



# Genomic Insights Into the *Archaea* Inhabiting an Australian Radioactive Legacy Site

Xabier Vázquez-Campos<sup>1\*</sup>, Andrew S. Kinsela<sup>2</sup>, Mark W. Bligh<sup>2</sup>, Timothy E. Payne<sup>3</sup>, Marc R. Wilkins<sup>1</sup> and T. David Waite<sup>2\*</sup>

<sup>1</sup> NSW Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW, Australia, <sup>2</sup> UNSW Water Research Centre, School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW, Australia, <sup>3</sup> Environmental Research Theme, Australian Nuclear Science and Technology Organisation, Kirrawee DC, NSW, Australia

## OPEN ACCESS

### Edited by:

Olga V. Golyshina,  
Bangor University, United Kingdom

### Reviewed by:

Nikolai Ravin,  
Research Center of Biotechnology  
(RAS), Russia  
Edward R. B. Moore,  
University of Gothenburg, Sweden

### \*Correspondence:

Xabier Vázquez-Campos  
xvazquezc@gmail.com  
T. David Waite  
d.waite@unsw.edu.au

### Specialty section:

This article was submitted to  
Biology of Archaea,  
a section of the journal  
Frontiers in Microbiology

Received: 29 June 2021

Accepted: 21 September 2021

Published: 18 October 2021

### Citation:

Vázquez-Campos X, Kinsela AS,  
Bligh MW, Payne TE, Wilkins MR and  
Waite TD (2021) Genomic Insights  
Into the *Archaea* Inhabiting an  
Australian Radioactive Legacy Site.  
Front. Microbiol. 12:732575.  
doi: 10.3389/fmicb.2021.732575

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 in a waste trench during a rainfall event have been previously investigated revealing a broad abundance of candidate and potentially undescribed taxa in this iron-rich, radionuclide-contaminated environment. Applying genome-based metagenomic methods, we recovered 37 refined archaeal MAGs, mainly from undescribed DPANN *Archaea* lineages without standing in nomenclature and ‘*Candidatus* Methanoperedenaceae’ (ANME-2D). Within the undescribed DPANN, the newly proposed orders ‘*Ca.* Gugararchaeales’, ‘*Ca.* Burarchaeales’ and ‘*Ca.* Anstonellales’, constitute distinct lineages with a more comprehensive central metabolism and anabolic capabilities within the ‘*Ca.* Micrarchaeota’ phylum compared to most other DPANN. The analysis of new and extant ‘*Ca.* Methanoperedens spp.’ MAGs suggests metal ions as the ancestral electron acceptors during the anaerobic oxidation of methane while the respiration of nitrate/nitrite via molybdopterin oxidoreductases would have been a secondary acquisition. The presence of genes for the biosynthesis of polyhydroxyalkanoates in most ‘*Ca.* Methanoperedens’ also appears to be a widespread characteristic of the genus for carbon accumulation. This work expands our knowledge about the roles of the *Archaea* at the LFLS, especially, DPANN *Archaea* and ‘*Ca.* Methanoperedens’, while exploring their diversity, uniqueness, potential role in elemental cycling, and evolutionary history.

**Keywords:** *Archaea*, DPANN, metagenomics, phylogenomics, ‘*Candidatus* Methanoperedens’, ANME-2d, radionuclides, pangenomics

## INTRODUCTION

The Little Forest Legacy Site (LFLS) is a low-level radioactive waste disposal site in Australia that was 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, and covered with a shallow (1 m layer) of local soil, as

was common practice at the time (Payne, 2012). The location of the LFLS site was partly chosen because of its clay and shale rich stratigraphy, intended to limit the migration of radionuclides from site. An unintended consequence of this, is that surface water infiltrates into the more porous trench waste form at a greater rate when compared to the surrounding geology, and during intense and prolonged rainfall events the contaminated water levels in trenches fill up, akin to a 'bathtub' (Payne et al., 2013, 2020). Aside from providing a mechanism enabling the export of contaminants, occurring when trench waters (infrequently) overflow out of the 'bathtub' to the surrounding environment (Payne et al., 2013), the periodic influx of oxic waters into these natural reducing zones results in transitory microbial population shifts associated with pronounced elemental cycling (Vázquez-Campos et al., 2017; Kinsela et al., 2021). Our previous research showed that the archaeal community in the LFLS waste trenches, while constituting a minor portion of the whole microbial community, included a number of potentially interesting members, either in terms of phylogeny and/or functionality (Vázquez-Campos et al., 2017).

The domain *Archaea* constitutes by far the most understudied domain of life. *Archaea* have been traditionally regarded as biological oddities and marginal contributors to the global geochemical elemental cycles, with most living in extreme environments (Dombrowski et al., 2019). However, the continuous development of cultivation-independent techniques in recent decades has completely changed this perception. For example, "*Thaumarchaeota*" are now known to be key components in the global nitrogen cycle (Kitzinger et al., 2019) and can even induce systemic resistance to pathogens in plants (Song et al., 2019). Some *Euryarchaeota*, which are capable of the anaerobic oxidation of methane (AOM), consume up to 80–90% of the methane produced in environments such as rice fields and ocean floors before this highly deleterious greenhouse gas can be released into the atmosphere (Reeburgh, 2007).

The DPANN *Archaea*, so-named from its constituent lineages when it was defined ('*Candidatus* Diapherotrites', '*Ca.* Parvarchaeota', '*Ca.* Aenigmarchaeota', '*Ca.* Nanohaloarchaeota', and '*Nanoarchaeota*') (Rinke et al., 2013), is an archaeal superphylum generally characterised by the small dimensions of its members (excluding '*Ca.* Altiarchaeota'). Their diminutiveness not only relates to their reduced physical size, many of which are capable of passing through a standard 0.2  $\mu\text{m}$  filter (Luef et al., 2015), but also due to their small genome sizes, i.e., many are below 1 Mbp and with an estimated maxima of  $\sim 1.5$  Mbp. Many of the DPANN with extremely reduced genomes lack biosynthetic pathways for amino acids, nucleotides, vitamins, and cofactors, which aligns with a host-dependent lifestyle (Dombrowski et al., 2019). Others, towards the higher end of the genome size scale, typically encode for some or most of the aforementioned pathways and are regarded as free-living. Irrespective of size, only a limited number of organisms from this lineage have been studied in detail (Waters et al., 2003; Baker et al., 2006; Castelle et al., 2015; Youssef et al., 2015; Golyshina et al., 2017).

Another *Archaea* taxon that has attracted substantial attention during the last few years are the anaerobic methane oxidisers ANME-2d (Mills et al., 2003) or AAA (AOM-associated *Archaea*) (Knittel and Boetius, 2009), now referred to as '*Ca.* Methanoperedens' (Haroon et al., 2013). These archaea have been purported to couple the oxidation of methane, at its source, with the reduction of nitrate (Haroon et al., 2013), nitrite (Arshad et al., 2015), Fe(III) (Ettwig et al., 2016; Cai et al., 2018), Mn(IV) (Ettwig et al., 2016; Leu et al., 2020), or even Cr(VI) (Lu et al., 2016), constituting an important environmental pathway in mitigating the release of methane into the atmosphere at a global scale.

Here we describe the results from the analysis of 37 new archaeal metagenome-assembled genomes (MAGs) with different degrees of completeness, derived from samples collected in a legacy radioactive waste trench at the LFLS during a redox cycling event. We propose 23 new *Candidatus* taxa (Murray and Schleifer, 1994; Murray and Stackebrandt, 1995) based on data analysed in this manuscript, as well as four missing taxa definitions. We present evidence of four undescribed DPANN lineages and six non-conspicuous '*Ca.* Methanoperedens spp.' genomes, while exploring their uniqueness and potential role in elemental cycling.

## MATERIALS AND METHODS

In order to better understand the archaeal contributions in LFLS groundwaters and their influence upon site biogeochemistry, raw sequencing reads from ENA project PRJEB14718 (Vázquez-Campos et al., 2017) were reanalysed using genome-based metagenomic methodologies. Briefly, samples were collected in triplicate over a period of 47 days at four time-points, i.e., 0 (SAMEA4074280–2), 4 (SAMEA4074283–5), 21 (SAMEA4074286–8), and 47 (SAMEA4074289–91) days, after an intense rainfall event that filled the trench. Thorough chemical and radiochemical analyses were conducted on the samples in order to fully characterise the geochemistry (Vázquez-Campos et al., 2017). Over this time period, the redox conditions transitioned from slightly oxic to increasingly reducing/anoxic, reflected in both the chemical analyses and the community profile, with a clear increase in obligate anaerobes on days 21 and 47 (Vázquez-Campos et al., 2017).

DNA sequencing was performed in a NextSeq 500 (Illumina) with a  $2 \times 150$ -bp high-output run.

## Recovery and Assessment of Community Genomes

QC and trimming of raw sequencing reads were performed with Trim Galore!<sup>1</sup>. Trimmed reads from each replicate were merged and treated as four individual samples from here on. Co-assembly of the trimmed reads was performed with MEGAHIT v1.0.2 (Li et al., 2015, 2016) (--kmin-1pass --k-min 27 --k-max 127 --k-step 10).

<sup>1</sup>[http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)

Contigs > 2.5 kbp were binned using CONCOCT (Alneberg et al., 2014) while limiting the number of clusters to 400 to avoid ‘fragmentation errors,’ within *anvi'o* v2.3 (Eren et al., 2015) following the metagenomics protocol as previously described (Eren et al., 2015). Bins of archaeal origin were manually refined in *anvi'o*. Completeness and redundancy were estimated with *anvi'o* single copy gene (SCG) collection, based on Rinke et al. (2013), as well as with CheckM v1.0.13 (Parks et al., 2015). In the case of CheckM, analyses were performed by both the standard parameters, with automatic detection of the marker gene set, and by forcing the universal *Archaea* marker gene collection.

## Phylogenomics and Phylogenetics

Prodigal v2.6.3 (-p meta) was used to predict proteins from the archaeal MAGs and 242 archaeal reference genomes (Supplementary Table 1). Predicted proteins were assigned to archaeal Clusters of Orthologous Groups (arCOGs) using the arNOG database of eggNOG v4.5.1 (Huerta-Cepas et al., 2016) using eggNOG-mapper v0.99.2 (Huerta-Cepas et al., 2017).

Markers for the archaeal phylogenomic tree (rp44) were selected from the universal and archaeal ribosomal protein list from Yutin et al. (2012) with the following criteria: (i) present in a minimum of 80% of the collection of reference genomes and (ii) have a limited level of duplications (standard deviation of copy number in genomes with given arCOG < 1.25). A total of 44 ribosomal proteins were selected (Supplementary Table 2).

Ribosomal proteins were individually aligned with MAFFT-L-INS-i v7.305b (Yamada et al., 2016) and concatenated. Concatenated alignment was trimmed with BMGE v1.2 (-m BLOSUM30 -g 0.5) (Criscuolo and Gribaldo, 2010) based on Lazar et al. (2017).

The concatenated archaeal protein tree was constructed using the posterior mean site frequency (PMSF) model (Wang et al., 2018) in IQ-TREE v1.5.5 (Nguyen et al., 2015), a faster and less resource-intensive approximation to the C10–C60 mixture models (Le et al., 2008), which are also a variant of Phylobayes' CAT model. Briefly, a base tree was inferred using a simpler substitution model such as LG. This tree was used as the base to infer the final tree under the optimum model suggested by IQ-TREE, LG+C60+F+I+G, with 1000 ultrafast bootstrap replicates (Minh et al., 2013).

Taxonomic identities of the recovered MAGs were initially also evaluated with GTDB-Tk v0.3.0 (Chaumeil et al., 2020) using GTDB r89 (Parks et al., 2020). The program was run with both *classify\_wf* and then with *de\_novo\_wf* in order to attain a better placement for the MAGs with a limited number of closely related references (Supplementary Table 3).

## Phylogenetic Placement of New Taxa in the DPANN Tree

In addition to the rp44 and GTDB-tk analyses, phylogenetic placement of the LFLS DPANN was investigated with additional methods based on a dataset containing all the representative genomes of DPANN from GTDB r202 (312, not including the LFLS MAGs already in r202), the 25 LFLS DPANN

MAGs, and 11 non co-generic representatives from ‘*Ca. Altiarchaeota*’ as outgroup.

Three methods were used for the DPANN-specific phylogenomic analysis: (i) concatenated proteins; (ii) partitioned analysis; and (iii) multispecies coalescence (MSC) model with ASTRAL (Zhang et al., 2018). For these, the protein sequences used for tree inference were derived from the 122 SCGs used for the archaeal GTDB (ar122) with certain caveats, as 29 proteins were present in <50% of the full dataset (348 genomes) and therefore automatically eliminated from the concatenated and partitioned trees. In all cases, each marker was aligned with MAFFT-L-INS-i v7.407 (Yamada et al., 2016).

- i) Concatenated protein (dpann-cat): a concatenated multiple sequence alignment (MSA) was generated as for the rp44 dataset using MAFFT and BMGE. The tree was generated using the PSMF model. Briefly, an initial tree was generated with FastTree v2.1.10 (-lg -gamma) (Price et al., 2010). The best model was searched with IQ-TREE v2.1.2 for all supported mixture models plus iterations of LG with C10–C60 models (Le et al., 2008). The final tree was created with model LG+C60+F+I+G and 1000 ultrafast bootstrap replicates (Hoang et al., 2018).
- ii) Partitioned analysis (dpann-part): the MSA derived from the concatenated protein tree was used as input where each partition was defined as per each protein. A total of 93 partitions with a length of 49-733 amino acids were included in the analysis. Substitution models were selected individually per partition. Support values in IQ-TREE were calculated based on 1000 ultrafast bootstrap replicates with nearest neighbour interchange optimisation and --sampling GENESITE.
- iii) Supertree with ASTRAL (dpann-astral): sequences for each marker were aligned with MAFFT-L-INS-i and trimmed with BMGE v1.2 with options defined above. Individual ML trees were generated for each marker with IQ-TREE v2.1.2 and the suggested model for each, with 1000 ultrafast bootstrap replicates with nearest neighbour interchange optimisation (Hoang et al., 2018). The resulting ML trees were used as input for ASTRAL-III v5.7.7 (Zhang et al., 2018) with multi-locus bootstrapping (-b -r 1000).

Prior iterations of these analyses performed with the DPANN representatives from GTDB r89 (dpann\_r89, 150 reference genomes and 10 ‘*Ca. Altiarchaeota*’ outgroup sequences) and dpann\_r89 plus 4 representatives from ‘*Ca. Fermentimicroarchaeales*’ (dpann\_r89F) were also included as part of the discussion on the placement of some of the LFWA lineages.

## Other Phylogenies

Phylogeny of NarG-like proteins was created using all proteins (416 sequences) as reference, from related orthologous groups from the EggNOG database, i.e., arCOG01497 (19), ENOG4102T1R (5), ENOG4102TDC (5), ENOG4105CRU (351), and ENOG4108JIG (36). Protein sequences were clustered with CD-HIT v4.6 (-c 0.9 -s 0.9 -g 1) (Li and Godzik, 2006) resulting in 276 sequences. Sequences were

aligned with MAFFT-L-INS-i v7.305b (Yamada et al., 2016). Alignment was trimmed with BMGE v1.2 (-g 0.9 -h 1 -m BLOSUM30) (Criscuolo and Gribaldo, 2010) and the tree built with IQ-TREE v1.5.5 under the recommended model (LG+R10) and 10,000 ultrafast bootstrap replicates (Nguyen et al., 2015).

Phylogeny of LysJ/ArgD and related proteins was created using all proteins (500 sequences) as reference from related orthologous groups from Swissprot. Protein sequences were clustered with CD-HIT v4.6 (with default parameters) (Li and Godzik, 2006) resulting in 256 representative sequences. Sequences were aligned with MAFFT-L-INS-i v7.305b (Yamada et al., 2016). Alignment was trimmed with BMGE v1.2 (with defaults) (Criscuolo and Gribaldo, 2010) and the tree built with IQ-TREE v1.5.5 under the recommended model (LG+F+I+G4) and 1,000 ultrafast bootstrap replicates (Nguyen et al., 2015).

## Functional Annotation

In addition to arCOGs, predicted proteomes were profiled with InterProScan v5.25-64.0 (Jones et al., 2014). InterProScan was run with options --disable-precalf --pathways --goterms --iprlookup and using the following databases: CDD v3.16 (Marchler-Bauer et al., 2017), Gene3D v4.1.0 (Lees et al., 2014), HAMAP v201701.18 (Pedruzzi et al., 2015), PANTHER v11.1 (Mi et al., 2013), Pfam v31.0 (Finn et al., 2014), PIRSF v3.02 (Wu et al., 2004), PRINTS v42.0 (Attwood et al., 1994), ProDom v2006.1 (Servant et al., 2002), ProSitePatterns and ProSiteProfiles v20.132 (Sigrist et al., 2013), SFLD v2 (Akiva et al., 2014), SMART v7.1 (Letunic et al., 2015), SUPERFAMILY v1.75 (Gough et al., 2001), and TIGRFAM v15.0 (Haft et al., 2013).

Key biogeochemical enzymes were identified based on signature hmm profiles in the InterProScan output or by custom hmm profiles not integrated in databases (Anantharaman et al., 2016; Dombrowski et al., 2017). Custom hmm profiles were searched with HMMER v3.1b2 (-E 1e-20) (Eddy, 2011). Proteins with single HMMER matches below previously established thresholds (Anantharaman et al., 2016; Dombrowski et al., 2017) were searched against Swissprot and checked against the InterProScan results to reduce the incidence of false negatives.

High-heme cytochromes ( $\geq 10$  binding sites per protein) were predicted using a modified motif\_search.py script<sup>2</sup> that uses CX(1,4)CH as well as the more canonical CXXCH heme-binding motif.

Carbohydrate-active enzymes were searched locally with hmmscan (HMMER v3.1b2) (Eddy, 2011) and the --domtblout output using a CAZy-based (André et al., 2014) hmm library v5 from dbCAN (Yin et al., 2012). Output was parsed as recommended with the hmmscan-parser.sh script.

Peptidases and proteases were annotated by similarity search of the predicted proteins against the MEROPS database v11.0 (blastp -max\_hsps 1 -evalue 1e-4 -seg yes -soft\_masking true -lcase\_masking -max\_target\_seqs 1) (Rawlings et al., 2014).

Transporters were annotated by similarity search against the TCDB blastp -max\_hsps 1 -evalue 1e-4 -seg yes -soft\_masking true -qcov\_hsp\_perc 50 -lcase\_masking -max\_target\_seqs 1 (database version downloaded on July 12, 2018) (Saier et al., 2014).

The subcellular location of the proteins was predicted with PSORTb v3.0.6 (Yu et al., 2010).

Prediction of tRNA and rRNA genes was performed with tRNAscan-SE 2.0 (Chan et al., 2019) and Barrnap v0.9<sup>3</sup> respectively. Total tRNAs are reported based on the 20 default aminoacyl-tRNAs plus the initiator tRNA (iMet).

## Genome Annotation

Metagenome assembled genomes were annotated with Prokka v1.12 (Seemann, 2014) with the --rfam option for the annotation of ncRNA genes. Annotations were imported into PathwayTools v22.0 (Karp et al., 2016) for modelling. Predicted proteins from each MAG were also processed with KofamScan v1.1.0 for KEGG annotation (Kanehisa et al., 2016; Aramaki et al., 2020).

## Multivariate Analysis of DPANN Genomes

DPANN MAGs from LFLS were compared with extant publicly available genomes based on functional and compositional characteristics. In addition to the genomes in the rp44 dataset, all (165) DPANN published genomes in IMG were downloaded (November 15, 2019). Genomes from rp44 and IMG were compared with MASH v2.2.2 (Ondov et al., 2016) to remove redundant genomes, resulting in 133 additional DPANN genomes to be added to the dataset (**Supplementary Table 4**).

Functional data included in the analysis included COG categories, PSORTb predictions, TCDB, MEROPS and CAZy main categories as percentage of the total proteins predicted in each genome. Compositional data included the proportion of each amino acid in the predicted proteome, GC content and coding density. Before analysis, all reference genomes with %C < 70% (based on CheckM Archaea) were removed.

Multivariate analysis was performed on zero-centred scaled data via Sparse Principal Component Analysis (sPCA) (Shen and Huang, 2008) as implemented in the MixOmics R package v6.13.21 (Rohart et al., 2017) and limiting the number of explanatory variables per component to five.

## Completeness of DPANN Metagenome-Assembled Genomes

Due to the limited completeness values obtained for DPANN genomes – even those in the literature reported to be assembled in single contigs (**Supplementary Table 5**), four different archaeal or prokaryotic SCG collections were examined to assess each of the constituent markers with regard to their suitability for being included as markers for reporting completeness of DPANN MAGs. In addition to the already calculated completeness based on Rinke (default in anvio until v5.5 – anvio Rinke, 162 SCGs) (Eren et al., 2021) and CheckM Archaea (149 SCGs)

<sup>2</sup><https://github.com/alexherns/biotite-scripts>

<sup>3</sup><https://github.com/tseemann/barrnap>

(Parks et al., 2015), two additional SCG libraries were utilised: anvio Archaea76 (76 SCG - current archaeal SCG in anvio) (Lee, 2019), and CheckM prokaryote (56 SCGs, forced prokaryote/root SCG collections from CheckM) (Parks et al., 2015).

An initial set of DPANN genomes based on the one used for the multivariate analysis were filtered at 50%C (193 genomes) and 75%C (101 genomes) based on completeness values obtained by CheckM Archaea. All individual markers in all SCG collections were evaluated for their prevalence in each set of genomes. Redundant markers across SCG collections were also compared in order to filter potential universal SCGs that might be represented by best/failed HMM profiles. Initial marker filtering required a prevalence of  $\geq 75\%$  across all DPANN genomes, and a prevalence within classic DPANN lineages of  $> 50\%$  to account for the limited number of highly complete genomes.

## Pangenomic Analysis

Pangenomic analysis of the '*Ca. Methanoperedens* spp.' genomes, as well as the six nearly complete reference '*Ca. Methanoperedens* spp.' was performed with anvio v7-dev (Eren et al., 2015, 2021) following the standard pangenomics workflow<sup>4</sup>. The reference genomes included in all comparisons and pangenome analysis were: '*Ca. Methanoperedens nitroreducens*' ANME-2D (type genome, JMIY00000000.1) (Haroon et al., 2013), '*Ca. Methanoperedens ferrireducens*' (PQAS00000000.1) (Cai et al., 2018), '*Ca. M. nitroreducens*' BLZ1 (LKCM00000000.1) (Arshad et al., 2015), and '*Ca. M. nitroreducens*' Vercelli (GCA\_900196725.1) (Vaksmas et al., 2017), '*Ca. M. manganicus*' (GCA\_012026835.1) (Leu et al., 2020), '*Ca. M. manganireducens*' (GCA\_012026795.1) (Leu et al., 2020). Protein-coding

<sup>4</sup><http://merenlab.org/2016/11/08/pangenomics-v2>

genes were clustered with an MCL inflation value of 6 (van Dongen and Abreu-Goodger, 2012).

Average Nucleotide Identity (ANI) and Average Amino acid Identity (AAI) were calculated with pyani v0.2.7 (Pritchard et al., 2015) and CompareM v0.0.23<sup>5</sup> respectively.

## Proposed Taxa

Details on etymology and nomenclature can be found in the Taxonomic Appendix. A summary of the proposed taxa can be found in **Table 1**. Additional details about the proposed species can be found in **Supplementary Table 6**.

Minimal quality criteria for denominating new candidate species was based on Chuvochina et al. (2019). Relaxed parameters were allowed for selected genomes when phylogenetic/taxonomic novelty was deemed of special relevance: completeness  $\geq 90\%$ , redundancy  $< 10\%$ ,  $\geq 18$  tRNA, and 16S rRNA gene sequence  $\geq 1000$  bp. MAGs passing those criteria, different to extant named species (ANI  $< 95\%$ ) and with no described higher taxonomic levels, were considered to describe new taxa.

Relative evolutionary divergence (RED) (Parks et al., 2018) values were generated with phylorank v0.1.10<sup>6</sup> for the dpann-cat tree based on the taxonomy levels from GTDB r202. Obtained RED values were used to circumscribe the limits of new and extant lineages from genus to class level. For lower taxonomic levels, AAI and ANI calculations were also used (Konstantinidis et al., 2017). AAI was used to evaluate delimitations between families, genera, and, to a lesser extent, species. ANI was only utilised for genus and species levels.

<sup>5</sup><https://github.com/dparks1134/CompareM>

<sup>6</sup><https://github.com/dparks1134/PhyloRank>

**TABLE 1** | Summary of new taxa proposed.

Phylum	Class	Order	Family	Genus	Species	Type
"Nanoarchaeota"	"Nanoarchaeia" class. nov. <sup>a</sup>	' <i>Candidatus</i> Tiddalikarchaeales' ord. nov.	' <i>Candidatus</i> Tiddalikarchaeaceae' fam. nov.	' <i>Candidatus</i> Tiddalikarchaeum' gen. nov.	' <i>Candidatus</i> Tiddalikarchaeum anstoanum' sp. nov.	LFW-252_1 <sup>b</sup>
' <i>Candidatus</i> Micrarchaeota'	' <i>Candidatus</i> Micrarchaeia' class. nov. <sup>a</sup>	' <i>Candidatus</i> Micrarchaeales' ord. nov. <sup>a</sup>	' <i>Candidatus</i> Micrarchaeaceae' fam. nov. <sup>a</sup>	' <i>Candidatus</i> Micrarchaeum'	' <i>Candidatus</i> Micrarchaeum acidiphilum'	ARMAN-2
		' <i>Candidatus</i> Norongarragalinales' ord. nov.	' <i>Candidatus</i> Norongarragalineae' fam. nov.	' <i>Candidatus</i> Norongarragalina' gen. nov.	' <i>Candidatus</i> Norongarragalina meridionalis' sp. nov.	LFW-144_1 <sup>b</sup>
		' <i>Candidatus</i> Gugararchaeales' ord. nov.	' <i>Candidatus</i> Gugararchaeaceae' fam. nov.	' <i>Candidatus</i> Gugararchaeum' gen. nov.	' <i>Candidatus</i> Gugararchaeum adminiculabundum' sp. nov.	LFW-121_3 <sup>b</sup>
		' <i>Candidatus</i> Burarchaeales' ord. nov.	' <i>Candidatus</i> Burarchaeaceae' fam. nov.	' <i>Candidatus</i> Burarchaeum' gen. nov.	' <i>Candidatus</i> Burarchaeum australiense' sp. nov.	LFW-281_7 <sup>b</sup>
		' <i>Candidatus</i> Anstonellales' ord. nov.	' <i>Candidatus</i> Anstonellaceae' fam. nov.	' <i>Candidatus</i> Anstonella' gen. nov.	' <i>Candidatus</i> Anstonella stagnisolia' sp. nov.	LFW-35 <sup>b</sup>
			' <i>Candidatus</i> Bilamarchaeaceae' fam. nov.	' <i>Candidatus</i> Bilamarchaeum' gen. nov.	' <i>Candidatus</i> Bilamarchaeum dharawalense' sp. nov.	LFW-283_2 <sup>b</sup>

Descriptions of the taxa in this table are included in the Taxonomic Appendix with additional details on the proposed type materials in **Supplementary Table 6**.

<sup>a</sup>new proposal to fill gaps in taxonomic ranks.

<sup>b</sup>obtained in this study.

## RESULTS

### Binning Results and Archaeal Community

The sequencing output totalled 236,581,129 paired-end reads and 71.1 Gbp of data across the whole dataset. The *de novo* co-assembly of sequence reads from the LFLS trench subsurface water samples (Vázquez-Campos et al., 2017) generated 187,416 contigs with a total length of 1.32 Gbp (based on 2.5 kbp cutoff).

Binning with CONCOCT generated 290 initial bins. A total of 21 bins with clear archaeal identity or with ambiguous identity (similar completeness scores for bacterial and archaeal SCG profiles) were further refined with *anvi'o*, producing 37 curated archaeal MAGs with completeness  $\geq 50\%$  (%C from here on; 22 of  $\geq 90\%$ C) and redundancy  $\leq 10\%$  (%R from here on) (**Table 2** and **Supplementary Table 8** for detailed completeness assessment).

Initial phylogenomic reconstruction was based on the 16 ribosomal proteins by Hug et al. (2016) with the exception of

**TABLE 2** | General characteristics of the genomes reconstructed in this study. MAGs in bold indicate named candidate species (see **Table 1**). Additional details can be found in **Supplementary Table 7**. Expanded completeness analysis is available in **Supplementary Table 8**.

MAG	Lineage	Contigs	Size (Mbp)	tRNA <sup>a</sup>	GC (%)	rRNA genes <sup>b</sup>	C%/R% <sup>c</sup>
LFW-28	Altiaarchaeota	151	2.41	23	49.9	Y/N/Y	92.6/0 <sup>e</sup>
LFW-252_2	Diapherotrites	64	1.05	21	40.6	Y/Y/Y	91.4/3.2
LFW-281_4	Diapherotrites	72	1.09	20	49.7	Y/Y/Y	71/5.4
LFW-144_2_1	Fermentimicrarchaeales	219	0.77	16	57.5	N/N/N	52.7/0
<b>LFW-252_1</b>	LFWA-I	56	1.16	21	36.5	Y/Y/Y	95.7/1.1
<b>LFW-144_1</b>	LFWA-II	49	0.93	20	57.5	Y/N/Y	93.5/1.1
LFW-156_1	LFWA-II	109	0.93	18	32.5	N/N/Y	90.3/8.6
<b>LFW-121_3</b>	LFWA-IIIa	76	1.47	21	49.8	Y/Y/Y	96.8/2.2
<b>LFW-281_7</b>	LFWA-IIIb	76	1.20	18	57.6	Y/N/Y	96.8/5.4
LFW-281_1_2	LFWA-IIIc	79	0.71	17	55.7	N/N/Y	50/2.5 <sup>e</sup>
<b>LFW-283_2</b>	LFWA-IIIc	63	1.27	20	40.0	Y/Y/Y	95.7/0
LFW-281_3_2	LFWA-IIIc	81	1.08	19	54.9	N/N/Y	78.5/6.5
LFW-281_5_4	LFWA-IIIc	104	0.64	15	55.4	N/N/Y	63.4/9.7
LFW-281_6_1	LFWA-IIIc	122	0.70	14	53.7	N/Y/N	59.1/9.7
LFW-29	LFWA-IIIc	68	1.30	21	62.2	N/N/Y	97.8/0
<b>LFW-35</b>	LFWA-IIIc	68	1.33	21	50.3	Y/Y/Y	97.8/2.2
LFW-242_1	LFWA-III related <sup>d</sup>	69	0.97	21	57.3	N/N/Y	93.5/6.5
LFW-46	LFWA-IV	85	1.20	18	43.8	N/N/Y	92.5/3.2
LFW-125_1	Micrarchaeales	97	0.82	20	54.7	Y/Y/N	84.9/0
LFW-165_1	Pacearchaeota	67	1.34	19	30.7	N/N/Y	94.6/0
LFW-170_1	Pacearchaeota	42	0.84	19	33.8	Y/Y/Y	91.4/0
LFW-170_3	Pacearchaeota	31	0.59	20	35.3	N/N/Y	92.5/2.2
LFW-262_2	Pacearchaeota	45	0.62	20	31.5	N/N/Y	74.2/8.6
LFW-262_5	Pacearchaeota	44	0.58	19	31.9	N/N/Y	88.2/8.6
LFW-273_1	Pacearchaeota	27	0.56	20	33.4	N/N/Y	93.5/1.1
LFW-252_3	Woearchaeota	101	1.04	20	42.3	N/N/Y	90.3/2.2
LFW-24	Methanoperedenaceae	169	2.92	21	40.3	N/N/Y	97.4/1.3
LFW-280_1_1	Methanoperedenaceae	207	1.69	20	44.4	N/N/N	56.9/0.7
LFW-280_2_2	Methanoperedenaceae	298	2.20	15	43.6	N/N/Y	73.3/4.6
LFW-280_3_1	Methanoperedenaceae	179	3.24	21	43.8	N/N/Y	99.4/1.3
LFW-280_3_2	Methanoperedenaceae	147	2.79	19	43.8	N/N/Y	96.1/2
LFW-280_4	Methanoperedenaceae	182	2.60	21	42.9	N/N/Y	98.3/0.8
LFW-151_1	Methanotrichaceae	401	1.75	15	53.2	N/N/N	63.2/4
LFW-151_2	Methanotrichaceae	390	2.28	16	50.6	N/N/Y	81.5/3.4
LFW-83_1	Methanotrichaceae	580	2.63	19	47.7	Y/Y/Y	63.1/2.6
LFW-68_2	Thermoplasmata	152	2.06	22	38.2	N/N/Y	91.4/5.4
LFW-283_4_5	Thaumarchaeota	152	1.15	18	37.0	N/N/N	56.2/7.8

<sup>a</sup>number of different tRNA found (21 includes starting *iMet*). Additional tRNA may include *Sec*, *Ile2*, or undetermined types.

<sup>b</sup>found rRNA genes, 16S/23S/5S.

<sup>c</sup>completeness (C%) and redundancy (R%) estimations. Values for DPANN and Thermoplasmata MAGs derived from the DPANN93 marker collection from this manuscript. Estimations for other MAGs obtained with CheckM lineage\_wf unless otherwise stated. Additional C%/R% estimations can be found in the **Supplementary Information**.

<sup>d</sup>likely to constitute a member of an undescribed order sister to all the LFWA-III lineages.

<sup>e</sup>based on Rinke et al. (2013).

rpL16 due to frequent misannotation (Laura Hug, pers. comm.) (data not shown) using the PhyloSift HMM profiles (Darling et al., 2014). The resulting trees showed relatively poorly resolved basal branches in addition to the absence of rpL22 genes from virtually all DPANN *Archaea*. We found this to also be the case with other known HMM profile-based databases such as TIGRFAM (Haft et al., 2013). As an example, the rpL14p in DPANN *Archaea* is often overlooked by the TIGR03673 even when the protein is present as per the more suitable model defined by arCOG04095. Subsequently, we used the more up to date arNOG and increased the number of phylogenetic markers to 44 universal and archaeal ribosomal proteins.

Phylogeny based on the 44 concatenated ribosomal proteins (rp44, **Figure 1**) showed that most metagenome assembled genomes (MAGs) from LFLS belonged to diverse DPANN lineages and *Methanomicrobia* ('*Ca. Methanoperedenaceae*' and *Methanotrichaceae*) (**Table 2**). Phylogenomic analysis based on the rp44 (**Figure 1**) showed a general tree topology largely consistent with current studies, e.g., '*Ca. Asgardarchaeota*' as sister lineage to TACK and '*Ca. Altiarchaeota*' as sister to DPANN (Zaremba-Niedzwiedzka et al., 2017; Dombrowski et al., 2019). 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 (Adam et al., 2017). The position of the MAGs in the archaeal phylogeny suggests four undescribed lineages (**Figure 1**) within DPANN, denoted LFWA-I to -IV. These lineages correspond to representatives of high-level taxa (order or above) that do not have either a current proposed name or have not been explored in detail in the literature. As a means of honouring the Australian Aboriginal community, and very especially, the traditional owners of the land where LFLS is located, many of the nomenclatural novelties proposed are based on terms from Aboriginal languages related to the site (mainly Dharawal), see Taxonomic Appendix. A summary of the nomenclatural proposals derived from this work are summarised in **Table 1**.

During the drafting of this manuscript, a genome-based taxonomy was developed for *Bacteria* and *Archaea* (GTDB) (Parks et al., 2018). For all proposed lineages, we have included the possible correspondence with the GTDB taxonomy (r202) wherever necessary using the GTDB nomenclature with prefixes, e.g., "p\_\_" for phylum, to avoid confusion with more "typically" defined lineages with the same names. It should be noted that due to the delay between the release of the pre-print and MAGs data, and the publication of this manuscript, 27 of the 37 MAGs from this work were included in the GTDB r202 as representatives.

## The Archaeal Community in Detail

Based on bin coverage information, the archaeal community was dominated by DPANN, esp. '*Ca. Pacearchaeota*' and LFWA-III (see below), and *Euryarchaeota*, esp. '*Ca. Methanoperedenaceae*,' with maximum relative abundance values of 79.7% at day 4, and 42.8% at day 47, respectively (**Supplementary Table 9**). Our previous analysis indicated that DPANN constituted a maximum of 55.8% of the archaeal community (Vázquez-Campos et al., 2017). This discrepancy could be an artefact derived from the different copy numbers of rRNA gene clusters, often limited to

one in DPANN and TACK, compared to other *Archaea* with more typically sized genomes containing multiple copies of the operon.

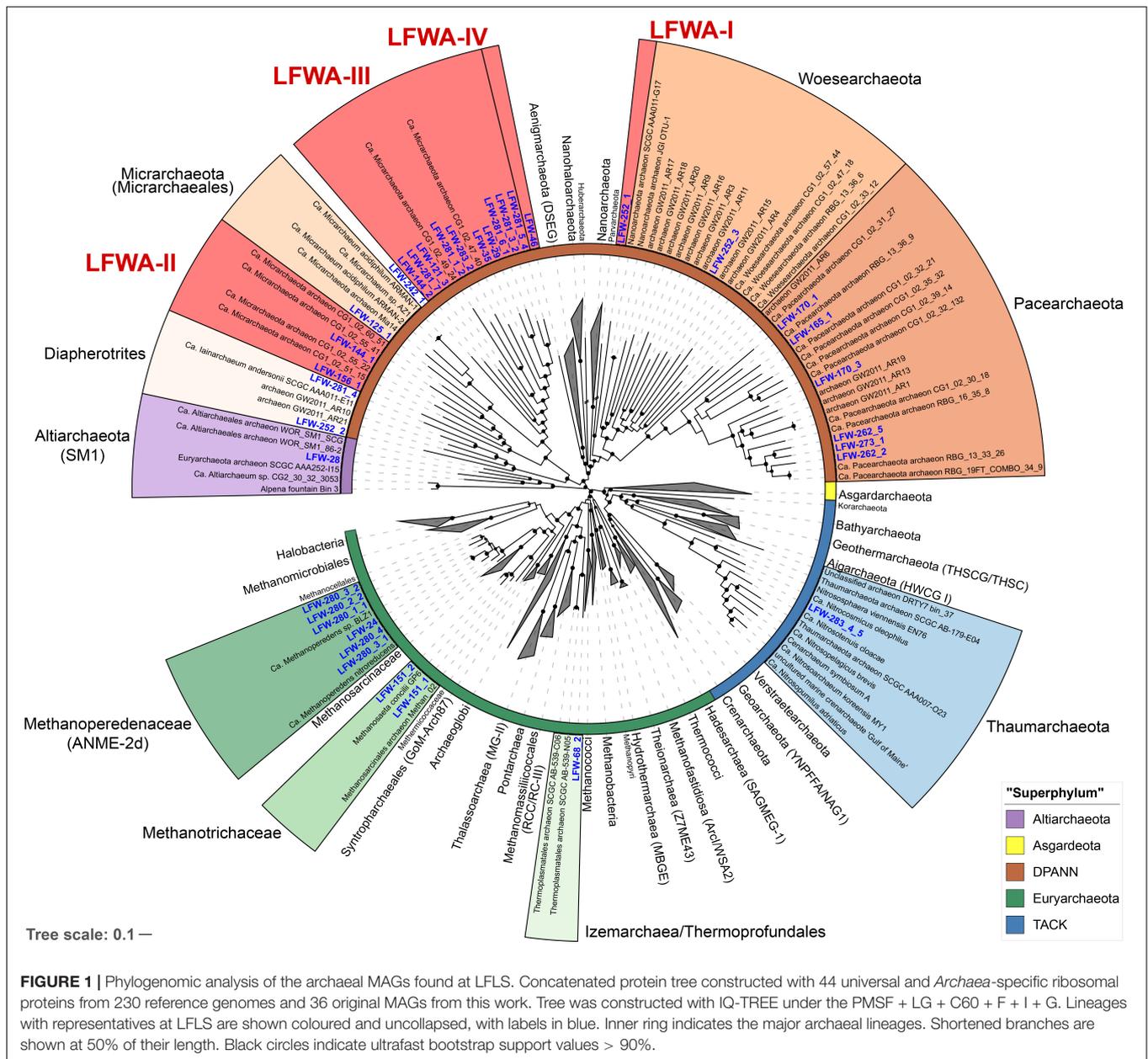
Based on the functional annotation of key proteins of biogeochemical relevance (**Figure 2** and **Supplementary Table 10**), the archaeal MAGs recovered from LFLS do not appear to play a major role in the sulfur cycle (lacking dissimilatory pathways, **Supplementary Table 10**). Regarding the nitrogen cycle, the presence of nitrogenases in some of the '*Ca. Methanoperedenens* spp.' MAGs (**Figure 2** and **Supplementary Table 10**) suggest that they may play a larger role in the fixation of N<sub>2</sub> than in the dissimilatory reduction of nitrate, particularly given the negligible nitrate concentrations measured (Vázquez-Campos et al., 2017). Another *Archaea* likely to be heavily involved in the nitrogen cycle, but through the degradation of proteins and D-amino acids, is the *Thermoplasmata* LFW-68\_2, based on the unusually high number of proteases, and specific proteases and transporters encoded in its genome (**Figure 2**, **Supplementary Figure 1**, and **Supplementary Table 11**; see **Supplementary Information** for details). The D-amino acids are one of the most recalcitrant components of necromass and are often found in bacterial cell walls as well as enriched in aged sediments as a result of amino acid racemisation (Lomstein et al., 2012; Lloyd et al., 2013). Thus, *Thermoplasmata* LFW-68\_2 may have a role in the remobilisation and remineralisation of organic matter in sediments.

Due to their special interest and relevance, the content below is focused on the new DPANN and the '*Ca. Methanoperedenens* spp.' from LFLS. Aspects related to other MAGs and other general observations regarding the whole archaeal community are contained within the **Supplementary Materials**.

## DPANN in Little Forest Legacy Site LFWA-I: Order '*Ca. Tiddalikarchaeales*'

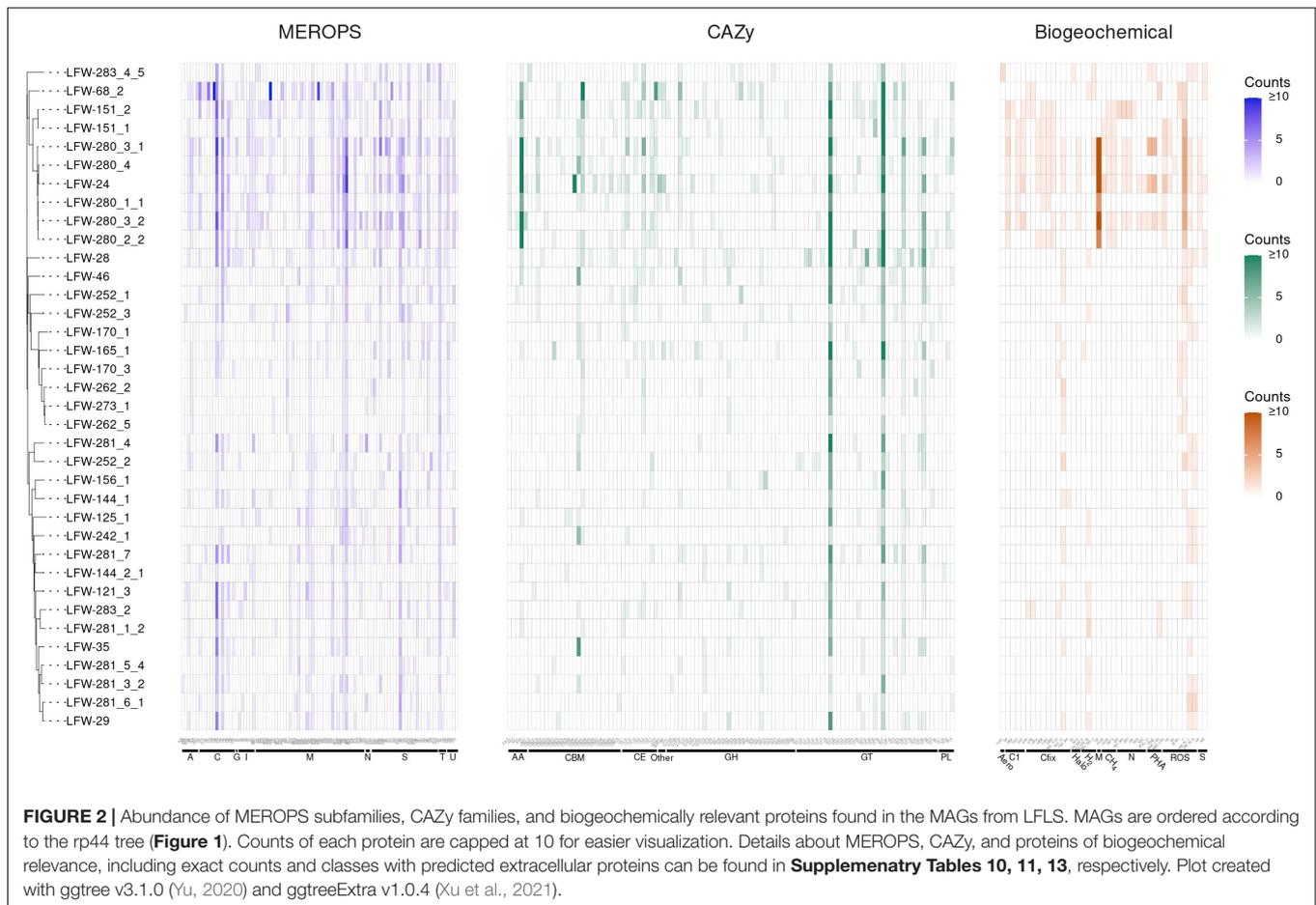
The LFWA-I lineage is a sister lineage to '*Ca. Parvarchaeota*' (ARMAN-4 and -5) (Baker et al., 2006; Rinke et al., 2013) in rp44 and contains a unique MAG, LFW-252\_1 ('*Ca. Tiddalikarchaeum anstoanum*') (**Figure 1** and **Supplementary Figure 2**). LFWA-I corresponds to the order and family o\_\_CG07-land f\_\_CG07-land within the c\_\_*Nanoarchaeia* as per GTDB. As with many other DPANN from the PANN branch (such as cluster 2 in Dombrowski et al., 2020) %C estimates with existing *Archaea* SCG collections are severely underestimated with barely 79.0%C reported by CheckM (best). The %C combined with the assembly size sets a complete genome size >1.5 Mbp, well above the expected genome size for a "*Nanoarchaeota*", and similar to that for many PANN MAGs (**Supplementary Table 8**). A dedicated SCG collection was built from extant HMM profiles by removing SCGs normally absent from all DPANN and/or from specific sub-lineages. This first attempt of a tailored SCG library for assessing the completeness of DPANN genomes (**Supplementary Table 12**), provides more sensible results not only for LFW-252\_1 (95.7%C/1.1%R) but also for DPANN assemblies in general, including some well-studied references (**Supplementary Table 5**).

The LFW-252\_1 lacks the biosynthetic capacity for almost all amino acids, and for the *de novo* enzymes biosynthesising



purine and pyrimidine, despite having a normal enzymatic representation for interconversion of purine nucleotides, and interconversion of pyrimidine nucleotides (Supplementary Figure 3). With the additional lack of a proper pentose biosynthesis pathway and only containing the last enzyme required for the biosynthesis of 5-phospho- $\alpha$ -D-ribose 1-diphosphate (PRPP), LFW-252\_1 is very likely to depend on a host, maybe aerobic or microaerophilic, although this is only based on the relative abundance patterns. Reactive oxygen species detoxification is mediated by superoxide reductase (SOR, neelaredoxin/desulfoferredoxin; TIGR00332) and rubredoxin. LFW-252\_1 is a subsurface decomposer of complex carbohydrates based on the CAZymes associated to the degradation of lignocellulosic material, including

carbohydrate esterases (CE1, CE12, and CE14),  $\beta$ -mannase (GH113), and even one vanillyl-alcohol oxidase (AA4) (Figure 2 and Supplementary Table 13). Its genome also harbours three different sialidases (GH33), at least two extracellular, making it one of the few non-pathogenic prokaryotes containing this enzyme (Buelow et al., 2016). In addition to the sialic acids as sources of C and N, LFW-252\_1 also contains two extracellular serine peptidases (S08A, subtilisin-like), indicating a possible active role in the degradation of proteins in the environment (Supplementary Table 11). Energy production is based on the Embden–Meyerhof–Parnas (EMP) pathway, or the ferredoxin-mediated oxidative decarboxylation of pyruvate (PorABC), 2-oxoglutarate (KorAB), and, likely, other oxoacids (OforAB), i.e., fermentation of carbohydrates



and amino acids. The EMP pathway, however, lacks the enzyme(s) required for the conversion of glyceraldehyde-3P (G3P) to 3-phosphoglycerate (3PG). Not a single subunit of the archaeal-type ATPase (A-ATPase), the electron transport chain (ETC) or the tricarboxylic acid (TCA) cycle were found. Pilus type IV (FlaIJK) but not archaeum is present.

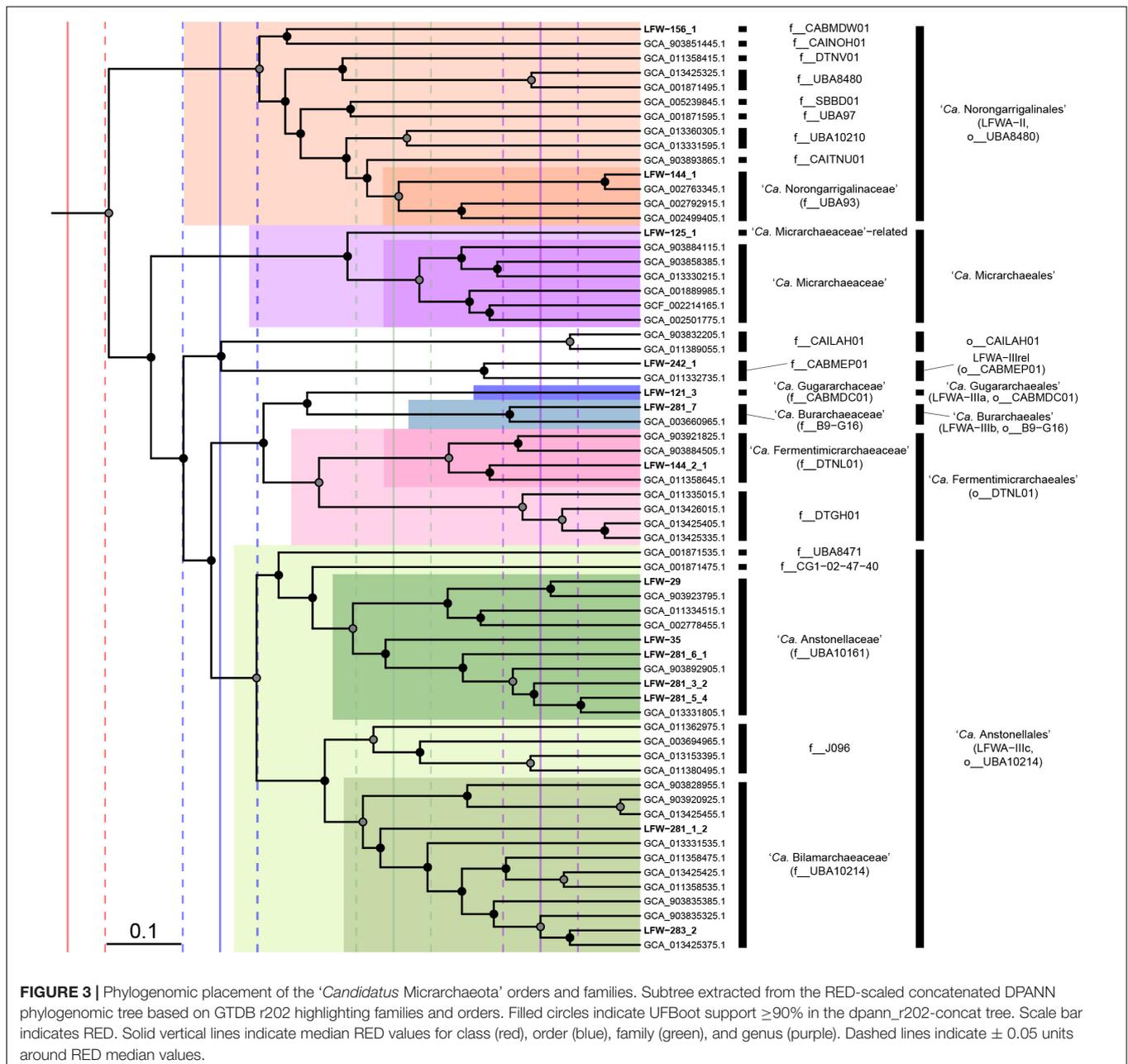
### LFWA-II: Order ‘*Ca. Norongarragalinales*’

The lineage LFWA-II is represented in LFLS by MAGs LFW-144\_1, and possibly LFW-156\_1, and it is consistently placed as a sister clade to all other ‘*Ca. Micrarchaea*’ (see taxonomic novelties; **Figure 3** and **Supplementary Figure 2**). Both MAGs have a similar size but different completeness estimates and GC content: 93.5%C/1.1%R and 57.5% GC for LFW-144\_1, and 90.3%C/8.6%R and 32.5% GC for LFW-156\_1. Each of the LFWA-II MAGs belong to different families within o\_\_UBA8480 (c\_\_*Micrarchaea*). ‘*Ca. Norongarragalina meridionalis*’ LFW-144\_1, is the most complete genome within this order and co-generic with the only representative on the placeholder family f\_\_0-14-0-20-59-11 in f\_\_UBA93. LFW-156\_1 is the sole representative of its own family (f\_\_CABMDW01).

Functionally, LFW-144\_1 is an auxotroph for most amino acids, displaying little capacity for amino acid interconversion

(**Supplementary Figure 4**). Conversely, it carries a complete pentose-phosphate pathway and has the capacity for *de novo* biosynthesis of purine and pyrimidine nucleotides as well as nicotinate. Only the fructose-bisphosphate aldolase was undetected from the EMP pathway. While it lacks extracellular CAZymes (**Figure 2**), it has an extracellular metalloproteinase M23B (lysostaphin-like). M23 peptidases are often used to break bacterial cell walls as either a defence mechanism (bacteriocin) or for nutrition (e.g., predation or parasitism). Peroxide detoxification is mediated by rubrerythrin (PF02915), but it lacks SOR or superoxide dismutase (SOD). Energy production is derived from the fermentation of pyruvate (PorABC). Archaeal-type ATPase and type-IV pili are present. No ETC or TCA cycle components are present.

Despite the similar assembly size (<3 kbp difference) and similar %C compared to LFW-144\_1, LFW-156\_1 presents an even more limited genome, lacking complete *de novo* nucleotide biosynthesis, vitamin and cofactor biosynthetic pathways, TCA cycle, and pentose phosphate pathways. The EMP is nearly complete missing only the glucose-6-P isomerase. Conversely, it has three different extracellular glycoside hydrolases (GH74, GH57, and GH23) and one protease (S08A). Superoxide is decomposed via Fe/Mn SOD. It can ferment D-lactate



(D-lactate dehydrogenase, LDH) but not pyruvate. Pilus type IV is present. Only four subunits of A-ATPase were detected (ABDI).

### LFWA-III Includes Several Distinct but Similar ‘*Ca. Micrarchaeota*’ Lineages: Orders ‘*Ca. Gugararchaeales*,’ ‘*Ca. Burarchaeales*,’ and ‘*Ca. Anstonellales*’

During the initial examination of the DPANN MAGs, a subset of ‘*Ca. Micrarchaeota*’ MAGs was differentiated from LFWA-II and ‘*Ca. Micrarchaeales*’ because of their biosynthetic capabilities. The lineage LFWA-III was the most diverse of all *Archaea* encountered at LFLS with 9 MAGs belonging to three different

families and 9 different genera based on AAI values and phylogeny (**Supplementary Tables 14, 15**).

Further detailed phylogenetic analyses of DPANN, suggest though that LFWA-III is actually several similar but distinct orders (**Figure 3**). Two of the possible orders, LFWA-IIIa and LFWA-IIIb, are poorly represented taxa whose position on the tree is easily affected by the taxa included and substitution model (**Supplementary Figures 5, 6**). For example, in the dpann\_r89F trees, LFWA-IIIa appears as sister of LFWA-IIIc (see below) and together constitute a sister lineage to LFWA-IIIb plus ‘*Ca. Fermentimicrarchaeales*’ (**Supplementary Figure 6**). However, in the concatenated and partitioned trees based on GTDB r202, LFWA-IIIa + LFWA-IIIb are sister to ‘*Ca.*

Fermentimicrarchaeales,' and these together sister to LFWA-IIIc. While based on the RED values from r202 would indicate LFWA-IIIa and LFWA-IIIb are likely to belong to the same order, their unstable position in respect to each other and to LFWA-IIIc and 'Ca. Fermentimicrarchaeales,' they would be considered independent orders, as the GTDB r202 also does. In the case of LFWA-IIIc, it seems to be topologically consistent across all trees. LFWA-IIIc is equivalent to a pre-existing order in GTDB, o\_\_UBA10214 (**Figure 3**).

LFWA-IIIa, or 'Ca. Gugararchaeales,' includes a single MAG (**Table 2**) of high quality: LFW-121\_3, 'Ca. Gugararchaeum adminiculabundum' LFW-121\_3 (96.8%C, 2.2%R, 49.8% GC). LFWA-IIIb, or 'Ca. Burarchaeales' also contains a single high-quality MAG, 'Ca. Burarchaeum australiense' LFW-281\_7 (96.8%C, 5.4%R, 57.6% GC). Both MAGs constitute the types of their respective families, 'Ca. Gugararchaeaceae' and 'Ca. Burarchaeaceae.' On the other hand, LFWA-IIIc or 'Ca. Anstonellales,' is the most diverse order in LFLS waters, with 7 MAGs from two different families, but only two of high quality: LFW-35, 'Ca. Anstonella stagnisolia' (97.8%C, 2.2%R, 62.2% GC), and LFW-283\_2, 'Ca. Bilamarchaeum dharawalense' (95.7%C, 0%R, 40.0% GC). 'Ca. Anstonellaceae' would be equivalent to f\_\_UBA10161, while 'Ca. Bilamarchaeaceae' would be equivalent to f\_\_UBA10214. In terms of abundance, LFWA-III, in general, constituted >25% of the archaeal community at any individual sampling day with 'Ca. Anstonella stagnisolia' LFW-35 being the most abundant *Archaea* at day 0 and the second most abundant during the remaining days (**Supplementary Table 9**).

### Comparative functional analysis

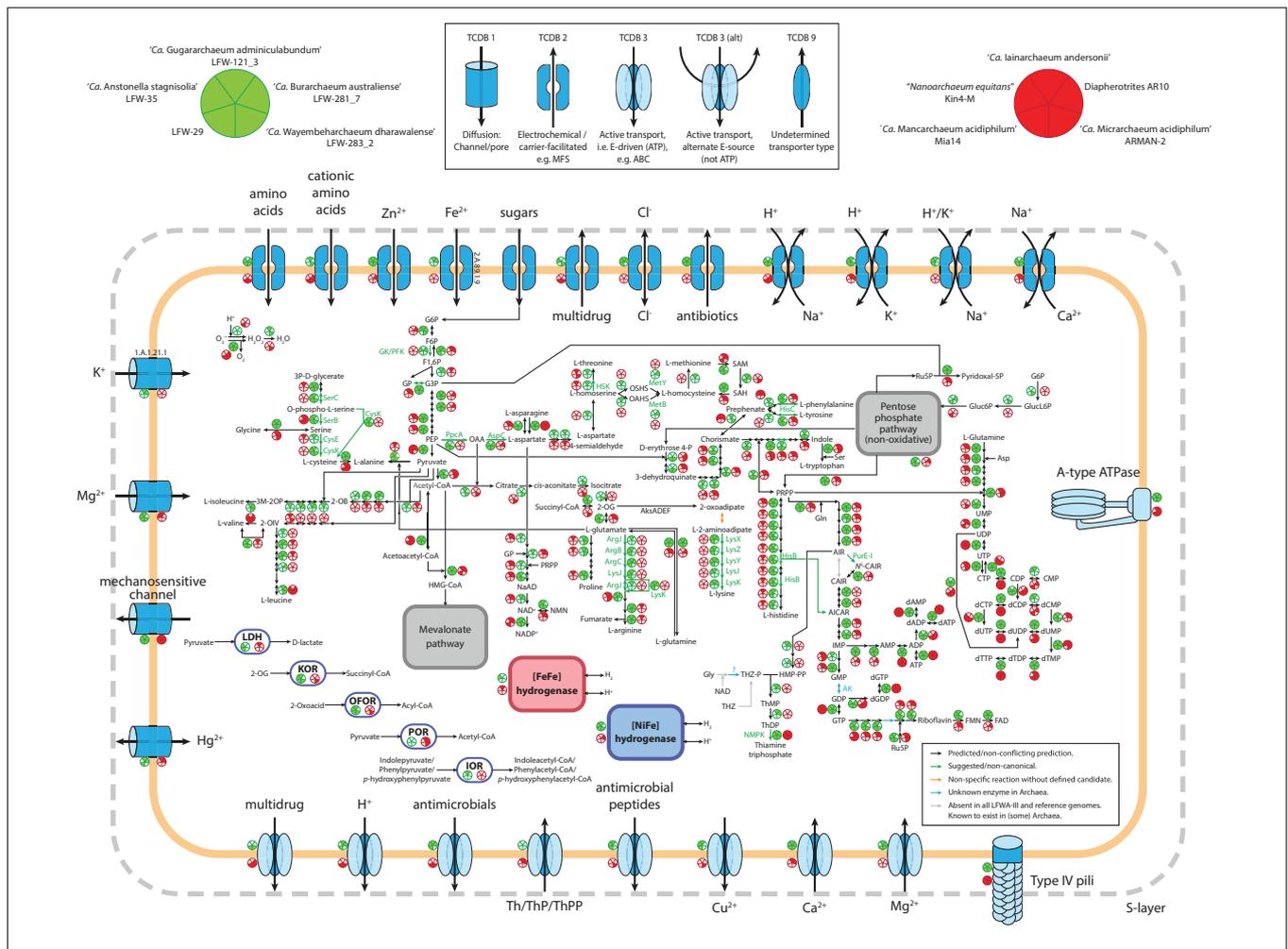
In addition to their high abundance and diversity, LFWA-III is also an interesting lineage at a functional level, constituting one of the few DPANN with broad anabolic capabilities including the *de novo* biosynthesis of the amino acids lysine, arginine, and cysteine as well as purine and pyrimidine nucleotides (**Figure 4**). The metabolic capabilities of the top five LFWA-III genomes (the four named species plus LFW-29) were compared with five of the better characterised and/or complete DPANN genomes (**Figure 4**): 'Ca. Iainarchaeum andersonii' (Youssef et al., 2015) and AR10 (Castelle et al., 2015) ('Ca. Diapherotrites'), 'Ca. Micrarchaeum acidiphilum' ARMAN-2 (Baker et al., 2006) and 'Ca. Mancarchaeum acidiphilum' Mia14 (Golyshina et al., 2017) ('Ca. Micrarchaeota'), and "*Nanoarchaeum equitans*" Kin4-M (Huber et al., 2002; Waters et al., 2003) ("*Nanoarchaeota*"). In general, both reference and LFWA-III MAGs have type-IV pili, A-ATPases (except "*Nanoarchaeum equitans*") and lack ETC and TCA cycle components (except ARMAN-2). In terms of core metabolism, the representative genomes of LFWA-III share more similarities with 'Ca. Diapherotrites' than with the reference 'Ca. Micrarchaeota,' e.g., *de novo* biosynthesis pathways for nucleotides, EMP pathway, *de novo* biosynthesis of nicotinate/nicotinamide or aromatic amino acids (**Figure 4**). The EMP pathway is also shared with other DPANN not shown in **Figure 4**, e.g., LFWA-II, LFWA-IV and 'Ca. Fermentimicrarchaeum limneticum,' while the *de novo* biosynthesis pathways for nucleotides are also shared with the latter. However, LFWA-III also contains other unique

features: pathways for the biosynthesis of L-arginine, L-lysine, L-cysteine, and L-histidine (only in AR10 in the reference genomes). LFW-121\_3 and LFW-281\_7 display most of the genes encoding the pathway for the biosynthesis of branched amino acids. Conversely, in the case of LFWA-IIIc, and likely also with 'Ca. Iainarchaeum andersonii,' would likely rely on the interconversion of L-valine and L-leucine. LFW-121\_3 in particular presents unique features that either distinguish it from other LFWA-III genomes and/or the reference genomes. Both LFW-121\_3 and 'Ca. Iainarchaeum andersonii' are the only MAGs able to convert L-aspartate to L-threonine based upon current evidence.

Another interesting difference relates to the thiamine acquisition (**Figure 4**). Both *Diapherotrites* genomes have transporters for thiamine and/or its phosphorylated forms but lack any enzyme known in the *de novo* pathway for its biosynthesis. Instead, LFW-121\_3 and LFW-281\_7, as well as 'Ca. Fermentimicrarchaeum limneticum,' are likely to be able to produce thiamine *de novo* provided a source for the thiazole branch of the thiamine pathway is available (see **Supplementary Information** for further discussion). None of the LFWA-IIIc MAGs have the biosynthetic pathway nor specific transporters for thiamine.

Like other DPANN, most LFWA-III are fermentative organisms. However, there are differences in the substrates used. Leaving aside LFW-281\_7, for which we couldn't detect any of the proteins mentioned next, all other LFWA-III and ARMAN-2 can ferment 2-oxoglutarate (KorAB), and other 2-oxoacids (OforAB). Pyruvate fermentation was detected in both *Diapherotrites* and ARMAN-2 but not in any LFWA-III. 'Ca. Anstonellaceae' MAGs (LFW-35, LFW-29) were the only ones with LDH, a differential characteristic from the other LFWA-III, but shared with 'Ca. Iainarchaeum andersonii.' LFW-121\_3 was the only one with a predicted capacity for the fermentation of aromatic amino acids amongst the 10 genomes compared (IorAB). LFW-281\_7 also lacks the [NiFe] group 3b hydrogenase present in the other LFWA-III and AR10. The [NiFe] group 3b hydrogenases are different from the [NiFe] group 4e found in 'Ca. Fermentimicrarchaeum limneticum' (Kadnikov et al., 2020). Both groups of hydrogenases are considered to be bidirectional under the proper conditions but they have some key differences. While the [NiFe] group 3b are cytosolic, O<sub>2</sub>-tolerant and couple the oxidation of NADPH to fermentative evolution of H<sub>2</sub>, the [NiFe] group 4e are membrane-bound, O<sub>2</sub>-sensitive and often associated to respiratory complexes (Søndergaard et al., 2016). Based on the lack of fermentative proteins and the associated hydrogenase, LFW-281\_7 is likely the only non-fermenting member of LFWA-III.

Regarding transporters (**Figure 4**), the high variability across individual MAGs makes generalisations difficult. Nonetheless, certain types of transporters show a distinctive distribution. Chloride channel proteins (TCDB 2.A.49.6.1), TCDB-2 antibiotic exporters (e.g., MarC family), H<sup>+</sup>,K<sup>+</sup>/Na<sup>+</sup> antiporters and Mg<sup>2+</sup> P-type ATPase (MgtA/MgtB, Mg<sup>2+</sup> import) appear unique to LFWA-III. Two transporters appear to be shared and specific for LFW-121\_3 and LFW-281\_7: the potassium channel protein (TCDB 1.A.1.21.1) and the uncharacterised TCDB 1.B.78.1.4 now



**FIGURE 4 |** Comparative visualization of the metabolism of LFWA-III representative MAGs from this study (green circles) and selected reference DPANN genomes (red circled). Each sector in the indicative circles indicates the presence or absence of a key protein in the represented genome. Extracellular enzymes are not shown for simplicity. Abbreviations: 2-OG, 2-oxoglutarate; 2-OIV, 2-oxoisovalerate; AICAR, 5-amino-1-(5-phospho-D-ribose)imidazole-4-carboxamide; AIR, 5-amino-1-(5-phospho-β-D-ribose)imidazole; ARP, 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione; CAIR, 5-amino-1-(5-phospho-D-ribose)imidazole-4-carboxylate; F1,6P, β-D-Fructose 1,6-bisphosphate; F6P, β-D-Fructose-6P; FAD, Flavin adenine dinucleotide; FMN, Flavin mononucleotide; G3P, Glycerolaldehyde-3P; G6P, Glucose-6P; Gluc6P: D-gluconate 6-phosphate; GlucL6P: D-glucono-1,5-lactone 6-phosphate; GP, Glycerone-P; HMP-PP, 4-amino-5-hydroxymethyl-2-methylpyrimidine diphosphate; IOR: indolepyruvate:ferredoxin oxidoreductase; KOR: 2-oxoglutarate:ferredoxin oxidoreductase; LDH: D-lactate dehydrogenase (NAD<sup>+</sup>); N5-CAIR, 5-(carboxyamino)imidazole ribonucleotide; NAD<sup>+</sup>, β-nicotinamide adenine dinucleotide; NADP<sup>+</sup>, β-nicotinamide adenine dinucleotide phosphate; NMN, β-nicotinamide D-ribonucleotide; OAA, Oxaloacetate; OAHS, O-acetyl-homoserine; OFOR: 2-oxoacid:ferredoxin oxidoreductase; OSHS, O-succinyl-homoserine; PEP, Phosphoenolpyruvate; POR: pyruvate:ferredoxin oxidoreductase; PRPP, 5-phospho-α-D-ribose 1-diphosphate; Ru5P, D-ribulose 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.

annotated as a VIT family protein (TCDB 2.A.89.1.9), often involved in iron homeostasis. Ca<sup>2+</sup>/Na<sup>+</sup> and H<sup>+</sup>/K<sup>+</sup> antiporters, and Ca<sup>2+</sup> P-type ATPase are common to both LFWA-III and *Diapherotrites*.

### LFWA-IV: Another Member of an Obscure DPANN Lineage

The MAG LFW-46 is the sole representative of the LFWA-IV lineage (Figure 1). While presenting good C/%R levels for a functional description (92.5% C, 3.2% R, 43.8% GC), no rRNA genes aside from a partial 5S were recovered even when searching the assembly graph directly (thus no name is proposed). All the

DPANN trees place LFW-46 as part of p\_EX4484-52, sister to ‘*Ca. Nanohaloarchaeota*’ (Supplementary Figure 2).

The LFW-46 genome is a typical example of a reduced, host-dependent DPANN (Supplementary Figure 7) with very limited amino acid interconversion pathways, lacking both the *de novo* nucleotide biosynthesis, the TCA cycle and the pentose phosphate pathways. It also lacks biosynthetic pathways for the production of vitamins and cofactors. However, LFW-46 differentiates itself by the relatively rich repertoire of transporters related to resistance, especially in comparison with the repertoire of other DPANN: multidrug, tetracycline, quinolones, arsenite, Cu<sup>+</sup> and Cu<sup>2+</sup>, and mono-/di-valent

organocations (e.g., quaternary ammonium compounds). It relies on the EMP pathway, but the enzyme(s) for the conversion of glyceraldehyde-3P to 3-phosphoglycerate could not be detected by KofamScan. However, arNOG annotation identified GapN (arCOG01252) related to the glyceraldehyde-3-phosphate dehydrogenase [NAD(P) +] from *Vulcanisaeta distributa* (seed *e*-Value:  $6.7 \cdot 10^{-101}$ ). The extracellular enzymes of LFW-46 include one protease (S08A), one glycoside hydrolase (GH57), and one polysaccharide lyase (PL9). ROS detoxification is mediated by neelaredoxin/desulfoferredoxin and rubrerythrin. LFW-46 ferments pyruvate for energy (PorABCD) and has no ETC components. Archaeal-type ATPase and type-IV pili are both present.

### Multivariate Comparison of DPANN Genomes

Sparse Principal Component Analyses (sPCA) of the functional (e.g., COG, and MEROPS annotations) and/or compositional (amino acid composition of predicted proteins, and GC%) profiles of selected *Archaea* and DPANN genomes (only  $\geq 75\%C$  based on CheckM) were performed in order to reveal the overall distinctiveness of LFWA lineages. In the all-*Archaea* dataset with compositional + functional data, LFWA-III clustered closer to other DPANN genomes (Supplementary Figure 8) yet positioned near to what could be referred to as a 'transition zone' between DPANN and normal-sized archaeal genomes (e.g., *Euryarchaeota* and TACK). Compositional and functional analysis provided an insight into the uniqueness of LFWA-III: displaying high GC% (usually  $\sim 50\%$  or greater,

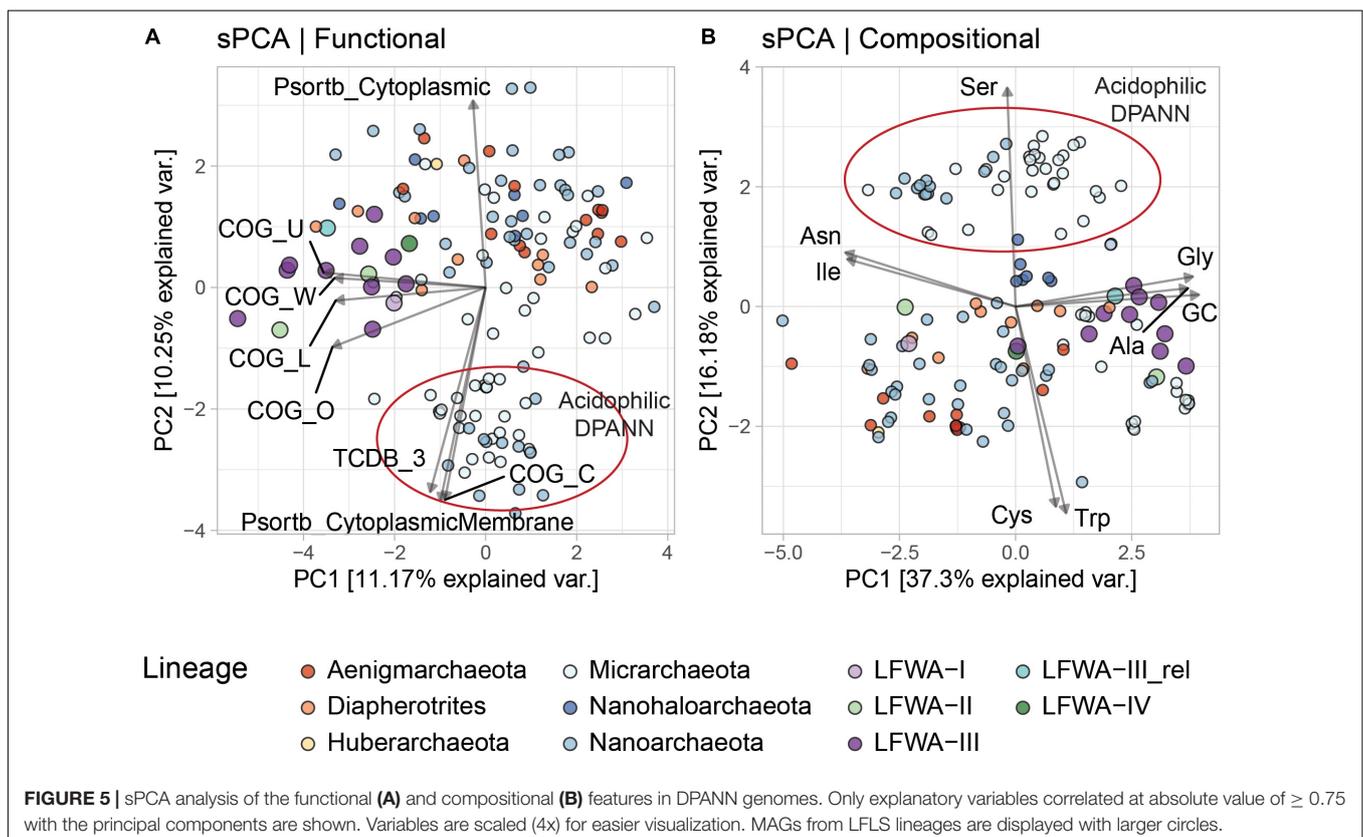
Supplementary Figure 9) and an overall above-average number of COG-annotated proteins.

Functional sPCA of the 140 DPANN genomes enabled differentiation of the metabolically rich (e.g., LFWA-II, LFWA-III, '*Ca. Diapherotrites*' and related) and acidophilic lineages from the other DPANN genomes (Figures 5A,B). This differentiation of LFWA-III was mainly shown in PC1 (Figure 5A; COG categories L, O, U, W) and PC3 (not shown; COG E, F, H and MEROPS C).

The high positive loadings in the compositional dataset along PC1 for Gly, Ala, and GC, and negative loadings for Ile and Asn may be due to the high GC content of most LFWA-III genomes (Figure 5B and Table 2), considering that those amino acids are encoded by GC-rich and AT-rich codons respectively (Grosjean and Westhof, 2016). Very high positive values along PC2 separated the acidophilic DPANN related to the acidophilic '*Ca. Micrarchaeota*' and '*Ca. Parvarchaeota*' from the other DPANN.

### Diverse and Abundant '*Ca. Methanoperedens* spp.'

The occurrence of three methanogenic MAGs from the '*Methanotrichaceae*' family (previously indicated to be the illegitimate *Methanosaetaceae*, Vázquez-Campos et al., 2017) 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*.' A total of six MAGs were



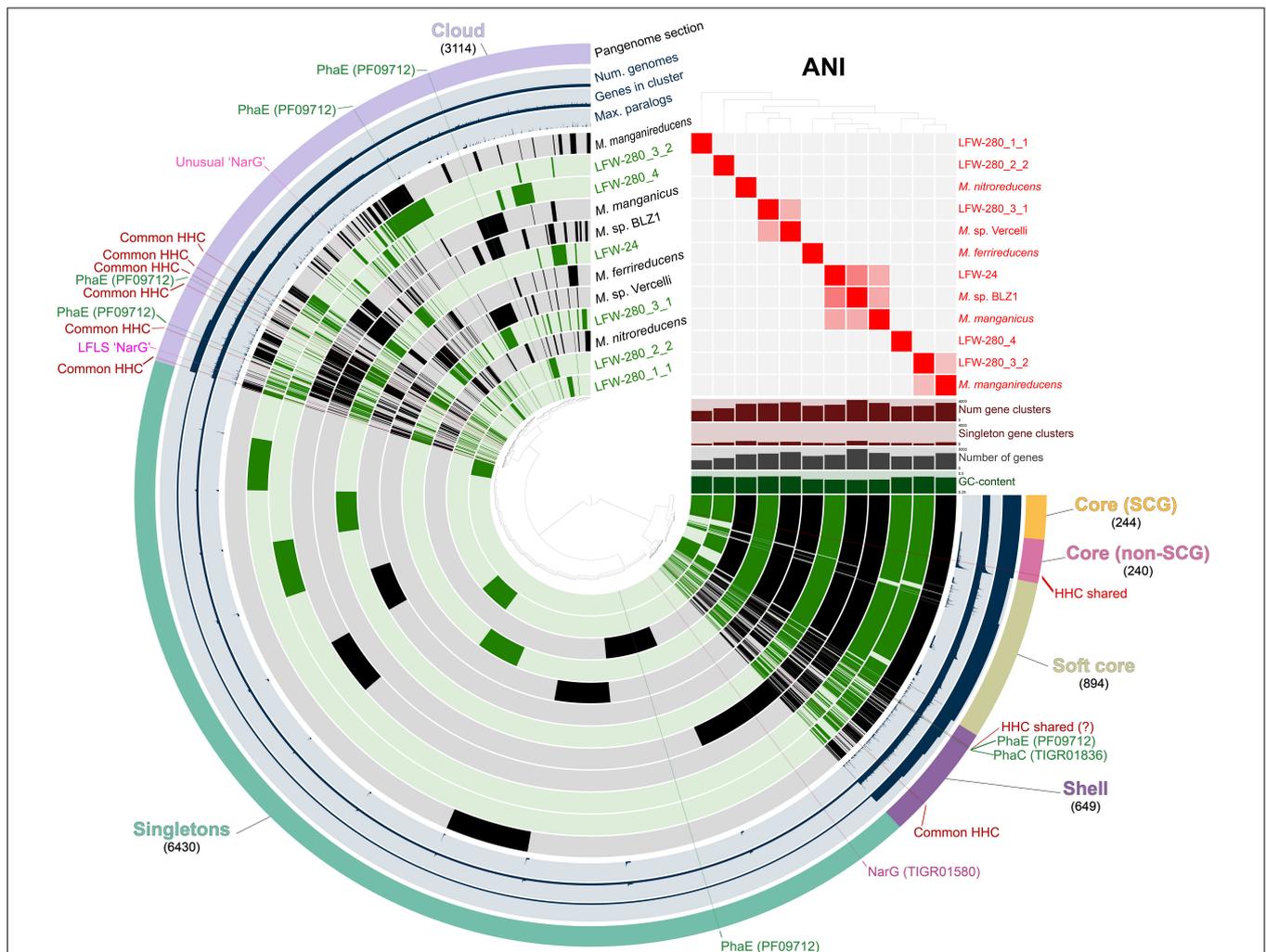
assembled in this study,  $4 \geq 95\%C$  (all  $< 10\%R$ ). Each of these MAGs belong to different species based on ANI and AAI scores (Figure 6 and Supplementary Tables 16, 17), constituting the most diverse set of '*Ca. Methanoperedens* spp.' genomes reconstructed to date in a single study and site, and all representing different species to previously sequenced MAGs.

The rich collection of '*Ca. Methanoperedens* spp.' recovered, in combination with selected published genomes, allows for the investigation of common aspects across the whole genus. A total of six reference genomes are included in all comparisons and pangenome analysis: '*Ca. Methanoperedens nitroreducens*' ANME-2D (Haroon et al., 2013), '*Ca. Methanoperedens ferrireducens*' (Cai et al., 2018), '*Ca. M. nitroreducens*' BLZ1 (Arshad et al., 2015), '*Ca. M. nitroreducens*' Vercelli (Vaksmas et al., 2017), '*Ca. M. manganicus*' (Leu et al., 2020), and '*Ca. M. manganireducens*' (Leu et al., 2020) (see Taxonomic Appendix and Supplementary Information). It should be noted

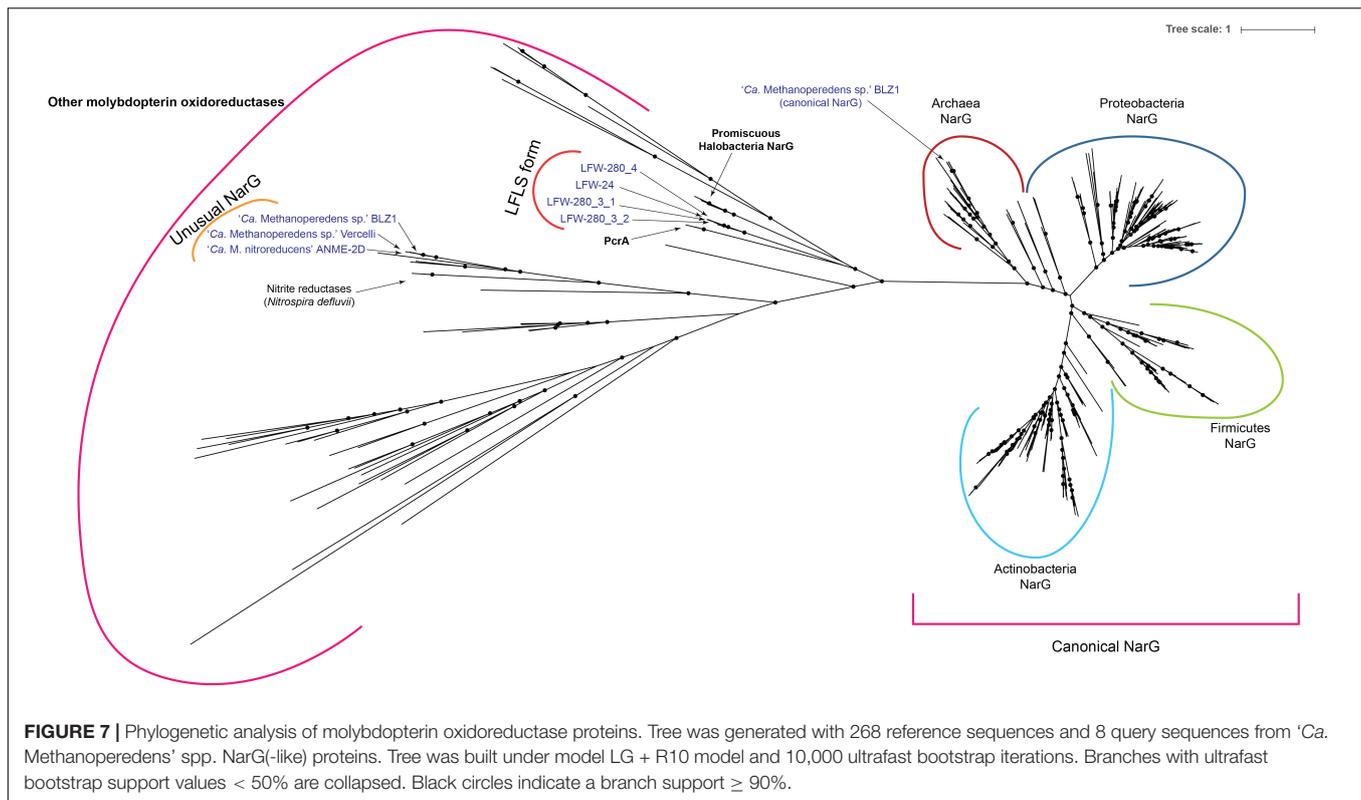
that based on ANI values, neither '*Ca. M. nitroreducens*' BLZ1 or '*Ca. M. nitroreducens*' Vercelli belong to the same species with '*Ca. Methanoperedens nitroreducens*' ANME-2D, and thus, to avoid confusion, we will refer to them as '*Ca. Methanoperedens* sp.' BLZ1 and '*Ca. Methanoperedens* sp.' Vercelli, respectively.

### Multiple Acquisition of Respiratory Molybdopterin Oxidoreductases

Most known '*Ca. Methanoperedens* spp.' encode some molybdopterin oxidoreductase enzyme, which are usually referred to as NarG. All the '*Ca. Methanoperedens* spp.' found at LFLS share a putative molybdopterin oxidoreductase that is not present in other members of the genus (Figure 7). Notably, this is different from any canonical or putative NarG (as is often suggested for other '*Ca. Methanoperedens* spp.') with no match to HMM profiles characteristic of these proteins but, instead,



**FIGURE 6 |** Pangenomic analysis of genus '*Candidatus Methanoperedens*.' Pangenome was generated with 6 '*Ca. Methanoperedens* spp.' MAGs from LFLS (green) and 6 reference genomes (black), accounting for a total of 39,271 protein-coding genes grouped in 11,571 gene clusters. Gene clusters generated with an MCL inflation value of 6.0. 'Common HHC' indicates high-heme cytochrome gene clusters present in at least 4 different MAGs. MAGs are ordered based on the presence-absence of gene clusters. ANI heatmap show values between 85 and 100%.



belonging to arCOG01497/TIGR03479 (DMSO reductase family type II enzyme, molybdopterine subunit).

Phylogenetic analysis (Figure 7) places the molybdopterine oxidoreductase form of LFLS ‘*Ca. Methanoperedens* spp.’ far from molybdopterine oxidoreductases from the reference ‘*Ca. Methanoperedens* spp.’ However, it is located closer to non-canonical NarG sequences from *Haloferax mediterranei* (I3R9M9) (Lledó et al., 2004) and other *Halobacteria*, as well as to the perchlorate reductase (PcrA) of *Dechloromonas aromatica* and *Azospira oryzae* (previously known as *Dechlorosoma suillum* strain PS) (He et al., 2019).

### High Heme Cytochromes Are Common in ‘*Ca. Methanoperedens* spp.’ Genomes

Metal respiration is usually associated with cytochromes with a high number of heme-binding sites ( $\geq 10$ ) or high heme cytochromes (HHC from here on). The screening of the ‘*Ca. Methanoperedens* spp.’ MAGs from LFLS revealed multiple HHC in all genomes, with a minimum of six copies for LFW-280\_1\_1 (68.5%, Figure 2 and Supplementary Table 10). All reference ‘*Ca. Methanoperedens* spp.’ genomes revealed similar abundances when considering their completeness, i.e.,  $\sim 10$  HHC per genome. While most HHC gene clusters are singletons (48/73 gene clusters), there is one highly conserved gene cluster in the core genome (GC\_00000082, median identity 75.5%) of ‘*Ca. Methanoperedens*’ containing 13 heme-binding sites. Another HHC that might be shared across all ‘*Ca. Methanoperedens*’ is the GC\_00001534 with 11 heme binding sites, although it is missing from the less complete LFLS MAGs and further

confirmation would be needed to assess if it could be part of the core genome and not an artefact due to low completeness. Although metal reduction has only been confirmed for a limited portion of the known ‘*Ca. Methanoperedens*’ that exist: ‘*Ca. Methanoperedens ferrireducens*’, ‘*Ca. Methanoperedens* sp.’ BLZ1, ‘*Ca. Methanoperedens manganireducens*’ and ‘*Ca. Methanoperedens manganicus*’ (Ettwig et al., 2016; Cai et al., 2018; Leu et al., 2020), our data suggests metal reduction might be a universal characteristic of ‘*Ca. Methanoperedens* spp.’

### Biosynthesis of Polyhydroxyalkanoates

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 (Reddy et al., 2003; Martínez-Gutiérrez et al., 2018). Genes involved in the biosynthesis of polyhydroxyalkanoates (PHA) were detected in three of the most complete ‘*Ca. 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 (Figure 2 and Supplementary Table 10).

In ‘*Ca. Methanoperedens* spp.’ the genes associated with the biosynthesis of PHA are usually present as an operon (Supplementary Figure 10A). 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 ‘*Ca. Methanoperedens*’ reference genomes included in this study (Haroon et al., 2013;

Arshad et al., 2015; Vaksmaa et al., 2017; Cai et al., 2018; Leu et al., 2020). However, while all PhaC proteins form a gene cluster in the pangenome analysis, PhaE does not (Figure 6). This could be related to their function: PhaC is a catalytic protein while PhaE ‘simply’ regulates PhaC (Koller, 2019), 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). In most instances, PHA biosynthesis genes have been found in putative operons (Supplementary Figure 9A).

## DISCUSSION

In this study we describe 23 DPANN *Archaea* candidate taxa (species to order), as well as proposing four supraspecific taxa to accommodate extant and newly proposed lineages according to the current taxonomy; proposed *Candidatus* names have been proposed (family to class), following the Linnaean system.

The archaeal community at LFLS, showed a clear dominance of DPANN and ‘*Ca. Methanoperedens* spp.’ with the data sets’ uniqueness highlighted by the unmatched MAGs (at species level) with any extant reference genomes. Amongst the DPANN MAGs recovered, a number of them constitute the first detailed description of lineages with placeholder names from large scale projects (Parks et al., 2017, 2020; Rinke et al., 2020).

### Self-Sustaining DPANN – Possibly More Common Than Previously Thought?

The DPANN archaea *sensu stricto* (i.e., not including ‘*Ca. Altiarchaeota*’) are often described as organisms with extremely reduced metabolic capacities and generally, with few exceptions, dependent on a host for the acquisition of essential metabolites such as vitamins, amino acids or even reducing equivalents (Dombrowski et al., 2019). During the metabolic reconstruction of individual archaeal genomes, it was observed that select DPANN from the LFWA-III group, exemplified by LFW-121\_3 (‘*Ca. Gugararchaeum adminiculabundum*’), possess near-complete pathways for the synthesis of amino acids, purine and pyrimidine nucleotides, riboflavin, and thiamine (Figure 4). These 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 pathways in the ‘*Ca. Gugararchaeales*’, ‘*Ca. Burarchaeales*’ and ‘*Ca. Anstonellales*’ MAGs had a limited number of gaps (Figure 4). These alleged ‘gaps,’ thoroughly discussed in the Supplementary Information with focus on LFW-121\_3, were often related to unusual enzymes poorly annotated in databases, or steps known to be possibly 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 perform those functions.

The discovery of additional DPANN taxa with a “rich” metabolism, at least compared with host-dependant

DPANN (e.g., “*Nanoarchaeum*,” ‘*Ca. Nanopusillus*,’ Mia14), highlights the metabolic diversity and widespread existence of free-living DPANN *Archaea*, at least within the *Diapherotrites/Micrarchaeota* lineage – DM branch or cluster 1 (Dombrowski et al., 2020) – as opposed to the PANN branch, suggesting a more generalised genome streamlining process/loss of function in the latter. Prior studies have remarked upon the generalised “limited metabolic capacities” in most DPANN (Dombrowski et al., 2019). However, many DPANN datasets are heavily enriched with lineages that are likely to be host-dependent, e.g., ‘*Ca. Pacearchaeota*’ or ‘*Ca. Woesearchaeota*’ (suggested to be orders in the GTDB), or derived from acidic environments, generally restricted to ARMAN lineages, esp. ‘*Ca. Micrarchaeaceae*.’ This study expands the phylum ‘*Ca. Micrarchaeota*’ with 3 orders, 5 families, 5 genera, and 5 species (Figure 3).

Several studies have reported that ultra-small *Bacteria* and *Archaea* can be lost using standard 0.22  $\mu\text{m}$  filters (Luef et al., 2015; Chen et al., 2018). However, this may only affect individual cells, while host-DPANN cell associations would be retained and overrepresented compared to free-living DPANN, which are not necessarily larger than the host-associated organisms. The high abundance of the putative free-living LFWA-III in the metagenomic samples from LFLS could be related to the elevated concentrations of labile colloidal Fe that precipitates at the redox interface following the infiltration of oxic-rainwaters. Acting in a similar manner as a flocculant in water treatment, iron (oxyhydr)oxide microaggregates have immense capacity to trap organic matter and microbial cells (Tang et al., 2016). This process has been used in protocols to recover and concentrate viruses from environmental samples without the need of ultrafiltration (John et al., 2011). As such, the colloidal Fe would also have improved the recovery of free-living DPANNs and any other ultra-small microorganisms that, by default would have been otherwise lost during the filtration with the 0.22  $\mu\text{m}$  filters used during sampling (Vázquez-Campos et al., 2017).

### ‘*Ca. Methanoperedens*’: Molybdopterin Oxidoreductases, Metal Reduction, and Polyhydroxyalkanoates Biosynthesis

‘*Ca. Methanoperedenaceae*’ archaea have been reported (or at least suggested) to be able to utilise a wide range of inorganic electron acceptors for AOM, including nitrate (Haroon et al., 2013), nitrite (Arshad et al., 2015), Fe(III) (Ettwig et al., 2016; Cai et al., 2018), Mn(IV) (Ettwig et al., 2016), and Cr(VI) (Lu et al., 2016). The type species of the genus ‘*Ca. Methanoperedens*,’ ‘*Ca. Methanoperedens nitroreducens*,’ receives its name from its ability to use nitrate as an electron acceptor (Haroon et al., 2013). However, the alpha subunit of this candidate respiratory nitrate reductase (NarG<sub>Mn</sub>) is a molybdopterin oxidoreductase different from any canonical or putative NarG, with no match with TIGRFAM, CDD or Pfam profiles characteristic of these proteins. This NarG<sub>Mn</sub> does not belong to the same orthologous group alongside other typical archaeal NarG (arCOG01497), but to ENOG4102T1R, which includes the well-characterised “dimethyl sulfoxide reductase subunit A” (DmsA) of several

*Halobacteria*. The NarG<sub>Mn</sub> has a high similarity (>80%) with proteins from both ‘*Ca. Methanoperedens* sp.’ BLZ1 (NarG<sub>1MB</sub>) and ‘*Ca. Methanoperedens* sp.’ Vercelli (NarG<sub>MV</sub>) and relates to the nitrite reductases of *Nitrospira defluvii* (Figure 7). While ‘*Ca. Methanoperedens* sp.’ Vercelli and ‘*Ca. Methanoperedens* sp.’ BLZ1 have only this unusual NarG, ‘*Ca. Methanoperedens* sp.’ BLZ1 harbours an additional, unrelated, canonical NarG (NarG<sub>2MB</sub>) based on the detection by the TIGRFAM profile (TIGR01580, arCOG01497).

Several of the NarG related to NarG<sub>LFLS</sub> (from *Halobacteria*) have been confirmed to have certain promiscuity (Figure 7) and are known to be able to use (per)chlorate as a substrate *in vivo* (Yoshimatsu et al., 2000; Oren et al., 2014). Although it is close to impossible to predict the actual substrates of the NarG<sub>LFLS</sub> (or PcrA<sub>LFLS</sub>), the phylogenetic analysis is, at least, suggestive of the potential utilisation of (per)chlorate. The oxidation of methane linked to the bioreduction of perchlorate has been proposed a number of times in the past (Luo et al., 2015; Xie et al., 2018; Wu et al., 2019). Recent works (Xie et al., 2018; Wu et al., 2019) have suggested the possible involvement of ANME archaea in (per)chlorate reduction based on chemical and amplicon data analyses, an assumption considered more likely given that nitrate was rarely measured in noticeable concentrations in the LFLS sampling trench (Vázquez-Campos et al., 2017). Indeed, nitrate concentrations were below the detection limit (<0.01 μM) before any ‘*Ca. Methanoperedens* spp.’ became relatively abundant. With LFLS historical disposal records indicating that quantities of perchlorate/perchloric acid were deposited in the vicinity of the sample location, it collectively suggests that ‘*Ca. Methanoperedens* spp.’ utilisation of nitrate is unlikely at the site. However, as no chlorite dismutase gene was found in any of the MAGs, the possibility that they may utilise perchlorate is also not well supported. Given the high iron concentrations, it is reasonable to assume that the AOM at LFLS mainly use Fe(III) electron acceptor due to its ready availability, although other redox-active metals detected in the trenches, e.g., Mn, might also be used (Vázquez-Campos et al., 2017).

The collective evidence indicates that nitrate reductase(-like) enzymes are rather widespread in ‘*Ca. Methanoperedens* spp.’ with the exception of ‘*Ca. Methanoperedens ferrireducens*’ from which no nitrate reductase candidate could be detected, aside from an orphan NapA-like protein (cd02754). Phylogenetic analysis of the NarG and similar proteins (Figure 7) suggests that these proteins might have been horizontally acquired at least three different times during the evolution of ‘*Ca. Methanoperedens*’. Congruently, the pangenomic analysis showed three different NarG variants in their respective gene clusters (Figures 6, 7).

### High-Heme Cytochromes

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 close proximity of the collected samples (Payne, 2012). 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 (~0.5–1 mM) (Vázquez-Campos et al., 2017).

Utilisation of metals (e.g., Fe, Mn, Cr, and U) as electron acceptors requires the presence of multi-heme cytochromes. Although multi-heme cytochromes are not intrinsically and exclusively employed for heavy metal reduction (e.g., decaheme DmsE for DMSO reduction), HHC are characteristic of microorganisms performing biological reduction of metals. In our case, HHC were found in all new and reference ‘*Ca. Methanoperedens*’, including a gene cluster from the core genome. This may well indicate that oxidised metal species were the ancestral electron acceptor to all ‘*Ca. Methanoperedens*’, rather than nitrate or other non-metallic oxyanion moieties.

Previous studies indicate that ‘*Ca. Methanoperedens* sp.’ BLZ1 and ‘*Ca. Methanoperedens ferrireducens*’ can both use the Fe(III) (oxyhydr)oxide ferrihydrite as electron acceptor for AOM (Ettwig et al., 2016; Cai et al., 2018). Reports showing the involvement of more crystalline forms of iron oxides for the 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 (Sivan et al., 2014). Based on the evidence to hand, it is not clear whether ‘*Ca. Methanoperedens*’ or other ANME could be responsible for these observations.

Tantalising questions remain as to the role of ‘*Ca. 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 (Ikeda-Ohno et al., 2014; Vázquez-Campos et al., 2017).

### Polyhydroxyalkanoates Biosynthesis

Variable nutrient levels can result in the microbial production of polyhydroxyalkanoates (PHA), which are carbon-rich, energetic polymers resulting from the polymerisation of hydroxyalkanoates (Reddy et al., 2003; Martínez-Gutiérrez et al., 2018). The biosynthesis of PHA is a widespread characteristic in many groups of aerobic *Bacteria* (Reddy et al., 2003; Martínez-Gutiérrez et al., 2018). However, very few strict anaerobes are able to synthesise PHA, being mostly limited to syntrophic bacteria (Spang et al., 2012). In *Archaea*, the biosynthesis of PHA is known to occur, particularly in many *Nitrososphaerales* (“*Thaumarchaeota*”) (Spang et al., 2012) and *Euryarchaeota*, where it has been traditionally limited to *Halobacteria* (Koller, 2019).

Key genes, as well as full operons, have been detected in several ‘*Ca. Methanoperedens*’ genomes, with recent experimental evidence identifying the production of PHA by ‘*Ca. Methanoperedens nitroreducens*’ (Cai et al., 2019). This suggests that the biosynthesis of PHA could be a widespread feature in ‘*Ca. Methanoperedens* spp.’ or even in all ‘*Ca. Methanoperedens*’, 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 (Karthikeyan et al., 2015).

It has been estimated that anaerobic methane oxidisers may consume up to 80–90% of the methane produced in certain environments, mitigating its release to the atmosphere (Reeburgh, 2007). The capacity of ‘*Ca. Methanoperedens* spp.’ to accumulate PHA might have implications for the further refinement of these estimates. The inference being that methane would not solely be used for energy production or to increase cell numbers, and that ‘*Ca. Methanoperedens* spp.’ could act as a ‘carbon-capture’ device, especially within carbon-rich anoxic environments whenever they are the main anaerobic methane oxidisers. This would likely be the case for the LFLS test trench, where ammonium, nitrate/nitrite and total dissolved nitrogen are limiting.

## CONCLUSION

While the *Archaea* inhabiting the LFLS trench water constitute a relatively small portion of its microbial community, they have important roles in biogeochemical cycling, especially with respect to methanogenesis, anaerobic methane oxidation and Fe(III) reduction. The diverse ‘*Ca. Methanoperedens* spp.’ are of special relevance due to their role in methane capture (and subsequent conversion to PHA), Fe cycling, and dissimilatory pathways for nitrate and, possibly, (per)chlorate reduction.

The broad phylogenetic representation of DPANN organisms, which have attracted particular attention in recent years, either for their unusual characteristics or their controversial evolutionary history, is another interesting feature of the LFLS archaeal community. The description and evaluation of the MAGs from lineages LFWA-I to IV constitutes a detailed first examination of several undescribed archaeal lineages.

The proposed LFWA-III lineages, ‘*Ca. Gugararchaeales*’, ‘*Ca. Burarchaeales*’, and ‘*Ca. Anstonellales*’ are amongst the most interesting. Foremost, they are together unusually diverse at LFLS, with 9 out of the 37 MAGs recovered in this study belonging to the 3 orders, with 4 families and 4 genera (and species) newly proposed within them. Furthermore, they have a cohesive central metabolism that likely spans the entire order with full pathways for the *de novo* biosynthesis of nucleotides, most amino acids, and several vitamins/cofactors. This is duly reflected in the proposed name ‘*Ca. Gugararchaeum adminiculabundum*’, proposed type for LFWA-IIIa, where the specific epithet translates as “self-supporting” relating to the possibility of not needing a symbiotic partner.

## TAXONOMIC APPENDIX

Due to several irregularities in the process of naming candidate lineages, the European Nucleotide Archive does not allow certain taxonomic names at taxonomic levels where they should exist. For example, the phylum ‘*Ca. Micrarchaeota*’, and its defining candidate species ‘*Ca. Micrarchaeum acidophilum*’,

have no defined nomenclature at class, order, or family levels. This and other issues related to the nomenclature of uncultivated *Bacteria* and *Archaea* have been previously discussed in the literature (Whitman, 2016; Konstantinidis et al., 2017, 2019; Oren and Garrity, 2018; Chuvochina et al., 2019; Rosselló-Móra and Whitman, 2019) and we will not provide further discussion on this matter.

Nonetheless, in order to cover these gaps, we feel obliged to suggest several of these intermediate nomenclatural levels, many of which are already covered in the GTDB but not proposed or described in the literature, in addition to the nomenclatural novelties intrinsic to this manuscript.

It should be noted that the use of genome sequences as type material is currently being considered to be included into the International Code on the Nomenclature of Prokaryotes Described from Sequence Data (ICNPDS, “SeqCode”) and is not accepted by ICNP (International Code of the Nomenclature of Prokaryotes).

### Description of ‘*Candidatus Tiddalikarchaeum*’ gen. nov.

‘*Candidatus Tiddalikarchaeum*’ (Ti.dda.lik.ar.chae’um. Gunai language, Tiddalik, frog from the Australian Aboriginal mythology; N.L. neut. n. *archaeum*, archaeon, from Gr. adj. *archaios* –ê –on, ancient; N.L. n. neut. *Tiddalikarchaeum*, the archaeon named after the greedy Aboriginal mythological Australian frog that burst with water, referring to the bathtub effect exhibited by the disposal trenches at the Little Forest Legacy Site).

The type species is ‘*Candidatus Tiddalikarchaeum anstoanum*.’ The genus appears as g\_\_CABMEV01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Tiddalikarchaeum anstoanum*’ sp. nov.

‘*Candidatus Tiddalikarchaeum anstoanum*’ (ans.to.a’num. N.L. neut. adj. from ANSTO, Australian Nuclear Science and Technology Organisation, institution managing the Little Forest Legacy Site).

The type material is the metagenome assembled genome (MAG) LFW-252\_1 (ERS2655302) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 1.16 Mbp in 56 contigs with an estimated completeness of 95.7%, redundancy of 1.1%, 16S, 23S, and 5S rRNA gene, and 21 tRNAs. The GC content of this MAG is 36.5%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_CABMEV01 sp902385255.

### Description of ‘*Candidatus Tiddalikarchaeaceae*’ fam. nov.

‘*Candidatus Tiddalikarchaeaceae*’ (Ti.dda.lik.ar.chae.a’ce.ae. N.L. neut. n. *Tiddalikarchaeum* a candidate genus; –aceae, ending to denote a family; N.L. fem. pl. n. *Tiddalikarchaeaceae*, the *Tiddalikarchaeum* candidate family).

The family ‘*Candidatus Tiddalikarchaeaceae*’ is circumscribed based on two independent concatenated protein phylogenies of

122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Tiddalikarchaeum*.’

The family is equivalent to f\_\_CG07-land in GTDB r89/r202 (Parks et al., 2017).

### Description of ‘*Candidatus Tiddalikarchaeales*’ ord. nov.

‘*Candidatus Tiddalikarchaeales*’ (Ti.dda.lik.ar.chae.a’les. N.L. neut. n. *Tiddalikarchaeum* a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Tiddalikarchaeales* the *Tiddalikarchaeum* candidate order).

The order ‘*Candidatus Tiddalikarchaeales*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Tiddalikarchaeum*.’

The order is equivalent to CG07-land from Probst et al. (2018) or o\_\_CG07-land in GTDB r89/r202 (Parks et al., 2017).

### Description of “*Nanoarchaeia*” class. nov.

“*Nanoarchaeia*” (Na.no.ar.chae’ia. N.L. neut. n. “*Nanoarchaeum*”, a genus; *-ia*, ending to denote a class; N.L. fem. pl. n. “*Nanoarchaeia*”, the “*Nanoarchaeum*” candidate class).

The class “*Nanoarchaeia*” is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The class “*Nanoarchaeia*” is defined as the most inclusive class that includes the genus “*Nanoarchaeum*”. The lineages ‘*Ca. Parvarchaeota*’ (o\_\_Parvarchaeales) (Rinke et al., 2013), ‘*Ca. Woesearchaeota*’ (o\_\_Woesearchaeales) and ‘*Ca. Pacearchaeota*’ (o\_\_Pacearchaeales) (Castelle et al., 2015), and order ‘*Candidatus Tiddalikarchaeales*’ (this work) are contained within “*Nanoarchaeia*”. The type genus is “*Nanoarchaeum*” (Huber et al., 2002).

This class is the only class within “*Nanoarchaeota*” Huber and Kreuter (2014), and equivalent to the c\_\_Nanoarchaeia in the GTDB r89/r202 (Parks et al., 2017).

### Description of ‘*Candidatus Gugararchaeum*’ gen. nov.

‘*Candidatus Gugararchaeum*’ (Gu.ga.rar.chae’um. Dharawal language, *gugara*, kookaburra – bird endemic to Australia, *Dacelo* spp.; N.L. neut. n. *archaeum*, archaeon, from Gr. adj. *archaios* –ê –on, ancient; N.L. neut. n. *Gugararchaeum*, the kookaburra archaeon, honouring the bird, common at the Little Forest Legacy Site, bird emblem of NSW, and typical across Australia).

The type species is ‘*Candidatus Gugararchaeum adminiculabundum*.’ The genus appears as g\_\_CABMDC01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Gugararchaeum adminiculabundum*’ sp. nov.

‘*Candidatus Gugararchaeum adminiculabundum*’ (ad.mi.ni.cu.la.bun’dum. L. adj. neut. self-supporting; in reference to the limited external requirements and its suggested independence from a host/symbiote, due to the predicted presence of pathways for the biosynthesis of amino acids, purines, pyrimidines, thiamine, and riboflavin).

The type material is the metagenome assembled genome (MAG) LFW-121\_3 (ERS2655302) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 1.47 Mbp in 76 contigs with an estimated completeness of 96.8%, redundancy of 2.2%, 16S, 23S, and 5S rRNA gene, and 21 tRNAs. The GC content of this MAG is 49.8%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_CABMDC01 sp902384795.

### Description of ‘*Candidatus Gugararchaeaceae*’ fam. nov.

‘*Candidatus Gugararchaeaceae*’ (Gu.ga.rar.chae.a’ce.ae. N.L. neut. n. *Gugararchaeum*, a candidate genus; *-aceae*, ending to denote a family. N.L. fem. pl. n. *Gugararchaeaceae*, the *Gugararchaeum* candidate family).

The family ‘*Candidatus Gugararchaeaceae*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Gugararchaeum*.’

The family appears in GTDB r202 (Parks et al., 2017) as f\_\_CABMDC01.

### Description of ‘*Candidatus Gugararchaeales*’ ord. nov.

‘*Candidatus Gugararchaeales*’ (Gu.ga.rar.chae.a’les. N.L. neut. n. *Gugararchaeum*, a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Gugararchaeales*, the *Gugararchaeum* candidate order).

The order ‘*Candidatus Gugararchaeales*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Gugararchaeum*.’

The order is equivalent to LFWA-IIIa in this manuscript and appears as o\_\_CABMDC01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Burarchaeum*’ gen. nov.

‘*Candidatus Burarchaeum*’ (Bu.rar.chae’um. Dharawal language, *buru*, kangaroo; N.L. neut. n. *archaeum*, archaeon, from Gr. adj. *archaios* –ê –on, ancient; N. L. neut. n. *Burarchaeum*, an archaeon from the land of the kangaroos, also faunal emblem of Australia).

The type species is ‘*Candidatus Burarchaeum australiense*.’ The genus appears as g\_\_CABMJK01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Burarchaeum australiense*’ sp. nov.

‘*Candidatus Burarchaeum australiense*’ (aus.tra.lien’sē. N.L. neut. adj. referring to Australia, country where its first genome was reconstructed).

The type material is the metagenome assembled genome (MAG) LFW-281\_7 (ERS2655318) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 1.20 Mbp in 76 contigs with an estimated completeness of 96.8%, redundancy of 5.4%, 16S and 5S rRNA gene, and 18 tRNAs. The GC content of this MAG is 57.6%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_CABMJK01 sp902386535.

### Description of ‘*Candidatus Burarchaeaceae*’ fam. nov.

‘*Candidatus Burarchaeaceae*’ (Bu.rar.chae.āce.ae. N.L. neut. n. *Burarchaeum*, a candidate genus; *-aceae*, ending to denote a family. N.L. fem. pl. n. *Burarchaeaceae*, the *Burarchaeum* candidate family).

The family ‘*Candidatus Burarchaeaceae*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Burarchaeum*.’

The family is equivalent to f\_\_B9-G16 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Burarchaeales*’ ord. nov.

‘*Candidatus Burarchaeales*’ (Bu.rar.chae.a’les. N.L. neut. n. *Burarchaeum*, a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Burarchaeales*, the *Burarchaeum* candidate order).

The order ‘*Candidatus Burarchaeales*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Burarchaeum*.’

The order is equivalent to LFWA-IIIc in this manuscript and o\_\_B9-G16 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Anstonella*’ gen. nov.

‘*Candidatus Anstonella*’ (Ans.to.ne’lla. N.L. dim. n. fem. *Anstonella* from ANSTO, Australian Nuclear Science and Technology Organisation, institution managing the Little Forest Legacy Site).

The type species is ‘*Candidatus Anstonella stagnisolia*.’ The genus appears as g\_\_CABMCJ01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Anstonella stagnisolia*’ sp. nov.

‘*Candidatus Anstonella stagnisolia*’ (s.tag.ni.so’lia. L. v. *stagnoverflow*; L. neut. n. *solium -a* tub, bathtub; N.L. adj. fem. *stagnisolia*, overflowing bathtub, in reference to the phenomenon described during heavy rainfalls at the Little Forest Legacy Site trenches).

The type material is the metagenome assembled genome (MAG) LFW-35 (ERS2655287) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 1.33 Mbp in 68 contigs with an estimated completeness of 97.8%, redundancy of 2.2%, 16S, 23S and 5S rRNA gene, and 21 tRNAs. The GC content of this MAG is 50.3%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_CABMCJ01 sp902384585.

### Description of ‘*Candidatus Anstonellaceae*’ fam. nov.

‘*Candidatus Anstonellaceae*’ (Ans.to.nel.la’ce.ae. N.L. neut. n. *Anstonella*, a candidate genus; *-aceae*, ending to denote a family; N.L. fem. pl. n. *Anstonellaceae*, the *Anstonella* candidate family).

The family ‘*Candidatus Anstonellaceae*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus Anstonella*.’

This family is equivalent to f\_\_UBA10161 in the GTDB r89/r202 (Parks et al., 2017).

### Description of ‘*Candidatus Bilamarchaeum*’ gen. nov.

‘*Candidatus Bilamarchaeum*’ (Bi.la.mar.chae’um. Dharawal language, *bilama*, freshwater turtle, in reference to their presence still nowadays in the creeks and rivers associated with the Little Forest Legacy Site; N.L. neut. n. *archaeum*, archaeon, from Gr. adj. *archaios -ē -on*, ancient; N.L. neut. n. *Bilamarchaeum*, an archaeon from the turtle lands).

The type species is ‘*Candidatus Bilamarchaeum dharawalense*.’ The genus appears as g\_\_CAILMU01 in GTDB r202 (Parks et al., 2017).

### Description of ‘*Candidatus Bilamarchaeum dharawalense*’ sp. nov.

‘*Candidatus Bilamarchaeum dharawalense*’ (dha.ra.wa.len’sē. N.L. neut. adj. pertaining to the Dharawal, traditional owners of the lands where the Little Forest Legacy Site is located).

The type material is the metagenome assembled genome (MAG) LFW-283\_2 (ERS2655319) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 1.27 Mbp in 63 contigs with an estimated completeness of 95.7%, redundancy of 0%, 16S, 23S and 5S rRNA gene, and 20 tRNAs. The GC content of this MAG is 40.0%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_CAILMU01 sp902386555.

## Description of ‘*Candidatus* Bilamarchaeaceae’ fam. nov.

‘*Candidatus* Bilamarchaeaceae’ (Bi.la.mar.chae.a’ce.ae. N.L. neut. n. *Bilamarchaeum*, a candidate genus; *-aceae*, ending to denote a family; N.L. fem. pl. n. *Bilamarchaeaceae*, the *Bilamarchaeum* candidate family).

The family ‘*Candidatus* Bilamarchaeaceae’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus* Bilamarchaeum.’

This family is equivalent to f\_\_UBA10214 in the GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Anstonellales’ ord. nov.

‘*Candidatus* Anstonellales’ (Ans.to.nel.la’les. N.L. neut. n. *Anstonella*, a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Anstonellales*, the *Anstonella* candidate order).

The order ‘*Candidatus* Anstonellales’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). It constitutes the most inclusive clade that includes the families ‘*Candidatus* Anstonellaceae,’ and ‘*Candidatus* Bilamarchaeaceae.’ The type genus is ‘*Candidatus* Anstonella.’

The order is equivalent to the lineage LFWA-IIIc in this manuscript, and o\_\_UBA10214 in the GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Micrarchaeaceae’ fam. nov.

‘*Candidatus* Micrarchaeaceae’ (Mi.crar.chae’a’ce.ae. N.L. neut. n. *Micrarchaeum*, a candidate genus; *-aceae*, ending to denote a family; N.L. fem. pl. n. *Micrarchaeaceae*, the *Micrarchaeum* candidate family).

The family ‘*Candidatus* Micrarchaeaceae’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). It is the most inclusive family that includes the genera ‘*Candidatus* Micrarchaeum’ and ‘*Candidatus* Mancarchaeum.’ The type genus is ‘*Candidatus* Micrarchaeum.’

This family is equivalent to f\_\_Micrarchaeaceae in the GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Micrarchaeales’ ord. nov.

‘*Candidatus* Micrarchaeales’ (Mi.crar.chae.a’les. N.L. neut. n. *Micrarchaeum*, a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Micrarchaeales*, the *Micrarchaeum* candidate order).

The order ‘*Candidatus* Micrarchaeales’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation

approach as per Parks et al. (2018). The type genus is ‘*Candidatus* Micrarchaeum.’

The order is equivalent to o\_\_Micrarchaeales in the GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Norongarragalina’ gen. nov.

‘*Candidatus* Norongarragalina’ (No.ron.ga.rra.ga.li’na. Dharawal language *Norongarragal*, Dharawal clan group who traditionally occupied the Menai/Lucas Heights area where the present study took place; L. fem. suff., *-ina*, pertaining or belonging to; N.L. dim. n. fem. *Norongarragalina*, the archaeon from the Norongarragal clan).

The type species is ‘*Candidatus* Norongarragalina meridionalis.’ This genus is equivalent to g\_\_0-14-0-20-59-11 in GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Norongarragalina meridionalis’ sp. nov.

‘*Candidatus* Norongarragalina meridionalis’ (me.ri.dio.na’lis. L. fem. adj. *meridionalis*, southern; referring to the Southern hemisphere, where its first genome was reconstructed).

The type material is the metagenome assembled genome (MAG) LFW-144\_1 (ERS265293) recovered from the groundwater of the Little Forest Legacy Site (NSW, Australia). The MAG consists of 0.93 Mbp in 49 contigs with an estimated completeness of 93.5%, redundancy of 1.1%, 16S and 5S rRNA gene, and 20 tRNAs. The GC content of this MAG is 57.5%.

The type material appears in GTDB r202 (Parks et al., 2017) as reference for s\_\_0-14-0-20-59-11 sp902384935.

## Description of ‘*Candidatus* Norongarragalinaeaceae’ fam. nov.

‘*Candidatus* Norongarragalinaeaceae’ (No.ron.ga.rra.ga.li.na’ce.ae. N.L. neut. n. *Norongarragalina*, a candidate genus; *-aceae*, ending to denote a family; N.L. fem. pl. n. *Norongarragalinaeaceae*, the *Norongarragalina* candidate family).

The family ‘*Candidatus* Norongarragalinaeaceae’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as that of its sole genus and species. The type genus is ‘*Candidatus* Norongarragalina.’

This family is equivalent to f\_\_0-14-0-20-59-11 in the GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus* Norongarragalinales’ ord. nov.

‘*Candidatus* Norongarragalinales’ (No.ron.ga.rra.ga.li.na’les. N.L. neut. n. *Norongarragalina*, a candidate genus; *-ales*, ending to denote an order; N.L. fem. pl. n. *Norongarragalinales*, the *Norongarragalina* candidate order).

The order ‘*Candidatus* Norongarragalinales’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). The description is the same as

that of its sole genus and species. The type genus is ‘*Candidatus Norongarragalina*.’

This order is equivalent to LFWA-II in this manuscript, and o\_\_UBA8480 in GTDB r89/r202 (Parks et al., 2017).

## Description of ‘*Candidatus Micrarchaeia*’ class. nov.

‘*Candidatus Micrarchaeia*’ (Mi.crar.chae’ia. N.L. neut. n. *Micrarchaeum*, a candidate genus; *-ia*, ending to denote a class; N.L. fem. pl. n. *Micrarchaeia*, the *Micrarchaeum* candidate class).

The class ‘*Candidatus Micrarchaeia*’ is circumscribed based on two independent concatenated protein phylogenies of 122 and 93 markers, and supported by the rank normalisation approach as per Parks et al. (2018). This class is defined as the most inclusive class that includes the orders ‘*Candidatus Micrarchaeales*,’ ‘*Candidatus Gugararchaeales*,’ ‘*Candidatus Anstonellales*,’ and ‘*Candidatus Norongarragalinales*.’ The type genus is ‘*Candidatus Micrarchaeum*.’

This class constitutes the only class within the phylum ‘*Candidatus Micrarchaeota*’ Baker and Dick (2013). It is equivalent to the c\_\_*Micrarchaeia* in the GTDB r89/r202 (Parks et al., 2017).

## DATA AVAILABILITY STATEMENT

Annotated assemblies are available at ENA under project PRJEB21808 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB21808>) and sample identifiers ERS2655284–ERS2655320. See details in **Supplementary Table 7**. The output of the different analyses, including the reference genomes, are available at Zenodo (doi: 10.5281/zenodo.3365725).

## AUTHOR CONTRIBUTIONS

XV-C conceptualized the study, carried out the data curation and formal analysis, investigated and visualized the data, performed the methodology, carried out the project administration,

## REFERENCES

- Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* 11, 2407–2425. doi: 10.1038/ismej.2017.122
- Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., Custer, A. F., Hicks, M. A., et al. (2014). The Structure–Function Linkage Database. *Nucleic Acids Res.* 42, D521–D530.
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Meth.* 11, 1144–1146.
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- André, I., Potocki-Véronèse, G., Barbe, S., Moulis, C., and Remaud-Siméon, M. (2014). CAZyme discovery and design for sweet dreams. *Curr. Opin. Chem. Biol.* 19, 17–24. doi: 10.1016/j.cbpa.2013.11.014

wrote the original draft, and wrote, reviewed, and edited the manuscript. ASK wrote the original draft and wrote, reviewed, and edited the manuscript. MWB wrote, reviewed, and edited the manuscript. MRW carried out the resource procurement, supervised the study, and wrote, reviewed, and edited the manuscript. TEP carried out the resource procurement, funding acquisition, and project administration and provided the site-specific expertise. TDW carried out the resource procurement, funding acquisition, and project administration and supervision, and wrote, reviewed, and edited the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

MRW acknowledges the support of the Australian Federal Government NCRIS scheme, via Bioplatforms Australia. MRW and XV-C acknowledge support from the New South Wales State Government RAAP scheme and the UNSW RIS scheme. ASK, TEP, and TDW acknowledge funding by the Australian Research Council (Discovery Project DP210103727) as well as from the Australian Nuclear Science and Technology Organisation (ANSTO).

## ACKNOWLEDGMENTS

We acknowledge Shane Ingrey and Raymond Ingrey from the Dharawal Language Program from the Gujaga Foundation for their assistance with the Dharawal language and with their views on aboriginal matters in general. We thank Maria Chuvochina for her comments on nomenclature.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.732575/full#supplementary-material>

- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859
- Arshad, A., Speth, D. R., de Graaf, R. M., Op, den Camp, H. J. M., Jetten, M. S. M., et al. (2015). A metagenomics-based metabolic model of nitrate-dependent anaerobic oxidation of methane by *Methanoperedens*-like Archaea. *Front. Microbiol.* 6:1423. doi: 10.3389/fmicb.2015.01423
- Attwood, T. K., Beck, M. E., Bleasby, A. J., and Parry-Smith, D. J. (1994). PRINTS—a database of protein motif fingerprints. *Nucleic Acids Res.* 22, 3590–3596.
- Baker, B. J., and Dick, G. J. (2013). Omic Approaches in Microbial Ecology: Charting the Unknown: Analysis of whole-community sequence data is unveiling the diversity and function of specific microbial groups within uncultured phyla and across entire microbial ecosystems. *Microbe* 8, 353–360.
- Baker, B. J., Tyson, G. W., Webb, R. I., Flanagan, J., Hugenholtz, P., Allen, E. E., et al. (2006). Lineages of acidophilic Archaea revealed by community genomic analysis. *Science* 314, 1933–1935.
- Buelow, H. N., Winter, A. S., Van Horn, D. J., Barrett, J. E., Gooseff, M. N., Schwartz, E., et al. (2016). Microbial Community Responses to Increased Water

- and Organic Matter in the Arid Soils of the McMurdo Dry Valleys. *Antarctica. Front. Microbiol.* 2016:7. doi: 10.3389/fmicb.2016.01040
- Cai, C., Leu, A. O., Xie, G.-J., Guo, J., Feng, Y., Zhao, J.-X., et al. (2018). A methanotrophic archaeon couples anaerobic oxidation of methane to Fe(III) reduction. *ISME J.* 12, 1929–1939.
- Cai, C., Shi, Y., Guo, J., Tyson, G. W., Hu, S., and Yuan, Z. (2019). Acetate production from anaerobic oxidation of methane via intracellular storage compounds. *Environ. Sci. Technol.* 53, 7371–7379. doi: 10.1021/acs.est.9b00077
- Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701.
- Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2019). tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *bioRxiv* 2019:614032.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/bt2848
- Chen, L.-X., Méndez-García, C., Dombrowski, N., Servín-Garcidueñas, L. E., Eloe-Fadros, E. A., Fang, B.-Z., et al. (2018). Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J.* 12, 756–775. doi: 10.1038/s41396-017-0002-z
- Chuvpochina, M., Rinke, C., Parks, D. H., Rappé, M. S., Tyson, G. W., Yilmaz, P., et al. (2019). The importance of designating type material for uncultured taxa. *Syst. Appl. Microbiol.* 42, 15–21. doi: 10.1016/j.syapm.2018.07.003
- Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210
- Darling, A. E., Jospin, G., Lowe, E., Iv, F. A. M., Bik, H. M., and Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. doi: 10.7717/peerj.243
- Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P., and Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* 366:fnz008. doi: 10.1093/femsle/fnz008
- 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. doi: 10.1186/s40168-017-0322-2
- Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J.-H., Minh, B. Q., et al. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat Commun* 11, 3939. doi: 10.1038/s41467-020-17408-w
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195.
- Eren, A. M., Esen, Ö.C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319. doi: 10.7717/peerj.1319
- Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* 6, 3–6. doi: 10.1038/s41564-020-00834-3
- Ettwig, K. F., Zhu, B., Speth, D., Keltjens, J. T., Jetten, M. S. M., and Kartal, B. (2016). Archaea catalyze iron-dependent anaerobic oxidation of methane. *PNAS* 113, 12792–12796.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230.
- Golyshina, O. V., Toshchakov, S. V., Makarova, K. S., Gavrillov, S. N., Korzhnikov, A. A., Cono, V. L., et al. (2017). 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. *Nat. Commun.* 8:60. doi: 10.1038/s41467-017-00104-7
- Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903–919. doi: 10.1006/jmbi.2001.5080
- Grosjean, H., and Westhof, E. (2016). An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* 44, 8020–8040. doi: 10.1093/nar/gkw608
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res.* 41, D387–D395. doi: 10.1093/nar/gks1234
- Haroon, M. F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., et al. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 500, 567–570.
- He, L., Zhong, Y., Yao, F., Chen, F., Xie, T., Wu, B., et al. (2019). Biological perchlorate reduction: which electron donor we can choose? *Environ. Sci. Pollut. Res.* 26, 16906–16922. doi: 10.1007/s11356-019-05074-5
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67. doi: 10.1038/417063a
- Huber, H., and Kreuter, L. (2014). "The Phylum Nanoarchaeota," in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), 311–318. doi: 10.1007/978-3-642-38954-2\_337
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 2016:16048.
- Ikeda-Ohno, A., Harrison, J. J., Thiruvoth, S., Wilsher, K., Wong, H. K. Y., Johansen, M. P., et al. (2014). Solution speciation of plutonium and americium at an Australian legacy radioactive waste disposal site. *Environ. Sci. Technol.* 48, 10045–10053. doi: 10.1021/es500539t
- John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., et al. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* 3, 195–202. doi: 10.1111/j.1758-2229.2010.00208.x
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kadnikov, V. V., Savvichev, A. S., Mardanov, A. V., Beletsky, A. V., Chupakov, A. V., Kokryatskaya, N. M., et al. (2020). Metabolic Diversity and Evolutionary History of the Archaeal Phylum "Candidatus Micrarchaeota" Uncovered from a Freshwater Lake Metagenome. *Appl. Environ. Microbiol.* 86, e2199–e2120. doi: 10.1128/AEM.02199-20
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462.
- Karp, P. D., Latendresse, M., Paley, S. M., Krummenacker, M., Ong, Q. D., Billington, R., et al. (2016). Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform.* 17, 877–890. doi: 10.1093/bib/bbv079
- Karthikeyan, O. P., Chidambarampadmavathy, K., Cirés, S., and Heimann, K. (2015). Review of sustainable methane mitigation and biopolymer production. *Crit. Rev. Environ. Sci. Technol.* 45, 1579–1610.
- Kinsela, A. S., Bligh, M., Vázquez-Campos, X., Sun, Y., Wilkins, M. R., Comarmond, M. J., et al. (2021). Biogeochemical Mobility of Contaminants from a Replica Radioactive Waste Trench in Response to Rainfall-induced Redox Oscillations. *Environ. Sci. Technol.* 55, 8793–8805. doi: 10.1021/acs.est.1c01604
- Kitzinger, K., Padilla, C. C., Marchant, H. K., Hach, P. F., Herbold, C. W., Kidane, A. T., et al. (2019). Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat. Microbiol.* 4:234. doi: 10.1038/s41564-018-0316-2
- Knittel, K., and Boetius, A. (2009). Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* 63, 311–334. doi: 10.1146/annurev.micro.61.080706.093130

- Koller, M. (2019). Polyhydroxyalkanoate biosynthesis at the edge of water activity—Haloarchaea as biopolyester factories. *Bioengineering* 6:34. doi: 10.3390/bioengineering6020034
- Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11:11.
- Konstantinidis, K. T., Rossello-Mora, R., and Amann, R. (2019). Moving the cataloguing of the “uncultivated majority” forward. *Syst. Appl. Microbiol.* 42, 3–4. doi: 10.1016/j.syapm.2018.12.001
- Lazar, C. S., Baker, B. J., Seitz, K. W., and Teske, A. P. (2017). Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISME J.* 11, 1118–1129.
- Le, S. Q., Gascuel, O., and Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24, 2317–2323.
- Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164.
- Lees, J. G., Lee, D., Studer, R. A., Dawson, N. L., Sillitoe, I., Das, S., et al. (2014). Gene3D: Multi-domain annotations for protein sequence and comparative genome analysis. *Nucleic Acids Res.* 42, D240–D245.
- Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949
- Leu, A. O., Cai, C., McIlroy, S. J., Southam, G., Orphan, V. J., Yuan, Z., et al. (2020). Anaerobic methane oxidation coupled to manganese reduction by members of the Methanoperedenaceae. *ISME J.* 14, 1030–1041.
- 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. doi: 10.1093/bioinformatics/btv033
- Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., et al. (2016). MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
- Lledó, B., Martínez-Espinosa, R. M., Marhuenda-Egea, F. C., and Bonete, M. J. (2004). Respiratory nitrate reductase from haloarchaeon *Haloferax mediterranei*: biochemical and genetic analysis. *Biochim. Biophys. Acta Gen. Subj.* 1674, 50–59. doi: 10.1016/j.bbagen.2004.05.007
- Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., et al. (2013). Predominant archaea in marine sediments degrade detrital proteins. *Nature* 496, 215–218. doi: 10.1038/nature12033
- Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jørgensen, B. B., and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 484, 101–104. doi: 10.1038/nature10905
- Lu, Y.-Z., Fu, L., Ding, J., Ding, Z.-W., Li, N., and Zeng, R. J. (2016). Cr(VI) reduction coupled with anaerobic oxidation of methane in a laboratory reactor. *Water Res.* 102, 445–452. doi: 10.1016/j.watres.2016.06.065
- Luef, B., Frischkorn, K. R., Wrighton, K. C., Holman, H.-Y. N., Birarda, G., Thomas, B. C., et al. (2015). Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* 6:6372. doi: 10.1038/ncomms7372
- Luo, Y.-H., Chen, R., Wen, L.-L., Meng, F., Zhang, Y., Lai, C.-Y., et al. (2015). Complete perchlorate reduction using methane as the sole electron donor and carbon source. *Environ. Sci. Technol.* 49, 2341–2349.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203.
- Martínez-Gutiérrez, C. A., Latisnere-Barragán, H., García-Maldonado, J. Q., and López-Cortés, A. (2018). Screening of polyhydroxyalkanoate-producing bacteria and PhaC-encoding genes in two hypersaline microbial mats from Guerrero Negro, Baja California Sur, Mexico. *PeerJ.* 6:e4780. doi: 10.7717/peerj.4780
- Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D377–D386. doi: 10.1093/nar/gks1118
- Mills, H. J., Hodges, C., Wilson, K., MacDonald, I. R., and Sobczyk, P. A. (2003). Microbial diversity in sediments associated with surface-breaching gas hydrate mounds in the Gulf of Mexico. *FEMS Microbiol. Ecol.* 46, 39–52. doi: 10.1016/S0168-6496(03)00191-0
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195.
- Murray, R. G. E., and Schleifer, K. H. (1994). Taxonomic Notes: A Proposal for Recording the Properties of Putative Taxa of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 44, 174–176. doi: 10.1099/00207713-44-1-174
- Murray, R. G. E., and Stackebrandt, E. Y. (1995). Taxonomic Note: Implementation of the Provisional Status Candidatus for Incompletely Described Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 45, 186–187. doi: 10.1099/00207713-45-1-186
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132. doi: 10.1186/s13059-016-0997-x
- Oren, A., Elvi Bardavid, R., and Mana, L. (2014). Perchlorate and halophilic prokaryotes: implications for possible halophilic life on Mars. *Extremophiles* 18, 75–80. doi: 10.1007/s00792-013-0594-9
- Oren, A., and Garrity, G. M. (2018). Uncultivated microbes—in need of their own nomenclature? *ISME J.* 12, 309–311.
- Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotech.* 2020, 1–8.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- 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. doi: 10.1101/gr.186072.114
- 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, 1533–1542.
- Payne, T. E. (2012). *Background report on the Little Forest Burial Ground legacy waste site*. Lucas Heights: Institute for Environmental Research, Australian Nuclear Science and Technology Organisation.
- Payne, T. E., Harrison, J. J., Cendon, D. I., Comarmond, M. J., Hankin, S., Hughes, C. E., et al. (2020). Radionuclide distributions and migration pathways at a legacy trench disposal site. *J. Environ. Radioact.* 211:106081. doi: 10.1016/j.jenvrad.2019.106081
- Payne, T. E., Harrison, J. J., Hughes, C. E., Johansen, M. P., Thiruvoth, S., Wilsher, K. L., et al. (2013). Trench ‘bathtubbing’ and surface plutonium contamination at a legacy radioactive waste site. *Environ. Sci. Technol.* 47, 13284–13293. doi: 10.1021/es403278r
- Pedrucci, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., et al. (2015). HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res.* 43, D1064–D1070. doi: 10.1093/nar/gku1002
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2015). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24.
- Probst, A. J., Ladd, B., Jarett, J. K., Geller-McGrath, D. E., Sieber, C. M. K., Emerson, J. B., et al. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* 3, 328–336. doi: 10.1038/s41564-017-0098-y
- Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–D509.
- Reddy, C. S. K., Ghai, R., Rashmi, and Kalia, V. C. (2003). Polyhydroxyalkanoates: an overview. *Bioresour. Technol.* 87, 137–146.
- Reeburgh, W. S. (2007). Oceanic methane biogeochemistry. *Chem. Rev.* 107, 486–513.

- Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Waite, D. W., Whitman, W. B., et al. (2020). A rank-normalized archaeal taxonomy based on genome phylogeny resolves widespread incomplete and uneven classifications. *bioRxiv* 2020:972265.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352
- Rohart, F., Gautier, B., Singh, A., and Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752
- Rosselló-Móra, R., and Whitman, W. B. (2019). Dialogue on the nomenclature and classification of prokaryotes. *Syst. Appl. Microbiol.* 42, 5–14. doi: 10.1016/j.syapm.2018.07.002
- Saier, M. H., Reddy, V. S., Tamang, D. G., and Västermark, Å (2014). The Transporter Classification Database. *Nucleic Acids Res.* 42, D251–D258.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069.
- Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D., et al. (2002). ProDom: Automated clustering of homologous domains. *Brief Bioinform.* 3, 246–251.
- Shen, H., and Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* 99, 1015–1034.
- Sigrist, C. J. A., Castro, E., de, Cerutti, L., Cuche, B. A., Hulo, N., et al. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347.
- Sivan, O., Antler, G., Turchyn, A. V., Marlow, J. J., and Orphan, V. J. (2014). Iron oxides stimulate sulfate-driven anaerobic methane oxidation in seeps. *PNAS* 111, E4139–E4147. doi: 10.1073/pnas.1412269111
- Søndergaard, D., Pedersen, C. N. S., and Greening, C. (2016). HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* 6:34212. doi: 10.1038/srep34212
- Song, G. C., Im, H., Jung, J., Lee, S., Jung, M.-Y., Rhee, S.-K., et al. (2019). Plant growth-promoting archaea trigger induced systemic resistance in *Arabidopsis thaliana* against *Pectobacterium carotovorum* and *Pseudomonas syringae*. *Environ. Microbiol.* 21, 940–948. doi: 10.1111/1462-2920.14486
- Spang, A., Poehlein, A., Offre, P., Zumbärgel, S., Haider, S., Rychlik, N., et al. (2012). The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ. Microbiol.* 14, 3122–3145. doi: 10.1111/j.1462-2920.2012.02893.x
- Tang, X., Zheng, H., Teng, H., Sun, Y., Guo, J., Xie, W., et al. (2016). Chemical coagulation process for the removal of heavy metals from water: a review. *Desalin. Water Treat* 57, 1733–1748.
- Vaksmas, A., Guerrero-Cruz, S., van Aalen, T. A., Cremers, G., Ettwig, K. F., Lütke, C., et al. (2017). Enrichment of anaerobic nitrate-dependent methanotrophic 'Candidatus Methanoperedens nitroreducens' archaea from an Italian paddy field soil. *Appl. Microbiol. Biotechnol.* 101, 7075–7084. doi: 10.1007/s00253-017-8416-0
- van Dongen, S., and Abreu-Goodger, C. (2012). "Using MCL to Extract Clusters from Networks," in *Bacterial Molecular Networks: Methods and Protocols* Methods in Molecular Biology, eds J. van Helden, A. Toussaint, and D. Thieffry (New York, NY: Springer), 281–295. doi: 10.1007/978-1-61779-361-5\_15
- Vázquez-Campos, X., Kinsela, A. S., Bligh, M., Harrison, J. J., Payne, T. E., and Waite, T. D. (2017). Response of microbial community function to fluctuating geochemical conditions within a legacy radioactive waste trench environment. *Appl. Environ. Microbiol.* 83:e00729. doi: 10.1128/AEM.00729-17
- Wang, H. C., Minh, B. Q., Susko, S., and Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235. doi: 10.1093/sysbio/syx068
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *PNAS* 100, 12984–12988. doi: 10.1073/pnas.1735403100
- Whitman, W. B. (2016). Modest proposals to expand the type material for naming of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 66, 2108–2112. doi: 10.1099/ijsem.0.000980
- Wu, C. H., Nikolskaya, A., Huang, H., Yeh, L.-S. L., Natale, D. A., Vinayaka, C. R., et al. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.* 32, D112–D114. doi: 10.1093/nar/gkh097
- Wu, M., Luo, J.-H., Hu, S., Yuan, Z., and Guo, J. (2019). Perchlorate bio-reduction in a methane-based membrane biofilm reactor in the presence and absence of oxygen. *Water Res.* 157, 572–578. doi: 10.1016/j.watres.2019.04.008
- Xie, T., Yang, Q., Winkler, M. K. H., Wang, D., Zhong, Y., An, H., et al. (2018). Perchlorate bioreduction linked to methane oxidation in a membrane biofilm reactor: Performance and microbial community structure. *J. Hazard Mater.* 357, 244–252. doi: 10.1016/j.jhazmat.2018.06.011
- Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., et al. (2021). ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Mole. Biol. Evol.* 38, 4039–4042. doi: 10.1093/molbev/msab166
- Yamada, K. D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251. doi: 10.1093/bioinformatics/btw412
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451.
- Yoshimatsu, K., Sakurai, T., and Fujiwara, T. (2000). Purification and characterization of dissimilatory nitrate reductase from a denitrifying halophilic archaeon. *Haloarcula marismortui*. *FEBS Lett.* 470, 216–220. doi: 10.1016/S0014-5793(00)01321-1
- Youssef, N. H., Rinke, C., Stepanauskas, R., Farag, I., Woyke, T., and Elshahed, M. S. (2015). Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. *ISME J.* 9, 447–460. doi: 10.1038/ismej.2014.141
- Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinformatics* 69:e96. doi: 10.1002/cpbi.96
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249
- Yutin, N., Puigbò, P., Koonin, E. V., and Wolf, Y. I. (2012). Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* 7:e36972.
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358.
- Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19:153. doi: 10.1186/s12859-018-2129-y

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Vázquez-Campos, Kinsela, Bligh, Payne, Wilkins and Waite. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.