

## MINI REVIEW

# Exploring the secrets of marine microorganisms: Unveiling secondary metabolites through metagenomics

Shaoyu Wang<sup>1,2</sup>  | Xinyan Li<sup>1,2</sup> | Weiqin Yang<sup>3</sup> | Ranran Huang<sup>1,2,4</sup>

<sup>1</sup>Institute of Marine Science and Technology, Shandong University, Qingdao, Shandong, China

<sup>2</sup>Qingdao Key Laboratory of Ocean Carbon Sequestration and Negative Emission Technology, Shandong University, Qingdao, China

<sup>3</sup>School of Computer Science and Technology, Shandong University, Qingdao, Shandong, China

<sup>4</sup>Global Ocean Negative Carbon Emissions (ONCE) Program Alliance, Qingdao, China

## Correspondence

Ranran Huang, Institute of Marine Science and Technology, Shandong University, 72 Binhai Road, Qingdao, Shandong 266237, China.  
Email: [huangrr@sdu.edu.cn](mailto:huangrr@sdu.edu.cn)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 32170065 and 31970113; Ocean Negative Carbon Emissions (ONCE) Program; Key Research and Development Program of Shandong Province, Grant/Award Number: 2020ZLYS04

## Abstract

Marine microorganisms are increasingly recognized as primary producers of marine secondary metabolites, drawing growing research interest. Many of these organisms are unculturable, posing challenges for study. Metagenomic techniques enable research on these unculturable microorganisms, identifying various biosynthetic gene clusters (BGCs) related to marine microbial secondary metabolites, thereby unveiling their secrets. This review comprehensively analyses metagenomic methods used in discovering marine microbial secondary metabolites, highlighting tools commonly employed in BGC identification, and discussing the potential and challenges in this field. It emphasizes the key role of metagenomics in unveiling secondary metabolites, particularly in marine sponges and tunicates. The review also explores current limitations in studying these metabolites through metagenomics, noting how long-read sequencing technologies and the evolution of computational biology tools offer more possibilities for BGC discovery. Furthermore, the development of synthetic biology allows experimental validation of computationally identified BGCs, showcasing the vast potential of metagenomics in mining marine microbial secondary metabolites.

## INTRODUCTION

The ocean, Earth's largest ecosystem, presents a complex ecological environment distinct from terrestrial ecosystems, characterized by high pressure, high salinity, and limited light (Nigam et al., 2019; Reuver et al., 2022). This unique marine setting necessitates specific adaptations in marine organisms, leading to distinctive metabolic patterns. Such distinctiveness harbours extensive potential for discovering new secondary metabolites (Bhatnagar & Kim, 2010; David et al., 2015; Jensen et al., 2015; Mehubub et al., 2014; Nigam et al., 2019; Subramani & Aalbersberg, 2013; Xiliang et al., 2021; Zhang et al., 2005). Marine organisms, which are home to an estimated 10 million species, constitute a significant

portion of the Earth's biomass. These organisms account for approximately 87% of the total carbon mass on the planet, highlighting their crucial role in the global carbon cycle and their substantial contribution to the Earth's overall biomass (Bar-On et al., 2018; Wang, 2016). Research in this area has already identified potential anticancer drugs and other bioactive substances in marine organisms, marking significant progress in medicine and biotechnology (Khalifa et al., 2019; Martins et al., 2014).

Marine microorganisms, constituting about 70% of the ocean's biomass, are an essential component of marine life (Bar-On et al., 2018). The intricate marine environment has driven the evolution of various unique metabolic pathways in these organisms. Traditional experimental methods, however, face limitations in

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Microbial Biotechnology* published by John Wiley & Sons Ltd.

comprehending the immense diversity of marine microorganisms. A vast number of undiscovered and uncultivable microbes remain in the ocean, posing significant challenges to revealing their secrets (Marmann et al., 2014; Romano et al., 2018).

Since its inception in 1998, metagenomics has undergone rapid development over the past 33 years. Initially focused on uncultivated soil microorganisms (Handelsman et al., 1998), the field, propelled by advances in sequencing technologies and bioinformatics, has broadened to encompass various disciplines (Breitwieser & Salzberg, 2019; Chiu & Miller, 2019; Gu et al., 2020; Key et al., 2017; Ko et al., 2022). This evolution in technology has significantly advanced the study of uncultivable marine microorganisms and opened up new avenues for exploring marine microbial life and their metabolites. The rapid progress of metagenomics provides an effective tool and framework for deepening our understanding of marine microorganisms and microbial communities across different ecosystems (Alma'abadi et al., 2015; Zhang et al., 2021). Such advancements offer scientists novel opportunities to expand the discovery of natural products, thereby playing a crucial role in uncovering unknown facets of marine microbiology (Coutinho et al., 2018; Kodzius & Gojoberi, 2015).

Investigating the secondary metabolites of marine microorganisms offer insights into their chemical communication, defence mechanisms, inter-microbial relationships, and food chain dynamics in marine ecosystems (Monti, 2023; Suryanarayanan, 2012; Teasdale et al., 2009). Viewing bioactive substances through an ecological lens provides a macroscopic perspective on the role of these compounds, foundational for marine chemical ecology research (Hay, 1996). Additionally, studies suggest that the count of biosynthetic gene clusters (BGCs) involved in producing secondary metabolites in microorganisms significantly exceeds the number of identified secondary metabolites (Hoshino et al., 2019; Knowles et al., 2022; Marmann et al., 2014; Zhuang & Zhang, 2021). Therefore, utilizing metagenomics to explore unexpressed or low-expressed BGCs is a crucial approach for discovering structurally novel and bioactive natural products from microorganisms. Consequently, employing metagenomics to examine unexpressed or low-expressed BGCs emerges as a key strategy for discovering structurally unique and bioactive natural products from microorganisms (Sekurova et al., 2019; Zhou et al., 2021).

## Multi-layered technological strategies in the study of novel secondary metabolites from marine microorganisms

### BGC databases

As research into secondary metabolites deepens, databases cataloguing various secondary metabolites have

been established. Currently, the analysis and comparison of BGCs primarily rely on these BGC databases. In 2013, Conway and Boddy (2012) founded the initial known product BGC database, ClusterMine360, now encompassing over 200 clusters from 185 compound families. This database also features a distinctive sequence library with over 10,000 polyketide synthase/non-ribosomal peptide synthetase structural domains (Conway & Boddy, 2012).

In 2015, Medema et al. (2015) launched the Minimum Information about a Biosynthetic Gene cluster (MIBiG) data standard, initially comprising 1170 BGC entries. MIBiG was updated to version 2.0 in 2019, incorporating 851 additional entries compared to its initial version (Kautsar et al., 2020). In 2022, the database progressed to version 3.0, adding 661 new BGCs and removing 69 entries due to low quality or redundancy (Terlouw et al., 2023). MIBiG is now extensively used in developing and evaluating various BGC prediction programmes.

With the advancement of computational biology, databases containing predicted BGCs derived from various software tools have also been compiled, complementing those based on actual products. In 2021, Kautsar, Blin, et al. (2021), Kautsar, van der Hooft, et al. (2021) created BiG-FAM, which includes predictions from 1,225,071 BGCs sourced from 209,206 publicly available microbial genomes and metagenome-assembled genomes. Such databases of BGCs are invaluable as foundational resources for discovering new secondary metabolites.

### Computational biology approach to mining more BGCs

With the ongoing advancement in secondary metabolite research, an increasing number of these compounds and their BGCs have been identified. As data accumulates, the application of artificial intelligence (AI) methods to the study of secondary metabolites has become increasingly feasible. This advancement allows AI technology to gradually permeate secondary metabolite research, offering more possibilities and new research pathways.

antiSMASH, introduced in 2011, has rapidly evolved into the most widely utilized tool for detecting and identifying BGCs in archaea, bacteria, and fungi (Blin et al., 2023). ClusterFinder was developed in 2014 and based on a dual-state Hidden Markov Model (HMM) (Cimermanic et al., 2014). DeepBGC, launched in 2019, employed a detailed process, including gene ORF detection, protein domain analysis, domain-level and protein-level BGC detection, post-processing, and classification. It set a new standard using the MIBiG 1.3 database, integrating advanced methods like BiLSTM and RF (Hannigan et al., 2019).

Machine learning-based programmes for predicting secondary metabolites have found extensive application across diverse research domains (Yuan et al., 2023). The ongoing development and enhancement of these tools equip researchers with efficient, precise, and interpretable methods, fuelling further progress in the exploration of secondary metabolites.

## Cluster analysis of BGC

For the effective identification of numerous BGCs from metagenomic datasets using computational biology, selecting a subset of representative and promising BGCs for further analysis is a more feasible approach. Recent studies have begun utilizing computational methods to group similar BGCs, using known BGCs of the same type to predict potential products, functions, and evolutionary histories of newly identified BGCs (Doroghazi et al., 2014; Duncan et al., 2015). This strategy may also reveal entirely new BGC types, offering directions for discovering novel secondary metabolites (Cimermanic et al., 2014; Nielsen et al., 2017). To facilitate this, Navarro-Muñoz et al. in Navarro-Muñoz et al., 2020 developed a computational workflow that integrates the “biosynthetic gene similarity clustering and prospecting engine” (BiG-SCAPE) and “core analysis of syntenic orthologues to prioritize natural product gene clusters” (CORASON). This workflow rapidly groups BGCs into gene cluster families (GCFs) and further categorizes them into gene cluster clans (Navarro-Muñoz et al., 2020). In 2021, BiG-SLiCE was introduced, significantly reducing the computational resources needed for grouping BGCs into GCFs and broadening the method's applicability (Kautsar, Blin, et al., 2021; Kautsar, van der Hooft, et al., 2021).

In summary, the evolution of BGC databases and computational biology has significantly propelled the study of secondary metabolites. These advancements have not only streamlined the identification and analysis of BGSs but also opened new avenues for discovering novel compounds. By leveraging advanced computational tools and methodologies, researchers can now delve deeper into the complex world of secondary metabolites, paving the way for ground-breaking discoveries in marine microbiology and beyond.

## Metagenomics reveals novel secondary metabolites from marine microorganisms

Metagenomics technology offers a ground-breaking perspective in understanding the potential of new secondary metabolites from marine microorganisms, given the vastness of the oceans (Figure 1). Paoli and colleagues have demonstrated the significant exploration

potential of these metabolites across the entire ocean. They compiled approximately 10,000 microbial genomes from cultured bacteria and single cells, and over 25,000 reconstructed metagenomes from more than 1000 seawater samples. Employing antiSMASH, they identified about 40,000 potential novel BGCs and discovered bacteria in the *Ca. Eudoremicrobiaceae* group, noted for their robust secondary metabolite synthesis capabilities (Paoli et al., 2022).

In their study, focusing on specific oceanic regions, Rego and colleagues assessed the diversity of metagenomic biosynthesis in the Arctic Ocean. They successfully recovered 149 ketosynthase (KS) domain operational taxonomic unit sequences, with 36% being unassignable to any known BGC. Furthermore, from 74 bacterial metagenome-assembled genomes, they extracted 179 BGCs, underscoring the Arctic Ocean's potential as a source of novel secondary metabolites (Rego et al., 2021).

In the Northwest Pacific, the study by Wang et al., 2023 identified 1195 BGCs in metagenomes obtained from surface seawater, with numerous terpene synthase clusters and Pfam domains for squalene/phytoene synthesis demonstrating the microbial biosynthetic potential of the region.

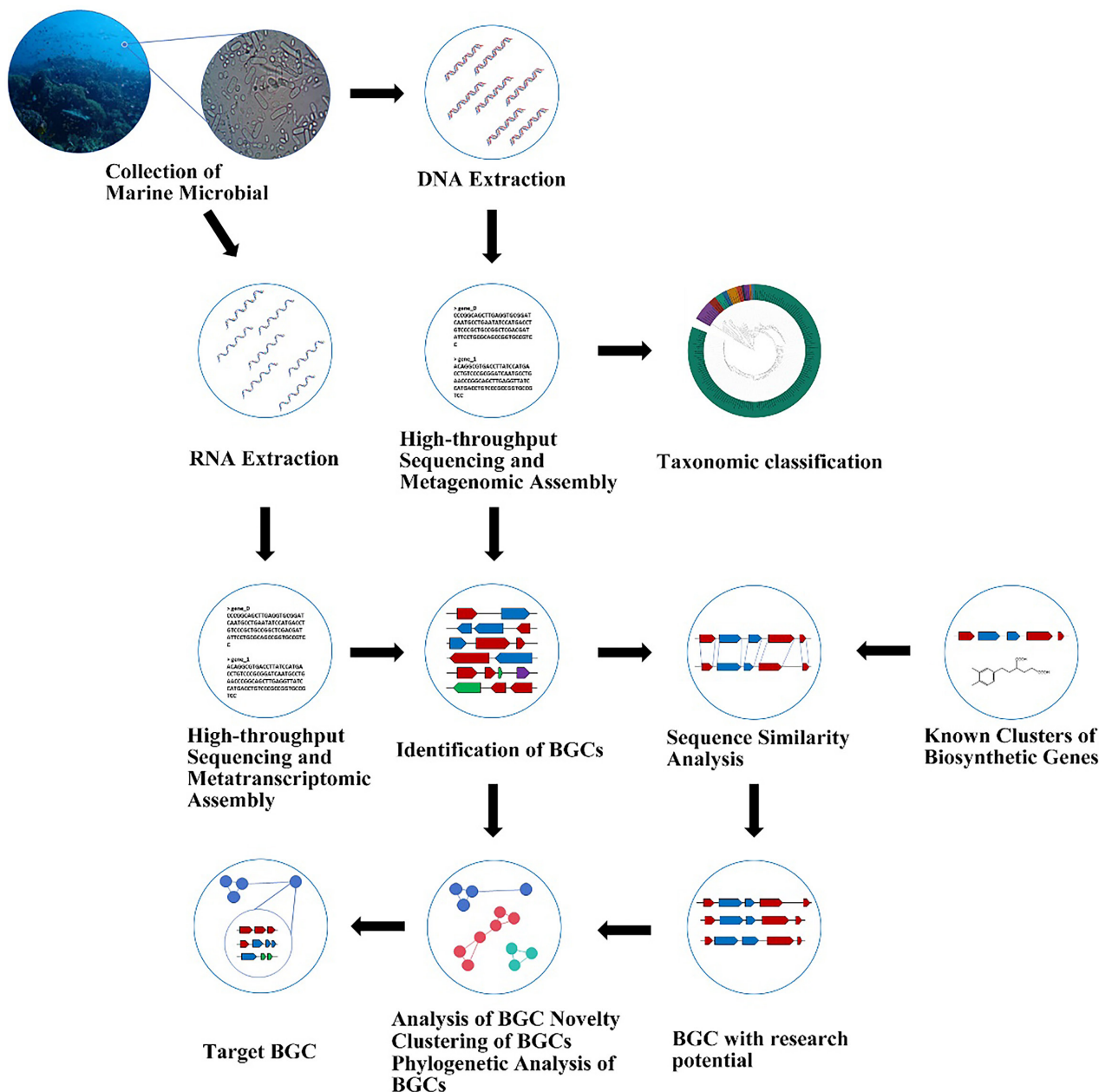
In the deep-sea sediments of the Gulf of Mexico, Fernández-López and their team investigated the metagenome for type I polyketide synthase (PKS) genes. They identified more than 2000 candidate genes encoding type I PKS domains, as well as potential biosynthetic pathways for various other secondary metabolites (Fernández-López et al., 2022).

Cuadrat and colleagues performed a metagenomic analysis on seawater influenced by upwelling off the Brazilian coast, resulting in the identification of 84 KS and 46 cyclization (C) domain sequences. This particular environment demonstrated significant potential for discovering novel secondary metabolites of biotechnological value (Cuadrat et al., 2015).

These metagenomic studies collectively highlight the vast potential for development across the entire ocean. The outcomes lay a solid groundwork for ongoing research into secondary metabolites derived from marine microorganisms. A substantial number of potential BGCs have been catalogued in databases, offering invaluable references for future studies and facilitating mutual validation.

## The application of metagenomics Technology in the Exploration of secondary metabolites within marine symbiotic systems

Symbiotic systems have emerged as a key area in the quest for novel secondary metabolites. Numerous studies highlight that these metabolites are prevalent



**FIGURE 1** Workflow for identifying and analysing marine microbial BGCs collection: Marine microbial samples are collected. DNA/RNA Extraction: DNA and RNA are extracted from samples. Sequencing and Assembly: High-throughput sequencing and assembly of metagenomic and metatranscriptomic data. Taxonomic Classification: Classification of species present in the samples. BGC Identification: Detection of BGCs from the assembled data. Sequence similarity analysis: Comparison of BGCs with known biosynthetic gene clusters. Known clusters: Identification of similarities with known BGCs. Research potential: identification of BGCs with novel research potential. Novelty and phylogenetic analysis: analysis of BGC novelty, clustering, and phylogenetic relationships.

within symbiotic systems, playing essential roles such as defence, communication, and regulation (Engel et al., 2015; Kaasalainen et al., 2012; Sagar et al., 2010). These secondary metabolites are not only vital for the survival of the organisms involved but also serve as a valuable source of biological resources for human exploration. However, culturing microorganisms from symbiotic systems in pure form poses significant challenges. Often, these organisms do not express crucial products in isolation (Bertrand et al., 2014; Woodhouse

et al., 2013), and co-cultivation presents its own difficulties (Lewis et al., 2021). Utilizing metagenomics technology, researchers can circumvent these cultivation challenges, directly investigating symbiotic systems with greater efficiency (Arumugam et al., 2011; Grube et al., 2015; Ni & Tokuda, 2013). Presently, extensive research focuses on using metagenomic methods to explore oceanic symbiotic systems, with the goal of deepening the mining and understanding of secondary metabolites.

## Marine sponge symbiotic systems

Marine sponges and their microbial symbiotic communities are recognized as prolific sources of diverse natural products (Taylor et al., 2007). In certain sponge species, microbes can comprise 40%–60% of the sponge's tissue volume (Fathima et al., 2022; Yang, Franco, & Zhang, 2019; Yang, Qian, et al., 2019). Extensive research has shown that these microbes, especially sponge-associated actinomycetes, are the primary producers of structurally varied and pharmacologically promising secondary metabolites (Jensen, 2016; Thomas et al., 2010; Yang, Franco, & Zhang, 2019; Yang, Qian, et al., 2019).

Current research on sponge symbiotic system metagenomes predominantly falls into two categories. One approach involves identifying potential BGCs from sponge metagenomes using bioinformatics software like antiSMASH. Subsequent studies focus on these identified BGCs, exploring their biosynthetic capabilities and the secondary metabolites they produce. For instance, Podell and colleagues reconstructed a nearly complete metagenome-assembled genome of the *Lamellodysidea herbacea*, revealing the secondary metabolite potential of 16 uncharacterized and uncultivated microbial groups associated with this sponge (Podell et al., 2020). Further metagenomic analyses, employing antiSMASH and MIBiG database comparisons, identified unique BGCs in various sponge symbiotic systems. Notably (Dat et al., 2023; Storey et al., 2020), *Acidobacterorita* and *Latescibacteria* symbiotic systems exhibited high biosynthetic potential (Loureiro et al., 2022). These unique BGCs likely contribute to novel synthetic pathways for unknown secondary metabolites. In one example, Nguyen and colleagues discovered a ribosomally synthesized and post-translationally modified peptide (RiPP) BGC in a marine sponge microbial community. This widely conserved BGC encodes a tryptophan brominase, notable for its independence from prior RiPP core peptide modifications. Shared biosynthetic motifs across different RiPP classes suggest the potential for diverse chemical space in RiPPs, underscoring the potential for discovering new secondary metabolites in sponges (Nguyen et al., 2021). El Samak et al.'s analysis of *Theonella* sp. metagenomics and metatranscriptomics identified numerous actively transcribed BGCs with antimicrobial, cytotoxic, and inhibitory properties. Complete clusters for the biosynthesis of theonellamide (Figure 2E) and misakinolide (Figure 2F) were found, along with clusters closely related to those producing bioactive natural products. The majority of these BGCs were novel, indicating *Theonella* as a key source for bioactive natural product discovery (El Samak et al., 2023).

Beyond the studies mentioned above, metagenomic methods can also be utilized for tracing back to the BGCs and actual producers of known secondary metabolites

produced by sponges. Calyculin A (1, Figure 2A), a cytotoxic compound discovered in the Japanese marine sponge *Discodermia calyx* (Ishihara et al., 1989), was studied by researchers who utilized specific primers for PCR amplification of the sponge's metagenomic DNA. This process identified a gene cluster containing transacyltransferase (AT) Type I polyketide synthase (PKS) as a potential BGC for calyculin A. Further analysis confirmed that this gene cluster originated from the symbiotic bacteria *Candidatus Entotheonella* (Wakimoto et al., 2014). This study not only guides the production of calyculin A but also demonstrates the efficacy of metagenomic mining as a powerful tool for isolating biosynthetic genes encoding specific trans-AT PKS from marine sponge metagenomes. In another study, Agarwal and colleagues utilized metagenomic techniques to identify the *hs\_bmp* BGC, responsible for polybrominated diphenyl ether production in sponges. Following this discovery, they successfully induced expression through vector construction and isolated various hydroxylated polybrominated diphenyl ethers, such as 2,4-dibromo-6-(2,4-dibromophenoxy)phenol (2), 3,5-dibromo-2-(3,5-dibromo-2-hydroxyphenoxy)phenol (3), and 6,6'-oxybis(2,4-dibromophenol) (4, Figure 2B). This research establishes a crucial foundation for further investigation into these compounds. The genetic tools developed for exploring the natural producers of polybrominated diphenyl ethers will facilitate monitoring of the populations responsible for their synthesis. Such advancements hold significant implications for policy formulation and human health (Agarwal et al., 2017). Piel and collaborators made significant discoveries in the metagenome of *Theonella swinhoei*, not only identifying AT PKS genes but also uncovering biosynthetic gene fragments related to onnamides (5, Figure 2C) and theopederins (6, Figure 2D). The genes they isolated encompass nearly the entire region necessary for synthesizing onnamide molecules, which are integral to producing anti-tumour marine compounds. These gene clusters have a substantial impact on the generation of both known and novel pederin-type anticancer agents, and their discovery carries important implications for marine biotechnology and ecology (Piel, Hui, Fusetani, & Matsunaga, 2004; Piel, Hui, Wen, et al., 2004).

These specific case studies have significantly broadened our understanding of the diversity of secondary metabolites produced by sponge symbiotic systems. Additionally, they offer essential insights that pave the way for future in-depth explorations of novel secondary metabolites derived from marine sponges.

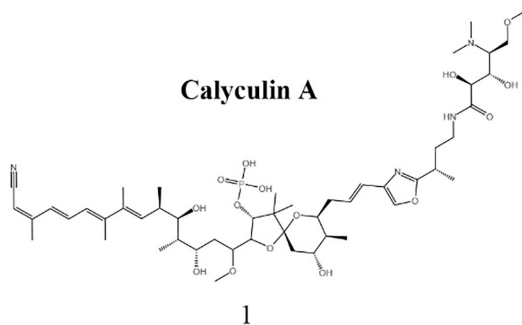
## Symbiotic systems of tunicates

Symbiotic systems dominated by tunicates, especially ascidians, have become a focal point in

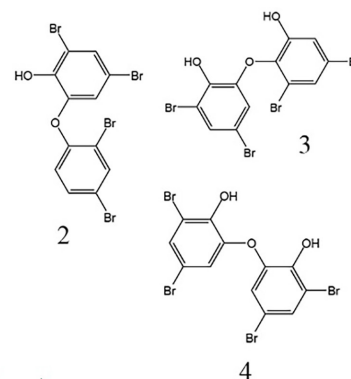
(A)



Calyculin A



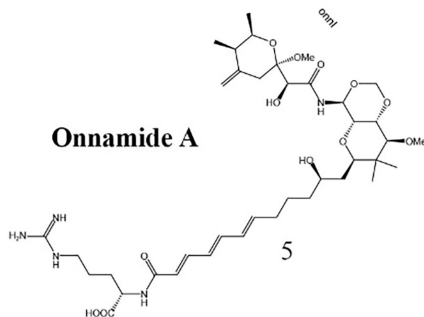
(B)



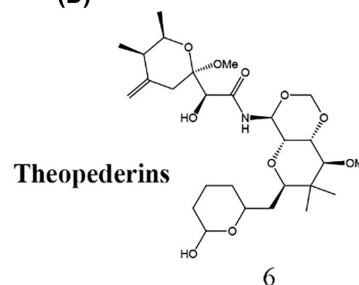
(C)



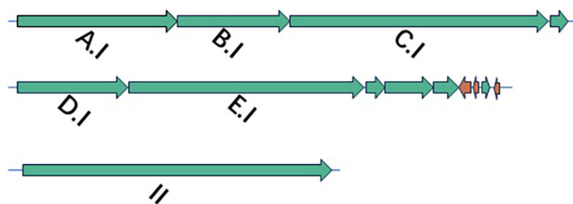
Onnamide A



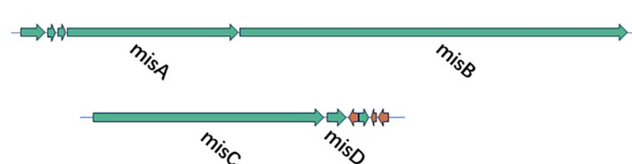
(D)



(E)



(F)



**FIGURE 2** Isolated BGCs and secondary metabolites from marine sponge symbiotic systems. (A) Calyculin A and its BGC detected in the Japanese marine sponge *Discodermia calyx* by metagenome mining and PCR amplification; (B) Various hydroxylated polybrominated diphenyl ethers detected in *Dysideidae* sponges by metagenome mining and PCR amplification; (C) Biosynthesis genes with partial sequences and Onnamide A detected in *Theonella swinhoei* by metagenome mining and PCR amplification; (D) Theopederins A detected in *Theonella swinhoei* by metagenome mining and PCR amplification; (E) Theonellamide BGC detected in Egyptian Red Sea sponge *Theonella* sp. by AntiSMASH; (F) Misakinolide BGC detected in Egyptian Red Sea sponge *Theonella* sp. by AntiSMASH.

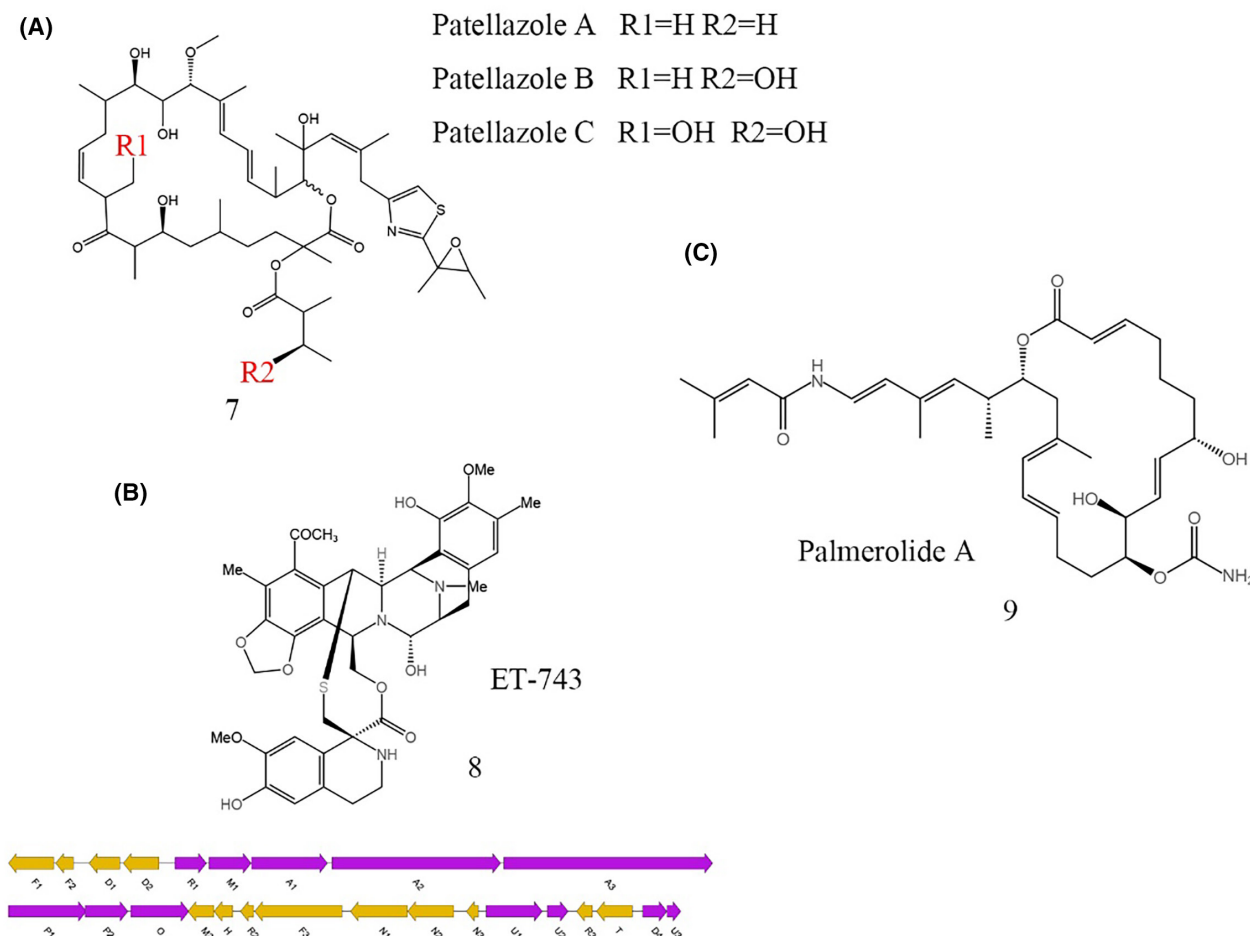
the exploration of secondary metabolites. Much like in sponge symbiotic systems, it is primarily the symbiotic bacteria within tunicates that

produce the majority of these secondary metabolites (Chen et al., 2017; Donia et al., 2011; Shenkar & Swalla, 2011).

The research on metagenomes of marine sponges and tunicates has shown remarkable similarities. Tianero and colleagues conducted a study on the metagenomes of 32 different tunicates, revealing species-specific microbiomes and secondary metabolites, highlighting the vast potential for secondary metabolite mining in tunicate symbiotic systems (Tianero et al., 2015). Similarly, Mohamed S. Donia and his team's research on *Lissoclinum patella*'s metagenome led to the isolation of the genome of its symbiotic microbe, *Prochloron didemni*, and examined its biosynthetic capabilities (Donia et al., 2011). This study also underscored the considerable potential for discovering secondary metabolites in tunicate symbiotic systems.

In their pursuit of exploring secondary metabolites in tunicates, Kwan and colleagues performed direct sequencing and assembly of the complete genome of *Candidatus Endolissoclinum faulkneri* derived from tunicate metagenomic DNA. Their research led to the identification of the biosynthetic model for patellazoles-like (7, Figure 3A) compounds. Furthermore, the

streamlined genomes of symbiotic bacteria they discovered indicated that secondary metabolism plays a crucial role in symbiotic interactions (Kwan et al., 2012). In another study, Rath and colleagues conducted metagenomic sequencing on the total DNA extracted from a tunicate symbiotic community. They focused on and successfully assembled a continuous 35kb fragment, which encompassed 25 genes central to the non-ribosomal peptide synthetase (NRPS) biosynthetic pathway of the ET-743 anticancer drug. This research confirmed that ET-743 (8, Figure 3B) is produced by the symbiotic bacterium *Candidatus Endoecteinascidia frumentensis*. Additionally, the study validated the expression of crucial proteins and enzymes involved in this pathway (Rath et al., 2011). In the study conducted by Nicole E. Avalon and colleagues, multiple functional trans-acyltransferase (AT) domains within the polyketide synthase-non-ribosomal peptide synthetase (PKS-NRPS) structures were identified in the Antarctic ascidian *Synoicum adareanum*. These findings pertain to the inferred BGC for palmerolide A. The research focused on the bioinformatic and mechanistic analysis



**FIGURE 3** Isolated BGC and secondary metabolites from tunicate symbiotic systems. (A) Patellazoles detected in *Lissoclinum patella* by metagenome mining and PCR amplification; (B) The anticancer agent ET-743 and its putative BGC isolated from *Ecteinascidia turbinata* by metagenome mining, codon usage bias analysis and metaproteomic analyses; (C) Palmerolide A isolated from *Synoicum adareanum* by metagenome mining and PCR amplification.

of the palmerolide PKS-NRPS biosynthetic pathway, highlighting the intricate relationship between the ascidian's microbiome and its secondary metabolite production (Avalon et al., 2021). Riesenfeld and colleagues successfully identified PKS genes within the metagenome of the tunicate *Synoicum adareanum*'s symbiotic system. Utilizing BLAST analysis for PKS genes, they ascertained that the encoded compound, palmerolide A (9, Figure 3C), recognized for its anti-melanoma properties, is likely synthesized by symbiotic microorganisms belonging to the  $\gamma$ -Proteobacteria class (Riesenfeld et al., 2008). Taken together, these studies underscore the crucial role of metagenomics in uncovering secondary metabolites in tunicates and elucidating their biosynthetic origins.

## Other symbiotic systems

Beyond the symbiotic systems of sponges and tunicates, there is notable potential for discovering new secondary metabolites in other symbiotic environments. Lu et al. delved into the metagenomics of microbial communities associated with large algae, predicting a significant number of polysaccharide utilization loci (4451) and BGCs (8810). Their findings revealed that compared to bacteria found in seawater and sediments, those in algal phycospheres possess larger genomes and a relatively higher number of BGCs (Lu et al., 2023). This study underscores the untapped potential of bacteria in algal phycospheres as a rich source for secondary metabolite synthesis.

The aforementioned studies highlight the crucial role of metagenomics in the investigation of secondary metabolites within marine symbiotic systems. By directly obtaining genomic information from environmental samples, metagenomic techniques allow for an in-depth analysis of complex microbial symbiotic communities. This technology not only reveals the genetic diversity and functional potential of symbiotic microorganisms but also aids in the identification and assembly of BGCs, providing insights into the pathways of secondary metabolite synthesis. Its high-throughput nature accelerates the discovery of new secondary metabolites, facilitating the efficient screening of potential bioactive molecules on a large scale.

## Challenges and prospects

The exploration of secondary metabolites in marine microorganisms represents a vital research area. The continual discovery and application of various active compounds from these organisms present significant challenges, particularly due to the difficulty of cultivating many of them in pure cultures. Additionally, the

high cost and complexity of spectroscopic and mass spectrometric methods for identifying secondary metabolites limit extensive research on new compounds. Therefore, in addressing these practical challenges, especially in the aspect of validation, the use of computational biology tools to mine BGCs of secondary metabolites from metagenomes has become a widely adopted strategy.

The widespread adoption of short-read sequencing technologies, known for their high accuracy and low cost, has significantly influenced metagenomic research. However, these methods have limitations in genome assembly, particularly in preserving crucial information within the flexible genome where most BGCs are found. This loss of information can significantly impact research on secondary metabolites. In response, there has been a shift towards long-read sequencing technologies, offering greater throughput and accuracy. For instance, Waschulin and colleagues, for example, recovered over 1400 full-length BGC sequences from Antarctic soil using third-generation sequencing (Waschulin et al., 2022). Mantri and team combined long and short reads in a hybrid assembly to study BGCs in the soil of Schönbuch Forest, southern Germany. This approach uncovered many unique BGCs not detectable through short-read assembly alone (Mantri et al., 2021). Similarly, Sánchez-Navarro and colleagues observed most contiguous BGCs in activated sludge using long-read metagenomic data, contrasting with the incomplete results from short-read data (Sánchez-Navarro et al., 2022). Huang and team identified 339 BGCs using a smaller set of long-read sequences, successfully classifying over 89% of them (Huang et al., 2023). These examples underscore the substantial benefits of third-generation sequencing technologies in preserving comprehensive BGC information.

With the rapid advancement of computer performance and AI, an increasing number of computational biology tools have been developed. In 2021, machine learning methods for BGC prediction and classification underwent significant advancements. GECCO, using Conditional Random Fields (CRF) for BGC prediction, established a more sophisticated benchmark with MIBiG 2.0 (Carroll et al., 2021). Yang et al. enhanced DeepBGC by developing DeepBGCpred, which identified more BGCs by expanding feature dimensions, implementing a sliding window approach, and incorporating negative class categories in Random Forest (RF) classification (Yang et al., 2021). In a similar vein, Liu et al. in 2022 upgraded DeepBGC to e-DeepBGC, employing comparable strategies to DeepBGCpred, conducting ablation experiments (Liu et al., 2022). In same year, Almeida et al. conducted evaluations of candidate BGCs in two fungal genomes. They employed Q-learning to refine BGC prediction results,



optimizing the prediction process specifically for fungal BGCs (Almeida et al., 2022). In 2023, Rios-Martinez and team introduced a novel self-supervised masked language model, primarily utilizing CARP's BiGCARP framework. This model, initially pre-trained on Pfam-level using mask LM CARP, was fine-tuned with a sliding window approach for prediction. Notably, this method achieved, for the first time, the simultaneous detection and classification of BGCs at the Pfam-level, marking a significant advancement in the field (Rios-Martinez et al., 2023). The continuous enhancement of algorithms in computational biology tools has led to the identification of more accurate BGCs, aiding in the discovery of novel secondary metabolites. These tools are accessible via the provided web addresses in the accompanying table (Table 1).

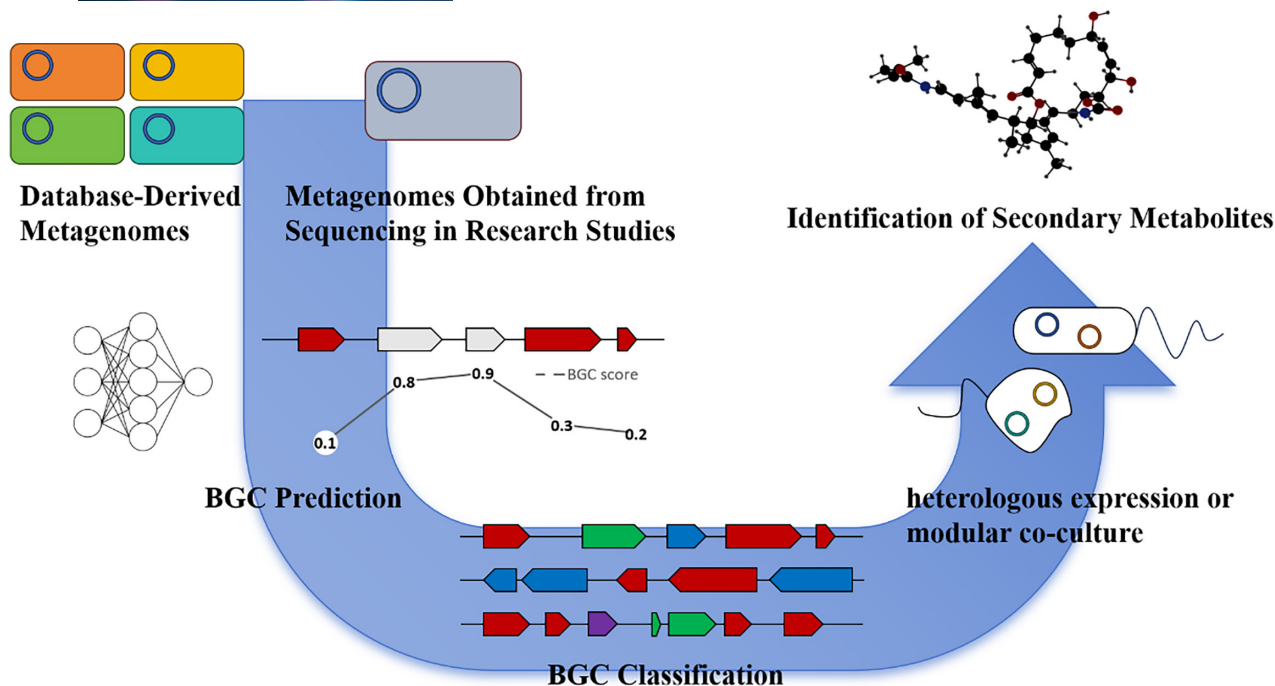
The use of computational biology tools has significantly streamlined the identification of potential secondary metabolites in metagenomic data, thereby reducing the complexity of research. However, as deep learning-based models begin to emerge, the validation of these models heavily relies on comparisons with predictions from traditional models like antiSMASH. Moreover, the claims of discovering new BGCs often lack experimental validation, making it difficult to determine whether these are false positives or genuinely new BGCs. This is a major shortcoming of most AI-based BGC studies, and we recommend that these studies be more closely linked to practical experiments. Additionally, existing BGC databases have significant issues and are not well-suited for training more complex models, such as transformers (Vaswani et al., 2017). On one hand, there are too few fully annotated BGC datasets. Current methods, such as DeepBGC, insert known BGCs into unannotated sequences to construct synthetic data, but this synthetic data does not accurately reflect real BGC environments. Furthermore, even these fully annotated BGCs are imbalanced in terms of categories,

which is detrimental to data-driven machine learning. Therefore, in the future, we hope for more comprehensive BGC databases and the development of models that integrate multiple features (as e-DeepBGC have done) to enhance BGC prediction. Simply transplanting natural language processing techniques does not lead to better performance and interpretability in BGC prediction (Vaswani et al., 2017).

BGCs identified in metagenomes through computational biology techniques often correspond to secondary metabolic pathways elusive to pure culture methods. Concurrently, the limitations in heterologous expression technologies pose challenges for accommodating complex BGCs in heterologous hosts, making the practical application and validation of these identified clusters a significant obstacle. The complexity of BGCs, along with transcriptional differences and codon usage variations between *E. coli* and original hosts, often hinders the discovery of new secondary metabolites through metagenomics. Identifying suitable heterologous expression hosts is challenging due to BGC complexity. Even with continuous improvements in *E. coli* expression vectors, new heterologous hosts, and genome engineering, effectively expressing complex BGCs remains difficult. Modular co-culture engineering, involving complete modularization of biosynthetic pathways with each module regulated in different hosts, emerges as a promising solution. This method, successfully applied in synthesizing natural products from plants and fungi, holds potential for heterologous production of marine secondary metabolites (Trindade et al., 2015; Zhou et al., 2021). Overcoming these challenges could simplify the process from identifying BGCs to elucidating new secondary metabolites (Figure 4). Given the immense potential in discovering secondary metabolites from marine microorganisms and the efficiency of metagenomics in identifying promising BGCs, there is significant growth potential for novel drugs derived from marine microorganisms.

**TABLE 1** Representative artificial intelligence approaches for the exploration of secondary metabolites.

| Software      | Main methods | Website   | References                  |
|---------------|--------------|---|-----------------------------|
| antiSMASH     | pHMM         | <a href="https://antismash.secondarymetabolites.org/">https://antismash.secondarymetabolites.org/</a>                   | Blin et al. (2023)          |
| ClusterFinder | HMM          | <a href="https://github.com/petercim/ClusterFinder">https://github.com/petercim/ClusterFinder</a>                       | Cimermanic et al. (2014)    |
| DeepBGC       | BiLSTM RF    | <a href="https://github.com/Merck/deepbgc">https://github.com/Merck/deepbgc</a>   | Hannigan et al. (2019)      |
| GECCO         | CRF          | <a href="https://gecco.embl.de/">https://gecco.embl.de/</a>   | Carroll et al. (2021)       |
| Deep-BGCpred  | BiLSTM RF    | <a href="https://github.com/pmabio/Deep-BGCpred">https://github.com/pmabio/Deep-BGCpred</a>                             | Yang et al., (2021)         |
| e-DeepBGC     | BiLSTM RF    |   | Liu et al. (2022)           |
| RLBGC         | Q-learning   | <a href="https://github.com/bioinfoUQAM/RL-bgc-components">https://github.com/bioinfoUQAM/RL-bgc-components</a>         | Almeida et al. (2022)       |
| BiGCARP       | mask LM CARP | <a href="https://github.com/microsoft/protein-sequence-models">https://github.com/microsoft/protein-sequence-models</a> | Rios-Martinez et al. (2023) |



**FIGURE 4** Workflow for predicting and identifying secondary metabolites using metagenomics. Database-derived metagenomes: Utilization of metagenomic data from existing databases. Metagenomes obtained from sequencing in research studies: Acquisition of metagenomic data from sequencing efforts in various research studies. BGC prediction: prediction of potential BGCs using BGC scoring. BGC classification: classification of predicted BGCs based on their types. Identification of secondary metabolites: detection of secondary metabolites produced by the identified BGCs. Heterologous expression or modular co-culture: validation of secondary metabolite functions through heterologous expression or modular co-culture techniques.

## AUTHOR CONTRIBUTIONS

**Shaoyu Wang:** Writing – original draft; writing – review and editing. **Xinyan Li:** Writing – review and editing. **Wei qin Yang:** Writing – review and editing. **Ranran Huang:** Writing – review and editing.

## ACKNOWLEDGEMENTS

This work was supported by Technology of the People's Republic of China; National Natural Science Foundation of China (32170065, and 31970113); Ocean Negative Carbon Emissions (ONCE) Program and Key Research and Development Program of Shandong Province (2020ZLYS04).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

## ORCID

Shaoyu Wang  <https://orcid.org/0009-0006-1477-5343>

## REFERENCES

- Agarwal, V., Blanton, J.M., Podell, S., Taton, A., Schorn, M.A., Busch, J. et al. (2017) Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nature Chemical Biology*, 13, 537–543.
- Alma'abadi, A.D., Gojobori, T. & Mineta, K. (2015) Marine metagenome as a resource for novel enzymes. *Genomics, Proteomics & Bioinformatics*, 13, 290–295.
- Almeida, H., Tsang, A. & Diallo, A.B. (2022) Improving candidate biosynthetic gene clusters in fungi through reinforcement learning. *Bioinformatics*, 38, 3984–3991.
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R. et al. (2011) Enterotypes of the human gut microbiome. *Nature*, 473, 174–180.
- Avalon, N.E., Murray, A.E., Daligault, H.E., Lo, C.C., Davenport, K.W., Dichosa, A.E.K. et al. (2021) Bioinformatic and mechanistic analysis of the Palmerolide PKS-NRPS biosynthetic pathway from the microbiome of an Antarctic ascidian. *Frontiers in Chemistry*, 9, 802574.
- Bar-On, Y.M., Phillips, R. & Milo, R. (2018) The biomass distribution on earth. *Proceedings of the National Academy of Sciences*, 115, 6506–6511.
- Bertrand, S., Bohni, N., Schnee, S., Schumpp, O., Gindro, K. & Wolfender, J.-L. (2014) Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnology Advances*, 32, 1180–1204.
- Bhatnagar, I. & Kim, S.-K. (2010) Immense essence of excellence: marine microbial bioactive compounds. *Marine Drugs*, 8, 2673–2701.
- Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M. et al. (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51, gkad344.
- Breitwieser, F.P. & Salzberg, S.L. (2019) Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*, 36, 1303–1304.
- Carroll, L.M., Larralde, M., Fleck, J.S., Ponnudurai, R., Milanese, A., Cappio, E. et al. (2021) Accurate de novo identification of biosynthetic gene clusters with GECCO. *Biorxiv*, 2021, 2003.442509.
- Chen, L., Fu, C. & Wang, G. (2017) Microbial diversity associated with ascidians: a review of research methods and application. *Symbiosis*, 71, 19–26.

- Chiu, C.Y. & Miller, S.A. (2019) Clinical metagenomics. *Nature Reviews Genetics*, 20, 341–355.
- Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Brown, L.C.W., Mavrommatis, K. et al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 158, 412–421.
- Conway, K.R. & Boddy, C.N. (2012) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Research*, 41, D402–D407.
- Coutinho, F.H., Gregoracci, G.B., Walter, J.M., Thompson, C.C. & Thompson, F.L. (2018) Metagenomics sheds light on the ecology of marine microbes and their viruses. *Trends in Microbiology*, 26, 955–965.
- Cuadrat, R.R., Cury, J.C. & Dávila, A.M. (2015) Metagenomic analysis of upwelling-affected Brazilian coastal seawater reveals sequence domains of type I PKS and modular NRPS. *International Journal of Molecular Sciences*, 16, 28285–28295.
- Dat, T.T.H., Steinert, G., Cuc, N.T.K., Cuong, P.V., Smidt, H. & Sipkema, D. (2023) Diversity of bacterial secondary metabolite biosynthetic gene clusters in three Vietnamese sponges. *Marine Drugs*, 21, 29.
- David, B., Wolfender, J.-L. & Dias, D.A. (2015) The pharmaceutical industry and natural products: historical status and new trends. *Phytochemistry Reviews*, 14, 299–315.
- Donia, M.S., Fricke, W.F., Partensky, F., Cox, J., Elshahawi, S.I., White, J.R. et al. (2011) Complex microbiome underlying secondary and primary metabolism in the tunicate-Prochloron symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 108, E1423–E1432.
- Doroghazi, J.R., Albright, J.C., Goering, A.W., Ju, K.-S., Haines, R.R., Tchaluikov, K.A. et al. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology*, 10, 963–968.
- Duncan, K.R., Crüsemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N. et al. (2015) Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chemistry & Biology*, 22, 460–471.
- El Samak, M., Zakeer, S., Hanora, A. & Solyman, S.M. (2023) Metagenomic and metatranscriptomic exploration of the Egyptian Red Sea sponge *Theonella* sp. associated microbial community. *Marine Genomics*, 70, 101032.
- Engel, P., Vizcaino, M.I. & Crawford, J.M. (2015) Gut symbionts from distinct hosts exhibit genotoxic activity via divergent Colibactin biosynthesis pathways. *Applied and Environmental Microbiology*, 81, 1502–1512.
- Fathima, A., Arafath, Y., Hassan, S., Bramhachari, P.V., Kiran, G.S. & Selvin, J. (2022) Novel insights into environmental niche adaptations and secondary metabolite biosynthesis potential of marine sponge microbiome. In: Veera Bramhachari, P. (Ed.) *Understanding the microbiome interactions in agriculture and the environment*. Singapore: Springer Nature Singapore, pp. 113–137.
- Fernández-López, M., Sánchez-Reyes, A., Barcelos, C., Sidón-Ceseña, K., Leite, R.B. & Lago-Lestón, A. (2022) Deep-Sea sediments from the southern Gulf of Mexico Harbor a wide diversity of PKS I genes. *Antibiotics*, 11, 887.
- Grube, M., Cernava, T., Soh, J., Fuchs, S., Aschenbrenner, I., Lassek, C. et al. (2015) Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics. *The ISME Journal*, 9, 412–424.
- Gu, Z., Wang, Y., Liu, Y., Yang, X. & Chen, F. (2020) Ancient environmental DNA and environmental archaeology. *Quaternary Sciences*, 40, 295–306.
- Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5, R245–R249.
- Hannigan, G.D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L. et al. (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Research*, 47, e110.
- Hay, M.E. (1996) Marine chemical ecology: what's known and what's next? *Journal of Experimental Marine Biology and Ecology*, 200, 103–134.
- Hoshino, S., Onaka, H. & Abe, I. (2019) Activation of silent biosynthetic pathways and discovery of novel secondary metabolites in actinomycetes by co-culture with mycolic acid-containing bacteria. *Journal of Industrial Microbiology and Biotechnology*, 46, 363–374.
- Huang, R., Wang, Y., Liu, D., Wang, S., Lv, H. & Yan, Z. (2023) Long-read metagenomics of marine microbes reveals diversely expressed secondary metabolites. *Microbiology Spectrum*, 11, e0150123.
- Ishihara, H., Martin, B.L., Brautigan, D.L., Karaki, H., Ozaki, H., Kato, Y. et al. (1989) Calyculin A and okadaic acid: inhibitors of protein phosphatase activity. *Biochemical and Biophysical Research Communications*, 159, 871–877.
- Jensen, P.R. (2016) Natural products and the gene cluster revolution. *Trends in Microbiology*, 24, 968–977.
- Jensen, P.R., Moore, B.S. & Fenical, W. (2015) The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. *Natural Product Reports*, 32, 738–751.
- Kaasalainen, U., Fewer, D.P., Jokela, J., Wahlsten, M., Sivonen, K. & Rikkinen, J. (2012) Cyanobacteria produce a high variety of hepatotoxic peptides in lichen symbiosis. *Proceedings of the National Academy of Sciences*, 109, 5886–5891.
- Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., Van Der Hoof, J.J. et al. (2020) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*, 48, D454–D458.
- Kautsar, S.A., Blin, K., Shaw, S., Weber, T. & Medema, M.H. (2021) BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Research*, 49, D490–D497.
- Kautsar, S.A., van der Hoof, J.J.J., de Ridder, D. & Medema, M.H. (2021) BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *GigaScience*, 10, g1aa154.
- Key, F.M., Posth, C., Krause, J., Herbig, A. & Bos, K.I. (2017) Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends in Genetics*, 33, 508–520.
- Khalifa, S.A.M., Elias, N., Farag, M.A., Chen, L., Saeed, A., Hegazy, M.-E.F. et al. (2019) Marine natural products: a source of novel anticancer drugs. *Marine Drugs*, 17, 491.
- Knowles, S.L., Raja, H.A., Roberts, C.D. & Oberlies, N.H. (2022) Fungal–fungal co-culture: a primer for generating chemical diversity. *Natural Product Reports*, 39, 1557–1573.
- Ko, K.K.K., Chng, K.R. & Nagarajan, N. (2022) Metagenomics-enabled microbial surveillance. *Nature Microbiology*, 7, 486–496.
- Kodzios, R. & Gojobori, T. (2015) Marine metagenomics as a source for bioprospecting. *Marine Genomics*, 24, 21–30.
- Kwan, J.C., Donia, M.S., Han, A.W., Hirose, E., Haygood, M.G. & Schmidt, E.W. (2012) Genome streamlining and chemical defense in a coral reef symbiosis. *Proceedings of the National Academy of Sciences*, 109, 20655–20660.
- Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z. & Ettema, T.J.G. (2021) Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, 19, 225–240.
- Liu, M., Li, Y. & Li, H. (2022) Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *Journal of Molecular Biology*, 434, 167597.
- Loureiro, C., Galani, A., Gavriilidou, A., Chaib de Mares, M., van der Oost, J., Medema, M.H. et al. (2022) Comparative metagenomic

- analysis of biosynthetic diversity across sponge microbiomes highlights metabolic novelty, conservation, and diversification. *mSystems*, 7, e00357-22.
- Lu, D.-C., Wang, F.-Q., Amann, R.I., Teeling, H. & Du, Z.-J. (2023) Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome*, 11, 126.
- Mantri, S.S., Negri, T., Sales-Ortells, H., Angelov, A., Peter, S., Neidhardt, H. et al. (2021) Metagenomic sequencing of multiple soil horizons and sites in close vicinity revealed novel secondary metabolite diversity. *mSystems*, 6, e0101821.
- Marmann, A., Aly, A.H., Lin, W., Wang, B. & Proksch, P. (2014) Co-cultivation—a powerful emerging tool for enhancing the chemical diversity of microorganisms. *Marine Drugs*, 12, 1043–1065.
- Martins, A., Vieira, H., Gaspar, H. & Santos, S. (2014) Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Marine Drugs*, 12, 1066–1101.
- Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K. et al. (2015) Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology*, 11, 625–631.
- Mehbub, M.F., Lei, J., Franco, C. & Zhang, W. (2014) Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. *Marine Drugs*, 12, 4539–4577.
- Monti, M. (2023) *Marine microbial communities in health and disease: An exploration of marine bacterial associations across space, time, and potential roles against pathogens in the context of the current global climate change scenario*. United States—Alabama: The University of Alabama, p. 177.
- Navarro-Muñoz, J.C., Selem-Mojica, N., Mallowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I. et al. (2020) A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*, 16, 60–68.
- Nguyen, N.A., Lin, Z., Mohanty, I., Garg, N., Schmidt, E.W. & Agarwal, V. (2021) An obligate peptidyl brominase underlies the discovery of highly distributed biosynthetic gene clusters in marine sponge microbiomes. *Journal of the American Chemical Society*, 143, 10221–10231.
- Ni, J. & Tokuda, G. (2013) Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnology Advances*, 31, 838–850.
- Nielsen, J.C., Grijseels, S., Prigent, S., Ji, B., Dainat, J., Nielsen, K.F. et al. (2017) Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nature Microbiology*, 2, 17044.
- Nigam, M., Suleria, H.A.R., Farzaei, M.H. & Mishra, A.P. (2019) Marine anticancer drugs and their relevant targets: a treasure from the ocean. *DARU Journal of Pharmaceutical Sciences*, 27, 491–515.
- Paoli, L., Ruscheweyh, H.-J., Forneris, C.C., Hubrich, F., Kautsar, S., Bhushan, A. et al. (2022) Biosynthetic potential of the global ocean microbiome. *Nature*, 607, 111–118.
- Piel, J., Hui, D., Fusetani, N. & Matsunaga, S. (2004) Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environmental Microbiology*, 6, 921–927.
- Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N. et al. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proceedings of the National Academy of Sciences*, 101, 16222–16227.
- Podell, S., Blanton, J.M., Oliver, A., Schorn, M.A., Agarwal, V., Biggs, J.S. et al. (2020) A genomic view of trophic and metabolic diversity in clade-specific *Lamellodysidea* sponge microbiomes. *Microbiome*, 8, 1–17.
- Rath, C.M., Janto, B., Earl, J., Ahmed, A., Hu, F.Z., Hiller, L. et al. (2011) Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743. *ACS Chemical Biology*, 6, 1244–1256.
- Rego, A., Fernandez-Guerra, A., Duarte, P., Assmy, P., Leão, P.N. & Magalhães, C. (2021) Secondary metabolite biosynthetic diversity in Arctic Ocean metagenomes. *Microbial Genomics*, 7, e000731.
- Reuver, M., Maher, J. & Wilson, A.M. (2022) Ocean restoration and the strategic plan of the marine microbiome. In: Stal, L.J. & Cretoiu, M.S. (Eds.) *The marine microbiome*. Cham: Springer International Publishing, pp. 731–766.
- Riesenfeld, C.S., Murray, A.E. & Baker, B.J. (2008) Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate *Synoicum adareanum*. *Journal of Natural Products*, 71, 1812–1818.
- Rios-Martinez, C., Bhattacharya, N., Amini, A.P., Crawford, L. & Yang, K.K. (2023) Deep self-supervised learning for biosynthetic gene cluster detection and product classification. *PLoS Computational Biology*, 19, e1011162.
- Romano, S., Jackson, S.A., Patry, S. & Dobson, A.D.W. (2018) Extending the “one strain many compounds” (OSMAC) principle to marine microorganisms. *Marine Drugs*, 16, 244.
- Sagar, S., Kaur, M. & Minneman, K.P. (2010) Antiviral lead compounds from marine sponges. *Marine Drugs*, 8, 2619–2638.
- Sánchez-Navarro, R., Nuhamunada, M., Mohite, O.S., Wasmund, K., Albertsen, M., Gram, L. et al. (2022) Long-read metagenome-assembled genomes improve identification of novel complete biosynthetic gene clusters in a complex microbial activated sludge ecosystem. *mSystems*, 7, e0063222.
- Sekurova, O.N., Schneider, O. & Zotchev, S.B. (2019) Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microbial Biotechnology*, 12, 828–844.
- Shenkar, N. & Swalla, B.J. (2011) Global diversity of Ascidiacea. *PLoS One*, 6, e20657.
- Storey, M.A., Andreassend, S.K., Bracegirdle, J., Brown, A., Keyzers, R.A., Ackerley, D.F. et al. (2020) Metagenomic exploration of the marine *SpongeMycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. *MBio*, 11, e02997-19.
- Subramani, R. & Aalbersberg, W. (2013) Culturable rare Actinomycetes: diversity, isolation and marine natural product discovery. *Applied Microbiology and Biotechnology*, 97, 9291–9321.
- Suryanarayanan, T.S. (2012) The diversity and importance of fungi associated with marine sponges. *Botanica Marina*, 55, 553–564.
- Taylor, M.W., Radax, R., Steger, D. & Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, 71, 295–347.
- Teasdale, M.E., Liu, J., Wallace, J., Akhlaghi, F. & Rowley, D.C. (2009) Secondary metabolites produced by the marine bacterium *Halobacillus salinus* that inhibit quorum sensing-controlled phenotypes in Gram-negative bacteria. *Applied and Environmental Microbiology*, 75, 567–572.
- Terlouw, B.R., Blin, K., Navarro-Munoz, J.C., Avalon, N.E., Chevrette, M.G., Egbert, S. et al. (2023) MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*, 51, D603–D610.
- Thomas, T.R.A., Kavlekar, D.P. & LokaBharathi, P.A. (2010) Marine drugs from sponge-microbe association—a review. *Marine Drugs*, 8, 1417–1468.
- Tianero, M.D., Kwan, J.C., Wyche, T.P., Presson, A.P., Koch, M., Barrows, L.R. et al. (2015) Species specificity of symbiosis and secondary metabolism in ascidians. *The ISME Journal*, 9, 615–628.
- Trindade, M., van Zyl, L.J., Navarro-Fernández, J. & Abd Elrazak, A. (2015) Targeted metagenomics as a tool to tap into marine

- natural product diversity for the discovery and production of drug candidates. *Frontiers in Microbiology*, 6, 890.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. et al. (2017) Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems*. Long Beach, California, USA: Curran Associates Inc, pp. 6000–6010.
- Wakimoto, T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T. et al. (2014) Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nature Chemical Biology*, 10, 648–655.
- Wang, S.-M. (2016) Historical story on natural medicinal chemistry: drugs from ocean. *Chinese Traditional and Herbal Drugs*, 47, 1629–1642.
- Wang, Y., Lin, H., Huang, R. & Zhai, W. (2023) Exploring the plankton bacteria diversity and distribution patterns in the surface water of northwest pacific ocean by metagenomic methods. *Frontiers in Marine Science*, 10, 1177401.
- Waschulin, V., Borsetto, C., James, R., Newsham, K.K., Donadio, S., Corre, C. et al. (2022) Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *The ISME Journal*, 16, 101–111.
- Woodhouse, J.N., Fan, L., Brown, M.V., Thomas, T. & Neilan, B.A. (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *The ISME Journal*, 7, 1842–1851.
- Xiliang, Y., Mengyao, R. & Qian, L. (2021) Research progress on secondary metabolites from marine-derived *Trichoderma* sp. in the past decade. *Guangxi Sciences*, 28, 440–450.
- Yang, C., Qian, R., Xu, Y., Yi, J., Gu, Y., Liu, X. et al. (2019) Marine Actinomycetes-derived natural products. *Current Topics in Medicinal Chemistry*, 19, 2868–2918.
- Yang, Q., Franco, C.M.M. & Zhang, W. (2019) Uncovering the hidden marine sponge microbiome by applying a multi-primer approach. *Scientific Reports*, 9, 6214.
- Yang, Z., Liao, B., Hsieh, C., Han, C., Fang, L. & Zhang, S. (2021) Deep-BGCpred: a unified deep learning genome-mining framework for biosynthetic gene cluster prediction. *bioRxiv*, 2011, 2015.468547.
- Yuan, Y., Shi, C. & Zhao, H. (2023) Machine learning-enabled genome mining and bioactivity prediction of natural products. *ACS Synthetic Biology*, 12, 2650–2662.
- Zhang, L., An, R., Wang, J., Sun, N., Zhang, S., Hu, J. et al. (2005) Exploring novel bioactive compounds from marine microbes. *Current Opinion in Microbiology*, 8, 276–281.
- Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L. et al. (2021) Advances in metagenomics and its application in environmental microorganisms. *Frontiers in Microbiology*, 12, 766364.
- Zhou, Q., Hotta, K., Deng, Y., Yuan, R., Quan, S. & Chen, X. (2021) Advances in biosynthesis of natural products from marine microorganisms. *Microorganisms*, 9, 2551.
- Zhuang, L. & Zhang, H. (2021) Utilizing cross-species co-cultures for discovery of novel natural products. *Current Opinion in Biotechnology*, 69, 252–262.

**How to cite this article:** Wang, S., Li, X., Yang, W. & Huang, R. (2024) Exploring the secrets of marine microorganisms: Unveiling secondary metabolites through metagenomics. *Microbial Biotechnology*, 17, e14533. Available from: <https://doi.org/10.1111/1751-7915.14533>