Microbial dark matter filling the niche in hypersaline microbial mats

Hon Lun Wong1,2, Fraser I. MacLeod1,2, Richard Allen White III2,3,4, Pieter T. Visscher2,5,6 and Brendan P. Burns1,2*

Abstract

Background: 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 microbes, jointly coined as ‘microbial dark matter’ (MDM), are hypothesised to play key roles in modern microbial mats.

Results: We reconstructed 115 metagenome-assembled genomes (MAGs) affiliated to MDM, spanning 42 phyla. This study reports for the first time 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 CO2 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.

Conclusions: This is the first study to reconstruct and describe in detail metagenome-assembled genomes (MAGs) affiliated with microbial dark matter in hypersaline microbial mats. Our data suggests that these microbial groups are major players in these systems. In light of our findings, we propose H2, ribose and CO/CO2 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 possessing 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, that cover diverging metabolic needs to fulfil energy demands [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]. 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, harbours 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 (68 ppt, up to 250 ppt during low tide), heatwaves (elevated water temperatures of 2–4 °C), high UV radiation (1750 µE/m²/s, up to 2000 µE/m²/s), oligotrophic waters (extremely low phosphorus concentration—6.46 µM), fluctuating tides and even seasonal cyclonic events (i.e. category 3 cyclone Olwyn) contribute to the mats being subjected to an extreme environment in Shark Bay [17–24]. 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 that 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

This study describes for the first time in detail metagenome-assembled genomes (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, Additional file 18: Table S1). 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 18: 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, Peregri nibacteria and ‘others’ (Fig. 1). Archaeal MAGs were classified into Asgard archaea, TACK, DPANN and Altia rcheales. It is still under debate whether Altia rcheales 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

Rhodopsin genes were identified in Asgard archaea, Parcubacteria, Bathyrachaeota 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 19: Table S2). This study expanded the range of phyla encoding schizorhodopsin, which has not been identified in Bathyrachaeota, 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 acquisition of such an evolutionary adaptation of this enigmatic group may be a feature in hypersaline, sunlit surface settings, with the ability to utilise light energy to counter the high cost of osmotic balance maintenance [31, 33, 34].

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

Most of the MDM MAGs harbour incomplete metabolic pathways despite the oligotrophic condition present in Shark Bay (Additional file 1: Supplementary Information), hence are suggested to be host-dependent [29, 36]. We screened the MAGs for diversity-generating retroelements (DGRs), which are fast-evolving proteins enabling host-dependent microorganisms to attach to the hosts’ surface [30, 37–39], and they were identified mostly in Parcubacteria and DPANN archaea (Figs. 1 and 2). Interestingly, despite having versatile metabolic pathways, DGRs were also identified in Asgard archaeal MAGs in this study, which has not been reported before (Fig. 1, Fig. 2 and Additional file 1: Supplementary Information). This may indicate Asgard archaea once resided in energy-limited environments which relies on symbiotic relationships [32, 37]. Parcubacteria and DPANN archaeal members lack biosynthetic capabilities and were suggested to have a symbiotic, host-dependent lifestyle, in which DGRs facilitate adaptation to rapidly changing environments, providing them with the tools to exploit symbiotic associations with their host [1, 29, 30, 40].

It is hypothesised that MDM may take advantage of the lack of virus defence systems observed here (CRISPR, BREX, DISARM and DNA phosphorothioation; Additional file 1: Supplementary Information and Additional file 20: Table S3), despite the high abundance of viruses and virus defence systems identified in Shark Bay in previous studies [24, 41]. Given the limited metabolic capacities and host-dependent lifestyle, these organisms may serve as ‘viral decoys’ to alleviate the load on the host’s viral defence system, and benefits from the viral DNA as a pentose and phosphorus source [42]. Furthermore, maintaining such systems is energetically costly [43, 44]. Another advantage of lacking such
systems is to avoid autoimmunity, in which the virus defence systems may inadvertently target its own genome due the lack of ability to distinguish between the MDM host genome from viruses [42, 43]. It is proposed that the synergy between the lack of viral defence systems and the presence of DGRs facilitate rapid screening and acquisition of biological functions for survival in these mat systems, allowing MDM with limited metabolic pathways to adapt to a dynamic and extreme environment.

Sulfur, nitrogen and carbon cycling in Shark Bay MDM
Despite having reduced metabolic capacity, various genes within sulfur, nitrogen and carbon cycles were found distributed among the Shark Bay microbial dark matter community. All Zixibacteria and Zixibacteria order GN15 (previously classified as candidatus phylum GN15) genomes harbour dissimilatory (bi)sulfite reductase (dsrAB) and adenylylsulfate reductase (aprAB) genes affiliated with dissimilatory sulfate reduction [45, 46].
MAGs of unclassified bacterium BM53Bbin04 (FCB Group) also harbour dsrAB, while Coatesbacteria encode for an aprA gene, suggesting that 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 dissipatory (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 20: Table S3), which is an essential physiological partner to dsrAB in sulfite reduction [46]. This suggests that 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 20: Table S3). Genes encoding nitrite reductase were found in all Asgard archaea MAGs except Heimdallarcheota (Fig. 2 and Additional file 20: 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 20: 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 (Additional file 1: Supplementary Information and Additional file 6: Figure S5). On the other hand, Parcubacteria, Microgenomatetes, Peregrinibacteria and DPANN archaea encode a lower range of GH enzymes, suggesting that 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 20: Table S3). This finding suggests that most MDM here 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. 4a and Additional File 7: Figure S6) and Bathyarchaeota (Fig. 4d)

Fig. 3 Putative involvement of Shark Bay MDM MAGs in carbon, sulfur, nitrogen and arsenic metabolisms. CAZy enzymes with different coloured 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.
and Additional File 10: Figure S9) MAGs, indicating their putative ability to fix and reduce CO₂ to acetyl-CoA through the Wood-Ljungdahl (WL) pathway (Fig. 2 and Additional file 20: Table S3). Generally, microorganisms can either use tetrahydrofolate (THF) or tetrahydromethanopterin (THMPT) as C₁ 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 Lokiarchaea MAG (Bin_186) harbours a complete anaerobic H₂-dependent THMPT-WL pathway, inferring the ability to fix CO₂ and H₂ into acetate [58] (Fig. 4a and Additional file 20: 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 20: 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 CO₂ assimilation**

Surprisingly, despite the reduced-sized genomes, 32 MAGs encode for ribulose biphosphate carboxylase (RuBisCo) (Fig. 2, Additional file 20: Table S3 and Additional file 21: Table S4). Given that 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 20: 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 20: Table S3), indicating the potential ability to incorporate CO₂ into nucleotide salvaging pathways [52, 65–67]. One Lokiarchaeota MAG (Bin_186) encodes for a type IIIa RuBisCo, which is known to fix CO₂ through the reductive hexulose-phosphate (RHP) cycle [65, 66]. Although it potentially lacks the ability to fix CO₂ 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).

![Fig. 5](image-url) 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 [64, 65], type IIIb [46], type IIIc [46] and type IV RuBisCo-like protein [66] were identified. Circular dots of different colors represent bootstrap values. RuBisCo sequences in this study and reference sequences are listed in Additional file 21: Table S4.
allowing for the potential production of methylene-
$\text{H}_2\text{MPT}$, which may play a role in replenishing the $\text{C}_1$
carriers in the THMPT-WL pathway [66].

One interesting finding is that Heimdallarchaeota (Bin_
120) contains RuBisCo at the basal position (Fig. 5), sug-
gesting that 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 sub-
stitute 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
$\text{CO}_2$) as augmented carbon sources into the central car-
bon metabolism, these non-autotrophic RuBisCo may
putatively free phosphate groups from nucleotides to sup-
plement 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 and Additional file 7-13:
Figure S6-S12).

**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 7-15: Figure S6-14 and
Additional file 20: Table S3). A total of 16 types of hydroge-
nases were identified against the HydDB database [69], in-
cluding 10 [NiFe] and 6 [FeFe] classes (Fig. 2, Additional
file 20: 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 (groups 3, 4
and [FeFe]) (Fig. 2).

Almost half of the Shark Bay MAGs harbour hydrogen-
ases associated with H$_2$ production, and of particular signifi-
cance, Parcubacteria and Woesearchaeota only en-
code H$_2$ producing hydrogenases ([NiFe]-3b and [FeFe]
Group A), which are fermentative in nature [70] (Fig. 2,
Fig. 4 and Additional file 20: 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]. Fur-
thermore, a global survey suggests that Woesearchaeota
form consortiums with hydrogenotrophic methanogens
by providing H$_2$ in exchange of nutrients [71]. It is sug-
gested 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 elec-
tron 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 that 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 hydrogemo-
 trophic 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). How-
ever, 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 irra-
diation 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 ex-
change 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, 64, 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
$\text{CO}_2$ are prominent currencies to augment energy
income.

**Conclusions**

This is the first study to reconstruct and describe in de-
tail 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 retro-
elements and novel ESPs identified in Asgard archaea. It
is suggested that Asgard archaea are not only organohe-
terotrophs 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 H₂ and carbon transformation in microbial mats. Various forms of RuBisCo were encoded, allowing putative CO₂ 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 evolutionarily 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, CO₂/CO and H₂ are suggested to be prominent currencies among MDM in these mats and were potentially a widespread phenomenon on early Earth [40, 86].

**Fig. 6** Proposed ecological roles of MDM in Shark Bay microbial mats. Orange rectangular indicates metabolic pathways, green oval represents MDM in Shark Bay microbial mats while blue circles represent genes found in a previous study [24] and the current study. CO₂/CO, H₂ and ribose are proposed as main energy currencies of Shark Bay MDM, enclosed in red boxes. Dashed arrows indicate putative metabolic exchange/microbial interactions. Within the dashed box are genes involved in environmental adaptation. HM hydrogenotrophic methanogens, WL pathway Wood-Ljungdahl pathway, 3HP/4HB pathway 3-hydroxypropionate/4-hydroxybutyrate pathway, EPS extracellular polymeric substance, DGR diversity-generating retroelements. A more detailed proposed ecological model of Shark Bay microbial mats combining MDM with other community members is illustrated in Additional File 17: Figure S16.
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 recover near-complete MAGs generated from the three binning programs [96]. CheckM was employed to examine the quality (completeness and contamination level of MAGs), and MAGs statistics were obtained through QUAST and tRNAscan-SE [97–99]. MAGs with at least medium quality (> 50% completeness, < 10% contamination) were selected in this study [27]. Subsequently, the taxonomy of MAGs was determined with GrDb-tk [100], in which only MAGs classified as microbial dark matter in previous literatures were chosen [1, 3–6, 12, 13, 25, 26].

Maximum-likelihood-based phylogenetic trees based on 16 concatenated ribosomal proteins (rpl2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and rps5, 8, 10, 17, 19) were constructed as described in Hug et al. (2013) [101]. Only MAGs with at least 8 ribosomal proteins were included in the analysis. Bin_245 (Bathyarchaeota) was not included in the phylogenetic tree (Fig. 1) as it has less than 8 ribosomal proteins. Phylosift version 1.0.1 was used to extract ribosomal proteins from the genomes [102]. Subsequently, ribosomal protein sequences were aligned using MAFFT version 7.310 [103]. BMGE was then used to remove gaps in the alignment with BLOSUM30 matrix and gap rate cut-off of 50% [104]. The resulting protein alignments were concatenated as described in Hug et al. (2013) [101]. The concatenated ribosomal proteins were then used to construct a phylogenetic tree using IQ-TREE version 1.6.1 with a total of 1000 bootstrap replicates, which the output file was visualised with iTOL [105, 106].

Functional annotation

Nucleotide contigs of metagenome-assembled genomes were translated to amino acid sequences by employing Prodigal version 2.6.3 [107]. Functional annotation was carried out using GhostKoala to assign amino acid contigs to MAGs against 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] and I-TASSER [112] 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 [113]. 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 [114].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40168-020-00910-0.

Additional file 1. Supplementary information. Supplementary text with additional information.

Additional file 2: Figure S1. Unrooted maximum-likelihood phylogenetic tree of putative rhodopsin in Shark Bay MDM MAGs. Maximum-likelihood phylogenetic tree constructed with rhodopsin gene found in the MDM MAGs with 1000 bootstrap replicates. Lokiarchaeota, Bath-yarchaeota, Uhrbacteria, Buchananbacteria and an unclassified archaeon encode rhodopsin clustered in the same group with the novel, recently discovered schizorhodopsin [7]. Circular dots of different colors represent
Additional file 3: Figure S2. Eukaryotic Signature Proteins (ESP) in the MAGs of Asgard archaea. MAGs were annotated using Interproscan [77] and GhostKoala [78] and confirmed using HHpred [79] and BLAST [80]. Shark Bay Asgard archaea were found to contain ESP likely involved in cytoskeleton dynamics, information processing, trafficking machinery, signalling systems as well as eukaryotic-like N-linked glycosylation. * indicates newly identified ESP. Dark and white boxes indicate presence and absence of genes respectively.

Additional file 4: Figure S3. Metabolic potential of FCB (Fibrobacteres-Chlorobi-Bacteroidetes) group bacteria. A metabolic map summarising the genomic potential and metabolic capacities of the 26 MAGs affiliated with the FCB group. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 5: Figure S4. Maximum-likelihood phylogenetic tree of dsrAB in Shark Bay MDM MAGs. Maximum-likelihood phylogenetic tree was constructed with reference dsrAB sequences from the dsrAB database [105], with 1000 bootstrap replications. dsrAB genes found in the present study are classified as reductive bacterial type dsrAB and are highlighted in green. Circular dots of different colors represent bootstrap values. dsrAB sequences found in the MDM MAGs are listed in Additional file 24: Table S7. Branches shaded red indicates reductive archaeal type dsrAB, yellow shade indicates oxidative bacterial type dsrAB, light green indicates Archaeoglobus lineages, light blue indicates Firmicutes lineages, light purple indicates Actinobacteria lineages, orange represents Nitrosipes lineages, purple represents Deltaproteobacteria lineages, green represent dsrAB in the present study and no shades represent uncultured/ environmental lineages.

Additional file 6: Figure S5. Color-coded table indicating major carbohydrate-active enzymes (CAZy) in MDM MAGs. X-axis indicates different types of glycoside hydrolase (GH) genes in the CAzy database and y-axis represent MAGs of microbial dark matter. White indicates absence of GH genes in the MAGs. Color panel on the left represents different groups of MDM MAGs according to Fig. 1.

Additional file 7: Figure S6. Metabolic potential of Asgard archaea. A metabolic map summarising the genomic potential and metabolic capacities of the 10 MAGs affiliated with Asgard archaea. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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.

Additional file 8: Figure S7. Metabolic potential of DPANN archaea. A metabolic map summarising the genomic potential and metabolic capacities of the 21 MAGs affiliated with DPANN archaea. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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.

Additional file 9: Figure S8. Metabolic potential of Parcubacteria. A metabolic map summarising the genomic potential and metabolic capacities of the 11 MAGs affiliated with Parcubacteria. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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.

Additional file 10: Figure S9. Metabolic potential of Bathyarchaeota (TACK archaea). A metabolic map summarising the genomic potential and metabolic capacities of the 3 MAGs affiliated with TACK archaea. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 11: Figure S10. Metabolic potential of Peregrinibacteria. A metabolic map summarising the genomic potential and metabolic capacities of the 5 MAGs affiliated with Peregrinibacteria. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 12: Figure S11. Metabolic potential of TAEarchaeae. A metabolic map summarising the genomic potential and metabolic capacities of the 3 MAGs affiliated with TAEarchaeae. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 13: Figure S12. Metabolic potential of other MDM bacteria. A metabolic map summarising the genomic potential and metabolic capacities of the 28 MAGs affiliated with other MDM bacteria. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 14: Figure S13. Metabolic potential of the Microgenomates. A metabolic map summarising the genomic potential and metabolic capacities of the 2 MAGs affiliated with Microgenomates. Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 15: Figure S14. Metabolic potential of the PVC (Planctomycetes-Verrucomicrobia-Chlamydiae) group bacteria. A metabolic map summarising the genomic potential and metabolic capacities of the 6 MAGs affiliated with Omnitrophia (OP3). Numbers represent specific genes in given pathways and the corresponding genes are listed in Additional file 20: 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; WL pathway, Wood-Ljungdahl pathway; PAPS, 3'-phosphoadenylyl sulfate; APS, Adenylyl sulfate.

Additional file 16: Figure S15. Maximum-likelihood phylogenetic tree of putative dehalogenase in Shark Bay MDM MAGs. Maximum-likelihood phylogenetic tree constructed with reductive dehalogenase domain (IPR028894) found in the MDM MAGs with 1000 bootstrap replications. Both reductive dehalogenase domain (IPR028894) and epoxyquiuosine reductase were found in Asgard archaea, KS81, Aminicenantes (OP8), Armatimonadetes (OP10), Xizibacteria and Batharyarchaeota. Although these MAGs encode both epoxyquiuosine reductase and reductive
dehalogenase domain, they cluster with homologous sequences of dehalo-
genase reductases. Thus it is unclear if the MDM community in Shark Bay can respire organohalides. Red shading indicates bona fide dehalo-
genases found in previous studies [6, 22], yellow shading indicates hom-
ologous sequences of dehalogenase reductases, and green shading
represent putative dehalogenase domains (IPR028894) in this study.

Additional file 17: Figure S16. Proposed ecological model of Shark
Bay mats (modified from Wong et al., 2018 [24], incorporating MDM).
This is a more detailed ecological model figure incorporating previous
findings with putative roles and interactions of MDM. Orange rectangular
boxes indicate putative metabolic pathways and putative functional roles
of MDM, blue rectangular boxes indicate pathways found in a previous
study (Wong et al., 2018 [24]) and white rectangular boxes represent
substrates/metabolites. Dark green ovals represent microbial dark matter
while light green ovals represent other microorganisms. Red boxes
encircle CO₂/CO, H₂ 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. Abbreviation: Chloro, Chloroflexi; Gemma,
Gemmatimonadetes; HM, hydrogenotrophic methanogens; N/C storage,
nitrogen/carbon storage; WL, pathway; Wood-Ljungdahl pathway; 3HP/4HBP pathway, 3-hydroxypropionate/4-hydroxybutyrate pathway; EPS,
extracellular polymeric substance; DGR, diversity-generating retroelements.

Additional file 18: Table S1. Genome statistics of 24 high quality MDM
MAGs and 91 medium quality MDM MAGs.

Additional file 19: Table S2. Rhodopsin sequences, BLAST results
and reference rhodopsin sequences used in Additional file 2: Figure S1.

Additional file 20: Table S3. Table indicating the presence and
absence of a wide range of genes involved in different metabolic
pathways. Green boxes indicate presence of genes while white boxes
indicate absence of genes.

Additional file 21: Table S4. Rubisco sequences, BLAST results and
reference Rubisco sequences used in Fig. 5.

Additional file 22: Table S5. Relative abundance of the bacterial
community in Shark Bay microbial mats. Green boxes indicate bacteria
affiliated with microbial dark matter.

Additional file 23: Table S6. Relative abundance of the archaeal
community in Shark Bay microbial mats. Green boxes indicate archaea
affiliated with microbial dark matter.

Additional file 24: Table S7. Dissimilatory sulfate reduction sequences
dsrAB identified in microbial dark matter MAGs in this study.

Additional file 25: Table S8. Reductive dehalogenase sequences
identified in microbial dark matter MAGs in this study.

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.

Authors’ contributions
HLW performed the major analyses of Shark Bay mat metagenomic data
and MDM MAGs and drafted the manuscript. FM conducted specific data
analyses in relation to Asgard archaea MAGs and RAW performed additional
data analyses in relation to binning and assembly. PTV conducted the
mat sampling and related field measurements, as well the writing of specific
manuscript sections. BPB designed, coordinated, and led the study, and
wrote specific manuscript sections. The author(s) read and approved the
final manuscript.

Availability of data and materials
MAGs were deposited in the National Center for Biotechnology Information
(NCBI) whole genome shotgun (WGS) projects, under BioProject ‘Novel
microbial dark matter shaping Shark Bay microbialites’ (PRJNA561032), with
accession numbers WJ100000000 to WJMT00000000 (Additional file 18: Table
S1). The MAGs are also available in Open Science Framework (https://osf.io/
bhcub/). All code is deposited on Github and freely accessible (https://
github.com/Handymanalan/Microbial-Dark-Matter-in-Shark-Bay).

Competing interests
The authors declare that they have no competing interests.

Author details
1School of Biotechnology and Biomolecular Sciences, The University of New
South Wales, Sydney 2052, Australia. 2Australian Centre for Astrobiology,
University of New South Wales, Sydney, Australia. 3RAW Molecular Systems
LLC, Spokane, WA, USA. 4Department of Bioinformatics and Genomics, The
University of North Carolina, Charlotte, NC, USA. 5Department of Marine
Sciences, University of Connecticut, Mansfield, USA. 6Biogeosciences, the
Université de Bourgogne Franche-Comté, Dijon, France.

Received: 24 June 2020 Accepted: 19 August 2020
Published online: 16 September 2020

References
1. Rinke C, Schwientek P, Szczynka A, Ivanova NN, Anderson U, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter.
Nature. 2013;499(7459):431–7.
2. Marcy Y, Ouvremy C, Bik EM, Løsekam T, Ivanova N, Martin HG, et al. Dissecting biological ‘dark matter’ with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci
U S A. 2007;104(11):8899–94.
3. Hedlund BP, Dodsworth JA, Murugapiran SK, Rinke C, Woyke T. Impact of single-cell genomics and metagenomics on the emerging view of extremophile ‘microbial dark matter’. Extremophiles. 2014;18(5):865–75.
4. Bruno A, Sandioni G, Rizzi E, Bernasconi M, Vicario S, Galimberti A, et al. Exploring the under-investigated ‘microbial dark matter’ of drinking water treatment plants. Sci Rep. 2017;7:44350.
5. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems. 2018;3(5):e00055–18.
6. Saw JH, Spang A, Zaremba-Niedzwiedzka K, Juzokaite L, Dodsworth JA, Murugapiran SK, et al. Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. Philos Trans Royal Soc B. 2015;370:1678–87.
7. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Blackstrom D, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017;541(7637):533–8.
8. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 2015;521:173–9.
9. Lacey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad Sci. 2016;113(21):5970–5.
10. Schulz F, Else-Fahdosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, et al. Towards a balanced view of the bacterial tree of life. Microbiome. 2017;5:140.
11. Makarova KS, Wolf YI, Koonin EV. Towards functional characterization of archeal genomic dark matter. Biochem Soc Trans. 2019;47(2):280–562.
12. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected
biogeochemical processes in an aquifer system. Nat Commun. 2016;7:13219.
13. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain bacteria.
Nature. 2015;23(2):208–11.
14. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017;541(7637):533–8.
15. Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ. Members of the candidate phyla radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. Microbiome. 2017;5(1):112.
16. Bernard G, Pathmanathan JS, Lannes R, Lopez P, Baptiste E. Microbial dark matter investigations: how microbial studies transform biological
knowledge and empirically sketch a logic of scientific discovery. Genome Biol Evol. 2018;10(1):707–15.
17. Burns BP, Goh F, Allen M, Neilan BA. Microbial diversity of extant stromatolites in the hypersaline marine environment of Shark Bay, Australia.
Environ Microbiol. 2004;6(10):1096–101.
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. Ruivind Y, White RA III, Neilsan BA, Burns BP. Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. ISME J. 2016;10(1):183–96.

20. Allen MA, Geh F, Burns BP, Neilsan BA. Bacterial, archaeal and eukaryotic diversity of smooth and puscular microbial mat communities in the hypersaline lagoon of Shark Bay. Geology. 2009;7:82–9.

21. Reinold M, Wong HL, MacLeod Fl, Metzler J, Thompson A, Burns BP. The vulnerability of microbial ecosystems in a changing climate: potential impact in Shark Bay. Life. 2019;9:7.

22. Wong HL, Smith DL, Visscher PT, Burns BP. Niche differentiation of bacterial communities at a kilometre scale in Shark Bay microbial mats. Sci Rep. 2015;5:16507.

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

24. Wong HL, White RA II, Visscher PT, Charlesworth JC, Vazquez-Comapos X, Burns BP. Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. ISME J. 2018;12:3618–39.

25. Selden L, Lloyd K, Wrightson KC. The bright side of microbial dark matter: lessons learned from the uncultivated majority. Curr Opin Microbiol. 2016;31:217–26.

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.

27. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy KP. DPANN radiations. Nat Rev Microbiol. 2018;16:629–37.

28. Arnold C. Core concepts: how diversity-generating retroelements promote mutation and adaptation in myriad microbes. Proc Natl Acad Sci. 2017;114(40):10509–11.

29. Nelson WC, Stegen JC. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol. 2015;6:713.

30. White RA III, Wong HL, Ruivind Y, Neilsan BA, Burns BP. Viral communities of Shark Bay modern stromatolites. Front Microbiol. 2018;9:1223.

31. Burstein D, Sun CL, Brown CY, Sharon I, Anasthanaran K, Probst AJ, et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun. 2016;7:10613.

32. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends Genet. 2010;26:335–40.

33. Vale PF, Lafforgue G, Gatchitch F, Gardan R, Moineau S, Gandon S. Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus. Proc R Soc B. 2015;282(1812):1270.

34. Romper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. ISME J. 2017;11:2319–33.

35. Dombrovski M, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat Commun. 2018;9(1):4999.

36. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs KU, Teske AP, et al. Genomic inference of the metabolism of cosmopolitan subsurface archaea, Hadesarchaea. Nat Microbiol. 2016;1(3):16002.

37. Wrighton KC, Thomas BC, Sharon I, Miller SS, Castelle CJ, Ver Berkmoes NC, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science. 2013;337(6102):1661–5.

38. Wrighton KC, Castelle CJ, Varaljay VA, Satogopan S, Brown CT, Wilkins MJ, et al. Rubisco of a nucleoside pathway known from archaea is found in diverse uncultivated phyla in bacteria. ISME J. 2016;10(11):1702–14.

39. Castelle CJ, Brown CT, Thomas BC, Williams RH, Banfield JF. Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the candidate phyla radiation. Sci Rep. 2017;7:40101.

40. MacLeod F, Kindler GS, Wong HL, Chen R, Burns BP. Asgard archaea: diversity, function, and evolutionary implications in a range of microbiomes. AIMS Microbiology. 2019;5(1):48–61.

41. Liu Y, Zhou Z, Pan J, Baker BJ, Gu JD, Li M. Comparative genomic inference suggests microtrophic lifestyle for Thorarchaeota. ISME J. 2018;12(4):1021–31.

42. Sousa FL, Martin WF. Biochemical fossils of the ancient transition from geoeconomics to bioenergetics in prokaryotic one carbon compound metabolism. Biochim Biophys Acta. 1837;2014:1ea59883.

43. Seitz KW, Lazar CS, Hinrichs KU, Teske AP, Baker BJ. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenes and sulfur reduction. ISME J. 2016;10(7):1696–705.

44. Sousa FL, Neukirchen S, Joshi AF, Lane N, Martin WF, Loksharchaeon is hydrogen dependent. Nat Microbiol. 2016;1(5):16034.

45. Hugler M, Severt SM. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Annu Rev Mar Sci. 2011;3:261–89.

46. Berg IA. Ecological aspects of distribution of different autotrophic CO2 fixation pathways. Appl Environ Microbiol. 2011;77(6):1925–36.

47. Ragsdale SW. Life with carbon monoxide. Curr Rev Biochem Mol. 2004;39:98–106.

48. King GM. Carbon monoxide as a metabolic energy source for extremely halophilic microbes: implications for microbial activity in Mars regolith. Proc Natl Acad Sci. 2015;112:4465–70.

49. Vourvourakis CD, Kharkov MV, Sorokin DY, Muyzer G. A metagenomic roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome. 2018;6(1):1–18.

50. Lázaro-García P, Moreira D. Cultured Asgard archaea shed light on eukaryogenesis. Cell. 2020;181(2):3389–402.

51. Sieber JR, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. Annu Rev Microbiol. 2013;67:6429–52.

52. Paul BG, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. Annu Rev Microbiol. 2013;67:6429–52.

53. Nelson WC, Stegen JC. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol. 2015;6:713.

54. White RA III, Wong HL, Ruivind Y, Neilsan BA, Burns BP. Viral communities of Shark Bay modern stromatolites. Front Microbiol. 2018;9:1223.

55. Arnold C. Core concepts: how diversity-generating retroelements promote mutation and adaptation in myriad microbes. Proc Natl Acad Sci. 2017;114(40):10509–11.

56. Nelson WC, Stegen JC. The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. Front Microbiol. 2015;6:713.

57. White RA III, Wong HL, Ruivind Y, Neilsan BA, Burns BP. Viral communities of Shark Bay modern stromatolites. Front Microbiol. 2018;9:1223.

58. Burstein D, Sun CL, Brown CY, Sharon I, Anastahnaran K, Probst AJ, et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun. 2016;7:10613.
