Stratified microbial communities in Australia's only anchialine cave are taxonomically novel and drive chemotrophic energy production via coupled nitrogen-sulphur cycling

Timothy M. Ghaly[1†], Amaranta Focardi[2†], Liam D. H. Elbourne[1,3], Brodie Sutcliffe[1], William Humphreys[4], Ian T. Paulsen[1,3]*, and Sasha G. Tetu[1,3]*

[1]School of Natural Sciences, Macquarie University, Sydney, Australia

[2]Climate Change Cluster (C3), University of Technology Sydney, Sydney, Australia

[3]ARC Centre of Excellence in Synthetic Biology, Macquarie University, Sydney, Australia

[4]School of Biological Sciences, University of Western Australia, Perth, Australia

[†] These authors contributed equally to this work

* Corresponding Authors:

Sasha G. Tetu, email: sasha.tetu@mq.edu.au

Ian T. Paulsen, email: ian.paulsen@mq.edu.au

The authors declare no competing interests

Keywords:

Chemolithotrophy, metabolic coupling, biogeochemical cycling, stratified water column, groundwater ecology, subterranean estuary, marine oxygen minimum zones

Abstract

Background

Anchialine environments, in which oceanic water mixes with freshwater in coastal aquifers, are characterised by stratified water columns with complex physicochemical profiles. These environments, also known as subterranean estuaries, support an abundance of endemic macro and microorganisms. There is now growing interest in characterising the metabolisms of anchialine microbial communities, which is essential for understanding how complex ecosystems are supported in extreme environments, and assessing their vulnerability to environmental change. However, the diversity of metabolic strategies that are utilised in anchialine ecosystems remains poorly understood.

Results

Here, we employ shotgun metagenomics to elucidate the key microorganisms and their dominant metabolisms along a physicochemical profile in Bundera Sinkhole, the only known continental subterranean estuary in the Southern Hemisphere. Genome-resolved metagenomics suggests that the communities are largely represented by novel taxonomic lineages, with 75% of metagenome-assembled genomes assigned to entirely new or uncharacterised families. These diverse and novel taxa displayed depth-dependent metabolisms, reflecting distinct phases along dissolved oxygen and salinity gradients. In particular, the communities appear to drive nutrient feedback loops involving nitrification, nitrate ammonification, and sulphate cycling. Genomic analysis of the most highly abundant members in this system suggests that an important source of chemotrophic energy is generated via the metabolic coupling of nitrogen and sulphur cycling.

Conclusion

These findings substantially contribute to our understanding of the novel and specialised microbial communities in anchialine ecosystems, and highlight key chemosynthetic pathways

48 that appear to be important in these energy-limited environments. Such knowledge is
49 essential for the conservation of anchialine ecosystems, and sheds light on adaptive processes
50 in extreme environments.

Introduction

The microbial communities of stratified aquatic systems serve as useful models for studying the relationships between metabolic strategies, water column depth, and physicochemistry. Stratified water columns, characterised by physical and chemical gradients, provide distinct niches for diverse assemblages of microbes, which, in turn, can support complex food webs in relatively extreme environments. Thus, unravelling the network of microbial metabolic strategies that link biogeochemical processes and trophic webs is important for understanding ecosystem functioning as well as evaluating ecosystem vulnerability [1].

Subterranean estuaries are stratified aquatic systems in which marine-derived groundwater mixes with meteoric freshwater in coastal aquifers [2]. These systems are globally distributed, and most commonly form in the porous limestone of karst coastlines [3]. They are characterised by water columns that exhibit stratified physicochemical profiles and low dissolved oxygen content [4]. Although they represent low-energy and extreme environments, subterranean estuaries can support complex ecosystems, which have been termed ‘anchialine’ [4]. The higher trophic levels of anchialine ecosystems largely comprise cave-adapted invertebrates with high rates of endemism [5, 6]. Earlier investigations into these anchialine food webs indicated that they may be supported, at least in part, by chemosynthetic microbes [7-9]. There is now growing interest in surveying the microbial communities that inhabit subterranean estuaries, and in particular, characterising their niche-adaptive metabolisms [1, 10]. Such endeavours are critical for assessing the vulnerability of anchialine ecosystems to environmental change.

Microbial ecology studies have revealed that anchialine ecosystems harbour highly diverse microbial assemblages. Examination of the prokaryotic community structure using 16S rRNA gene amplicon sequencing has been undertaken for several anchialine systems,

including those found in Eastern Adriatic Sea Islands [11], Sansha Yongle Blue Hole in the South China Sea [12], Indonesian anchialine lakes [13], Blackwood Sinkhole in the Bahamas [14], and coastal aquifers of the Yucatán Peninsula, Mexico [10, 15]. These sites all revealed a high degree of taxonomic richness spanning functionally diverse microbial groups. Brankovits D, *et al.* [10] combined 16S rRNA gene sequencing with respiratory quinone biomarker analysis to infer the metabolic phenotypes of an anchialine water column, which contained a mixture of methanotrophs, heterotrophs, photoautotrophs, and nitrogen and sulphur cycling chemolithotrophs. They identified methane and dissolved organic carbon as key microbial energy sources that support higher trophic levels of the anchialine food web. Though, comparison between the microbial communities within coastal and in-land sinkholes of the same region (Yucatán Peninsula) show that the dominant metabolic strategies can differ significantly between different sinkholes along the same aquifer network [15].

Bundera Sinkhole, located in the karstic coast of Cape Range Peninsula in north-western Australia, is the only known continental anchialine system in the Southern Hemisphere. The sinkhole, which is the only opening to the subterranean estuary, is located 1.7 km inland from the Indian Ocean. The water column exhibits strong vertical stratification in its physicochemical profile, with decreasing dissolved oxygen and increasing salinity with depth, and polymodal peaks of inorganic nitrogen and sulphur compounds [16-18]. A range of endemic eukaryotes have been discovered in Bundera Sinkhole, including copepods, remipeds, and polychaetes [19-22]. Chemical profiling suggests that this trophic web may be supported by microbial chemosynthesis [16].

Microbial studies of Bundera Sinkhole using flow cytometry and 16S rRNA gene sequencing have shown the microbial communities to be stratified along the depth profile [17, 18, 23]. A diverse range of prokaryotes have been identified in the water column, comprising 67 identifiable bacterial and archaeal phyla [18]. Although community profiling

suggests that a range of chemolithotrophic metabolisms are present throughout the water column, the high level of taxonomic novelty has made it difficult to infer the metabolic functions of many of the most abundant members [18]. Here, we employed shotgun metagenomic sequencing across a depth profile in Bundera Sinkhole to elucidate the metabolisms of these novel microbial communities. We identified key depth-dependent chemotrophic metabolic pathways, including coupled nitrogen-sulphur cycling, that may be driving nutrient feedback loops in this system. To the best of our knowledge, this is the first whole metagenomic sequencing approach of any anchialine ecosystem, and represents important findings that can help us to better understand microbial metabolic and biogeochemical processes in these unique environments.

Methods

Sample collection, DNA extraction and sequencing

Water samples were collected from Bundera Sinkhole as previously described [18]. Briefly, this involved pumping water samples from depths of 2, 8, 17, 18, 22, and 28 m between the 29th of June and the 1st of July 2015 for metagenomic analysis. For depths of 8 m and below, samples were collected using four previously installed boreholes (Fig. 1b). Physicochemical data, including salinity, dissolved oxygen (DO), dissolved organic carbon (DOC), ammonia (NH_3), nitrate (NO_3^-), and sulphate (SO_4^{2-}) measurements were obtained from our previous study [18]. For metagenomic analysis, ~4 L water samples were pre-filtered using 60 µm filters (Millipore Type NY60), and then passed through 0.2 µm Sterivex™ filters. The 0.2 µm filters with captured microbial cells were cut from their casing, and DNA extractions carried out using the PowerWater® DNA Isolation kit (MO BIO Laboratories, Inc., Carlsbad, USA),

according to the manufacturer's protocol. Metagenomic libraries were prepared for duplicate biological replicates from each depth using the Illumina TruSeq DNA Library Preparation Kit, according to the manufacturer's protocol, and sequenced on the Illumina HiSeq 2000 platform (High-Output v4).

Metagenomic assembly and functional annotation

Raw reads were trimmed and quality filtered using Trimmomatic v 0.38 [24], and assembled with metaSPAdes v 3.13.0 [25] with default parameters. Quality of the assembly for each sample was assessed with QUAST v 5.0.2 using the metaQUAST option [26], and contigs shorter than 1 kb were removed from the assemblies. Open reading frames (ORFs) and translated protein sequences were predicted using Prodigal v2.6.3 [27] in metagenomic mode [parameter: -p meta]. ORFs from all samples were pooled and dereplicated at 98% nucleotide identity using CD-HIT v4.8.1 [28, 29] [parameters: -c 0.98 -n 10 -d 0 -t 0 -M 0]. The relative abundance of ORFs in each sample was calculated using the transcripts per million (TPM) method with CoverM v0.6.1 (https://github.com/wwood/CoverM) in contig mode [parameters: contig -t 24 --coupled -m TPM].

Translated protein sequences of the dereplicated ORFs were functionally annotated using METABOLIC v4.0 [30], by implementing the METABOLIC-G workflow with default parameters. The METABOLIC software identifies metabolic and biogeochemical traits by integrating several hidden Markov model (HMM) databases, comprising KOfam [31] (containing KEGG HMMs [32]), TIGRfam [33], Pfam [34], and custom [35] HMM databases.

MAG binning and quality control

To improve sequencing depth, the replicate metagenome samples were co-assembled using MEGAHIT v1.2.9 [36, 37], with contig coverage calculated using Bowtie 2 v2.3.2 [38]. Co-assembled contigs were then binned using METABAT 2 v2.2.15 [39] with default parameters within Anvi'o v6.2 [40]. The resulting MAGs were then manually refined in Anvi'o. The completion and contamination of MAGs were estimated with CheckM v1.2.1 [41] using lineage-specific marker sets [parameters: lineage_wf -t 24]. MAG chimerism was assessed using GUNC v1.0.5 [42] with default parameters. Only MAGs that passed the GUNC chimerism check, had an estimated completion greater than 50%, and had an estimated contamination less than 10% were retained for further analysis. These represent the completion and contamination MIMAG criteria for high- and medium-quality MAGs [43].

MAGs taxonomy and functional annotation

MAG taxonomy was assigned using GTDB-Tk v2.1.1 [44, 45] [parameters: classify_wf --cpus 24] with release R207_v2 of the Genome Taxonomy Database (GTDB) [46-49]. We inferred domain-specific phylogenies using concatenated protein alignments generated by GTDB-Tk, which were based on the BAC120 [50] and AR53 [51] protein marker sets. The phylogenies were inferred from the alignments using a maximum-likelihood approximation employed by FastTree v2.1.10 [52, 53]. We applied a WAG substitution model with branch lengths rescaled to optimise the Gamma20 likelihood, and 1,000 resamples [parameters: -gamma -wag]. The inferred phylogenies were visualised using the ggtree v2.4.2 [54] and ggtreeExtra v1.7.0.990 [55] R packages.

MAGs were functionally annotated using METABOLIC v4.0 [30], by implementing the METABOLIC-C workflow with default parameters. The relative abundance of MAGs in

each sample was calculated using the TPM method with CoverM v0.6.1 (https://github.com/wwood/CoverM) in genome mode [parameters: genome -t 24 --coupled -m TPM]. Four MAGs that were highly abundant, having TPM values greater than 50 in at least one sample, were further profiled for nitrogen cycling genes using the NCycDB [56]. DIAMOND v2.0.15 [57] was used to query MAG proteins against the NCycDB with a minimum E-value of 1e-05 [parameters: blastp -p 8 -k 1 -e 1e-5], and filtered using an amino acid identity cut-off of 70%.

Statistical analyses

Beta-diversity analyses of the whole metagenomes, key metabolic genes, and MAG phyla were assessed using non-metric multidimensional scaling (NMDS) based on Bray-Curtis distances using the *vegdist* and *metaMDS* functions from the vegan v2.5-7 R package [58]. Groupings inferred from the NMDS ordination were compared with PERMANOVA using the pairwiseAdonis v0.4 R package [59], which uses the vegan functions, *vegdist* and *adonis*, to calculate inter-group differences in a pairwise fashion.

Results and Discussion

Bundera Sinkhole, Australia's only deep water anchialine system, supports a complex trophic web with an abundance of endemic micro- and macroorganisms. Previous chemical and community profiling using 16S rRNA gene sequencing suggest that this ecosystem may be sustained by microbial chemosynthesis [18, 23]. However, the high degree of taxonomic novelty, with associated uncertainty of metabolic functions, has limited our understanding of the dominant metabolic pathways in this system. Here, we employed shotgun metagenomic

sequencing to investigate the distribution of key metabolic genes and to identify the biogeochemical cycling potential of the stratified microbial communities in Bundera Sinkhole.

Microbial metabolic profiles are associated with water depth and physicochemistry

Bundera sinkhole exhibited a highly stratified water column with a marked physicochemical profile (Supplementary Table 1). The only oxic depth sampled was at 2 m, which had a dissolved oxygen (DO) concentration of 2.75 mg/L, and had the lowest salt concentration, with a salinity of 18.69 practical salinity units (PSS). The 8 m depth, representing the sinkhole's halocline [16, 17], had a DO (0.86 mg/L) relatively higher than the samples from 17-28 m depths, and an intermediate salinity of 25.46 PSS. The lower depths, encompassing the 17-28 m samples, had lower levels of DO (0.28-0.47 mg/L) and higher salinity (31.41-32.35 PSS). Polymodal peaks of dissolved organic carbon (DOC), ammonia, nitrate and sulphate were observed along the water column (Supplementary Table 1).

Clear distinctions in microbial metabolic strategies were observed at different depths (Fig. 2). Microbial communities sampled from the 17, 18, 22, and 28 m depths exhibited similar metabolic gene diversity profiles, which differed from the 2 m and 8 m communities (Fig. 2b; PERMANOVA, P=0.04). The 2 m and 8 m metabolic profiles form distinct clusters based on NMDS analysis (Fig. 2b), although this separation was not determined to be significantly different (PERMANOVA, p=0.33), likely due to the limited statistical power of this comparison. The same clustering is observed for the beta-diversity of all genes detected in the metagenomes (Fig. 2a). Since genes were de-replicated at 98% nucleotide identity, clustering of all genes is more likely to reflect the taxonomic composition of the samples. Thus, both taxonomic and functional composition of the sinkhole appear to cluster according

to salinity and oxygen concentrations. These same depth clusters are observed from16S rDNA amplicon sequencing of the sinkhole [18].

Autotrophic CO_2 fixation strategies differed by depth (Fig. 2c), likely in response to oxygen levels and percentage of incident light. The CBB cycle, which utilises the CO_2 fixation enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) by photo- and chemo-autotrophs, was depth-dependent. Two main forms of RuBisCO are known to be involved in the classical CBB cycle [60]. Surface samples, particularly those from the 2 m depth, were characterised by a greater relative abundance of the Form I RuBisCO compared to other depths (and other C-fixation strategies), presumably from a greater abundance of photoautotrophs. While the relative abundance of form II RuBisCO, which is adapted to low-O_2 conditions [61], had an opposite trend, with greater relative abundance at lower depths. The relative abundance of genes that drive the reverse TCA cycle and Wood-Ljungdahl pathway increased with depth, which are the hypoxic regions of this system (Supplementary Table 1). Similar trends have been observed in hypoxic and anoxic zones of stratified water columns [62, 63].

The relative abundance of marker genes for different pathways involved in carbon metabolism also corresponded to a depth gradient (Fig. 2c). Methanol and formaldehyde oxidation (C1 metabolism), decreased with depth, while the methane monooxygenase gene, *mmoB*, involved in the first step of methane metabolism, increased with depth. Similar patterns of carbon metabolism genes have been observed over an oxygen gradient in a permanently stratified lake [63]. Arsenic and selenium cycling genes also corresponded to a depth gradient (Fig. 2c). In particular, the abundance of genes involved in dissimilatory (respiratory) arsenate and selenate reduction increased with depth. Both arsenate and selenate can be utilised in anerobic respiration for energy production [64, 65], explaining their greater

relative abundance at hypoxic depths. These elements can thus provide additional energy sources for facultative or obligate anaerobes at the lower depths of the sinkhole.

Pathways for the complete cycling of nitrogen (N) and sulphur (S) compounds were observed in the sinkhole (Fig. 2d), with diverse N and S cycling reactions present at different depths (Fig. 2c). Several key N and S cycling genes were strongly correlated with concentrations of ammonia, nitrate, and sulphate (Fig. 3; Supplementary Table 3), highlighting these as key environmental parameters. To infer the direction of these correlations and to identify nutrient feedback loops, we examined whether the correlated genes were involved in either the production or substrate utilisation of these chemical compounds. Marker genes for N cycling that correlated with ammonia concentrations were all involved in pathways that produced ammonia (Fig. 3a-c; Supplementary Table 3). These included: *napA*, encoding a nitrate reductase, involved in the first step of the dissimilatory nitrate reduction to ammonia (DNRA) pathway, reducing nitrate to nitrite; and *nirB* and *nrfA*, both encoding nitrite reductases, involved in the second step of the DNRA pathway, further reducing nitrite to ammonia. Similarly, N cycling marker genes that correlated with nitrate concentrations were all involved in nitrate production (Fig. 3d,e; Supplementary Table 3). These included *amoA*, encoding an ammonia monooxygenase, involved in the first step of nitrification, oxidising ammonia to nitrite; and *nxrA*, encoding a nitrite oxidoreductase, involved in the final step of nitrification, oxidising nitrite to nitrate. We also found that the relative abundance of both *amoA* and *nxrA* are negatively correlated with the concentration of dissolved organic carbon (DOC) (Fig. S1; Supplementary Table 4), suggesting that chemolithotrophic nitrification is an important metabolic pathway when available organic carbon is limited. Thus, microbial communities and environmental concentrations of DOC, ammonia and nitrate are apparently linked in a feedback loop involving nitrification (ammonia to nitrate) and DNRA (nitrate to ammonia) pathways.

The S cycling marker genes, *sat* and *sdo,* were significantly correlated with sulphate concentrations (Fig. 3f,g; Supplementary Table 3), and are involved in the utilisation and production of sulphate, respectively. *sat* encodes a sulphate adenylyltransferase that coverts sulphate to adenosine-5′-phosphosulfate (APS) [66]. *sdo* encodes a sulphur deoxygenase which oxidises glutathione persulphide (GSSH). Sulphite is the first product of SDO activity via GSSH oxidation, which then leads to the non-enzymatic production of sulphate (likely from auto-oxidation of sulphite) [67]. Thus, sulphate concentrations in Bundera Sinkhole are likely being driven by, as well as shaping, the microbial communities in a sulphate-feedback loop.

Taxonomically novel and functionally diverse prokaryotes inhabit the sinkhole

Bundera Sinkhole harbours considerable microbial diversity, so to gain better insight into the metabolic potential of the novel and abundant microbial species, we employed genome-resolved metagenomic analysis. We generated 180 medium- to high-quality MAGs from the twelve co-assembled metagenomes (median completion = 88.75%, median contamination = 0.93%; Supplementary Table 5). These comprised 150 bacterial MAGs from 20 phyla, with the remaining 30 MAGs from 3 archaeal phyla (Fig. 4). The composition of prokaryotic phyla differed significantly by water depth, with distinct phyla found at 2 m, 8 m, and 17-28 m depths (Fig. S2; Supplementary Table 6), reflecting the same groupings as the gene-based clusters. This is supported by 16S rDNA amplicon sequencing of Bundera Sinkhole communities [18], which suggests similar depth-dependent composition of microbial taxa.

The communities inhabiting Bundera Sinkhole are taxonomically novel, with 75% of MAGs assigned to entirely new or uncharacterised families that lack cultured representatives. In the Genome taxonomy Database (GTDB), newly delineated taxa are allocated with

alphanumeric placeholder labels. Using GTDB nomenclature, we found that 64% of MAGs were assigned to families with such placeholder labels, and a further 11% of MAGs could not be assigned to any family (Supplementary Table 5). Even at the class level, almost a quarter of all MAGs in this system were assigned to placeholder-labelled lineages. Such taxonomic novelty is likely driven by niche adaptation to the distinctive geomorphological and physicochemical properties of anchialine ecosystems.

The suite of MAGs assembled from Bundera Sinkhole provides an ideal opportunity to assess the functional potential of these diverse and novel taxa. The relative abundance of MAG-related functions associates with water depth (Fig. 5), as observed with the gene-based functional analysis. We found that the number of MAGs that have the genetic potential for each key metabolic reaction varied considerably, as does their relative abundance at different depths.

We found that the taxonomy of carbon metabolism varied based on the carbon substrate (Fig. 5). For example, one-carbon (C1) molecules (e.g., methanol, formaldehyde, formate, and carbon monoxide) are largely metabolised by Proteobacteria, while complex carbon molecules (e.g., cellulose, chitin, starch, and other oligo- and poly-saccharides) are metabolised by bacteria from a wider range of phyla.

The taxonomy of autotrophic microbes differed based on the CO_2 fixation strategy (Fig. 5). Photo- and chemo-autotrophs that utilise RuBisCO as part of the carbon-fixing CBB cycle were almost all Proteobacteria (80%). A much more diverse range of bacteria and archaea had the genetic potential for utilising the reverse TCA (Patescibacteria, Nanoarchaeota, Campylobacterota, Myxococcota, Bacteroidota) and Wood-Ljungdahl (Planctomycetota, Desulfobacterota, Chloroflexota, Verrucomicrobiota, Nitrospirota, Bdellovibrionota) pathways for carbon fixation.

For the most part, N and S cycling pathways were performed by Proteobacteria (Fig. 5). As described above, both the DNRA and nitrification processes appear to be important N cycling pathways that drive a nitrogen-feedback loop in this system. The DNRA pathway, involving nitrate reduction to nitrite, which is then further reduced to ammonia, is largely driven by Proteobacteria (Fig. 5). The reverse of this process, nitrification, involves ammonia oxidation to nitrite, which is further oxidised to nitrate. Here, the final nitrification step (nitrite oxidation) is predominately driven by Myxococcota, and to a lesser extent, Planctomycetota, Marinisomatota, and Nitrospinota (Fig. 5). However, the first step in nitrification (ammonia oxidation), mediated by ammonia monooxygenases, was not detected in any MAG, despite their presence in the gene-based analysis (Fig. 2c). Therefore, to identify the taxa involved in ammonia oxidation, we queried the genes annotated as *amoA* (encoding the ammonia monooxgenase, alpha subunit) against NCBI's nr database using BLASTP. Three *amoA* genes were detected among the set of de-replicated genes. All three were identified as archaeal, belonging to the NCBI phylum Thaumarchaeota (classified in the GTDB as class Nitrososphaeria – phylum Thermoproteota [48]). Thus, the nitrogen-feedback loop that cycles between ammonia and nitrate is driven by distinct prokaryotes – predominately those belonging to Proteobacteria, Myxococcota, and Archaea. The aforementioned sulphate-feedback loop, associated with sulphate reduction (*sat*) and sulphur oxidation (*sdo*) processes, is also largely driven by Proteobacteria (Fig. 5).

Given the large metabolic contribution of Proteobacteria to this system, we further investigated their functional potential at lower taxonomic levels (Fig. 6). We found that the most important contributors to key metabolic reactions (based on read coverage) are species from less well characterised proteobacterial lineages. In particular, bacteria belonging to the gammaproteobacterial orders PS1 (n=1) and GCF-002020875 (n=7) were key contributors to carbon fixation (CBB cycle), and nitrogen and sulphur cycling (Fig. 6). The single PS1 MAG

belongs to the genus *Thioglobus*, which encompass members of the sulphur-oxidising marine SUP05 clade of Gammaproteobacteria. *Thioglobus* comprises a handful of cultured representatives which consist of chemoauto- and hetero-trophic bacteria that grow under aerobic and anaerobic conditions, and are assumed to contribute to denitrification [68-71]. The seven MAGs assigned to the order GCF-002020875, which lacks any cultured representatives, all belong to the same family, also designated GCF-002020875. Of these, four MAGs belong to the genus *Thiopontia*, while the other three MAGs were unclassified at the genus level. There are five species representative MAGs for *Thiopontia* (GCA_018671205.1, GCA_018658305.1, GCA_018648825.1, GCA_013349825.1, GCA_014384675.1), all of which were assembled from hypoxic saline water metagenomes [72-74] (NCBI BioProject Accessions: PRJNA630981, PRJNA632036, and PRJNA649215), suggesting that these bacteria are specific to this environmental niche.

Bundera Sinkhole has one to two highly abundant MAGs at each depth

Four highly abundant MAGs (with TPM values >50 in at least one sample) were dominant at different depths (Fig. 7). These included two gammaproteobacterial MAGs, one assigned at the family level (family GCF-002020875), and a *Thioglobus* sp., which were highly abundant at the 2 m and 8 m depths, respectively. A Marinisomatota MAG (order Marinisomatales) was highly abundant across all lower-depth samples (17-28 m). An archaeal MAG, *Nitrosopumilus* sp., was also abundant across the lower-depth samples, particularly, at the 22 m depth.

The GCF-002020875 MAG (MAG-172), which comprised ~9% of the metagenomic reads from the 2 m samples (Fig. 8), represents a novel gammaproteobacterial lineage, having no classification below the family level. It encodes several enzymes that would enable it to

374 utilise sulphur as an energy source. However, it also carries genes for complex carbon
375 degradation, suggesting it has the potential for both thioauto- and hetero-trophy. It also has
376 the genetic potential to mediate two steps in the denitrification pathway (nitrite reduction to
377 nitric oxide, and nitrous oxide reduction to N_2 gas).

378 The highly abundant *Thioglobus* MAG (MAG-2) represents a major component of the
379 8 m community, comprising 26% of the reads from the 8 m samples (Fig. 8). It encodes
380 several enzymes that suggest it also has the capacity for both thioauto- and hetero-trophy. It
381 appears to be an important mediator of sulphur cycling, encoding several sulphur
382 transformation pathways, and carries marker genes for the complete denitrification pathway,
383 converting nitrate to N_2 gas, via nitrite, nitric oxide, and nitrous oxide intermediates. Both
384 dominant MAGs at the 2 m and 8 m depths possess the genetic potential for several sulphur
385 cycling pathways as well as denitrification (Fig. 8). In marine oxygen minimum zones, a
386 denitrification pathway linking reduced sulphur compounds to the loss of bioavailable
387 nitrogen represents an important mode of metabolic coupling [75-78]. These two dominant
388 MAGs are likely mediating this linking of sulphur cycling and denitrification in the shallower
389 waters of the sinkhole.

390 In the deeper layers (17-28 m), two MAGs were highly abundant. One of these,
391 MAG-107, belongs to the genus *Nitrosopumilus*, which comprise a group of ammonia-
392 oxidising Archaea [79]. Given their important ecological role in ammonia oxidation, we
393 searched this MAG for the marker gene for ammonia oxidation, *amoA*, encoding the
394 ammonia monooxygenase alpha subunit. Surprisingly, *amoA* was not detected in this MAG.
395 However, as described above, we detected three archaeal *amoA* genes from the complete set
396 of de-replicated metagenomic genes. One of these was predicted to belong to the genus
397 *Nitrosopumilus* (100% query cover and 98.61% amino acid identity to *Nitrosopumilus* AmoA
398 [NCBI accession WP_141977518.1]), and its relative abundance is almost perfectly

399 correlated (r^2 = 0.97) with that of MAG-107, suggesting it to be indeed a component of its
400 genome. The failure for the *amoA* gene to be binned with MAG-107, is possibly due to the
401 several ribosomal protein genes co-located on the same contig (*rpl32e*, *rpl19e*, *rpl10*, *rpl12*,
402 *rpl21e*, *rps17e*, *rps11*, *rps15*, *rps3ae*), which are often difficult to bin because of their
403 differential codon usage patterns that have been optimised for rapid translation [80]. Besides
404 ammonia oxidation, this MAG also had the genetic potential for several nitrate reduction
405 pathways, as well as sulphite production (Fig. 8).
406 The Marinisomatales MAG, MAG-158, represents the other dominant MAG at the
407 lower depths. This MAG belongs to the phylum Marinisomatota, also commonly known as
408 Marinimicrobia. These bacteria are widespread in the global oceans, and are particularly
409 abundant in sub-euphotic oxygen minimum zones [75], which correspond to the samples that
410 MAG-158 was most abundant. Out of the four dominant MAGs, MAG-158 had the lowest
411 estimated genome completeness (57.14%), partially obscuring detailed analysis of its
412 metabolism. Nevertheless, we detected several enzymes involved in selenium and arsenic
413 cycling, as well as nitrate reduction (representing the first step in denitrification) (Fig. 8).
414 Previous analyses of these bacteria indicate that they are important drivers of denitrification
415 and sulphur cycling in hypoxic and anoxic seawater [75, 81], suggesting that this MAG might
416 also be involved in coupled sulphur-nitrogen cycling in the sinkhole.
417

418 **Conclusion**

419

420 Here, we characterised the metabolic and biogeochemical cycling potential of the microbial
421 communities inhabiting Bundera Sinkhole. We found that the microbial communities, largely
422 represented by novel taxonomic lineages, display depth-dependent metabolisms. Key
423 metabolic genes group into three depth-specific clusters that reflect distinct phases along the

dissolved oxygen and salinity gradients. In particular, chemotrophic metabolisms that couple nitrogen and sulphur cycling appear to be characteristic of the dominant members in this ecosystem. These data support the idea that microbial chemosynthesis is sustaining the higher trophic levels in the sinkhole. To the best of our knowledge, this is the first whole metagenomic analysis of an anchialine ecosystem, and thus presents key findings that contribute to our understanding of ecosystem functions in subterranean estuaries.

Understanding the diversity of metabolic strategies utilised by anchialine microbial communities can provide important insights into how trophic webs are supported in these unique ecosystems. This is particularly important given the high endemism of anchialine species and the potential vulnerability of these ecosystems to global environmental change and other anthropogenic influences [1]. Identifying the key microbial members and biogeochemical process is critical for the conservation of anchialine ecosystems.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Raw metagenomic sequence data are available in the NCBI SRA Database under BioSample Accessions SAMN32209613-SAMN32209624, from the BioProject PRJNA911846.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by Australian Research Council Laureate Fellowship #FL140100021 (I.T.P).

Authors' contributions

T.M.G conducted the data analyses and wrote the manuscript draft. A.F and L.D.H.E conducted data analyses. B.S performed the experimental work. W.H collected the water samples and was involved in the project design. I.T.P and S.G.T were involved in project design and management. All authors contributed to the final editing of the manuscript.

Acknowledgements

TMG would like to thank Mary Ghaly for support and discussions regarding the manuscript.

References

1. Mejía-Ortız LM, Chavez-Solıs EM, Brankovits D: **Editorial: The effects of environmental change on anchialine ecosystems.** *Frontiers in Marine Science* 2022, **9:**1029027.
2. Moore WS: **The subterranean estuary: a reaction zone of ground water and sea water.** *Marine Chemistry* 1999, **65:**111-125.
3. Moore WS: **The effect of submarine groundwater discharge on the ocean.** *Annual Review of Marine Science* 2010, **2:**59-88.
4. Bishop RE, Humphreys WF, Cukrov N, Žic V, Boxshall GA, Cukrov M, Iliffe TM, Kršinić F, Moore WS, Pohlman JW: **'Anchialine' redefined as a subterranean estuary in a crevicular or cavernous geological setting.** *Journal of Crustacean Biology* 2015, **35:**511-514.
5. van Hengstum PJ, Cresswell JN, Milne GA, Iliffe TM: **Development of anchialine cave habitats and karst subterranean estuaries since the last ice age.** *Scientific Reports* 2019, **9:**11907.
6. Calderón-Gutiérrez F, Sánchez-Ortiz CA, Huato-Soberanis L: **Ecological patterns in anchialine caves.** *PLoS One* 2018, **13:**e0202909.
7. Pohlman JW: **The biogeochemistry of anchialine caves: progress and possibilities.** *Hydrobiologia* 2011, **677:**33-51.

8. Pohlman JW, Iliffe TM, Cifuentes LA: **A stable isotope study of organic cycling and the ecology of an anchialine cave ecosystem.** *Marine Ecology Progress Series* 1997, **155:**17-27.
9. Sarbu SM, Kane TC, Kinkle BK: **A chemoautotrophically based cave ecosystem.** *Science* 1996, **272:**1953-1955.
10. Brankovits D, Pohlman JW, Niemann H, Leigh MB, Leewis MC, Becker KW, Iliffe TM, Alvarez F, Lehmann MF, Phillips B: **Methane- and dissolved organic carbon-fueled microbial loop supports a tropical subterranean estuary ecosystem.** *Nature Communications* 2017, **8:**1835.
11. Kajan K, Cukrov N, Cukrov N, Bishop-Pierce R, Orlić S: **Microeukaryotic and prokaryotic diversity of anchialine caves from Eastern Adriatic Sea Islands.** *Microbial Ecology* 2022, **83:**257-270.
12. He H, Fu L, Liu Q, Fu L, Bi N, Yang Z, Zhen Y: **Community structure, abundance and potential functions of bacteria and archaea in the Sansha Yongle Blue Hole, Xisha, South China Sea.** *Frontiers in Microbiology* 2019, **10:**2404.
13. Cleary D, Polónia A: **Bacterial and archaeal communities inhabiting mussels, sediment and water in Indonesian anchialine lakes.** *Antonie van Leeuwenhoek* 2018, **111:**237-257.
14. Risley CA, Tamalavage AE, van Hengstum PJ, Labonté JM: **Subsurface microbial community composition in anchialine environments is influenced by original organic carbon source at time of deposition.** *Frontiers in Marine Science* 2022**:**480.
15. Suárez-Moo P, Remes-Rodríguez CA, Márquez-Velázquez NA, Falcón LI, García-Maldonado JQ, Prieto-Davó A: **Changes in the sediment microbial community structure of coastal and inland sinkholes of a karst ecosystem from the Yucatan Peninsula.** *Scientific Reports* 2022, **12:**1110.
16. Humphreys W: **Physico-chemical profile and energy fixation in Bundera Sinkhole, an anchialine remiped habitat in north-western Australia.** *Journal of the Royal Society of Western Australia* 1999, **82:**89-98.
17. Seymour J, Humphreys W, Mitchell J: **Stratification of the microbial community inhabiting an anchialine sinkhole.** *Aquatic Microbial Ecology* 2007, **50:**11-24.
18. Elbourne LDH, Sutcliffe B, Humphreys W, Focardi A, Saccò M, Campbell MA, Paulsen IT, Tetu SG: **Unravelling stratified microbial assemblages in Australia's only deep anchialine system, the Bundera Sinkhole.** *Frontiers in Marine Science* 2022, **9:**872082.
19. Yager J, Humphreys W: ***Lasionectes exleyi*, sp, nov., the first remipede crustacean recorded from Australia and the Indian Ocean, with a key to the world species.** *Invertebrate Systematics* 1996, **10:**171-187.
20. Danielopol DL, Baltanás A, Humphreys WF: ***Danielopolina kornickeri* sp. n.(Ostracoda, Thaumatocypridoidea) from a western Australian anchialine cave: morphology and evolution.** *Zoologica Scripta* 2000, **29:**1-16.
21. Jaume D, Humphreys WF: **A new genus of epacteriscid calanoid copepod from an anchialine sinkhole on northwestern Australia.** *Journal of Crustacean Biology* 2001, **21:**157-169.
22. Wilson RS, Humphreys WF: ***Prionospio thalanji* sp. nov.(Polychaeta: Spionidae) from an anchialine cave, Cape Range, northwest Western Australia.** *Records of the Western Australian Museum Supplement* 2001, **64:**e113.
23. Humphreys W, Tetu S, Elbourne L, Gillings M, Seymour J, Mitchell J, Paulsen I: **Geochemical and microbial diversity of Bundera sinkhole, an anchialine system in the eastern Indian ocean.** *Natura Croatica: Periodicum Musei Historiae Naturalis Croatici* 2012, **21:**59-63.

24. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30:**2114-2120.
25. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA: **metaSPAdes: a new versatile metagenomic assembler.** *Genome Research* 2017, **27:**824-834.
26. Mikheenko A, Saveliev V, Gurevich A: **MetaQUAST: evaluation of metagenome assemblies.** *Bioinformatics* 2016, **32:**1088-1090.
27. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification.** *BMC Bioinformatics* 2010, **11:**119.
28. Li W, Godzik A: **CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22:**1658-1659.
29. Fu L, Niu B, Zhu Z, Wu S, Li W: **CD-HIT: accelerated for clustering the next-generation sequencing data.** *Bioinformatics* 2012, **28:**3150-3152.
30. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, Karaoz U, Anantharaman K: **METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks.** *Microbiome* 2022, **10:**33.
31. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H: **KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold.** *Bioinformatics* 2020, **36:**2251-2252.
32. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Research* 2000, **28:**27-30.
33. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O: **TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes.** *Nucleic Acids Research* 2006, **35:**D260-D264.
34. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al: **Pfam: the protein families database.** *Nucleic Acids Research* 2013, **42:**D222-D230.
35. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, et al: **Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system.** *Nature Communications* 2016, **7:**13219.
36. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph.** *Bioinformatics* 2015, **31:**1674-1676.
37. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W: **MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices.** *Methods* 2016, **102:**3-11.
38. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nature Methods* 2012, **9:**357-359.
39. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ* 2019, **7:**e7359.
40. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO: **Anvi'o: an advanced analysis and visualization platform for 'omics data.** *PeerJ* 2015, **3:**e1319.
41. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Research* 2015, **25:**1043-1055.

42. Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, Schmidt TS, Bork P: **GUNC: detection of chimerism and contamination in prokaryotic genomes.** *Genome Biology* 2021, **22:**1-19.
43. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, et al: **Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.** *Nature Biotechnology* 2017, **35:**725-731.
44. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH: **GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.** *Bioinformatics* 2019, **36:**1925-1927.
45. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH: **GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database.** *Bioinformatics* 2022**:**btac672.
46. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P: **A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life.** *Nature Biotechnology* 2018, **36:**996-1004.
47. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P: **A complete domain-to-species taxonomy for Bacteria and Archaea.** *Nature Biotechnology* 2020, **38:**1079-1086.
48. Rinke C, Chuvochina M, Mussig AJ, Chaumeil P-A, Davín AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P: **A standardized archaeal taxonomy for the Genome Taxonomy Database.** *Nature Microbiology* 2021, **6:**946-959.
49. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P: **GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy.** *Nucleic Acids Research* 2021, **50:**D785-D794.
50. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW: **Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life.** *Nature Microbiology* 2017, **2:**1533-1542.
51. Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, Rinke C, Spang A: **Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution.** *Nature Communications* 2020, **11:**3939.
52. Price MN, Dehal PS, Arkin AP: **FastTree: computing large minimum evolution trees with profiles instead of a distance matrix.** *Molecular Biology and Evolution* 2009, **26:**1641-1650.
53. Price MN, Dehal PS, Arkin AP: **FastTree 2–approximately maximum-likelihood trees for large alignments.** *PLoS One* 2010, **5:**e9490.
54. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y: **ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data.** *Methods in Ecology and Evolution* 2017, **8:**28-36.
55. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, Tang W, Feng T, Chen M, Zhan L, et al: **ggtreeExtra: Compact visualization of richly annotated phylogenetic data.** *Molecular Biology and Evolution* 2021, **38:**4039-4042.
56. Tu Q, Lin L, Cheng L, Deng Y, He Z: **NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes.** *Bioinformatics* 2019, **35:**1040-1048.
57. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND.** *Nature Methods* 2015, **12:**59-60.

58. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, et al: **vegan: Community ecology package. R package version 2.5–7. https://CRAN.R-project.org/package=vegan.** 2022.
59. Martinez Arbizu P: **pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4. https://github.com/pmartinezarbizu/pairwiseAdonis.** 2020.
60. Berg IA: **Ecological aspects of the distribution of different autotrophic CO2 fixation pathways.** *Applied and Environmental Microbiology* 2011, **77:**1925-1936.
61. Badger MR, Bek EJ: **Multiple Rubisco forms in Proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle.** *Journal of Experimental Botany* 2008, **59:**1525-1541.
62. Peura S, Buck M, Aalto SL, Morales SE, Nykänen H, Eiler A: **Novel autotrophic organisms contribute significantly to the internal carbon cycling potential of a boreal lake.** *MBio* 2018, **9:**e00916-00918.
63. Jaffe AL, Bardot C, Le Jeune A-H, Liu J, Colombet J, Perrière F, Billard H, Castelle CJ, Lehours A-C, Banfield JF: **Variable impact of geochemical gradients on the functional potential of bacteria, archaea, and phages from the permanently stratified Lac Pavin.** *bioRxiv* 2022**:**2022.2007.2018.500538.
64. Stolz JF, Oremland RS: **Bacterial respiration of arsenic and selenium.** *FEMS Microbiology Reviews* 1999, **23:**615-627.
65. Oremland RS, Stolz JF: **The acology of arsenic.** *Science* 2003, **300:**939-944.
66. Rabus R, Venceslau SS, Woehlbrand L, Voordouw G, Wall JD, Pereira IA: **A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes.** *Advances in Microbial Physiology* 2015, **66:**55-321.
67. Rohwerder T, Sand W: **The sulfane sulfur of persulfides is the actual substrate of the sulfur-oxidizing enzymes from *Acidithiobacillus* and *Acidiphilium* spp.** *Microbiology* 2003, **149:**1699-1710.
68. Hawley AK, Brewer HM, Norbeck AD, Paša-Tolić L, Hallam SJ: **Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes.** *Proceedings of the National Academy of Sciences* 2014, **111:**11395-11400.
69. Marshall KT, Morris RM: **Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade.** *The ISME Journal* 2013, **7:**452-455.
70. Shah V, Chang BX, Morris RM: **Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium.** *The ISME Journal* 2017, **11:**263-271.
71. Spietz RL, Lundeen RA, Zhao X, Nicastro D, Ingalls AE, Morris RM: **Heterotrophic carbon metabolism and energy acquisition in Candidatus *Thioglobus singularis* strain PS1, a member of the SUP05 clade of marine Gammaproteobacteria.** *Environmental Microbiology* 2019, **21:**2391-2401.
72. Lin H, Ascher DB, Myung Y, Lamborg CH, Hallam SJ, Gionfriddo CM, Holt KE, Moreau JW: **Mercury methylation by metabolically versatile and cosmopolitan marine bacteria.** *The ISME Journal* 2021, **15:**1810-1825.
73. Uzun M, Alekseeva L, Krutkina M, Koziaeva V, Grouzdev D: **Unravelling the diversity of magnetotactic bacteria through analysis of open genomic databases.** *Scientific Data* 2020, **7:**252.
74. Villanueva L, von Meijenfeldt FAB, Westbye AB, Yadav S, Hopmans EC, Dutilh BE, Damsté JSS: **Bridging the membrane lipid divide: bacteria of the FCB group**

superphylum have the potential to synthesize archaeal ether lipids. *The ISME Journal* 2021, **15:**168-182.

75. Hawley AK, Nobu MK, Wright JJ, Durno WE, Morgan-Lang C, Sage B, Schwientek P, Swan BK, Rinke C, Torres-Beltrán M, et al: **Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients.** *Nature Communications* 2017, **8:**1507.

76. Stewart FJ, Ulloa O, DeLong EF: **Microbial metatranscriptomics in a permanent marine oxygen minimum zone.** *Environmental Microbiology* 2012, **14:**23-40.

77. Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, Ranjan P, Sarode N, Malmstrom RR, Padilla CC, et al: **SAR11 bacteria linked to ocean anoxia and nitrogen loss.** *Nature* 2016, **536:**179-183.

78. Louca S, Hawley AK, Katsev S, Torres-Beltran M, Bhatia MP, Kheirandish S, Michiels CC, Capelle D, Lavik G, Doebeli M, et al: **Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone.** *Proceedings of the National Academy of Sciences* 2016, **113:**E5925-E5933.

79. Prosser JI, Nicol GW: **Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment.** *Environmental Microbiology* 2008, **10:**2931-2941.

80. Mise K, Iwasaki W: **Unexpected absence of ribosomal protein genes from metagenome-assembled genomes.** *ISME Communications* 2022, **2:**118.

81. Bertagnolli AD, Padilla CC, Glass JB, Thamdrup B, Stewart FJ: **Metabolic potential and in situ activity of marine Marinimicrobia bacteria in an anoxic water column.** *Environmental Microbiology* 2017, **19:**4392-4416.

Figure captions

Fig. 1. Location and sampling map of the Bundera Sinkhole. (**a**) Location of the Bundera Sinkhole in the Cape Range Peninsula, Western Australia. (**b**) Topology of the sinkhole and sampling points for shotgun metagenomic sequencing. Figure panels are adapted from Elbourne LDH, *et al.* [18].

Fig. 2. Relative abundance and diversity of key metabolic and biogeochemical cycling genes in Bundera Sinkhole. (**a**-**b**) Non-linear multidimensional scaling (NMDS) based on Bray-Curtis distances of normalised read counts for (**a**) whole metagenomes (with genes dereplicated at 98% nucleotide identity) and (**b**) key metabolic genes (TPM sums) displayed in panel **c**. In panel **a**, NMDS points that represent replicate samples lie on top of each other, as do those representing all samples from 17, 18, 22, and 28 m depths. The NMDS groupings (circles, triangles, and squares) represent samples with similar levels of dissolved oxygen (DO) and salinity (Supplementary Table 1). In both NDMS plots, the grouping of samples from 17, 18, 22, and 28 m depths (squares) is supported by PERMANOVA ($p=0.04$; Supplementary Table 2). (**c**) Relative abundance of key metabolic marker genes within each sample. Colour scale is displayed as $\log_{10}(\text{TPM} +1)$ to account for TPM values of zero. Gene names are displayed to the left of the heatmap, and the reactions that they facilitate are on the right. (**d**) Visualisation of microbial nitrogen and sulphur cycling pathways present in Bundera Sinkhole. Chemical compounds that represent either the substrate or product of a reaction are boxed, with oxidation states shown in parentheses.

Fig. 3. Correlations between chemical compound concentrations and genes involved in their cycling. Nitrogen and sulphur cycling genes whose relative abundance (TPM) are strongly correlated ($r^2 > 0.6$) with the environmental concentrations of (**a**-**c**) ammonia (NH_3), (**d**-**e**) nitrate (NO_3^-), and (**f**-**g**) sulphate (SO_4^{2-}). Plots coloured red represent genes involved in pathways that produce the corresponding chemical compound, either directly (**b**,**c**,**e**) or indirectly, via an intermediate compound (**a**,**d**,**g**). Correlation between *sat* gene relative abundance and SO4 concentrations (**f**) is coloured blue to indicate the gene's involvement in SO4 substrate utilisation. Shaded regions represent the 95% confidence interval of the fitted linear model. A full list of r^2 and p-values for all evaluated nitrogen and sulphur cycling gene correlations is presented as Supplementary Table 3.

Fig. 4. **Domain-specific phylogenies of MAGs from Bundera Sinkhole.** Tips of the trees are coloured by their assigned phylum. Heatmaps display the relative abundance of MAGs in each of the duplicate samples collected from six depths (from inner to outer rings: 2 m, 8 m, 17 m, 18 m, 22 m, and 28 m).

Fig. 5. Key metabolic and biogeochemical cycling traits of MAGs in Bundera Sinkhole. From left to right: the numbers of MAGs that carry genetic markers (listed in Supplementary Table 7) for each functional trait are displayed by numerals, and represented visually by the size of the circles; the average relative abundance (TPM) for corresponding MAGs at each depth are displayed by the blue heatmap; and the proportion of MAGs assigned to each phylum is represented by the red heatmap. Archaeal phyla are denoted with asterisks.

Fig. 6. Metabolic functions associated with proteobacterial MAGs. MAGs are grouped according to their taxonomic class (left) and order (middle). Width of curved lines indicate the relative contribution, based on read coverage, of proteobacterial orders (middle) to a given metabolic reaction (right).

Fig. 7. Relative abundance of MAGs in Bundera Sinkhole. Phyla of MAGs are displayed to the left of the heatmap. Archaeal phyla are denoted with asterisks. The four most abundant MAGs, having a TPM value greater than 50 in any one sample, are denoted on the right.

Fig. 8. Metabolisms of the most highly abundant MAGs in Bundera Sinkhole. Estimated genome completeness is displayed within square brackets under each MAG ID. Pie charts indicate the proportion of reads at each depth that map to the four MAGs. Metabolic reactions are labelled in red text, proteins mediating those reactions are labelled in black text, and the reaction products/substrates are labelled in blue text. Bar charts indicate the dissolved oxygen (DO) and salinity at each depth. In MAG-107, ammonia oxidation is displayed as a dashed arrow, as the *amoA* gene was not originally binned with this MAG. However, it was included here after detecting an *amoA* gene, taxonomically classified as *Nitrosopumilus*, that had a relative abundance almost perfectly correlated ($r^2 = 0.97$) with that of MAG-107.

Supplementary Figures

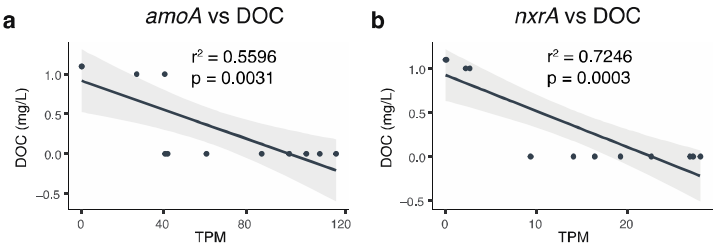


Fig. S1. Correlation between dissolved organic carbon and nitrification. (**a**) Correlation between the relative abundance (TPM) of the marker gene for ammonia oxidation (*amoA*; first step of nitrification) and dissolved organic carbon (DOC) concentration. (**b**) Correlation between the relative abundance of the marker gene for nitrite oxidation (*nxrA*; final step of nitrification) and DOC concentration. Shaded regions represent the 95% confidence interval of the fitted linear model. A full list of r^2 and p-values for all evaluated nitrogen and sulphur cycling gene correlations is presented as Supplementary Table 4.

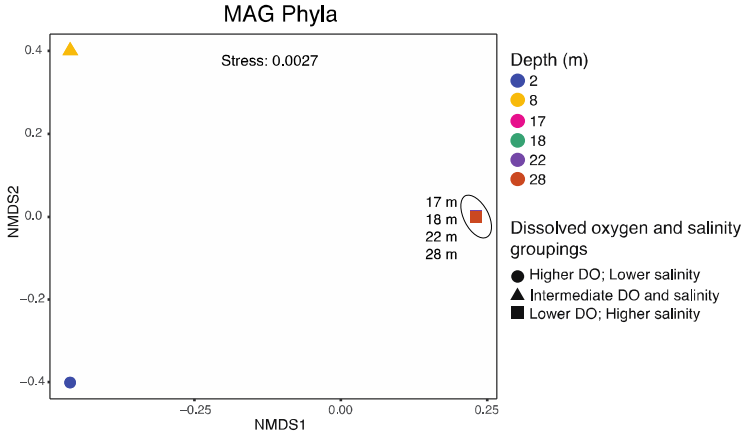
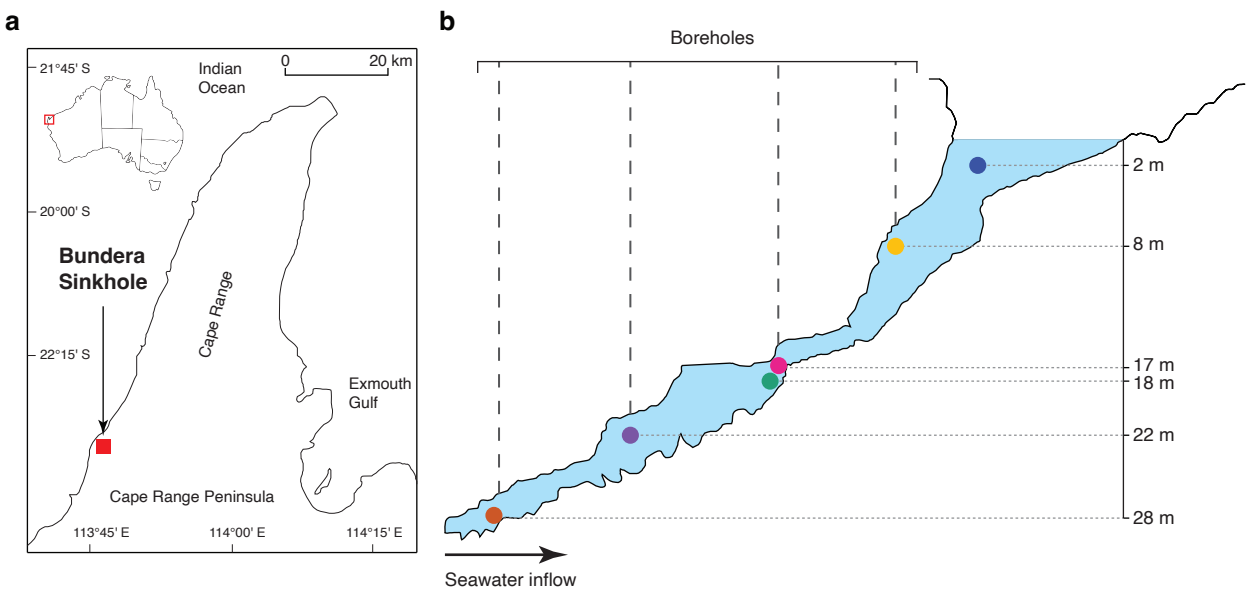
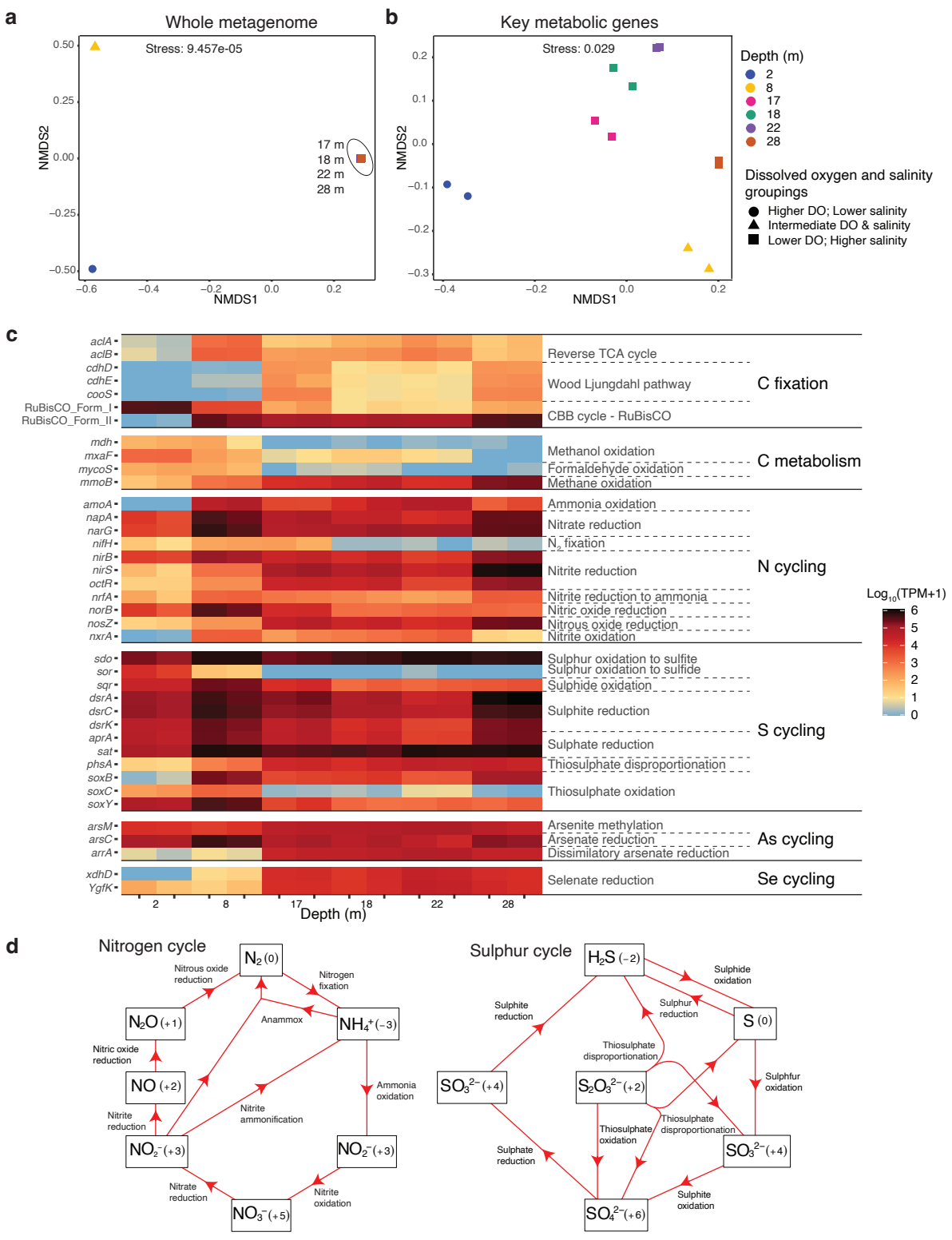
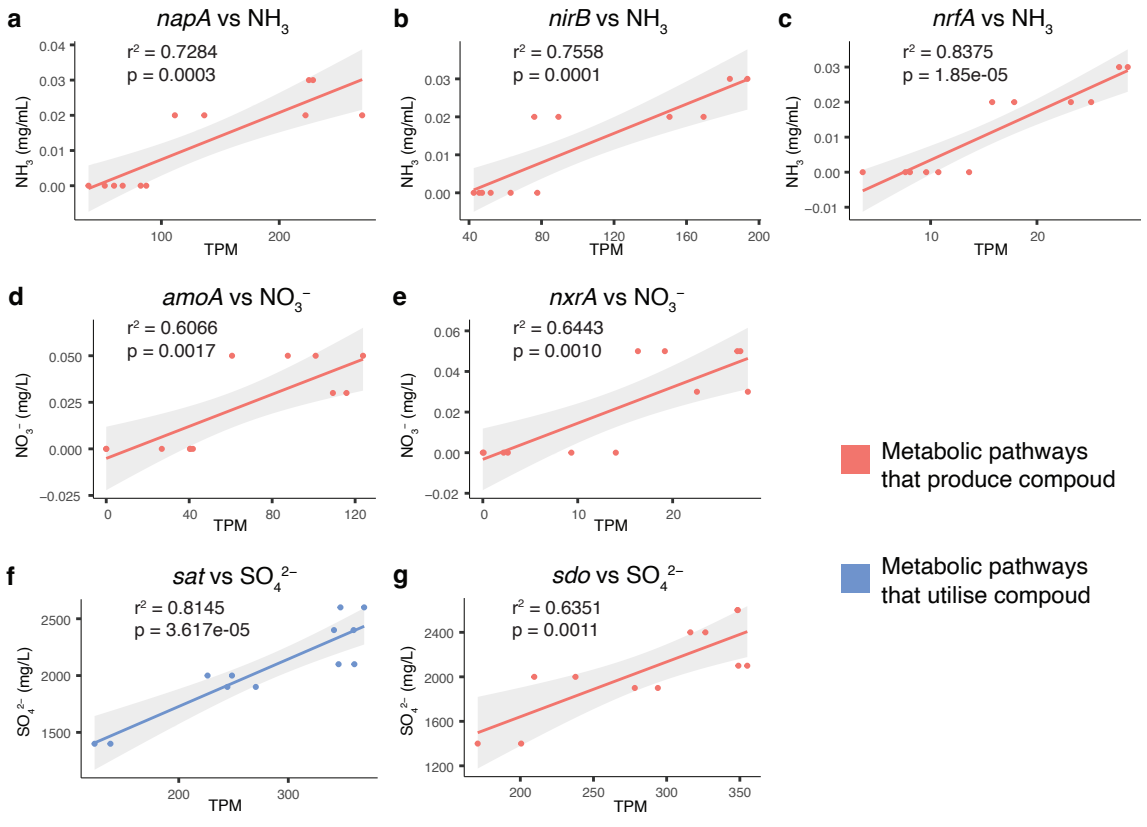
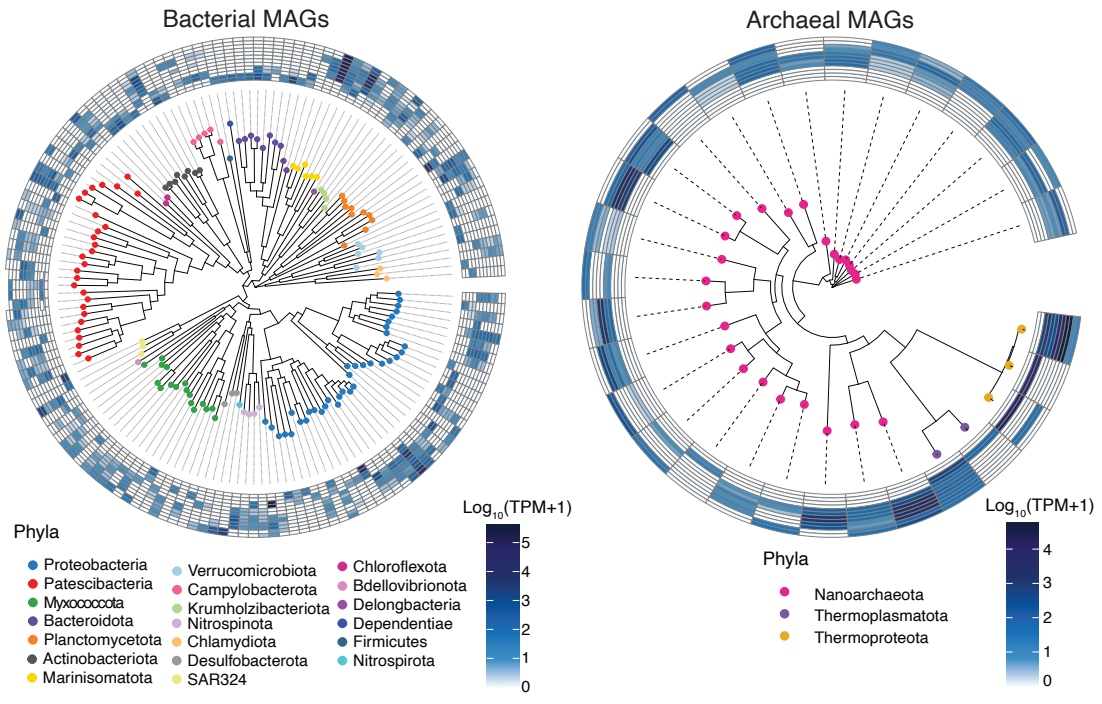


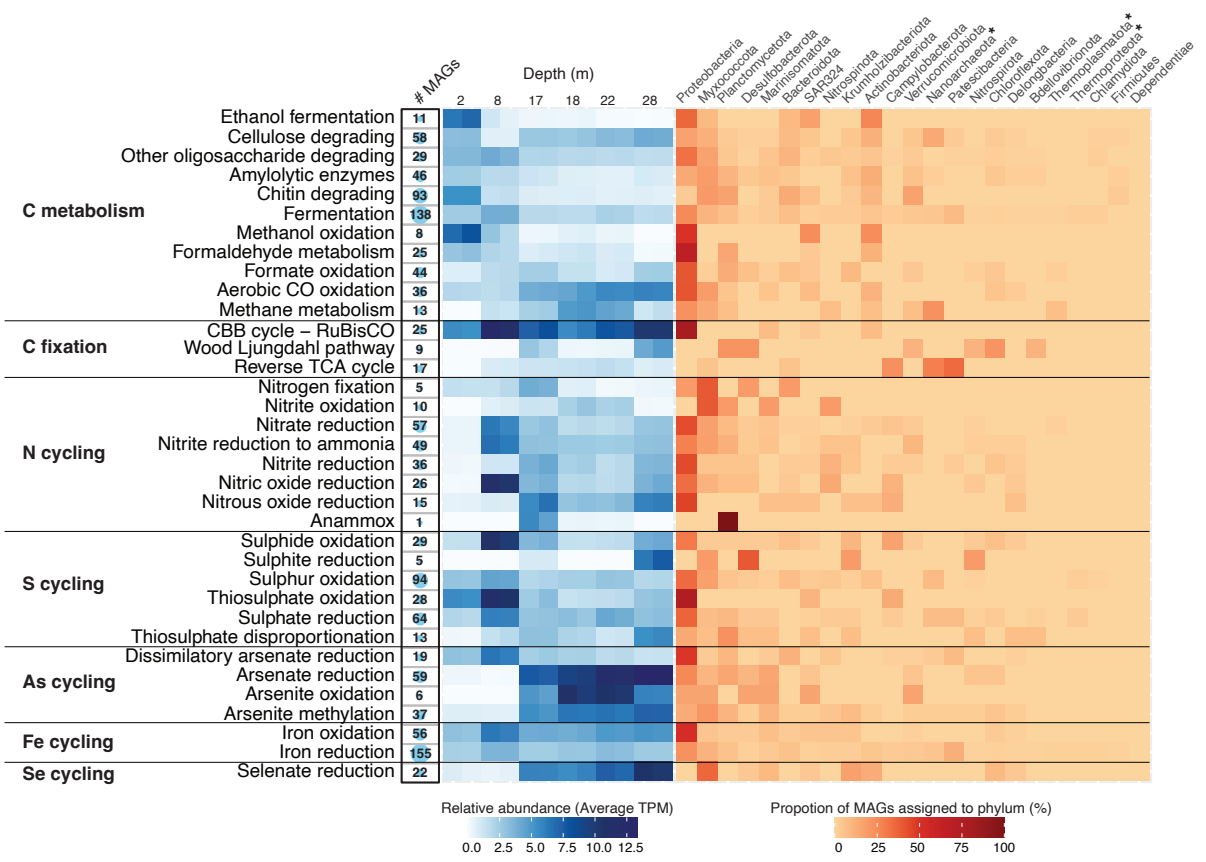
Fig. S2. Beta-diversity of MAG phyla in the Bundera sinkhole. Non-linear multidimensional scaling (NMDS) based on Bray-Curtis distances of normalised read counts for MAG phyla. NMDS points that represent replicate samples lie on top of each other, as do those representing all samples from 17, 18, 22, and 28 m depths. The groupings (circles, triangles, and squares) represent samples with similar levels of dissolved oxygen (DO) and salinity (Supplementary Table 1). The grouping of samples from 17, 18, 22, and 28m depths (squares) is supported by PERMANOVA (p=0.046; Supplementary Table 6).

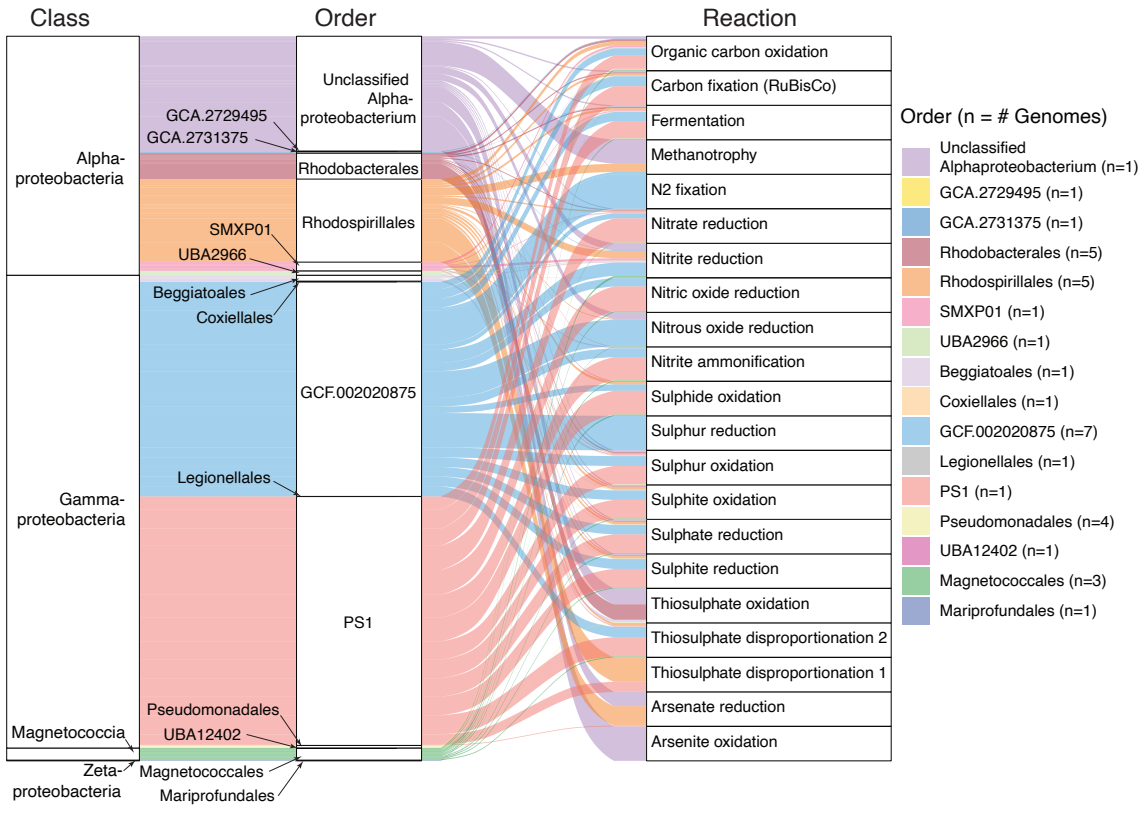

a
21°45' S
Indian
Ocean
0
20 km
20°00' S
Bundera
Sinkhole
22°15' S
Cape Range
Exmouth
Gulf
Cape Range Peninsula
113°45' E
114°00' E
114°15' E
b
Boreholes
2 m
8 m
17 m
18 m
22 m
28 m
Seawater inflow

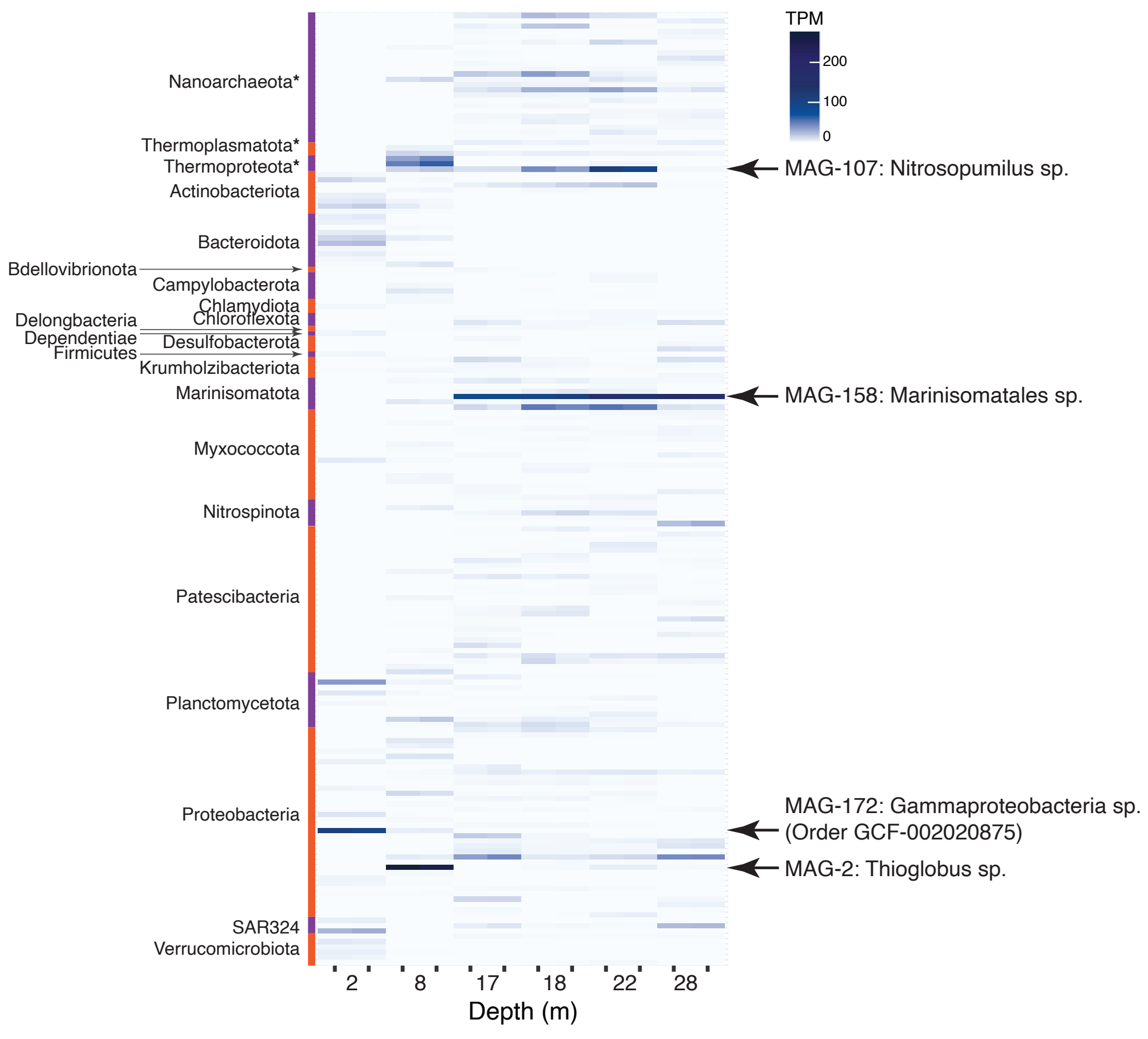

a
Whole metagenome
Stress: 9.457e-05
NMDS2
NMDS1
17 m
18 m
22 m
28 m
b
Key metabolic genes
Stress: 0.029
NMDS2
NMDS1
Depth (m)
2
8
17
18
22
28
Dissolved oxygen and salinity groupings
Higher DO; Lower salinity
Intermediate DO & salinity
Lower DO; Higher salinity
c
aclA
aclB
cdhD
cdhE
cooS
RuBisCO_Form_I
RuBisCO_Form_II
Reverse TCA cycle
Wood Ljungdahl pathway
CBB cycle - RuBisCO
C fixation
mdh
mxaF
mycoS
mmoB
Methanol oxidation
Formaldehyde oxidation
Methane oxidation
C metabolism
amoA
napA
narG
nifH
nirB
nirS
octR
nrfA
norB
nosZ
nxrA
Ammonia oxidation
Nitrate reduction
N2 fixation
Nitrite reduction
Nitrite reduction to ammonia
Nitric oxide reduction
Nitrous oxide reduction
Nitrite oxidation
N cycling
sdo
sor
sqr
dsrA
dsrC
dsrK
aprA
sat
phsA
soxB
soxC
soxY
Sulphur oxidation to sulfite
Sulphur oxidation to sulfide
Sulphide oxidation
Sulphite reduction
Sulphate reduction
Thiosulphate disproportionation
Thiosulphate oxidation
S cycling
arsM
arsC
arrA
Arsenite methylation
Arsenate reduction
Dissimilatory arsenate reduction
As cycling
xdhD
YgfK
Selenate reduction
Se cycling
2
8
17
18
22
28
Depth (m)
Log10(TPM+1)
6
5
4
3
2
1
0
d
Nitrogen cycle
N2 (0)
Nitrous oxide reduction
Nitrogen fixation
N2O (+1)
Anammox
NH4+ (-3)
Nitric oxide reduction
NO (+2)
Ammonia oxidation
Nitrite reduction
Nitrite ammonification
NO2- (+3)
NO2- (+3)
Nitrate reduction
Nitrite oxidation
NO3- (+5)
Sulphur cycle
H2S (-2)
Sulphide oxidation
Sulphite reduction
Sulphur reduction
S (0)
Thiosulphate disproportionation
SO3 2- (+4)
S2O3 2- (+2)
Sulphfur oxidation
Thiosulphate oxidation
Thiosulphate disproportionation
Sulphate reduction
SO3 2- (+4)
Sulphite oxidation
SO4 2- (+6)

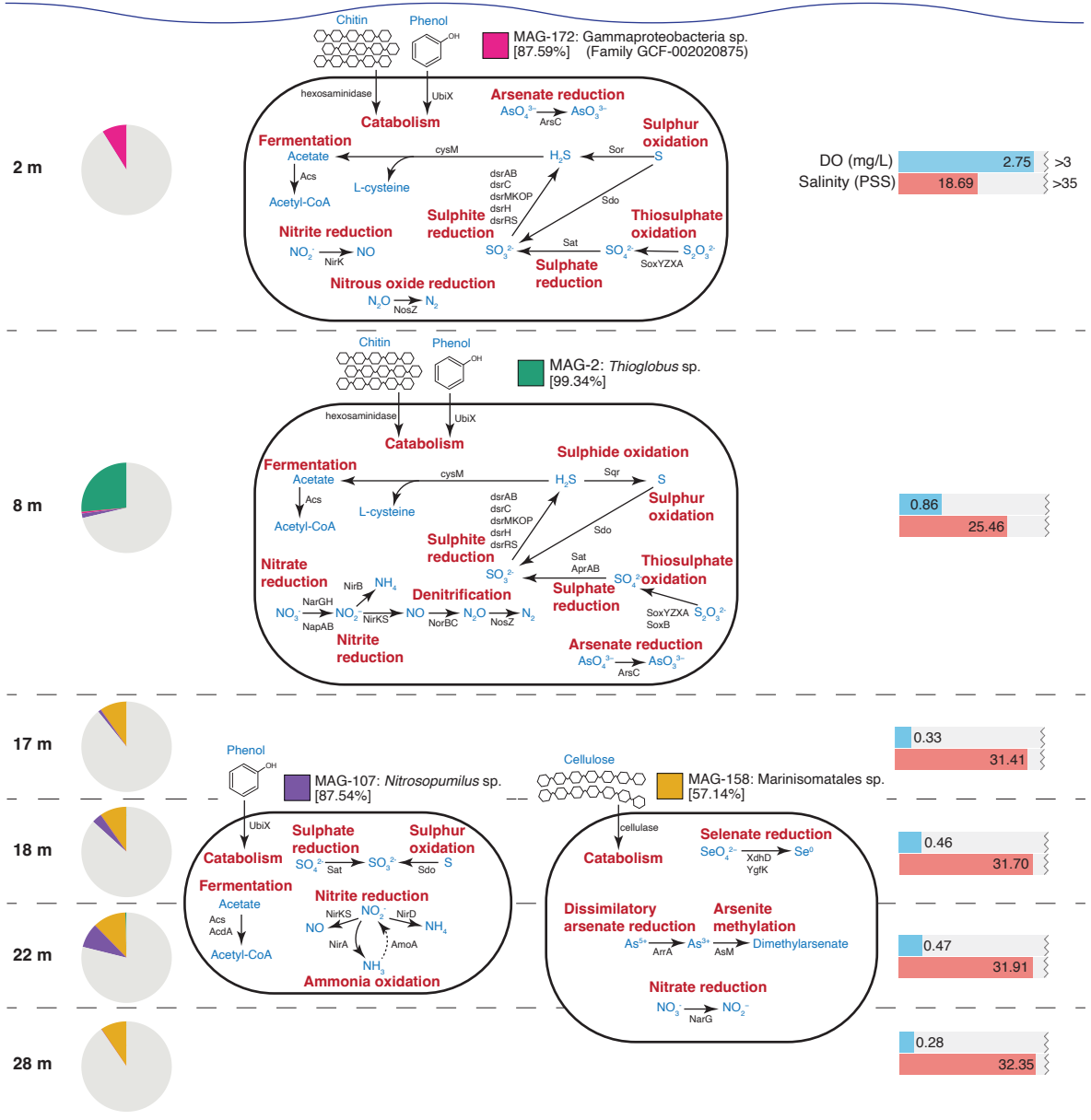

a
napA vs NH3
r2 = 0.7284
p = 0.0003
NH3 (mg/mL)
TPM
b
nirB vs NH3
r2 = 0.7558
p = 0.0001
NH3 (mg/mL)
TPM
c
nrfA vs NH3
r2 = 0.8375
p = 1.85e-05
NH3 (mg/mL)
TPM
d
amoA vs NO3−
r2 = 0.6066
p = 0.0017
NO3− (mg/L)
TPM
e
nxrA vs NO3−
r2 = 0.6443
p = 0.0010
NO3− (mg/L)
TPM
f
sat vs SO4 2−
r2 = 0.8145
p = 3.617e-05
SO4 2− (mg/L)
TPM
g
sdo vs SO4 2−
r2 = 0.6351
p = 0.0011
SO4 2− (mg/L)
TPM
Metabolic pathways that produce compoud
Metabolic pathways that utilise compoud


Bacterial MAGs
Archaeal MAGs
Phyla
Proteobacteria
Patescibacteria
Myxococcota
Bacteroidota
Planctomycetota
Actinobacteriota
Marinisomatota
Verrucomicrobiota
Campylobacterota
Krumholzibacteriota
Nitrospinota
Chlamydiota
Desulfobacterota
SAR324
Chloroflexota
Bdellovibrionota
Delongbacteria
Dependentiae
Firmicutes
Nitrospirota
Log10(TPM+1)
5
4
3
2
1
0
Phyla
Nanoarchaeota
Thermoplasmatota
Thermoproteota
Log10(TPM+1)
4
3
2
1
0


MAGs
Depth (m)
2
8
17
18
22
28
Proteobacteria
Myxococcota
Planctomycetota
Desulfobacterota
Marinisomatota
Bacteroidota
SAR324
Nitrospinota
Krumholzibacteriota
Actinobacteriota
Campylobacterota
Verrucomicrobiota
Nanoarchaeota*
Patescibacteria
Nitrospirota
Chloroflexota
Deongbacteria
Bdellovibrionota
Thermoplasmatota*
Thermoproteota*
Chlamydiota
Firmicutes
Dependentiae
C metabolism
Ethanol fermentation 11
Cellulose degrading 58
Other oligosaccharide degrading 29
Amylolytic enzymes 46
Chitin degrading 93
Fermentation 138
Methanol oxidation 8
Formaldehyde metabolism 25
Formate oxidation 44
Aerobic CO oxidation 36
Methane metabolism 13
C fixation
CBB cycle – RuBisCO 25
Wood Ljungdahl pathway 9
Reverse TCA cycle 17
N cycling
Nitrogen fixation 5
Nitrite oxidation 10
Nitrate reduction 57
Nitrite reduction to ammonia 49
Nitrite reduction 36
Nitric oxide reduction 26
Nitrous oxide reduction 15
Anammox 1
S cycling
Sulphide oxidation 29
Sulphite reduction 5
Sulphur oxidation 94
Thiosulphate oxidation 28
Sulphate reduction 64
Thiosulphate disproportionation 13
As cycling
Dissimilatory arsenate reduction 19
Arsenate reduction 59
Arsenite oxidation 6
Arsenite methylation 37
Fe cycling
Iron oxidation 56
Iron reduction 155
Se cycling
Selenate reduction 22
Relative abundance (Average TPM)
0.0 2.5 5.0 7.5 10.0 12.5
Propotion of MAGs assigned to phylum (%)
0 25 50 75 100


Class
Order
Reaction
Alpha-proteobacteria
Gamma-proteobacteria
Magnetococcia
Zeta-proteobacteria
GCA.2729495
GCA.2731375
SMXP01
UBA2966
Beggiatoales
Coxiellales
Legionellales
Pseudomonadales
UBA12402
Magnetococcales
Mariprofundales
Unclassified Alpha-proteobacterium
Rhodobacterales
Rhodospirillales
GCF.002020875
PS1
Organic carbon oxidation
Carbon fixation (RuBisCo)
Fermentation
Methanotrophy
N2 fixation
Nitrate reduction
Nitrite reduction
Nitric oxide reduction
Nitrous oxide reduction
Nitrite ammonification
Sulphide oxidation
Sulphur reduction
Sulphur oxidation
Sulphite oxidation
Sulphate reduction
Sulphite reduction
Thiosulphate oxidation
Thiosulphate disproportionation 2
Thiosulphate disproportionation 1
Arsenate reduction
Arsenite oxidation
Order (n = # Genomes)
Unclassified Alphaproteobacterium (n=1)
GCA.2729495 (n=1)
GCA.2731375 (n=1)
Rhodobacterales (n=5)
Rhodospirillales (n=5)
SMXP01 (n=1)
UBA2966 (n=1)
Beggiatoales (n=1)
Coxiellales (n=1)
GCF.002020875 (n=7)
Legionellales (n=1)
PS1 (n=1)
Pseudomonadales (n=4)
UBA12402 (n=1)
Magnetococcales (n=3)
Mariprofundales (n=1)


TPM
200
100
0
Nanoarchaeota*
Thermoplasmatota*
Thermoproteota*
Actinobacteriota
Bacteroidota
Bdellovibrionota
Campylobacterota
Chlamydiota
Chloroflexota
Delongbacteria
Dependentiae
Desulfobacterota
Firmicutes
Krumholzibacteriota
Marinisomatota
Myxococcota
Nitrospinota
Patescibacteria
Planctomycetota
Proteobacteria
SAR324
Verrucomicrobiota
MAG-107: Nitrosopumilus sp.
MAG-158: Marinisomatales sp.
MAG-172: Gammaproteobacteria sp.
(Order GCF-002020875)
MAG-2: Thioglobus sp.
2
8
17
18
22
28
Depth (m)


2 m
Chitin
Phenol
MAG-172: Gammaproteobacteria sp. [87.59%] (Family GCF-002020875)
hexosaminidase
UbiX
Catabolism
Arsenate reduction
AsO4 3- → AsO3 3- (ArsC)
Fermentation
Acetate
Acs
Acetyl-CoA
L-cysteine
cysM
H2S
Sor
S
Sulphur oxidation
dsrAB
dsrC
dsrMKOP
dsrH
dsrRS
Sdo
Sulphite reduction
Thiosulphate oxidation
Nitrite reduction
NO2- → NO (NirK)
SO3 2-
Sat
SO4 2-
S2O3 2-
SoxYZXA
Sulphate reduction
Nitrous oxide reduction
N2O → N2 (NosZ)
DO (mg/L) 2.75 >3
Salinity (PSS) 18.69 >35
8 m
Chitin
Phenol
MAG-2: Thioglobus sp. [99.34%]
hexosaminidase
UbiX
Catabolism
Sulphide oxidation
Fermentation
Acetate
Acs
Acetyl-CoA
L-cysteine
cysM
H2S
Sqr
S
Sulphur oxidation
dsrAB
dsrC
dsrMKOP
dsrH
dsrRS
Sdo
Sulphite reduction
Nitrate reduction
NO3- → NO2- (NarGH, NapAB)
NO2- → NH4 (NirB)
NO2- → NO (NirKS)
NO → N2O (NorBC)
N2O → N2 (NosZ)
Denitrification
Nitrite reduction
SO3 2-
Sat
AprAB
SO4 2-
Thiosulphate oxidation
Sulphate reduction
SoxYZXA
SoxB
S2O3 2-
Arsenate reduction
AsO4 3- → AsO3 3- (ArsC)
0.86
25.46
17 m
Phenol
MAG-107: Nitrosopumilus sp. [87.54%]
Cellulose
MAG-158: Marinisomatales sp. [57.14%]
0.33
31.41
18 m
UbiX
Catabolism
Sulphate reduction
SO4 2- → SO3 2- (Sat)
Sulphur oxidation
S (Sdo)
Fermentation
Acetate
Acs
AcdA
Acetyl-CoA
Nitrite reduction
NO2- → NO (NirKS)
NO2- → NH4 (NirD)
NO → NH3 (NirA)
NH4 → NH3 (AmoA)
Ammonia oxidation
cellulase
Catabolism
Selenate reduction
SeO4 2- → Se0 (XdhD, YgfK)
0.46
31.70
22 m
Dissimilatory arsenate reduction
As5+ → As3+ (ArrA)
Arsenite methylation
As3+ → Dimethylarsenate (AsM)
Nitrate reduction
NO3- → NO2- (NarG)
0.47
31.91
28 m
0.28
32.35