Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Remarkable biosynthetic capacity of Arctic hydrothermal biofilms

Runar Stokke

Runar.Stokke@uib.no

University of Bergen https://orcid.org/0000-0002-6056-7119

Thuc Nguyen

University of Bergen https://orcid.org/0000-0003-1089-5920

Ida Helene Steen

Univ Bergen https://orcid.org/0000-0001-5903-6276

Article

Keywords:

Posted Date: September 29th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7619139/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on September 3rd, 2026. See the published version at https://doi.org/10.1038/s41467-026-77384-5.

Abstract

Biosynthetic gene clusters (BGCs) which encode diverse secondary metabolites are ubiquitous across microbiomes and support critical ecological functions. They also serve as attractive targets for new drug discovery. Here we combined genome resolved metagenomics with long-read Nanopore RNA sequencing, yielding 1016 bacterial and 124 archaeal medium-to-high quality MAGs from previously unchartered microbial communities in Arctic hydrothermal vent biofilms. We identified 2965 BGCs from 870 metagenome-assembled genomes (MAGs) comprising a distinctive and rich diversity of BGCs, with ribosomally synthesized and post-translationally modified peptides (RiPPs) predominating across all samples. RiPPs and non-ribosomal peptide synthetases (NRPs) show strong antimicrobial potential, supported by transcriptomic expression. Terpenes, though less expressed, contribute to microbial signaling and defense. Notably, we identify hydrogen cyanide (HCN) synthesis pathways in archaeal genomes, challenging the view that cyanogenesis is restricted to bacteria and eukaryotes. Our findings demonstrate that microbial adaptation to extreme environments favors RiPP-based biosynthesis and that HCN may play a role in archaeal ecological interactions. Moreover, microbial communities in Arctic hydrothermal vent biofilms provide a rich reservoir of unique bioactive compounds, with implications for novel drug discovery.

Introduction

Antibiotic resistance is a growing global crisis, responsible for over 1.27 million deaths annually as of 2019[1]. This situation raises the need to find new antibiotics, particularly those derived from natural products, as many approved antibiotics have natural origins[1–3]. These bioactive compounds are typically synthesised by BGCs, which are physically co-located sets of genes encoding the enzymes, transporters, and regulatory elements required for secondary metabolite biosynthesis[4]. BGCs play important ecological roles by mediating chemical interactions within microbial communities[5]. Their products, known as secondary metabolites, can function as antimicrobials, toxins, siderophores, or signalling molecules, facilitating competition, cooperation, or adaptation in diverse environments[5]. BGCs are recognized as an important source of clinically valuable natural products, such as antibiotics (e.g., streptomycin, erythromycin), anticancer agents (e.g., doxorubicin), and immunosuppressants (e.g., rapamycin)[3]. Due to their essential biological activities, BGC-derived metabolites remain central to drug discovery efforts. In response to rising antibiotic resistance, genome mining of uncultivated microorganisms, particularly those from underexplored environments, is increasingly employed to identify novel BGCs with therapeutic potential.

Deep-sea hydrothermal vents are extreme environments characterised by high pressure, elevated temperatures, toxic chemical composition, and complete absence of sunlight. Unlike phototrophic ecosystems, life in hydrothermal vents is fueled by chemosynthetic primary production[6]. Microbial communities associated with vents exhibit extraordinary phylogenetic and functional diversity, with many taxa endemic to these environments[6–8]. Prokaryotic BGCs from various marine environments,

including water columns, deep-sea sediments, cold-seep sediments, and bacterial mats, have been shown to encode a wide array of novel secondary metabolites, many with antimicrobial properties[9–14]. In contrast, the microbial communities inhabiting deep-sea hydrothermal vent biofilms remain largely unexplored in this context, despite their biodiversity and potential to yield previously unrecognized biosynthetic pathways.

Biofilm-forming microorganisms play important ecological roles in vent ecosystems by structuring microbial assemblages and mediating nutrient flux[15]. Given the spatial and resource limitations within biofilms, microbial interactions in marine biofilms occur through various mechanisms, including cooperation, communication, and competition[16]. The production and secretion of secondary metabolites, such as toxins and antibiotics, are common strategies employed by microbes to compete within biofilm communities[16,17]. Despite this, deep-sea hydrothermal vent biofilms have been largely overlooked as a source of secondary metabolites, and to date, BGCs in biofilm communities from Arctic Mid-ocean Ridges (AMOR) hydrothermal systems have not been systematically described.

In this study, we explore the biodiversity and biosynthetic potential of biofilm microbial communities from the Loki's Castle, Aegir's Vent Field, and Seven Sisters vent field along the AMOR. By integrating Nanopore long-read metatranscriptomic data with BGC predictions from MAGs, we assess the transcriptional activity and site-specific expression profiles of BGCs across distinct vent ecosystems. Biofilm communities were taxonomically diverse, with bacteria dominating across all samples. Common deep-sea hydrothermal vent phyla such as *Pseudomonadota*, *Bacteriota*, *Aquificota*, and *Campylobacteota* were consistently present[14]. RiPPs represented the most prevalent BGC class, both at the genomic and transcriptomic levels, and were detected in both bacterial and archaeal lineages. Comparative analysis with the MIBiG database revealed that a large proportion of predicted BGCs do not cluster with any known reference, underscoring the uniqueness of hydrothermal vent biofilms as rich reservoirs of previously uncharacterized secondary metabolites. Notably, long-read transcriptomics data confirmed that many of these BGCs are actively expressed *in situ*, providing direct evidence that microbes within these extreme environments produce these secondary metabolites as part of their ecological strategy. This is the first comprehensive report on the microbial communities and biosynthetic capabilities of the Aegir's vent field, and the first observation of the potential for HCN synthesis in the domain Archaea, specifically from an *Asgardarchaeota*-affiliated MAG. Together, our findings reveal the ecological importance of secondary metabolism in hydrothermal vent biofilms and highlight their untapped potential for natural product discovery.

Results

Microbial Composition Across Biofilm Communities

To resolve the microbial diversity and biosynthetic capacity of Arctic hydrothermal vent biofilms, we performed an integrated metagenome- and metatranscriptome-based study of 9 unique hydrothermal

biofilms (Fig. 1, Table 1). The biofilms were collected from three distinct hydrothermal vent fields along AMOR, the shallow water Seven Sisters vent field, the deep-sea Aegir's vent field and the deep-sea Loki's Castle vent field. From the biofilms, we recovered 1147 MAGs. In addition, long-read nanopore metatranscriptomic data were obtained from 7 of the biofilms. Sequencing information is given in Supplementary Tables 1 and 2.

Table 1
Samples and collection sites used in the study.

Sample ID	Vent field	Metagenome ID	Metatranscriptome ID	Depth (m)	BioProject
GS17-BB	Loki's Castle Vent Field (Barite field)	M33	c33	2390	PRJNA1301837
GS21-106-ROV435A-BS	Seven Sisters Vent Field	M42	NA	165	PRJNA1301830
GS21-ROV-106-ROV432	Aegir's Vent Field	M24	c24	2288	PRJNA1301805
GS20-ROV9-BS1	Aegir's Vent Field	M25	c25	2288	PRJNA881934
GS20-ROV9-BS2	Aegir's Vent Field	M26	NA	2288	PRJNA1301797
GS22-107-ROV569-BS1	Aegir's Vent Field	M36	c36	2306	PRJNA1301809
GS22-107-ROV569-BS2	Aegir's Vent Field	M37	c37	2308	PRJNA1301812
GS22-107-ROV579-BS1	Aegir's Vent Field	M38	c38	2304	PRJNA1301820
GS22-107-ROV579-BS2	Aegir's Vent Field	M39	c39	2317	PRJNA1301824

After de-replication, 1140 medium-to-high quality bacterial and archaeal MAGs (> 2000 bp) were retained (Supplementary Table 3). Among them, 124 were classified using GTDB-tk with database version R220 into nine archaeal phyla, while 1016 MAGs belonged to 49 bacterial phyla. Within bacteria, *Pseudomonadota* was the most prevalent phylum, detected in all samples (327 MAGs), followed by *Bacteroidota* (166) (Fig. 2A). *Thermoproteota* was the dominant archaeal phylum, present in six of the nine metagenomes (Fig. 2B). Detailed taxonomic information is provided in the supplementary materials (Supplementary Table 3). Sample M25 exhibited the highest biodiversity, containing 254 MAGs from 25 bacterial and 8 archaeal phyla (Fig. 2A and 2B). Bacterial MAGs dominated all samples, but archaeal abundance was noticeably high in M25 and M39, with 60 and 24 archaeal MAGs, respectively. The archaeal taxa present suggest an elevated temperature in these samples, supported by the measured sediment temperature using a live temperature probe connected to the remotely operating vehicle (ROV).

Just below the biofilm in M39, the temperature was measured at 33°C, and at 10 cm sediment depth, the temperature was elevated to 96°C. The recently described phylum JALSQH01 from deep-sea hydrothermal deposits[7] was classified into three MAGs from sample M25. This suggests that this new taxon is more widely distributed and specific to the hydrothermal vent ecosystem.

Relative abundance analysis based on mapped reads to MAGs using CoverM, revealed that *Pseudomonadota* dominated seven of the nine biofilm communities, while *Campylobacterota* prevailed in samples M26 and M36. Notably, *Aquificota* demonstrates significant relative abundance in samples M25 (3.53%), M26 (11.85%), and M42 (10.56%). *Bacteroidota* is also an important bacterial phylum within the biofilm communities, contributing more than 1% to the total relative abundance in all samples except M26. Detailed information on the relative abundance of each phylum is provided in the supplementary materials (Supplementary Table 3). The presence of well-known deep-sea groups such as *Pseudomonadota*, *Campylomonadota*, *Bacteroidota*, *Aquificota*, *Chloroflexota*, and *Planctomycetota* aligns with other studies on deep-sea microbial communities [7,18–20]. *Desulfobacterota* was detected at > 1% relative abundance in samples M37, M38, and M39, suggesting adaptation to sulfur-rich environments. Sample M39 exhibited the highest diversity of this phylum, with 22 MAGs classified as *Desulfobacteria*. M25 showed the greatest diversity within the archaeal phylum *Thermoproteota*, with 38 MAGs classified to this phylum, accounting for 6.75% of its metagenome. 29 of these MAGs belonged to *Sulfolobales*, a group specialised in sulfur metabolism and known for thriving in high-temperature, acidic environments [21]. *Halobacteriota* was another archaeal phylum frequently observed in our biofilms, exhibiting a relative abundance of 1.2% in sample M38, where 4 MAGs were identified, and 2.32% in M39, with 8 MAGs detected.

Community dissimilarity analysis based on Bray-Curtis and Sørensen indices revealed distinct clustering among the samples. Bray-Curtis dissimilarities, which account for relative abundance, showed that sample M26 was significantly different from the others (0.78–0.90), with high bootstrap confidence intervals (upper CI > 0.95 in most pairwise comparisons) (Supplementary Fig. 1). In contrast, samples such as M24, M36, and M39 exhibited lower dissimilarity values (e.g., M24 and M39, 0.49), indicating more similar community structures. This was supported by Sørensen dissimilarities, which are based on the presence or absence of taxa, with M26 showing high dissimilarity (e.g., M26 and M37, 0.84; M26 and M42, 0.82), suggesting a distinct taxonomic composition (Supplementary Fig. 2). Most Sørensen p-values were not statistically significant ($p > 0.1$). These results indicate that M26 harbours a microbial community that is compositionally distinct from the other samples. Meanwhile, samples M24, M36, and M39 appear to form a more cohesive group in terms of both shared taxa and relative abundances.

Linking diversity and transcription of BGCs across hydrothermal biofilms

Metagenomic analysis of nine hydrothermal biofilm samples revealed a total of 2,965 BGCs across 870 MAGs (Fig. 3A). The dominant phylum, *Pseudomonadota*, encoded the largest number of BGCs (n =

1,258), followed by *Bacteroidota* (n = 335) and *Myxococcota* (n = 277). BGCs were distributed across 45 of 49 bacterial and 6 of 9 archaeal phyla, with *Nitrospirota* exhibiting the highest BGC density. Only 11.97% (n = 355) of BGCs were complete, while the majority (n = 2,610) were fragmented. RiPPs were the most abundant BGC class (n = 1,151), followed by Other (n = 638), Terpene (n = 394), NRP (n = 383), Polyketide (n = 236), Hybrid (n = 162), and Saccharide (n = 1) (Fig. 3B, Supplementary Table 4). RiPP-like clusters, reclassified from bacteriocins in antiSMASH v6.0 [22], were the most frequently predicted subtype (n = 633) (Fig. 3B; Supplementary Fig. 3A). RiPPs were consistently dominant across all biofilm samples (Fig. 3C), suggesting a conserved ecological role. In comparison, archaeal MAGs encoded 50 BGCs across six phyla, with lower diversity and representation limited to RiPP, Other, and Hybrid classes (Fig. 4).

To assess functional expression, metatranscriptomic profiling using long-read nanopore sequencing was successful on seven of the nine biofilm samples. A total of 517 BGCs were actively transcribed across 266 MAGs, spanning 28 microbial phyla: 24 bacterial (Fig. 5). and 4 archaeal (Fig. 6). Expression patterns closely mirrored metagenomic abundance, whereby *Pseudomonadota*, which encoded the highest number of BGCs, also exhibited the highest transcriptional activity (n = 265 expressed BGCs), followed by *Bacteroidota* (n = 79) and *Desulfobacterota* (n = 37). Sample M39 showed the highest overall BGC expression (n = 205 clusters) (Fig. 5B). The MAG Barite_M33_B89 (*Pseudomonadota*, candidate genus *JABSOG01*) exemplified the link between genomic potential and transcriptional output. This MAG encoded 107 BGCs, spanning NRPS (n = 74), polyketides (n = 15), RiPPs (n = 8), Hybrids (n = 6), and Other (n = 4), and exhibited the highest number of expressed clusters (n = 26), positioning it as a key biosynthetic contributor within the biofilm community.

RiPPs were the most transcriptionally active BGC class (n = 158), consistent with their genomic prevalence (Fig. 5B and 6B). Expression levels, quantified as transcripts per million (TPM), are detailed in Supplementary Table 5. Statistical analysis (Kruskal–Wallis with Dunn's post hoc test) revealed significant differences in expression among BGC classes (Supplementary Fig. 6B). RiPPs showed significantly higher expression than Hybrid ($p = 0.0004$), Polyketide ($p = 0.0005$), and Terpene ($p = 0.0134$) clusters after Bonferroni correction. No significant differences were observed among Hybrid, NRP, Other, Polyketide, and Terpene groups ($p > 0.05$), indicating a distinct transcriptional profile for RiPPs. Notably, a non-alpha poly-amino acid (NAPAA) BGC from MAG Aegir_M37_B22 (*Pseudomonadota*, candidate genus *HyVt-477*) exhibited the highest transcriptional abundance (TPM = 80929), despite being the only NAPAA BGC detected, suggesting a specialized ecological function (Supplementary Table 5).

Functional Roles of Expressed BGCs in Biofilm Adaptation

Several BGCs linked to microbial adaptation were both genomically encoded and transcriptionally active. Aryl polyene (APE) BGCs (n = 253), known to enhance oxidative stress tolerance and biofilm formation[23], were consistently expressed across all metatranscriptomes, particularly in *Pseudomonadota*,

Bacteroidota, *Campylobacterota*, *Desulfobacterota*, and *Myxococcota* (Fig. 7, Supplementary Fig. 3B, Supplementary Table 5). Ectoine BGCs (n = 14), primarily from *Pseudomonadota*, were detected in eight metagenomes and actively transcribed in M38 and M39, suggesting roles in osmotic and thermal stress mitigation [24]. BGCs associated with nutrient acquisition and regulation included phosphonates (n = 17), thioamitides (n = 8), and siderophores (n = 2) [9,25]. Quorum sensing and biofilm regulation were supported by homoserine lactone (hserlactone) BGCs (n = 100) and ranthipeptides (n = 51) [26–28]. While hserlactone BGCs were present in all metagenomes, transcriptional activity was restricted to four metatranscriptomes (M24, M25, M33, M39; Supplementary Table 5) and confined to *Pseudomonadota* (Fig. 7), indicating phylum-specific regulatory roles.

Hydrothermal biofilms harbored a diverse array of antimicrobial BGCs, many of which were actively transcribed. (Fig. 7) RiPP subclasses included thiopeptides (n = 23) [29], lanthipeptides (n = 20) [20,30], lasso peptides (n = 8) [31], sactipeptides (n = 6) [32], and linear azol(in)e-containing peptides (LAPs; n = 4) [33]. Additional antimicrobial classes included β-lactones (n = 73) [34], N-acyl amino acids (n = 65) [35], resorcinols (n = 34) [36], NRPS (n = 203) [37], and polyketide synthases (T1PKS: n = 78; T3PKS: n = 96) [38,39]. The MAG Barite_M33_B89 (phylum *Pseudomonadota*, candidate genus *JABSOG01*) was a major contributor to antimicrobial potential, expressing BGCs across multiple classes (Supplementary Table 5). These BGCs spanned five classes: NRPS (74), polyketides (15), RiPPs (8), Hybrids (6), and Other (4). Given that NRPs, polyketides, and RiPPs are renowned to produce diverse antimicrobial compounds[9,40,41], this understudied genus, JABSOG01, represents a compelling candidate for future studies on novel natural bioactive compounds. The MAG Aegir_M24_B72, classified within the phylum *Bacteroidota*, contained the longest BGC (105.77 kb; Supplementary Table 4, predicted to encode a novel NRPS–T1PKS hybrid.

Network Analysis of Biosynthetic Gene Clusters

To evaluate the uniqueness of BGCs recovered from biofilm communities, we performed a comprehensive network analysis comparing BGCs identified in nine biofilm samples with those curated in the MIBiG database (version 3.1). In total, 5,450 BGCs were analyzed, comprising 2,497 reference BGCs from MIBiG and 2,953 BGCs from our metagenomic dataset. These BGCs were clustered into 4,953 gene cluster families (GCFs), of which 4,643 were singletons and 310 contained two or more BGCs (Fig. 8A). Strikingly, only a single GCF included BGCs from both the biofilm-derived dataset and the MIBiG reference set, underscoring the pronounced divergence between these two sources. This minimal overlap highlights the distinct biosynthetic repertoire of biofilm-associated microorganisms and suggests the presence of novel metabolic pathways not represented in existing databases. The high proportion of singleton GCFs further supports the hypothesis that biofilm communities harbor a largely untapped reservoir of unique biosynthetic potential. Pairwise cluster sharing analysis revealed substantial variability in BGC composition among biofilm samples. For instance, samples M24 and M39 shared 75 GCFs, as did M25 and M39, while M24 and M25 shared 50 GCFs (Fig. 8B). Sample M36

exhibited moderate overlap with M39 (46 shared GCFs), M25 (25), and M24 (22), indicating partial biosynthetic similarity. In contrast, M26 was markedly distinct, sharing only 9 GCFs with M39 and between 2–4 GCFs with other samples. Notably, only a single GCF was shared between MIBiG and the biofilm samples M36, M37, and M38, further reinforcing the novelty of the biofilm-derived BGCs.

Previously unrecognized biosynthetic capacities in Archaea

A total of 62 BGCs associated with hydrogen cyanide (HCN) synthesis were identified, including one in the archaeal MAG Aegir_M39_B184, taxonomically assigned to the class *Heimdallarchaeia* within the phylum *Asgardarchaeota* (Supplementary Fig. 4A and 4B). HCN is a well-characterized virulence factor predominantly found in bacterial taxa such as *Pseudomonas* spp. [42]. Its biosynthesis is mediated by HCN synthase, encoded by the canonical *hcnABC* operon, which catalyzes the oxidative decarboxylation of glycine to yield carbon dioxide and HCN [43], a process termed cyanogenesis. This pathway is tightly regulated by environmental conditions, including iron and phosphate availability, as well as microaerobic conditions [43]

In contrast to the tripartite *hcnABC* operon of *Pseudomonas fluorescens* (BGC0002345, MIBiG 3.1), the putative HCN synthase in Aegir_M39_B184 is encoded by two genes. Structural modelling using the AlphaFold3 tripartite webserver and comparative superimposition with *hcnA* and *hcnB* from *P. fluorescens* (Supplementary Fig. 5), supports the presence of a complete and potentially functional HCN synthase. Transcriptomic data confirmed expression of both genes (Supplementary Table 5). Comparative genomic analysis using antiSMASH v7.0 and the MIBiG 3.1 database revealed a homologous BGC architecture in another *Asgardarchaeota* genome, *Candidatus Prometheoarchaeum syntrophicum* strain MK-D1 (NZ_CP042905.1; Supplementary Fig. 4C), further supporting the potential for archaeal cyanogenesis.

Additionally, the MAG Aegir_M39_B193, classified within the phylum *Thermoproteota*, harbors a BGC predicted to encode the biosynthesis of 2-deoxystreptamine, a central scaffold of aminoglycoside antibiotics such as streptomycin, kanamycin, and gentamicin [44]. This compound is predominantly associated with bacterial lineages, particularly within the phyla *Actinomycetota* and *Bacillota* [45–47]. The presence of aminoglycoside biosynthetic pathways in archaeal genomes is exceedingly rare and, to our knowledge, has not previously been reported in hyperthermophilic archaea such as *Thermoproteota* [48].

Discussion

Deep-sea hydrothermal microbiomes, shaped by extreme environmental pressures and intense microbial competition, represent promising reservoirs of novel bioactive compounds. Despite their potential, the biosynthetic landscape of Arctic deep-sea hydrothermal biofilms remains largely unexplored. In this study, we investigated nine biofilms from three Arctic hydrothermal vent sites, analyzing BGC diversity across 1,140 archaeal and bacterial MAGs. Our findings underscore the biosynthetic richness of these

biofilms and highlight their potential as untapped sources of antimicrobial agents. The microbial consortia inhabiting Arctic hydrothermal biofilms exhibit remarkable taxonomic and metabolic diversity, encompassing a broad spectrum of bacterial and archaeal phyla. BGCs serve as adaptive molecular tools, conferring selective advantages that promote microbial persistence and competitive success in these extreme environments. Notably, the M26 biofilm displayed a distinct community composition despite originating from the same vent field (Aegir's vent field) as six other samples, suggesting that biofilm assembly is influenced not only by environmental parameters but also by intra-community interactions. Diversity metrics based on abundance and presence/absence data support niche specialization, as shown by the unique BGC repertoire identified in M26. These patterns point to a metabolically distinct microbial consortium structured within this biofilm. The average number of BGCs per MAG (2.60) in the AMOR biofilms exceeds the global marine average, emphasizing the influence of habitat on secondary metabolism [13]. For instance, Antarctic soils yielded 1,417 bacterial BGCs with no archaeal BGCs, whereas cold seep communities contained 238 archaeal BGCs among 487 MAGs, predominantly encoding RiPPs [9,49]. These observations suggest that selective pressures in deep-sea ecosystems favor the retention and expression of RiPP BGCs as a survival strategy. RiPPs emerged as the most abundant BGC class across all nine biofilms, contrasting with other extreme environments such as Antarctic soils and Whalers Bay sediments, where terpenes predominate [49,52]. This divergence suggests that adaptation to hydrothermal conditions may favor RiPP biosynthesis, potentially due to their roles in microbial antagonism and communication. The consistent dominance of RiPPs across samples, rather than being driven by outliers, indicates community-wide selection for these compounds. This pattern aligns with global surveys of marine microbial genomes, which also report RiPPs and terpenes as the most prevalent BGC classes [13,53]. Moreover, RiPP BGCs in our dataset encode a diverse array of antimicrobial peptides, many of which have previously demonstrated *in vitro* antimicrobial activity [13].

From an ecological perspective, aryl polyene BGCs, encoding widespread bacterial pigments [23], were consistently detected and transcriptionally active across all samples, indicating their potential role in microbial fitness and biofilm stability. Phosphonate BGCs, which contribute to marine phosphorus cycling [25], were identified in 18 MAGs across five biofilms and seven bacterial phyla. Despite their frequent occurrence in marine and cold seep prokaryotic genomes[9,10], transcriptomic activity was restricted to the Barite field biofilm (phylum *Bacteroidota*), likely reflecting site-specific geochemical conditions. In contrast, siderophore BGCs were rare, with only two detected (M25 and M26, phylum *Pseudomonadota*) and no detected expression, suggesting limited iron acquisition strategies in these biofilms.

Thioamitide BGCs, previously reported in the archaeal phylum *Halobacteriota* from cold seeps [9], were also identified in our dataset, spanning bacterial phyla *Actinomycetota* and *Pseudomonadota*, as well as archaeal phyla *Halobacteriota* (orders ANME-1 and *Alkanophagales*). These clusters mediate thioamidation, a biochemical process replacing an amide oxygen with sulfur [50]. Prior studies proposed that *Halobacteriota* members utilize thioamide pathways to modify methyl-coenzyme M reductase,

facilitating anaerobic oxidation of methane (AOM) [9]. Consistent with this, MAG Aegir_M38_B90 (ANME-1) exhibited transcriptomic expression of a thioamitide BGC, indicating active AOM metabolism. Another ANME-1 MAG (Aegir_M38_B4) expressed ranthipeptides, metabolites previously considered bacterial and only recently reported in archaea [28,51], expanding their known distribution and suggesting a role in ANME-1 adaptability. Collectively, these findings demonstrate that BGC-encoded metabolites contribute to microbial resilience in extreme environments and influence biofilm dynamics through selective pressures, interspecies signaling, and physiological protection. Furthermore, NRPS BGCs, another prominent class of antimicrobial producers, were detected in eight of nine biofilms. M33 was notably enriched, harboring 91 of the 203 NRPS clusters. Transcriptomic data revealed active expression of 60 NRPS BGCs, including 30 from M33, affiliated with diverse hydrothermal vent phyla such as *Pseudomonadota*, *Planctomycetota*, *Campylobacterota*, *Chloroflexota*, *Calditrichota*, *Bacteroidota*, and candidate phylum KSB1. These data suggest that NRPS-derived metabolites play a significant role in microbial competition and niche establishment. Terpenes, though less abundant than RiPPs and NRPS, were detected in all metagenomes (394 BGCs), exclusively in bacterial taxa. Their known roles in microbial signaling and defense [54], coupled with transcriptomic evidence from 56 BGCs, mostly in M39 and absent in M37, highlight their functional relevance and underscore the metabolic heterogeneity among biofilm communities [13].

The discovery of HCN biosynthetic pathways in archaeal genomes expands the known metabolic repertoire of this domain and challenges the long-standing assumption that cyanogenesis is restricted to bacteria and eukaryotes. This suggests that archaea may utilize HCN as a chemical mediator in ecological interactions, particularly within the competitive and chemically dynamic environment of deep-sea biofilms. Given HCN's established role in microbial antagonism and defense [55], its presence in archaea raises intriguing questions about convergent evolution or horizontal gene transfer events. The broader taxonomic distribution of cyanogenic BGCs observed here, extending beyond traditional producers such as *Pseudomonas* and *Cyanobacteria* [56], indicates that HCN synthesis may be a more widespread and ecologically relevant trait in marine microbial communities. The expression of these BGCs in multiple biofilm samples further supports their functional role *in situ*, suggesting that cyanogenesis contributes to niche differentiation, microbial competition, and defense strategies in hydrothermal ecosystems. These findings prompt a reevaluation of archaeal roles in chemically mediated interactions and highlight the need for further investigation into the evolutionary origins and regulatory mechanisms of archaeal cyanogenesis. Future studies integrating comparative genomics, transcriptomics, and metabolomics will be essential to elucidate the functional significance of HCN production and its impact on biofilm community dynamics.

In summary, the diversity and expression of BGCs across AMOR biofilms reflect finely tuned adaptive strategies to the physicochemical pressures of hydrothermal vent ecosystems. The broad spectrum of predicted bioactive compounds, many with antimicrobial properties, positions these biofilm-associated microbiomes as a valuable and underexplored resource for biotechnological and pharmaceutical innovation.

Material and Methods

Sampling sites and collection

Biofilm samples were collected at three hydrothermal locations, both from shallow water and deep-sea, along the AMOR. In the current study 9 biofilms were collected, one from the shallow water Seven Sisters vent field at 130 mbsl [57], one from the diffuse venting at the Loki's Castle vent field (Barite field) situated at 2390 mbsl [58], and seven samples from the AVF at 2200 mbsl [59] (Table 1). Samples were collected using the Ægir6000 ROV on board the R/V G.O. SARS, equipped with a biosyringe (a hydraulic sampling cylinder) connected to the ROV manipulator arm. Sampling was conducted during sampling campaigns in 2017, 2020, 2021 and 2022 to AMOR led by the Centre for Deep Sea Research at the University of Bergen. Samples retrieved onboard R/V G.O. SARS were centrifuged at 6000 rcf for 5 minutes and the supernatant were removed. Aliquots were frozen in liquid nitrogen and stored at -80°C until processing.

Genome-resolved metagenomics

DNA was extracted using FastDNA Spin Kit for Soil (MP Biomedicals) from frozen samples and sent to the Norwegian Sequencing Centre (University of Oslo, Norway) for shotgun metagenomic sequencing. A 150-bp paired-end sequencing was performed using an Illumina NovaSeq S4 flow cell. Raw reads were scanned and filtered for quality, duplication rate, and adapter contamination using fastp (v0.23.2) [60].

Filtered reads from individual biofilms were assembled into contigs by MEGAHIT (v1.2.9) [61] with parameters *--min-contig-len 2000 --presets meta-large*. Contigs with lengths longer than 2,000 bp were subjected to automatic MAG reconstruction using metaWRAP (v1.3.2) [62] which implements the combinatorial use of MaxBin2 (v2.2.6) [63], CONCOCT (v1.0.0) [64], and MetaBat2 (v2.12.1) [65] for MAG refinement. The quality of each MAG was estimated with CheckM1 (v1.0.12) and CheckM2 (1.1.0) and the low-quality genomes following the MIMAG standard (completeness < 50% and contamination > 10%) were removed from downstream analysis. MAGs with completeness ≥ 90% and contamination ≤ 5% were defined as near-complete genomes, MAGs with completeness ≥ 70% and contamination ≤ 10% defined as high-quality, and MAGs with completeness ≥ 50% and contamination ≤ 10% were defined as medium-quality[66].

De-replication of MAGs from all 9 biofilms were performed at the species level using dRep (v3.4.0) [67] with parameter settings *-l 40000 -comp 50 -con 10 --S_ani 98*. Abundance of de-replicated MAGs was estimated by mapping filtered reads from each biofilm sample to MAGs using CoverM (https://github.com/wwood/CoverM). Taxonomic classification was performed using the Genome Taxonomy Database [68,69] via the Genome Taxonomy Database Toolkit [70] (GTDB-Tk, v2.4.0 – GTDB R220 taxonomy) with the classify_wf using default parameters.

RNA extraction, complementary DNA synthesis, and Nanopore sequencing

RNA was extracted using the Quick-DNA/RNA Miniprep Kit (Zymo Research, USA), and subsequent RNA purification steps utilized HighPrep RNA Elite beads (MagBio Genomics Inc., USA). Ribosomal RNA (rRNA) was removed from total RNA samples for all samples, except for c25, with the Pan-Prokaryote riboPOOL (Bacteria and Archaea) (siTOOLs Biotech GmbH, Germany). Polyadenylation was performed using *E. coli* Poly(A) Polymerase (New England Biolabs, USA).

Complementary DNA (cDNA) synthesis was carried out using the Ligation Sequencing Kit V14, and sequencing was performed on the MinION Mk1C device (Oxford Nanopore Technologies plc, UK). The full description of the adapted protocol following the Nanopore sequencing guidelines for microbial metatranscriptomics is available at protocols.io (https://www.protocols.io/private/5CE72DE2F50B11EF9F770A58A9FEAC02). Sequencing utilized the Oxford Nanopore Flow Cell R10.4.1 (FLO-MIN114, Oxford Nanopore Technologies, UK).

Biosynthetic gene clusters; screening and expression profiling

Gene annotation was conducted on MAG sequences using the Prokka annotation pipeline (version 1.14.6)[71]. The dataset was divided into two separate runs for archaeal and bacterial MAGs. The following parameters were applied: *--addgenes*, *--force*, *--increment* 10, --compliant, *--centre* NorSeq, *--norrna*, and *--notrna*. Secondary metabolite BGCs were identified and analyzed using antiSMASH[72] (version 7.0) with the following options: *--cc-mibig, --cb-general, --cb-knownclusters, --cb-subclusters, --asf, --pfam2go, --smcog-trees, --rre*, and *--genefinding-tool* prodigal. Predicted BGCs were subjected to downstream analysis with BiG-SLiCE[73] (version 2.0.0) to extract features, construct GCF models, and assign GCFs.

A network analysis of recovered BGCs was performed using BiG-SCAPE v1.1.5 [74]. The program was executed in '--auto' mode. Networks were computed with the 'mix' and 'mibig' options, which analyse all classes together and incorporate MIBIG v3.1 [75] BGCs into the network. The '--include singletons' option was enabled to visualize singleton BGCs. Multiple raw distance cutoff values (ranging from 0.1 to 1.0) were tested, and networks generated with a 0.3 cutoff were selected for further analysis. Lower cutoff values are more suitable for grouping BGCs that produce identical compounds, whereas higher cutoffs provide a broader view of related product families. The resulting sequence similarity matrices were visualized using Cytoscape v3.10.3 [76]. Additionally, the network annotation table generated by BiG-SCAPE was included to differentiate between the two sets of BGCs and to categorise different BGC classes.

Nanopore sequencing reads were basecalled using Dorado (version 0.5.1) (https://github.com/nanoporetech/dorado) with low-quality reads removed (*--min-qscore 10*) and adapter sequences trimmed (*--trim*). Ribosomal RNA sequences were filtered out using SortMeRNA[77]. The resulting cDNA sequences were mapped against their corresponding metagenomic datasets using BiG-MAP[78] to profile the expression of BGCs. Mapping results from these samples were first reported in

Reads Per Kilobase of transcript per Million mapped reads (RPKM) and then converted to TPM for downstream analyses, using the conversion formula from Zhao et al. (2020)[79].

Abundance estimation of BGCs from metagenomes and metatranscriptomes

The relative abundance of biosynthetic core genes across different metagenomes was estimated using the BiG-MAP.map module with the Bowtie2 aligner[78]. The mapping function was used to map filtered paired-end reads from all nine biofilm samples onto a set of non-redundant, representative gene clusters for each GCF. Quality-filtered, preprocessed cDNA reads obtained from Nanopore sequencing were mapped against the biosynthetic core genes of BGCs predicted from the seven corresponding samples using the BiG-MAP.map module with the *-U* option (for unpaired reads). Mapping results were reported in RPKM, the default output of BiG-MAP. Subsequently, the results were converted to Genes Per Million (GPM) for metagenomic data and TPM for metatranscriptomic data, using the conversion formula from Zhao et al. (2020)[79], for statistics.

Statistical analyses and data visualisations

All statistical analyses were conducted using Python (version 3.13.2). Principal Component Analysis (PCA) of microbial communities across biofilm samples was performed using the scikit-learn package. Bray-Curtis dissimilarities based on relative abundances of MAGs were calculated using the pdist function from the scipy package. Kruskal-Wallis tests comparing the expression of different BGC classes were performed using the kruskal function from the scipy package. Post-hoc Dunn's tests to identify significant pairwise differences among groups were conducted with the scikit-posthocs package, using TPM for each BGC class across samples. Data visualisations were created using RAWGraphs[80] and Python's Matplotlib library, with all source data provided in the supplementary materials.

Declarations

Data availability

Metadata of all samples and metagenomes in the study is deposited at NCBI under BioProjects listed in Table 1. Data not included in the manuscript and the associated supplementary material can be made available upon request, where possible/permitted. The source code and data for all statistical analyses is available on GitHub (https://github.com/trongthucnguyen/DeepSeaQuence_biofilms).

Acknowledgements

We thank the ROV Ægir6000 operator team and the R/V G.O. SARS crew for assistance during sampling campaigns to the Arctic Mid-Ocean Ridges in 2017, 2020, 2021 and 2022. Bioinformatic analyses were performed on resources provided by Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. This study is supported by the project DeepSeaQuence funded

by the Norwegian Research Council (Grant No. 315427) and University of Bergen through the Centre for Deep Sea research (Grant No. TMS2020TMT13).

Author information

Thuc Trong Nguyen and R.S designed the study. Thuc Trong Nguyen, R.S and I.H.S collected samples at the Arctic Mid-Ocean Ridges and contributed to formal analysis. Thuc Trong Nguyen and R.S conducted bioinformatic analysis and visualization. Thuc Trong Nguyen wrote the manuscript, and R.S and I.H.S reviewed and revised the manuscript.

References

1. Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).
2. Hemmerling, F. & Piel, J. Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. *Nat. Rev. Drug Discov.* **21**, 359–378 (2022).
3. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
4. Medema, M. H. *et al.* Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
5. Sharrar, A. M. *et al.* Bacterial Secondary Metabolite Biosynthetic Potential in Soil Varies with Phylum, Depth, and Vegetation Type. *mBio* **11**, (2020).
6. Jannasch, H. W. & Taylor, C. D. DEEP-SEA MICROBIOLOGY. *Annu. Rev. Microbiol.* **38**, 487–487 (1984).
7. Zhou, Z., St. John, E., Anantharaman, K. & Reysenbach, A.-L. Global patterns of diversity and metabolism of microbial communities in deep-sea hydrothermal vent deposits. *Microbiome* **10**, 241 (2022).
8. Dahle, H. *et al.* Energy Landscapes in Hydrothermal Chimneys Shape Distributions of Primary Producers. *Front. Microbiol.* **9**, 1570 (2018).
9. Dong, X. *et al.* A vast repertoire of secondary metabolites potentially influences community dynamics and biogeochemical processes in cold seeps. *Sci. Adv.* **10**, eadl2281 (2024).
10. Paoli, L. *et al.* Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111–118 (2022).
11. Chen, R. *et al.* Discovery of an Abundance of Biosynthetic Gene Clusters in Shark Bay Microbial Mats. *Front. Microbiol.* **11**, 1950 (2020).
12. Mara, P. *et al.* Metagenomic profiles of archaea and bacteria within thermal and geochemical gradients of the Guaymas Basin deep subsurface. *Nat. Commun.* **14**, 7768 (2023).
13. Wei, B. *et al.* Global analysis of the biosynthetic chemical space of marine prokaryotes. *Microbiome* **11**, 144 (2023).
14. Zhou, Z., St. John, E., Anantharaman, K. & Reysenbach, A.-L. Global patterns of diversity and metabolism of microbial communities in deep-sea hydrothermal vent deposits. *Microbiome* **10**,

(2022).

15. O'Brien, C. E. *et al.* Microbial biofilms associated with fluid chemistry and megafaunal colonization at post-eruptive deep-sea hydrothermal vents. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **121**, 31–40 (2015).
16. Qian, P.-Y., Cheng, A., Wang, R. & Zhang, R. Marine biofilms: diversity, interactions and biofouling. *Nat. Rev. Microbiol.* **20**, 671–684 (2022).
17. Goel, N., Fatima, S. W., Kumar, S., Sinha, R. & Khare, S. K. Antimicrobial resistance in biofilms: Exploring marine actinobacteria as a potential source of antibiotics and biofilm inhibitors. *Biotechnol. Rep.* **30**, e00613 (2021).
18. Acinas, S. G. *et al.* Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* **4**, 604 (2021).
19. Mara, P. *et al.* Metagenomic profiles of archaea and bacteria within thermal and geochemical gradients of the Guaymas Basin deep subsurface. *Nat. Commun.* **14**, 7768 (2023).
20. Vander Roost, J., Thorseth, I. H. & Dahle, H. Microbial analysis of Zetaproteobacteria and co-colonizers of iron mats in the Troll Wall Vent Field, Arctic Mid-Ocean Ridge. *PLOS ONE* **12**, e0185008 (2017).
21. Liu, L.-J. *et al.* Physiology, Taxonomy, and Sulfur Metabolism of the Sulfolobales, an Order of Thermoacidophilic Archaea. *Front. Microbiol.* **12**, 768283 (2021).
22. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
23. Johnston, I. *et al.* Identification of essential genes for Escherichia coli aryl polyene biosynthesis and function in biofilm formation. *Npj Biofilms Microbiomes* **7**, 56 (2021).
24. Hermann, L. *et al.* The ups and downs of ectoine: structural enzymology of a major microbial stress protectant and versatile nutrient. *Biol. Chem.* **401**, 1443–1468 (2020).
25. Lockwood, S., Greening, C., Baltar, F. & Morales, S. E. Global and seasonal variation of marine phosphonate metabolism. *ISME J.* **16**, 2198–2212 (2022).
26. Shiner, E. K., Rumbaugh, K. P. & Williams, S. C. Interkingdom signaling: Deciphering the language of acyl homoserine lactones. *FEMS Microbiol. Rev.* **29**, 935–947 (2005).
27. Fuqua, C. & Greenberg, E. P. Listening in on bacteria: acyl-homoserine lactone signalling. *Nat. Rev. Mol. Cell Biol.* **3**, 685–695 (2002).
28. Chen, Y. *et al.* The SCIFF-Derived Ranthipeptides Participate in Quorum Sensing in Solventogenic Clostridia. *Biotechnol. J.* **15**, (2020).
29. Chan, D. C. K. & Burrows, L. L. Thiopeptides: antibiotics with unique chemical structures and diverse biological activities. *J. Antibiot. (Tokyo)* **74**, 161–175 (2021).
30. Ayikpoe, R. S. *et al.* A scalable platform to discover antimicrobials of ribosomal origin. *Nat. Commun.* **13**, 6135 (2022).

31. Jangra, M. *et al.* A broad-spectrum lasso peptide antibiotic targeting the bacterial ribosome. *Nature* **640**, 1022–1030 (2025).
32. Benjdia, A. & Berteau, O. Radical SAM Enzymes and Ribosomally-Synthesized and Post-translationally Modified Peptides: A Growing Importance in the Microbiomes. *Front. Chem.* **9**, (2021).
33. Travin, D. Y. *et al.* Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. *Nat. Commun.* **10**, (2019).
34. Robinson, S. L., Christenson, J. K. & Wackett, L. P. Biosynthesis and chemical diversity of β-lactone natural products. *Nat. Prod. Rep.* **36**, 458–475 (2019).
35. Brady, S. F. & Clardy, J. Long-Chain N-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA. *J. Am. Chem. Soc.* **122**, 12903–12904 (2000).
36. Romagnoli, C. *et al.* Antidermatophytic Action of Resorcinol Derivatives: Ultrastructural Evidence of the Activity of Phenylethyl Resorcinol against Microsporum gypseum. *Molecules* **21**, 1306 (2016).
37. Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**, 275 (2003).
38. Peters, L. *et al.* Polyketide synthase-derived sphingolipids mediate microbiota protection against a bacterial pathogen in C. elegans. *Nat. Commun.* **16**, (2025).
39. Navarro-Muñoz, J. C. & Collemare, J. Evolutionary Histories of Type III Polyketide Synthases in Fungi. *Front. Microbiol.* **10**, (2020).
40. Sayed, A. M. *et al.* Extreme environments: microbiology leading to specialized metabolites. *J. Appl. Microbiol.* **128**, 630–657 (2020).
41. Ortega, M. A. & van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31–44 (2016).
42. Chowdhury, N. & Bagchi, A. Structural Insight into the Gene Expression Profiling of the hcn Operon in Pseudomonas aeruginosa. *Appl. Biochem. Biotechnol.* **182**, 1144–1157 (2017).
43. Laville, J. *et al.* Characterization of the hcnABC Gene Cluster Encoding Hydrogen Cyanide Synthase and Anaerobic Regulation by ANR in the Strictly Aerobic Biocontrol Agent Pseudomonas fluorescens CHA0.
44. Park, S. R., Park, J. W., Ban, Y. H., Sohng, J. K. & Yoon, Y. J. 2-Deoxystreptamine-containing aminoglycoside antibiotics: Recent advances in the characterization and manipulation of their biosynthetic pathways. *Nat Prod Rep* **30**, 11–20 (2013).
45. Distler, J., Mansouri, K., Mayer, G., Stockmann, M. & Piepersberg, W. Streptomycin biosynthesis and its regulation in Streptomycetes. *Gene* **115**, 105–111 (1992).
46. Pissowotzki, K., Mansouri, K. & Piepersberg, W. Genetics of streptomycin production in Streptomyces griseus: molecular structure and putative function of genes strELMB2N. *Mol. Gen. Genet. MGG* **231**, 113–123 (1991).

47. Ota, Y. *et al.* Butirosin-biosynthetic Gene Cluster from Bacillus circulans. *J. Antibiot. (Tokyo)* **53**, 1158–1167 (2000).
48. Krause, K. M., Serio, A. W., Kane, T. R. & Connolly, L. E. Aminoglycosides: An Overview. *Cold Spring Harb. Perspect. Med.* **6**, a027029 (2016).
49. Waschulin, V. *et al.* Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *ISME J.* **16**, 101–111 (2022).
50. Liu, A. *et al.* Functional elucidation of TfuA in peptide backbone thioamidation. *Nat. Chem. Biol.* **17**, 585–592 (2021).
51. Song, Z. *et al.* Decoding the Chemical Language of Ribosomally Synthesized and Post-Translationally Modified Peptides from the Untapped Archaea Domain. *Angew. Chem. Int. Ed.* **64**, (2025).
52. Medeiros, W. *et al.* Unlocking the biosynthetic potential and taxonomy of the Antarctic microbiome along temporal and spatial gradients. *Microbiol. Spectr.* **12**, e00244-24 (2024).
53. Chen, J. *et al.* Global marine microbial diversity and its potential in bioprospecting. *Nature* **633**, 371–379 (2024).
54. Wei, X. *et al.* Exploring and expanding the natural chemical space of bacterial diterpenes. *Nat. Commun.* **16**, (2025).
55. Díaz-Rueda, P., Morales De Los Ríos, L., Romero, L. C. & García, I. Old poisons, new signaling molecules: the case of hydrogen cyanide. *J. Exp. Bot.* **74**, 6040–6051 (2023).
56. Anand, A. *et al.* Biological hydrogen cyanide emission globally impacts the physiology of both HCN-emitting and HCN-perceiving *Pseudomonas*. *mBio* **14**, e00857-23 (2023).
57. A. Marques, A. F. *et al.* The Seven Sisters Hydrothermal System: First Record of Shallow Hybrid Mineralization Hosted in Mafic Volcaniclasts on the Arctic Mid-Ocean Ridge. *Minerals* **10**, 439 (2020).
58. Steen, I. H. *et al.* Novel Barite Chimneys at the Loki's Castle Vent Field Shed Light on Key Factors Shaping Microbial Communities and Functions in Hydrothermal Systems. *Front. Microbiol.* **6**, (2016).
59. Pedersen, R. B. *et al.* Discovery of a black smoker vent field and vent fauna at the Arctic Mid-Ocean Ridge. *Nat. Commun.* **1**, 126 (2010).
60. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).
61. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–1676 (2015).
62. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
63. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).

64. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
65. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
66. The Genome Standards Consortium *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
67. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
68. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
69. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* btz848 (2019) doi:10.1093/bioinformatics/btz848.
70. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
71. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
72. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
73. Kautsar, S. A., van der Hooft, J. J. J., de Ridder, D. & Medema, M. H. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *GigaScience* **10**, giaa154 (2021).
74. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
75. Terlouw, B. R. *et al.* MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **51**, D603–D610 (2023).
76. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
77. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
78. Pascal Andreu, V. *et al.* BiG-MAP: an Automated Pipeline To Profile Metabolic Gene Cluster Abundance and Expression in Microbiomes. *mSystems* **6**, e00937-21 (2021).
79. Zhao, S., Ye, Z. & Stanton, R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* **26**, 903–909 (2020).
80. Mauri, M., Elli, T., Caviglia, G., Uboldi, G. & Azzi, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. in *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* 1–5 (ACM,

Cagliari Italy, 2017). doi:10.1145/3125571.3125585.

Figures

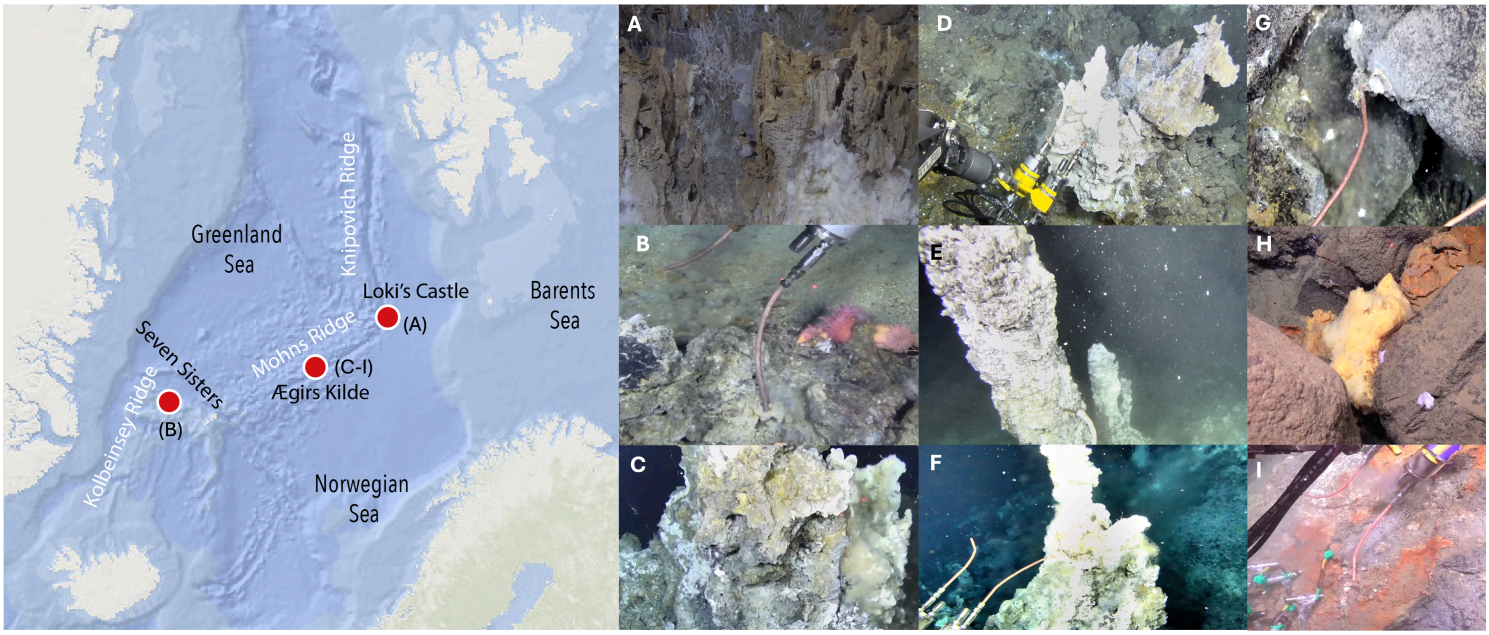

Figure 1

Biofilms collected from three distinct vent fields at AMOR (map left): **A;** the Barite Field within the Loki's Castle Vent Field (Table 1; Metagenome ID M33), **B;** the Seven Sisters Vent Field (Table 1; Metagenome ID M42), **C-I;** intra vent field sampling of biofilms from the AEGIR Vent Field (Table 1; Metagenome ID M24, M25, M26, M36, M37, M38 and M39). Samples were collected during four sampling campaigns to AMOR onboard the research vessel G.O SARS (Summer of 2017, 2020, 2021 and 2022).

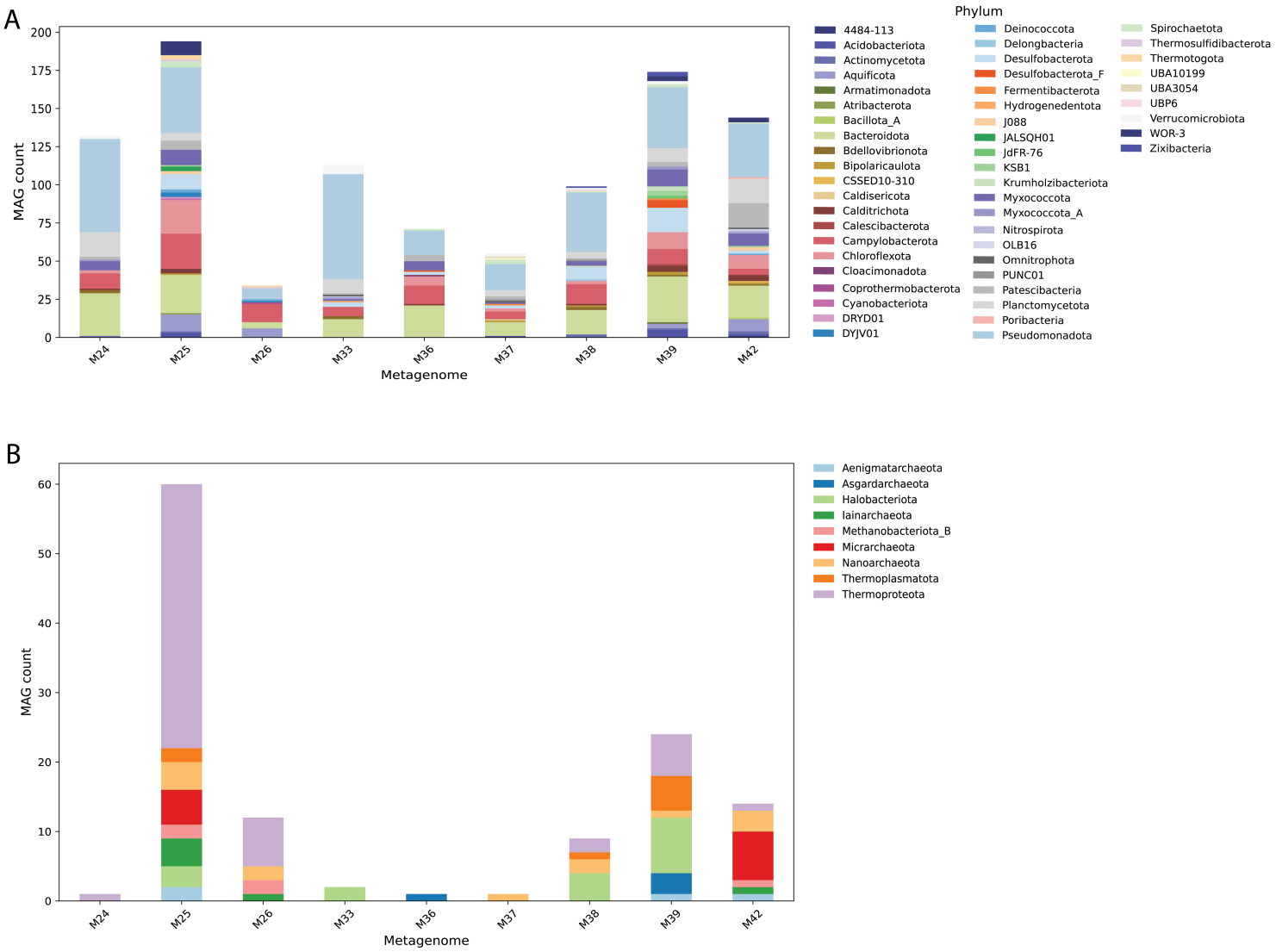

Figure 2

Distribution of microbial phyla in AMOR biofilms. The number of MAGs identified in each of the nine biofilms, grouped by phylum within: (A) Bacteria; (B) Archaea. Only medium- to high-quality MAGs, as defined by the MIMAG standards, were included. Taxonomic classifications are based on the Genome Taxonomy Database (GTDB v. R220).



Figure 3

Diversity of BGC classes and predicted products in AMOR biofilms showing (A) Abundance of individual BGC classes, (B) Predicted product types, with circle sizes indicating relative abundance, (C) Composition of BGC classes across nine metagenomes, excluding saccharides. BGCs are categorized based on the biosynthetic classes of their core enzymes, as assigned by antiSMASH v7.0 and BiG-SLiCE v2.0.0

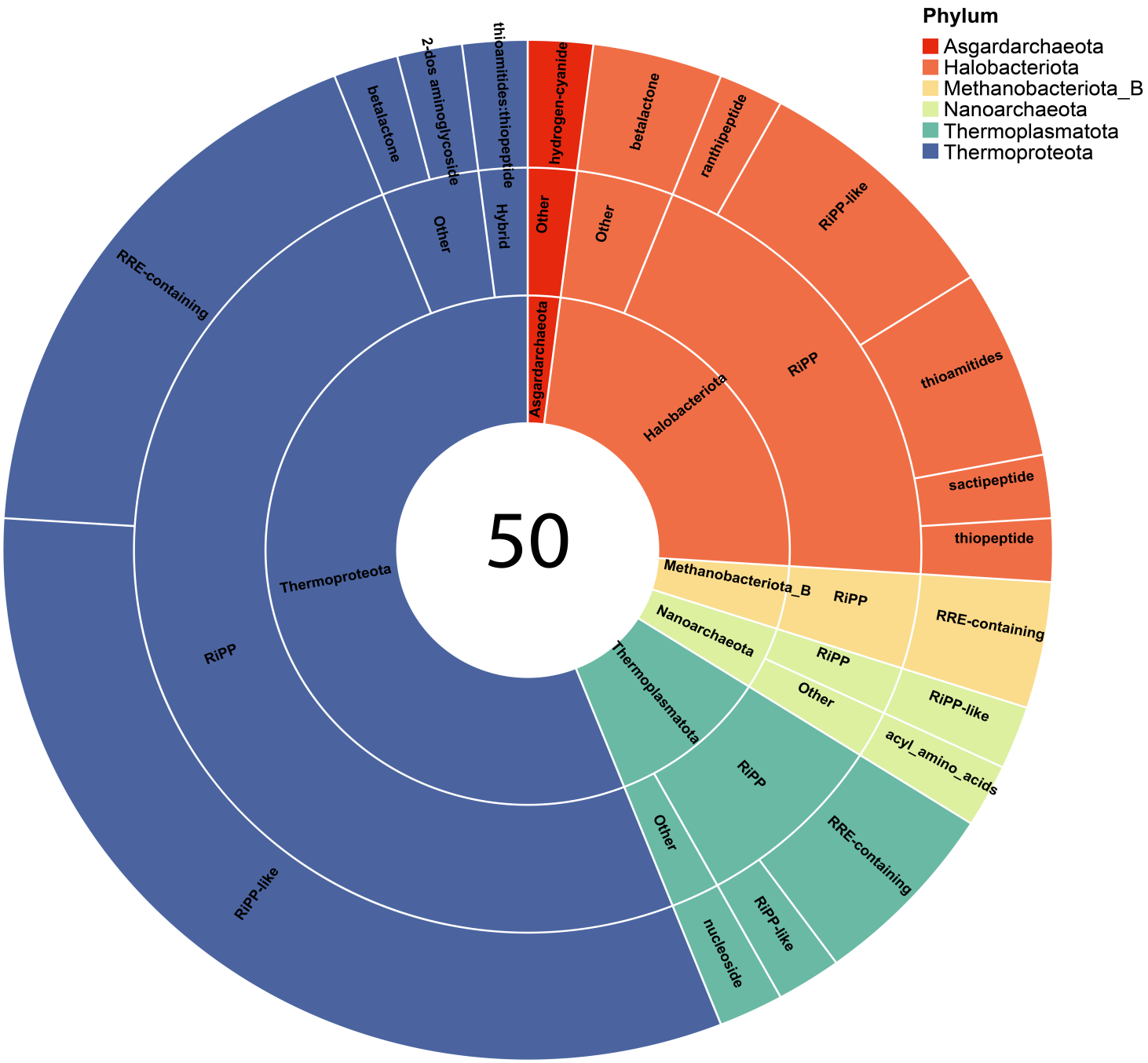

Figure 4

Distribution of BGCs and predicted metabolites in archaeal MAGs. The number and types of BGCs identified across archaeal phyla from nine biofilms, categorized by BGC class and predicted metabolite. Taxonomic classifications are based on GTDB release R220. BGC classes and predicted products were assigned using antiSMASH v7.0 and BiG-SLiCE v2.0.

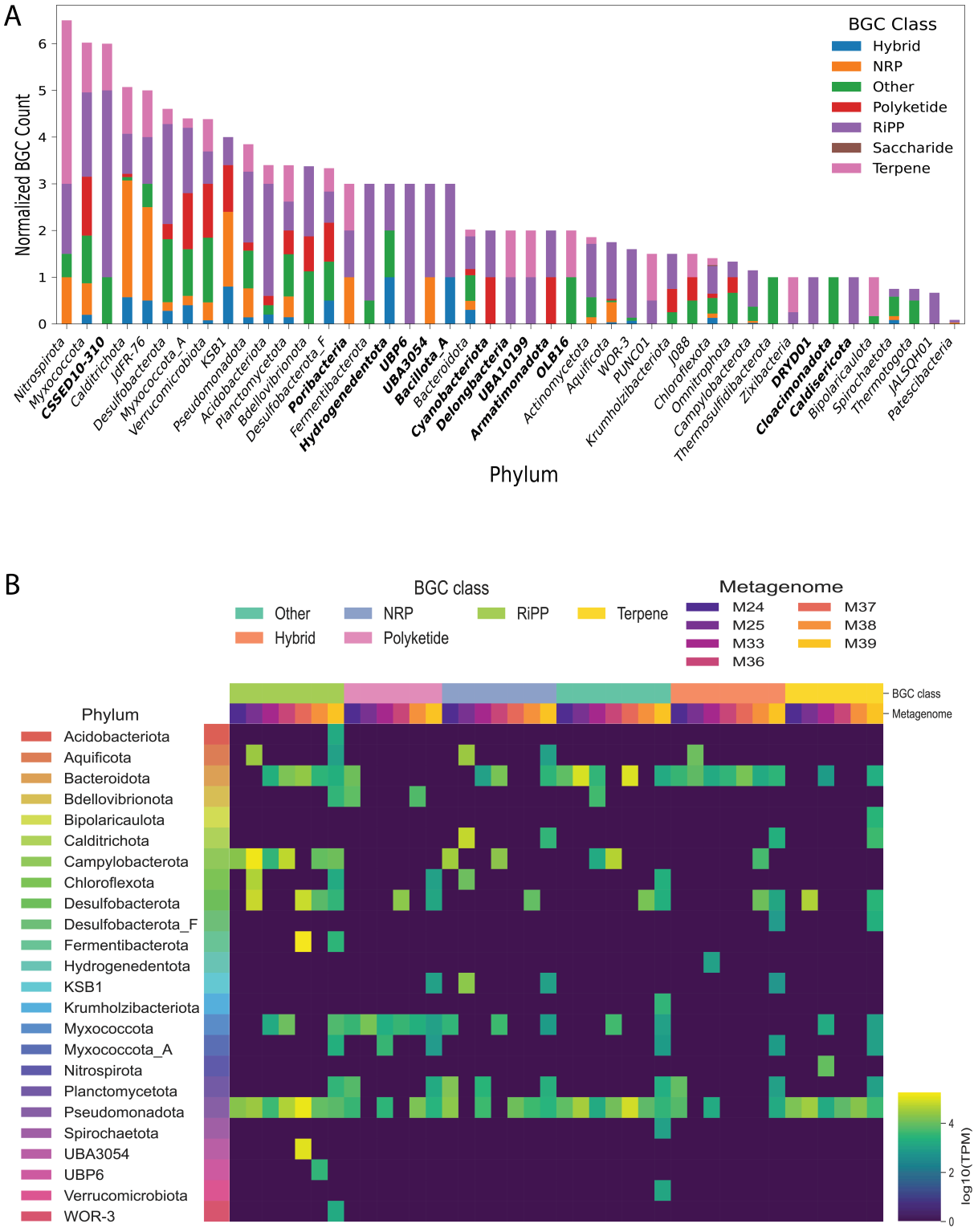

Figure 5

BGC richness across Bacterial phyla in the AMOR biofilms showing (A) normalized BGC counts across bacterial phyla in all samples, with phyla represented by a single MAG indicated in bold. Taxonomic classifications are based on GTDB release R220, and (B) expression profiles of bacterial BGC classes in AMOR biofilms. Heatmap show the expression levels of predicted BGC classes across the seven AMOR biofilms. Expression levels were estimated using BiG-MAP v1.0.0 and normalized as average $\log_{10}$-

transformed of TPM values. BGC detection and classification were performed using antiSMASH v7.0 and BiG-SLiCEv2.0.0

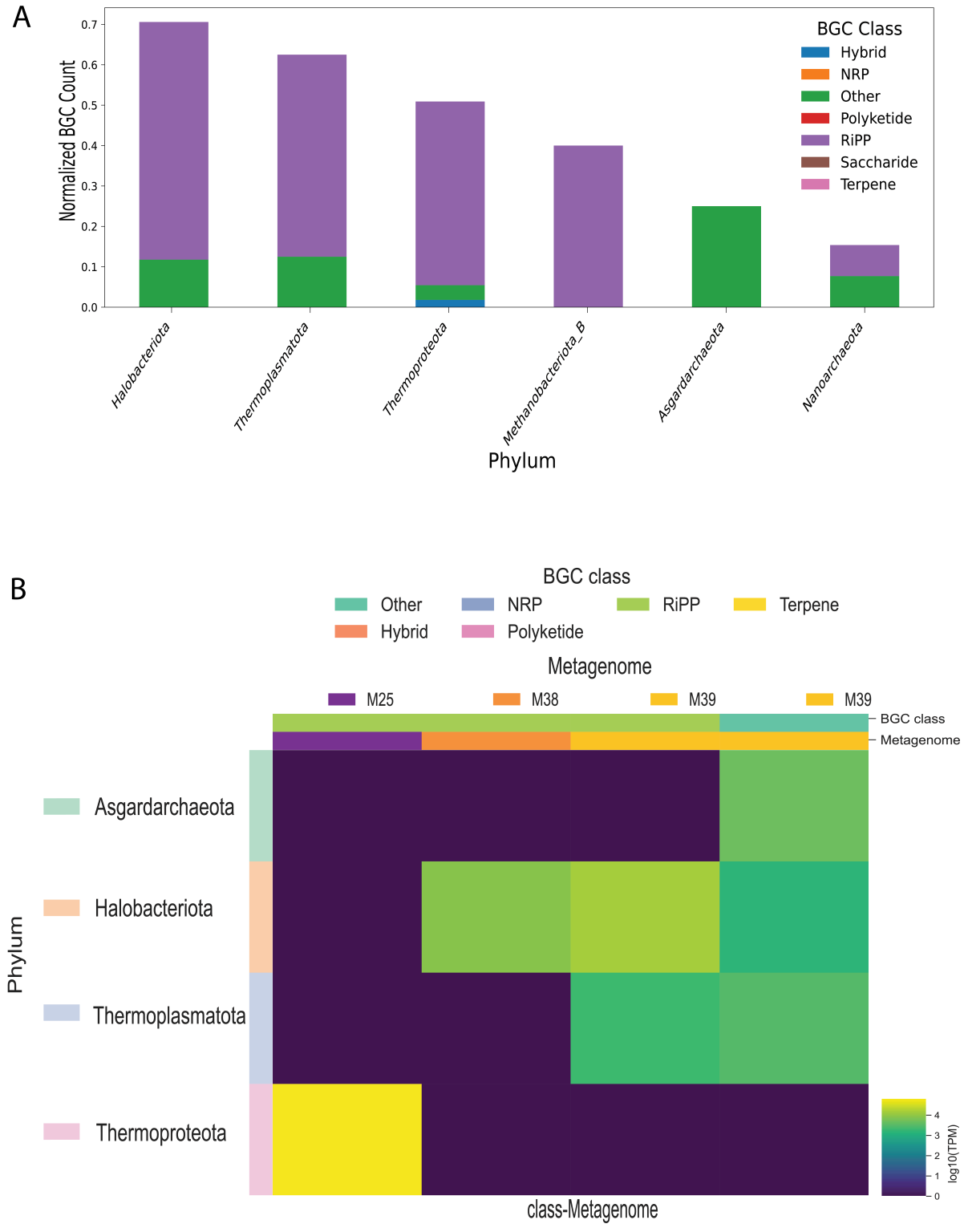

Figure 6

BGC richness across Archaeal phyla in AMOR biofilms showing (A) normalized BGC counts across archaeal phyla in all samples, and (B) expression profiles of archaeal BGC classes in the AMOR biofilms.

Heatmap show the expression levels of predicted BGC classes across the seven AMOR biofilms. Expression levels were estimated using BiG-MAP v1.0.0 and normalized as average log_{10}-transformed of TPM values. Taxonomic classifications are based on the GTDB release R220, BGC detection and classification were performed using antiSMASH v7.0 and BiG-SLiCEv2.0.0

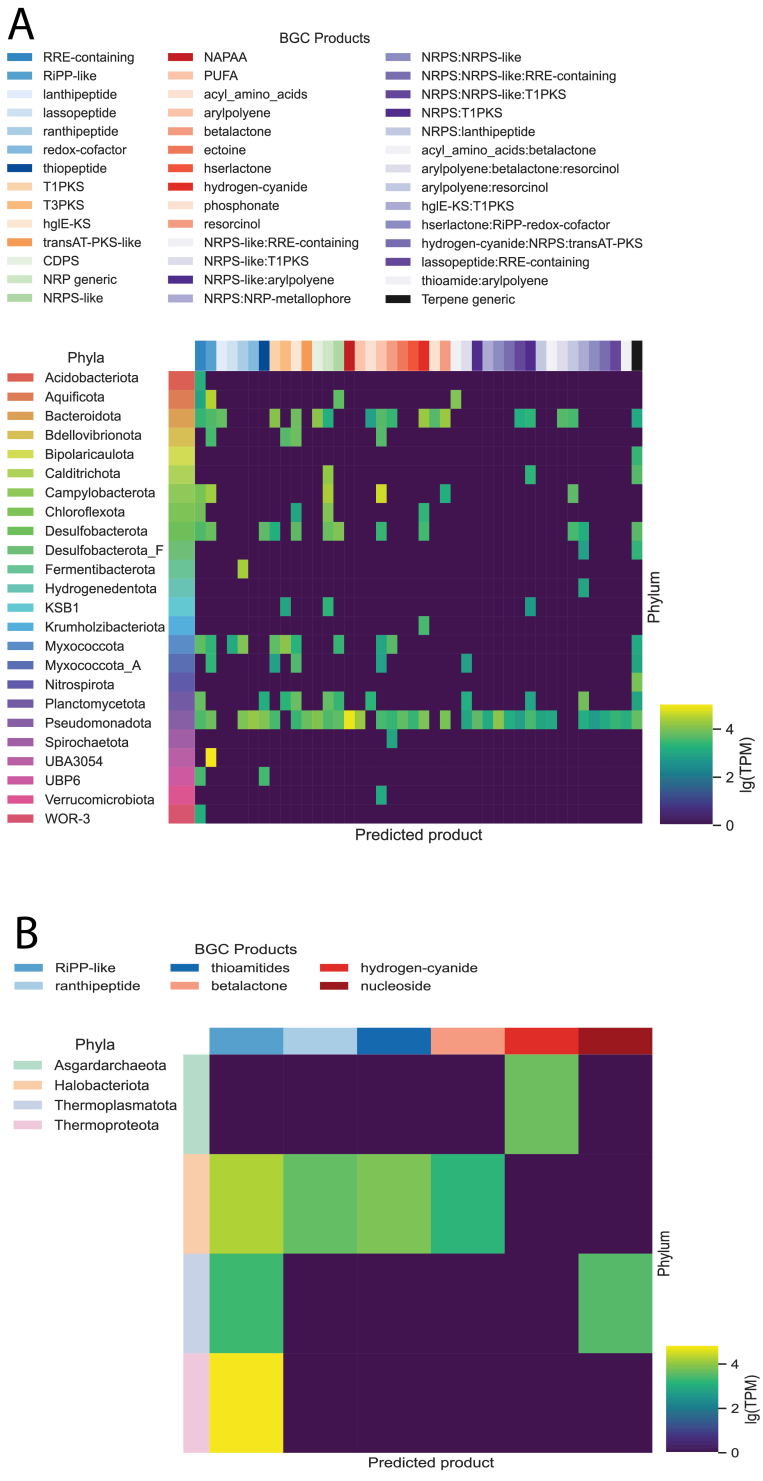

Figure 7

Expression profiles of BGC products in AMOR biofilms. Heatmaps show the expression levels of predicted BGC products across seven AMOR biofilms, separated by domain in (A) Bacteria, and (B) Archaea. Expression levels were estimated using BiG-MAP v1.0.0 and normalized as average log_{10}-transformed of TPM values.

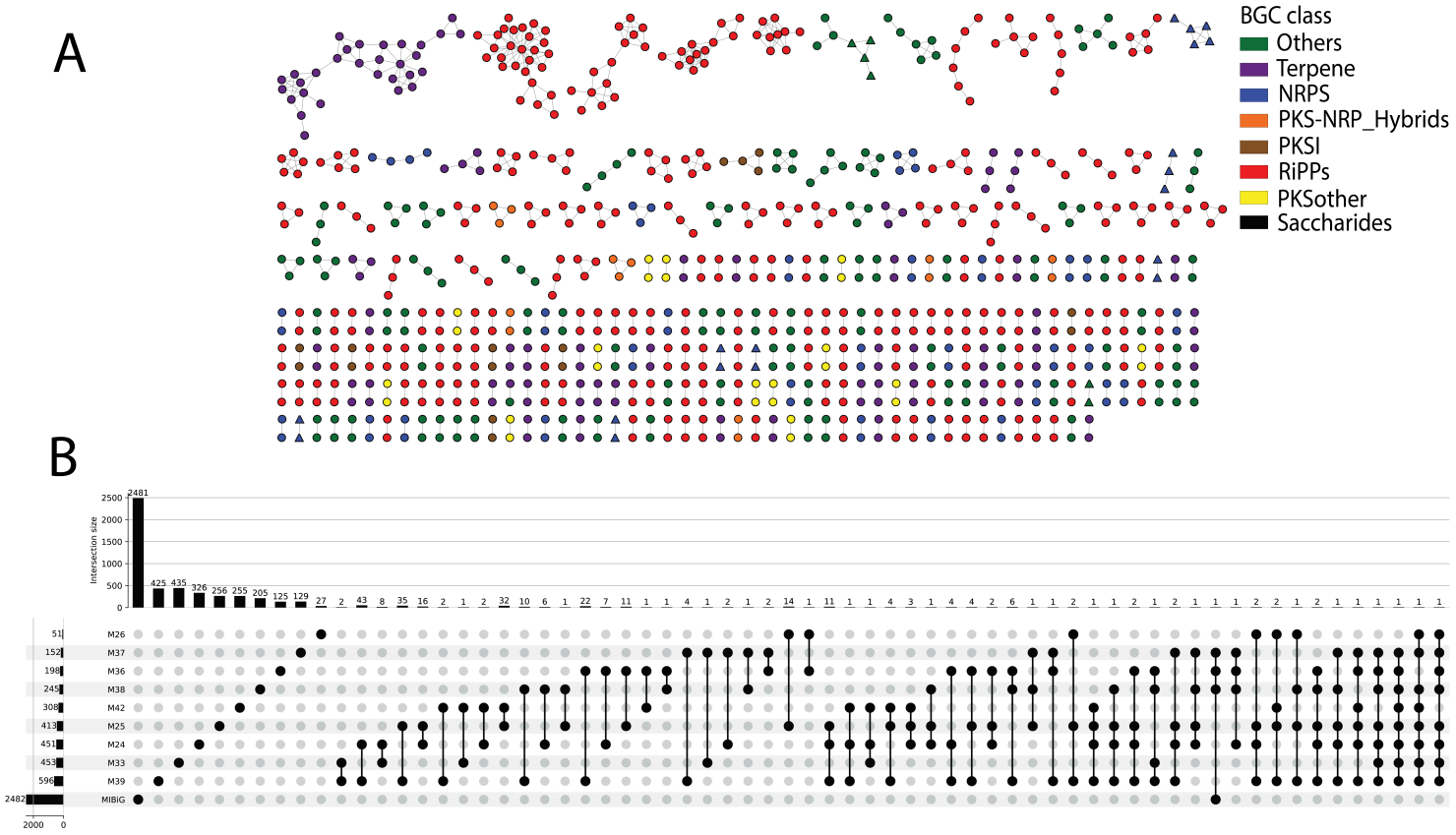

Figure 8

Cluster analysis of BGCs from AMOR biofilms. (A) Sequence similarity network of GCFs identified using BiG-SCAPE v1.1.5 (similarity cutoff = 0.3). Each node represents a BGC, coloured by class as assigned by BiG-SCAPE. Circular nodes indicate BGCs from AMOR biofilms, and triangular nodes represent reference BGCs from the MIBiG v3.1 database. Singletons are excluded for clarity. (B) UpSet plot showing the number of shared GCFs across biofilms, based on the sequence similarity network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.png

- FigureS5.png
- FigureS6.png
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx