Discovery of an Abundance of Biosynthetic Gene Clusters in Shark Bay Microbial Mats

Ray Chen[1,2], Hon Lun Wong[1,2], Gareth S. Kindler[1,2], Fraser Iain MacLeod[1,2], Nicole Benaud[1], Belinda C. Ferrari[1,2] and Brendan P. Burns[1,2]*

[1] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia, [2] Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW, Australia

Microbial mats are geobiological multilayered ecosystems that have significant evolutionary value in understanding the evolution of early life on Earth. Shark Bay, Australia has some of the best examples of modern microbial mats thriving under harsh conditions of high temperatures, salinity, desiccation, and ultraviolet (UV) radiation. Microorganisms living in extreme ecosystems are thought to potentially encode for secondary metabolites as a survival strategy. Many secondary metabolites are natural products encoded by a grouping of genes known as biosynthetic gene clusters (BGCs). Natural products have diverse chemical structures and functions which provide competitive advantages for microorganisms and can also have biotechnology applications. In the present study, the diversity of BGC were described in detail for the first time from Shark Bay microbial mats. A total of 1477 BGCs were detected in metagenomic data over a 20 mm mat depth horizon, with the surface layer possessing over 200 BGCs and containing the highest relative abundance of BGCs of all mat layers. Terpene and bacteriocin BGCs were highly represented and their natural products are proposed to have important roles in ecosystem function in these mat systems. Interestingly, potentially novel BGCs were detected from Heimdallarchaeota and Lokiarchaeota, two evolutionarily significant archaeal phyla not previously known to possess BGCs. This study provides new insights into how secondary metabolites from BGCs may enable diverse microbial mat communities to adapt to extreme environments.

Keywords: biosynthetic gene cluster, microbial mat, secondary metabolite, natural product, genome mining, metagenomics

INTRODUCTION

Microorganisms in the environment can produce a wide range of secondary metabolites with diverse chemical structures (Katz and Baltz, 2016). The diversity in chemical structures enable these natural products to perform a variety of functions. Many natural products today are important components of human medicine and industry by acting as antibacterials, antitumor agents, insecticides, and many more (Schwecke et al., 1995; Brautaset et al., 2000; Waldron et al., 2001; Liu et al., 2016). Approximately 70% of the anti-infective drugs used in human medicine have been derived from natural products (Newman and Cragg, 2016). Secondary metabolites are not

directly associated with the growth of microorganisms but are known to provide benefits to the host producer by acting as growth inhibitors against rival bacteria, metal transporting agents, and quorum sensing molecules (Demain and Fang, 2000).

Biosynthetic gene clusters (BGCs) are a locally clustered group of two or more genes that together encode a biosynthetic pathway for the production of a secondary metabolite (Medema et al., 2015). Different structural classes of BGCs exist including non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), terpenes, and bacteriocins. NRPS and PKS are popular targets for natural product discovery as they are known to synthesize a diversity of antibiotics and immunosuppressants with enormous pharmaceutical potential (Schwecke et al., 1995; Tillett et al., 2000; Wang et al., 2014). Condensation (C) domains from NRPS clusters are functionally active protein sequences that catalyze the amide bond formation, an important step in peptide elongation (Bloudoff and Schmeing, 2017). Similarly, ketosynthase (KS) domains catalyze the condensation reaction in PKS clusters (Wang et al., 2014). These domains are suitable targets in genomic analysis as they are highly conserved and can be used to distinguish between different NRPS/PKS natural product pathways (Ziemert et al., 2012).

Previous work to uncover biosynthetic systems in uncultivated microorganisms have relied on cloning environmental DNA into host organisms to screen for function (Rondon et al., 2000). Other studies have utilized degenerate PCR primers to search for BGCs without the need for cloning (Charlop-Powers et al., 2014, 2015). However, primers become increasingly difficult to design for large gene families and can fail to amplify novel sequences of interest (Linhart and Shamir, 2005). Using a metagenomics workflow, we can reconstruct near-complete genomes *de novo*. This enables for potentially novel BGCs to be identified in both culturable and unculturable microorganisms as well as infer phylogenies and their ecological contribution in specific environments. antiSMASH is an *in silico* pipeline offering detection and analysis of many BGC types (Blin et al., 2017b). This pipeline was previously validated against a database of 473 verified BGCs with a high reported accuracy of 97.7% (Medema et al., 2011). BGC detection pipelines have significantly advanced our understanding of a range of ecologically and evolutionarily significant environments where natural product discovery has substantial potential, however, one environment where there is a knowledge gap in terms of BGCs are microbial mats.

Microbial mats are organo-sedimentary multilayered ecosystems that are host to a diverse community of microorganisms (Prieto-Barajas et al., 2018). Recent work has demonstrated microbial mats emerged as far back as 3.7 billion years (Nutman et al., 2016), and thus these systems hold substantial evolutionary significance. Shark Bay in Western Australia has extensive microbial mat systems, and the microbial communities have been studied in detail at the taxonomic and functional levels using a range of methods (Leuko et al., 2007; Allen et al., 2008; Wong et al., 2017, 2018; White et al., 2018, 2020). Mats located in Shark Bay are constantly exposed to high UV radiation, hypersalinity, and desiccation stresses (Burns et al., 2009; Wong et al., 2015). Recent metagenomic studies have shown microbial networks and feedback loops

are key in maintaining ecosystem function and stability in modern stromatolites (Prieto-Barajas et al., 2018), and it has been suggested the Shark Bay systems harbor microorganisms and functional genes that might have been prevalent in ancient microbial mats (Wong et al., 2018). Our data has shown that metabolic specialization may allow switching metabolic pathways at specific depth horizons in microbial mats, and these findings are part of a changing paradigm regarding the capability of microbial networks to function under seemingly inhospitable conditions. However, to date few studies have delineated the presence and potential significance of BGCs in microbial mats, despite the potential for biotechnological applications in these systems (Prieto-Barajas et al., 2018). The present study thus aimed to uncover the diversity of BGCs from microbial mats in Shark Bay, search for potential BGCs in understudied microorganisms, and identify the ecological functions of secondary metabolites in this particular ecosystem, via in-depth analyses of metagenome assembled genomes (MAGs) using antiSMASH and NaPDoS pipelines.

MATERIALS AND METHODS

Microbial Mat Sampling, Metagenomic Sequencing, and Data Availability

Sampling and metagenomic sequencing of Shark Bay microbial mats was undertaken previously (Wong et al., 2015, 2018). Briefly, triplicate mat samples were aseptically taken in April 2013 from Nilemah (26°27′336″S, 114°05.762″E), sliced into ten 2 mm layers, and total community DNA extracted. After quality control and library preparation, shotgun sequencing on an Illumina NextSeq 500 platform was performed in duplicate on each layer.

Quality Control, Assembly, and Binning

The quality of the reads was assessed with FastQC version 0.11.6 and low-quality bases (per base sequence quality <28) were removed with Trimmomatic version 0.36 (Bolger et al., 2014; Andrews, 2020). All sequencing files from the ten layers were co-assembled (minimum kmer 27, with incremental kmer of 10) with Megahit version 1.1.1 as described (Li et al., 2015; Stewart et al., 2018). After contigs assembly, only contigs with length more than 2000 bp were retained to avoid ambiguous gene annotation and binning errors from shorter contigs. BWA-MEM version 0.7.7 was used to map reads back to the assembled contigs (Li and Durbin, 2009). MAGs were constructed with MetaBAT2 version 2.12.1 (Kang et al., 2015). The completeness and contamination values of each MAG was assessed by CheckM version 1.0.12 (Parks et al., 2015). High-quality MAGs (>90% completeness and <5% contamination) were then selected for taxonomic assignment and BGC detection as recommended in the antiSMASH pipeline (Blin et al., 2017a).

Taxonomic Assignment and Phylogenetics

Taxonomic assignment to each high-quality MAG was done using GTDB-tk version 0.2.1. Furthermore, fifteen ribosomal

proteins from the large and small subunit (rpL2, rpL3, rpL4, rpL5, rpL6, rpL14, rpL15, rpL18, rpL22, rpL24, rpS3, rpS8, rpS10, rpS17, and rpS19) were then generated for all high-quality MAGs using Phylosift version 1.0.1 (Darling et al., 2014). The fifteen ribosomal proteins listed above are known to be ubiquitous in bacteria and archaea and hence suitable for building a phylogenetic tree to represent various bacterial and archaeal groups in an environmental (Darling et al., 2014). Sequences for each ribosomal protein were merged together and then aligned using MAFFT version 7.310, a multiple sequence alignment program (Katoh et al., 2002). Any gaps in the aligned sequences were removed using BMGE version 1.12 (Criscuolo and Gribaldo, 2010). After removing gaps in the aligned sequences, all aligned ribosomal protein sequences were merged to create a single file. A phylogenetic tree was constructed by maximum likelihood method based on 1000 bootstrap replications, via IQ-TREE version 1.6.9 (Nguyen et al., 2014). The tree was visualized and annotated using iTOL version 3.5.3 (Letunic and Bork, 2019).

Detection of BGCs in Shark Bay Mat Metagenomes

The overall file size of each metagenome layer was reduced to improve processing and detection of BGCs over 1 kb which was previously recommended (Blin et al., 2017a). To achieve this, contigs less than 1 kb in length were removed from each metagenome file. Annotation and labeling of relevant genomic features on contigs were done using Prokka version 1.13 (Seemann, 2014). Prediction of BGCs in mat metagenomes was performed using antiSMASH version 4.2.0 with minimum contig length at 1000 bp and the minimal detection option selected so that only BGCs are detected (Blin et al., 2017b). Predicted BGCs in each metagenome were compiled and the top ten highly abundant BGCs in each mat layer were represented in a bar chart generated using Rstudio with ggplot2 and RColorBrewer packages.

Comparison of Bacteriocin Clusters Between Microbial Mat Layers

Bacteriocin gene clusters detected in the top (0–2 mm), middle (8–10 mm), and bottom microbial mat layer (18–20 mm) were selected for further analysis due to their abundance and distribution throughout each mat layer. Ninety-five protein sequences from bacteriocin clusters were extracted from antiSMASH output and were blasted against the non-redundant protein sequences database with NCBI BLASTP (cut-off e-value $> 10^{-6}$, amino acid identity cut-off $<30\%$, and bit-score cut-off <50). The top three hits for each protein sequence were represented in **Supplementary Table S5**.

Detection and Analysis of BGCs in MAGs

From the group of high-quality MAGs, fifteen candidate MAGs of different assigned phyla were selected for BGC detection (**Supplementary Table S1**). The MAGs were passed through antiSMASH version 4.2.0 with minimum contig length at 1000 bp and all analysis options selected except for the ClusterFinder

algorithm. The KnownClusterBlast algorithm was also selected in order to compare predicted BGCs to the Minimum Information about a Biosynthetic Gene cluster (MIBIG) database (Medema et al., 2015). The MIBIG database contains BGC entries that have been experimentally characterized and can be used to infer natural products encoded by detected BGCs. A heatmap representing MAGs and the number of BGCs detected was constructed using Rstudio with ggplot2 and reshape2 packages.

Analysis of C and KS Domains From NRPS and PKS Clusters

To identify potential natural product pathways from NRPS and PKS gene clusters, the NaPDoS pipeline was used to compare C and KS domain sequences to a domain database of previously characterized natural products (Ziemert et al., 2012). The C and KS domains from NRPS and PKS detected in the fifteen candidate MAGs were extracted from antiSMASH output and subsequently analyzed using the NaPDoS web server with default settings. A phylogenetic tree was constructed by maximum likelihood using NaPDoS for all C domains. A separate phylogenetic tree for all KS domains was also constructed by maximum likelihood using NaPDoS. The trees were then visualized and annotated using iToL.
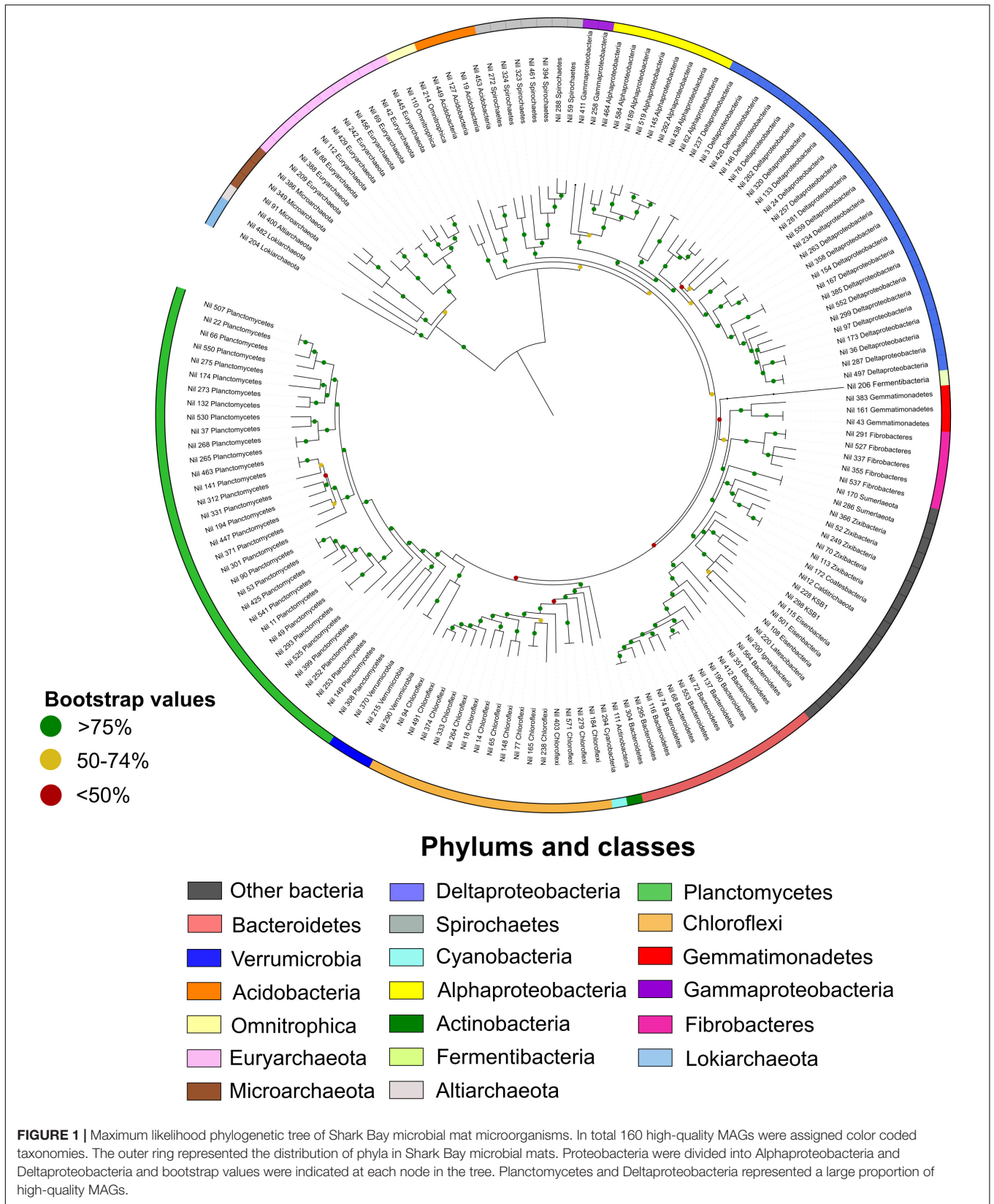
RESULTS AND DISCUSSION

Metagenomic Data Summary of Shark Bay Microbial Mats

A total of 761 MAGs were recovered from the binning process. Of these, 160 MAGs were classified as high-quality drafts ($>90\%$ completeness and $<5\%$ contamination) according to recently established standards (Bowers et al., 2017).

Microbial Mat Community Composition

A total of 27 phyla were identified based on phylogenetic analyses of the mat metagenomes (**Figure 1**). Planctomycetes represented 22% of MAGs, followed by Deltaproteobacteria (16%), and Chloroflexi (11%). This is in support of previous metagenomic studies and total 16S rRNA community data have shown that microbial mats are dominated by Deltaproteobacteria, Cyanobacteria, Planctomycetes, Spirochaetes, Chloroflexi, and Bacteroidetes (Wong et al., 2015, 2018).

Microbial dark matter – the vast array of microorganisms in the environment with no cultured representatives – MAGs such as "Candidatus Zixibacteria," "Candidatus Eisenbacteria," "Candidatus Sumerlaeota," and "Candidatus KSB1" were also uncovered here (**Figure 1** and **Supplementary Table S1**). The metabolism and functions of these latter bacterial groups are poorly understood but were suggested to have minimal genome capacities and require hosts for symbiotic lifestyles (Castelle et al., 2018). Two MAGs assigned as Lokiarchaeota (90.89% completeness and 9.76% contamination) and Heimdallarchaeota (95.33% completeness and 9.81% contamination) were also uncovered and assessed for potential BGCs as the secondary metabolism of Asgard archaea remains largely unexplored.

FIGURE 1 | Maximum likelihood phylogenetic tree of Shark Bay microbial mat microorganisms. In total 160 high-quality MAGs were assigned color coded taxonomies. The outer ring represented the distribution of phyla in Shark Bay microbial mats. Proteobacteria were divided into Alphaproteobacteria and Deltaproteobacteria and bootstrap values were indicated at each node in the tree. Planctomycetes and Deltaproteobacteria represented a large proportion of high-quality MAGs.

The metabolic functions of Lokiarchaeota have recently been proposed in a number of studies (Sousa et al., 2016; Wong et al., 2018), however, prior to the present study it has not been shown whether this phylum possess the capacity for secondary metabolism.

BGC Abundance Comparison in Different Microbial Mat Layers

The metagenomes from each microbial mat layer depth were analyzed with antiSMASH to compare their biosynthetic potential. A total of 1477 BGCs were detected with each layer having at least over 100 BGCs (**Figure 2**). Deeper mat layers were relatively less enriched with BGCs although the overall types of BGCs detected in each layer were not very different. A similar trend was also observed in a recent soil study where genomes from shallow samples were more enriched with BGCs compared to genomes from deeper samples (Sharrar et al., 2020). Natural product discovery efforts have mainly focused on marine and soil environments due to their rich biodiversity and biosynthetic potential. The BGC abundance observed in our dataset were found to be as high as those found in soil ecosystems and more enriched compared to a metagenomics lake study (Crits-Christoph et al., 2018; Cuadrat et al., 2018). We report the distribution of BGCs in Shark Bay microbial mats comparable with BGCs recovered from soil and water ecosystems thus supporting Shark Bay microbial mats as a possible new source of untapped biosynthetic potential.
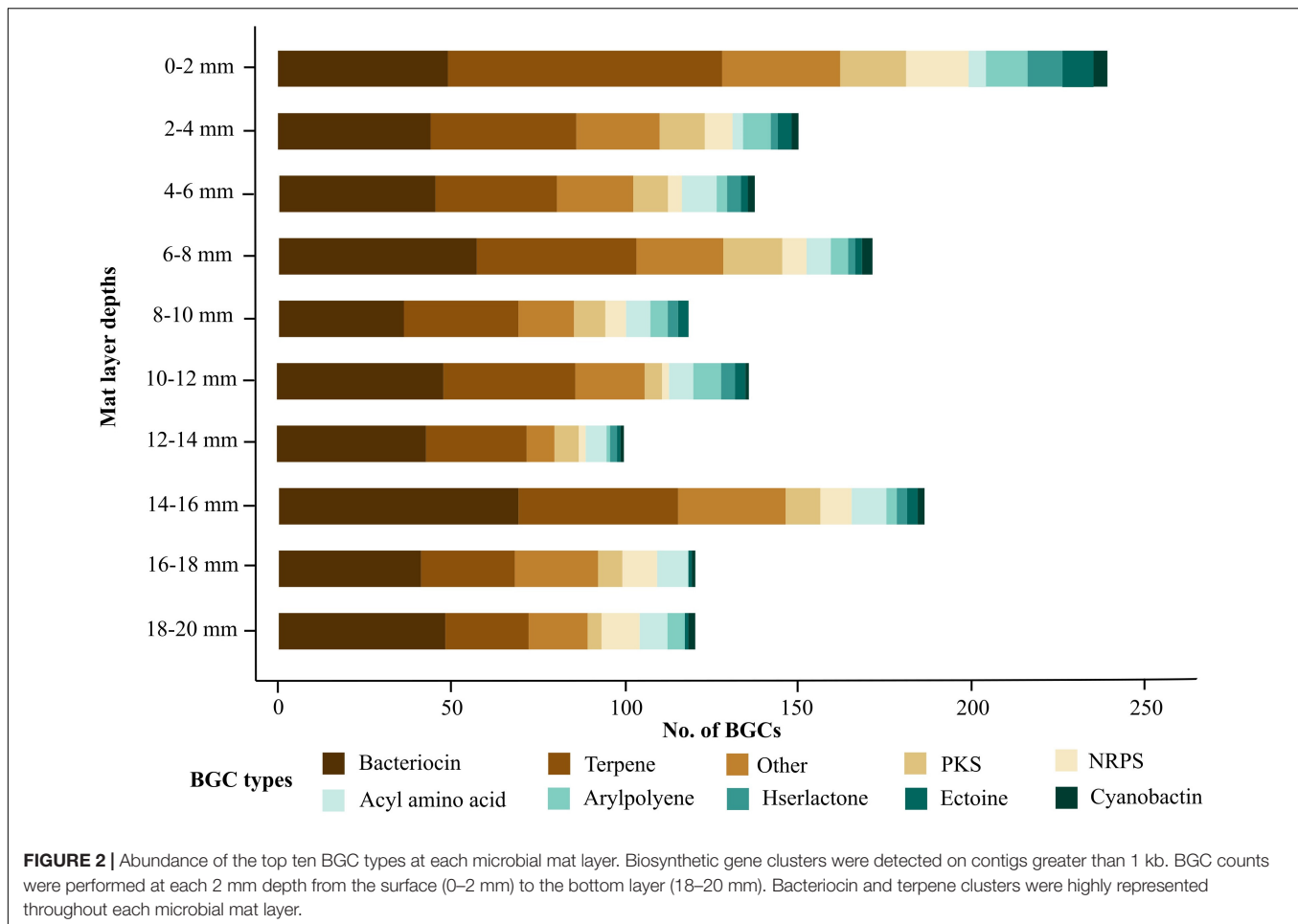
The surface layer had over 200 BGCs and contained the highest amount of BGCs of all mat layers (**Figure 2**). A previous 16S rRNA survey of Shark Bay microbial mats revealed the surface layer being dominated and over-represented by Alpha-, Gammaproteobacteria, Bacteroidetes, and Cyanobacteria (Wong et al., 2015). A recent study also found BGCs from Gammaproteobacteria were enriched in shallow sections of soil (Sharrar et al., 2020). Gammaproteobacteria and Cyanobacteria are known prolific secondary metabolite producers and Bacteroidetes have previously been reported to encode for a range of bacteriocins thus supporting the relatively high count of BGCs at the surface layer (Wang et al., 2014; Lopetuso et al., 2019; Reddy et al., 2020). The high distribution of BGCs at the microbial mat surface may also be attributed to microbial adaptations against harsh environmental conditions due to the surface layer being directly exposed to high UV radiation and desiccation stress (Wong et al., 2018). In this study, terpene gene clusters represented up to 33% of the BGCs in the microbial mat surface (**Figure 2**). Terpene gene clusters are responsible for producing a wide range of functionally diverse secondary metabolites such as carotenoids providing colorful pigments and geosmin, an odorous metabolite produced mainly by Cyanobacteria (Pattanaik and Lindberg, 2015). A function of geosmin has been associated with the removal of excess metabolites when microorganisms are subjected to stress events (Watson, 2003), and a similar role could be occurring in microbial mats. Photosynthetic Cyanobacteria have previously been observed to synthesize several carotenoids and syctonemin, a secondary metabolite synthesized in response to UV exposure

(Balskus et al., 2011; Galasso et al., 2017). Photosynthetic Cyanobacteria in Shark Bay microbial mats have previously been shown to produce syctonemin (D'Agostino et al., 2019), and the high representation of Cyanobacteria at the surface layer supports the relatively high number of surface terpene gene clusters observed in this study (Burns et al., 2005; Fisher et al., 2018). We hypothesized that high UV exposure at the surface layer of Shark Bay microbial mats have led to adaptations in Cyanobacteria encoding for additional terpene gene clusters. These gene clusters can produce a range of UV protective secondary metabolites which can absorb incoming UV radiation and subsequently shield the wider microbial mat community from harmful UV radiation.

A previous clone library survey of Shark Bay microbial mats first proposed the genetic potential for secondary metabolism in this ecosystem due to the identification of genes from NRPS and PKS encoding for enzymes responsible for the production hepatotoxins and antibiotics (Burns et al., 2005). However, the identification of gene clusters from this study was limited by the specificity of primers which allowed for only PKS/NRPS fragments to be amplified and cloned. Cultured strains of Cyanobacteria were also required in this study thus greatly limiting the search for BGCs. In comparison to the present study, we greatly expand the classes of BGCs detected in Shark Bay microbial mats and provide a greater resolution of BGC abundance and diversity present in this ecosystem.

There were a total of 77 NRPS and 101 PKS clusters detected from all layers combined, likely indicating microorganisms in all layers may encode for a range of secondary metabolites conferring benefits to the host or microbial community (**Figure 2**). In comparison, genomes of *Bacillus paralicheniformis* recovered from the hypersaline microbial mats of Rabigh Harbor Lagoon by the Red Sea in Saudi Arabia were found to be enriched with 480 BGCs and indicate the potential adaptations undertaken by microorganisms to survive the relatively hot and hypersaline ecosystems (Othoum et al., 2018). Here we report a relatively high abundance of NRPS and PKS clusters which may be attributed by adaptations undertaken by the microbial community to survive the extreme conditions in Shark Bay.

Bacteriocin and terpene gene clusters dominated in all layers and contributed over 20% of the total BGCs detected (**Figure 2**). Other studies have also observed similar trends of bacteriocin and terpene gene clusters dominating environments such as lakes, agricultural groundwater, and soil (Ludington et al., 2017; Cuadrat et al., 2018; Sharrar et al., 2020). The abundance of bacteriocins and terpenes in different environments suggests the secondary metabolites encoded by these gene clusters may play important roles in the survival and adaptation of the microbial community. In the present study, we found terpene gene clusters still represented approximately 20% of the BGCs recovered from deeper microbial mat layers despite exposure to significantly lower levels of light and UV radiation (**Figure 2**). Notably, some Cyanobacteria species and carotenoids encoded by terpene gene clusters were previously reported to be abundant in deeper layers of Shark Bay microbial mats hence supporting photosynthetic bacteria undergoing adaptations to lower light conditions (Wong et al., 2015; Fisher et al., 2018). Here we hypothesize that

FIGURE 2 | Abundance of the top ten BGC types at each microbial mat layer. Biosynthetic gene clusters were detected on contigs greater than 1 kb. BGC counts were performed at each 2 mm depth from the surface (0–2 mm) to the bottom layer (18–20 mm). Bacteriocin and terpene clusters were highly represented throughout each microbial mat layer.

the proportion of terpene gene clusters recovered from deeper layers of microbial mats may be contributed by a population of Cyanobacteria adapted to low light conditions. Furthermore, bacteriocin gene clusters were also widely distributed throughout microbial mat depths thus providing an opportunity to assess the protein sequences of these bacteriocin gene clusters for potential novelty.
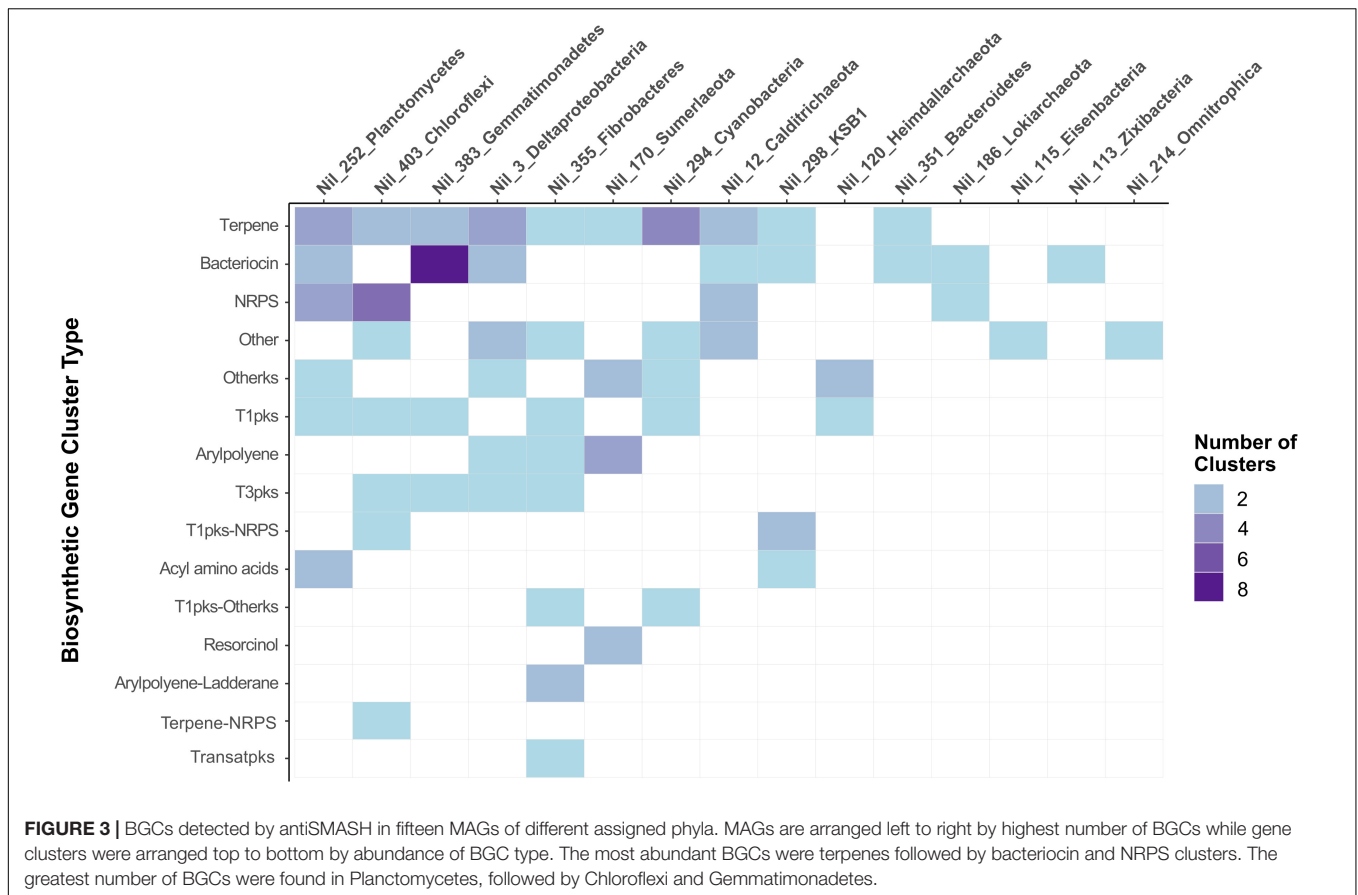
Bacteriocin Novelty in Microbial Mats

From the 95 protein sequences analyzed, a nitrogen-fixation protein (Nif11) represented 20% of matches (**Supplementary Table S5**). The Nif11 protein was previously reported to be associated with photosynthetic Cyanobacteria and Proteobacteria (Jacobson et al., 1989; Kaneko et al., 2001). While Nif11 plays a role in the nitrogen cycle, the identification of this protein within bacteriocin gene clusters is unusual and may suggest this protein could be repurposed for secondary metabolism. One particular study also reported evidence of Nif11 serving as natural product precursors (Haft et al., 2010). Previous studies have reported the widespread distribution of bacteriocin and their capacity to function as colonizing peptides, signaling molecules, or as killing peptide molecules (Dobson et al., 2012). Here we highlight the abundance and potential novelty of

bacteriocin clusters in Shark Bay microbial mats which were previously not reported in this ecosystem. Future transcriptomic work will help determine expression levels of bacteriocin and indicate if fluctuating temperatures and hypersaline conditions may induce higher bacteriocin production. Notably, hypothetical proteins and domains of unknown function were the top hits in approximately 38% of sequences (**Supplementary Table S5**), inferring potentially novel bacteriocins present. With bacteriocin clusters dominating most microbial mat layers, we also report the potential for novel bacteriocin clusters. Unculturable microorganisms may also be contributing to the largely unknown portion of putative BGC sequences yet to be annotated and characterized (Bernard et al., 2018).

Detection of BGCs in MAGs With Assigned Taxonomies

Biosynthetic gene cluster analysis was carried out on two archaeal MAGs and thirteen bacterial MAGs to link their taxonomies with their putative secondary metabolism (**Supplementary Table S2**). Terpene and bacteriocin BGCs were most abundant and occurred in ten and eight phyla, respectively (**Figure 3**). The greatest number of BGCs were detected in Planctomycetes (Nil_252), Chloroflexi (Nil_403), and Gemmatimonadetes (Nil_383). The

FIGURE 3 | BGCs detected by antiSMASH in fifteen MAGs of different assigned phyla. MAGs are arranged left to right by highest number of BGCs while gene clusters were arranged top to bottom by abundance of BGC type. The most abundant BGCs were terpenes followed by bacteriocin and NRPS clusters. The greatest number of BGCs were found in Planctomycetes, followed by Chloroflexi and Gemmatimonadetes.

Planctomycetes and Chloroflexi phyla are commonly found in soil and marine sponges and share characteristics with highly bioactive Actinobacteria such as large genomes and complex life cycles (Graça et al., 2016). Studies on Chloroflexi inhabiting marine sponges have detected novel NRPS and PKS as well as terpene clusters in their genomes suggesting their suitability for natural product discovery (Siegl and Hentschel, 2010; Bayer et al., 2018). An antiSMASH survey of the Planctomycetes genome also recovered numerous ladderane, bacteriocin, terpene, and fatty acid BGCs (Ludington et al., 2017). Similarly, we found several NRPS, PKS and terpene gene clusters in Planctomycetes and Chloroflexi which further highlights the high biosynthetic potential of these phyla (**Figure 3**).

Gemmatimonadetes represents one of many poorly understood bacterial phyla with only a few strains successfully cultured (Zeng et al., 2016). While commonly found in soil, few reference genomes are available and previous studies found evidence of some BGCs present in their genome (Wang et al., 2014; Crits-Christoph et al., 2018). From the twelve BGCs detected in Gemmatimonadetes, eight bacteriocins were detected suggesting this phylum may contain a greater biosynthetic capacity than previously thought (**Figure 3**). Similarly, a recent metagenomics soil study described bacteriocins being the prominent gene cluster detected in the genomes of Gemmatimonadetes (Sharrar et al., 2020). We report the prevalence of bacteriocin clusters in the genome of

Gemmatimonadetes recovered from hypersaline microbial mats which appear to share similar biosynthetic capacities seen in the genomes of Gemmatimonadetes inhabiting soil environments.

Microbial phyla affiliated with microbial dark matter possessing biosynthetic potential include "Candidatus KSB1" and "Candidatus Sumerlaeota." Terpene and PKS clusters were detected in these phyla and their KS domains were subjected to further analysis in NaPDoS (**Figure 3**). To our knowledge, their secondary metabolisms have not been previously reported. Only one BGC was detected in "Candidatus Zixibacteria," "Candidatus Omnitrophica," and "Candidatus Eisenbacteria." This is in contrast with a recent study which found a "Candidatus Eisenbacteria" genome encoding for NRPS and PKS domains which are comparable with some Actinobacteria genomes which are well-known for their incredible biosynthetic capacity (Wohlleben et al., 2016). In another study, the genomes of "Candidatus Omnitrophica" revealed terpene clusters involved in the non-mevalonate pathway (Ludington et al., 2017). Due to the lack of BGCs detected in these phyla, more work may be needed to recover more genomes from Shark Bay microbial mats before assessing the biosynthetic potential of these phyla. Nonetheless, we found several genomes containing NRPS or PKS including Cyanobacteria, Planctomycetes, Chloroflexi, Fibrobacteres, "Candidatus Sumerlaeota," "Candidatus KSB1," Lokiarchaeota and Heimdallarchaeota (**Figure 3**). The C and KS domains were extracted from these gene clusters and were

subjected to NaPDoS analysis to link NRPS and PKS systems to potential natural products.

Prediction of NRPS and PKS Products Encoded by Microbial Mat Microorganisms

The threshold for determining functional gene novelty remains debatable. According to NaPDoS, sequences with hits below 85% may indicate the domain of interest may contribute to natural products that have not yet been characterized (Ziemert et al., 2012). We found an overwhelming majority of C and KS domain sequences having sequence identity scores well below 85% to their predicted natural product (**Supplementary Tables S2, S3**). It was previously shown that shotgun metagenomics may lead to fragments of NRPS/PKS gene clusters which may account for the large number of low identity scores (Meleshko et al., 2019). On the contrary, scores below the threshold may imply potential novelty in NRPS and PKS products, although further studies would be required to assess the extent of this novelty.

Several KS domain sequences from Planctomycetes and Cyanobacteria aligned with poly-unsaturated fatty acid (PUFA) compounds from *Azotobacter* and *Shewanella* genus with up to 54 and 64% identity, respectively (**Figure 4** and **Supplementary Table S3**). PUFA compounds are well-known for their photoprotective properties against harsh UV radiation which damage the cell wall and membranes of microorganisms. A previous study employing bioinformatics and chemical analyses on the Shark Bay microbial mat community revealed a higher proportion of UV screening compounds present in the upper layer of microbial mats and found the abundance of Cyanobacteria influences the occurrence of scytonemin, a secondary metabolite with photoprotective properties (D'Agostino et al., 2019; Pathak et al., 2019). We report Planctomycetes and Cyanobacteria as the dominating phyla and main producers of PUFA on the surface layer of Shark Bay microbial mats which provide the microbial community shielding against high UV radiation.

Surprisingly, no KS domains from Cyanobacteria's PKS cluster matched to any toxin products despite numerous studies on microcystins which are a major class of toxins synthesized by Cyanobacteria (**Figure 4**; Bláha et al., 2009). In contrast, a metagenomics study on Cyanobacteria dominated microbial mats from the Eel River Network in California identified the natural product anatoxin-a from numerous Cyanobacterial genomes (Bouma-Gregson et al., 2019). We report the absence of Cyanobacteria gene clusters encoding for toxins in our dataset. Recovery of additional Cyanobacterial genomes for analysis would provide stronger support for these observations.

The secondary metabolism of the Chloroflexi phylum is generally understudied with recent studies only beginning to uncover their secondary metabolism (Sharrar et al., 2020). Initial surveys of Chloroflexi's genome reveal the presence of some NRPS and PKS clusters (Siegl and Hentschel, 2010; Wang et al., 2014; Bayer et al., 2018). In the present study, we found 17 of 28 C domains from Chloroflexi aligned closely with syringomycin (**Figure 5** and **Supplementary Table S4**). Syringomycin is

a phytotoxin compound belonging to the same family of lipopeptides which exhibit antimicrobial and biosurfactant properties (Raaijmakers et al., 2010). Biosurfactants are functionally versatile natural products sought after in medicine and biotechnology. Of ecological significance, biosurfactants enable microorganisms to grow on water-immiscible substrates by reducing the surface tension between interfaces thus facilitating greater nutrient uptake (Desai and Banat, 1997). This is especially relevant in microbial mats as a complex microbial community develops at the interfaces of water and biosedimentary minerals (Allen et al., 2009). This may also contribute to Chloroflexi being the most abundant bacteria in Shark Bay microbial mats given the oligotrophic environment (Wong et al., 2015). Here we expand on the secondary metabolism of Chloroflexi and report a potential NRPS cluster from Chloroflexi which may encode for a novel syringomycin-like product with potential ecological roles in enabling greater nutrient uptake and metabolism in microbial communities.

Analysis of KS and C Domains From Understudied Phyla

Ketosynthase domains from Heimdallarchaeota showed alignment to aflatoxin, an antifungal compound (**Figure 4**). PUFA domains were also the nearest hits according to NaPDoS (**Supplementary Table S3**). It is not clear if there is an association with Heimdallarchaeota and antifungal secondary metabolism as this was not previously reported. According to BLASTP, acetyltransferase, and beta-ketoacyl synthase specific to Heimdallarchaeota had 72% and 81% identity to KS domains from our Heimdallarchaeota MAG, respectively (**Supplementary Table S3**). This implies the presence of PKS clusters in the Heimdallarchaeota genome as acetyltransferase and beta-ketoacyl synthase are key biosynthetic components (Cummings et al., 2014). This study reports a potentially novel PKS cluster in Heimdallarchaeota's genome which may encode for a product sharing some similarity to PUFA or antifungal compounds.

Phylogenetic analyses revealed that the C domain from Lokiarchaeota did not align with any natural product although mycosubtilin, an antifungal compound is reported as the nearest hit at 28% similarity (**Figure 5** and **Supplementary Table S4**). A recent study found evidence of Asgard archaeal genomes possessing transcriptional regulators which are components of putative archaeal offense systems and toxin-antitoxin systems (Makarova et al., 2019). This seems to suggest that Asgard archaeal genomes have the capacity to encode for complex systems. While inconclusive at this stage, the lack of similarity to a natural product may suggest that Lokiarchaeota's genome may hold an unusually novel NRPS-like machinery not yet characterized. Here we describe the possibility of secondary metabolism in Lokiarchaeota which has not been previously shown.

A KS domain from "Candidatus KSB1" aligned closely with yersiniabactin, a siderophore molecule involved in transporting iron across cell membranes (**Figure 4**; Perry and Fetherston, 2011). Siderophores can also bind to a variety of metals to supply
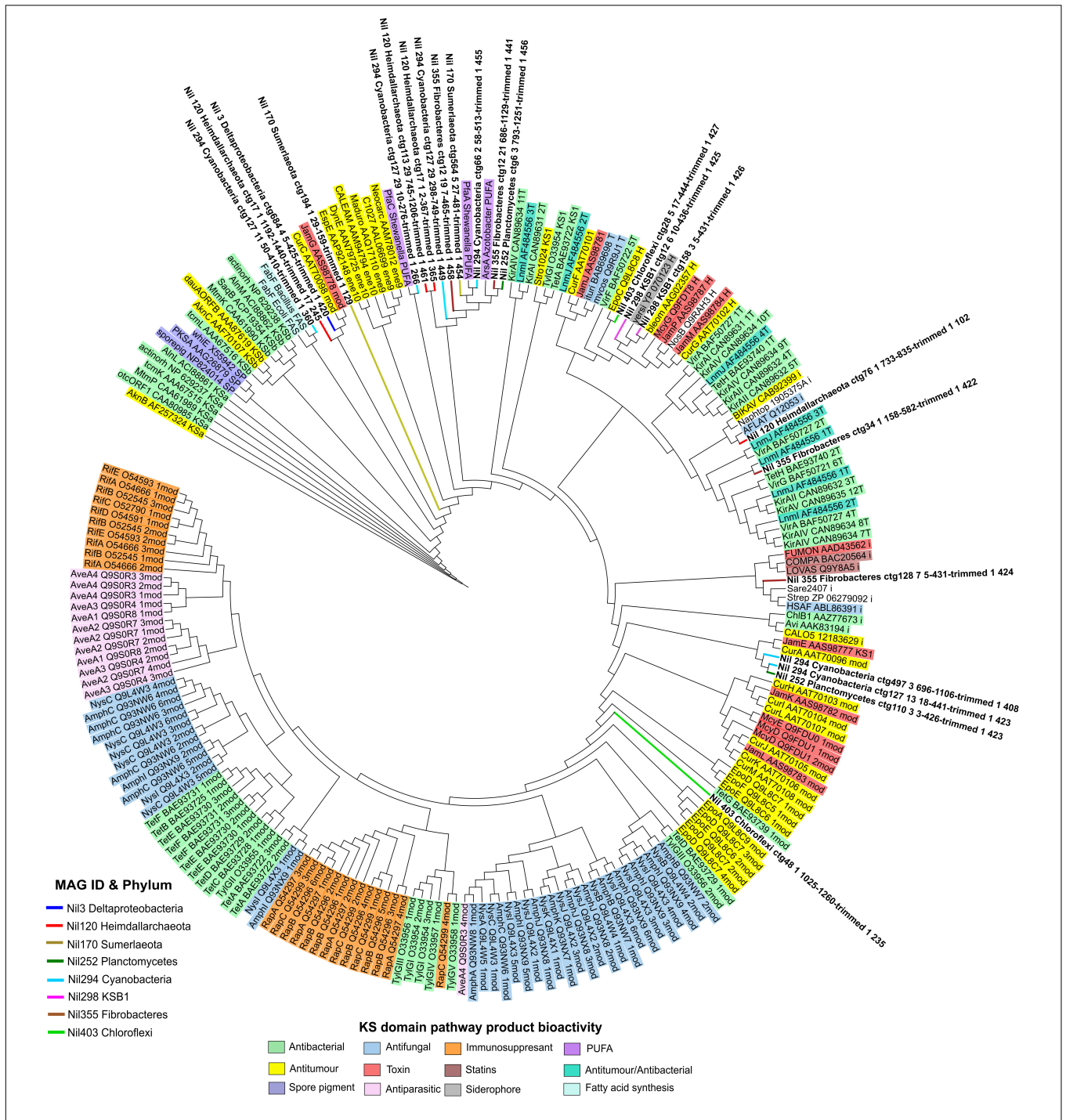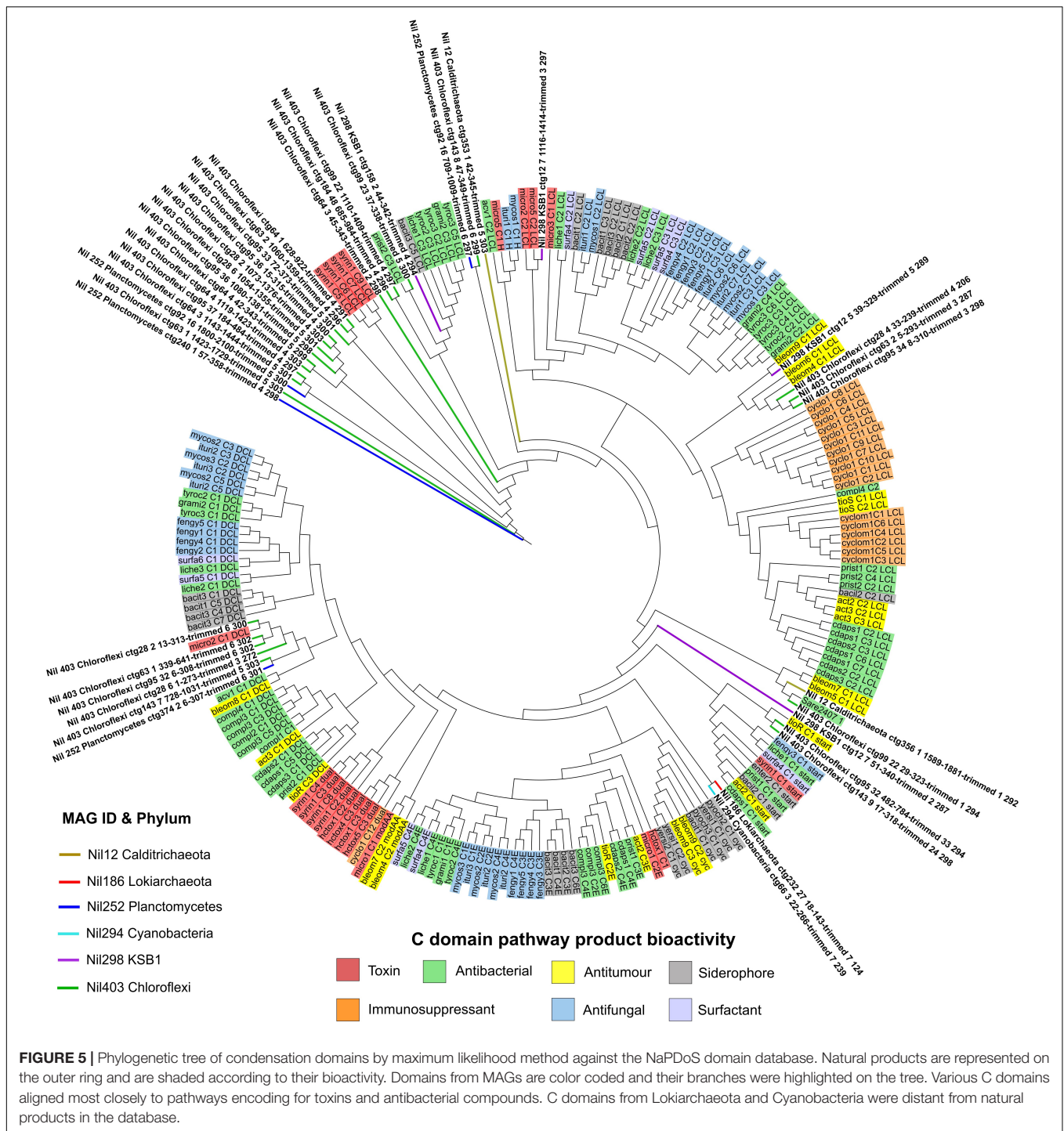
FIGURE 4 | Phylogenetic tree of ketosynthase domains by maximum likelihood method against the NaPDoS domain database. Natural products are represented on the outer ring and are shaded according to their bioactivity. Domains from MAGs are color coded and their branches are highlighted on the tree. Several KS domains from Deltaproteobacteria, Spirochaetes and Chloroflexi aligned closely to antitumor compounds. KS domains from Heimdallarchaeota were distant from any natural products in the database with PUFA compounds being the nearest. Several KS domains from Fibrobacteres, Planctomycetes, and Sumerlaeota aligned near poly-unsaturated fatty acid (PUFA) compounds.

essential elements to microorganisms (Ahmed and Holmström, 2014). Iron is an essential element for the growth of most microorganisms by acting as a catalyst in enzymatic processes such as oxygen metabolism and for stabilizing the polysaccharide

matrix during biofilm formation (Ahmed and Holmström, 2014). The identification of a KS domain in KSB1 with similarity to a siderophore compound suggests this phylum may play an important role in scavenging metals and biofilm formation in

FIGURE 5 | Phylogenetic tree of condensation domains by maximum likelihood method against the NaPDoS domain database. Natural products are represented on the outer ring and are shaded according to their bioactivity. Domains from MAGs are color coded and their branches were highlighted on the tree. Various C domains aligned most closely to pathways encoding for toxins and antibacterial compounds. C domains from Lokiarchaeota and Cyanobacteria were distant from natural products in the database.

microbial mats. Observations at this stage are speculative and will require future investigation to validate the presence of these PKS clusters in this candidate phylum.

Ketosynthase domains from Fibrobacteres and "Candidatus Sumerlaeota" aligned with PUFA from the *Shewanella* and *Azotobacter* genus with 54% and 53% similarity, respectively (**Figure 4** and **Supplementary Table S3**). PUFA are physiologically important compounds that take part in

facilitating the function of individual membrane proteins in bacteria (Yoshida et al., 2016). They have been observed to provide bacteria with membrane-shielding function against reactive oxygen species (ROS; Nishida et al., 2006). Consequently, high UV exposure and hypersaline conditions in the Shark Bay environment can induce greater amounts of ROS such as peroxide and hydroxyl radicals which are toxic to microbial communities (Wong et al., 2018). Due to the low similarity

score, we propose Fibrobacteres and Sumerlaeota may encode potentially novel PUFA compounds that may serve as a defense mechanism against elevated levels of ROS in microbial mats. This study also expands our knowledge of the understudied and unculturable phyla Fibrobacteres and Sumerlaeota which were previously not known to encode for PUFA.

CONCLUSION

Microorganisms living in Shark Bay microbial mats are exposed to harsh environmental conditions that require numerous survival responses. Microbial mat microorganisms may encode a wealth of natural products that serve various survival benefits and have been exploited by society in the field of medicine and biotechnology. Using a metagenomics approach, this study reports the abundance and variety of BGCs recovered from microbial mat microorganisms with the potential for some novel BGCs encoding for new natural products in deep-branching archaea and microbial dark matter. Microbial mat microorganisms that are exposed to harsh environmental stressors. This study has increased our understanding of microbial adaptation in the extreme environment of Shark Bay microbial mats and provides a framework for future studies mapping secondary metabolism pathways of microbial mat

microorganisms and their potential to encode for new natural products of interest.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.mg-rast.org/linkin.cgi?project=mgp81948, 4762868.3 to 4762965.3.

AUTHOR CONTRIBUTIONS

RC performed antiSMASH analyses on metagenomic data and drafted the manuscript. HW, GK, and FM conducted assembly, binning, and other specific data analyses. NB and BF provided detailed data analyses and interpretation. BB designed and led the study, and wrote specific manuscript sections. All authors edited the final manuscript.

SUPPLEMENTARY MATERIAL

REFERENCES

Ahmed, E., and Holmström, S. J. M. (2014). Siderophores in environmental research: roles and applications. *Microb. Biotechnol.* 7, 196–208. doi: 10.1111/1751-7915.12117

Allen, M. A., Goh, F., Burns, B. P., and Neilan, B. A. (2009). Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. *Geobiology* 7, 82–96. doi: 10.1111/j.1472-4669.2008.00187.x

Allen, M. A., Goh, F., Leuko, S., Echigo, A., Mizuki, T., Usami, R., et al. (2008). *Haloferax elongans* sp. nov. and Haloferax mucosum sp. nov., isolated from microbial mats from Hamelin Pool, Shark Bay, Australia. *Int. J. Syst. Evol. Microbiol.* 58, 798–802. doi: 10.1099/ijs.0.65360-0

Andrews, S. (2020). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed January 2, 2020).

Balskus, E. P., Case, R. J., and Walsh, C. T. (2011). The biosynthesis of cyanobacterial sunscreen scytonemin in intertidal microbial mat communities. *FEMS Microbiol. Ecol.* 77, 322–332. doi: 10.1111/j.1574-6941.2011.01113.x

Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L., and Hentschel, U. (2018). Marine Sponges as Chloroflexi hot spots: genomic insights and high-resolution visualization of an abundant and diverse symbiotic clade. *mSystems* 3:e00150-18. doi: 10.1128/mSystems.00150-18

Bernard, G., Pathmanathan, J. S., Lannes, R., Lopez, P., and Bapteste, E. (2018). Microbial dark matter investigations: how microbial studies transform biological knowledge and empirically sketch a logic of scientific discovery. *Genome Biol. Evol.* 10, 707–715. doi: 10.1093/gbe/evy031

Bláha, L., Babica, P., and Maršálek, B. (2009). Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdiscip. Toxicol.* 2, 36–41. doi: 10.2478/v10102-009-0006-2

Blin, K., Kim, H. U., Medema, M. H., and Weber, T. (2017a). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinform.* 20, 1103–1113. doi: 10.1093/bib/bbx146

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al. (2017b). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36–W41. doi: 10.1093/nar/gkx319

Bloudoff, K., and Schmeing, T. M. (2017). Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta ProteinsProteomics* 1865(11 Part B), 1587–1604. doi: 10.1016/j.bbapap.2017.05.010

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouma-Gregson, K., Olm, M. R., Probst, A. J., Anantharaman, K., Power, M. E., and Banfield, J. F. (2019). Impacts of microbial assemblage and environmental conditions on the distribution of anatoxin-a producing cyanobacteria within a river network. *ISME J.* 13, 1618–1634. doi: 10.1038/s41396-019-0374-3

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893

Brautaset, T., Sekurova, O. N., Sletta, H., Ellingsen, T. E., StrLm, A. R., Valla, S., et al. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* 7, 395–403. doi: 10.1016/s1074-5521(00)00120-4

Burns, B. P., Anitori, R., Butterworth, P., Henneberger, R., Goh, F., Allen, M. A., et al. (2009). Modern analogues and the early history of microbial life. *Precambrian Res.* 173, 10–18.

Burns, B. P., Seifert, A., Goh, F., Pomati, F., Jungblut, A.-D., Serhat, A., et al. (2005). Genetic potential for secondary metabolite production in stromatolite communities. *FEMS Microbiol. Lett.* 243, 293–301. doi: 10.1016/j.femsle.2004.12.019

Castelle, C. J., Brown, C. T., Anantharaman, K., Probst, A. J., Huang, R. H., and Banfield, J. F. (2018). Biosynthetic capacity, metabolic variety and unusual

biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645. doi: 10.1038/s41579-018-0076-2

Charlop-Powers, Z., Owen, J. G., Reddy, B. V., Ternei, M. A., and Brady, S. F. (2014). Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3757–3762. doi: 10.1073/pnas.1318021111

Charlop-Powers, Z., Owen, J. G., Reddy, B. V. B., Ternei, M. A., Guimarães, D. O., de Frias, U. A., et al. (2015). Global biogeographic sampling of bacterial secondary metabolism. *eLife* 4:e05048. doi: 10.7554/eLife.05048

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210

Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C., and Banfield, J. F. (2018). Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558, 440–444. doi: 10.1038/s41586-018-0207-y

Cuadrat, R. R. C., Ionescu, D., Dávila, A. M. R., and Grossart, H.-P. (2018). Recovering genomics clusters of secondary metabolites from lakes using genome-resolved metagenomics. *Front. Microbiol.* 9:251. doi: 10.3389/fmicb.2018.00251

Cummings, M., Breitling, R., and Takano, E. (2014). Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiol. Lett.* 351, 116–125. doi: 10.1111/1574-6968.12365

D'Agostino, P. M., Woodhouse, J. N., Liew, H. T., Sehnal, L., Pickford, R., Wong, H. L., et al. (2019). Bioinformatic, phylogenetic and chemical analysis of the UV-absorbing compounds scytonemin and mycosporine-like amino acids from the microbial mat communities of Shark Bay, Australia. *Environ. Microbiol.* 21, 702–715. doi: 10.1111/1462-2920.14517

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A. I. V., Bik, H. M., and Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. doi: 10.7717/peerj.243

Demain, A. L., and Fang, A. (2000). The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* 69, 1–39. doi: 10.1007/3-540-44964-7_1

Desai, J. D., and Banat, I. M. (1997). Microbial production of surfactants and their commercial potential. *Microbiol. Mol. Biol. Rev. MMBR* 61, 47–64. doi: 10.1128/.61.1.47-64.1997

Dobson, A., Cotter, P. D., Ross, R. P., and Hill, C. (2012). Bacteriocin production: a probiotic trait? *Appl. Environ. Microbiol.* 78, 1–6. doi: 10.1128/AEM.05576-11

Fisher, A., Wangpraseurt, D., Larkum, A. W. D., Johnson, M., Kühl, M., Chen, M., et al. (2018). Correlation of bio-optical properties with photosynthetic pigment and microorganism distribution in microbial mats from Hamelin Pool, Australia. *FEMS Microbiol. Ecol.* 95:fiy219. doi: 10.1093/femsec/fiy219

Galasso, C., Corinaldesi, C., and Sansone, C. (2017). Carotenoids from marine organisms: biological functions and industrial applications. *Antioxidants* 6:96. doi: 10.3390/antiox6040096

Graça, A. P., Calisto, R., and Lage, O. M. (2016). Planctomycetes as novel source of bioactive molecules. *Front. Microbiol.* 7:1241. doi: 10.3389/fmicb.2016.01241

Haft, D. H., Basu, M. K., and Mitchell, D. A. (2010). Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol.* 8:70. doi: 10.1186/1741-7007-8-70

Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., et al. (1989). Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. *J. Bacteriol.* 171, 1017–1027. doi: 10.1128/jb.171.2.1017-1027.1989

Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., et al. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res.* 8, 205–213. doi: 10.1093/dnares/8.5.205

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. doi: 10.7717/peerj.1165

Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Katz, L., and Baltz, R. H. (2016). Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* 43, 155–176. doi: 10.1007/s10295-015-1723-5

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Leuko, S., Allen, M. A., Goh, F., Burns, B. P., Walter, M. R., and Neilan, B. A. (2007). Analysis of intergenic spacer region length polymorphisms to investigate the halophilic archaeal diversity of stromatolites and microbial mats. *Extremophiles* 11, 203–210. doi: 10.1007/s00792-006-0028-z

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Linhart, C., and Shamir, R. (2005). The degenerate primer design problem: theory and applications. *J. Comput. Biol.* 12, 431–456. doi: 10.1089/cmb.2005.12.431

Liu, X. F., Xiang, L., Zhou, Q., Carralot, J.-P., Prunotto, M., Niederfellner, G., et al. (2016). Actinomycin D enhances killing of cancer cells by immunotoxin RG7787 through activation of the extrinsic pathway of apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10666–10671. doi: 10.1073/pnas.1611481113

Lopetuso, L. R., Giorgio, M. E., Saviano, A., Scaldaferri, F., Gasbarrini, A., and Cammarota, G. (2019). Bacteriocins and bacteriophages: therapeutic weapons for gastrointestinal diseases? *Int. J. Mol. Sci.* 20:183. doi: 10.3390/ijms20010183

Ludington, W. B., Seher, T. D., Applegate, O., Li, X., Kliegman, J. I., Langelier, C., et al. (2017). Assessing biosynthetic potential of agricultural groundwater through metagenomic sequencing: a diverse anammox community dominates nitrate-rich groundwater. *PLoS One* 12:e0174930. doi: 10.1371/journal.pone.0174930

Makarova, K. S., Wolf, Y. I., Karamycheva, S., Zhang, D., Aravind, L., and Koonin, E. V. (2019). Antimicrobial peptides, polymorphic toxins, and self-nonself recognition systems in archaea: an untapped armory for intermicrobial conflicts. *mBio* 10:e00715-19. doi: 10.1128/mBio.00715-19

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346. doi: 10.1093/nar/gkr466

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11:625. doi: 10.1038/nchembio.1890

Meleshko, D., Mohimani, H., Tracanna, V., Hajirasouliha, I., Medema, M. H., Korobeynikov, A., et al. (2019). BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. *Genome Res.* 29, 1352–1362. doi: 10.1101/gr.243477.118

Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi: 10.1021/acs.jnatprod.5b01055

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nishida, T., Orikasa, Y., Watanabe, K., and Okuyama, H. (2006). The cell membrane-shielding function of eicosapentaenoic acid for *Escherichia coli* against exogenously added hydrogen peroxide. *FEBS Lett.* 580, 6690–6694. doi: 10.1016/j.febslet.2006.11.030

Nutman, A. P., Bennett, V. C., Friend, C. R., Van Kranendonk, M. J., and Chivas, A. R. (2016). Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537, 535. doi: 10.1038/nature19355

Othoum, G., Bougouffa, S., Razali, R., Bokhari, A., Alamoudi, S., Antunes, A., et al. (2018). In silico exploration of Red Sea *Bacillus* genomes for natural product biosynthetic gene clusters. *BMC Genomics* 19:382. doi: 10.1186/s12864-018-4796-5

Pathak, J., Pandey, A., Maurya, P. K., Rajneesh, R., Sinha, R. P., and Singh, S. P. (2019). Cyanobacterial secondary metabolite scytonemin: a potential photoprotective and pharmaceutical compound. *Proc. Natl. Acad. Sci. India B Biol. Sci.* [Epub ahead of print]. doi: 10.1007/s40011-019-01134-5

Pattanaik, B., and Lindberg, P. (2015). Terpenoids And Their Biosynthesis In Cyanobacteria. *Life* 5, 269–293. doi: 10.3390/life5010269

Parks, D. H., Imelfortm, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. *Genome Res.* 25, 1043–1055.

Perry, R. D., and Fetherston, J. D. (2011). Yersiniabactin iron uptake: mechanisms and role in Yersinia pestis pathogenesis. *Microbes Infect.* 13, 808–817. doi: 10.1016/j.micinf.2011.04.008

Prieto-Barajas, C. M., Valencia-Cantero, E., and Santoyo, G. (2018). Microbial mat ecosystems: structure types, functional diversity, and biotechnological application. *Electron. J. Biotechnol.* 31, 48–56. doi: 10.1016/j.ejbt.2017.11.001

Raaijmakers, J. M., De Bruijn, I., Nybroe, O., and Ongena, M. (2010). Natural functions of lipopeptides from Bacillus and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol. Rev.* 34, 1037–1062. doi: 10.1111/j.1574-6976.2010.00221.x

Reddy, G. K., Leferink, N. G. H., Umemura, M., Ahmed, S. T., Breitling, R., Scrutton, N. S., et al. (2020). Exploring novel bacterial terpene synthases. *PLoS One* 15:e0232220. doi: 10.1371/journal.pone.0232220

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., et al. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66, 2541–2547. doi: 10.1128/aem.66.6.2541-2547.2000

Schwecke, T., Aparicio, J. F., Molnar, I., Konig, A., Khaw, L. E., Haydock, S. F., et al. (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7839–7843. doi: 10.1073/pnas.92.17.7839

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sharrar, A. M., Crits-Christoph, A., Méheust, R., Diamond, S., Starr, E. P., and Banfield, J. F. (2020). Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *mBio* 11:e00416-20. doi: 10.1128/mBio.00416-20

Siegl, A., and Hentschel, U. (2010). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ. Microbiol. Rep.* 2, 507–513. doi: 10.1111/j.1758-2229.2009.00057.x

Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N., and Martin, W. F. (2016). Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* 1:16034. doi: 10.1038/nmicrobiol.2016.34

Stewart, R. D., Auffret, M. D., Warr, A., Wiser, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9:870. doi: 10.1038/s41467-018-03317-6

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B. A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chem. Biol.* 7, 753–764. doi: 10.1016/S1074-5521(00)00021-1

Waldron, C., Matsushima, P., Rosteck, P. R. Jr., Broughton, M. C., Turner, J., Madduri, K., et al. (2001). Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa. *Chem. Biol.* 8, 487–499. doi: 10.1016/S1074-5521(01)00029-1

Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9259–9264. doi: 10.1073/pnas.1401734111

Watson, S. B. (2003). Cyanobacterial and eukaryotic algal odour compounds: signals or by-pdoucts? A review of their biological activity. *Phycologia* 42, 332–350. doi: 10.2216/i0031-8884-42-4-332.1

White, R. A. III,, Wong, H. L., Ruvindy, R., Neilan, B. A., and Burns, B. P. (2018). Viral communities of Shark Bay modern stromatolites. *Front. Microbiol.* 2018:1223. doi: 10.3389/fmicb.2018.01223

White, R. A. III, Visscher, P. T., and Burns, B. P. (2020). Between a rock and a soft place: viral role in stromatolite formation. *Trends Microbiol.* [Epub ahead of print]. doi: 10.1016/j.tim.2020.06.004

Wohlleben, W., Mast, Y., Stegmann, E., and Ziemert, N. (2016). Antibiotic drug discovery. *Microb. Biotechnol.* 9, 541–548. doi: 10.1111/1751-7915.12388

Wong, H. L., Smith, D.-L., Visscher, P. T., and Burns, B. P. (2015). Niche differentiation of bacterial communities at a millimeter scale in Shark Bay microbial mats. *Sci. Rep.* 5:15607. doi: 10.1038/srep15607

Wong, H. L., Visscher, P. T., White, I. I. I. R. A., Smith, D. L., Patterson, M. M., and Burns, B. P. (2017). Dynamics of archaea at fine spatial scales in Shark Bay mat microbiomes. *Sci. Rep.* 7:46160.

Wong, H. L., White, R. A., Visscher, P. T., Charlesworth, J. C., Vázquez-Campos, X., and Burns, B. P. (2018). Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. *ISME J.* 12, 2619–2639. doi: 10.1038/s41396-018-0208-8

Yoshida, K., Hashimoto, M., Hori, R., Adachi, T., Okuyama, H., Orikasa, Y., et al. (2016). Bacterial long-chain polyunsaturated fatty acids: their biosynthetic genes, functions, and practical use. *Mar. Drugs* 14:94. doi: 10.3390/md14050094

Zeng, Y., Baumbach, J., Barbosa, E. G., Azevedo, V., Zhang, C., and Koblížek, M. (2016). Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environ. Microbiol. Rep.* 8, 139–149. doi: 10.1111/1758-2229.12363

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7:e34064. doi: 10.1371/journal.pone.0034064

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