Microbial dark matter filling the niche in hypersaline microbial mats

Hon Lun Wong[1,2], Fraser I. MacLeod[1,2], Richard Allen White III[2-4], Pieter T. Visscher[2,5,6], Brendan P. Burns[1,2*]

[1]School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, Australia

[2]Australian Centre for Astrobiology, University of New South Wales, Sydney, Australia

[3]RAW Molecular Systems LLC, Spokane, WA, United States

[4]Plant Pathology, Washington State University, Pullman, WA, United States

[5]Department of Marine Sciences, University of Connecticut, USA

[6]Biogeosciences, the Université de Bourgogne Franche-Comté, Dijon, France

*Corresponding author:

Dr Brendan P. Burns

School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, 2052, Australia.

Phone: 612 93853659.

Email: brendan.burns@unsw.edu.au

Abstract

Shark Bay, Australia, harbours one of the most extensive and diverse systems of living microbial mats, that are proposed to be analogs of some of the earliest ecosystems on Earth. These ecosystems have been shown to possess a substantial abundance of uncultivable microorganisms. These enigmatic groups - 'microbial dark matter' (MDM) - are hypothesised to play key roles in microbial mats. We reconstructed 115 metagenome-assembled genomes (MAGs) affiliated to MDM, spanning 42 phyla within the bacterial and archaeal domains. We classified bacterial MDM from the PVC group, FCB group, Microgenomates, Parcubacteria, and Peregrinibacteria, as well as a high proportion of archaeal MDM under the TACK, DPANN, Altiarchaeales, and Asgard archaea. The latter includes the first putative Heimdallarchaeota MAG obtained from any microbial mat system. This study reports novel microorganisms (Zixibacterial order GN15) putatively taking part in dissimilatory sulfate reduction in surface hypersaline settings, as well as novel eukaryote signature proteins in the Asgard archaea. Despite possessing reduced-size genomes, the MDM MAGs are capable of fermenting and degrading organic carbon, suggesting a role in recycling organic carbon. Several forms of RuBisCo were identified, allowing putative CO_2 incorporation into nucleotide salvaging pathways, which may act as an alternative carbon and phosphorus source. High capacity of hydrogen production was found among Shark Bay MDM. Putative schizorhodopsins were also identified in Parcubacteria, Asgard archaea, DPANN archaea, and Bathyarchaeota, allowing these members to potentially capture light energy. Diversity-generating retroelements were prominent in DPANN archaea that likely facilitate the adaptation to a dynamic, host-dependent lifestyle. In light of our findings, we propose H_2, ribose and CO/CO_2 as the main energy currencies of the MDM community in these mat systems.

INTRODUCTION

A vast 'known-unknown' and even 'unknown-unknown', many microorganisms have yet to be unlocked from a majority of Earth's ecosystems. These uncultured microbial community members represent a vast untapped and uncharacterised resource of biological information, representing the 'microbial dark matter' (MDM) of many microbial ecosystems [1, 2]. Despite the majority of these unexplored lineages have reduced genome sizes and minimal biosynthetic capacity, they were proposed to represent a major uncharacterised portion of microbial diversity and inhabiting every possible metabolic niche [1-5].

Advances in microbial dark matter research have the potential to alter our understanding of key tenets of evolutionary principles, such as the current debate over the phylogenetic position of the Asgard archaea, and the idea that the eukaryotic cell emerged from within the archaeal domain [6-8]. It is estimated that at least 80% of environmental genomic content could be considered as 'genomic dark matter' [5, 9-11], with the majority found in subsurface environments [4, 12-15]. There is indeed a risk that many novel and unexpected organisms and/or their gene products may simply be lost in large datasets. Therefore, this prompted the desire to uncover unknown genes, functions, and ecological roles of these novel groups in other microbial ecosystems [16], of particular interest for the present study, hypersaline microbial mats.

Shark Bay, in Western Australia, harbors one of the most extensive (and diverse) microbial mat systems in the world that are analogs of some of the earliest ecosystems on Earth. Hypersalinity, heatwaves, high UV radiation, oligotrophic waters, fluctuating tides, and even cyclonic events, contribute to the mats being subjected to an extreme environment in Shark

Bay [17-21]. Indeed recent findings have shown that these mats are subjected to extreme fluctuations over a diel cycle as tides change [22-24], both at the surrounding microenvironment level (large changes in temperature, salinity, pH) as well as at the metabolic level (rapid changes in O and S levels reflective of changing rates of photosynthesis, respiration, and sulfur metabolisms). Despite recent advances made in our understanding of Shark Bay mat taxonomic and functional complexity [17, 19, 20, 22-24], the diversity and ecological role of MDM in these evolutionarily significant systems is unknown. Amplicon sequencing revealed that MDM comprises over 15% of the bacterial community and over half of the archaeal population in Shark Bay mats [22, 23]. Although these results have indicated Shark Bay microbial mats are a huge genetic pool of novel lineages, their functional role(s), including how they adapt to such an extreme environment and their putative interactions with other microorganisms, are still unknown.

We hypothesize that microbial dark matter in microbial mats may be key in nutrient cycling, symbioses, and overall health of these systems under extreme environmental conditions. In this study we have unravelled in detail the metabolic potential and capacities of this enigmatic group of novel microorganisms in Shark Bay mats.

RESULTS AND DISCUSSION

Microbial dark matter metagenome-assembled genomes (MAGs). This study describes for the first time in detail MAGs associated with microbial dark matter in hypersaline microbial mats. In total, 115 MAGs were found in Shark Bay mat metagenomic data [24], spanning 42 phyla within the bacterial and archaeal domains (Fig. 1). MAGs that were classified as part of microbial dark matter in previous literature were chosen in this study [1,

3-6, 12, 13, 25 ,26]. Genome statistics are provided in Additional file 13: Table S1. Of the 115 MAGs, 24 high-quality MAGs were obtained (> 90% completeness, < 5% contamination, encoding at least 18 out of 20 amino acids) and the remaining were of medium quality MAGs (> 50% completeness, < 10% contamination) based on recently established standards [27]. Although one Heimdallarchaeota (Bin_120) and one Lokiarchaeota (Bin_186) MAG had slightly over 10% contamination levels (both with 10.75%), they are included in this study due to high completeness (> 85%), as well as this being the first Heimdallarchaeota MAG obtained from any microbial mat system. Of the 91 medium quality MAGs, 65 have > 70% completeness. Bacterial MAGs were further classified into the PVC group, FCB group, Microgenomates, Parcubacteria, Peregrinibacteria, and 'others' (Fig. 1). Archaeal MAGs were classified into Asgard archaea, TACK, DPANN, and Altiarchaeales. It is still under debate whether Altiarchaeales should be placed within the DPANN superphylum [28-30], so these were placed as a separate group in the current analyses.

Distribution of novel rhodopsins and eukaryotic signature proteins (ESPs). Rhodopsin genes were identified in Asgard archaea, Parcubacteria, Bathyarchaeota, and DPANN archaeal MAGs. Phylogenetic analysis of the rhodopsins showed affiliation with schizorhodopsins, which was recently found in Asgard archaea in a microoxic niche setting [31] (Additional file 2: Figure S1 and Additional file 14: Table S2). This study expanded the range of phyla encoding schizorhodopsin, which has not been identified in Bathyarchaeota, DPANN archaea, and Parcubacteria previously. The novel rhodopsin may infer light sensitivity in these microbial groups with a recent study indicating Asgard archaeal schizorhodopsins as light-driven H^+ pumps [32]. Rhodopsins in MDM (Saccharibacteria; Asgard archaea) have only been found in hypersaline environments to date, suggesting the

122 acquisition of such an evolutionary adaptation of this enigmatic group may be a feature in
123 hypersaline, sunlit surface settings, with the ability to utilize light energy to counter the high
124 cost of osmotic balance maintenance [31, 33, 34].

125

126 Eukaryotic signature proteins (ESPs) were found distributed in the ten Asgard archaeal
127 MAGs (Additional file 1: Supplementary Information and Additional file 3: Figure S2),
128 hinting a close evolutionary relationship between Asgard archaea and eukaryotes [7, 8, 32,
129 35]. ESPs in Shark Bay Asgard archaeal MAGs are putatively involved in cytoskeleton
130 dynamics, information processing, trafficking machinery, signalling systems and N-linked
131 glycosylation [7, 8]. Novel ESPs were identified affiliated with information processing and
132 GTP binding proteins belonging to ARF, RAS, RAB and RAN families in Asgard archaea for
133 the first time, suggesting that this archaeal superphylum possess a diverse range of
134 eukaryotic-like signalling systems (Additional file 3: Figure S2).

135

136 **Limited metabolic capacities and putative host-dependent lifestyle**. Most of the MDM
137 MAGs harbor incomplete metabolic pathways despite the oligotrophic condition present in
138 Shark Bay (Additional file 1: Supplementary Information), hence are suggested to be host-
139 dependent [29, 36]. We screened the MAGs for diversity-generating retroelements (DGRs),
140 which are fast-evolving proteins enabling host-dependent microorganisms to attach to the
141 hosts' surface [30, 37-39], and they were identified mostly in Parcubacteria and DPANN
142 archaea (Fig. 1 and Fig. 2). Interestingly, despite having versatile metabolic pathways, DGRs
143 were also identified in Asgard archaeal MAGs in this study, which has not been reported
144 before (Fig. 1, Fig. 2 and Additional file 1: Supplementary Information). This may indicate
145 Asgard archaea once resided in energy-limited environments which relies on symbiotic

146 relationships [32, 37]. Parcubacteria and DPANN archaeal members lack biosynthetic
147 capabilities and were suggested to have a symbiotic, host-dependent lifestyle, in which DGRs
148 facilitate adaptation to rapidly changing environments, providing them with the tools to
149 exploit symbiotic associations with their host [1, 29, 30, 40].

150

151 It is hypothesised that MDM may take advantage of the lack of virus defence systems
152 observed here (CRISPR, BREX, DISARM and DNA phosphorothioation; Additional file 1:
153 Supplementary Information and Additional file 15: Table S3), despite the high abundance of
154 viruses and virus defence systems identified in Shark Bay in previous studies [24, 41]. Given
155 the limited metabolic capacities and host-dependent lifestyle, these organisms may serve as
156 'viral decoys' to alleviate the load on the host's viral defence system, and benefits from the
157 viral DNA as a pentose and phosphorus source [42]. Another advantage is to avoid
158 autoimmunity, in which the virus defence systems may attack the host [42, 43]. Furthermore,
159 maintaining such systems is energetically costly [44]. It is proposed that the synergy between
160 the lack of viral defence systems and the presence of DGRs facilitate rapid screening and
161 acquisition of biological functions for survival in these mat systems, allowing MDM with
162 limited metabolic pathways to adapt to a dynamic and extreme environment.

163

164 **Sulfur, nitrogen, and carbon cycling in Shark Bay MDM**. Despite having reduced
165 metabolic capacity, various genes within sulfur, nitrogen, and carbon cycles were found
166 distributed among the Shark Bay microbial dark matter community. All Zixibacteria and
167 Zixibacteria order GN15 (previously classified as candidatus phylum GN15) genomes
168 harbour dissimilatory (bi)sulfite reductase (*dsrAB*) and adenylylsulfate reductase (*aprAB*)
169 genes affiliated with dissimilatory sulfate reduction [45, 46] (Additional file 4: Figure S3 and

Additional file 15: Table S3). MAGs of unclassified bacterium BMS3Bbin04 (FCB Group) also harbour *dsrAB*, while Coatesbacteria encode for an *aprA* gene, suggesting these bacteria may play a role in the sulfur cycle. To confirm their role(s) in the sulfur cycle, *dsrAB* genes were analysed against the *dsrAB* database [47]. The genes were classified as a reductive bacterial type, confirming their likely function as dissimilatory (bi)sulfite reductase (Additional file 5: Figure S4). The *dsrAB* genes clustered with uncultured lineages in estuarine environments and interestingly, one arctic *dsrAB* lineage, suggesting a specific lineage in Shark Bay (Additional file 5: Figure S4). Furthermore, *dsrC* genes were found co-localised on the same contigs in the *dsrAB* encoding MAGs (Additional file 15: Table S3), which is an essential physiological partner to *dsrAB* in sulfite reduction [46]. This suggests Zixibacteria (and order GN15) potentially partake in dissimilatory sulfate reduction not only in deep subsurface [46, 48, 49], but also in surface hypersaline environments. Genes *dsrEFH* were also identified in 40 MAGs in the present study, expanding the lineages taking part in sulfur cycling (Additional file 1: Supplementary Information). Previous studies in these mat systems have indicated a putative surface anoxic niche with high rates of sulfur cycling [22, 24], and potentially microbial dark matter could be contributing to this phenomenon.

Only a limited number of genes involved in nitrogen cycling were identified in MDM (Fig. 1, Fig. 3 and Additional file 15: Table S3). Genes encoding nitrite reductase were found in all Asgard archaea MAGs except Heimdallarchaeota (Fig. 2 and Additional file 15: Table S3). The co-occurrence of CO dehydrogenase and nitrite reductase suggests that Asgard archaea may potentially couple CO oxidation to nitrite reduction [50], allowing them to derive energy from an oligotrophic environment (Fig. 2, Fig. 4 and Additional file 15: Table S3).

To infer the capacity of carbohydrate degradation in microbial mat microbial dark matter, we analysed MAGs for carbohydrate-active enzymes (CAZy). Asgard archaea and MAGs affiliated within the FCB group have the broadest cassette of glycoside hydrolase (GH) genes, (hemicellulose, amylase, animal and plant polysaccharides), indicating a highly flexible metabolic capacity for carbon acquisition (Fig. 3, Additional file 1: Supplementary Information and Additional file 6: Figure S5). On the other hand, Parcubacteria, Microgenomates, Peregrinibacteria, and DPANN archaea encode a lower range of GH enzymes, suggesting these members could scavenge readily degraded carbohydrates through their potential symbiotic hosts or partners. Most of the microorganisms identified in this study are likely capable of fermenting various carbon sources into formate, acetate, lactate, and ethanol **(**Fig. 3, Fig. 4 and Additional file 15: Table S3). This finding suggests that most MDM MAGs undergo anoxic carbon transformation, corroborating with previous studies [13, 14, 40, 51-53].

Genes encoding anaerobic carbon monoxide dehydrogenase (*cooSF*) and acetyl-CoA synthase (*cdhDE*, *acsB*) were identified in FCB group MAGs (Modulibacteria, KSB1, Fibrobacteres) (Additional file 4: Figure S3), Asgard archaea (Heimdall-, Loki-, Thorarchaeota) (Fig. 4) and Bathyarchaeota (Additional file 7: Figure S6) MAGs, indicating their putative ability to fix and reduce CO_2 to acetyl-CoA through the Wood-Ljundahl (WL) pathway (Fig. 2 and Additional file 15: Table S3). Generally, microorganisms can either use tetrahydrofolate (THF) or tetrahydromethanopterin (THMPT) as C_1 carriers in the WL pathway [54, 55]. Usually, bacteria utilise the THF-WL pathway while THMPT-WL pathways are mostly found in archaea [54-57]. One Lokiarchaeota MAG (Bin_186) harbours a complete anaerobic H_2-dependent THMPT-WL pathway, inferring the ability to fix CO_2 and H_2 into acetate [58] (Fig. 4 and Additional file 15: Table S3). All other Asgard archaea in

the Shark Bay mats contained most genes for both WL pathways, though incomplete genomes may account for the absence of these genes (Additional file 15: Table S3).

The presence of the Wood-Ljungdahl pathway is suggested to be a result of energy limitation since it is energetically inexpensive compared to other carbon fixation pathways [45, 58, 59]. Up to a quarter of the MAGs in the present study encode for CO dehydrogenase, allowing CO to be putatively utilised [60, 61]. Given the high UV radiation the Shark Bay mats are exposed to, CO may be produced through photo-degradation and subsequently oxidised as an alternative carbon source for energy conservation [24, 62, 63].

Furthermore, based on the observed genomic repertoires, Asgard archaea in these mats are putatively heterotrophic acetogens, encode for a complete beta-oxidation pathway and may take part in the carbon fixing 4-hydroxybutyrate pathway (Additional file 1: Supplementary Information). The MDM community has scattered genes in other carbon metabolisms but encode peptidases, putatively facilitating scavenging organic carbon in their oligotrophic environment (Additional file 1: Supplementary Information).

Nucleotide salvaging and putative CO_2 assimilation. Surprisingly, despite the reduced-sized genomes, 32 MAGs encode for ribulose biphosphate carboxylase (RuBisCo) (Fig. 2, Additional file 15: Table S3 and Additional file 16: Table S4). Given not all types of ribulose biphosphate carboxylase undergo carbon fixation, a phylogenetic tree was constructed to examine the variety of RuBisCo in these mat metagenomes. The MDM MAGs appear to harbour bacterial and archaeal type III, type IIIa, type IIIb, type IIIc, and type IV RuBisCo (Fig. 5). Furthermore, all MAGs have incomplete CBB (Calvin-Benson-Bassham) cycle

(Additional file 15: Table S3). This suggests that these microorganisms are involved in the AMP nucleotide salvaging pathway, while MAGs harbouring type IV RuBisCo are involved in methionine salvage pathways [64, 65]. Interestingly, 22 out of the 32 MAGs with RuBisCo also encode both AMP phosphorylase (*deoA*) and R15P isomerase (*e2b2*) (Additional file 15: Table S3), indicating the potential ability to incorporate CO_2 into nucleotide salvaging pathways [52, 65-67]. One Lokiarchaeota MAG (Bin_186) encodes for a type IIIa RuBisCo, which is known to fix CO_2 through the reductive hexulose-phosphate (RHP) cycle [65, 66]. Although it potentially lacks the ability to fix CO_2 due to the absence of homologs to genes encoding phosphoribulokinase, this MAG encodes for a fused bifunctional enzyme 3-hexulose-6-phosphate synthase/formaldehyde-activating enzyme (*fae-hps*) allowing for the potential production of methylene-H_4MPT, which may play a role in replenishing the C_1 carriers in the THMPT-WL pathway [66].

One interesting finding is that Heimdallarchaeota (Bin_120) contains RuBisCo at the basal position (Fig. 5), suggesting it may possess RuBisCo as an early-evolved form. The widespread distribution of RuBisCo among MDM in Shark Bay mats implies the use of ribose to substitute upper glycolysis, as some of the key genes in this pathway are missing [52] (Additional file 1: Supplementary Information). Other than feeding ribose (and putatively CO_2) as augmented carbon sources into the central carbon metabolism, these non-autotrophic RuBisCo may putatively free phosphate groups from nucleotides to supplement the extremely limited phosphorus in Shark Bay found in previous studies [24, 68]. In this study the genomes harbouring RuBisCo were identified across 14 MDM phyla, suggesting that ribose may be a prominent currency among microbial dark matter in hypersaline microbial mats (Fig. 4, Fig. 5, Additional file 7-11: Figure S6-S10).

High capacity for hydrogen production among Shark Bay MDM. A total of 267 hydrogenases were detected in 81 out of 115 MDM MAGs, implying prominent hydrogen metabolism in MDM even with minimal genomes (Fig. 4, Additional file 4: Figure S3, Additional file 7-11: Figure S6-10 and Additional file 15: Table S3). A total of 16 types of hydrogenases were identified against the HydDB database [69], including 10 [NiFe] and 6 [FeFe] classes (Fig. 2, Additional file 15: Table S3 and Additional file 1: Supplementary Information). Most of the hydrogenases identified are putatively involved in H_2 uptake (Group 1), H_2 consumption/production, fermentative H_2-evolving, and H_2 sensing (Group 3, 4 and [FeFe]) (Fig. 2).

Almost half of the Shark Bay MAGs harbour hydrogenases associated with H_2 production, and of particular significance, Parcubacteria and Woesearchaeota only encode H_2 producing hydrogenases ([NiFe]-3b and [FeFe] Group A), which are fermentative in nature [70] (Fig. 2, Fig. 4 and Additional file 15: Table S3). H_2 production is potentially an important energy currency in these mats as hydrogenotrophic methanogenesis was found to be the prominent mode of methane production [23]. Furthermore, a global survey suggests that Woesearchaeota form consortiums with hydrogenotrophic methanogens by providing H_2 in exchange of nutrients [71]. It is suggested that these MAGs (especially among Parcubacteria and Woesearchaeota) support and complement H_2/CO_2 methanogenesis in Shark Bay microbial mats.

One-third of the MAGs (43 out of 115) encodes for 3b and 3c hydrogenases, which play essential roles as electron donors and H_2 production during hydrogenogenic fermentation and

Wolfe cycle of methanogenesis [72, 73]. Of particular interest, the presence of the WL pathway along with hydrogenase group 3b and 3c in Asgard archaea suggests this group are putatively lithoautotrophs that use H_2 as electron donors [58, 74]. With a range of CAZy enzymes distributed among Shark Bay MDM (Additional file 6: Figure S5), these microorganisms likely participate in anoxic carbon transformations and hydrogen turnover [12, 14, 40, 75-77]. Therefore, MDM in these systems may act as a 'recycler' in the mats to recycle organic carbon from dead cells, employing hydrogenogenic or hydrogenotrophic metabolisms.

Energy Currencies of MDM. As described in an earlier study [24], it is likely that the WL-pathway is the main mode of carbon fixation in these mats, and the surface phototrophic consortia produce the energy and organic carbon for the rest of the microbial community [22, 78]. Various adaptation strategies to the hypersalinity, limited phosphorus, and high copper concentration were described [24] (Additional file 1: Supplementary Information). However, given the oligotrophic nature of Shark Bay waters [79], MDM in these mats that lacks the metabolic capacity may utilise alternative carbon sources to augment nutrient intake. First, it is proposed that due to the high UV irradiation in Shark Bay, photo-degradation of surface organic matter may provide CO as an alternative carbon source [62, 63]. Secondly, the widespread hydrogenases among MDM may contribute to the hydrogen turnover in exchange of nutrients as high rates of hydrogenotrophic methanogenesis were measured and detected in these mats [23]. Thirdly, RuBisCo found in the MDM MAGs is proposed to fix CO_2 alongside nucleotide salvaging, which is subsequently fed into glycolysis, maximising energy yield [52, 66, 80, 81]. It is therefore proposed in an ecological context, MDM occupies metabolic niches in Shark Bay microbial mats where ribose, H_2, CO and CO_2 are prominent currencies to augment energy income.

Conclusions. This is the first study to reconstruct and describe in detail high-quality genomes affiliated with microbial dark matter in microbial mats. This study reports the novel uncultured bacterial phyla Zixibacteria (including order GN15) and an unidentified bacterium (Bin_338) as likely participants in dissimilatory sulfate reduction in surface hypersaline settings, as well as diversity generating retroelements and novel ESPs identified in Asgard archaea. It is suggested that Asgard archaea are not only organoheterotrophs, but also putatively lithoautotrophs that have more versatile metabolic capacities than the other groups of MDM, possessing both THMPT- and THF-WL pathways, RuBisCo and schizorhodopsin. For the other MDM groups, although possessing minimal genomes and the lack of complete biosynthetic pathways, they are potentially capable of degrading and fermenting organic carbon and are suggested to play a role in H_2 and carbon transformation in microbial mats. Various forms of RuBisCo were encoded, allowing putative CO_2 incorporation into nucleotide salvaging pathways, acting as an alternative carbon and phosphorus source. Despite possessing minimal genomes, DGRs were prominent in Parcubacteria and DPANN archaea to likely adapt to a dynamic, host-dependent environment. Under the oligotrophic environment in Shark Bay, MDM needs to exploit every opportunity for energy generation, such as harbouring scattered genes of various nutrient cycles to fill in metabolic gaps or function in "filling the niches" and as is the case for some other ecological systems [12, 30, 82] (Additional file 1: Supplementary Information). On the other hand, MDM in Shark Bay may be shaping the mat environment through their various metabolic capacities in a process called niche construction, modifying their own and each other's niches and functional roles in the ecosystem [16, 83, 84]. It has in fact been recently suggested that early ecosystems such as microbial mats were not nutrient starved but rather limited by electron donor/acceptor availability [85], thus the ability for these ecosystems to maximize energy yielding capacities

is evolutionary advantageous. A conceptual ecological model of MDM in Shark Bay mats is shown in Fig. 6, proposing the MDM serves to fill in the metabolic gaps. Ribose, CO_2/CO, and H_2 are suggested to be prominent currencies among MDM in these mats and were potentially a widespread phenomenon on early Earth [40, 86].

MATERIALS AND METHODS

Sampling and metagenomic sequencing. Microbial mat sampling from Shark Bay was performed in a previous study [22], and DNA extraction and sequencing of total community DNA are described previously [24]. Metagenomes were analysed from smooth mats in the present study. The Fastq sequencing data files obtained from the Illumina NextSeq platform detailed by Wong and colleagues [24] were used in the present study for detailed analysis of microbial dark matter.

Assembly, binning and phylogenetic analyses. Low-quality bases (per base sequence quality < 28) from each sequencing file were trimmed and examined using Trimmomatic (version 0.36) and FastQC (version 0.11.6) respectively [87, 88]. All sequencing files from the ten layers of smooth mats were co-assembled as described [89] (minimum kmer 27, with incremental kmer set as 10) using Megahit version 1.1.1 [90]. Subsequently, all contigs with length less than 2000 bp were removed to avoid ambiguous contig annotation of shorter contigs. This step is to avoid misinterpretation of novel annotated contigs. An alignment algorithm, BWA-MEM (version 0.7.7), was used to map reads back to the assembled contigs [91]. SAMtools version 1.3.1 was used to convert SAM files to binary format BAM files [92]. MetaBAT2 (Version 2.12.1), MaxBin2 (Version 2.2.3) and CONCOCT (Version 1.0.0) were applied for metagenomic binning [93-95]. Subsequently, DAS Tool was used to refine and

365 filter lower quality MAGs generated from the three binning programs [96]. CheckM was
366 employed to examine the quality (completeness and contamination level of MAGs), and
367 MAGs statistics were obtained through QUAST and tRNAScan-SE [97-99]. MAGs with at
368 least medium quality (>50% completeness, <10% contamination) were selected in this study
369 [27]. Subsequently, the taxonomy of MAGs was determined with GtDb-tk [100], in which
370 only MAGs classified as microbial dark matter in previous literatures were chosen [1, 3-6,
371 12 ,13, 25, 26].

372

373 Maximum-likelihood-based phylogenetic trees based on 16 concatenated ribosomal proteins
374 (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) were constructed as described
375 in Hug et al (2013) [101]. Only MAGs with at least 8 ribosomal proteins were included in the
376 analysis. Bin_245 (Bathyarchaeota) was not included in the phylogenetic tree (Fig. 1) as it
377 has less than 8 ribosomal proteins. Phylosift version 1.0.1 was used to extract ribosomal
378 proteins from the genomes [102]. Subsequently, ribosomal protein sequences were aligned
379 using MAFFT version 7.310 [103]. BMGE was then used to remove gaps in the alignment
380 with BLOSUM30 matrix and gap rate cut-off of 50% [104]. The resulting protein alignments
381 were concatenated as described in Hug et al (2013) [101]. The concatenated ribosomal
382 proteins were then used to construct a phylogenetic tree using IQ-TREE version 1.6.1 with a
383 total of 1000 bootstrap replicates, which the output file was visualised with iTOL [105, 106].

384

385 **Functional annotation**. Nucleotide contigs of metagenome-assembled genomes were
386 translated to amino acid sequences by employing Prodigal version 2.6.3 [107]. Functional
387 annotation was carried out using GhostKoala to assign amino acid contigs to MAGs against
388 the KEGG database [108]. InterProScan version 5.25-64.0 was employed to annotate protein

domains of MAGs to PFAM and TIGRfam databases, with cutoff value <$1e^{-10}$ [109]. Contigs were annotated against the CAZy database to identify carbohydrate-active enzymes in the MAGs [75]. DGRScan was used to identify diversity-generating retroelements (DGR) among the MAGs [110]. Hydrogenase sequences derived from KEGG and PFAM databases were extracted and annotated against HydDB to further classify hydrogenases [69]. ESP and rhodopsin sequences were submitted to HHPred [111] to confirm their identity.

Phylogenetic analysis of RuBisCo, Rhodopsin and *dsrAB*. To determine the type of RuBisCo identified in MAGs presented in this study, RuBisCo sequences were downloaded from NCBI and ggkBase [52] (http://ggkbase.berkeley.edu). Reference sequences of RuBisCo and rhodopsins are listed in Additional file 16: Table S4 and Additional file 14: Table S2. Reference sequences of dissimilatory sulfate reduction *dsrAB* were obtained from the dsrAB reference database [47]. RuBisCo, rhodopsin, and *dsrAB* sequences were aligned with MAFFT version 7.310 [103], with gaps subsequently removed by UGENE [112]. IQ-TREE version 1.6.1 was employed to construct a phylogenetic tree with a total of 1000 bootstrap replicates and visualised with iTOL [105, 106]. To further confirm the identity of RuBisCo and rhodopsins, the sequences were annotated against the HHpred [111] and the BLAST database [113].

ACKNOWLEDGEMENTS

The authors would like to thank the Ramaciotti Centre for Functional Genomics for valuable technical support. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

412 **REFERENCES**

413 1. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al.
414 Insights into the phylogeny and coding potential of microbial dark matter. Nature.
415 2013;499(7459):431-437.
416 2. Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, Martin HG, et al. Dissecting
417 biological "dark matter" with single-cell genetic analysis of rare and uncultivated
418 TM7 microbes from the human mouth. Proc. Natl. Acad. Sci. USA. 2007;104:11889-
419 11894.
420 3. Hedlund BP, Dodsworth JA, Murugapiran SK, Rinke C, Woyke T. Impact of single-
421 cell genomics and metagenomics on the emerging view of extremophile 'microbial
422 dark matter'. Extremeophiles. 2014;18(5):865-875.
423 4. Bruno A, Sandionigi A, Rizzi E, Bernasconi M, Vicario S, Galimberti A, et al.
424 Exploring the under-investigated "microbial dark matter" of drinking water treatment
425 plants. Sci. Reps. 2017;7:44350.
426 5. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured
427 microbial cells dominate Earth microbiomes. mSystems. 2018;3(5):e00055-18.
428 6. Saw JH, Spang A, Zaremba-Niedzwiedzka K, Juzokaite L, Dodsworth JA,
429 Murugapiran SK, et al. Exploring microbial dark matter to resolve the deep archaeal
430 ancestry of eukaryotes. Philos Trans Royal Soc B. 2015;370:1678-1687.
431 7. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L,
432 Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular
433 complexity. Nature. 2017;541(7637):353-358.
434 8. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martjin J, Lind AE, et al.
435 Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature.
436 2015;521:173-179.

437 9. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad
438 Sci. 2016;113(21):5970-5975.
439 10. Schulz F, Eloe-Fadrosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, et al.
440 Towards a balanced view of the bacterial tree of life. Microbiome. 2017;5:140.
441 11. Makarova KS, Wolf YI, Koonin EV. Towards functional characterization of archaeal
442 genomic dark matter. Biochemc Soc Trans. 2019;BST20280560.
443 12. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al.
444 Thousands of microbial genomes shed light on interconnected biogeochemical
445 processes in an aquifer system. Nat Comm. 2016;7:13219.
446 13. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual
447 biology across a group comprising more than 15% of domain bacteria. Nature.
448 2015;523:208-211.
449 14. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Small
450 genomes and sparse metabolism of sediment-associated bacteria from four candidate
451 phyla. mBio. 2013;4(5):e00708-00713.
452 15. Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ.
453 Members of the candidate phyla radiation are functionally differentiated by carbon-
454 and nitrogen-cycling capabilities. Microbiome. 2017;5(1):112.
455 16. Bernard G, Pathmanathan JS, Lannes R, Lopez P, Bapteste E. Microbial dark matter
456 investigations: how microbial studies transform biological knowledge and empirically
457 sketch a logic of scientific discovery. Genome Biol Evol. 2018;10(3):707-715.
458 17. Burns BP, Goh F, Allen M, Neilan BA. Microbial diversity of extant stromatolites in
459 the hypersaline marine environment of Shark Bay, Australia. Environ Microbiol.
460 2004;6(10):1096-1101.

18. Wong HL, Ahmed-Cox A, Burns BP. Molecular ecology of hypersaline microbial mats: current insights and new directions. Microorganisms. 2016;4(1):6.

19. Ruvindy R, White III RA, Neilan BA, Burns BP. Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. ISME J. 2016;10(1):183-196.

20. Allen MA, Goh F, Burns BP, Neilan BA. Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geology. 2009;7:82-96.

21. Reinold M, Wong HL, MacLeod FI, Meltzer J, Thompson A, Burns BP. The vulnerability of microbial ecosystems in a changing climate: Potential impact in Shark Bay. Life 2019;9:71.

22. Wong HL, Smith DL, Visscher PT, Burns BP. Niche differentiation of bacterial communities at a millimetre scale in Shark Bay microbial mats. Sci Reps. 2015;5:15607.

23. Wong HL, Visscher PT, White III RA, Smith DL, Patterson MM, Burns BP. Dynamics of archaea at fine spatial scales in Shark Bay mat microbiomes. Sci Reps. 2017;7:46160.

24. Wong HL, White III RA, Visscher PT, Charlesworth JC, Vázquez-Campos X, Burns BP. Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. ISME J. 2018;12:2619-2639.

25. Solden L, Lloyd K, Wrighton KC. The bright side of microbial dark matter: lessons learned from the uncultivated majority. Curr Opin Microbiol. 2016;31:217-226.

26. Thrash JC, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN, et al. Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico “Dead Zone”. mBio. 2017;8(5):e01017-17.

27. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TB, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35(8):725-731.

28. Spang A, Caceres EF, Ettem, TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science. 2017;357(6351):eaaf3883.

29. Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell. 2018;172(6):1181-1197.

30. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 2018;16:629-645.

31. Bulzu PA, Andrei AŞ, Salcher MM, Mehrshad M, Inoue K, Kandori H, et al. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. Nat Microbiol. 2019;4:1129-1137.

32. Inoue K, Tsunoda SP, Singh M, Tomida S, Hososhima S, Konno M, et al. Schizorhodopsins: A family of rhodopsins from Asgard archaea that function as light-driven inward H^+ pumps. Science Advances. 2020;6(15):eaaz2441.

33. Pushkarev A, Inoue K, Larom S, Flores-Uribe J, Singh M, Konno M, et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature. 2018;558:595-599.

34. Vavourakis CD, Andrei AŞ, Mehrshad M, Ghai R, Sorokin DY, Muyzer G. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome. 2018;6(1):1-18.

35. López-García P, Moreira D. Cultured Asgard archaea shed light on eukaryogenesis. Cell. 2020;181(2):3389-3402.

511 36. Sieber, JR, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the
512 paradigm for anaerobic metabolic cooperation. Annu Rev Microbiol. 2013;66:429-
513 452.
514 37. Paul BG, Bagby SC, Czornyj E, Arambula D, Handa S, Sczyrba A, et al. Targeted
515 diversity generation by intraterrestrial archaea and archaeal viruses. Nat Commun.
516 2015;6:6585.
517 38. Paul BG, Burstein D, Castelle CJ, Handa S, Arambula D, Czornyj E, et al.
518 Retroelement-guided protein diversification abounds in vast lineages of bacteria and
519 archaea. Nat Microbiol. 2017;2(6):17045.
520 39. Arnold C. Core concepts: How diversity-generating retroelements promote mutation
521 and adaptation in myriad microbes. Proc Natl Acad Sci. 2017;114(40):10509-10511.
522 40. Nelson WC, Stegen JC. The reduced genomes of Parcubacteria (OD1) contain
523 signatures of a symbiotic lifestyle. Front Microbiol. 2015;6:713.
524 41. White III RA, Wong HL, Ruvindy R, Neilan BA, Burns BP. Viral communities of
525 Shark Bay modern stromatolites. Front Microbiol. 2018;9:1223.
526 42. Burstein D, Sun CL, Brown CY, Sharon I, Anantharaman K, Probst AJ, et al. Major
527 bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat
528 Commun. 2016;7:10613.
529 43. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: gene
530 regulation or autoimmunity? Trends Genet. 2010;26:335-340.
531 44. Vale PF, Lafforgue G, Gatchitch F, Gardan R, Moineau S, Gandon S. Costs of
532 CRISPR-Cas-mediated resistance in Streptococcus thermophilus. Proc R Soc B.
533 2015;282(1812):1270.
534 45. Momper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a
535 deep terrestrial subsurface fluid microbial community. ISME J. 2017;11:2319-2333.

536 46. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, et al.
537 Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME
538 J. 2018;12:1715-1728.
539 47. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. Phylogenetic and environmental
540 diversity of DsrAB-type dissimilatory (bi) sulfite reductases. ISME J.
541 2015;9(5):1152-1165.
542 48. Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman K, et al.
543 Potential for microbial H_2 and metal transformations associated with novel bacteria and
544 archaea in deep terrestrial subsurface sediments. ISME J. 2017;11(8):1915-1929.
545 49. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and
546 biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat Commun.
547 2018;9(1):4999.
548 50. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs KU, Teske AP, et al. Genomic
549 inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. Nat
550 Microbiol. 2016;1(3):16002.
551 51. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, et al.
552 Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial
553 phyla. Science. 2013;337(6102):1661-1665.
554 52. Wrighton KC, Castelle CJ, Varaljay VA, Satagopan S, Brown CT, Wilkins MJ, et al.
555 RuBisCo of a nucleoside pathway known from archaea is found in diverse
556 uncultivated phyla in bacteria. ISME J. 2016;10(11):2702-2714.
557 53. Castelle CJ, Brown CT, Thomas BC, Williams KH, Banfield JF. Unusual respiratory
558 capacity and nitrogen metabolism in a Parcubacterium (OD1) of the candidate phyla
559 radiation. Sci Reps. 2017;7:40101.

560 54. MacLeod F, Kindler GS, Wong HL, Chen R, Burns BP. Asgard archaea: Diversity,
561 function, and evolutionary implications in a range of microbiomes. AIMS
562 Microbiology. 2019;5(1):48-61.
563 55. Liu Y, Zhou Z, Pan J, Baker BJ, Gu JD, Li M. Comparative genomic inference
564 suggests mixotrophic lifestyle for Thorarchaeota. ISME J. 2018;12(4):1021-1031.
565 56. Sousa FL, Martin WF. Biochemical fossils of the ancient transition from
566 geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism.
567 Biochim Biophys Acta. 2014;1837:964-981.
568 57. Seitz KW, Lazar CS, Hinrichs KU, Teske, AP, Baker BJ. Genomic reconstruction of a
569 novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and
570 sulfur reduction. ISME J. 2016;10(7):1696-1705.
571 58. Sousa FL, Neukirchen S, Allen JF, Lane N, Martin WF. Lokiarchaeon is hydrogen
572 dependent. Nat Microbiol. 2016;1(5):16034.
573 59. Hügler M, Sievert SM. Beyond the Calvin cycle: autotrophic carbon fixation in the
574 ocean. Ann. Rev. Mar. Sci. 3, 261-289 (2011).
575 60. Berg, I. A. (2011). Ecological aspects of distribution of different autotrophic CO_2
576 fixation pathways. Appl Environ Microbiol. 2011;7(6):1925-1936.
577 61. Ragsdale SW. Life with carbon monoxide. Crit Rev Biochem Mol. 2004;39:P165-195.
578 62. King GM. Carbon monoxide as a metabolic energy source for extremely halophilic
579 microbes: implications for microbial activity in Mars regolith. Proc Natl Acad Sci.
580 2015;112:4465-4470.
581 63. Vavourakis CD, Ghai R, Rodriguez-Valera F, Sorokin DY, Tringe SG, Hugenholtz P,
582 et al. Metagenomic insights into the uncultured diversity and physiology of microbes
583 in four hypersaline soda lake brines. Front Microbiol. 2016;7:211.

584 64. Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. Function, structure, and
585 evolution of the RuBisCo-like proteins and their RuBisCo homologs. Microbiol Mol
586 Biol Rev. 2007;71(4):576-599.
587 65. Ashida H. RuBisCo-like proteins as the enolase enzyme in the methionine salvage
588 pathway: Functional and evolutionary relationships between RuBisCo-like proteins
589 and photosynthetic RuBsiCo. J Exp Bot. 2008;59(7):1543-1554.
590 66. Kono T. A RuBisCo-mediated carbon metabolic pathway in methanogenic archaea.
591 Nat Commun. 2017;8:14007.
592 67. Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. Lateral gene transfer shapes the
593 distribution of RuBisCo among candidate phyla radiation bacteria and DPANN
594 archaea. Mol Biol Evol. 2018;36(3):435-446.
595 68. Atkinson MJ. Low phosphorus sediments in a hypersaline marine bay. Estuar Coast
596 Shelf Sci. 1987;24(3):335-347.
597 69. Søndergaard D, Pedersen CN, Greening C. HydDB: a web tool for hydrogenase
598 classification and analysis. Sci Reps. 2016;6:34212.
599 70. Greening C, Geier R, Wang C, Woods LC, Morales SE, McDonald MJ, et al. Diverse
600 hydrogen production and consumption pathways influence methane production in
601 ruminants. ISME J. 2019;13:2617-2632.
602 71. Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, et al. Insights into ecology,
603 evolution, and metabolism of the widespread Woesearchaeotal lineages. Microbiome.
604 2018;6(1):102.
605 72. Bryant FO, Adams MW. Characterization of hydrogenase from the hyperthermophilic
606 archaebacterium, Pyrococcus furiosus. J Biol Chem. 1989;264(9):5070-5079.
607 73. Thauer RK. The Wolfe cycle comes full circle. Proc Natl Acad Sci.
608 2012;109(38):15084-15085.

609 74. Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, et al. Proposal
610 of the reverse flow model for the origin of the eukaryotic cell based on comparative
611 analyses of Asgard archaeal metabolism. Nat Microbiol. 2019;4:1138-1148.
612 75. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The
613 carbohydrate-active enzymes database (CAZy) in 2013. Nucl Acids Res.
614 2013;42(D1):D490-D495.
615 76. Wrighton KC, Castelle CJ, Wilkins MJ, Hug LA, Sharon I, Thomas BC, et al.
616 Metabolic interdependencies between phylogenetically novel fermenters and
617 respiratory organisms in an unconfined aquifier. ISME J. 2014;8:1452-1463.
618 77. Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, et al.
619 Genomic expansion of domain archaea highlights roles for organisms from new phyla
620 in anaerobic carbon cycling. Curr Biol. 2015;25(6):690-701.
621 78. Fisher A, Wangpraseurt D, Larkum AW, Johnson M, Kühl M, Chen M. et al.
622 Correlation of bio-optical properties with photosynthetic pigment and microorganism
623 distribution in microbial mats from Hamelin Pool, Australia. FEMS Microbiol Ecol.
624 2018;95(1):fiy219.
625 79. Burns BP, Anitori R, Butterworth P, Henneberger R, Goh F, Allen MA, et al. Modern
626 analogues and the early history of microbial life. Precambrian Res. 2009;173(1-4):10-
627 18.
628 80. Aono R, Sato T, Yano A, Yoshida S, Nishitani Y, Miki K, et al. Enzymatic
629 characterization of AMP phosphorylase and ribose-1, 5-bisphosphate isomerase
630 functioning in an archaeal AMP metabolic pathway. J Bacteriol. 2012;194(24):6847-
631 6855.

632 81. Techtmann SM, Colman AS, Robb FT. "That which does not kill us only makes us
633 stronger": the role of carbon monoxide in thermophilic microbial consortia. Environ
634 Microbiol. 2012;11(5):1027-1037.
635 82. Dombrowski N, Seitz KW, Teske AP, Baker BJ. Genomic insights into potential
636 interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal
637 sediments. Microbiome. 2017;5(1):106.
638 83. Odling-Smee FJ, Laland KN, Feldman MW. Niche construction. Am Nat.
639 1996;147:641-648.
640 84. Laland K, Matthews B, Feldman MW. An introduction to niche construction theory.
641 Evol Ecol. 2016;30:191-202.
642 85. Ward LW, Rasmussen B, Fischer WW. Primary productivity was limited by electron
643 donors prior to the advent of oxygenic photosynthesis. J Geophys Res. 2019;124:211.
644 86. Schönheit P, Buckel W, Martin WF. On the origin of heterotrophy. Trends Microbiol.
645 2016;24(1):12-25.
646 87. Andrews C. FastQC: a quality control tool for high throughput sequence data.
647 Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
648 (2010).
649 88. Bolger A, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
650 sequence data. Bioinformatics. 2014;30(15):2114-2120.
651 89. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al.
652 Assembly of 913 microbial genomes from metagenomic sequencing of the cow
653 rumen. Nat Commun. 2018;9:870.
654 90. Li D, Liu CM, Luo R, Sadakane K, Lam TW. Megahit: an ultra-fast single-node
655 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
656 Bioinformatics. 2015;31(10):1674-1676.

657 91. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
658 transform. Bioinformatics. 2009;25:1754-1760.
659 92. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
660 alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-2079.
661 93. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
662 metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144.
663 94. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately
664 reconstructing single genomes from complex microbial communities. PeerJ.
665 2015;3:e1165.
666 95. Wu YW, Simmons BA, Singer SW. MaxBIN 2.0: an automated binning algorithm to
667 recover genomes from multiple metagenomic datasets. Bioinformatics.
668 2015;32(4):605-607.
669 96. Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of
670 genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat
671 Microbiol. 2018;3:836-843.
672 97. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing
673 the quality of microbial genomes recovered from isolates, single cells and
674 metagenomes. Genome Res. 2015;25(7):1043-1055.
675 98. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for
676 genome assemblies. Bioinformatics. 2013;29(8):1072-1075.
677 99. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer
678 RNA genes in genomic sequence. Nucl Acids Res. 1997;25(5):955-964.
679 100. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil
680 PA, et al. A standardized bacterial taxonomy based on genome phylogeny
681 substantially revises the tree of life. Nat Biotechnol. 2018;36:996-1004.

682 101. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et
683 al. Community genomic analyses constrain the distribution of metabolic traits across
684 the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome.
685 2013;1:22.
686 102. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. Phylosift:
687 phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:e243.
688 103. Katoh K, Standley DM. MAFFT multiple sequence alignment software
689 version 7: improvements in performance and usability. Mol Biol Evol.
690 2013;30(4):772-780.
691 104. Criscuolo A, Gribaldo S. BMGE (Block mapping and gathering with entropy):
692 a new software for selection of phylogenetic informative regions from multiple
693 sequence alignments. BMC Evol Biol. 2010;10(1):210.
694 105. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for
695 phylogenetic tree display and annotation. Bioinformatics. 2006;23(1):127-128.
696 106. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and
697 effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol
698 Biol Evol. 2014;32(1):268-274.
699 107. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
700 prokaryotic gene recognition and translation initiation site identification. BMC
701 Bioinformatics. 2010;11(1):119.
702 108. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG
703 tools for functional characterization of genome and metagenome sequences. J Mol
704 Biol. 2016;428(4):726-731.

705 109. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al.
706 InterProScan 5: genome-scale protein function classification. Bioinformatics.
707 2014;30(9):1236-1240.
708 110. Ye Y. Identification of diversity-generating retroelements in human
709 microbiomes. Int J Mol Sci. 2014;15(8):14234-14246.
710 111. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A
711 completely reimplemented MPI bioinformatics toolkit with a new HHpred sever at its
712 core. J Mol Biol. 2018;S0022-2836(17):30587-30589.
713 112. Okonechnikov K, Golosova O, Fursov M, Ugene Team. Unipro UGENE: a
714 unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-1167.
715 113. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al.
716 Gapped BLAST and PSI-BLAST: a new generation of protein database search
717 programs. Nucl Acids Res. 1997;25(17):3389-3402.

718

719 **Figures**

720 **Figure 1. Phylogenetic tree of novel MAGs of the MDM in Shark Bay microbial mats.**
721 Maximum likelihood phylogenetic tree of up to 16 concatenated ribosomal proteins (rpL2, 3,
722 4, 5, 6, 14, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) was constructed. Bin_245
723 (Bathyarchaeota) is not included in the tree as it has less than 8 ribosomal proteins. MAGs
724 assigned to different groups are highlighted in different colors on the outer circular stripe.
725 Circles represent genes for various nutrient cycles present in the MAGs (detailed in
726 Additional file 15: Table S3).

727

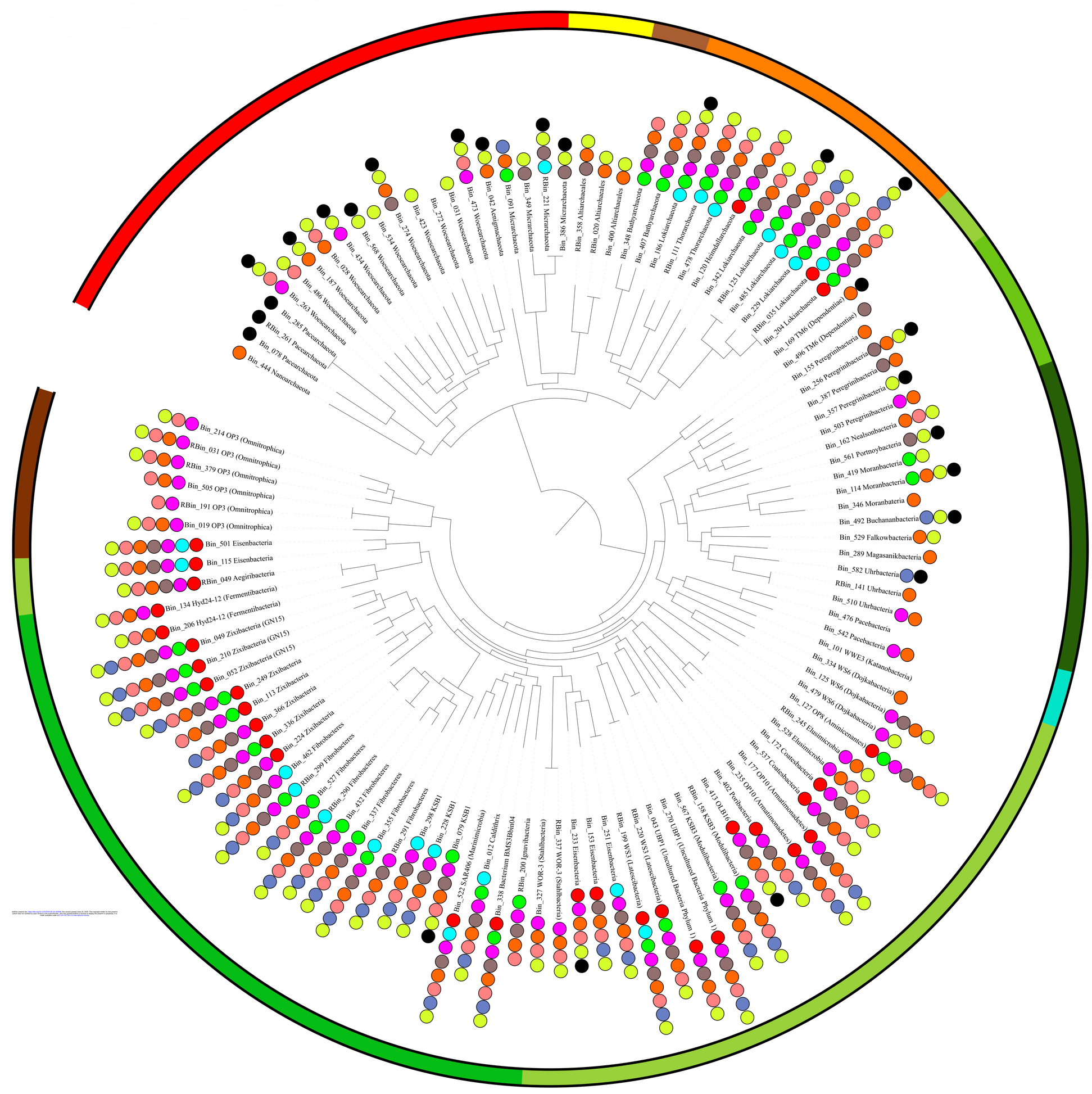
728 **Figure 2. Color-coded table indicating major functional genes identified in Shark Bay**
729 **mat novel microbiome MAGs.** X-axis indicates specific genes likely involved in either
730 nutrient cycling or environmental adaptation and y-axis indicates the specific microbial dark
731 matter MAGs. Key: Grey indicates the partial pathways identified in carbon, sulfur and
732 nitrogen cycles; white indicates the genes and associated pathways are absent. Color panel on
733 the left represents different groups of MDM MAGs according to Figure 1. ASR, assimilatory
734 sulfate reduction; DSR, dissimilatory sulfate reduction; SOX, sulfur oxidation; DNRA,
735 dissimilatory nitrate reduction; ANR, assimilatory nitrate reduction; rTCA, reverse
736 tricarboxylic cycle; WL pathway, Wood-Ljungdahl pathway; Inorg P, inorganic phosphorus;
737 Alkal Phos, alkaline phosphatase; copAB/cusAB, copper efflux systems; opu, osmoprotectant
738 transport system; pro, glycine betaine/proline transport system; bet, choline/glycine/proline
739 betaine transport protein; uvr, exinuclease; cph, cyanophycin; gltBD/glnAE, ammonium
740 assimilation; mtr, tetrahydromethanopterin S-methyltransferase; hdr, heterodisulfide
741 reductase; mttB, trimethylamine-corrinoid protein co-methyltransferase; DGR, diversity-
742 generating retroelements.

Figure 3. Putative involvement of Shark Bay MDM MAGs in carbon, sulfur, nitrogen and arsenic metabolisms. CAZy enzymes with different colored arrows representing various groups of glycoside hydrolase corresponding to Additional file 6: Figure S5. Numbers indicate the quantity and percentage of MAGs encoding for the nitrogen/sulfur cycles and metabolic pathways.

Figure 4. Metabolic potential of Asgard archaea, DPANN archaea, Parcubacteria and Microgenomates in the Shark Bay systems. A metabolic map summarising the genomic potential and metabolic capacities of MAGs affiliated with: **A.** Asgard archaea MAGs **B.** DPANN archaea MAGs **C.** Parcubacteria MAGs **D.** Microgenomates MAGs. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 15: Table S3. Different colors in the square boxes represent different numbers of MAGs encoding the genes, while white square boxes indicate the absence of the genes. TCA, tricarboxylic acid cycle; THF, tetrahydrofolate; THMPT, tetrahydromethanopterin; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Figure 5. Unrooted maximum-likelihood phylogenetic tree of RuBisCo genes in Shark Bay MDM MAGs. Maximum-likelihood phylogenetic tree constructed with RuBisCo gene found in the MDM MAGs with 1000 bootstrap replications. Archaeal and bacterial type III, type IIIa [66, 67], type IIIb [46], type IIIc [46] and type IV RuBisCo-like protein [64] were identified. Circular dots of different colors represent bootstrap values. RuBisCo sequences in this study and reference sequences are listed in Additional file 16: Table S4.

Figure 6. Proposed ecological model of Shark Bay mats (modified from Wong et al., 2018 [24], incorporating MDM). Green rectangular boxes indicate metabolic pathways, yellow rectangular boxes indicate putative pathway and orange rectangular boxes represent putative functional roles of MDM in these mats. Dark green ovals represent microorganisms and their corresponding functions [24] while orange ovals represent microbial dark matter. Red boxes encircle CO_2/CO, H_2 and ribose that are proposed as main energy currencies of Shark Bay MDM. Dashed arrows indicate putative metabolic exchange/microbial interactions. Dashed box includes genes involved in environmental adaptation. Chloro, Chloroflexi; Gemma, Gemmatimonadetes; HM, hydrogenotrophic methanogens; N/C storage, nitrogen/carbon storage; WL pathway, Wood-Ljungdahl pathway; 3HP/4HB pathway, 3-hydroxypropionate/4-hydroxybutyrate pathway; EPS, extracellular polymeric substance; DGR, diversity-generating retroelements.



DPANN Archaea
Altiarchaeales
TACK Archaea
Asgard Archaea
Peregrinibacteria
Parcubacteria Superphylum
Microgenomates Superphylum
Other
FCB
PVC

Sulfur cycle
Nitrogen cycle
Carbon fixation
Phosphorus intake
Viral defense systems
Copper efflux systems
Arsenic metabolism
Hypersalinity adaptation
Hydrogenase
DGRs

Tree scale: 0.5

MAGs

Omnitrophica (PVC Group)

FCB Group

Others

Microgenomates Superphylum

Parcubacteria Superphylum

Peregrinibacteria

Asgard Archaea

TACK Archaea

Altiarchaeales

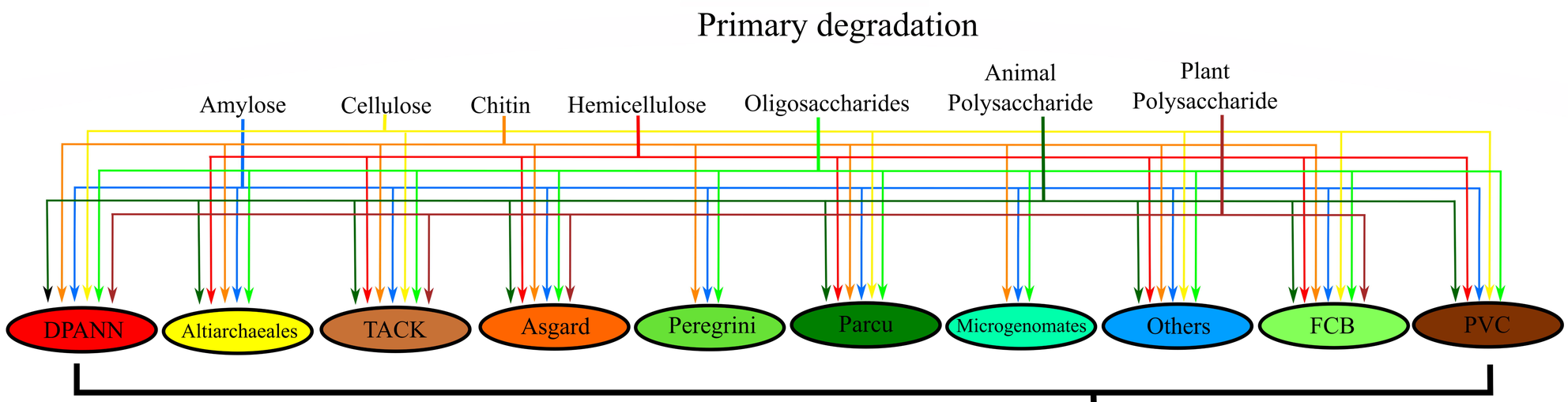
DPANN Archaea

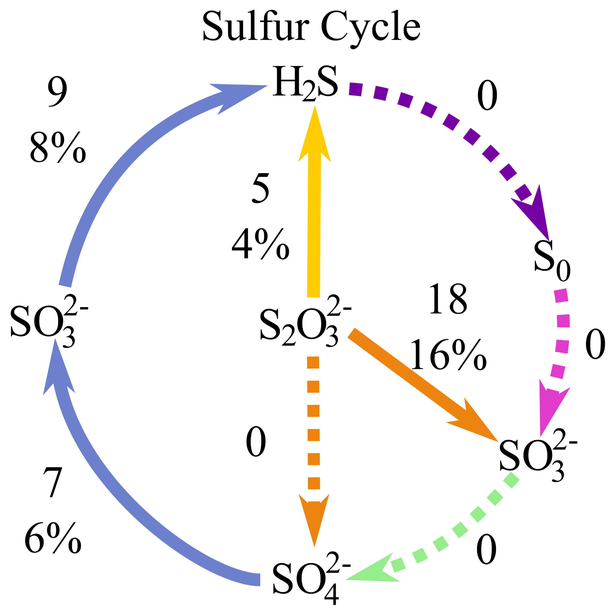
Bin_214 OP3 (Omnitrophica)
RBin_031 OP3 (Omnitrophica)
RBin_379 OP3 (Omnitrophica)
Bin_505 OP3 (Omnitrophica)
RBin_191 OP3 (Omnitrophica)
Bin_019 OP3 (Omnitrophica)
Bin_522 SAR406 (Marinimicrobia)
Bin_134 Hyd24-12 (Fermentibacteria)
Bin_206 Hyd24-12 (Fermentibacteria)
Bin_049 Zixibacteria (GN15)
Bin_210 Zixibacteria (GN15)
Bin_052 Zixibacteria (GN15)
Bin_224 Zixibacteria
Bin_249 Zixibacteria
Bin_113 Zixibacteria
Bin_366 Zixibacteria
Bin_336 Zixibacteria
Bin_298 KSB1
Bin_228 KSB1
Bin_079 KSB1
Bin_338 Bacterium BMS3Bbin04
RBin_200 Ignavibacteria
Bin_012 Caldithrix
Bin_462 Fibrobacteres
RBin_299 Fibrobacteres
Bin_527 Fibrobacteres
RBin_290 Fibrobacteres
Bin_432 Fibrobacteres
Bin_337 Fibrobacteres
Bin_355 Fibrobacteres
RBin_291 Fibrobacteres
RBin_049 Aegiribacteria
Bin_327 WOR-3 (Stahlbacteria)
RBin_337 WOR-3 (Stahlbacteria)
RBin_199 WS3 (Latescibacteria)
RBin_220 WS3 (Latescibacteria)
Bin_043 UBP1
Bin_270 UBP1
Bin_413 OLB16
Bin_402 Poribacteria
Bin_115 Eisenbacteria
Bin_153 Eisenbacteria
Bin_233 Eisenbacteria
Bin_251 Eisenbacteria
Bin_501 Eisenbacteria
Bin_528 Elusimicrobia
RBin_245 Elusimicrobia
Bin_127 OP8 (Aminicenantes)
Bin_177 OP10 (Armatimonadetes)
Bin_235 OP10 (Armatimonadetes)
Bin_567 KSB3 (Candidatus Vecturithrix)
RBin_158 KSB3 (Candidatus Vecturithrix)
Bin_537 Coatesbacteria
Bin_172 Coatesbacteria
Bin_169 TM6 (Dependentiae)
Bin_469 TM6 (Dependentiae)
Bin_479 WS6 (Dojkabacteria)
Bin_125 WS6 (Dojkabacteria)
Bin_334 WS6 (Dojkabacteria)
Bin_101 WWE3 (Katanobacteria)
Bin_542 Pacebacteria
Bin_476 Pacebacteria
Bin_510 Uhrbacteria
RBin_141 Uhrbacteria
Bin_582 Uhrbacteria
Bin_289 Magasanikbacteria
Bin_529 Falkowbacteria
Bin_429 Buchananbacteria
Bin_346 Moranbacteria
Bin_114 Moranbacteria
Bin_419 Moranbacteria
Bin_516 Portnoybacteria
Bin_162 Nealsonbacteria
Bin_503 Peregrinibacteria
Bin_357 Peregrinibacteria
Bin_387 Peregrinibacteria
Bin_256 Peregrinibacteria
Bin_155 Peregrinibacteria
Bin_120 Heimdallarchaeota
Bin_204 Lokiarchaeota
RBin_035 Lokiarchaeota
Bin_229 Lokiarchaeota
Bin_485 Lokiarchaeota
RBin_125 Lokiarchaeota
Bin_186 Lokiarchaeota
Bin_342 Lokiarchaeota
Bin_478 Thorarchaeota
RBin_111 Thorarchaeota
Bin_407 Bathyarchaeota
Bin_348 Bathyarchaeota
Bin_245 Bathyarchaeota
Bin_400 Altiarchaeales
RBin_020 Altiarchaeales
RBin_358 Altiarchaeales
Bin_386 Micrarchaeota
RBin_221 Micrarchaeota
Bin_349 Micrarchaeota
Bin_091 Micrarchaeota
Bin_042 Aenigmarchaeota
Bin_473 Woesearchaeota
Bin_031 Woesearchaeota
Bin_272 Woesearchaeota
Bin_423 Woesearchaeota
Bin_274 Woesearchaeota
Bin_534 Woesearchaeota
Bin_568 Woesearchaeota
Bin_434 Woesearchaeota
Bin_028 Woesearchaeota
Bin_187 Woesearchaeota
Bin_486 Woesearchaeota
Bin_263 Woesearchaeota
Bin_285 Pacearchaeota
RBin_261 Pacearchaeota
Bin_078 Pacearchaeota
Bin_444 Nanoarchaeota

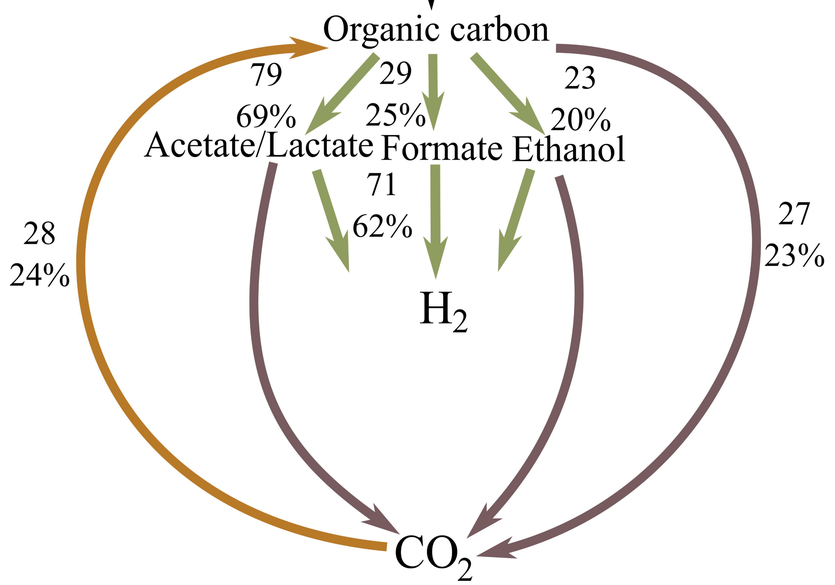
ASR
DSR
SOX
DNRA
Nitrogen fixation
Denitrification
ANR
RuBisCo
3HP Pathway
4HB Pathway
rTCA
WL Pathway
Inorg P Transportation
Alkal Phos
Phosphonatase
Polyphosphate
CRISPR
DISARM
BREX
Cd/Zn/Co efflux ATPase
CopAB
cusAB
Arsenate Reductase
Arsenite Efflux
opu/pro/betT
uvrABCD
cphAB
gltBD/glnAE
mtrAH/hdrABC/mttB
[NiFe] Group 1a hydrogenase
[NiFe] Group 1c hydrogenase
[NiFe] Group 3b hydrogenase
[NiFe] Group 3c hydrogenase
[NiFe] Group 3d hydrogenase
[NiFe] Group 4a hydrogenase
[NiFe] Group 4b hydrogenase
[NiFe] Group 4e hydrogenase
[NiFe] Group 4g hydrogenase
[NiFe] Group 4l hydrogenase
[FeFe] Group A hydrogenase
[FeFe] Group B hydrogenase
[FeFe] Group C1 hydrogenase
[FeFe] Group C2 hydrogenase
[FeFe] Group C3 hydrogenase
DGR
Rhodopsin

Pathways/Genes

Pathway partially present
Pathway/genes absent
Sulfur metabolism
Nitrogen metabolism
Carbon fixation
Phosphorus intake
CRISPR/DISARM/BREX associated genes
Heavy metal resistance
Arsenic resistance
Osmoprotection
UV radiation resistance
Cyanophycin associated genes
Ammonium assimilation
C_1 metabolism
Hydrogenases
DGRs
Rhodopsins

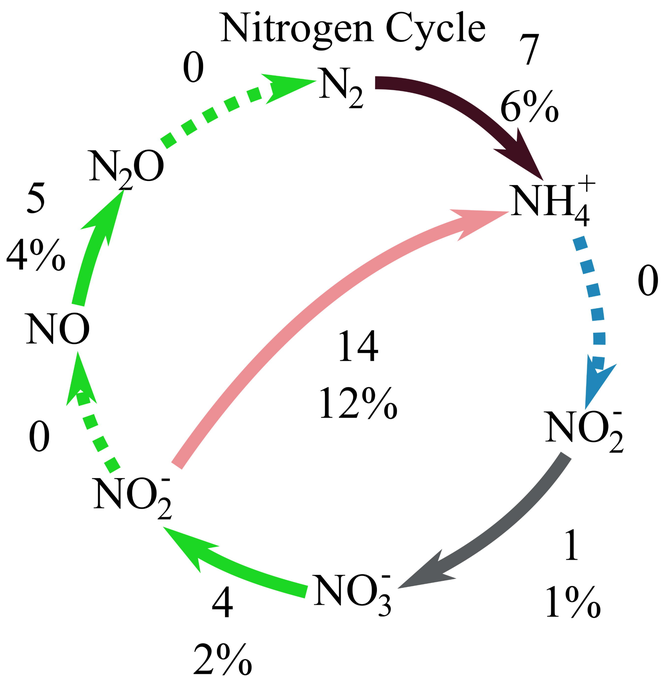






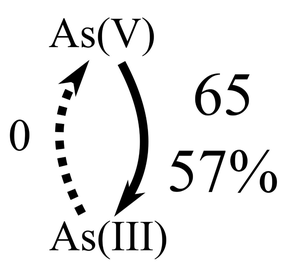
Sulfur cycle
Sulfate reduction
Sulfide oxidation
Sulfur oxidation
Sulfite oxidation
Thiosulfate oxidation
Thiosulfate reduction

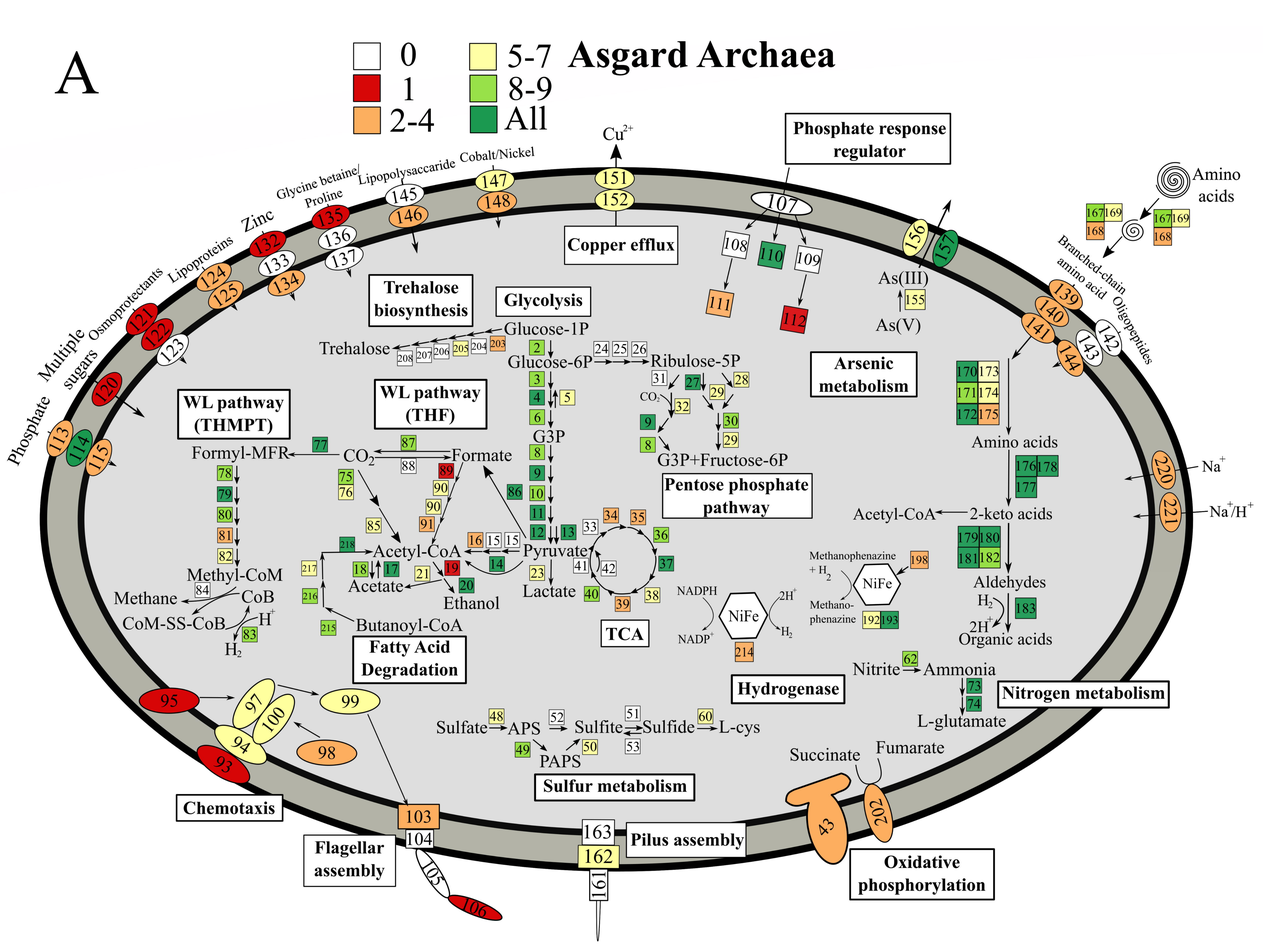
Carbon cycle
Fermentation
Carbon fixation
Organic carbon oxidation

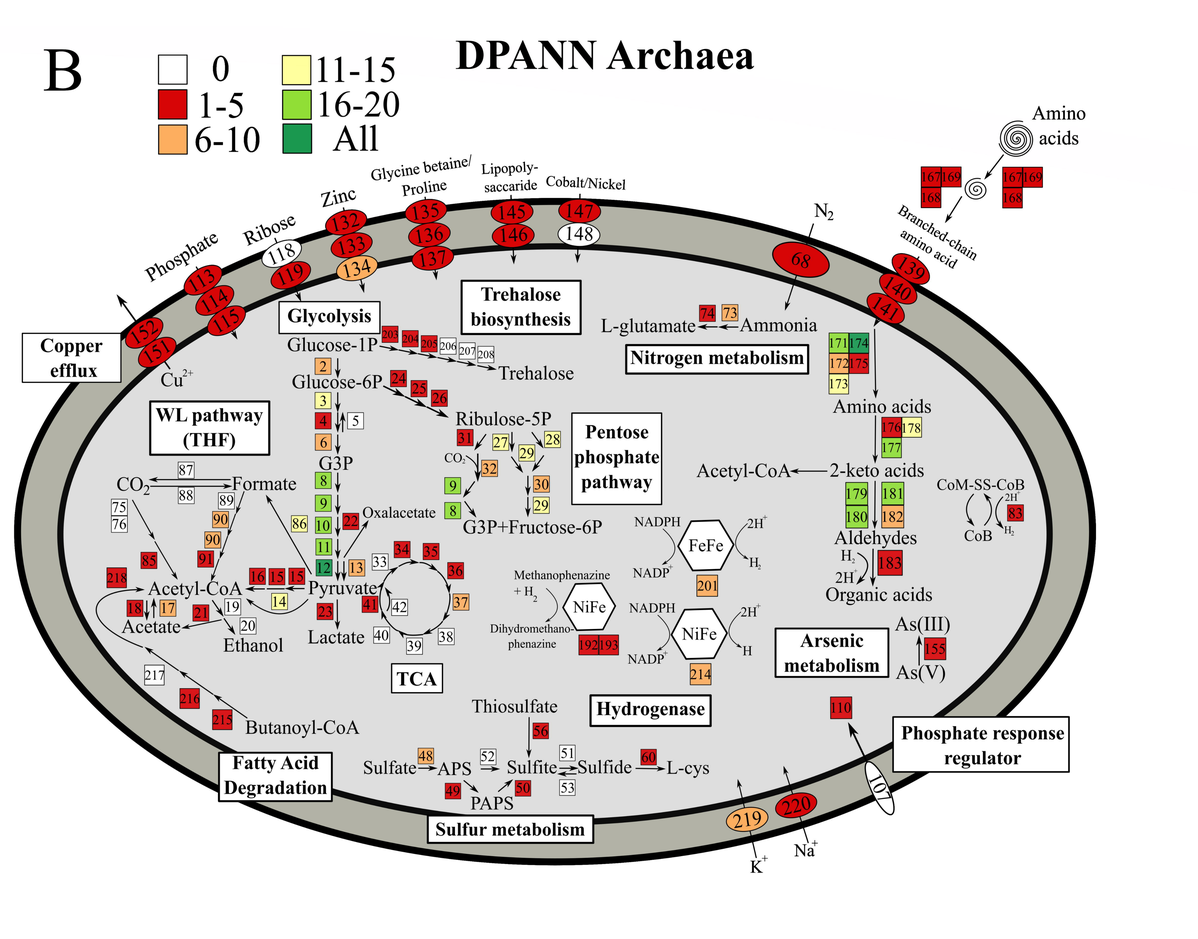


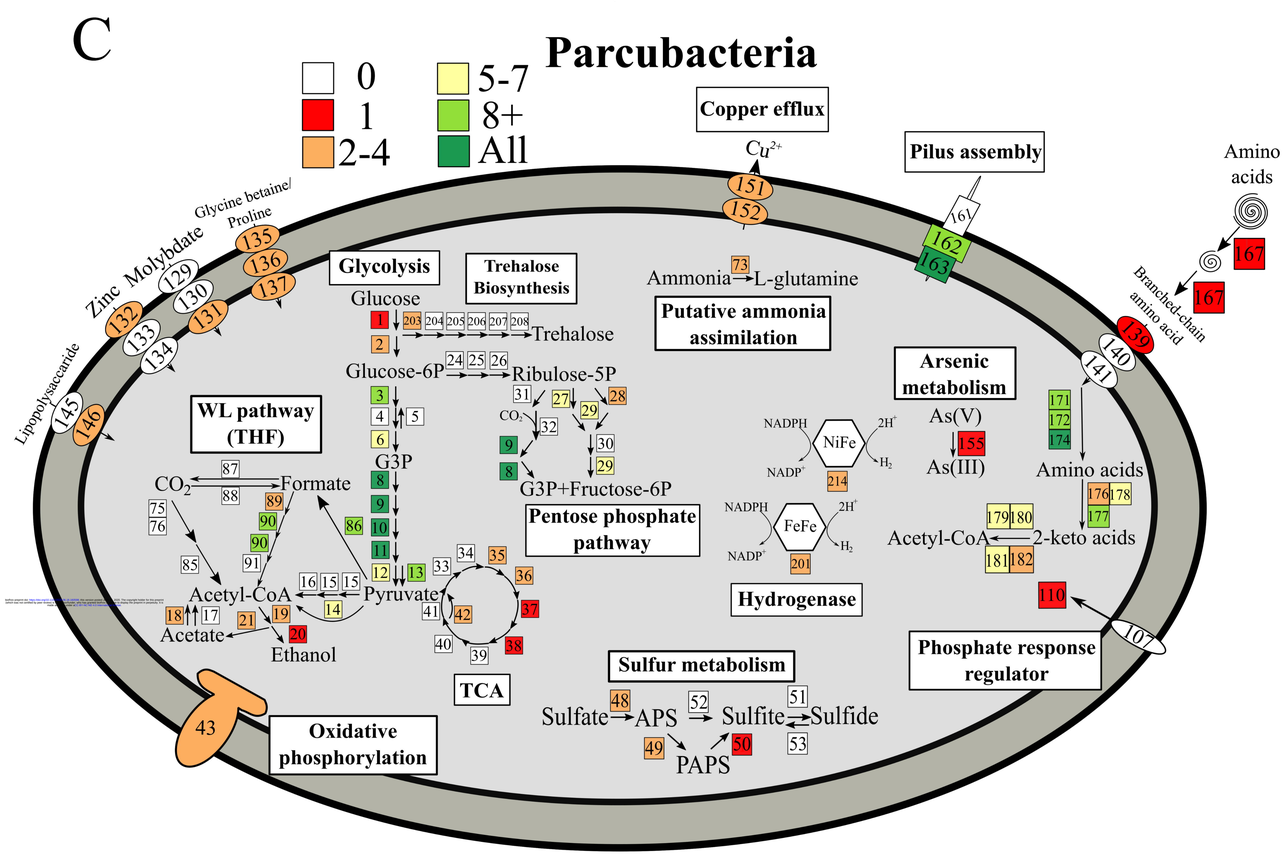
Nitrogen cycle
Denitrification
Nitrogen fixation
Ammonia oxidation
Nitrite oxidation
DNRA

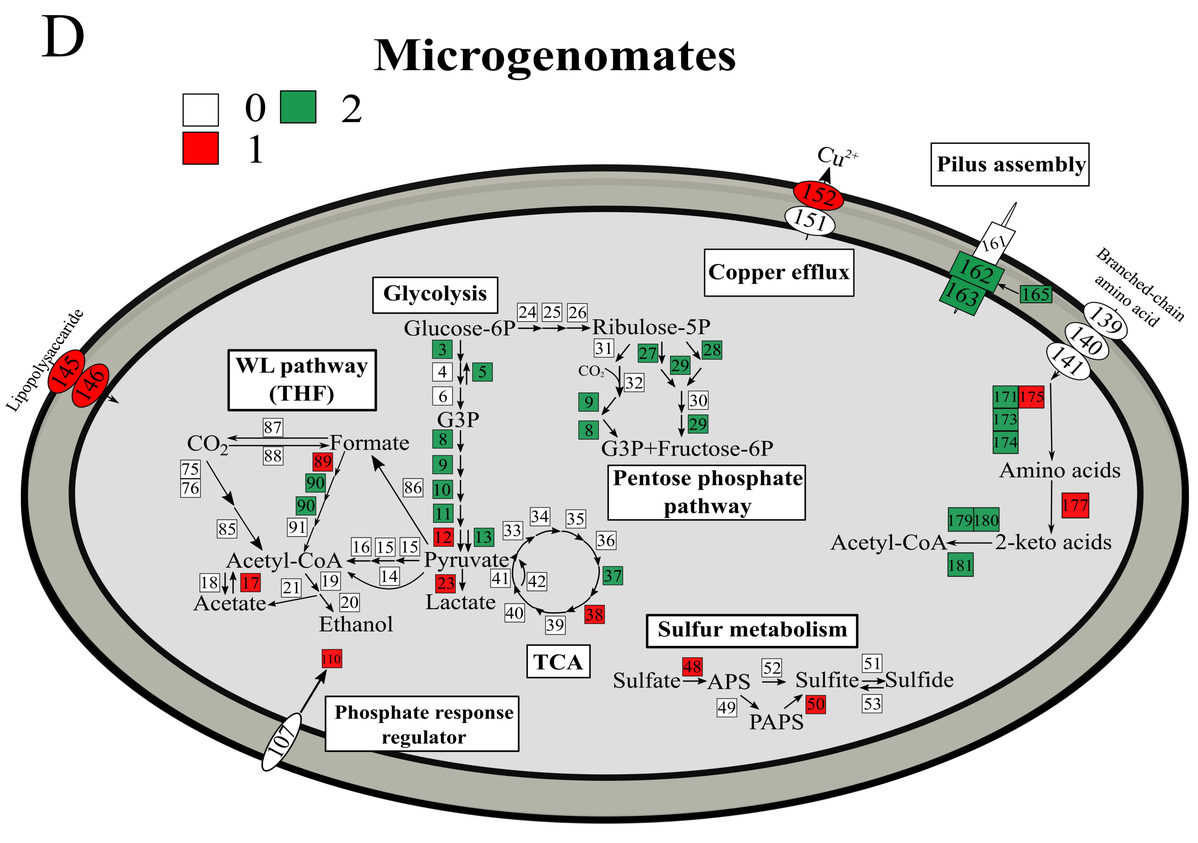
Arsenic metabolism

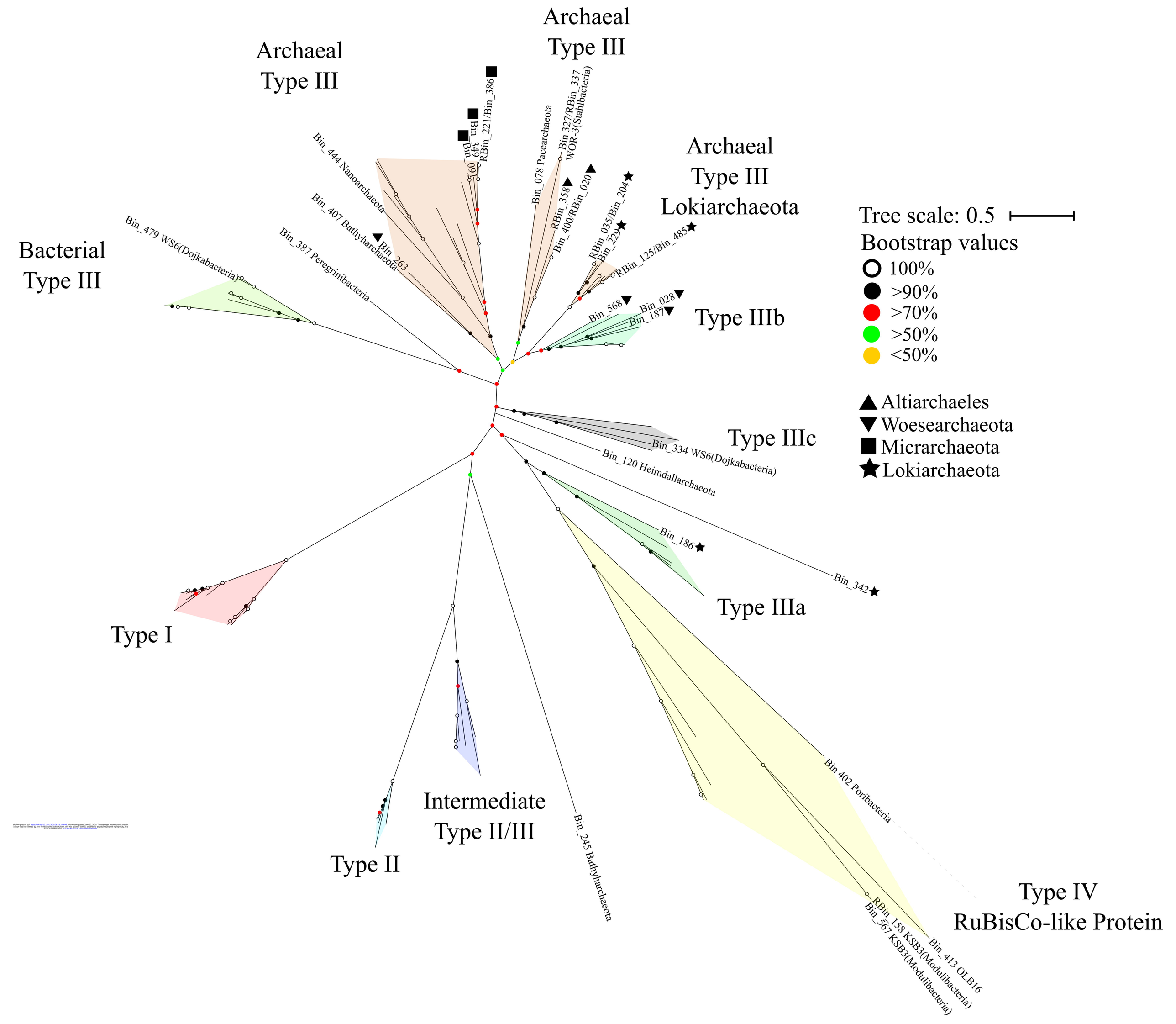



A
Asgard Archaea
0
1
2-4
5-7
8-9
All
Cu2+
Copper efflux
Phosphate response regulator
Amino acids
Branched-chain amino acid
Oligopeptides
Glycine betaine/ Proline
Zinc
Lipoproteins
Cobalt/Nickel
Lipopolysaccaride
Multiple sugars
Osmoprotectants
Phosphate
Trehalose biosynthesis
Glycolysis
Glucose-1P
Trehalose
Glucose-6P
Ribulose-5P
G3P
G3P+Fructose-6P
Arsenic metabolism
As(III)
As(V)
WL pathway (THMPT)
WL pathway (THF)
Formyl-MFR
CO2
Formate
Methyl-CoM
Methane
CoB
CoM-SS-CoB
H2
Acetyl-CoA
Acetate
Ethanol
Pyruvate
Lactate
Butanoyl-CoA
Fatty Acid Degradation
Pentose phosphate pathway
TCA
Amino acids
2-keto acids
Aldehydes
Organic acids
Methanophenazine + H2
Methano-phenazine
NiFe
NADPH
NADP+
Hydrogenase
Nitrite
Ammonia
Nitrogen metabolism
L-glutamate
Sulfate
APS
PAPS
Sulfite
Sulfide
L-cys
Sulfur metabolism
Succinate
Fumarate
Oxidative phosphorylation
Pilus assembly
Chemotaxis
Flagellar assembly
Na+
Na+/H+


B
DPANN Archaea
0
1-5
6-10
11-15
16-20
All
Amino acids
Branched-chain amino acid
Glycine betaine/ Proline
Zinc
Lipopoly-saccaride
Cobalt/Nickel
Phosphate
Ribose
Copper efflux
Cu2+
Trehalose biosynthesis
Glycolysis
Glucose-1P
Trehalose
Glucose-6P
Ribulose-5P
G3P
G3P+Fructose-6P
WL pathway (THF)
Pentose phosphate pathway
CO2
Formate
Oxalacetate
Acetyl-CoA
Acetate
Ethanol
Pyruvate
Lactate
Butanoyl-CoA
Fatty Acid Degradation
TCA
N2
L-glutamate
Ammonia
Nitrogen metabolism
Amino acids
2-keto acids
Aldehydes
Organic acids
CoM-SS-CoB
CoB
Methanophenazine + H2
Dihydromethano-phenazine
NiFe
FeFe
NADPH
NADP+
Hydrogenase
Arsenic metabolism
As(III)
As(V)
Thiosulfate
Sulfate
APS
PAPS
Sulfite
Sulfide
L-cys
Sulfur metabolism
Phosphate response regulator
K+
Na+


C
Parcubacteria
0
1
2-4
5-7
8+
All
Copper efflux
Cu2+
Pilus assembly
Amino acids
Branched-chain amino acid
Glycine betaine/ Proline
Zinc
Molybdate
Lipopolysaccaride
Glycolysis
Trehalose Biosynthesis
Glucose
Trehalose
Glucose-6P
Ribulose-5P
G3P
G3P+Fructose-6P
Putative ammonia assimilation
Ammonia
L-glutamine
Arsenic metabolism
As(V)
As(III)
WL pathway (THF)
CO2
Formate
Acetyl-CoA
Acetate
Ethanol
Pyruvate
Lactate
Pentose phosphate pathway
TCA
Amino acids
2-keto acids
NADPH
NADP+
NiFe
FeHe
Hydrogenase
Sulfur metabolism
Sulfate
APS
PAPS
Sulfite
Sulfide
Phosphate response regulator
Oxidative phosphorylation


D
Microgenomates
0
1
2
Cu2+
Copper efflux
Pilus assembly
Amino acids
Branched-chain amino acid
Lipopolysaccaride
Glycolysis
Glucose-6P
Ribulose-5P
G3P
G3P+Fructose-6P
WL pathway (THF)
Pentose phosphate pathway
CO2
Formate
Acetyl-CoA
Acetate
Ethanol
Pyruvate
Lactate
TCA
Amino acids
2-keto acids
Sulfur metabolism
Sulfate
APS
PAPS
Sulfite
Sulfide
Phosphate response regulator



Sun

$CO_2/CO/H_2$

UV

CPR

DPANN/ Asgard archaea

AMP Nucleotide salvaging pathway

SO_4^{2-}

GN15/ Zixibacteria

Chloro/ Gemma

Delta

H_2S

N_2

CO_2 O_2

Oxygenic Photosynthesis

N_2

H_2S

SO_4^{2-}

Asgard archaea

H_2 H^+

Cyanobacteria

Bacteroidetes

N_2O

NO

NH_4^+

Ammonia Assimilation

Organic matter/ EPS

CO CO_2

WL Pathway

Ribose

Cyanobacteria

Dynamic carbon source?

EPS

CO_2

3HP/4HB Pathway

Acetyl-CoA

N/C Storage

Cyanophycin

Organic Carbon

Sulfur Assimilation

Polysulfide

Scavenging and recycling organic carbon

Alpha

Delta

Gamma

Plancto

Polyhydroxyalkanoate

Heterotrophs

CPR/ DPANN

Assimilatory Sulfate Reduction

H_2S

APS

SO_4^{2-}

Nitrification?

NO_3^- NO_2^-

Denitrification

CO_2/H_2/Trimethylamine

mvh *hdr* *mtr* *mtt*

H_2

HM

DPANN archaea

Nutrients

Methanogenesis?

CH_4

High salinity/ UV

Phosphate intake

pstB *pho* *phn*

opu *bet* *uvr*

Chemotaxis

che

K^+/Na^+

trk *nhaA*

Adaptive response

DGR

Signalling

cas *pglZ* *drm* *brxC*

Viral defense CRISPR/BREX/ DISARM

ars *cop* *czc*

Heavy metal resistance

LasI *LuxR* *AhlD* *quiP*