.*, 2000). Genes encoding all four subunits of the (sulf)hydrogenase in bins DG-70 and DG-70-1 were found to be similar to the genes belonging to the H-II of *P. furiosus* (Supplementary Figure S6). The Theionarchaea could also use thiosulfate as an electron acceptor, as bin DG-70-1 contains a *phsB* gene encoding a subunit of a thiosulfate reductase (Heinzinger *et al.*, 1995).

Bins DG-70 and DG-70-1 have the potential for detrital protein uptake and degradation to pyruvate and acetyl-CoA (Figure 2), similar to RC-III and MBG-D archaea. Bins DG-70 and DG-70-1 contain genes encoding all subunits of a pyruvate dehydrogenase complex, converting pyruvate derived from amino acid degradation to acetyl-CoA. Like the RC-III archaea, bins DG-70 and DG-70-1 include genes encoding a pyruvate-formate lyase and pyruvate-formate lyase activating enzymes, probably induced by anoxic conditions as shown previously for *E. coli* (Sawers and Bock, 1988). In addition, bins DG-70

and DG-70-1 have genes encoding a formate dehydrogenase, which would dispose of the toxic formate produced by conversion of pyruvate to acetyl-CoA via the pyruvate-formate lyase (Figure 2).

The carbohydrate metabolism of the Theionarchaea appears to be limited. Bins DG-70 and DG-70-1 contain one gene encoding a sugar permease; and their partial EMP pathways are lacking genes involved in the first steps of glycolysis (Figure 2). Bins DG-70 and DG-70-1 also contain genes involved in the methylglyoxal pathway, a low-energy yielding bypass of the lower EMP that does not produce ATP and is typically activated under phosphate limiting conditions (Weber et al., 2005). One possible fate of acetyl-CoA obtained by either protein or fructose/ glucose degradation is conversion to acetate via an ATP-producing acetyl-CoA synthase (EC:6.2.1.13); the acetate produced this way, or taken up directly via an cation/acetate symporter, can ultimately be fermented to ethanol (Figure 2). The Theinarchaeoa could also convert acetate to acetyl-CoA as bins DG-70 and DG-70-1 have an acs gene encoding an acetyl-CoA synthase (EC:6.2.1.1). As bins DG-70 and DG-70-1 have genes encoding a complete TCA cycle, acetyl-CoA can enter this pathway, hence producing ATP and amino acid biosynthesis precursors.

Bins DG-70 and DG-70-1 contain genes encoding a complete reductive acetyl-CoA (Wood–Ljungdahl) pathway using the methanogenic C_1-carrier tetra-methanopterin, as well as genes involved in coenzyme F420 biosynthesis and genes encoding a coenzyme F420-reducing hydrogenase (Figure 2, Supplementary Table S4). The phylogenetic tree based on amino acid sequences of CdhA placed the Theionarchaea CdhA sequences within a cluster containing mainly methanogens and Archaeoglobus (Supplementary Figure S4). In contrast, the MBG-D genome encodes a pathway using tetrahydrofolate, the typical C_1-carrier in acetogenesis; and its CdhA sequences form a distinct cluster lacking methanogens (Supplementary Figure S4). These carriers have been suggested to functionally distinct, based on energy metabolism, thermodynamics of both reactions and biosynthetic pathways of the two carriers (Maden, 2000). The tetrahydrofolate pathway plays a major role in biosynthesis of purines and regeneration of methionine, although in acetogenic bacteria most of the C1 flux was shown to be used for energy generation. In the reductive direction, the tetrahydrofolate pathway consumes ATP, whereas the redox reactions involved in the methanopterin pathway generate electrochemical ion gradients (Deppenmeier et al., 1996) leading to chemiosmotic synthesis of ATP (Maden, 2000).

The Theionarchaea potentially have two ways of obtaining reduced ferredoxin that may substitute for each other. In addition to containing a HdrABC-MvhADG complex, a likely source of reduced ferredoxin, bins DG-70 and DG-70-1 also have eight subunits of an energy-converting hydrogenase B (ehb). These subunits have been suggested to constitute membrane proteins forming a large electron transfer complex (Tersteegen and Hedderich, 1999; Major et al., 2010) catalysing energy-driven reduction of low-potential electron carriers, such as ferredoxin. The expression levels of hydrogenase B in Methanobacterium thermoautotrophicum cells increased significantly in response to H_2 limiting conditions (Tersteegen and Hedderich, 1999), suggesting that the gene expression levels of ferredoxin-reducing enzyme systems in the Theionarchea are environmentally modulated.
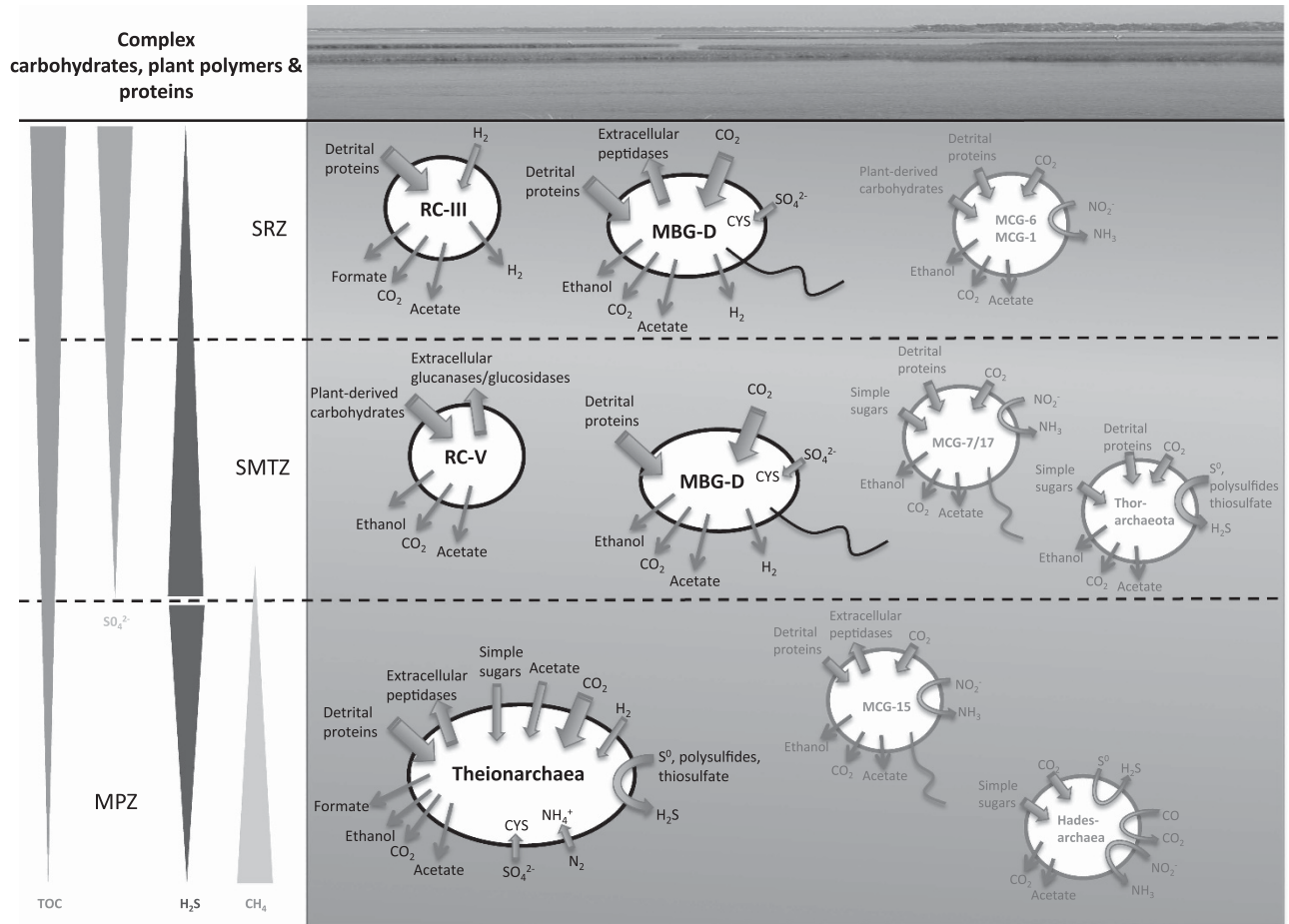
The Theionarchaea can potentially fix nitrogen to ammonium, as suggested by the detection of a nif operon encoding subunits of a nitrogenase and accessory proteins (Figure 2; Supplementary Table S4). Nitrogen fixation could allow Theionarchaea to use N_2 as sole nitrogen source in nitrogen-limiting conditions. In addition, bins DG-70 and DG-70-1 have genes encoding proteins (NtrC, NtrY) involved in the Ntr two-component system which regulates nitrogen fixation (Gussin et al., 1986; Figure 2). Similarly to the MBG-D genomes, the theionarchaeal bins DG-70 and DG-70-1 possess numerous genes encoding one- and two-component systems, six genes encoding CRISPR exonucleases, genes encoding enzymes involved in DNA repair and the uspA gene (Figure 2; Supplementary Table S4).

Habitat specificity and predicted biogeochemical roles of the WOR archaea

Estuaries such as the WOR are known for harboring highly active heterotrophic and autotrophic microbial communities as they are continuously replenished in organic matter and nutrients from both land and sea (Poremba et al., 1999). Thus, the roles of archaea in these complex microbial ecosystems extend beyond the uptake and degradation of detrital proteins (Ouverney and Fuhrman, 2000; Kemnitz et al., 2005; Lloyd et al., 2013). Although eight of the nine reconstructed WOR archaeal genomes in this study—belonging to the RC-III, MBG-D and Theionarchaea—encode the capability for extracellular protein degradation, they also show additional capabilities that link them to carbohydrate degradation, acetogenesis and sulfur reduction.

In contrast to the MBG-D and Theionarchaea genomes, the RC-III genome recovered from the shallow SRZ layers did not contain any genes encoding extracellular peptidases. It is possible that the RC-III archaea depend on other archaea to predigest detrital proteins into oligopeptides that are small enough for transmembrane uptake (Figure 3). Compatible with such a metabolic limitation, molecular surveys have detected RC-III archaea mostly in nutrient-rich rhizosphere habitats (Kemnitz et al., 2005), whereas the MBG-D are found in a wider range of aquatic benthic environments with limited carbon substrates and energy sources (Teske and Sorensen, 2008; Lloyd et al., 2013).

Figure 3 Predicted general metabolisms of the WOR archaea in relation to the geochemical layer from which they were recovered. In addition, the genome-based metabolic predictions for the Bathyarchaeota subgroups −6, −1, −7/17 and −15 (Lazar et al., 2015), the Thorarchaeota (Seitz et al., 2016) and the Hadesarchaea (Baker et al., 2016) were added. TOC, total organic carbon; MCG, miscellaneous crenarchaeotal group (Bathyarchaeota); CYS, cysteine. WOR picture modified from http://www.learnnc.org/lp/editions/cede_black waterriver/54.

Metabolic predictions for the RC-V archaea, based on the genome reconstructed from the sulfate-depleted sediment layers of the SMTZ in the WOR estuary, indicate that these archaea seem to possess limited metabolic capacities and are probably obligate fermenters of carbohydrates (Figure 3). Although the RC-V archaea appear usually as a non-dominant archaeal group in molecular surveys (Jurgens et al., 2000; Glissmann et al., 2004), they were sufficiently abundant in the sampled WOR sediments for genomic reconstruction. As no sedimentary habitat has been shown so far to be exclusively dominated by the RC-V archaea, it is likely that they are generalists sharing the fermentative ecological niche with a wide range of other microorganisms. Consistent with this interpretation, high concentrations of total organic carbon in the range of 4–6% were measured in the WOR sediment samples used for genomic reconstruction (Lazar et al., 2015). An average δ^{13}C-total organic carbon of −25‰ indicated a major contribution of terrestrial C4 plant-derived organic matter in the sampled WOR

sediments (Meador et al., 2015). Thus, the availability of marine and terrestrial sources of organic carbon could allow for the coexistence of many different types of fermenters in the WOR estuary sediments, including the RC-V archaea.

In contrast to the carbohydrate-dependent RC-V archaea, the MBG-D archaea lack genes involved in carbohydrate uptake and assimilation. On the basis of the observed set of genes, cellular carbon would originate from protein degradation, complemented most likely by heterotrophic CO_2 recycling through the Wood–Ljungdahl pathway. The MBG-D genome recovered from the SMTZ might possess a more limited capacity for extracellular degradation of proteins compared with the MBG-D genome reconstructed from the shallow SPZ layers, and its core metabolism seems to be protein degradation either by respiration or fermentation.

Genomic analysis of the Theionarchaea highlighted their repertoire of metabolic strategies to deal with substrate- and nutrient-limiting conditions. The Theionarchaea genomes show potential for

using and converting sugar phosphates and amino acids to acetyl-CoA and further to acetate with their ATP-producing acetyl-CoA synthase; they also have the potential for taking up and oxidizing acetate by converting it to acetyl-CoA and inserting it into their complete TCA cycle. Hence these archaea could scavenge extracellular acetate when acetate-producing carbon sources become depleted, an adaptation strategy also known as 'acetate switch' (Wolfe, 2005). The Theionarchaea could also cope with phosphate limiting conditions by activating the methylglyoxal pathway, a low-energy bypass of the lower EMP pathway; they could adapt to nitrogen-limiting conditions by fixing atmospheric N_2; and could respond to H_2-limiting conditions by using a specialized energy-converting hydrogenase B for CO_2 fixation. Finally, the Theionarchaea genome features genes encoding enzymes involved in elemental sulfur or thiosulfate reduction. As a product of sulfide oxidation, thiosulfate has been shown to be an important intermediate for oxidative as well as reductive sulfur transformations in marine sediments (Jorgensen and Bak, 1991). Hence, the Theionarchaea could play an important role in sulfur cycling, especially in strongly reduced marine sediment environments.

Conclusions

Overall, the archaeal genomes described here suggest a diversified spectrum of metabolic pathways and biogeochemical niches. On the basis of the genomic evidence in the WOR bins, the RC-III archaea take up and ferment low-molecular-weight proteins that have to be pre-digested from larger precursors by other community members; the RC-V archaea appear to be obligate carbohydrate fermenters, whereas the MBG-D archaea lack uptake systems for carbohydrates and cover their carbon needs by protein and peptide assimilation. The Theionarchaea employ complex strategies to use proteins and sugars as carbon sources for their sulfur-reducing metabolism, and for adapting to nitrogen, phosphate and hydrogen limitation. This extended genomic repertoire suggests that benthic archaea have diversified roles in marine sedimentary environments, and are not limited to detrital protein degradation (Lloyd et al., 2013) or acetogenesis (He et al., 2016). A flexible complement of metabolic pathways could allow these archaea to survive and to thrive in a wide range of benthic and sedimentary habitats.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Baker BJ, Lazar CS, Teske AP, Dick GJ. (2015). Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**: 14.

Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs K-U, Teske AP et al. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat Microbiol* **1**: 16002.

Bara MTF, Lima AL, Ulhoa CJ. (2003). Purification and characterization of an exo-L-1,3-glucanase produced by *Trichoderma asperellum*. *FEMS Microbiol Lett* **213**: 81–85.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.

Battchikova N, Eisenhut M, Aro E-M. (2011). Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim Biophys* **1807**: 935–944.

Biddle JF, Lipp JS, Lever MA, Llyod KG, Sorensen KB, Anderson R et al. (2006). Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *PNAS* **103**: 3846–3851.

Bräsen C, Esser D, Rauch B, Siebers B. (2014). Carbohydratemetabolism in archaea: current Insights into unusual enzymes and pathways and their regulation. *Microbiol Mol Biol Rev* **78**: 89–175.

Cárdenas JP, Martínez V, Covarrubias P, Holmes DS, Quatrini R. (2009). Predicted CO/CO2 fixation in Ferroplasma spp. via a novel chimaeric pathway. *Adv Mat Res* **71-73**: 219–222.

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 Commun* **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. *Curr Microbiol* **25**: 690–701.

Chaban B, Ng SYM, Kanbe M, Saltzman I, Nimmo G, Aizawa S-I et al. (2007). Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon *Methanococcus maripaludis*. *Mol Microbiol* **63**: 596–609.

Criscuolo A, Gribaldo S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**: 210.

1128

Darling AE, Jospin G, Lowe E, Matsen FA IV, Bik HM, Eisen JA. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**: e243.

Deppenmeier U, Muller V, Gottschalk G. (1996). Pathways of energy conservation in methanogenic archaea. *Arch Microbiol* **165**: 149–163.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Biology* **10**: R85.

Dobbek H, Svetlitchnyi V, Gremer L, Huber R, Meyer O. (2001). Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] Cluster. *Science* **293**: 1281–1285.

Fiala G, Stetter KO. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 °C. *Arch Microbiol* **145**: 56–61.

Glissmann K, Chin K-J, Casper P, Conrad R. (2004). Methanogenic pathway and archaeal community structure in the sediment of eutrophic lake Dagow: effect of temperature. *Microb Ecol* **48**: 389–399.

Grosskopf R, Stubner S, Liesack W. (1998). Novel euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. *Appl Environ Microbiol* **64**: 4983–4989.

Gussin GN, Ronson CW, Ausubel FM. (1986). Regulation of nitrogen fixation genes. *Ann Rev Genet* **20**: 567–591.

He Y, Li M, Perumal V, Feng X, Fang J, Xie J *et al.* (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol* **1**: 16035.

Heinzinger NK, Fujimoto NY, Clark MA, Moreno MS, Barett EL. (1995). Sequence analysis of the phs operon in Salmonella typhimurium and the contribution of thiosulfate reduction to anaerobic energy metabolism. *J Bacteriol* **177**: 2813–2820.

Huelsenbeck JP, Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Jorgensen BB, Bak F. (1991). Pathways and microbiology of thiosulfate transformations and sulfate reduction in a marine sediment (Kattegat, Denmark). *Appl Environ Microbiol* **57**: 847–856.

Jurgens G, Glöckner F-O, Amann R, Saano A, Montonen L, Likolammi M *et al.* (2000). Identification of novel Archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent *in situ* hybridization. *FEMS Microbiol Ecol* **34**: 45–56.

Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *PNAS* **109**: 16213–16216.

Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Kelley CA, Martens CS, Chanton JP. (1990). Variations in sedimentary carbon remineralization rates in the White Oak River Estuary, North Carolina. *Limnol Oceanogr* **35**: 372–383.

Kelley CA, Martens CS, Ussler W III. (1995). Methane dynamics across a tidally flooded riverbank margin. *Limnol Oceanogr* **40**: 1112–1129.

Kemnitz D, Kolb S, Conrad R. (2005). Phenotypic characterization of rice cluster III archaea without prior isolation by applying quantitative polymerase chain reaction to an enrichment culture. *Environ Microbiol* **7**: 553–565.

Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, Wollherr A *et al.* (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *PNAS* **107**: 13087–13092.

Land LA, Paull CK. (2001). Thermal gradients as a tool for estimating groundwater advective rates in a coastal estuary: White Oak River, North Carolina, USA. *J Hydrol* **248**: 198–215.

Lazar CS, Biddle JF, Meador TB, Blair N, Hinrichs K-U, Teske AP. (2015). Environmental controls on intra-group diversity of the uncultured benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in anoxic sediments of the White Oak River estuary (North Carolina, USA). *Environ Microbiol* **17**: 2228–2238.

Lazar CS, Baker BJ, Seitz K, Hyde AS, Dick GJ, Hinrichs K-U *et al.* (2016). Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ Microbiol* **18**: 1200–1211.

Lipp JS, Morono Y, Inagaki F, Hinrichs K-U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**: 991–994.

Lloyd KG, Alperin MJ, Teske A. (2011). Environmental evidence for net methane production and oxidation in putative ANaerobic MEthanotrophic (ANME) archaea. *Environ Microbiol* **13**: 2548–2564.

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.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Ma K, Schicho RN, Kelly RM, Adams MWW. (1993). Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: Evidence for a sulfur-reducing hydrogenase ancestor. *PNAS* **90**: 5341–5344.

Ma K, Weiss R, Adams MWW. (2000). Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. *J Bacteriol* **182**: 1864–1871.

Maden BEH. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochem J* **350**: 609–629.

Major TA, Liu Y, Whitman WB. (2010). Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*. *J Bacteriol* **192**: 4022–4030.

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

Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y *et al.* (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123–D129.

Martens CS, Goldhaber MB. (1978). Early diagenesis in transitional sedimentary environments of the White Oak River Estuary, North Carolina. *Limnol Oceanogr* **23**: 428–441.

Meador TB, Bowles M, Lazar CS, Zhu C, Teske A, Hinrichs K-U. (2015). The archaeal lipidome in estuarine sediment dominated by members of the Miscellaneous Crenarchaeotal Group. *Environ Microbiol* **17**: 2441–2458.

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.

Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD *et al.* (2000). Genome sequence of Halobacterium species NRC-1. *PNAS* **22**: 12176–12181.

Nystrom T, Neidhardt FC. (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol Microbiol* **11**: 537–544.

Ouverney CC, Fuhrman JA. (2000). Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**: 4829–4833.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ* **2**: e554v551.

Patenge N, Berendes A, Engelhardt H, Schuster SC, Oesterhelt D. (2001). The fla gene cluster is involved in the biogenesis of flagella in *Halobacterium salinarum*. *Mol Microbiol* **41**: 653–663.

Peng Y, Leung HCM, Yiu SM, Chin FCL. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Poremba K, Tillmann U, Hesse K-J. (1999). Tidal impact on planktonic primary and bacterial production in the German Wadden Sea. *Helgoland Mar Res* **53**: 19–27.

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

Sapra R, Bagramyan K, Adams MWW. (2003). A simple energy-conserving system: proton reduction coupled to proton translocation. *PNAS* **100**: 7545–7550.

Sawers G, Bock A. (1988). Anaerobic regulation of pyruvate formate-lyase from Escherichia coli K-12. *J Bacteriol* **170**: 5330–5336.

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

Schicho RN, Ma K, Adams MWW, Kelly RM. (1993). Bioenergetics of sulfur reduction in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **175**: 1823–1830.

Schuchmann K, Müller V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev* **12**: 809–821.

Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J* **10**: 1696–1705.

Smith MR, Lequerica JL. (1985). Methanosarcina mutant unable to produce methane or assimilate carbon from acetate. *J Bacteriol* **164**: 618–625.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**: 1449–1452.

Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) Software Version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Tersteegen A, Hedderich R. (1999). *Methanobacterium thermoautotrophicum* encodes two multisubunit membrane-bound [NiFe] hydrogenases. *Eur J Biochem* **264**: 930–943.

Teske A, Sorensen KB. (2008). Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J* **2**: 3–18.

Ultsch A, Moerchen F. *ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM.* Technical Report, Department of Mathematics and Computer Science, University of Marburg: Germany, 2005. Report no. 46.

Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach A-L. (1999). Population structure and phylogenetic characterization of marine bentic Archaea in deep-sea sediments. *Appl Environ Microbiol* **65**: 4375–4384.

Weber J, Kayser A, Rinas U. (2005). Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. II. Dynamic response to famine and feast, activation of the methylglyoxal pathway and oscillatory behaviour. *Microbiology* **151**: 707–716.

Wiles AM, Cai H, Naider F, Becker JM. (2006). Nutrient regulation of oligopeptide transport in *Saccharomyces cerevisiae*. *Microbiology* **152**: 3133–3145.

Wolfe AJ. (2005). The acetate switch. *Microbiol Mol Biol Rev* **69**: 12–50.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)