Genomic insights into the Archaea inhabiting an Australian radioactive legacy site

Xabier Vázquez-Campos[1*], Andrew S. Kinsela[2,3], Mark W. Bligh[2,3], Timothy E. Payne[3], Marc R. Wilkins[1] & T. David Waite[2*]

[1] NSW Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales 2052, Australia

[2] UNSW Water Research Centre and School of Civil and Environmental Engineering, The University of New South Wales, Sydney, New South Wales 2052, Australia

[3] Institute for Environmental Research, Australian Nuclear Science and Technology Organisation, Locked Bag 2001, Kirrawee DC, New South Wales 2232, Australia

* Corresponding authors:

Running title: Archaea from Little Forest

Abstract

During the 1960s, small quantities of radioactive materials were co-disposed with chemical waste at the Little Forest Legacy Site (LFLS, Sydney, Australia). The microbial function and population dynamics during a rainfall event using shotgun metagenomics has been previously investigated. This revealed a broad abundance of candidate and potentially undescribed taxa in this iron-rich, radionuclide-contaminated environment.

Here, applying genome-based metagenomic methods, we recovered 37 refined archaeal bins (≥50% completeness, ≤10% redundancy) from 10 different major lineages. They were, for the most part, included in 4 proposed lineages within the DPANN supergroup (LFWA-I to IV) and *Methanoperedenaceae*.

The new *Methanoperedens* spp. bins, together with previously published data, suggests a potentially widespread ability to use nitrate (or nitrite) and metal ions as electron acceptors during the anaerobic oxidation of methane by *Methanoperedens* spp.

While most of the new DPANN lineages show reduced genomes with limited central metabolism typical of other DPANN, the candidate species from the proposed LFWA-III lineage show some unusual features not often present in DPANN genomes, i.e. a more comprehensive central metabolism and anabolic capabilities. While there is still some uncertainty about the capabilities of LFW-121_3 and closely related archaea for the biosynthesis of nucleotides *de novo* and amino acids, it is to date the most promising candidate to be a *bona fide* free-living DPANN archaeon.

Keywords: Archaea, DPANN, metagenomics, genomics, phylogenomics, ANME-2d

Introduction

The Little Forest Burial Ground, now termed the Little Forest Legacy Site (LFLS), was a low-level radioactive waste (LLRW) disposal site active between 1960 and 1968. During this period, small quantities of both short and long-lived radionuclides (including plutonium and americium) were disposed of in three-metre deep, unlined trenches, as was common practice at the time [1]. Although the waste material was overtopped with locally sourced geological fill, the legacy trenches remain to this day orders of magnitude more porous than the surrounding clay and shale strata. Consequently, during extended rainfall events, atmospheric and surface waters fill up the legacy trenches and periodically (depending on preceding conditions) result in the trench water discharging into the surface and near-surface soils in a 'bath-tub'-like mechanism [2]. This mechanism has been proposed as a major process for exporting plutonium and americium from the legacy trenches [2].

Frequent filling and discharge events within the LFLS trenches and changes to the redox conditions affect the chemistry as well as the resident microbial community [3]. When oxygen-laden rainwater enters the trenches, the predominantly reducing redox conditions of the trench water switch to being partially oxic, altering a range of biogeochemical processes which impact upon contaminant mobility [3]. Repeated redox oscillations in mesocosm-like experiments across a range of depositional environments have been shown to alter iron and sulfur mineral phase solubility and crystallinity [4, 5], processes closely coupled to organic carbon mineralisation [6] that lead to variable rates of entrained contaminant mobilisation and retardation [7, 8].

The domain Archaea constitutes by far the most understudied domain of life. Archaea have been traditionally regarded as biological oddities and marginal players in the global geochemical elemental cycles, with most living in extreme environments [9]. However, the continuous development of cultivation-independent techniques in recent decades has completely changed this conception. For example, *Thaumarchaeota* Archaea are known now to be key elements in the global nitrogen cycle [10] and can even induce systemic resistance to pathogens in plants [11]; *Euryarchaeota* capable of anaerobic oxidation of methane (AOM) may consume up to 80-90% of the methane produced in

64 environments such as rice fields and ocean floors before it is released into the atmosphere [12]. Our
65 previous research showed that the archaeal community in the LFLS trenches was mainly composed of
66 methanogens and anaerobic methane oxidisers (ANME-2d) from the *Methanomicrobia* class and
67 DPANN archaea [3]. While they constituted a minor component of the community, they still may
68 have an important role in facilitating redox cycling and therefore impacting upon contaminant
69 mobilisation.

70 Here we describe the analysis of 37 new archaeal genomes with different degrees of completeness,
71 derived from samples collected in a legacy radioactive waste trench at the LFLS during a redox
72 cycling event. We present evidence of four new DPANN lineages and six non-conspecific
73 *Methanoperedens* sp. genomes, while exploring their potential role in the biogeochemical cycles and
74 uniqueness.

75 Material and Methods

76 Data source and experimental details

77 Raw sequencing reads from ENA project PRJEB14718 [3] were reanalysed, this time using
78 genome-based metagenomic methodologies in order to better understand the contributions of the
79 Archaea in the LFLS groundwater to the biogeochemistry of the site. Briefly, samples were collected
80 in triplicate over a period of 47 days at 4 time-points (0, 4, 21 and 47 days) after an intense rainfall
81 event that completely filled the trench. Thorough chemical and radiochemical analyses were
82 conducted on the samples in order to understand the biogeochemistry of the site [3]. Over this time
83 period, the redox conditions became increasingly reducing/anoxic and it was reflected in both the
84 chemical analyses and the community profile, with a clear increase in obligate anaerobes on days 21
85 and 47 [3].

Recovery and assessment of community genomes

QC and pre-processing of raw sequencing reads were performed with Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Co-assembly of the trimmed reads from all samples was performed with MEGAHIT v1.0.2 [13, 14].

Contigs >2.5kbp were binned with CONCOCT [15] within Anvi'o v2.3 [16]. Bins of archaeal origin were manually refined. Completeness and redundancy were estimated with Anvi'o single copy gene (SCG) collection, based on Rinke *et al.* [17] as well as with CheckM v1.0.13 [18]. Further details are available in the Supplementary Information (SI).

Phylogenomics and phylogenetics

Prodigal v2.6.3 (-p meta) was used to predict proteins from the archaeal bins and 242 archaeal reference genomes (Table S1). Predicted proteins were assigned to archaeal Clusters of Orthologous Groups (arCOGs) using the arNOG database of eggNOG v4.5.1 [19] using eggNOG-mapper v0.99.2 [20]. An original set of 44 arCOGs (rp44, Table S2) combining universal and archaeal-specific ribosomal proteins was created for the phylogenomic analysis based, to some extent, on Yutin et al. [21]. Further details are provided in SI.

Ribosomal proteins were individually aligned with MAFFT-L-INS-i v7.305b [22] and concatenated. Concatenated alignment was trimmed with BMGE v1.2 [23] based on [24]. The concatenated archaeal protein tree was constructed using the posterior mean site frequency (PMSF) model [25] in IQ-TREE v1.5.5 [26], using a base tree built under an LG model and a final tree under LG+C60+F+I+G with 1000 ultrafast bootstrap replicates [27].

Phylogeny of NarG-like proteins was created using as reference all proteins (416) from related orthologous groups from the EggNOG database, i.e.: arCOG01497, ENOG4102T1R, ENOG4102TDC, ENOG4105CRU and ENOG4108JIG. Protein sequences were clustered with CD-HIT v4.6 [28] and aligned with MAFFT-L-INS-i v7.305b [22]. Alignment was trimmed with BMGE

v1.2 [23] and tree build with IQ-TREE v1.5.5 under the recommended model (LG+R10) and 10,000 ultrafast bootstrap replicates [26].

Functional annotation

In addition to arCOGs, predicted proteomes were profiled with InterProScan v5.25-64.0 [29]. Key biogeochemical enzymes were identified based on signature hmm profiles in the InterProScan output or by custom hmm profiles not integrated in available databases [30, 31]. Custom hmm profiles were searched with HMMER v3.1b2 (-E 1e-20) [32]. Proteins with single HMMER matches below previously established thresholds [30, 31] were searched against Swissprot and checked against the InterProScan results to reduce the incidence of false negatives.

High-heme cytochromes (≥10 binding sites per protein) were predicted using a modified motif_search.py script (https://github.com/alexherns/biotite-scripts) that uses CX(1,4)CH as well as the canonical CXXCH heme-binding motif.

Carbohydrate-active enzymes were searched locally with hmmscan (HMMER v3.1b2) [32] using the CAZy-based [33] hmm library v5 from dbCAN [34].

Peptidases/proteases and transporters were annotated by similarity search of the predicted proteins (blastp) against the MEROPS database v11.0 [35] and TCDB (downloaded on 12 July 2018) [36], respectively.

The subcellular location of the proteins was predicted with PSORTb v3.0.6 [37].

Prediction of tRNA and rRNA genes was performed with tRNAscan-SE 2.0 [38] and Barrnap v0.9 (https://github.com/tseemann/barrnap) respectively.

Genome annotation

Metagenome assembled genomes were annotated with Prokka v1.12 [39]. Annotations were imported into PathwayTools v22.0 [40] for modelling. Selected proteomes were also submitted to GhostKOALA for KEGG annotation [41, 42].

Pangenomic analyses

Pangenomic analysis of the found '*Ca.* Methanoperedens spp.' genomes, as well as the nearly complete genomes of '*Ca.* M. nitroreducens' ANME-2D (JMIY00000000.1) [29], '*Ca.* M. ferrireducens' (PQAS00000000.1) [43], '*Ca.* M. nitroreducens' BLZ1 (also known as MPEBLZ) (LKCM00000000.1) [44], and '*Ca.* M. nitroreducens' Vercelli (GCA_900196725.1) [45], was performed with Anvi'o v5.4.0 [16] following the standard pangenomics workflow (http://merenlab.org/2016/11/08/pangenomics-v2). Protein-coding genes were clustered with an MCL inflation value of 6 [46]. Note that neither '*Ca.* M. nitroreducens' BLZ1 or '*Ca.* M. nitroreducens' Vercelli are actual '*Ca.* M. nitroreducens' based on ANI/AAI values and therefore we will refer to them as '*Ca.* Methanoperedens sp.' BLZ1 and '*Ca.* Methanoperedens sp.' Vercelli respectively to avoid confusion. New species names for these two MAGs are proposed in the SI.

For the pangenomic analysis of the LFWA-III lineage, only genomes ≥70%C from the rp44 dataset were considered, i.e. 6 newly obtained and 2 from references. Protein-coding genes were clustered with an MCL inflation values of 1.0, 1.5 and 2.0 [46].

Average Nucleotide Identity (ANI) and Average Amino acid Identity (AAI) were calculated with pyani [47] and CompareM v0.0.23 (https://github.com/dparks1134/CompareM) respectively, for both the *Methanoperedens* and LFWA-III genome datasets.

Proposed taxa

Proposed species and genera were registered in the Digital Protologue (http://imedea.uib-csic.es/dprotologue). Details on etymology, nomenclature and Digital Protologue registration numbers can be found in Table S3. Additional nomenclatural proposal changes above genus level are described in the Supplementary Information (see Nomenclatural novelties section). A summary of the proposed candidate species can be found in Table 1.

Data availability

Annotated assemblies are available at ENA under project PRJEB21808 and sample identifiers ERS2655284-ERS2655320. See details in Table S4. The output of the different analyses, including the reference genomes, are available at Zenodo (doi:10.5281/zenodo.3365725).

The methods section is described in further detail in the SI.

Results and Discussion

Binning results and archaeal community

De novo assembly and binning from LFLS trench subsurface water samples [3] generated 290 initial bins. Bins with clear archaeal identity and bins with ambiguous identity (similar completeness scores for bacterial and archaeal SCG profiles) were further refined with Anvi'o, producing a total of 37 draft archaeal genomes with completeness ≥50% (20 of which ≥70%C) and redundancy ≤10% (Table 2).

Phylogeny based on the 44 concatenated ribosomal proteins (Figure 1) showed that most metagenome assembled genomes (MAGs) belonged to diverse DPANN lineages and *Methanomicrobia* (*Methanoperedenaceae* and *Methanotrichaceae*) (Table 2). Phylogenomic analysis based on the rp44 (Figure 1) showed a tree topology largely consistent with current studies, e.g. *Asgardarchaeota* as

sister lineage to TACK and *Altiarchaeota* as sister to DPANN [9, 48]. Most high level branches showed UF bootstrap support values >90%, with the exception of the very basal Euryarchaeota, which is known to be difficult to resolve [49]. The position of the MAGs in the archaeal phylogeny, together with the 16S rRNA gene identities with described or proposed taxa, suggest four new lineages (Figure 1), denoted LFWA-I to -IV. LFWA-I lineage is a sister lineage to *Parvarchaeota* (ARMAN-4 and 5) [50, 51] and contains a unique MAG, LFW-252_1 ('*Ca.* Tiddalikarchaeum anstoanum') (Figure S1). LFWA-II is sister to the *Micrarchaeota*+LFWA-III clade and contains two MAGs, including the type LFW-144_1 ('*Ca.* Wianamattarchaeum fermentum') (Figure S2). LFWA-III is the closest branch to *Micrarchaeota* s. stricto, contains 11 MAGs from this work and is typified by LFW-121_3 ('*Ca.* Gugararchaeum adminiculabundum') (Figure 2). Lastly, LFWA-IV is the closest relative to *Aenigmarchaeota*, containing an unnamed MAG, LFW-46 (missing all rRNA genes aside from a partial 5S) (Figure S3). In a way to honour the traditional owners of the land where LFLS is based, many of the nomenclatural novelties proposed are based on terms from Aboriginal languages related to the site (Guringai and Dharawal), see Table S3 and SI.

During the drafting of this manuscript, a genome-based taxonomy was developed for Bacteria and Archaea (GTDB) [52]. Based on the GTDB (r89), LFWA-I would correspond to the order/family "o__CG07-land; f__CG07-land" (class *Nanoarchaeia*); LFWA-II, to "o__UBA8480" (class *Micrarchaeia*); LFWA-III, primarily to "o__UBA10214" (class *Micrarchaeia*); and, LFWA-IV could remain a new class within the yet-to-be-named phylum "p__EX4484-52".

The lineage LFWA-III was the most diverse of all Archaea encountered at LFLS with 11 bins belonging to 10 different genera based on AAI values (Tables S6 and S7). While this lineage contains the expected conserved SCG in their putative core genome (178-184 gene clusters), it included barely any other gene clusters (Figure 3 and Table S8). This was also evident regarding the singletons: >89% of the gene clusters were singletons covering >66% of all protein-coding genes, regardless of the inflation values used for clustering. The large proportion of singletons relates to the presumably fast evolutionary rates in DPANN Archaea [9] though the core genes appear to be far less affected as evidenced by their absence in the cloud or singleton pangenome gene sets.

The occurrence of methanogenic bins (3) from the *Methanotrichaceae* family (previously indicated to be the illegitimate *Methanosaetaceae* [3]) were principally, and unsurprisingly, detected during the highly anoxic phase (117-147 mV Standard Hydrogen Electrode). More interestingly, though, was the recovery of MAGs belonging to the genus '*Ca.* Methanoperedens' (simply *Methanoperedens* from here on) a group of Archaea capable of AOM, previously known as ANME-2d [53] or AAA (AOM-associated Archaea) [54]. A total of six MAGs were assembled in this study, four ≥90%C. Each of these MAGs belong to different species based on ANI and AAI scores (Figure 4, Table S9 and Table S10), constituting the most diverse set of *Methanoperedens* spp. genomes reconstructed to date in a single study, and all different to previously sequenced MAGs.

Based on coverage information, DPANN bins, especially those from the *Pacearchaeota* and LFWA-III groups, dominated the archaeal community with a maximum relative abundance of 82.5% of the entire archaeal community by day 4 (Table S11). *Euryarchaeota* bins (*Methanotrichaceae* and, especially, *Methanoperedenaceae*) do not become an important constituent prior to day 47, when all *Euryarchaeota* account for 42.8% of the archaeal community. Our previous analysis indicated that DPANN constituted a maximum of 55.8% of the archaeal community [3]. This discrepancy could be an artefact derived from the different copy numbers of rRNA gene clusters in most DPANN (likely limited to one) compared to other Archaea with more typically sized genomes containing multiple copies of the operon.

Carbon metabolism

The central carbon metabolism of the majority of Archaea in the LFLS trench waters was either host-dependent glycolytic/fermentative (DPANN) or methane cycling-related (methanogens and AOM). The predominant glycolytic pathway was suggested to be the Embden-Meyerhof-Parnas pathway, based on the lack of 2-keto-3-deoxygluconate kinase (KDG kinase) and 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPG aldolase), along with the absence of key Entner-Doudoroff pathway-enzymes in every MAG, excluding LFW-68_2.

Extracellular CAZymes in the archaeal bins were diverse: a total of 12 different glycoside hydrolase (GH), 9 carbohydrate binding module (CBM), 5 carbohydrate esterase (CE), and 3 polysaccharide lyase (PL) families were detected (Table S12). Proteins containing CBM from the CBM40 and CBM44 families were especially abundant.

Interestingly, within LFW-68_2, a total of 19 extracellular proteins were predicted to contain a CBM44 domain and 18 of those also contained a polycystic kidney disease (PKD) domain (Table S13). Despite its name, the PKD domains are relatively widespread, mediating protein-protein or protein-carbohydrate interactions in proteins such as collagenases and chitinases, and are also thought to be important components in proteins mediating cell to cell interactions in *Metazoa* [55]. Only a limited number of CBM44-containing proteins have been well-characterised to date (only 18 sequences, 2 characterised in the CAZy website, 1 August 2019). The CBM44-PKD domain combination has previously been described in the bifunctional cellulase/xyloglucanase of *Clostridium thermocellum* [56] where it binds cellulose or β1,4-glucans (branched or not, including xyloglucan), and the carbohydrate metabolism protein BT2081 of *Bacteroides thetaiotaomicron* [57]. However, it is likely that most of the CBM44-PKD containing proteins (12 out of the 18) are acting as proteases rather than on carbohydrates, given their similarity with proteases in the MEROPS database, mainly M09B (Table S13). Further details into the detritivorous/proteolytic lifestyle of LFW-68_2 are explored in the Protein degradation / catabolism section.

Carbon fixation is explored in detail in the SI.

The biosynthesis of polyhydroxyalkanoates is a common feature of *Methanoperedens* spp.

Polyhydroxyalkanoates (PHA) are carbon-rich, energetic polymers that result from the polymerisation of hydroxyalkanoates, e.g. 3-hydroxybutyrate, produced by numerous microorganisms under unfavourable conditions, particularly in association with unbalanced nutrient levels [58, 59]. The biosynthesis of PHA is a widespread characteristic in many groups of aerobic Bacteria, including photosynthetic [58] and methanotrophic bacteria [59]. However, very few strict anaerobes are able to

synthesise PHA, being mostly limited to syntrophic bacteria [60]. Genes involved in the biosynthesis of polyhydroxyalkanoates (PHA) were detected in three of the most complete *Methanoperedens* spp. genomes, i.e. LFW-24, LFW-280_3_1 and LFW-280_3_2, as well as in the *Thaumarchaeota* LFW-283_4_5 (Table S14).

In *Methanoperedens* spp., the genes associated with the biosynthesis of PHA appear grouped as a well conserved gene cluster (Figure 5A). In Archaea, the biosynthesis of PHA is known to occur, mainly, in many *Nitrososphaerales* (*Thaumarchaeota*) [60] and *Euryarchaeota*, where it has been traditionally limited to *Halobacteria* [61]. Although not explicitly reported in their respective manuscripts, PhaC (PHA synthase type III heterodimeric, TIGR01836) and PhaE (PHA synthase subunit E, PF09712) were predicted in all the *Methanoperedens* reference genomes included in this study [43–45, 62], and it was not until 2019 that experimental evidence was generated for the production of PHA by *Methanoperedens nitroreducens* [63]. However, while all PhaC proteins form a gene cluster in the pangenome analysis, the PhaE do not (Figure 4). This could be related to their function: PhaC is a catalytic protein while PhaE 'simply' regulates PhaC [61], so any sequence modifications in PhaC might have a more critical effect on the biosynthesis of PHA. Both proteins, PhaC and PhaE have also been predicted in three of the most complete new genomes in the present study (LFW-24, LFW-280_3_1 and LFWA-280_3_2). This suggests that the biosynthesis of PHA could be a widespread feature in *Methanoperedens* spp. or even in all *Methanoperedenaceae*, indicating a possible, generalised, role in the accumulation of excess carbon at times when the carbon source (methane) is much more abundant than other nutrients and/or trace elements [64].

It has been estimated that the anaerobic methane oxidisers may consume up to 80-90% of the methane produced in certain environments avoiding its release to the atmosphere [12]. The capacity of *Methanoperedens* spp. to accumulate PHA might have implications for the calculation of these estimates. The inference being that methane would not solely be used for energy production or to increase cell numbers, and that *Methanoperedens* spp. could act as a 'carbon-capture' device, especially within carbon-rich anoxic environments whenever they are the main anaerobic methane

277 oxidisers. This would likely be the case for the LFLS test trench, where ammonium, nitrate/nitrite and
278 total dissolved nitrogen are limiting.

279 *Methanoperedens* spp. acquired respiratory nitrate reductases at
280 least three different times

281 *Methanoperedenaceae* archaea have been reported (or at least suggested) to be able to utilise a wide
282 range of inorganic electron acceptors for AOM, e.g. nitrate [62], nitrite [44], Fe(III) [43, 65], Mn(IV)
283 [65] or even Cr(VI) [66]. The type species of the genus *Methanoperedens*, '*Ca.* Methanoperedens
284 nitroreducens', receives its name due to its ability to use nitrate as an electron acceptor [62]. However,
285 the alpha subunit of this candidate respiratory nitrate reductase ($NarG_{Mn}$) is a molybdopterin
286 oxidoreductase different from any canonical or putative NarG, with no match with TIGRFAM, CDD
287 or Pfam profiles characteristic of these proteins. The $NarG_{Mn}$ does not belong to the same orthologous
288 group alongside other typical archaeal NarG (arCOG01497), but to ENOG4102T1R, which includes
289 the well-characterised "dimethyl sulfoxide reductase subunit A" (DmsA) of several *Halobacteria*. The
290 $NarG_{Mn}$ has a high similarity (>80%) with proteins from both *Methanoperedens* sp. BLZ1
291 ($NarG1_{MBLZ1}$) and *Methanoperedens* sp. Vercelli ($NarG_{MV}$), and relates to the nitrite reductases of
292 *Nitrospira defluvii* (Figure 6). While *Methanoperedens* sp. Vercelli and *M. nitroreducens* have only
293 this unusual NarG, *Methanoperedens* sp. BLZ1 harbours an additional, unrelated, canonical NarG
294 ($NarG2_{MBLZ1}$) based on the detection by the TIGRFAM profile (TIGR01580, arCOG01497). All the
295 *Methanoperedens* spp. found at LFLS have similar putative $NarG_{LFLS}$, not present in the other
296 genomes, that belong to arCOG01497 and are identified as TIGR03479 (DMSO reductase family type
297 II enzyme, molybdopterin subunit). Interestingly, the putative $NarG_{LFLS}$ not only relates to the NarG
298 from *Haloferax mediterranei* (I3R9M9) [61] and other Halobacteria, but also to the PcrA of
299 *Dechloromonas aromatica* and *Azospira oryzae* (previously known as *Dechlorosoma suillum* strain
300 PS) [62]. It is worth mentioning that several of the NarG related to $NarG_{LFLS}$ have been confirmed to
301 have certain promiscuity (Figure 6) and known to be able to use (per)chlorate as substrate *in vivo* [67,
302 68]. Although it is close to impossible to predict the actual potential substrates of the $NarG_{LFLS}$ (or

303 $PcrA_{LFLS}$), the phylogenetic analysis is, at least, suggestive of the potential utilisation of (per)chlorate.
304 However, no chlorite dismutase gene was found in any of the LFLS *Methanoperedens* spp. genomes,
305 raising some uncertainty about this possibility.

306 Based on the observations made above, nitrate reductase(-like) enzymes are rather widespread in
307 *Methanoperedens* spp., with the exception of '*Ca.* M. ferrireducens', from which no nitrate reductase
308 candidate could be detected, aside from an orphan NapA-like protein (cd02754). Phylogenetic
309 analysis of the NarG and similar proteins (Figure 6) suggests that these proteins might have been
310 acquired at least three different times during the evolution of *Methanoperedens*. At the same time, the
311 pangenomic analysis showed the three different NarG variants in their respective gene clusters (Figure
312 4 and Figure 6).

313 Nitrate was rarely measured in noticeable concentrations in the LFLS test trench [3]. With a
314 maximum concentration of 0.55 μM during partially oxidising conditions (after the rainfall event), it
315 fell below detection limit (<0.01 μM) before any *Methanoperedens* became relatively abundant.
316 Historical disposal records of the waste deposited into the LFLS trenches list quantities of
317 perchlorate/perchloric acid as being disposed in the vicinity of the sample location. Despite its highly
318 oxidising capacity, perchlorate has been shown to be persistent in the environment [69].

319 The oxidation of methane linked to the bioreduction of perchlorate has been proposed a number of
320 times in the past [70–72]. Recent work by Xie *et al.* [71] and Wu *et al.* [72] suggest the possible
321 involvement of ANME archaea based on chemical and amplicon data analyses. Our results indicate
322 that this might be a reality, although further studies are needed to confirm this.

323 While the *Nitrosotalea*-related MAG LFW-283_4_5 would be expected to contribute to the nitrogen
324 cycle as ammonia oxidising archaea, the recovered bin is partial, so we were not able to predict any
325 proteins involved in this process (Table S14). No other MAG showed any indication of dissimilatory
326 pathways for inorganic nitrogen compounds.

MAG LFW-68_2 is a versatile protein-degrading Archaea

In settings where assimilable nitrogen can be scarce, as is the case at LFLS, being able to fix N_2 can be advantageous. Dinitrogenase components I and II were only detected in the *Methanoperedens* spp. MAGs with ≥90%C and also within a single methanogen (LFW-151_2, Table S14). While the dinitrogenase in LFW-151_2 can be easily defined as [MoFe]-nitrogenase, the nitrogenases of the *Methanoperedens* spp. MAGs do not match any specific type based on the metal centre. This is concordant with previous work on the metal centre of nitrogenases where the anaerobic methane oxidising Archaea were classified as "undefined" type nitrogenases [74] and likely belonging to the Nif-D lineage [75].

However, nitrogen fixation is an energetically costly process. Sediments accumulate detrital organic matter that can provide a source of carbon and nitrogen. Some archaeal lineages such as *Thorarchaeota* [73] and MBG-D (Marine Benthic Group D) [74] are well equipped to exploit detrital material. Similar to the related MBG-D archaea SCGC AB-539-N05 (ALXL00000000.1) and SCGC AB-539-C06 (AOSH00000000.1) [74], LFW-68_2 possess an expanded repertoire of peptidases. With 59.84 MEROPS matches per Mbp (compared to ~50/Mbp for the aforementioned MBG-D and 25-35/Mbp for most other archaea), LFW-68_2 is the archaeal genome with the highest density of peptidase-coding genes (Figure S4). The collection of peptidases in the genome of LFW-68_2 is especially expanded in C25.001 cysteine peptidases (15 copies), M09.002 metallopeptidases (17 copies), and M08.020 metallopeptidases (8 copies). A total of 26 extracellular peptidases provides LFW-68_2 with an exhaustive machinery for the degradation of detrital proteins.

The proteolytic capabilities of LFW-68_2 are even more exceptional than just a broad repertoire of proteases. We predicted an extracellular protease from the S12 family (PID 281199). The family S12, along with S11 and S13, constitute the SE clan which includes peptidases with specialised roles in the bacterial cell-wall metabolism, i.e. they act on D-amino acids [75]. A cytoplasmic M19 dipeptidase (PID 342444) was also predicted and is one of the few dipeptidases able to cleave dipeptides regardless of whether the C-terminal residue is a D- or L- isomer [76]. Lastly, we predicted a putative

operon (PID 515476-85, Figure 5B) containing homologues to the peptide-binding substrate binding protein-dependent ABC transport system (Dpp/Ddp/Opp/App/Gsi), used for the import of D,D-amino acids [77]. In contrast to other well-known examples where the operons are limited to a single gene per protein in the transport system, the organisation of the peptide-binding transport operon in LFW-68_2, i.e. DdpFDCBAFFCBA (Figure 5B), suggests that a duplication/rearrangement/recombination event occurred somewhere during its evolutionary history. The above three points suggest that LFW-68_2 might not have a role in the degradation of proteins but in the degradation of D-amino acid-containing organic matter.

The D-amino acid containing necromass is one of the most recalcitrant components, often found in bacterial cell walls, as well as enriched in aged sediments, as a product of amino acid racemisation [74, 78]. The presence of the three aforementioned components (i.e. extracellular protease S12, M19 dipeptidase and oligo-/di-peptide transport system) support the hypothesis that LFW-68_2 is a detritivorous Archaea able to feed on bacterial cell walls and recalcitrant biomass.

While all other MAGs have some kind of amino acid transporters (MFS and/or ABC), the transporter systems for oligopeptides are mainly limited to App (min 4 aa chains) in *Methanomicrobia*.

No evidence of sulfur dissimilatory pathways was found in any of the recovered genomes.

High heme cytochromes are common in *Methanoperedens* spp. genomes

As described above, field measurements revealed that nitrogen compounds were limited in the LFLS trenches and, while the concentration of dissolved Mn doubled during anoxic periods, the iron concentration was 200-400 times higher than Mn at any point in time [3]. Analysis of historical disposal records revealed that over 760 (intact and partially corroded) steel drums were deposited within the legacy trenches at LFLS, including in the vicinity of the collected samples [79]. The unabated redox oscillations which have occurred in the trenches over the last 60+ years and the resultant impact upon the steel drums, have likely contributed to the elevated concentrations of

soluble iron observed (~0.5 to 1 mM) [3]. Our field data indicates that during rainfall events when oxygen-laden rainwater enters the reducing trench water, soluble iron, i.e. Fe(II), is oxidised to insoluble Fe(III) oxides ferrihydrite and lepidocrocite [80]. However, these initially formed, meta-stable iron oxides are readily removed from the water column, via reductive dissolution to Fe(II), or transformed, via Fe(II)-catalysed remineralisation, to more crystalline Fe(III) oxides including goethite [81]. Being less susceptible to microbial reduction, it is plausible that crystalline Fe(III) minerals persist in reducing trench waters over periods between rainfall events, when Fe(III) oxides could be reformed.

Utilisation of metals (e.g. Fe, Mn, Cr, U) as electron acceptors requires the presence of multi-heme cytochromes. Although multi-heme cytochromes are not intrinsically exclusively employed for the reduction of heavy metals (e.g. decaheme DmsE for DMSO reduction), multi-heme cytochromes with a high number (≥10) of heme binding sites (HHC, high heme cytochromes) are generally exclusive of microorganisms implicated in the biological reduction of metals. In our case, the screening of the *Methanoperedens* spp. MAGs from LFLS revealed multiple HHCs in all genomes with a minimum of 6 copies for LFW-280_1_1 (68.5%C, Table S14). All reference *Methanoperedens* spp. genomes revealed similar abundances when considering their relative completeness, i.e., ~10 HHC per genome. Although metal reduction has only been confirmed for *M. ferrireducens* and *Methanoperedens* spp. BLZ1 [43, 65], our data suggests metal reduction to be a universal, or at least widespread, characteristic of *Methanoperedens* spp.

Previous studies indicate that *Methanoperedens* spp. BLZ1 and *M. ferrireducens* can both use ferrihydrite as electron acceptor for AOM [43, 65]. Reports showing the involvement of more crystalline forms of iron oxides in AOM are scarce, with the possibility that these organisms may drive a cryptic sulfate reduction cycle rather than a more direct Fe(III)-dependent AOM [82]. Nonetheless, it is not clear if *Methanoperedenaceae* archaea or other ANME would be responsible for these observations.

Questions remain as to the role of *Methanoperedens* spp. with regard to the ultimate fate of iron within the legacy trenches; an important consideration given the central role that iron plays in the mobilisation/retardation of key contaminants plutonium and americium [3, 83].

A self-sustaining DPANN?

The DPANN archaea *sensu stricto* (not including *Altiarchaeota*) are generally described as organisms with extremely reduced metabolic capacities and generally dependant on a host for the acquisition of essential metabolites such as vitamins, amino acids or even reducing equivalents [9].

During the examination of the metabolic reconstruction of the individual archaeal genomes, it was observed that some DPANN from the LFWA-III lineage, exemplified by LFW-121_3, possess nearly complete pathways for the synthesis of amino acids, purine and pyrimidine nucleotides, riboflavin and thiamine (Figure 2). The predictions derived from either Pathway Tools (via Prokka annotations) or KEGG were unusual given that when a DPANN genome lacks the biosynthetic capacity for an amino acid, all enzymes for that pathway are typically missing. However, several of the LFW-121_3 pathways had a limited number of gaps (Figure 2). These alleged ‘gaps’, thoroughly discussed in the SI, were often related to, e.g., unusual enzymes poorly annotated in databases, or steps known to be possible to be performed by bifunctional or promiscuous enzymes. Manual curation with predicted functions from other databases combined with literature searches was able to fill some of those gaps or, at least, provide reasonable candidate proteins that may carry on those functions.

Conclusion

While the Archaea inhabiting the LFLS trench water constitute a relatively small portion of its microbial community, they have important roles in the biogeochemical cycles: methanogenesis, anaerobic methane oxidation, Fe(III) reduction, and protein degradation. The diverse *Methanoperedens* spp., *Methanotrichaceae*, and *Thermoplasmata* are of special relevance due to their

426 roles in organic matter turnover (C and N), Fe cycling and dissimilatory pathways for nitrate and
427 maybe even use of (per)chlorate as electron acceptor.

428 The broad phylogenetic representation of DPANN organisms, which have attracted special attention
429 in the last few years either for their unusual characteristics or their controversial evolutionary history,
430 is another interesting feature of the LFLS archaeal community. The proposed LFWA-III lineage is
431 amongst the most interesting. Not only is this because they are unusually diverse or predicted to have
432 any special metabolic functions *per se*, but because they have a proper central metabolism. While
433 there is still some uncertainty about the capabilities of LFW-121_3 and closely related archaea for the
434 biosynthesis of nucleotides *de novo* and amino acids, it is to date the most promising candidate to be
435 the first *bona fide* free-living DPANN archaeon. This is duly reflected in the proposed name '*Ca*.
436 Gugararchaeum adminiculabundum`, where the specific epithet translates as "self-supporting" relating
437 to the possibility of not needing a symbiotic partner.

438 Acknowledgements

439 MRW acknowledges the support of the Australian Federal Government NCRIS scheme, via
440 Bioplatforms Australia, and the NSW State Government RAAP scheme.

441 Conflict of Interest

442 The authors declare that they have no conflict of interest.

443 Contribution statement

444 XVC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project
445 administration, Visualization, Writing – original draft, Writing – review & editing. ASK: Writing –
446 original draft, Writing – review & editing. MWB: Writing – review & editing. MRW: Resources,

Writing – review & editing. TEP: Resources, Funding acquisition, Project administration. TDW: Resources, Funding acquisition, Project administration, Supervision, Writing – review & editing.

References

1. Payne TE. Background report on the Little Forest Burial Ground legacy waste site. 2012. Institute for Environmental Research, Australian Nuclear Science and Technology Organisation, Lucas Heights, N.S.W. (AU).

2. Payne TE, Harrison JJ, Hughes CE, Johansen MP, Thiruvoth S, Wilsher KL, et al. Trench 'bathtubbing' and surface plutonium contamination at a legacy radioactive waste site. *Environ Sci Technol* 2013; **47**: 13284–13293.

3. Vázquez-Campos X, Kinsela AS, Bligh M, Harrison JJ, Payne TE, Waite TD. Response of microbial community function to fluctuating geochemical conditions within a legacy radioactive waste trench environment. *Appl Environ Microbiol* 2017; **83**: e00729.

4. Noël V, Boye K, Kukkadapu RK, Li Q, Bargar JR. Uranium storage mechanisms in wet-dry redox cycled sediments. *Water Res* 2019; **152**: 251–263.

5. Winkler P, Kaiser K, Thompson A, Kalbitz K, Fiedler S, Jahn R. Contrasting evolution of iron phase composition in soils exposed to redox fluctuations. *Geochim Cosmochim Acta* 2018; **235**: 89–102.

6. Bhattacharyya A, Campbell AN, Tfaily MM, Lin Y, Kukkadapu RK, Silver WL, et al. Redox fluctuations control the coupled cycling of iron and carbon in tropical forest soils. *Environ Sci Technol* 2018; **52**: 14129–14139.

7. Borch T, Kretzschmar R, Kappler A, Cappellen PV, Ginder-Vogel M, Voegelin A, et al. Biogeochemical Redox Processes and their Impact on Contaminant Dynamics. *Environ Sci Technol* 2010; **44**: 15–23.

8. Wilkins MJ, Livens FR, Vaughan DJ, Beadle I, Lloyd JR. The influence of microbial redox cycling on radionuclide mobility in the subsurface at a low-level radioactive waste storage site. *Geobiology* 2007; **5**: 293–301.

9. Dombrowski N, Lee J-H, Williams TA, Offre P, Spang A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett* 2019; **366**: fnz008.

10. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol* 2019; **4**: 234.

11. Song GC, Im H, Jung J, Lee S, Jung M-Y, Rhee S-K, et al. Plant growth-promoting archaea trigger induced systemic resistance in *Arabidopsis thaliana* against *Pectobacterium carotovorum* and *Pseudomonas syringae*. *Environ Microbiol* 2019; **21**: 940–948.

12. Reeburgh WS. Oceanic methane biogeochemistry. *Chem Rev* 2007; **107**: 486–513.

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

14. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 2016; **102**: 3–11.

15. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.

16. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015; **3**: e1319.

17. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013; **499**: 431–437.

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

19. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 2016; **44**: D286–D293.

20. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 2017; **34**: 2115–2122.

21. Yutin N, Puigbò P, Koonin EV, Wolf YI. Phylogenomics of prokaryotic ribosomal proteins. *PLOS ONE* 2012; **7**: e36972.

22. Yamada KD, Tomii K, Katoh K. Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 2016; **32**: 3246–3251.

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

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

25. Wang HC, Minh BQ, Susko S, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol* 2018; **67**: 216–235.

26. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 2015; **32**: 268–274.

27. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013; **30**: 1188–1195.

28. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; **22**: 1658–1659.

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

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

31. Dombrowski N, Seitz KW, Teske AP, Baker BJ. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* 2017; **5**: 106.

531 32. Eddy SR. Accelerated profile HMM searches. *PLOS Comput Biol* 2011; **7**: e1002195.

532 33. André I, Potocki-Véronèse G, Barbe S, Moulis C, Remaud-Siméon M. CAZyme discovery and
533 design for sweet dreams. *Curr Opin Chem Biol* 2014; **19**: 17–24.

534 34. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated
535 carbohydrate-active enzyme annotation. *Nucleic Acids Res* 2012; **40**: W445–W451.

536 35. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic
537 enzymes, their substrates and inhibitors. *Nucleic Acids Res* 2014; **42**: D503–D509.

538 36. Saier MH, Reddy VS, Tamang DG, Västermark Å. The Transporter Classification Database.
539 *Nucleic Acids Res* 2014; **42**: D251–D258.

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

543 38. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: Improved Detection and Functional
544 Classification of Transfer RNA Genes. *bioRxiv* 2019; 614032.

545 39. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.

546 40. Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, et al. Pathway
547 Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief*
548 *Bioinform* 2016; **17**: 877–890.

549 41. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for
550 functional characterization of genome and metagenome sequences. *J Mol Biol* 2016; **428**: 726–
551 731.

552 42. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource
553 for gene and protein annotation. *Nucleic Acids Res* 2016; **44**: D457–D462.

554 43. Cai C, Leu AO, Xie G-J, Guo J, Feng Y, Zhao J-X, et al. A methanotrophic archaeon couples
555 anaerobic oxidation of methane to Fe(III) reduction. *ISME J* 2018; **12**: 1929–1939.

556 44. Arshad A, Speth DR, de Graaf RM, Op den Camp HJM, Jetten MSM, Welte CU. A
557 metagenomics-based metabolic model of nitrate-dependent anaerobic oxidation of methane by
558 *Methanoperedens*-like Archaea. *Front Microbiol* 2015; **6**: 1423.

559 45. Vaksmaa A, Guerrero-Cruz S, van Alen TA, Cremers G, Ettwig KF, Lüke C, et al. Enrichment
560 of anaerobic nitrate-dependent methanotrophic '*Candidatus* Methanoperedens nitroreducens'
561 archaea from an Italian paddy field soil. *Appl Microbiol Biotechnol* 2017; **101**: 7075–7084.
562 46. van Dongen S, Abreu-Goodger C. Using MCL to Extract Clusters from Networks. In: van
563 Helden J, Toussaint A, Thieffry D (eds). *Bacterial Molecular Networks: Methods and*
564 *Protocols*. 2012. Springer New York, New York, NY, pp 281–295.
565 47. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in
566 diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2015;
567 **8**: 12–24.
568 48. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, et al.
569 Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017; **541**: 353–
570 358.
571 49. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. The growing tree of Archaea: new
572 perspectives on their diversity, evolution and ecology. *ISME J* 2017.
573 50. Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, et al. Lineages of
574 acidophilic Archaea revealed by community genomic analysis. *Science* 2006; **314**: 1933–1935.
575 51. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the
576 phylogeny and coding potential of microbial dark matter. *Nature* 2013; **499**: 431–437.
577 52. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A
578 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of
579 life. *Nat Biotechnol* 2018; **36**: 996–1004.
580 53. Mills HJ, Hodges C, Wilson K, MacDonald IR, Sobecky PA. Microbial diversity in sediments
581 associated with surface-breaching gas hydrate mounds in the Gulf of Mexico. *FEMS Microbiol*
582 *Ecol* 2003; **46**: 39–52.
583 54. Knittel K, Boetius A. Anaerobic oxidation of methane: progress with an unknown process. *Annu*
584 *Rev Microbiol* 2009; **63**: 311–334.

585 55. Jing H, Takagi J, Liu J, Lindgren S, Zhang R, Joachimiak A, et al. Archaeal surface layer
586 proteins contain β propeller, PKD, and β helix domains and are related to metazoan cell surface
587 proteins. *Structure* 2002; **10**: 1453–1464.
588 56. Najmudin S, Guerreiro CIPD, Carvalho AL, Prates JAM, Correia MAS, Alves VD, et al.
589 Xyloglucan is recognized by carbohydrate-binding modules that interact with β-glucan chains. *J*
590 *Biol Chem* 2006; **281**: 8815–8828.
591 57. Yeh AP, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, et al. Structure of
592 *Bacteroides thetaiotaomicron* BT2081 at 2.05 Å resolution: the first structural representative of
593 a new protein family that may play a role in carbohydrate metabolism. *Acta Crystallogr Sect F*
594 2010; **66**: 1287–1296.
595 58. Reddy CSK, Ghai R, Rashmi, Kalia VC. Polyhydroxyalkanoates: an overview. *Bioresour*
596 *Technol* 2003; **87**: 137–146.
597 59. Martínez-Gutiérrez CA, Latisnere-Barragán H, García-Maldonado JQ, López-Cortés A.
598 Screening of polyhydroxyalkanoate-producing bacteria and PhaC-encoding genes in two
599 hypersaline microbial mats from Guerrero Negro, Baja California Sur, Mexico. *PeerJ* 2018; **6**:
600 e4780.
601 60. Spang A, Poehlein A, Offre P, Zumbrägel S, Haider S, Rychlik N, et al. The genome of the
602 ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and
603 environmental adaptations. *Environ Microbiol* 2012; **14**: 3122–3145.
604 61. Koller M. Polyhydroxyalkanoate biosynthesis at the edge of water activity-Haloarchaea as
605 biopolyester factories. *Bioengineering* 2019; **6**: 34.
606 62. Haroon MF, Hu S, Shi Y, Imelfort M, Keller J, Hugenholtz P, et al. Anaerobic oxidation of
607 methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 2013; **500**: 567–570.
608 63. Cai C, Shi Y, Guo J, Tyson GW, Hu S, Yuan Z. Acetate production from anaerobic oxidation of
609 methane via intracellular storage compounds. *Environ Sci Technol* 2019; Just accepted.
610 64. Karthikeyan OP, Chidambarampadmavathy K, Cirés S, Heimann K. Review of sustainable
611 methane mitigation and biopolymer production. *Crit Rev Environ Sci Technol* 2015; **45**: 1579–
612 1610.

613 65. Ettwig KF, Zhu B, Speth D, Keltjens JT, Jetten MSM, Kartal B. Archaea catalyze iron-
614 dependent anaerobic oxidation of methane. *Proc Natl Acad Sci* 2016; **113**: 12792–12796.
615 66. Lu Y-Z, Fu L, Ding J, Ding Z-W, Li N, Zeng RJ. Cr(VI) reduction coupled with anaerobic
616 oxidation of methane in a laboratory reactor. *Water Res* 2016; **102**: 445–452.
617 67. Yoshimatsu K, Sakurai T, Fujiwara T. Purification and characterization of dissimilatory nitrate
618 reductase from a denitrifying halophilic archaeon, *Haloarcula marismortui*. *FEBS Lett* 2000;
619 **470**: 216–220.
620 68. Oren A, Elevi Bardavid R, Mana L. Perchlorate and halophilic prokaryotes: implications for
621 possible halophilic life on Mars. *Extremophiles* 2014; **18**: 75–80.
622 69. Brown GM, Gu B. The Chemistry of Perchlorate in the Environment. In: Gu B, Coates JD (eds).
623 *Perchlorate: Environmental Occurrence, Interactions and Treatment*. 2006. Springer US,
624 Boston, MA, pp 17–47.
625 70. Luo Y-H, Chen R, Wen L-L, Meng F, Zhang Y, Lai C-Y, et al. Complete perchlorate reduction
626 using methane as the sole electron donor and carbon source. *Environ Sci Technol* 2015; **49**:
627 2341–2349.
628 71. Xie T, Yang Q, Winkler MKH, Wang D, Zhong Y, An H, et al. Perchlorate bioreduction linked
629 to methane oxidation in a membrane biofilm reactor: Performance and microbial community
630 structure. *J Hazard Mater* 2018; **357**: 244–252.
631 72. Wu M, Luo J-H, Hu S, Yuan Z, Guo J. Perchlorate bio-reduction in a methane-based membrane
632 biofilm reactor in the presence and absence of oxygen. *Water Res* 2019; **157**: 572–578.
633 73. Seitz KW, Lazar CS, Hinrichs K-U, Teske AP, Baker BJ. Genomic reconstruction of a novel,
634 deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction.
635 *ISME J* 2016; **10**: 1696–1705.
636 74. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, et al. Predominant
637 archaea in marine sediments degrade detrital proteins. *Nature* 2013; **496**: 215–218.
638 75. Laskar A, Chatterjee S, Roy A, Kumar Dey S, Mandal C. Molecular modeling and structural
639 analysis of five SE clan (S12 family) serine proteases. *Asian J Biotechnol* 2011; **3**: 435–448.

640 76. Hooper NM. Membrane dipeptidase. In: Rawlings ND, Salvesen G (eds). *Handbook of*
641 *Proteolytic Enzymes*, 3rd ed. 2013. pp 1670–1673.
642 77. Maqbool A, Levdikov VM, Blagova EV, Hervé M, Horler RSP, Wilkinson AJ, et al.
643 Compensating stereochemical changes allow murein tripeptide to be accommodated in a
644 conventional peptide-binding protein. *J Biol Chem* 2011; **286**: 31512–31521.
645 78. Lomstein BA, Langerhuus AT, D'Hondt S, Jørgensen BB, Spivack AJ. Endospore abundance,
646 microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 2012; **484**:
647 101–104.
648 79. Payne TE, Kinsela AS, Rowling B, Hughes CE, Hankin S, Wilsher KL, et al. Application of an
649 Experimental Trench to Evaluate Hydrogeological Factors Impacting Radionuclide Migration at
650 a Legacy Waste Disposal Site. *J Environ Radioact* 2019; Submitted.
651 80. Kinsela AS, Jones AM, Bligh MW, Pham AN, Collins RN, Harrison JJ, et al. Influence of
652 dissolved silicate on rates of Fe(II) oxidation. *Environ Sci Technol* 2016; **50**: 11663–11671.
653 81. Boland DD, Collins RN, Miller CJ, Glover CJ, Waite TD. Effect of solution and solid-phase
654 conditions on the Fe(II)-accelerated transformation of ferrihydrite to lepidocrocite and goethite.
655 *Environ Sci Technol* 2014; **48**: 5477–5485.
656 82. Sivan O, Antler G, Turchyn AV, Marlow JJ, Orphan VJ. Iron oxides stimulate sulfate-driven
657 anaerobic methane oxidation in seeps. *Proc Natl Acad Sci* 2014; **111**: E4139–E4147.
658 83. Ikeda-Ohno A, Harrison JJ, Thiruvoth S, Wilsher K, Wong HKY, Johansen MP, et al. Solution
659 speciation of plutonium and americium at an Australian legacy radioactive waste disposal site.
660 *Environ Sci Technol* 2014; **48**: 10045–10053.
661

Figures and tables

Table 1. Summary of new candidate species proposed.

Proposed species name	Current name	Current lineage name (GTDB)	Proposed type
'*Ca*. Tiddalikarchaeum anstoniae'	Uncultured archaeon LFW-252_1 [a]	Nanoarchaeota; Nanoarchaeia; CG07-land	ERS2655302
'*Ca*. Wianamattarchaeum fermentum'	Uncultured archaeon LFW-144_1 [a]	Micrarchaeota; Micrarchaeia; UBA8480	ERS2655293
'*Ca*. Gugararchaeum adminiculabundum'	Uncultured archaeon LFW-121_3 [a]	Micrarchaeota; Micrarchaeia; UBA10214	ERS2655291
'*Ca*. Wayembeharchaeum dharawalense'	Uncultured archaeon LFW-283_2 [a]	Micrarchaeota; Micrarchaeia; UBA10214	ERS2655319
'*Ca*. Burarchaeum australiense'	Uncultured archaeon LFW-281_7 [a]	Micrarchaeota; Micrarchaeia; UBA10214	ERS2655318
'*Ca*. Anstonella stagnisolia'	Uncultured archaeon LFW-35 [a]	Micrarchaeota; Micrarchaeia; UBA10214	ERS2655287
'*Ca*. Methanoperedens vercellense'	'*Ca*. Methanoperedens nitroreducens' Vercelli	Methanoperedenaceae	GCA_900196725
'*Ca*. Methanoperedens batavicum'	'*Ca*. Methanoperedens nitroreducens' BLZ1 (or MPEBLZ)	Methanoperedenaceae	GCA_001317315

[a]: obtained in this study.

666 Table 2. General stats of the genomes reconstructed in this study.

667 MAGs in bold indicate named candidate species. Additional details can be found in Table S4.

668 Expanded completeness analysis are available in Table S5.

MAG	Lineage	Contigs	Size (Mbp)	tRNA[a]	GC (%)	rRNA genes[b]	C% / R%[c]
LFW-28	Altiarchaeota	151	2.41	23	49.9	Y/N/Y	91.4 / 4.9
LFW-252_2	Diapherotrites	64	1.05	21	40.6	Y/Y/Y	75.3 / 1.9
LFW-281_4	Diapherotrites	72	1.09	20	49.7	Y/Y/Y	64.2 / 4.9
LFW-252_1	LFWA-I	56	1.16	21	36.5	Y/Y/Y	75.3 / 1.9
LFW-144_1	LFWA-II	49	0.93	20	57.5	Y/N/Y	85.8 / 2.5
LFW-156_1	LFWA-II	109	0.93	18	32.5	N/N/Y	72.8 / 7.4
LFW-121_3	LFWA-III	76	1.47	21	49.8	Y/Y/Y	93.2 / 3.7
LFW-144_2_1	LFWA-III	219	0.77	16	57.5	N/N/N	61.7 / 6.8
LFW-242_1	LFWA-III	69	0.97	21	57.3	N/N/Y	77.8 / 8.6
LFW-281_1_2	LFWA-III	79	0.71	17	55.7	N/N/Y	50 / 2.5
LFW-281_3_2	LFWA-III	81	1.08	19	54.9	N/N/Y	75.3 / 9.3
LFW-281_5_4	LFWA-III	104	0.64	15	55.4	N/N/Y	54.9 / 4.3
LFW-281_6_1	LFWA-III	122	0.70	14	53.7	N/Y/N	52.5 / 6.2
LFW-281_7	LFWA-III	76	1.20	18	57.6	Y/N/Y	90.1 / 3.1
LFW-283_2	LFWA-III	63	1.27	20	40.0	Y/Y/Y	85.8 / 2.5
LFW-29	LFWA-III	68	1.30	21	62.2	N/N/Y	88.3 / 1.9
LFW-35	LFWA-III	68	1.33	21	50.3	Y/Y/Y	87 / 3.7
LFW-46	LFWA-IV	85	1.20	18	43.8	N/N/Y	83.3 / 4.9
LFW-125_1	Micrarchaeota	97	0.82	20	54.7	Y/Y/N	66.7 / 1.2
LFW-165_1	Pacearchaeota	67	1.34	19	30.7	N/N/Y	67.3 / 1.2
LFW-170_1	Pacearchaeota	42	0.84	19	33.8	Y/Y/Y	64.2 / 1.2
LFW-170_3	Pacearchaeota	31	0.59	20	35.3	N/N/Y	63 / 1.2
LFW-262_2	Pacearchaeota	45	0.62	20	31.5	N/N/Y	51.2 / 6.2
LFW-262_5	Pacearchaeota	44	0.58	19	31.9	N/N/Y	61.7 / 5.6
LFW-273_1	Pacearchaeota	27	0.56	20	33.4	N/N/Y	61.1 / 0.6
LFW-252_3	Woesearchaeota	101	1.04	20	42.3	N/N/Y	69.1 / 3.1
LFW-24	Methanoperedenaceae	169	2.92	21	40.3	N/N/Y	93.8 / 5.6
LFW-280_1_1	Methanoperedenaceae	207	1.69	20	44.4	N/N/N	68.5 / 4.3
LFW-280_2_2	Methanoperedenaceae	298	2.20	15	43.6	N/N/Y	74.7 / 4.9
LFW-280_3_1	Methanoperedenaceae	179	3.24	21	43.8	N/N/Y	95.7 / 9.9
LFW-280_3_2	Methanoperedenaceae	147	2.79	19	43.8	N/N/Y	90.7 / 9.9
LFW-280_4	Methanoperedenaceae	182	2.60	21	42.9	N/N/Y	92 / 9.9
LFW-151_1	Methanotrichaceae	401	1.75	15	53.2	N/N/N	68.5 / 8
LFW-151_2	Methanotrichaceae	390	2.28	16	50.6	N/N/Y	83.3 / 3.7
LFW-83_1	Methanotrichaceae	580	2.63	19	47.7	Y/Y/Y	62.4 / 8.6
LFW-68_2	Thermoplasmata	152	2.06	22	38.2	N/N/Y	85.2 / 6.2
LFW-283_4_5	Thaumarchaeota	152	1.15	18	37.0	N/N/N	50.6 / 3.1

669

[a]: number of different tRNA found (21 includes starting iMet). Additional tRNA may include Sec, Ile2, or undetermined types.

[b]: found rRNA genes, 16S/23S/5S.

[c]: completeness (C%) and redundancy (R%) estimations based on Anvi'o. Additional estimations based on CheckM can be found in the SI.

Figure 1. Phylogenomic analysis of the archaeal bins found at LFLS.

Concatenated protein tree constructed with 44 universal and Archaea-specific ribosomal proteins from 230 reference genomes and 36 original MAGs from this work. Tree was constructed with IQ-TREE under the PMSF+LG+C60+F+I+G. Lineages with representatives at LFLS are shown coloured and uncollapsed, with labels in blue. Inner ring indicates the major archaeal lineages. Shortened branches are shown at 50% of their length. Black circles indicate ultrafast bootstrap support values >90%.

Figure 2. Model of a LFWA-III archaeon cell.

Figure is based on the metabolic modelling of LFW-121_3 ('*Ca.* Gugararchaeum adminiculabundum'). Abbreviations: 2-OG, 2-oxoglutarate; 2-OIV, 2-oxoisovalerate; AICAR, 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide; AIR, 5-amino-1-(5-phospho-β-D-ribosyl)imidazole; ARP, 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione; CAIR, 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxylate; F1,6P, β-D-Fructose 1,6-bisphosphate; F6P, β-D-Fructose-6P; FAD, Flavin adenine dinucleotide; FMN, Flavin mononucleotide; G3P, Glyceraldehyde-3P; G6P, Glucose-6P; GP, Glycerone-P; HMP-PP, 4-amino-5-hydroxymethyl-2-methylpyrimidine diphosphate; N5-CAIR, 5-(carboxyamino)imidazole ribonucleotide; NAD+, β-nicotinamide adenine dinucleotide; NADP+, β-nicotinamide adenine dinucleotide phosphate; NMN, β-nicotinamide D-ribonucleotide; OAA, Oxaloacetate; OAHS, O-acetyl-homoserine; OSHS, O-succinyl-homoserine; PEP, Phosphoenolpyruvate; PRPP, 5-phospho-α-D-ribose 1-diphosphate; R5P, D-ribose 5-phosphate; SAH, S-adenosyl-L-homocysteine; SAM, S-adenosyl-L-methionine; ThDP, Thiamine diphosphate; ThMP, Thiamine monophosphate; THZ, 4-methyl-5-(β-hydroxyethyl)thiazolium; THZ-P, 4-methyl-5-(β-hydroxyethyl)thiazolium phosphate.

Figure 3. Pangenomic analysis of lineage LFWA-III (Gugararchaeaceae).

Pangenome generates with 8 genomes from the LFWA lineage, with a total of 10,238 protein-coding genes grouped in 7,584 gene clusters. Gene clusters generated with an MCL inflation value of 2.0.

Figure 4. Pangenomic analysis of genus *Methanoperedens.*

Pangenome generates with 10 *Methanoperedens* spp. genomes, accounting for a total of 31,974 protein-coding genes grouped in 10,525 gene clusters. Gene clusters generated with an MCL inflation value of 6.0.

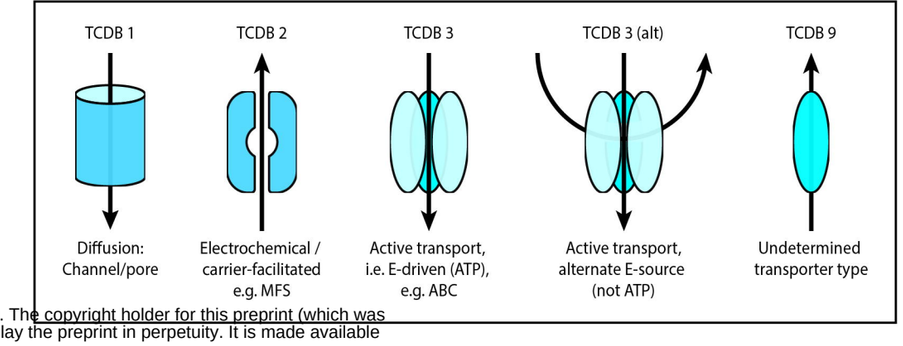
Figure 5. Schematic diagram of operons of interest.

A) Biosynthetic gene cluster for polyhydroxyalkanoates in *Methanoperedens* spp. from LFLS, one of the clusters in LFW-280_3_1 as an example. B) D,D-amino acid import/cleavage operon of the archaeon LFW-68_2.

Figure 6. Phylogenetic analysis of molybdopterin oxidoreductase proteins.

Tree was generated with 268 reference sequences and 8 query sequences from '*Ca.* Methanoperedens' spp. NarG(-like) proteins. Tree was built under model LG+R10 model and 10,000 ultrafast bootstrap iterations. Branches with ultrafast bootstrap support values <50% are collapsed. Black circles indicate a branch support ≥90%.

"Superphylum"
Altiarchaeota
Asgardeota
DPANN
Euryarchaeota
TACK
LFWA-IV
LFWA-III
LFWA-I
LFWA-II
Micrarchaeota
Diapherotrites
Altiarchaeota (SM1)
Aenigmarchaeota (DSEG)
Nanohaloarchaeota
Huberarchaeota
Nanoarchaeota
Parvarchaeota
Woesearchaeota
Pacearchaeota
Asgardaeota
Korarchaeota
Bathyarchaeota
Geothermarchaeota (THSCG/THSC)
Aigarchaeota (HWCG I)
Thaumarchaeota
Verstraetearchaeota
Geoarchaeota (YNPFFA/NAG1)
Crenarchaeota (SAGMEG-1)
Hadesarchaea (ArcI/NSA2)
Thermococci
Methanofastidiosa (ArcI/NSA2)
Theionarchaea (Z7ME43)
Hydrothermarchaea (MBGE)
Methanopyri
Methanobacteria
Methanococci
Izemarchaea/Thermoprofundales
Thalassoarchaea (MG-II)
Pontarchaea
Methanomassiliicoccales (RCC/RC-III)
Archaeoglobi
Syntropharchaeales (GoM-Arch87)
Methanotrichaceae
Methanosarcinaceae
Methanoperedenaceae (ANME-2d)
Halobacteria
Methanomicrobiales
Methanocellales
Tree scale: 0.1 —



Extracellular proteases

S08A

I39

Extracellular glycosidases

amino acids
Cl^-
Zn^{2+}
Mn^{2+}
K^+
sugars
SO_3^{2-}
TCA
H^+
H^+/K^+
K^+
Na^+
Na^+
pyruvate

NH_4^+
Na^+/K^+
Mg^{2+}
mechanosensitive channel
H_2O
H_2O
Hg^{2+}

$GDP+P_i$
Fe^{2+}
GTP
Pb^{2+}

$O_2^{\cdot-}$ → H_2O_2 → O_2

G6P ↔ F6P ↔ F1,6P ↔ GP ↔ G3P (GK/PFK)

3P-D-glycerate → Serine (SerB, SerC) → Glycine
Serine → L-cysteine (CysE, CysK) → L-alanine (CsdB) → Pyruvate (ALAT)

PEP ↔ OAA (PEPCK) → L-aspartate (AspC) ← L-asparagine
L-aspartate → L-aspartate 4-semialdehyde → L-homoserine (HSK) → L-threonine
L-homoserine → L-homocysteine (OSHS, OAHS; MetY, MetB) ← L-methionine → SAM → SAH

Pyruvate → Acetyl-CoA → Citrate → *cis*-aconitate → Isocitrate → 2-OG → Succinyl-CoA
2-OG → 2-oxoadipate (AksADEF) ↔ L-2-aminoadipate (LysX, LysZ, LysY, LysJ, LysK) → L-lysine

Pyruvate → L-isoleucine → 2-OB
Pyruvate → L-valine → 2-OIV → L-leucine

L-glutamate ↔ L-glutamine
L-glutamate → Proline (ArgJ, ArgB, ArgC, LysJ, ArgJ)
L-glutamate → Fumarate ← Arginine

Glycerone-P → PRPP → NAD^+ → $NADP^+$ → NMN

Prephenate → L-phenylalanine (HisC), L-tyrosine
Chorismate → Indole (HisA) → Ser → L-tryptophan

Pentose phosphate pathway
R5P → Pyridoxal-5P (Gln → Glu)

L-Glutamine → Asp

PRPP → Gly, Gln, Glu, Gln → AIR (PurE-I) → N^5-CAIR → CAIR → Fumarate → AICAR → IMP → AMP → ADP → Adenine; dAMP → dADP ↔ dATP
IMP ← Asp
GMP, dGTP, AK, GDP → dGDP, GTP
PRPP → HisB → L-Histidine

UMP, CMP, dCMP → dUMP → dTMP
UDP, CDP → dCDP → dUDP → dTDP
NH_4^+
UTP, CTP (Gln, Glu) → dCTP → dUTP → dTTP

Gly → THZ-P (?) ← NAD, THZ → HMP-PP → ThMP → ThDP → Thiamine triphosphate

GTP → ARP → Riboflavin → FMN → FAD
Ru5P

[NiFe] 3b hydrogenase → H_2

V-type ATPase

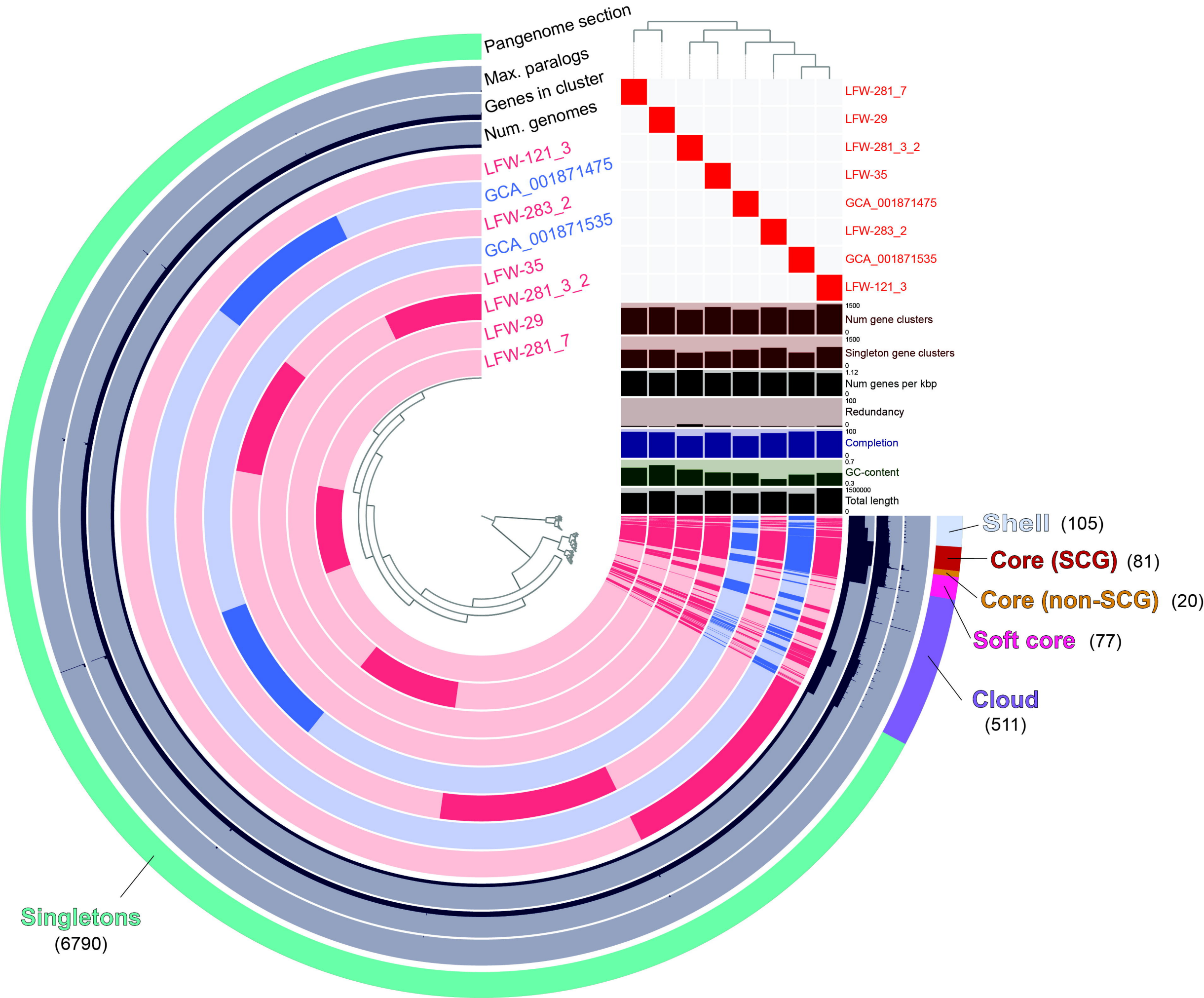
PP_i Na^+ P_i

cobalamin
Ca^{2+}
sugars
Mg^{2+}
branched amino acids
taurine

Type IV pili

S-layer

- Predicted/non-conflicting prediction.
- Suggested/non-canonical.
- Absent but found in other LFWA-III genomes.
- Non-specific reaction without defined candidate.
- Unknown enzyme in Archaea.
- Absent in all LFWA-III. Known to exist in (some) Archaea.


Pangenome section
Max. paralogs
Genes in cluster
Num. genomes
LFW-121_3
GCA_001871475
LFW-283_2
GCA_001871535
LFW-35
LFW-281_3_2
LFW-29
LFW-281_7
LFW-281_7
LFW-29
LFW-281_3_2
LFW-35
GCA_001871475
LFW-283_2
GCA_001871535
LFW-121_3
1500
Num gene clusters
0
1500
Singleton gene clusters
0
1.12
Num genes per kbp
0
100
Redundancy
0
100
Completion
0
0.7
GC-content
0.3
1500000
Total length
0
Shell (105)
Core (SCG) (81)
Core (non-SCG) (20)
Soft core (77)
Cloud
(511)
Singletons
(6790)

PhaE (PF09712)

LFLS NarG

Shell (1032)

PhaC (TIGR01835)

PhaE (PF09712)

Soft core
(926)

Core (non-SCG)
(234)

Core (SCG)
(248)

Cloud
(2090)

Unusual NarG

PhaE (PF09712)

PhaE (PF09712)

Singletons
(5995)

NarG
(TIGR01580)

PhaE
(PF09712)

PhaE (PF09712)

Pangenome section

Max. paralogs

Genes in cluster

Num. genomes

LFW-280_4

LFW-280_1_1

LFW-280_3_2

LFW-280_2_2

M. nitroreducens

M. vercellense

LFW-280_3_1

M. batavicum

LFW-24

M. ferrireducens

M. ferrireducens

LFW-24

M. batavicum (BLZ1)

LFW-280_3_1

M. vercellense (Vercelli)

M. nitroreducens

LFW-280_2_2

LFW-280_3_2

LFW-280_1_1

LFW-280_4

4600
Singleton gene clusters
0

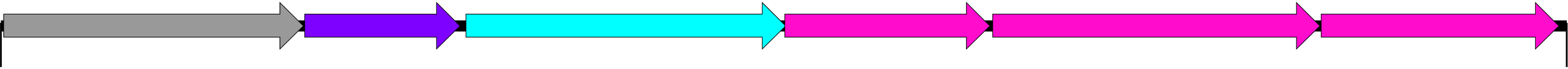
4600
Num gene clusters
0

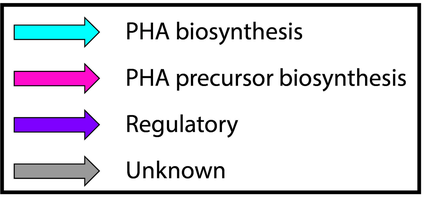
100
Redundancy
0

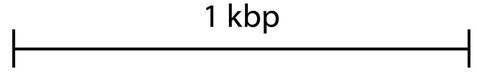
100
Completion
0

0.5
GC-content
0

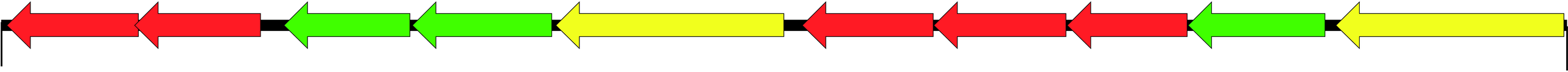
A.

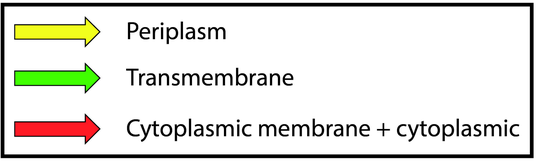


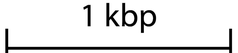


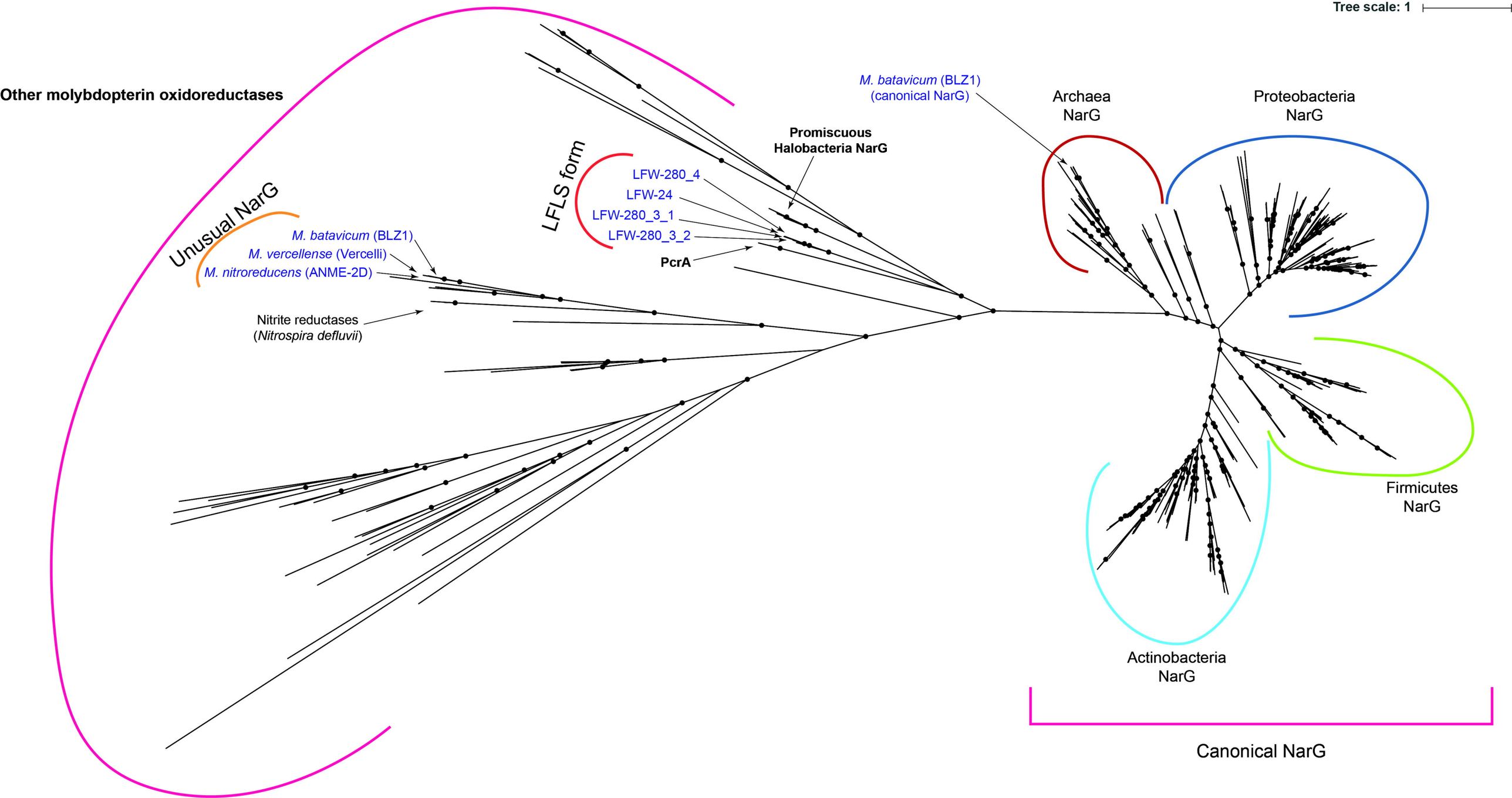


B.








Tree scale: 1
Other molybdopterin oxidoreductases
Unusual NarG
M. batavicum (BLZ1)
M. vercellense (Vercelli)
M. nitroreducens (ANME-2D)
Nitrite reductases
(Nitrospira defluvii)
LFLS form
LFW-280_4
LFW-24
LFW-280_3_1
LFW-280_3_2
PcrA
Promiscuous
Halobacteria NarG
M. batavicum (BLZ1)
(canonical NarG)
Archaea
NarG
Proteobacteria
NarG
Firmicutes
NarG
Actinobacteria
NarG
Canonical NarG