Title:
Metabolic strategies of marine subseafloor Chloroflexi inferred from genome reconstructions

Running title:
Metabolic strategies of subseafloor Chloroflexi

Authors:
Maeva Fincker[a], Julie A. Huber[b], Victoria J. Orphan[c], Michael S. Rappé[d], Andreas Teske[e], Alfred M. Spormann[a,f]

Affiliations:
[a] Stanford University, Department of Civil and Environmental Engineering, Stanford, CA, USA
[b] Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
[c] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[d] Hawaii Institute of Marine Biology, University of Hawaii at Manoa, Kaneohe, HI, USA
[e] Department of Marine Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[f] Stanford University, Department of Chemical Engineering, Stanford, CA, USA

Corresponding author:
Alfred M. Spormann
Clark Center, E250
318 Campus Drive,
Stanford CA 94305
USA
spormann@stanford.edu
(650) 724-4927

Originality-Significance Statement.

This work provides the first comprehensive genomic insights into Chloroflexi microorganisms that dominate many marine subsurface environments globally. Our data revealed a near-ubiquitous core metabolic pattern based on so far unrecognized features of the microbial Wood-Ljungdahl pathway for CO_2 reduction in conjunction with diverse organic matter degradation pathways. Our genomic and geochemical observations provide a metabolic explanation for the world-wide presence of these microorganisms in subsurface environments at high frequency, as well as the first insights into how deepsea environments may select for microorganisms with specific metabolic traits that are uncommon in terrestrial, high flux environments. The presented findings have significant relevance for oceanography for understanding deep ocean and marine subsurface sediments, for microbial biology for understanding emerging patterns of metabolism, and for global ecology for gaining insights into key metabolic fluxes, activities and natural selection in subsurface environments.

SUMMARY

Uncultured members of the Chloroflexi phylum are highly enriched in numerous subseafloor environments. Their metabolic potential was evaluated by reconstructing 31 Chloroflexi genomes from 6 different subseafloor habitats. The near ubiquitous presence of enzymes of the Wood-Ljungdahl pathway, electron bifurcation, and ferredoxin-dependent transport-coupled phosphorylation indicated anaerobic acetogenesis was central to their catabolism. Most of the genomes simultaneously contained multiple degradation pathways for complex carbohydrates, detrital protein, aromatic compounds, and hydrogen, indicating the coupling of oxidation of chemically-diverse organic substrates to ubiquitous CO_2 reduction. Such pathway combinations may confer a fitness advantage in subseafloor environments by enabling these Chloroflexi to act as primary fermenters and acetogens in one microorganism without the need for syntrophic H_2 consumption. While evidence for catabolic oxygen respiration was limited to two phylogenetic clusters, the presence of genes encoding putative reductive dehalogenases throughout the phylum expanded the phylogenetic boundary for potential organohalide respiration past the Dehalococcoidia class.

INTRODUCTION

Members of the Chloroflexi phylum have been repeatedly identified throughout marine subseafloor environments – including sediments, surficial and deep, as well as crustal and hydrothermal fluids – across depth, local redox and geochemical gradients, and habitat types (Inagaki et al. 2006; Durbin and Teske 2011; Reveillaud et al. 2016; Jungbluth et al. 2016). In marine sediments, Chloroflexi can accumulate with depth up to a frequency of 75%, based on Chloroflexi 16S rRNA gene frequency (Parkes et al. 2014; Durbin and Teske 2011). Subphylum-level analyses revealed taxonomically-similar Chloroflexi members across marine environments with varied local oligotrophy and carbon conditions (Petro et al. 2019), indicating potential selection for common metabolic traits in both organic-rich and –poor environments. However, our understanding of the metabolic processes of these Chloroflexi and their ecophysiological acclimation/adaptation has so far been limited by inadequate *in situ* characterization and pure culture isolates (Parkes et al. 2014).

The Chloroflexi phylum contains metabolically diverse microorganisms, including photoautotrophs (e.g., *Chloroflexus aurantiacus*), aerobic organotrophs (e.g., *Thermomicrobium*), obligate organohalide respirers (e.g., *D. mccartyi* and other Dehalococcoidia Class members), as well as iron and nitrate reducers (e.g., *Ardenticatena maritima*). Marine anoxic subsurface and fluid communities often have been found to contain Chloroflexi members that are only distantly related to organohalide-respiring Dehalococcoidia (Group 1 in

Fig. 1A) (Löffler et al. 2013; Parkes et al. 2014) and Anaerolineae, Caldilinea (Group 2 in Fig. 1A), whose cultured isolates are organoheterotrophs (Yamada et al. 2006). The relative enrichment of Dehalococcoidia-related Chloroflexi in these environments is particularly intriguing as isolated terrestrial Dehalococcoidia members are highly specialized to an exclusive life style of reductive dehalogenation using a unique but ill-understood mechanism of anaerobic energy conservation (Fincker and Spormann 2017). The metabolic diversity of extant Chloroflexi coupled with the lack of related subseafloor isolates makes it difficult to accurately predict ecophysiological traits for subseafloor Chloroflexi.

We reconstructed and analyzed the metabolic potential of 31 near-complete Chloroflexi genomes from marine habitats metagenomes from 6 locations worldwide, including coastal, estuarine, and methane seep sediments as well as hydrothermal vent fluid, in order to predict catabolic processes by identifying energy-conserving strategies that may explain the enrichment of Chloroflexi in these environments. The resulting metabolic reconstructions hint at similar metabolic strategies and trade-offs that are irrespective of local redox stratification and oligotrophic conditions, and that may prove beneficial for sedimentary microorganisms in a structured, transport-limited environment such as the deep-sea subsurface. We also discovered a link between marine and terrestrial, as well as cultured and uncultured, Chloroflexi via the presence of reductive dehalogenases and the potential for an organohalide-respiring lifestyle in the marine subseafloor.

[Insert Figure 1]

RESULTS

Genome origins, statistics and phylogeny.

Six marine samples representing sediments as well as crustal and hydrothermal fluids were selected as representatives of diverse subseafloor environments in a range of geographic locations with varied depth, geochemistries and local environments (Supplementary Fig. S1 and Supplementary Table S1). Samples included shallow sediment cores from organic-rich coastal and estuarine from sites in Aarhus Bay (A) in Denmark (Starnawski et al. 2017) and the White Oak River Estuary (W) in North Carolina (Baker et al. 2015), as well as those from methane seep sediments at Santa Monica Mounds (S) off the coast of California (Skennerton et al. 2017) and those from hydrothermal sediments in the Guaymas Basin (G) near Mexico (Dombrowski et al. 2017). Samples also included anoxic, warm basaltic crustal fluids from drilled boreholes located in the flanks of the Juan de Fuca Ridge (J) (Jungbluth et al. 2016); and warm, mixed diffuse hydrothermal vent fluids from the ultramafic vent field Von Damm in the Mid-Cayman Rise (M) in the Caribbean Sea (Galambos et al. 2019). All samples were derived from anoxic environments, and sediment samples covered the sulfate reduction zone, the methane-to-sulfate transition zone and the methanogenic zone (Supplementary Table S1). Reads from 32 Illumina sequencing libraries from the 6 sampling sites were quality controlled, trimmed, and merged before assembly using MegaHit. The resulting contigs were binned using 3 binning software, and consensus bins were generated (see Methods). Overall, the metagenomic analysis pipeline yielded 31 high-quality Chloroflexi genomes that were more than 75% complete and with less

than 5% contamination according the CheckM single-copy gene analysis (Supplementary Table S2).

Although frequently used for phylogenetic identification, 16S rRNA gene sequences are notoriously difficult to assemble from metagenomes (Blazejak and Schippers 2010), and partial 16S rRNA genes were present in only 9 of the 31 genomes. Instead, PhyloPhlan (Segata et al. 2013) and other methods relying on the concatenation and alignments of conserved prokaryotic proteins have been shown to accurately resolve phylogenetic placements of new genomes even in the absence of 16S rRNA gene sequences. PhyloPhlan analysis confirmed the genomes belonged to the Chloroflexi phylum (Fig. 1). More specifically, 17 genomes from sediments from White Oak River Estuary, Aarhus Bay, Santa Monica Mounds sediments and fluids from Juan de Fuca Ridge belonged to a wide clade encompassing the Class Dehalococcoidia, known to contain organohalide respirers (Group 1 on Fig. 1). Genome W693 clustered closely with *Dehalococcoides* species, making it the sequenced marine subseafloor Chloroflexi genome most closely related to terrestrial organohalide-respiring Dehalococcoidia. The other 14 genomes formed a separate group comprising the Anaerolineae, Ardenticatenia and Thermoflexia clades (Group 2 in Fig. 1). Six genomes of the Anaerolineae clade were found in sediments from Guaymas Basin, Aarhus Bay, White Oak River Estuary and fluids from Mid-Cayman Rise, respectively. Four Thermoflexia genomes were reconstructed from the Guaymas Basin sediments and Juan de Fuca Ridge fluids. The last four genomes, recovered from Guaymas Basin sediments, and Mid Cayman Rise and Juan de Fuca Ridge fluids, were related to the

Ardenticatenia and Caldilineae clade. Notably, the genomes analyzed in this study belonged to Chloroflexi phylogenetic clades that have been repeatedly found to be enriched in 16S gene libraries or metagenomic studies of subseafloor environments (Inagaki et al. 2006; Wasmund et al. 2014; Nunoura et al. 2013), making them a representative sample of the phylogenetic diversity of subseafloor Chloroflexi.

Evidence for a catabolic role for the widespread Wood-Ljungdahl pathway.

Analysis of the genomes revealed the widespread presence of genes encoding the Wood-Ljungdahl pathway (WLP). The WLP is a linear pathway reducing two CO_2 to form one acetate in four two-electron reduction steps (Fig. 2B). The WLP operates only under anoxic conditions and can serve multiple metabolic functions. The pathway can function i) as a reductive anabolic pathway for CO_2 assimilation to acetyl-CoA (Ragsdale and Pierce 2008) in methanogenic archaea and acetogens; ii) as a reductive catabolic pathway as shown in model acetogenic bacteria *Moorella thermoacetica* and *Acetobacterium woodii* (Schuchmann and Müller 2016); and iii) as an acetate oxidation pathway in some sulfate-reducing microorganisms (Müller, Sun, and Schnürer 2013) in the presence of a suitable high-redox potential external electron acceptor or under syntrophic conditions. A near-complete set of enzymes for a functional WLP was present in all Group 1 genomes and the majority of Group 2 genomes (Fig. 2D and Supplementary Table S4).

Although no unambiguous genomic marker has been reported to indicate whether the WLP is operating anabolically or catabolically, common features exist in the genomes of acetogenic bacteria relying on a reductive catabolic WLP for energy conservation including enzyme complexes participating in electron bifurcation and membrane-bound ferredoxin-dependent transport-coupled phosphorylation (Nobu, Narihiro, Rinke, et al. 2015) (Fig. 2A and 2B). Both *A. woodii* and *M. thermoacetica* conserve energy by coupling reduced ferredoxin oxidation to proton- or sodium-translocation, respectively. *A. woodii* expresses a seven-subunit, membrane-spanning Rnf complex (Ferredoxin:NAD$^+$ oxidoreductase) that is encoded by the *rnfA-E* genes (Müller et al. 2008). Alternatively, the genome of *M. thermoacetica* encodes a membrane-bound energy conserving hydrogenase (Ech) as well as a heterodisulfide-reductase-associated ion-translocating Fd:NADH oxidoreductase (Hdr-IFO); both complexes participate in generating an electrochemical proton gradient (Nobu, Narihiro, Rinke, et al. 2015; Pierce et al. 2008; Nobu, Narihiro, Hideyuki, et al. 2015). The genomes of A089 and W693 contain homologous *rnfA-E* genes in an operon that mirrors the operon structure of *A. woodii*. Moreover, Hdr-IFO homologous complexes are present in 15 genomes, which may allow them to conserve energy during electron transfer from reduced ferredoxin to NAD$^+$. These ion-translocating complexes are critical during hydrogenotrophic (H$_2$ and CO$_2$ as catabolic substrates) acetogenesis when net ATP formation by transport-coupled phosphorylation is the only possible mechanism for net ATP synthesis. However, under heterotrophic conditions, oxidation of organic substrates may yield ATP via substrate-level phosphorylation, alleviating the need for exclusive energy

conservation coupled to ferredoxin oxidation. The absence of either of these complexes, however, does not rule out a catabolic function of the WLP as they are also missing from the genome of *Acetonema longum*, a known acetogenic bacteria (Nobu, Narihiro, Rinke, et al. 2015).

[Insert Figure 2]

Another key feature of acetogenic bacteria is the presence of electron bi/confurcating complexes coordinating the stoichiometries and energetics of electron flux from redox reactions in the oxidative branch of catabolism ($E'_{Fd/reduced\ Fd}$= -500mV, $E'_{NAD+/NADH}$= -280mV, $E'_{NADP+/NADPH}$= -380mV) with the idiosyncratic four reduction steps of the WLP as well as net conservation of redox energy ($E^{o'}_{CO2/CO}$ = -520mV, $E^{o'}_{CO2/Formate}$ = -430mV, $E^{o'}_{Methenyl-THF/Methylene-THF}$ = -295mV (Schuchmann and Müller 2014), $E^{o'}_{Methylene-THF/Methyl-THF}$ = -200mV (Wohlfarth and Diekert. 1991)) (Fig. 2BCD). Specifically, *M. thermoacetica* uses NfnAB, a bifurcating NADH-dependent ferredoxin:NADP$_+$ oxidoreductase, to generate NADPH necessary in the first reduction step of the WLP to form formate (Huang et al. 2012), and homologous NfnAB complexes were present in most Group 1 genomes. Alternatively, *Thermacetogenium phaeum*, a known acetogenic bacteria, as well as genome W589, contain PntAB, a NAD(P)$^+$ transhydrogenase that relies on reverse electron transport to interconvert NADH and NADPH (Nobu, Narihiro, Rinke, et al. 2015).

Overall, the co-occurrence of a complete set of enzymes of the WLP, of membrane-spanning, ferredoxin-dependent ion-translocating enzyme complexes, and of electron bifurcation/reverse

electron transport enzymes in a majority of the genomes is a strong indication that the WLP plays a central role in the catabolism of subseafloor Chloroflexi.

Enzymatic redundancy in the WLP and evidence for metabolic rate/yield trade-off

The electron carrier requirement and overall energetic efficiency of the four reduction steps of the WLP in acetogenic bacteria depends on the type of dehydrogenases/reductases present and their ability to participate in electron bifurcation. Electron bifurcation is a recently discovered mode of metabolic energetic conversion where a flavin-based protein complex couples an endergonic reduction of ferredoxin or ferredoxin-equivalent molecule to an exergonic reduction reaction. *M. thermoacetica* encodes a hexaheteromeric methylene-tetrahydrofolate reductase (MetF) that is proposed to couple the exergonic reduction of methylene-tetrahydrofolate with NADH to the endergonic reduction of a yet unknown low redox potential electron acceptor via electron-bifurcation (Mock et al. 2014), leading to the formation of methyl-tetrahydrofolate and a reduced ferredoxin equivalent. In conjunction with an energy-conserving hydrogenase (Ech) coupling the oxidation of this ferredoxin-equivalent electron carrier to proton translocation, the redox energy released during oxidation would result in the net conservation of 0.5 ATP per acetate formed (Mock et al. 2014). In contrast, *A. woodii*, which uses a non-bifurcating NADH-dependent MetF, dissipates this redox energy as heat. (Ragsdale and Ljungdahl 1984; Schuchmann and Müller 2014). Therefore, we explored the genetic context of the *metF* genes found in the 31 Chloroflexi genomes to identify bifurcating vs. non-bifurcating enzymes operating the WLP.

Analysis of the 31 subseafloor Chloroflexi genomes identified 45 MetF genes, with up to four copies in W589 genome (Fig. 3 and Supplementary Table S4). Operons containing genes encoding homologues to the MetVF-MvhD-HdrABC complex from *M. thermoacetica* were identified in seven Group 2 genomes, indicating that in these microorganisms, the reduction of methylene-tetrahydrofolate is potentially coupled to the formation of reduced ferredoxin for subsequent ferredoxin-dependent energy conservation. Most Group 1 genomes only encoded the *metFV* part of the above complex. This simpler architecture, also found in *C. ljungdahlii*, may still be involved in bifurcation (Nagarajan et al. 2013) or may interact with different enzymatic partners (e.g. a [NiFe] hydrogenase, or homologues to the NADH-interacting part of Complex I, NuoEFG). Alternatively, the absence of a bifurcating MetF may reflect a rate-over-yield strategy in Chloroflexi where dissipating Gibbs free energy as heat may enable a faster reaction rate instead of a higher energetic efficiency (Mondeel et al. 2016). The presence of both types of MetF in a single genome, such as in the genome of S90, may also give a microorganism the capacity to switch between rate- and yield-mode for maximizing either energetic yield or overall pathway rate depending on environmental conditions. However, no obvious trend existed between the type of MetF found in a genome and the local geochemical conditions.

A similar analysis of the operon structure of formate dehydrogenases (FdhA), catalyzing the first reduction step in the WLP from CO_2 to formate, revealed the presence of a bifurcating enzyme complex comprising a FdhA, HylB and HylC in 20 genomes (Supplementary Fig. S4 and Table S4). HylB and HylC are iron-sulfur containing protein subunits that, in *Clostridium*

acidiurici (Wang et al. 2013) and *Desulfotomaculum thermosubterraneum* (Visser et al. 2013), are part of a NAD^+- and ferredoxin-dependent confurcating formate dehydrogenase complex. Using a confurcating FdhA, the concurrent oxidation of reduced ferredoxin and NADH could drive the reduction of two CO_2 to form two formate, which increases the thermodynamic efficiency of this reaction compared to a non-confurcating FdhA requiring one reduced ferredoxin per CO_2 reduction.

[Insert Figure 3]

Aromatic compound metabolism.

Near-complete anaerobic benzoyl-CoA degradation pathways to acetate were present in 8 genomes (Fig. 1 and Supplementary Table S5). The key step of anaerobic benzoyl-CoA degradation is the reductive dearomatization of benzoyl-CoA to cyclohex-(di)ene-carbonyl-CoA, catalyzed by a benzoyl-CoA reductase (BCR) (Supplementary Fig. S3). Class I BCRs hydrolyze two ATP equivalents for the activation of two electrons derived from ferredoxin to reductively overcome the aromaticity of benzoyl-CoA, whereas class II BCRs catalyze the same reaction in an ATP-independent way. It has been hypothesized that class II BCRs instead rely on either membrane potential or electron bifurcation to activate the ferredoxin-derived electrons for aromatic ring reduction (Buckel and Thauer 2013). Correlating with the significant difference in energy requirement, class II BCRs are typically found in obligate anaerobes (e.g. *Geobacter metallireducens*) and class I BCRs are only found in facultative anaerobes (e.g. *Azoarcus* spp.),

with the exception of *Ferroglobus placidus* (Carmona et al. 2009; Holmes et al. 2011; Durante-Rodríguez et al. 2018).

Genes encoding class I BCR were present in genomes W156, W167, W281 and W589, and class II BCR genes in genomes A328 and W1033. These genomes were found across depth-driven geochemical gradients, and the presence of either type of BCR did not correlate with local geochemical conditions. Surprisingly, genomes W111 and W403 encoded genes for both a class I and a class II BCR, making them the first known genomes to contain both classes of BCR. The presence of class I BCRs in these predicted strictly anaerobic Chloroflexi is puzzling as its high ATP cost conflicts with the paucity of energy recovered in the WLP. Nevertheless, class I BCR have been previously identified in Chloroflexi from the Peru margin and, therefore, appear to be not unique in subseafloor Chloroflexi (Sewell, Kaster, and Spormann 2017; Wasmund et al. 2016).

In the anaerobic benzoyl-CoA degradation pathway, the product of benzoyl-CoA reduction, cyclohex-(di)ene-carbonyl-CoA, is further converted and oxidized to either 3-hydroxypimelyl-CoA or pimelyl-CoA following a modified β-oxidation pathway, and is eventually oxidized to three acetyl-CoA (Supplementary Fig. S3) (Carmona et al. 2009). All genes necessary for the formation of 3-hydroxypimelyl-CoA were accounted for in 6 of the 8 genomes containing a BCR. The final conversion of hydroxypymelyl-CoA to acetyl-CoA via glutaryl-CoA is catalyzed by enzymes involved in the metabolism of the dicarboxylates. Most bacteria typically contain

multiple candidate genes participating in dicarboxylate metabolism, and the genomes in this study were no exception.

Activation of aromatic substrates is typically initiated by the formation of an aryl-coenzyme A thioester (Carmona et al. 2009). The absence of benzoate-CoA ligase in any of the genomes indicates that free benzoate is unlikely to be the primary substrate for aromatic compound degradation in these Chloroflexi. Benzoyl-CoA is a funneling intermediate in the degradation pathway of a large number of aromatic compounds (Carmona et al. 2009) including 4-hydroxybenzoate. Genome A328 encodes a 4-hydroxybenzoate-CoA ligase, and its BCR operon was directly followed by genes encoding a 4-hydroxybenzoyl-CoA reductase making 4-hydroxybenzoate a potential substrate. Alternatively, putative phenylacetate-CoA ligases were also present in 8 genomes but no other gene associated with phenylacetate degradation was identified.

Complex carbohydrate degradation.

One of the major differences between Group 1 and Group 2 genomes was the number of genes predicted to participate in complex carbohydrate degradation and metabolism. A large range of predicted glycohydrolases involved in polysaccharide metabolism was found in Group 2 genomes, especially in J4, whereas carbohydrate metabolism in Group 1 was mostly restricted to glycolysis and a partial pentose-phosphate pathway (Supplementary Table S6).

In canonical glycolysis, glucose is first activated to fructose-1,6-bisphosphate by the successive action of hexokinase, isomerase and phosphofructokinase, before being cleaved into glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate via a fructose-bisphosphate aldolase (Supplementary Fig. S3). Interestingly, fructose-bisphosphate aldolase and phosphofructokinase were missing from 11 genomes. Instead, some of these genomes encoded a fructose-6-phosphate aldolase and a dihydroxyacetone kinase, an enzyme combination that yields the same glycolytic products but 'delays' the consumption of one ATP until after the first glyceraldehyde-3P is formed. Another missing enzyme in half of the genomes is pyruvate kinase, which catalyze the last step of glycolysis. It is unlikely pyruvate kinase was missed during reconstruction in 50% of the genomes given that median genome completeness was 83%, hence some subseafloor Chloroflexi must utilize a different enzyme to form pyruvate during glycolysis. Finally, pyruvate can be funneled into diverse anabolic pathways or turned into acetyl-CoA via either a pyruvate dehydrogenase or pyruvate:flavodoxin oxidoreductase in AC_1 and AC_2 genomes (see section on Respiratory Metabolism), or via a pyruvate:ferredoxin oxidoreductase as found in all other genomes.

Import and metabolism of carbohydrates in Group 2 genomes is predicted to be more versatile. The median count of glycohydrolases in Group 1 and Group 2 genomes was 4 and 16 respectively, with a maximum of 60 glycohydrolases in genome J4. Furthermore, Group 2 genomes encode on average five carbohydrate transporters while most Group 1 genomes did not contain any. The microorganism J4 in particular seems to specialize in carbohydrate degradation

as genes needed to metabolize fructose, xylose, arabinose, rhamnose, fucose, sucrose, cellulose and starch were present in its genome. Comparison with other isolated or environmental Chloroflexi indicated that carbohydrate degradation is a metabolic trait shared by most Group 2 Chloroflexi (Imachi et al. 2014; Fullerton, Moyer, and Löffler 2016; McIlroy et al. 2017; Dodsworth et al. 2014; Kale et al. 2013) and some Group 1 (Hug et al. 2013).

Unusual amino acid degradation pathways.

Another trait shared by some isolated Group 2 Chloroflexi is their ability to degrade amino acids (Dodsworth et al. 2014; Kawaichi et al. 2013; Podosokorskaya et al. 2013). Analysis of the 31 genomes revealed that the potential for anaerobic amino acid degradation was not however restricted to Group 2 (Supplementary Table S7), and the distribution of potentially accessible substrates did not seem to match phylogenetic clustering. Overall, oxidative pathways to degrade 13 different amino acids to central intermediates could be found.

Interestingly, the glycine and histidine degradation pathways found in the genomes interact directly with the WLP. The glycine cleavage pathway found in 16 genomes (Supplementary Fig. S3) is based on the glycine cleavage complex which yields 5,10-methenyl-THF as a direct intermediate in the WLP. Similarly, the histidine degradation pathway encoded in 8 genomes depends on a glutamate formimidoyltransferase to catalyze the key step in transformation of histidine to glutamate and 5,10-methenyl-THF. The latter can be funneled into the WLP while the former can be further oxidized to succinyl-CoA by the successive action of glutamate

dehydrogenase and 2-oxoglutarate:ferredoxin oxidoreductase - both enzymes are present in the genomes of the 8 histidine-degrading Chloroflexi. This pathway, usually restricted to eukaryotic organisms (Tabor and Wyngarden 1959), produces direct intermediates to the WLP and provides additional support for the hypothesis that acetogenesis via the WLP is a key marker of subseafloor Chloroflexi.

Respiratory metabolism restricted to two phylogenetic groups.

[Insert Figure 4]

Aerobic metabolism involves high redox potential cellular electron carriers such as ubiquinone and cytochromes, and a terminal oxygen reductase (Anraku 1988). Additionally, in chemoorganotrophic microorganisms, the oxidative citric acid cycle, as indicated by the presence of a citrate (*Si*)-synthase and a 2-oxo-glutarate dehydrogenase, is often present for complete oxidation of acetyl-CoA and for transferring reducing equivalents to NAD(P)$^+$ for oxidative phosphorylation (Schägger 2002). With the exception of a 2-oxo-glutarate dehydrogenase, these aerobic respiration gene markers were present in nine Group 2 genomes clustering in two distinct phylogenetic clusters (Fig. 4, Supplementary Table S3), indicating a potential for chemoorganotrophic aerobic respiration.

The first Aerobic Cluster (AC_1) was composed of deep-branching genomes from the *Caldilinea* clade whereas the second Aerobic Cluster (AC_2) formed a sister branch to the *Anaerolineae* clade. AC_1 and AC_2 differed in the predicted composition of their respiratory chains. While complete sets of genes for complex I (NADH-dehydrogenase), complex II (succinate

dehydrogenase complex) and cytochrome *bd* oxidases were present in both groups, genes

encoding cytochrome *c* oxidases and alternative complex III (Supplementary Fig. 1) were only

found in AC_1. Cytochrome *bd* oxidases generally exhibit high oxygen affinity and are

expressed in *E. coli* under low oxygen conditions (K_m for O_2 = 3 - 8 nM) (Alexeeva,

Hellingwerf, and Teixeira de Mattos 2002; D'mello, Hill, and Poole 1996). In contrast,

cytochrome *c* oxidases are the more typical terminal oxidases in most obligate aerobic

microorganisms (Ostermeier, Iwata, and Michel 1996). Alternative complexes III enable aerobic

microorganisms to tune growth yield by varying the stoichiometry of proton translocation to the

oxidation of cytochrome c during respiration (Refojo, Teixeira, and Pereira 2012). Consistent

with our findings, the closest cultured relative of AC_1, *Litorilinea aerophila* (Kale et al. 2013),

was isolated from intertidal hot springs and is strictly aerobic, whereas *Aggregatilinea*

lenta (Nakahara et al. 2019), recently isolated from a continuous-flow reactor and closest relative

to AC_2, was shown to be a facultative aerobe. Additionally, RBG-9, a Group 2 Chloroflexi

from uranium-contaminated aquifer sediments, was also predicted to be aerobic with

fermentative capabilities (Hug et al. 2013). Our results also demonstrated that close phylogenetic

proximity to either AC_1 or AC_2 may be an accurate indicator for the capacity to respire

oxygen as isolated aerobic or facultatively anaerobic Chloroflexi from diverse environments also

clustered within AC_1 and AC_2. The presence of O_2-reduction pathways in genomes from

sampling locations with low or no oxygen is interesting and may reflect the environmental

conditions these subseafloor Chloroflexi encountered before their sedimentation if these

microorganisms had been seeded from oxic pelagic environments. Alternatively, oxygen reduction capabilities could serve as a protective mechanism for aerotolerance during temporary exposure to oxygen, as is the case in some sulfate-reducing bacteria (Lamrabet et al. 2011).

Instead of the expected 2-oxoglutarate dehydrogenase functioning in the oxidative citric acid cycle, all AC_1 and AC_2 genomes encoded a 2-oxoglutarate:ferredoxin oxidoreductase, an enzyme enabling reversible 2-oxoglutarate oxidation which is typically found in anaerobic microorganisms or facultative anaerobes as well as in some aerobic Archaea (Baughn 2009; Zhang et al. 1996). When combined with the low oxygen concentration at the sampling locations - estimated to be null in Juan de Fuca samples (Lin et al. 2012) and rapidly depleted at the sediment-water interface in Guaymas Basin sediments (Teske et al. 2016), respectively - this observation raises the possibility that oxygen respiration may not be the main reductive catabolic pathway for these microbes, and might represent merely one ecophysiological strategy. Furthermore, the potential for aerobic respiration did not correlate with local geochemistry and was found across sedimentary geochemical gradients. In contrast to AC_1 and AC_2 genomes that contained a citrate (*Si*)-synthase, most Group 1 genomes encoded a gene with high homology to citrate (*Re*)-synthase of *D. mccartyi* CBDB1 (Marco-Urrea et al. 2011). (*Re*)-citrate synthases, first identified in *Clostridium kluyveri* (Gottschalk and Barker 1966), have so far only been found in strictly anaerobic microorganisms (Li et al. 2007). Their presence in Group 1 genomes therefore suggests a strictly anaerobic lifestyle for these Chloroflexi.

No compelling evidence for anaerobic respiration was found in any of the 31 genomes. Key genes for sulfate, sulfite, and/or nitrate reduction were absent. Genome J4 contained a putative tetrathionate reductase and a nitrous oxide reductase, and genome G439 revealed a putative nitrite reductase (Supplementary Table 4). Either enzyme could act as terminal reductase in a respiratory chain under anoxic conditions. Furthermore, genes encoding additional molybdopterin-dependent oxidoreductases were found in a number of genomes, but their low sequence similarity with any characterized protein made functional inference difficult (Supplementary Fig. S2). Based on phylogeny, *Ardenticatena maritima* (Kawaichi et al. 2013), a facultatively aerobic nitrate and iron-reducing member of the Chloroflexi isolated from a coastal hydrothermal field, is relatively close to genomes J4, G439 and AC_2, hinting that a strictly anaerobic respiratory metabolism may not be widespread in Group 2 subseafloor Chloroflexi.

Reductive dehalogenases present throughout the phylum.

Reductive dehalogenases catalyze reductive elimination of a halide from an organohalide compounds. They act as membrane-associated periplasmic terminal reductases during anaerobic organohalide respiration, and are the key enzymes of the sole catabolic pathway in members of the Dehalococcoidia Class (Group 1) that couples hydrogen oxidation to organohalide reduction and energy conservation. Catabolic reductive dehalogenases are typically found in an *rdhAB* operon encoding a large catalytic unit RdhA and a membrane anchor RdhB. Reductive dehalogenases can also be used by a wide variety of anaerobic or microaerophilic

microorganisms to access recalcitrant organohalide compounds, which can then be used in central metabolism (Fincker and Spormann 2017; Lohner and Spormann 2013).

Nineteen reductive dehalogenase genes were found in 12 reconstructed genomes, with a maximum of four genes in genome G519 (Supplementary Table S8). Three *rdhA* genes found in genome G519 were directly upstream of a predicted *rdhB*, and two of them also encoded a TAT signal, marking these proteins for export through the cytoplasmic membrane. Three *rdhA* sequences from genomes G519, J4 and W167 also contained N-terminal transmembrane helices, which have been hypothesized to act as membrane anchors, negating the need for a separate *rdhB* (Atashgahi 2019). Reductive dehalogenases appear widespread across subseafloor Chloroflexi: their presence was not restricted to the genomes belonging to the Dehalococcoidia class, and no obvious phylogenic pattern could explain their phylogenetic distribution. The new Chloroflexi RdhA sequences scatter across phylogenetic boundaries, indicating that lateral gene transfer may play a role in the acquisition of reductive dehalogenases, as has been hypothesized in *Dehalococcoides* (Krajmalnik-Brown et al. 2007; McMurdie et al. 2011) and *Desulfurospirillum* (Fig. 5) (Maillard, Regeard, and Holliger 2005).

[Insert Figure 5]

Interestingly, the typical divide observed between *Dehalococcoidia* and non-*Dehalococcoidia* *rdhA* sequences was broken by a sequence found in Group 2 genome G519. The divergence between *Dehalococcoidia* sequences and other *rdhA* has been attributed to the difference in electron transport partner involved in reductive dehalogenation (Fincker and Spormann 2017).

Dehalococcoidia-type reductive dehalogenases are quinone-independent and part of an enzymatic electron transport complex comprised of an uptake hydrogenase, a molybdopterin-enzyme homologous to a formate dehydrogenase and a reductive dehalogenase (Kublik et al. 2016). The G519 genome only contained a group 3c hydrogenase, a type of soluble hydrogenase with undiscovered function (Vignais et al. 2001). In the absence of a typical uptake hydrogenase (hydrogenase group 1 or 4, see Supplementary Table S9) or formate dehydrogenase-like protein in G519, the identity of the electron donating partner interacting with the *Dehalococcoidia*-type RdhA remains unknown. Hence, the potential for reductive dehalogenation appears to be widespread in the Chloroflexi phylum and not restricted to the Dehalococcoidia branch.

Organohalide respiration is thought to be an important catabolic pathway in subseafloor microbial communities (Futagami et al. 2009), and it may indeed be present in some subseafloor Chloroflexi but the pathway may interact with hydrogen and other electron donors in yet unidentified ways given that typical group 1 uptake hydrogenases found in *D. mccartyi* are absent in most genomes (Supplementary Table S9). Alternatively, the reductive dehalogenases found could function to funnel halogenated organic compounds into the previously discovered pathways for organic matter degradation as found in *S. oneidensis* (Lohner and Spormann 2013).

DISCUSSION

Our comprehensive study revealed the near-ubiquitous presence of a functional catabolic WLP in conjunction with oxidation pathways of H_2 and diverse organic compounds in globally distributed Chloroflexi in subseafloor environments. These data generalize previous genomic reconstruction from three single Chloroflexi cells from the Peruvian margin (Dsc genomes in Fig. 1) (Sewell, Kaster, and Spormann 2017; Kaster et al. 2014) and Chloroflexi from terrestrial sediments (Hug et al. 2013).

In electron acceptor-limited anoxic environments with a high carbon flux such as the rumen of ruminants, anaerobic digestors, or some freshwater sediments, degradation of organic matter is typically mediated by metabolically interdependent consortia of anaerobic microorganisms. These consortia include primary fermenters, syntrophs, acetogenic, and methanogenic microorganisms (Fig. 6A), and reflect a speciation of the overall degradation pathway into short pathways in distinct physiological groups driven by competition in high flux environments (Pfeiffer, Schuster, and Bonhoeffer 2001; Pfeiffer and Bonhoeffer 2004). Catabolism, in particular of carbohydrates, of specialized fermenting microorganisms depends to a large extent on protons as electron acceptor. In fact, removing oxidative electrons, e.g. from pyruvate oxidation to acetyl-CoA, from the collective cellular pool of fermentative electrons as H_2 enables a more efficient resource utilization, since acetyl-CoA can be converted to ATP and acetate. However, such metabolic mode of primary and secondary fermenters requires H_2 consumptions by hydrogentrophic microorganisms such as acetogens or methanogens.

By using ubiquitous CO_2 as electron acceptor, WLP-containing Chloroflexi can oxidize their organic substrates more efficiently, e.g., to acetyl-CoA, relative to fast-growing strictly fermentative Prokaryotes (Fig. 6B). Moreover, using the CO_2-reductive WLP to balance fermentative catabolic electrons instead of syntrophic interspecies H_2 transfer eliminates the need for spatial proximity to H_2-consuming microorganisms. Such spatial proximity, or absence thereof, may become a critical limitation in fitness of strictly fermenting microorganisms in low flux, slow mixing environments such as marine subsurface sediments, and opens a niche for fermenting microbes with an internal CO_2-reductive pathway (Fig. 6 B). Finally, the multiple copies of MetF and formate dehydrogenases present in many of these Chloroflexi may accommodate the variation in cognate electron carriers for the specific reductases of the WLP (e.g. NADH-dependent MetF vs. NADH- and ferredoxin-dependent MetF). They may also provide important metabolic flexibility by accommodating multiple combinations of electron carriers involved in the oxidation of diverse organic substrates. In addition, the presence of energy-conserving and non-energy-conserving MetF enables Chloroflexi with multiple metF copies to switch between a rate and a yield mode; a property that can be beneficial in mixing-limited environments. The presence of overall longer pathways in the subseafloor Chloroflexi and the absence of cross-feeding agrees with the theory predicting that longer cellular pathways leading to complete substrate degradation are being selected for in environments where resource concentrations are low and growth is slow (Pfeiffer and Bonhoeffer 2004). Heterotrophic acetogenesis therefore emerged as widespread metabolism in subseafloor Chloroflelxi; while

most of the metabolism of acetogenic bacteria has been studied in the context of microbial H_2 scavenging and utilization, this comprehensive analysis of subseafloor Chloroflexi revealed that an organotrophic fermentative life style with internal CO_2 reduction seems more likely and the predominant mode of 'acetogens' in such environments (Schuchmann and Müller 2016).

No obvious trend existed between predicted metabolism and local geochemical conditions, suggesting that the identified metabolic traits may be beneficial in the context of other constraints set by marine subsurface environments. It is possible that other traits, such a decelerated death rate, contribute to the predominant presence of Chloroflexi in subsurface sediments. Notably, the relative composition of sedimentary microbial communities, including the increase in Chloroflexi frequency, is determined in the first anoxic layers during sedimentation, just below the more dynamic bioturbation zone. This relative composition persists as it is buried with time with increasing restriction on chemical transport, microbial migration, and energy availability (Petro et al. 2019). Thus, it is unlikely that the genomic patterns and predicted catabolic phenotype of subseafloor Chloroflexi are the result of convergent adaptive evolution as they reoccur independently in different marine environments worldwide (Petro et al. 2019). Finally, the Chloroflexi studied here belong to the same taxonomic groups that are found enriched in the deep marine subsurface, hence the highlighted metabolic strategies may be shared by the marine subsurface Chloroflexi in energy-limited environments.

[Insert Figure 6]

EXPERIMENTAL PROCEDURES

Data availability

Genome sequences have been submitted to the GenBank databases under the Bioproject PRJNA605778.

Sampling locations and sequencing libraries

Six subseafloor sampling locations that had previously been identified as containing Chloroflexi and where metagenomic sequencing data was available were selected for this study. Characteristics of the sampling locations and sediments pore fluids are reported in Supplementary Table S1, along with the accession numbers of the raw sequencing data obtained at each location. In total, 32 metagenomic libraries were assembled and analyzed.

Quality control, *de novo* assembly and binning of metagenomic sequences

Raw reads from each metagenomic library were quality trimmed using BBDuk (December 2017), and forward and reverse reads were merged if they overlapped using BBMerge (March 2017). All clean reads originating from the same sampling sites were aggregated and co-assembled with MegaHit (v1.1.2) (Li et al. 2015) using default parameters, and only contigs longer than 1 kbp were kept for downstream analysis. Contig-read alignments were computed by mapping the clean reads back to the assembled contigs using Bowtie2 (v2.3.4.1) (Langmead and Salzberg 2012) before being input to the jgi_summarize_bam_contig_depths script included with

the Metabat2 distribution (v2.12.1) (Kang et al., n.d.) to estimate and coverage and tetranucleotide frequency. To limit contamination and produce high-quality genomes, our binning strategies followed 3 steps. First, contigs were binned in parallel with Concoct (v0.4.1) (Alneberg et al. 2013), Metabat2 (v2.12.1) (Kang et al., n.d.) and Maxbin (v2.2.5) (Wu, Simmons, and Singer 2015) using the precomputed coverage and tetranucleotide frequencies. Then, a non-redundant set of bins combining the results of the 3 binning algorithms was obtained with DAS Tool (v1.1.0) (Sieber et al. 2018) and only Chloroflexi bins more then 75% complete and with less than 5% contamination, as assessed by CheckM (v1.0.11) (Parks et al. 2015), were kept (bin phylum was inferred from the CheckM lineage workflow and completeness was estimated with the taxonomic-specific workflow using the phylum Chloroflexi taxon_set). Finally, bins were refined with RefineM (v0.0.23) (Parks et al. 2017) using both the genomic property pipeline as well as the taxonomic assignment pipeline. Contigs from the resulting bins were scaffolded using Geneious (v10.2) assembler and reevaluated for contamination and completeness with CheckM, yielding 31 high-quality metagenomic-assembly genomes (referred simply as genomes in the text).

Taxonomic analysis

Genomes from isolated Chloroflexi and closely related environmental Chloroflexi were downloaded from NCBI and ggkBase. In the absence of 16S rRNA gene in most of the genomes, PhyloPhlan (v0.99) (Segata et al. 2013) was used to impute the taxonomy of the 31 new Chloroflexi genomes by placing them on the tree of life using a concatenated alignment of up to

400 of the most conserved proteins in prokaryotes. Additionally, PhyloPhlan computed a bootstrapped phylogenetic tree of the new genomes and their closely related isolated or environmental Chloroflexi.

Metagenome annotation and functional gene analysis

Protein coding sequence (CDS) were determined with Prokka (v1.12) (Seemann 2014) using a custom protein database containing all Chloroflexi proteins available in the NCBI NR database (February 2019). For each CDS, functional annotation resulted from the combination of sequence similarity searches against the KEGG database using BlastKoala (Kanehisa, Sato, and Morishima 2016) and Hidden-Markov Models from PFAM and TIGRFAM using the trusted cut-off score of each model. Additionally, orthologous groups of proteins across the 31 genomes were computed using OrthoMCL (v2.0.9) (Li, Stoeckert, and Roos 2003) (with default parameters). When only some proteins in an orthologous group shared an annotation and the others were labelled as "hypothetical proteins", the annotated function was chosen as final annotation for all proteins in the group. In addition, when proteins with different functional annotations were grouped together, manual curation was used to assess annotation likelihood using additional BlastP (Camacho et al. 2009) search similarity and InterProScan domain search (Jones et al. 2014; Mitchell et al. 2019). CDS and predicted functions were also subject to manual inspection to merge the functional annotations of the different analysis described above, and additional curation using InterPro domain search or sequence-similarity search against the NCBI NR database or other custom BlastP databases was employed when needed. Specifically, a

BlastP database containing the *bzdNOPQWXY* genes from *Azoarcus sp.* strain CIB and the

bamABCDEFGHIRQ genes from *G. metallireducens* was created to search for BCR genes.

Glycohydrolases, glycotransferase and polysaccharide lyase were predicted using the dbCAN2

server (Zhang et al. 2018) and only functions identified with at least 2 tools were considered.

Functional annotations were organized into pathways from MetaCyc (Caspi et al. 2018), KEGG

(Kanehisa and Goto 2000) or literature.

Protein phylogenies

All protein phylogenetic trees were constructed as follows. Amino acid sequences of interest

were aligned using MUSCLE (Edgar 2004) and manually masked and curated to focus on shared

amino acid positions. The refined alignment was re-aligned with MUSCLE with the refinement

flag option. We then identified the best amino acid substitution model for this alignment with

RaxML (Stamatakis 2014) (using the -m PROTGAMMAAUTO flag). 30 maximum-likelihood

trees were computed under the chosen model (RaxML -#30) and the best tree was displayed with

support for each branch calculated from bootstrap analysis (RaxML -f a -# autoMRE).

Reductive dehalogenase phylogeny

Reductive dehalogenase sequences were selected from the reductive dehalogenase Pfam database

(PF13486) by clustering all sequences in the Pfam database at a 90% identity at the amino acid

level using CD-HIT (http://weizhongli-lab.org/cd-hit/servers.php) and choosing a representative

sequence for each cluster. A tree combining all resulting sequences with the novel Chloroflexi

ones was then computed as described above and is available in Supplementary Figure S4. To make Figure 5, nodes from the full tree were downsampled by taxonomy to only show 20% of the sequences per taxa.

Operon structure analysis

All CDS within a 5 gene window upstream and downstream of every MetF gene were extracted and the orthologous groups to which they belonged identified. Orthologous groups containing less than 3 genes were discarded to focus on shared protein functionality around metF genes. Presence and absence of protein from the selected orthologous groups in the gene window around each metF gene was recorded and the data clustered using binary distance. Varying the size of the gene window or the number of times a protein function needed to appear in total around all metF genes didn't change substantially the clustering pattern. A similar analysis was carried out for formate dehydrogenase genes with a window size of 7.

REFERENCES

Alexeeva, S., Hellingwerf, K.J., and Teixeira de Mattos, M.J. (2002) Quantitative assessment of
oxygen availability: perceived aerobiosis and its effect on flux distribution in the respiratory
chain of Escherichia coli. *J Bacteriol* **184**: 1402–1406.

Alneberg, J., Bjarnason, B.S., Bruijn, I. de, Schirmer, M., Quick, J., Ijaz, U.Z., et al. (2013)
CONCOCT: Clustering contigs on coverage and composition. *arXiv preprint arXiv:13124038*.

Anraku, Y. (1988) Bacterial electron transport chains. *Ann Rev Biochem* **57**: 101–132.

Atashgahi, S. (2019) Discovered by genomics: putative reductive dehalogenases with N-terminus
transmembrane helixes. *FEMS Microbiol Ecol* **95**: 13219.

Baker, B.J., Lazar, C.S., Teske, A.P., and Dick, G.J. (2015) Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**: 14.

Baughn, S.J.A.V., Anthony D. AND Garforth (2009) An anaerobic-type α-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of mycobacterium tuberculosis. *PLOS Pathog* **5**: 1–10.

Blazejak, A. and Schippers, A. (2010) High abundance of JS-1- and Chloroflexi-related Bacteria in deeply buried marine sediments revealed by quantitative, real-time PCR. *FEMS Microbiol Ecol* **72**: 198–207.

Buckel, W. and Thauer, R.K. (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**: 94–113.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A., et al. (2009) Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. *Microbiol Mol Biol Rev* **73**: 71–133.

Caspi, R., Billington, R., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., et al. (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* **46**: D633-D639.

D'mello, R., Hill, S., and Poole, R.K. (1996) The cytochrome bd quinol oxidase in escherichia coli has an extremely high oxygen affinity and two oxygen-binding haems: Implications for regulation of activity in vivo by oxygen inhibition. *Microbiology* **142**: 755–763.

Dodsworth, J.A., Gevorkian, J., Despujos, F., Cole, J.K., Murugapiran, S.K., Ming, H., et al. (2014) Thermoflexus hugenholtzii gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the Chloroflexi, Thermoflexia classis nov., and description of Thermoflexaceae fam. nov. and Thermoflexales ord. nov. *Int J Syst Evol Microbiol* **64**: 2119–2127.

Dombrowski, N., Seitz, K.W., Teske, A.P., and Baker, B.J. (2017) Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* **5**: 106.

Durante-Rodríguez, G., Gómez-Álvarez, H., Blázquez, B., Fernández-Llamosas, H., Martín-Moldes, Z., Sanz, D., et al. (2018) Chapter 13 Anaerobic Pathways for the Catabolism of Aromatic Compounds. In *Lignin valorization: Emerging approaches*. The Royal Society of Chemistry, pp. 333–390.

Durbin, A.M. and Teske, A. (2011) Microbial diversity and stratification of south pacific abyssal marine sediments. *Environ Microbiol* **13**: 3219–3234.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Fincker, M. and Spormann, A.M. (2017) Biochemistry of Catabolic Reductive Dehalogenation. *Ann Rev Biochem* **86**: 357–386.

Fullerton, H., Moyer, C.L., and Löffler, F.E. (2016) Comparative Single-Cell Genomics of Chloroflexi from the Okinawa Trough Deep-Subsurface Biosphere. *Appl Environ Microbiol* **82**: 3000–3008.

Futagami, T., Morono, Y., Terada, T., Kaksonen, A.H., and Inagaki, F. (2009) Dehalogenation activities and distribution of reductive dehalogenase homologous genes in marine subsurface sediments. *Appl Environ Microbiol* **75**: 6905–6909.

Galambos, D., Anderson, R.E., Reveillaud, J., and Huber, J.A. (2019) Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. *Environ Microbiol* **21**: 4395–4410.

Gottschalk, G., and Barker, H.A. (1966) Synthesis of glutamate and citrate by Clostridium kluyveri. A new type of citrate synthase. *Biochemistry* **5**: 1125-1133.

Holmes, D.E., Risso, C., Smith, J.A., and Lovley, D.R. (2011) Anaerobic oxidation of benzene by the hyperthermophilic archaeon Ferroglobus placidus. *Appl Environ Microbiol* **77**: 5926–5933.

Huang, H., Wang, S., Moll, J., and Thauer, R.K. (2012) Electron Bifurcation Involved in the Energy Metabolism of the Acetogenic Bacterium Moorella thermoacetica Growing on Glucose or H2 plus CO2. *J Bacteriol* **194**: 3689–3699.

Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., et al. (2013) Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.

Imachi, H., Sakai, S., Lipp, J.S., Miyazaki, M., Saito, Y., Yamanaka, Y., et al. (2014) Pelolinea submarina gen. nov., sp. nov., an anaerobic, filamentous bacterium of the phylum Chloroflexi isolated from subseafloor sediment. *Int J Syst Evol Microbiol* **64**: 812–818.

Inagaki, F., Nunoura, T., Nakagawa, S., Teske, A., Lever, M., Lauer, A., et al. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc Natl Acad Sci U S A* **103**: 2815–2820.

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

Jungbluth, S.P., Bowers, R.M., Lin, H.-T., Cowen, J.P., and Rappé, M.S. (2016) Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. *ISME J* **10**: 2033–2047.

Kale, V., Björnsdóttir, S.H., Frijónsson, Ó.H., Pétursdóttir, S.K., Ómarsdóttir, S., and Hreggvisson, G.Ó. (2013) Litorilinea aerophila gen. nov., sp. nov., an aerobic member of the class Caldilineae, phylum Chloroflexi, isolated from an intertidal hot spring. *Int J Syst Evol Microbiol* **63**: 1149–1154.

Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726–731.

Kang, D., Li, F., Kirton, E.S., Thomas, A., Egan, R.S., An, H., and Wang, Z. (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.

Kaster, A.-K., Mayer-Blackwell, K., Pasarelli, B., and Spormann, A.M. (2014) Single cell genomic study of *Dehalococcoidetes* species from deep-sea sediments of the Peruvian Margin. *ISME J* **8**: 1831–1842.

Kawaichi, S., Ito, N., Kamikawa, R., Sugawara, T., Yoshida, T., and Sako, Y. (2013) Ardenticatena maritima gen. nov., sp. nov., a ferric iron- and nitrate-reducing bacterium of the

phylum 'Chloroflexi' isolated from an iron-rich coastal hydrothermal field, and description of Ardenticatenia classis nov. *Int J Syst Evol Microbiol* **63**: 2992–3002.

Krajmalnik-Brown, R., Sung, Y., Ritalahti, K.M., Saunders, F.M., and Löffler, F.E. (2007) Environmental distribution of the trichloroethene reductive dehalogenase gene (tceA) suggests lateral gene transfer among Dehalococcoides. *FEMS Microbiol Ecol* **59**: 206–214.

Kublik, A., Deobald, D., Hartwig, S., Schiffmann, C.L., Andrades, A., Bergen, M. von, et al. (2016) Identification of a multi-protein reductive dehalogenase complex in Dehalococcoides mccartyi strain CBDB1 suggests a protein-dependent respiratory electron transport chain obviating quinone involvement. *Environ Microbiol* **18**: 3044–3056.

Lamrabet, O., Pieulle, L., Aubert, C., Mouhamar, F., Stocker, P., Dolla, A. and Brasseur, G., (2011) Oxygen reduction in the strict anaerobe Desulfovibrio vulgaris Hildenborough: characterization of two membrane-bound oxygen reductases. *Microbiology*, **157**: 2720-2732.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* **9**: 357.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.

Li, F., Hagemeier, C.H., Seedorf, H., Gottschalk, G., and Thauer, R.K. (2007) Re-citrate synthase from Clostridium kluyveri is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J Bacteriol* **189**: 4299-4304.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**: 2178–2189.

Lin, H.T., Cowen, J.P., Olson, E.J., Amend, J.P. and Lilley, M.D. (2012) Inorganic chemistry, gas compositions and dissolved organic carbon in fluids from sedimented young basaltic crust on the Juan de Fuca Ridge flanks. *Geochim Cosmochim Acta* **85**: 213–227.

Lohner, S.T. and Spormann, A.M. (2013) Identification of a reductive tetrachloroethene dehalogenase in Shewanella sediminis. *Philos Trans R Soc Lond B Biol Sci* **368**: 20120326–20120326.

Löffler, F.E., Yan, J., Ritalahti, K.M., Adrian, L., Edwards, E.A., Konstantinidis, K.T., et al. (2013) Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov. and family Dehalococcoidaceae fam. nov., within the phylum Chloroflexi. *Int J Syst Evol Microbiol* **63**: 625–635.

Maillard, J., Regeard, C., and Holliger, C. (2005) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of Desulfitobacterium hafniense strain TCE1. *Environ Microbiol* **7**: 107–117.

Marco-Urrea, E., Paul, S., Khodaverdi, V., Seifert, J., Bergen, M. von, Kretzschmar, U., and Adrian, L. (2011) Identification and characterization of a re-citrate synthase in Dehalococcoides strain CBDB1. *J Bacteriol* **193**: 5171-5178.

McIlroy, S.J., Karst, S.M., Nierychlo, M., Dueholm, M.S., Albertsen, M., Kirkegaard, R.H., et al. (2016) Genomic and *in situ* investigations of the novel uncultured Chloroflexi associated with 0092 morphotype filamentous bulking in activated sludge. *ISME J* **10**: 2223–2234.

McIlroy, S.J., Kirkegaard, R.H., Dueholm, M.S., Fernando, E., Karst, S.M., Albertsen, M., and Nielsen, P.H. (2017) Culture-Independent Analyses Reveal Novel Anaerolineaceae as Abundant Primary Fermenters in Anaerobic Digesters Treating Waste Activated Sludge. *Front Microbiol* **8**: 1134.

McMurdie, P.J., Hug, L.A., Edwards, E.A., Holmes, S., and Spormann, A.M. (2011) Site-Specific Mobilization of Vinyl Chloride Respiration Islands by a Mechanism Common in Dehalococcoides. *BMC Genomics* **12**: 287.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**: D351-D360.

Mock, J., Wang, S., Huang, H., Kahnt, J., and Thauer, R.K. (2014) Evidence for a Hexaheteromeric Methylenetetrahydrofolate Reductase in Moorella thermoacetica. *J Bacteriol* **196**: 3303–3314.

Mondeel, T.D.G.A., Rehman, S., Zhang, Y., Verma, M., Dürre, P., Barberis, M., and Westerhoff, H.V. (2016) Maps for when the living gets tough: Maneuvering through a hostile energy landscape. *IFAC-PapersOnLine* **49**: 364–370.

Müller, B., Sun, L., and Schnürer, A. (2013) First insights into the syntrophic acetate-oxidizing bacteria a genetic study. *Microbiologyopen* **2**: 35–53.

Müller, V., Imkamp, F., Biegel, E., Schmidt, S., and Dilling, S. (2008) Discovery of a ferredoxin:NAD+-oxidoreductase (Rnf) in Acetobacterium woodii: a novel potential coupling site in acetogens. *Ann N Y Acad Sci* **1125**: 137–146.

Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D.R., Ebrahim, A., and Zengler, K. (2013) Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. *Microb Cell Fact* **12**: 118.

Nakahara, N., Nobu, M.K., Takaki, Y., Miyazaki, M., Tasumi, E., Sakai, S., et al. (2019) Aggregatilinea lenta gen. nov., sp. nov., a slow-growing, facultatively anaerobic bacterium isolated from subseafloor sediment, and proposal of the new order Aggregatilineales ord. nov. within the class Anaerolineae of the phylum Chloroflexi. *Int J Syst Evol Microbiol* **69**: 1185–1194.

Nobu, M.K., Narihiro, T., Hideyuki, T., Qiu, Y.L., Sekiguchi, Y., Woyke, T., et al. (2015) The genome of Syntrophorhabdus aromaticivorans strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol* **17**: 4861–4872.

Nobu, M.K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S.G., Woyke, T., and Liu, W.T. (2015) Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J* **9**: 1710–1722.

Nunoura, T., Hirai, M., Miyazaki, M., Kazama, H., Makita, H., Hirayama, H., et al. (2013) Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine Chloroflexi bacterium from a Chloroflexi-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for Thermomarinilinea lacunofontalis gen. nov., sp. nov. *Microbes Environ* **28**: 228–235.

Ostermeier, C., Iwata, S., and Michel, H. (1996) Cytochrome c oxidase. *Curr Opin Struct Biol* **6**: 460–466.

Parkes, R.J., Cragg, B., Roussel, E., Webster, G., Weightman, A., and Sass, H. (2014) A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Mar Geol* **352**: 409–425.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., et al. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**: 1533.

Petro, C., Zäncker, B., Starnawski, P., Jochum, L.M., Ferdelman, T.G., Jørgensen, B.B., et al. (2019) Marine deep biosphere microbial communities assemble in near-surface sediments in aarhus bay. *Front Microbiol* **10**: 758.

Pfeiffer T., and Bonhoeffer S. (2004) Evolution of cross-feeding in microbial populations. *Am Nat*. **163**:E126–E135.

Pfeiffer, T., Schuster, S., and Bonhoeffer, S. (2001) Cooperation and competition in the evolution of atp-producing pathways. *Science* **292**: 504–507.

Pierce, E., Xie, G., Barabote, R.D., Saunders, E., Han, C.S., Detter, J.C., et al. (2008) The complete genome sequence of Moorella thermoacetica (f. Clostridium thermoaceticum). *Environ Microbiol* **10**: 2550–2573.

Podosokorskaya, O.A., Bonch-Osmolovskaya, E.A., Novikov, A.A., Kolganova, T.V., and Kublanov, I.V. (2013) Ornatilinea apprima gen. nov., sp. nov., a cellulolytic representative of the class Anaerolineae. *Int J Syst Evol Microbiol* **63**: 86–92.

Ragsdale, S.W., and Ljungdahl, L.G. (1984) Purification and properties of NAD-dependent 5, 10-methylenetetrahydrofolate dehydrogenase from Acetobacterium woodii. *J Biol Chem* **259**: 3499-3503.

Ragsdale, S.W. and Pierce, E. (2008) Acetogenesis and the wood–ljungdahl pathway of co2 fixation. *Biochim Biophys Acta* **1784**: 1873–1898.

Refojo, P.N., Teixeira, M., and Pereira, M.M. (2012) The Alternative complex III: properties and possible mechanisms for electron transfer and energy conservation. *Biochim Biophys Acta* **1817**: 1852–1859.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, S., et al. (2016) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the mid-cayman rise. *Environ Microbiol* **18**: 1970–1987.

Schägger, H. (2002) Respiratory chain supercomplexes of mitochondria and bacteria. *Biochim Biophys Acta* **1555**: 154–159.

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

Schuchmann, K. and Müller, V. (2016) Energetics and Application of Heterotrophy in Acetogenic Bacteria. *Appl Environ Microbiol* **82**: 4056–4069.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Segata, N., Börnigen, D., Morgan, X.C., and Huttenhower, C. (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**: 2304.

Sewell, H.L., Kaster, A.-K., and Spormann, A.M. (2017) Homoacetogenesis in Deep-Sea Chloroflexi, as Inferred by Single-Cell Genomics, Provides a Link to Reductive Dehalogenation in Terrestrial Dehalococcoidetes. *mBio* **8**: e02022–17.

Sieber, C.M., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**: 836.

Skennerton, C.T., Chourey, K., Iyer, R., Hettich, R.L., Tyson, G.W., and Orphan, V.J. (2017) Methane-Fueled Syntrophy through Extracellular Electron Transfer: Uncovering the Genomic Traits Conserved within Diverse Bacterial Partners of Anaerobic Methanotrophic Archaea. *mBio* **8**: 725.

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

Starnawski, P., Bataillon, T., Ettema, T.J.G., Jochum, L.M., Schreiber, L., Chen, X., Lever, M.A., et al. (2017) Microbial community assembly and evolution in subseafloor sediment. *Proc Natl Acad Sci U S A* **114**: 2940-2945.

Tabor, H. and Wyngarden, L. (1959) The enzymatic formation of formiminotetrahydrofolic acid, 5,10-methenyltetrahydrofolic acid, and 10-formyltetrahydrofolic acid in the metabolism of formiminoglutamic acid. *J Biol Chem* **234**: 1830–1846.

Teske, A., Beer, D. de, McKay, L.J., Tivey, M.K., Biddle, J.F., Hoer, D., et al. (2016) The Guaymas Basin Hiking Guide to Hydrothermal Mounds, Chimneys, and Microbial Mats: Complex Seafloor Expressions of Subsurface Hydrothermal Circulation. *Front Microbiol* **7**: 75.

Vignais, P.M., Billoud, B., and Meyer, J. (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* **25**: 455-501.

Visser, M., Worm, P., Muyzer, G., Pereira, I.A.C., Schaap, P.J., Plugge, C.M., et al. (2013) Genome analysis of Desulfotomaculum kuznetsovii strain 17 T reveals a physiological similarity with Pelotomaculum thermopropionicum strain SI T. *Stand Genomic Sci* **8**: 69–87.

Wang, S., Huang, H., Kahnt, J., and Thauer, R.K. (2013) Clostridium acidurici Electron-Bifurcating Formate Dehydrogenase. *Appl Environ Microbiol* **79**: 6176–6179.

Wasmund, K., Cooper, M., Schreiber, L., Lloyd, K.G., Baker, B.J., Petersen, D.G., et al. (2016) Single-Cell Genome and Group-Specific dsrAB Sequencing Implicate Marine Members of the Class Dehalococcoidia (Phylum Chloroflexi) in Sulfur Cycling. *mBio* **7**: e00266–16.

Wasmund, K., Schreiber, L., Lloyd, K.G., Petersen, D.G., Schramm, A., Stepanauskas, R., et al. (2014) Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. *ISME J* **8**: 383–397.

Wohlfarth, G., and Diekert, G. (1991) Thermodynamics of methylenetetrahydrofolate reduction to methyltetrahydrofolate and its implications for the energy metabolism of homoacetogenic bacteria. *Arch Microbiol* **155**: 378-381.

Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2015) MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607.

Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H., and Kamagata, Y. (2006) *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. *Int J Syst Evol Microbiol* **56**: 1331–1340.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018) DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95–W101.

Zhang, Q., Iwasaki, T., Wakagi, T., and Oshima, T. (1996) 2-Oxoacid:Ferredoxin Oxidoreductase from the Thermoacidophilic Archaeon, Sulfolobus sp. Strain 71. *J Biochem* **120**: 587–599.

FIGURE LEGENDS

Figure 1

Reconstructed phylogeny and metabolism of the 31 subseafloor Chloroflexi. A) Maximum likelihood phylogeny of the reconstructed Chloroflexi genomes and their closest relatives based on PhyloPhlan alignment of conserved bacterial proteins and using the VT + Gamma model. Branch support values > 50 and > 95 are denoted by an empty and black diamonds, respectively.

See Methods for details on alignment and phylogeny construction. Sequences from this study are colored according to their sampling locations, sequences derived from metagenomic studies are marked with a star (see (Hug et al. 2013) for RBG genomes, and (Sewell, Kaster, and Spormann 2017) for Dsc genomes), and sequences from isolated Chloroflexi are unmarked. B) Predicted metabolism of the subseafloor Chloroflexi and their closest isolated relatives. Circles with a cross denote experimental data from culture of *A. thermolimosa* (Yamada et al. 2006), *P. submarina* (Imachi et al. 2014), *B. fermentans* (McIlroy et al. 2017), Candidatus *P. breve* (McIlroy et al. 2016), *T. hugenholtzii* (Dodsworth et al. 2014), *L. aerophila* (Kale et al. 2013), *A. martima* (Kawaichi et al. 2013) and *D. mccartyi* (Löffler et al. 2013). Genomes belonging to the two Aerobic Clusters defined in the text are boxed.

Figure 2

Overview of the Wood-Ljungdahl pathway (WLP) as a catabolic reductive pathway in acetogenic bacteria and subseafloor Chloroflexi. A) Heterotrophic acetogenesis can be divided in three modules. The oxidation of organic substrates (e.g. glucose) releases CO_2, and reducing equivalents in the form of NAD(P)H or reduced ferredoxin, as well as ATP via substrate level phosphorylation. In the presence of a membrane-spanning Rnf, Hdr-IFO or Ech complex, ferredoxin oxidation can be coupled to energy conservation. The WLP acts as an electron sink for 8 reducing equivalents per 2 CO_2. B) The ATP-neutral WLP is a linear pathway reducing 2 CO_2 to methyl-tetrahydrofolate (THF) and CO, respectively, before condensation to acetyl-CoA.

The latter is subsequently converted to acetate and ATP, which is consumed in the initial formyltetrahydrofolate synthase reaction. C) Genes encoding key complexes for electron bifurcation or reverse electron transport such as NfnAB, Fdh, PntAB and potentially MetF were found in the genomes. D) Full gene sets for the WLP were found in the majority of the genomes. Abbreviations: FdhA: formate dehydrogenase, Fhs: formyltetrahydrofolate synthase, FolD: bifunctional methenyl-tetrahydrofolate cyclohydrolase and reductase, MetF: methylene-tetrahydrofolate reductase, CODH / ACS: carbon monoxide dehydrogenase / acetyl-CoA synthase, ACS-I: acetyl-CoA synthetase.

Figure 3

Shared methylenetetrahydrofolate reductase operon architecture. Sequence names in bold gray and bold black come from Branch 1 and Branch 2 genomes respectively. NCBI accession number for MetF sequences from model acetogenic bacteria: *A. longum* - WP_004092543.1, *A. woodii* - WP_014355325.1, *C. ljungdahlii* - WP_013240365.1, *M. thermoacetica* - WP_011392705.1. See methods for clustering of operon architecture.

Figure 4

Key genetic markers for aerobic metabolism. Genomes are ordered according to their phylogeny. Genomes encoding the potential for oxygen respiration cluster in two groups defined as Aerobic Cluster 1 and 2 (AC_1 and AC_2).

Figure 5

Maximum likelihood phylogeny of reductive dehalogenases found in the genomes using the PMB gamma amino acid substitution model. Sequences from isolated members of the Dehalococcoidia Class are marked with a star. Reductive dehalogenase sequences have been recovered from a wide range of phyla and environments, including activated sludge, compost, contaminated-aquifer sediments, deep-sea sediments and forest soil. All known obligate organohalide respirers belong to the Dehalococcoidia Class, with the exception of the Firmicute *Dehalobacter restrictus* (Fincker and Spormann 2017). Sequences shown here were subsampled from Supplementary Fig. S5.

Figure 6

Proposed metabolism of Chloroflexi in anoxic marine subsurface environments. A) In terrestrial anoxic (methanogenic) high carbon-flux environments, organic matter is degraded by consortia of anaerobic microorganisms with short catabolic pathways which are favored under competitive conditions. In such consortia contain primary fermenters, secondary fermenters (syntrophs),

acetogenic and methanogenic microorganisms, the overall organic matter degradation depends on a flux of electrons between microorganisms often in the form of interspecies H_2 or formate transfer. The acetogens involved, carrying the WLP, are often specialized in H_2 consumption. The overall pathway for conversion of organic matter to acetate is speciated into distinct microorganisms. B) Subseafloor Chloroflexi contain pathways of fermenters and acetogens integrated in one microorganism in conjunction with the CO_2-reductive Wood Ljungdahl pathway which enables a more efficient use of a catabolic substrate. Such efficient metabolism is independent of an obligate interspecies H_2/formate transfer to H_2/formate-consuming microorganisms, because the WLP consumes electrons from oxidative pathways internally during CO_2 reduction to acetate. Such integrated metabolism with an overall longer pathway may be more advantageous in a carbon- and mixing-limited environment.

SUPPLEMENTARY MATERIAL

Supplementary Figure S1. Location of the sampling sites. The number of reconstituted genomes at each site is reported in parenthesis.

Supplementary Figure S2. Maximum likelihood tree of the DMSO reductase protein superfamily. Genomes are colored according to their phylogenetic group and clades of interest are highlighted. Most proteins from the Chloroflexi genomes did not share a high enough sequence similarity to infer their function.

Supplementary Figure S3. Pathways described in the text.

Supplementary Figure S4. Shared formate dehydrogenase operon architecture. Sequence names in bold gray and bold black come from Branch 1 and Branch 2 genomes respectively. Gene names in italic are from model acetogenic bacteria. NCBI accession number for sequences from *D. thermosubterraneus*: SHJ75222.1, *B. producta*: WP_054352386., *C. acidiurici*: WP_014969038 and *C. ljungdahlii*: WP_013238658.1.

Supplementary Figure S5. Maximum likelihood tree of the reductive dehalogenase protein family.

Supplementary Table S1. Description of sampling locations and associated metagenomic sequencing libraries.

Supplementary Table S2. Assembly and Genome statistics.

Supplementary Table S3. Gene ID of genes involved in aerobic respiratory metabolism.

Supplementary Table S4. Gene ID of genes associated with the WLP.

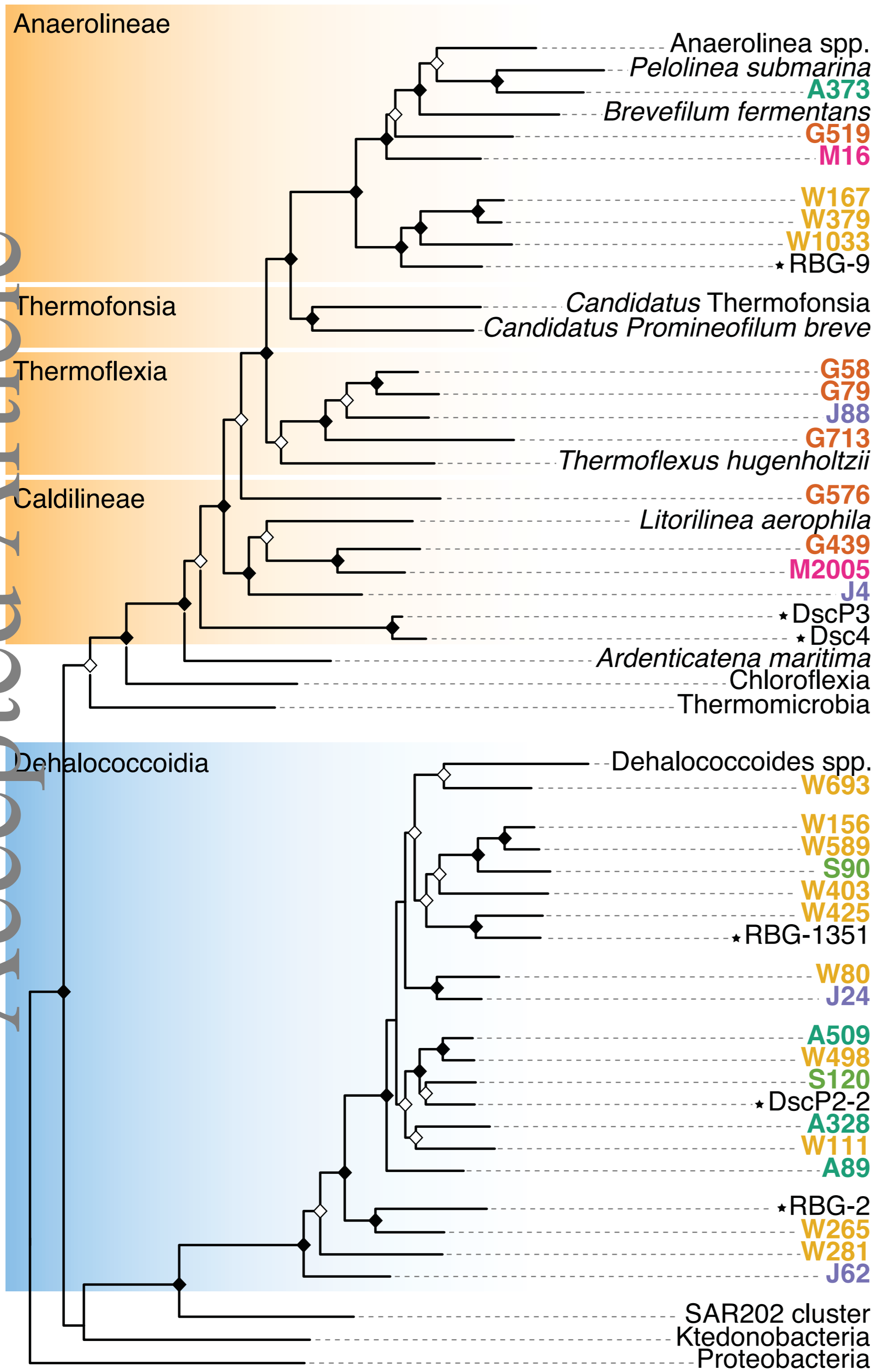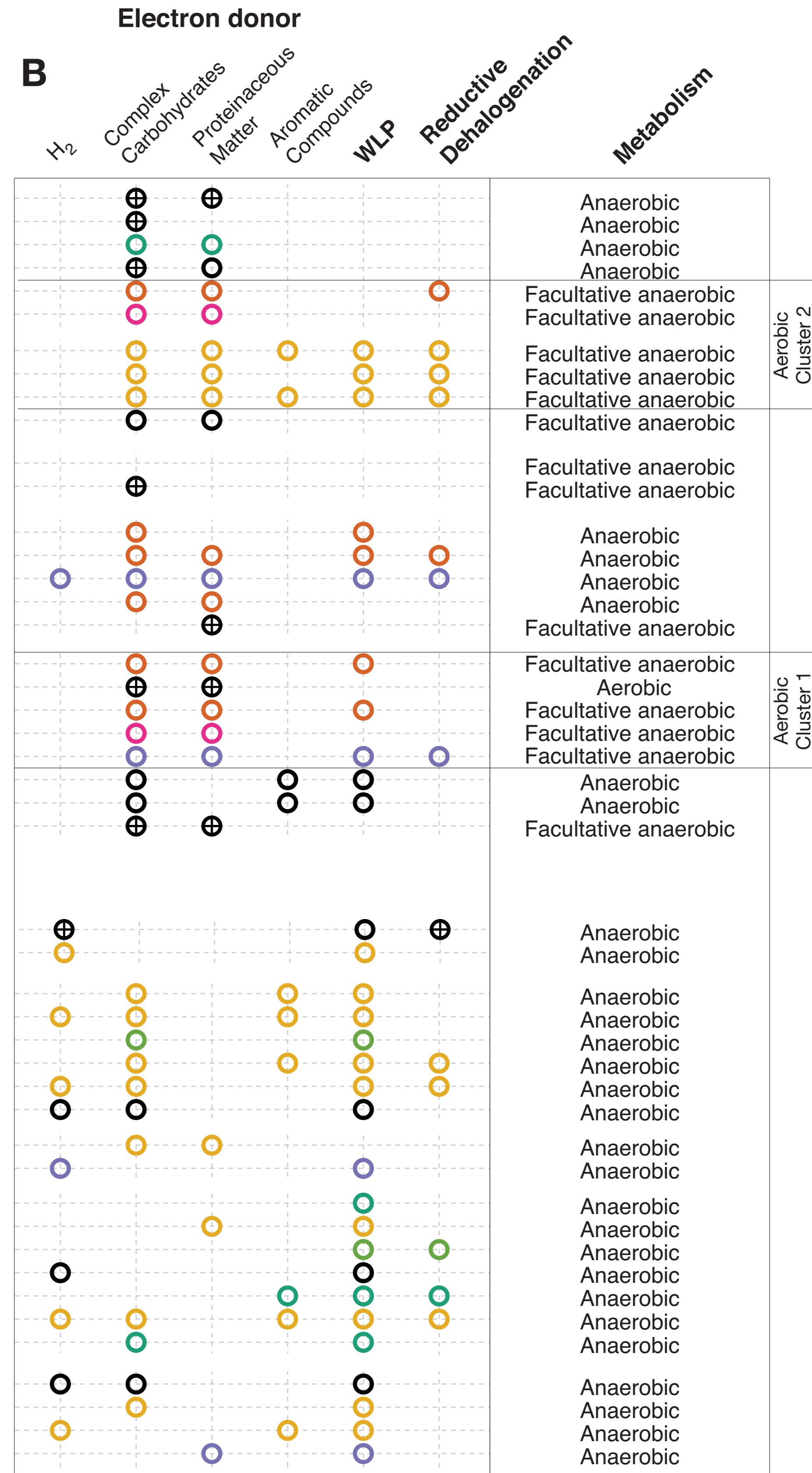**Supplementary Table S5.** Gene ID of genes involved in aromatic compound degradation.

Supplementary Table S6. Gene ID of genes involved in carbohydrate metabolism.

Supplementary Table S7. Gene ID of the genes involved in amino acid degradation.

Supplementary Table S8. Characteristics of the novel reductive dehalogenase sequences from the genomes.

Supplementary Table S9. Gene ID of hydrogenase genes.

A Catabolism in acetogenic bacteria

Oxidative Branch

Organic compound
- n Fd^{2-}
- m NADH
- p NADPH
- r CO$_2$

Acetyl-CoA → ATP

Acetate

Cation Gradient Generation

2 Na$^+$/H$^+$ 2 Na$^+$/H$^+$

Rnf / Hdr-Ifo   Ech

Fd^{2-} → NADH   Fd^{2-} → H$_2$

Reductive Branch - WLP

2 CO$_2$
- x Fd^{2-}
- y NADH
- z NADPH

Acetate

B Wood-Ljungdahl Pathway (WLP)

CO$_2$ CO$_2$

2 [H] → **Formate dehydrogenase - FdhA**

Formate

THF + ATP → **Formyl-THF synthase - Fhs**
ADP + Pi

Formyl-THF

Formyl-THF cyclohydrolase - FolD
H$_2$O

Methenyl-THF

2 [H] → **Methylene-THF dehydrogenase - FolD**

Methylene-THF

2 [H] → **Methylene-THF reductase - MetF**

Methyl-THF   CO

2 [H]

CODH / ACS
CoA

THF

Acetyl-CoA

Phospho-transacetylase   ADP + Pi
ACS-I

Acetyl-phosphate

Acetate kinase   ATP + CoA

Acetate

C Redox balancing enzyme complexes

Fd^{2-} → Fd   NfnAB   2 NADP$^+$ → 2 NADPH
NADH → NAD$^+$

m H$^+$ PntAB   NADH → NAD$^+$ / NADPH → NADP$^+$

2 NADH → 2 NAD$^+$ MetF   Fd → Fd^{2-} / Methylene-THF → Methyl-THF

Fd^{2-} → Fd   Fdh   2 CO$_2$ → 2 formate
NADH → NAD$^+$
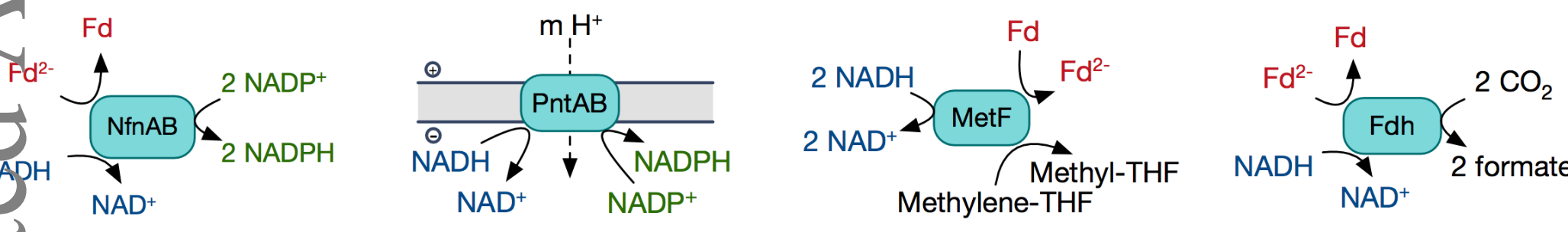
D WLP-associated genes in the genomes

count: 1, 2, 3, 4, 5

Phylogenetic tree (colored by taxonomic group)

gil960425208lreflWP_058317067.1l
gil972927093lgblKUJ77558.1l
gil960418993lreflWP_058310926.1l
gil497514669lreflWP_009828867.1l
gil849133745lreflWP_047995705.1l
gil750538686lreflWP_040820555.1l
gil931626697lemblCTQ57787.1l
gil746056058lreflWP_039124764.1l
gil495870429lreflWP_008595008.1l
gil931419475lgblKPK05944.1l
gil135555204lgblEBI39761.1l
W403_00806
W425_00659
W403_02978
W111_01559
gil643377553lreflWP_025205949.1l
gil643376909lreflWP_025205305.1l
gil553954273lgblAGY78736.1l
gil409021351lgblAFV03382.1l
gil648464677lreflWP_026156428.1l
gil648464917lreflWP_026156668.1l
gil643377521lreflWP_025205917.1l
gil504769700lreflWP_014956802.1l
gil918025442lreflWP_052315992.1l
W167_02463
G519_01141
J4_03630
gil780817349lgblKJS21523.1l
J88_02050
G79_02090
gil974226593lgblKUO67246.1l
gil808567446lemblCEP66756.1l
gil820748469lreflWP_046737363.1l
gil966393596lreflWP_058439793.1l
G519_01361
G519_00952
W1033_00986
gil410598001lgblEKQ52598.1l
gil167044780lgblABZ09449.1l
gil496312080lreflWP_009021258.1l
gil262117675ldbjlBAI47796.1l
gil262117720ldbjlBAI47826.1l
gil262117756ldbjlBAI47850.1l
gil960229032lgblKTB47891.1l
gil500355981lreflWP_011928759.1l
gil262117672ldbjlBAI47794.1l
gil822531487lreflWP_046961237.1l
gil752683582lreflWP_041331532.1l
gil262117687ldbjlBAI47804.1l
gil966390807lreflWP_058437571.1l
gil499628103lreflWP_011308837.1l
gil561138592lreflWP_023652806.1l
gil299528338lgblADJ26806.1l
gil752552061lreflWP_041223303.1l
gil553954906lgblAGY79043.1l
gil163644650lgblABY28318.1l
gil163644674lgblABY28330.1l
gil499629179lreflWP_011309913.1l
gil505220225lreflWP_015407327.1l
gil553954843lgblAGY79014.1l
gil502982668lreflWP_013217644.1l
gil499239657lreflWP_010937197.1l
gil736351256lreflWP_034377053.1l
gil452111224lgblAGG06956.1l
gil553953918lgblAGY78560.1l
gil124039603ldbjlBAF45397.1l
gil262117684ldbjlBAI47802.1l
gil820748252lreflWP_046737146.1l
gil966391758lreflWP_058438522.1l
gil966391762lreflWP_058438526.1l
gil499629242lreflWP_011309976.1l
gil499629251lreflWP_011309985.1l
A328_01037
gil919193982lreflWP_052749483.1l
gil553954910lgblAGY79045.1l
G519_01357
gil299528329lgblADJ26797.1l
gil3002555lgblAAC60788.1l
gil495789518lreflWP_008514097.1l
gil974222140lgblKUO63051.1l
gil917071391lreflWP_051678103.1l
gil974219278lgblKUO60296.1l
gil491788352lreflWP_005599817.1l
gil654652787lreflWP_028113772.1l
gil746201201lreflWP_039258541.1l
gil648290750lreflWP_026071899.1l
gil643377529lreflWP_025205925.1l
gil504604938lreflWP_014792040.1l
gil652986928lreflWP_027239508.1l
gil134493029lgblEBB67317.1l
gil671754831lgblAll87706.1l
gil142842448lgblEDA99384.1l
gil496312165lreflWP_009021343.1l
gil917735424lreflWP_052249649.1l
gil136280620lgblEBN10068.1l
S120_01545
S120_01544
S120_00922
G79_01539
W379_01907

Legend:

- α-proteobacteria
- β-proteobacteria
- Chloroflexi
- δ/ε-proteobacteria
- Euryarchaeota
- Firmicutes
- γ-proteobacteria
- Environmental sequence
- Novel Chloroflexi sequence
- Group 1 genome
- Group 2 genome