



Review

Strategies for Natural Products Discovery from Uncultured Microorganisms

Khorshed Alam , Muhammad Nazeer Abbasi, Jinfang Hao, Youming Zhang * and Aiyong Li * 

Helmholtz International Lab for Anti-Infectives, Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China; mikrokhoshed@mail.sdu.edu.cn (K.A.); nazeer.abbasi1990@gmail.com (M.N.A.); h15935152510@163.com (J.H.)
* Correspondence: zhangyouming@sdu.edu.cn (Y.Z.); ayli@sdu.edu.cn (A.L.)

Abstract: Microorganisms are highly regarded as a prominent source of natural products that have significant importance in many fields such as medicine, farming, environmental safety, and material production. Due to this, only tiny amounts of microorganisms can be cultivated under standard laboratory conditions, and the bulk of microorganisms in the ecosystems are still unidentified, which restricts our knowledge of uncultured microbial metabolism. However, they could hypothetically provide a large collection of innovative natural products. Culture-independent metagenomics study has the ability to address core questions in the potential of NP production by cloning and analysis of microbial DNA derived directly from environmental samples. Latest advancements in next generation sequencing and genetic engineering tools for genome assembly have broadened the scope of metagenomics to offer perspectives into the life of uncultured microorganisms. In this review, we cover the methods of metagenomic library construction, and heterologous expression for the exploration and development of the environmental metabolome and focus on the function-based metagenomics, sequencing-based metagenomics, and single-cell metagenomics of uncultured microorganisms.

Keywords: natural products; uncultured microorganisms; metagenomics; extreme ecosystem; heterologous expression; eDNA; microbial dark matter; environmental metabolome; single cell sequencing



Citation: Alam, K.; Abbasi, M.N.; Hao, J.; Zhang, Y.; Li, A. Strategies for Natural Products Discovery from Uncultured Microorganisms.

Molecules **2021**, *26*, 2977. <https://doi.org/10.3390/molecules26102977>

Academic Editor: Rosa M. Durán-Patrón

Received: 23 March 2021
Accepted: 13 May 2021
Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Natural Products and Uncultured/Uncharacterized Microbes

Microbial natural products (NPs) are specialized metabolites, commonly treated as unprecedented important sources and perform a prominent impact, especially in pharmaceutical research and production [1–3]. Systematic exploration and biosynthetic elucidation of NPs is therefore an important turn against studying complex biodiversity and leveraging those chemical varieties for use in medicine, food, farming, and environmental safety [4,5]. However, due to the rediscovery of NPs in high rate by conventional methods, microorganisms ever seemed to have been depleted sources for following NP-mining for almost a century [6]. The latest boom of microbial genome sequences indicates a huge unexplored treasure trove of novel NPs [7]. Many new NPs have been uncovered through the mining of microbial genetic data by following recently developed synthetic biology-inspired strategies and tools. However, the genomic period has shown that cultivated bacteria constitute only a small portion of the overall known biomass of bacteria.

Microorganisms can inhabit as free living organisms or symbiotic associations with multicellular hosts or other unicellular organisms in various ecological habitats, from deepest seas to arid deserts and the guts of the multicellular organisms, even some symbiotic or parasitic systems [8]. It has been estimated that 4×10^6 different microbial taxa and 10^9 cells are present per ton and per gram of soil samples, while microbial soil diversity is estimated to have between 3000 and 11,000 genomes per gram of the soils [9]. However, less than 1% of the microbes of the ecosystems can be accessible through cultivation techniques and more than 99 % cannot be studied, restricting our knowledge of microbial morphology,

metabolism, genetics, and population ecology of uncultured microorganisms [10]. Maintaining or imitating natural growth conditions for these uncultured microorganisms is still a challenge [9]. Thus, it has been a challenge to identify the NPs from those uncultured or uncharacterized microbes for a long time.

1.2. Culture-Independent Tools for the Study of Uncultured Microorganisms Related to NP Accumulation

The advent of next-generation sequencing technology and comprehensive use of metagenomics-related techniques has made uncultured microorganisms accessible for functional research: (i) function-based metagenomics based on library construction, heterologous expression and functional screening, (ii) sequencing-based metagenomics [11], and (iii) single-cell metagenomics [12] offer alternate, culture-independent approaches, allowing researchers to discover complex microbial flora, novel taxa, and explore dark or hidden NPs from these uncultured microorganisms [13,14]. Insights into unusual enzymology in the context of NP biosynthesis suggested that the huge diversity of NPs awaits discovery [15]. Combining synthetic biological approaches with these metagenomics techniques will promote their wider application into the exploration of NPs from these uncultured microorganisms [14].

1.2.1. Evaluation of Microbial Heterogeneity and Abundance in Environmental Samples

To evaluate the heterogeneity and abundance of microbes with potentials to produce NPs in environmental samples, 16S rRNA gene sequencing has been used to classify uncultured microbial diversity. Moreover, some other functional genes have demonstrated higher accuracy for the analysis of genetic variation in the normal bacterial populations including genes *soxB* and *amoA* as unique genes to sulfur oxidizing bacteria and ammonia oxidizing microbes, and genes encoding conserved domains in polyketide synthetases (PKS) or nonribosomal peptide synthetases (NRPS) for NP-producing bacteria, and so on [12,16–20].

Because metagenomics avoids the isolation and propagation of organisms and does not entail prior understanding of the organisms present in the samples, its application facilitates researchers to look at the broad range of individual genes and whole operons for biosynthetic or degradative pathways, and thus promotes the biochemical characterization of diverse natural samples and the discovery of unique bioactive compounds and enzymes from environmental DNA samples [21]. In the first microbial metagenomics project, Venter and colleagues reported 1800 genomic organisms, comprising 148 new species phylotypes within water bodies of the Sargasso Sea [22].

1.2.2. Metagenomics-Related Approaches

Since 1998, metagenomics has appeared as a concept and been developed into important tools for gaining insight into the morphology and genetics of uncultured species [21].

For function-based metagenomics, the construction of a metagenomic library in a suitable host is the first step by cloning fragmented environmental DNA (eDNA) into expression systems and then introducing these metagenomic clones into a heterologous host [23]. Next, screening targets of interest genes or gene clusters could be conducted using function-based screening. Biosynthetic pathways of NPs of uncultured microorganisms can be obtained for the detection of compounds through the design and deployment of a metagenomic library, metabolic screens, and color screens (Figure 1).

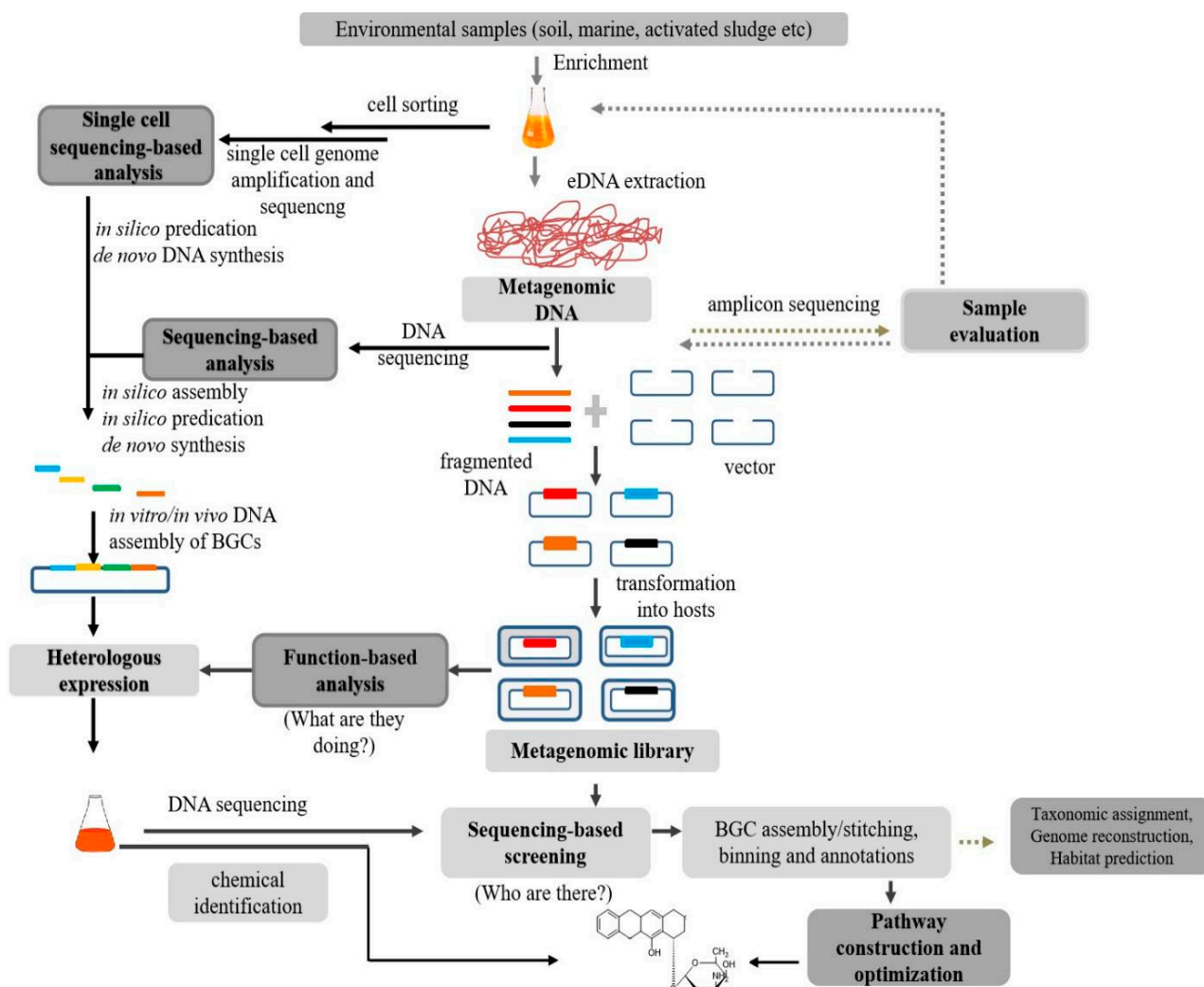


Figure 1. Graphical representation of approaches for NP discovery from environmental samples. Enrichment and evaluation of environmental samples are necessary steps prior to metagenomics analysis. Function-based metagenomics analysis includes library construction, heterologous expression, and function-based screening. Sequencing-based metagenomics analysis requires eDNA sequencing, DNA synthesis and assembly, and heterologous expression of target clones. Single-cell metagenomics analysis includes cell sorting, single cell genomic DNA amplification, genome sequencing, and DNA synthesis and assembly of NP biosynthetic gene clusters (BGCs) of interest, then expression and chemical identification.

For sequencing-based metagenomics, total eDNA isolation and sequencing, pathway predication and designing, and DNA assembly are required, which follow heterologous expression of target clones (Figure 1).

Many excellent reviews on function-based metagenomics [24], sequencing-based metagenomics [25], and their applications in biotechnology are available. In particular, the number of a particular species in the metagenomic library is determined by factors such as genome size, genome copy number, within-species heterogeneity, and relative frequency of the species as well as biases in DNA extraction and sequencing.

Single-cell metagenomics is also an effective tool for insights into the taxonomy, physiology, and biochemistry of specific uncultured species in environmental samples [12,26–28] (Figure 1). It depends on the sorting of cells present in environmental samples, single-cell genomic DNA isolation, amplification and sequencing, and target BGC assembly and expression. Recently, it has been used for the exploration of NPs from different environmental samples such as marine animal tissues.

2. Function-Based Metagenomics for Exploration of NPs from Different Environmental Samples

This metagenomic analysis focused on the cloning of eDNA into expression vectors and their propagation in suitable hosts, accompanied by an evaluation of recombinant action. Metagenomic gene libraries may be used to identify novel enzymes encoded by a single gene or a small operon, whereas large insert libraries are necessary for the extraction of large BGCs that encode complex pathways containing multiple genes for the production of NPs [29].

2.1. Sample Preparation and Evaluation, Library Construction, and Heterologous Expression

Generally, it is necessary to enrich and evaluate environmental samples prior to the creation of the metagenomic library to improve the performance of gene screening through metagenomic analysis (Figure 1). Aided by next generation sequencing, PCR amplification of particular functional genes using degenerated primers was used to study the genetic variation in the environmental samples while amplification of 16S rRNA by PCR was applied to evaluate the genetic variety in prospective environmental samples for the development of metagenomic libraries.

Additionally, a microarray technology was also used to evaluate and disseminate functional genes through DNA array technology. Suppressive subtractive hybridization (SSH) was used to assess microbial heterogeneity and functional discrepancies in genetic information among different environments. Stable isotope sampling probing (SIP) was used to identify DNA and RNAs of each gene [30]. After that, the tagged DNA and RNAs can be isolated using density gradient centrifugation. Multiple displacement amplification (MDA) has been applied to improve the environmental yield of DNA from low-biomass environmental samples [23].

In order to increase the possibility of extracting NP complete gene clusters during the construction of metagenomic libraries, the library size and gene cluster structure need to be considered carefully before library construction. Thus, methods divided as direct or indirect extraction procedures for separating DNA from environmental samples like soils/sediments have been developed [31]. Direct DNA isolation methods depend on the in situ lysis of microbial cells in the specimens before retrieval and extraction of DNA. Indirect DNA isolation techniques have been established to shield DNA from mechanical stress, which can cause DNA trimming and allow large fragments to be obtained.

Short insert libraries will not promote the identification of larger gene clusters (20–200 kb) or operons for NPs. To address this restriction, bacterial artificial chromosome (BAC) vectors have been acclaimed for creating large-insert libraries (up to 200 kb).

Escherichia coli strain is a better host for the cloning and expression of Gram-negative bacterial enzymes, but not perfect for the cloning and expression of Gram-positive bacterial BGCs with high GC content. In order to improve the biosynthetic variability and expression-related problems defined in function-based metagenomic studies, a variety of study teams have established expression mechanisms focused on non-*E. coli* hosts.

The Brady group evaluated the capacity to produce similar biosynthetic genes with pigmented or antibacterial phenotypes for a variety of host strains [32,33]. In most instances, only the hosts where genes were initially extracted or close-phylogenetically hosts were recorded to show activity or functional expression of foreign genes. This finding highlights the important role of developing alternate hosts in the exploration of NPs powered by metagenomics. *Streptomyces lividans* (*S. lividans*), *S. albus*, *Pseudomonas putida* (*P. putida*), *Ralstonia metallidurans*, *Agrobacterium tumefaciens*, *Burkholderia graminis*, and *Caulobacter vibrioides* bacteria have been used as eDNA recipient hosts [34].

To increase heterologous expression of target genes or BGCs located in the metagenomic libraries, some regulators were introduced into heterologous hosts. For example, introduction of heterologous sigma factors could significantly improve expression of eDNA [35]. The *trp* operon and genes employed in the biosynthesis of new antibiotics were identified using *S. lividans*, *R. leguminosarum*, and *P. putida*. *R. leguminosarum* was used as

a suitable host for the expression of functional genes. *S. lividans* was often used for the expression of DNA derived from the soil. Five novel molecules, terragins A–E, along with siderophoric nocardamine, have been evaluated for these strains. *P. putida* is not a suitable host for screening the inhibition zone [32].

Advanced BAC shuttle vectors were constructed for high-throughput conjugation of big threads of eDNA from *E. coli* to *S. lividans* and *P. putida*. Martinez observed that genes responsible for the production of various antibiotics in *E. coli*, *P. putida*, and *S. lividans* were expressed differently using this shuttle vector [36]. A variety of polyketides were obtained from *R. metallidurans*, which were new [37] and two were amides gained by antibacterial screening [32].

2.2. Function-Based Screening

Function-based screening in a metagenomic study is a potent, but challenging technique that requires the creation of an expression library first, and next the measurement of a particular reaction with a substrate to evaluate the biochemical and metabolic functions of relevance [38]. A variety of antibiotics (Table 1), hydrolytic enzymes, antibiotic-resistance genes, Na(Li)/H transporters, and several other functions have been successfully identified using function-based metagenomics [23]. Furthermore, it permitted the identification of genes with specific behavior encoding enzymes, which are entirely new sequence forms with no known counterparts. Some extreme ecosystems have been chosen for study and allow for the detection of enzymes with the expected properties [39]. For instance, it might be predicted that thermal spring metagenome will contain unique genes that encode thermostable enzymes. The method's benefit is that it needs no sequence analysis of the genes of interest, rendering it capable of identifying completely new groups of genes for unknown or unknown roles.

Table 1. Antibiotics discovered by function-based screening of metagenomic clones.

Compound	Habitat	Library Type	Reference
Fasamycin A and B	Soil	Cosmid	[40]
Indirubin	Soil	Fosmid	[41]
Terragine	Soil	Cosmid	[42]
Turbomycins A and B	Soil	BAC	[43]
Violacein	Soil	Cosmid	[40]

2.2.1. Approaches for Function-Based Screening

Function-based screening is capable of the identification of new gene groups that encode new BGCs for natural compounds and functional enzymes. The identification of clones with changed phenotypes is one of the simplest screenings. These optical displays may be focused on pigmentation (color), development of inhibition zones around clones of metagenomic library forming on bacterial or fungal test lawns or blood agar plates as a result of biocatalytic conversion (production of antimicrobial or hemolytic factors) [34]. Along with turbomycin A (1, Figure 2) originally identified from a fungus, the red broad-spectrum antibiotic turbomycin B (2, Figure 2) [43] was extracted, as an example of a new NP uncovered by color/hemolytic screening. *N*-acyl amino acid derivatives and structurally related drugs as well as palmitoyl putrescine were among the first NPs identified using inhibition zone-based screening [44–46]. However, initially, metagenomic libraries had a low sensitivity and low performance based on function-based screening.

The advent of hybrid practical “intracellular” screening methodologies to distinguish clones of concern like METREX (metabolite-regulated expression) [47], SIGEX (substrate-induced gene expression) [48], and PIGEX (product-induced gene expression) [49] has sped up the extraction of new biocatalysts from microbial communities. Fluorescence-activated cell sorting (FACS) or fluorescence microscopy was used for these screening methods. FACS has a wide variety of uses in high-throughput metagenomic clone screening since it can detect biological activity within a single cell [50].

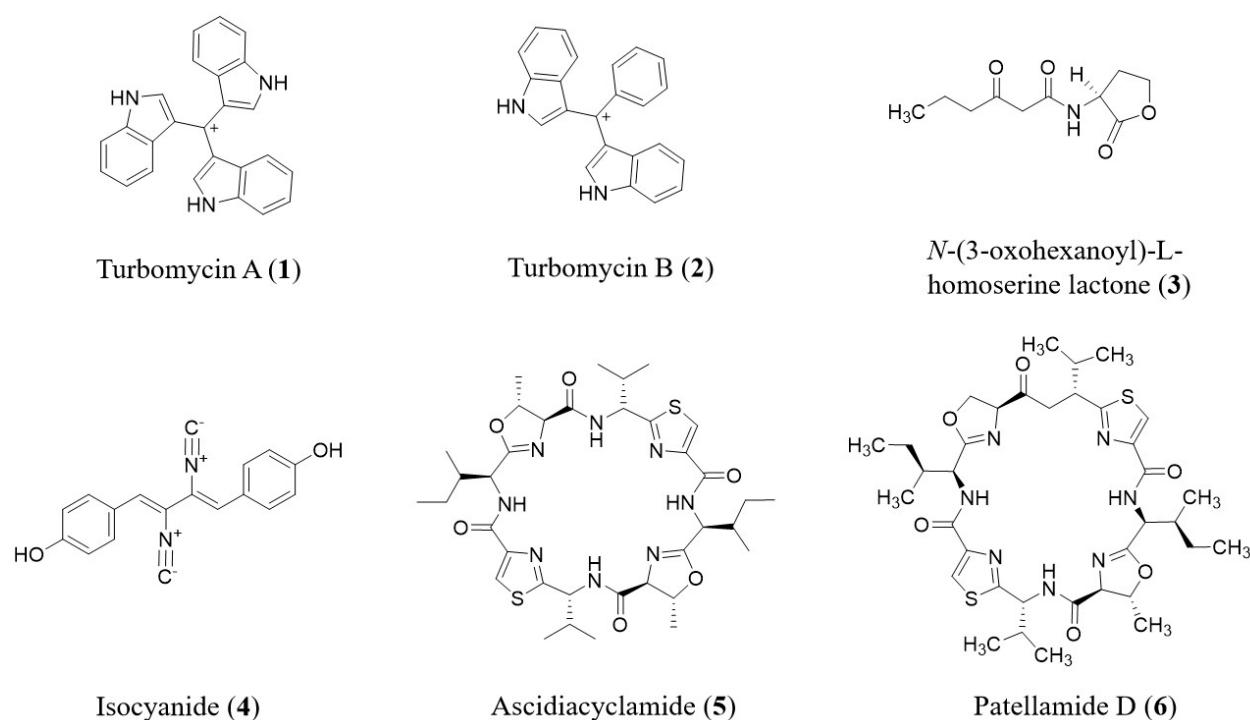


Figure 2. Natural products discovered by function-based screening of metagenomic-library clones.

METREX, a framework wherein metagenomic DNA is embedded with a biosensor in a host cell, which can trigger bacterial quorum sensing [47,51]. The METREX system was used to retrieve quorum-sensing inducer-generating metagenomic clones from soil and activated sludge metagenomic libraries [38]. Green fluorescent protein (GFP) expressed by the metagenomic clone in the host cell can be observed by using fluorescent microscopy or a spectrophotometer. This technique succeeded in the preliminary detection of the recognized *N*-(3-oxohexanoyl)-*L*-homoserine lactone (3, Figure 2) and an unknown indoxyl degradation agent developed by a DNA library from soil and gypsy moth, *Lymantria dispar* [47,52].

SIGEX is another intracellular method of screening, based upon the understanding that catabolic gene expression is normally triggered by appropriate substrates and, in many cases, regulated by regulatory elements. A shotgun cloning with an operon-trap *gfp*-expression vector was developed to allow the collection of positive clones in liquid cultures using fluorescence-activated cell sorting. The cloning of new aromatic hydrocarbon-induced genes from a groundwater metagenomic library was illustrated using this screening platform [53].

Some studies have reported lately on high efficient screening of novel enzymes and natural compounds from a metagenomic library using microfluidic technique, despite the fact that the traditional isolation of active clones was done using an *in vitro* method [54–57]. Hosokawa et al. unveiled individual cells embedded in gel microdroplets (GMDs) for the screening of lipolytic enzyme genes from the soil-based metagenomic library [58]. Such findings showed the capacity of microfluid systems as an effective method for detecting new metagenomic enzymes, and minimizing time and cost.

In 2011, Jiang et al. noticed a new β -glucosidase gene (*bg11D*) with lipolytic activity (renamed Lip1C) that was detected by a function-based screening of a soil-based metagenomic library [59]. Lipase and esterase continue to be the most specific enzyme activities using the function-based screening of diverse metagenomic libraries [60,61]. Some of the early metabolites showed intriguing changes that lead to enamine and enolether moieties. Isocyanide (4, Figure 2) is another antibiotic with a unique structure [62].

Chemical analysis such as high-performance liquid chromatography-mass spectrometry (HPLC-MS), is an effective yet time-consuming version of a function-based screening

for identifying compounds in samples [42,63]. The screening time can be reduced by first assessing pools of clones. Biosynthetic genes for patellamide D (6, Figure 2) and ascidiacyclamide (5, Figure 2) have been discovered using a BAC library developed from the DNA of uncultured tunicate-derived cyanobacterial symbionts [63].

Although the results of its phenotypic screening of *E. coli*-based libraries have been promising in terms of innovative enzymology, the rate of positive clones and the structural variability of the molecules identified are usually limited. Brady et al. observed antimicrobial properties in one out of every 10,000 to 20,000 eDNA cosmid clones that express antibiotic proteins and ribosomal peptides [45]. An antifungal screening using *S. cerevisiae* revealed only one successful clone out of over 110,000 fosmid clones [64]. The understanding that heterologous expression of eDNA can lead to the exploration of new fascinating biochemistry was among the most intriguing aspects of these studies.

2.2.2. Limitations of Function-Based Screening

Function-based screening is a simple system that allows researchers to access an immense genetic resource in a microbial population through the expression of cloned genes from metagenomic origin in heterologous hosts, without knowledge of the origin of the microbe or original gene sequence, or NP structure, or composition of the target proteins. However, the lack of direct and powerful detection methods for certain products and a limited selection of hosts for the expression of certain foreign genes or gene clusters have been major drawbacks in function-based screening.

First, the variety of functional activities is restricted to current screening methods like antibiotics or hydrolytic enzyme identification. Second, it is hard to find a suitable host able to post-translationally alter apo proteins and heterologously express all genes of one BGC. In particular, those genes derived from evolutionarily distant uncultured species cannot be effectively expressed in standard hosts as *E. coli*, due to divergent usage of codon, incompatible regulatory factors, lack of biosynthetic precursors, and a drawback to the extraction of full information from function-based screening [38]. Recently, new transformation systems using various microbes with additional platforms of gene expression and a variety of protein secretion mechanisms have been reported. Next, a constraint in size for metagenomic libraries was introduced. A fosmid library of 50,000 clones with an average insert size of 40 kbp equals around 500 bacterial genomes (4 Mbp size), even while 1 g of soils could constitute thousands of varying microbial species [65].

The carrier protein (CP) domains of multimodular nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) enzymes act as anchors to the developing natural product chain. In standard *E. coli* expression hosts, CPs are unresponsive until they are translated into holo domains by attaching a 4-phosphopantetheinyl unit [66]. Moreover, *E. coli* lacks the ability to synthesize methylmalonyl-CoA (coenzyme A), a necessary component for the development of several methyl-branched polyketides, like antibiotic erythromycin [67]. *E. coli* can be transformed into an effective producer of complex PKSs and NRPSs by the incorporation of genes for Phosphopantetheinyl transferases (PPTs) and for branched PKSs methylmalonyl-CoA [67]. However, due to the large size of the most complex PKSs and NRPSs gene loci, which outstrips the normal fosmid and sometimes even BAC inserts, no efficient functional exploration methods for these compounds are known to date.

3. Sequencing-Based Metagenomics

In contrast to function-based metagenomics, sequencing-based metagenomics does not need heterologous expression to distinguish metagenomic clones of consideration. Additionally, even if NP pathways are not observed in the library host, they can be identified. This method avoids the difficulties of heterologous expression by using sequence analysis to classify gene clusters of interest. After that, heterologous expression experiments can be used to produce molecules based on the target pathways. Some examples of NPs screened by such a sequencing-based approaches were identified from symbiotic

systems (Table 2, Figure 3), where most of the microorganisms cannot live freely, but in a host-dependent mode.

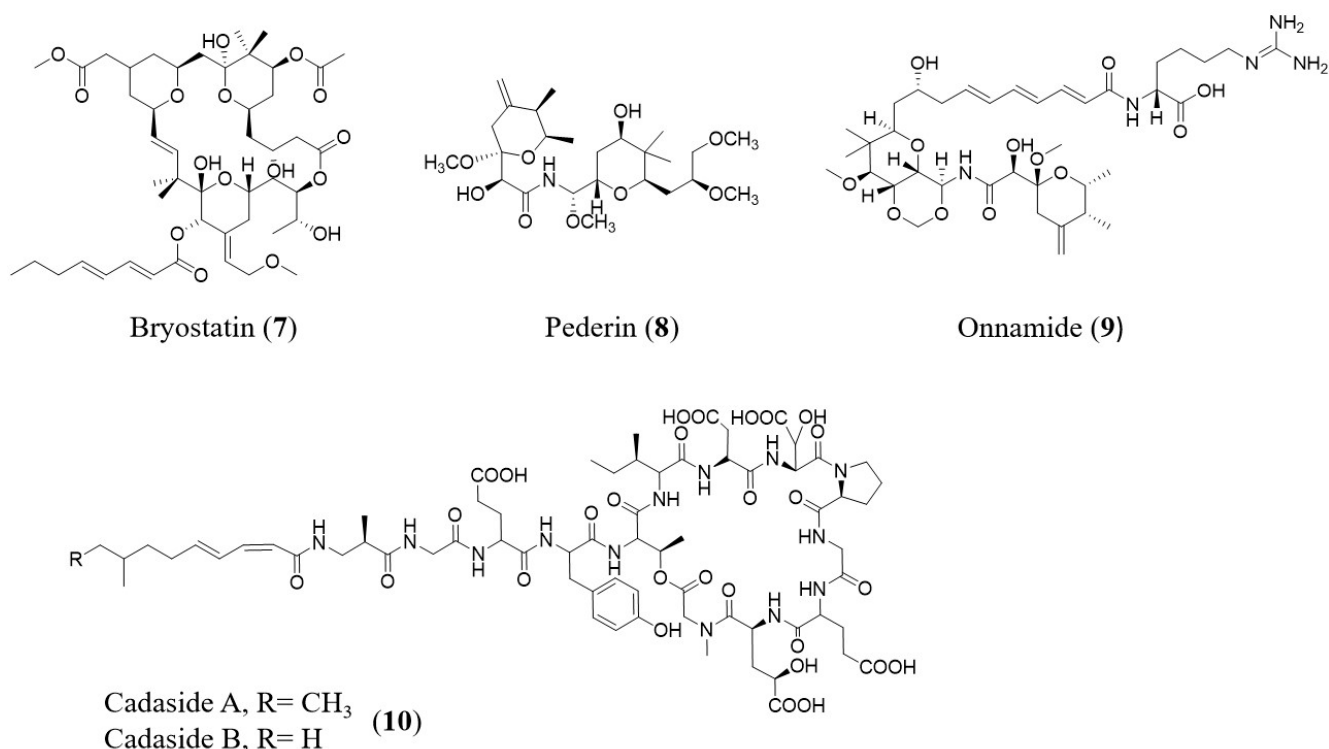


Figure 3. Natural products identified/studied by sequencing-based approaches.

Early implementations of sequencing-based approaches employed primers specific to genes preserved in comparatively limited NP biosynthesis families. Several of the latest studies have been carried out to increase such endeavors in order to identify a wider variety of retained biosynthetic NP domains with an important emphasis on domains found among biosynthetically modified PKS and NRPS genes [68–71]. For example, for large multimodular PKS and NRPS clusters that typically surpass the size of fosmid or even BAC inserts, a sequencing-based screening is the best suitable approach.

Another important example is the discovery of cadasides (CDE) from a soil sample: the Brady SF group used NRPS adenylation (AD) domain sequencing to guide the identification, recovery, and cloning of the *cde* BGC, which led to the production of cadasides A and B (13, Figure 3), a subfamily of acidic lipopeptides. They found that sequencing of AD domains from diverse soils revealed that these sequences predicted to arise from cadaside-like gene clusters are predominantly found in soils containing high levels of calcium carbonate [72] (Table 2).

Sequencing analysis driven by the discovery of phylogenetic markers is a powerful technique first introduced by the DeLong group, which provided the first genomic sequence related to the 16S rRNA gene of an uncultured archaeon [73]. Targeted gene sequencing or shotgun metagenome sequencing are two commonly used platforms in microbial metagenomics [74]. When targeting genomics, eDNA is extracted and purified using highly effective DNA isolation and purification methods. The selected genes are multiplied with oligonucleotide markers and sequencing adaptors, specifically designed sequences for the combined various specimens [75]. In a shotgun sequencing method, the metagenomic DNA is segmented, finally reconstructed, and connected to adaptors that enable template replication and eventual sequencing to produce a large number of short readings that can be subsequently constructed *in silico*, guiding to the creation of NP BGCs of interest using different computational tools and synthetic biology methods [75].

Table 2. Natural products identified/studied by sequencing-based approaches.

Compound	Sources	Mode of Action	Reference
Bryostatin #	<i>Bugula neritina</i>	Antitumor	[76]
Pederin #	<i>Paederus fuscipes</i>	Antibacterial	[77,78]
Onnamide #	<i>Theonella swinhoei</i>	Antitumor	[79]
Cadasides	Soil sample	Antibacterial	[72]

Note: # indicating these compounds were isolated via sequencing-based screening from uncultivated microorganisms that live in symbiotic systems.

Degenerate primers that target certain typical biosynthesis domains are constructed to create intricate PCR amplicons from gene clusters contained in specific microbial communities or metagenomic libraries. Specific next generation sequencing (NGS) readings obtained from such PCR amplicons are referred to as natural product sequence tags (NPSTs) [11]. The data provided by these studies have resulted in new bioinformatic technologies, which can affirm and arrange NPST databases. Although a range of bioinformatics tools have already been designed to mine full sequenced genomes for secondary metabolite biosynthetic clusters, these programs typically need much greater DNA fragment inputs than metagenomic sequencing efforts provide.

Programs like eSNaPD (Environmental Surveyor of Natural Product Diversity) [71] and NaPDoS (Natural Product Domain Search) [80] were designed to evaluate very large NPST datasets. Such bioinformatics tools interpret NPSTs with standard libraries of sequences collected from the BGCs to determine the output of the gene clusters and the possible chemical structures of the environmental samples or libraries. This method is related to the reconstruction of whole species within the ecosystem using 16S rRNA sequences [81].

Two routes for environmental NP mining are established by the phylogenetic assembly of BGCs (Figure 1):

(i) Constituents of characterized NPs may be mined through the pursuit of NPST-associated gene clusters that are closely linked to biosynthetic domains that encode recognized natural products [72,82–84]. Comparative searches of these near and distant NPSTs by using bioinformatic tools are easy. However, when the required sequence alignment markers are experimentally determined, their functions have been proven to be reliable indicators of biosynthetic pathways and compound production [81]. eSNaPD has NPST analysis features that can organize the data to enable archiving and comparative study of NPST datasets from a number of environments. Such approaches for investigating biosynthetic complexity in environmental and metagenomic library collections are low-cost and computationally intensive. However, only a small number of PCR amplicons are sequenced and then computed instead of entire genomes.

(ii) In a random sequence (shotgun sequencing), larger fragments of DNA are broken into shorter fragments and randomly sequenced for multiple rounds (Figure 1). After separation, the entire sequenced are re-assembled using several overlapping of the sequenced fragments [85]. Necessary steps required for the shotgun sequencing are (i) high-quality metagenomic DNA isolation, (ii) random metagenomic DNA fragmentation (ultra-sonication), (iii) size fractionation (electrophoresis), (iv) metagenomic library construction, (v) paired-end sequencing, and (vi) assembly in silico of sequenced segments. Random metagenomic sequence is preferred for characterizing the genome content of bacteria, archaea, and viruses and their secondary metabolic pathways [86]. This method allows complete genomic DNA to be recovered, possibly from sources without in vitro cultivation. Random sequencing has benefits over other methods in that it speeds up the gene mapping and takes less effort to map. Conversely, it mechanistically lacks positive sorting and recognition of the biological functionaries' active classes. Furthermore, certain drawbacks are often related when applied to metagenomes composed of multiple repeated DNA sequences, and thus provide inexact evidence for the genome sequenced [87].

With the cost of DNA synthesis falling, synthetic biology will substitute the desire to isolate cells from the source environments, in order to have access to heterologous expression of BGCs. Calcium-dependent antibiotic malacidins active against multidrug-resistant pathogens are commonly encrypted in soil microbiomes and discovered via a sequencing-based metagenomic strategy [88,89].

However, additional quality controls are needed because the risk of chimeras in metagenomic sequence clusters is always present, particularly with fast-evolving and repeated genes in BGCs that encode modular polyketide synthases, for example. Particularly, metagenomics is a transient research area in its present 'short read-based' nature, since the technological constraints that currently characterize it are likely to be overcome quickly. Despite the expected new technologies in the coming years, the shotgun sequencing method for metagenomic BGC classification may completely substitute tag sequencing [90].

Nevertheless, many BGCs surpass the normal insert size of the cosmid, and thus many overlapping DNA cosmid clones covering the entire BGCs must be rearranged into a bacterial artificial chromosome (BAC) (stitching) by the commonly employed transformation-associated recombination process (TAR) or Red/ET-mediated recombineering [91]. Subsequently, the reassembled BGC may be moved from the yeast/*E. coli* to separate bacterial hosts for heterologous expression and further pathway optimization (Figure 1). As a result, this technique is more a heterologous language dependent on a metagenomic library than expression of metagenomic samples.

4. Single-Cell Metagenomics

As opposed to sequencing-based metagenomics, single-cell metagenomics analyzes the genomes of individual cells retrieved from the environmental samples, which examines the overall population DNA with subsequent bioinformatics-based binning of obtained sequences among individual microbial organisms [26–28]. One significant benefit of this strategy is that it is new and effective for analyzing uncultured microorganisms in environmental samples and determining whether metabolites as well as other genes are linked with phylogenetic knowledge, which has been challenging for general metagenomic techniques. Thus, it is considered as a powerful complement to the general metagenomics techniques (Figure 1).

4.1. Cell Sorting and Single-Cell Sequencing

Isolation of single cells from environmental samples is the first step of single-cell metagenomics, generally depending on flow cytometry, microfluidic instruments, and micromanipulation [92].

PCR can then be applied specifically on the cells in order to connect the feature with taxonomy (digital PCR). For example, after screening cells with a multiplex microfluidic system, genes from primary metabolism in a termite gut culture were linked in this way with a spirochete symbiont [93].

In some cases, it is indeed helpful to amplify a cell's genome before additional study, a technique known as whole-genome amplification (WGA). It is possible due to bacteriophage polymerase, which produces microgram quantities of DNA in only one reaction, linking multiple displacement amplification (MDA) [94]. This method revealed the origins of an uncommon community of PKSs found in microbial–sponge interactions and supposed to play a role in the biosynthesis of methyl-branched fatty acids [95,96]. Single cells were processed in such experiments by dissociating sponge tissue and filtering bacteria into 96-well plates using flow cytometry [97,98].

It was also noted that some WGA approaches set up in human genetic study could provide methodological directions for future single-cell genome sequencing. For instance, MALBAC (Multiple Annealing and Looping Based Amplification Cycles) was exemplified for single-cell genomic sequencing [99,100]. It required specifically-designed primers having a 27-nt common sequence, followed by eight random nucleotides. Using these primers for PCR amplification of the single cell genome, the resultant amplicons could form

loops due to the presence of the two complementary terminals, which may be resistant to subsequent amplification and hybridization. Combining next-generation sequencing, this approach was expected to be applied into uncultured microorganisms.

4.2. Screening of NP Gene Clusters in Single-Cell Metagenomics

Single-cell metagenomics as an important tool for studying environmental bacteria's lifestyles has a lot of potentials in NP analysis [100–102]. The availability of single-cell genome sequences provides rich genetic resources for NP exploration from uncultured microorganisms. For example, study of uncultured sponge-symbiotic *Entotheonella* single-bacterial genomics confirms a rich resource of different specialized metabolites in sponges [96].

Marine sponges were considered to have plenty of microbes with giant potentials to produce NPs. Following MDA, DNA samples were screened that revealed the PKSs belong to members of the rare candidate phylum “*Poribacteria*” [103], which is extensively connected with sponges but almost missing in other habitats [97]. The expression of the PKS genes was confirmed by partial sequencing of the genome of single cell amplified poribacteria [98]. This was the earliest research to use single-cell metagenomics to chart the positions of NP genes. Although the construction of chimeric amplicons during amplification may render genomic assembly difficult, there have been reports of entire genomes being assembled from single cells [104].

The drawbacks of this method are related to single cell isolation and lysis, contamination susceptibility, and imperfect and unevenly amplifying single-cell genomes (single amplified genomes, SAG) [18]. As a consequence, sequenced SAGs are commonly characterized by a series of contigs that only occupy a small portion of the whole genome [26]. The combination of data from a sequence of several SAGs from cells of a single species might overcome such challenges [13]. Rinke et al. sequenced 201 SAGs comprising 29 uncultured lineages of “microbial dark matter” as an example of active implementation of single-cell metagenomics [105,106].

Using single cell genome sequences containing NP BGCs, the combining of DNA synthesis and BGC assembly with heterologous expression will be applied to establish biosynthetic pathways of target NPs hidden in uncultured microorganisms [90].

Both single-cell metagenomics and sequencing-based metagenomics yield DNA sequences from uncultured microorganisms, but which are radically distinct. SAGs from single-cell metagenomics are genomes of individual organisms, while sequencing-based metagenomics bins are aggregate genomes of a genetically heterogeneous community. The latter is ineffective since it does not allow for the full genetic capacity of various strains of the same species present in a community [18].

5. Conclusions

Different environmental habitats have different ecosystems and biodiversity, and have different potentials for NP discovery. For example, in animal guts, many NPs were accumulated by gut microorganisms related to many physiological processes of animals [107], and metagenomic analysis revealed a high diversity in animal gut microbiome [108]. Strategies for multi-dimensional visualization of metagenomic data have been developed [109]. An Influential international project (The Earth Microbiome Project) launched in August 2010 has been attempting to classify worldwide microbial diversity in terms of phylogeny and function [110].

Though the number of NPs isolated from environmental samples is still extremely low, a high number of NP BGCs have been identified from uncultivated microorganisms that will still be great treasure troves for NP discovery in this field [111,112]. Among the approaches used in study on environmental samples or uncultured microorganisms, metagenomics provides innovative tools for discovering natural resources that has the ability to access historically undiscovered biosynthetic diversity. Latest advancements in bioinformatics techniques and sequencing technologies (e.g., pyrosequencing, nanopore se-

quencing) have enabled the analysis of whole microbial populations made up of unknown organisms by metagenomics.

In the coming years, there will be a combination of fast-developing long-read technologies and new bioinformatics methods that will enable scientists to acquire complete genome sequences of all or complete gene clusters of NPs, but the least influential members of the microbial population by scanning them directly from the broader metagenomic data collection. In the near future, hybrid methods that incorporate assembly and tag sequences may also be of big benefit; in such a hybrid technique, integrated metagenomic contigs may be used to classify new candidate groups of BGCs for which primers are then built for further investigation through tag sequences, phylogenomic analysis, and cosmid sequences.

Thus, metagenomics is opening up unique opportunities for the identification of bioactive compounds and several valuable compounds are anticipated to be identified and exploited in the near future. Interdisciplinary studies into bioactive compounds synthesized by NRPS and PKS will continue to be broadening in order to access the wide range of secondary metabolites obtained by microorganisms (including uncultured or yet-uncultured species). Moreover, metatranscriptomics, metaproteomics, and metabolomics techniques may help researchers better understand how microbes respond in a community [113].

Even by only changing the cultivation conditions or establishing novel screening methods, uncultured microorganisms could become culturable and accessible, and this is still proven to be reliable for the discovery of some new NPs, exemplified by the exploration of teixobactin, considered as a new generation of antibiotic against multi-drug resistant (MDR) pathogens [114]. Single-cell genome mining of NPs for insight into environmental samples or uncultured microorganisms was exemplified with the biosynthesis of the extraordinary complex hypermodified polytheonamide-type compounds, which BGC sequence was identified originally from a sponge-symbiotic uncultivated species in *Entotheonella*. The availability of such a BGC sequence guided the discovery of similar biosynthetic pathways in some culturable bacteria, suggesting that some BGCs are subject to extensive horizontal gene transfer between culturable and uncultivated sources [115].

With advances in the synthetic biological methods used in DNA synthesis, DNA assembly, and direct cloning, and refactoring of large BGCs, combining of updated bioinformatics, combinatorial chemistry, and synthetic biological methods with metagenomic approaches would further promote the development or man-made modification of potentially necessary medicinal compounds from uncultured microorganisms.

Author Contributions: K.A.: Draft writing; M.N.A. and J.H.: Figure drawing, Y.Z.: Draft planning and organization; A.L.: Draft organization and manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This study was supported by the National Key R&D Program of China (No. 2019YFA0905700, 2018YFA0900400) and the Shandong Province Natural Science Foundation (ZR2017MC031). We are thankful to the National Natural Science Foundation of China (31170050), the Open Project Program of the State Key Laboratory of Bio-based Material and Green Papermaking (KF201825), and the 111 Project (B16030) for financial support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

BGC	Biosynthetic Gene Cluster
NP	Natural Products
eDNA	Environmental DNA
BAC	Bacterial Artificial Chromosome
SAG	Single Amplified Genomes
NPST	Natural Product Sequence Tags
CP	Carrier Protein
eSNaPD	Environmental Surveyor of Natural Product Diversity
NaPDoS	Natural Product Domain Search

References

1. Kalkreuter, E.; Pan, G.; Cepeda, A.J.; Shen, B. Targeting bacterial genomes for natural product discovery. *Trends Pharmacol. Sci.* **2020**, *41*, 13–26. [[CrossRef](#)] [[PubMed](#)]
2. 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.* **2020**, *83*, 770–803. [[CrossRef](#)] [[PubMed](#)]
3. Shen, B. A new golden age of natural products drug discovery. *Cell* **2015**, *163*, 1297–1300. [[CrossRef](#)]
4. Kim, E.; Moore, B.S.; Yoon, Y.J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **2015**, *11*, 649–659. [[CrossRef](#)] [[PubMed](#)]
5. Smanski, M.J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M.A.; Voigt, C.A. Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **2016**, *14*, 135–149. [[CrossRef](#)]
6. Xu, F.; Wu, Y.; Zhang, C.; Davis, K.M.; Moon, K.; Bushin, L.B.; Seyedsayamdost, M.R. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nat. Chem. Biol.* **2019**, *15*, 161–168. [[CrossRef](#)]
7. Baltz, R.H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 573–588. [[CrossRef](#)]
8. Wilson, M.C.; Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* **2013**, *20*, 636–647. [[CrossRef](#)]
9. Chaudhary, D.K.; Khulan, A.; Kim, J. Development of a novel cultivation technique for uncultured soil bacteria. *Sci. Rep.* **2019**, *9*, 1–11. [[CrossRef](#)]
10. Vartoukian, S.R.; Palmer, R.M.; Wade, W.G. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol. Lett.* **2010**, *309*, 1–7.
11. Katz, M.; Hover, B.M.; Brady, S.F. Culture-independent discovery of natural products from soil metagenomes. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 129–141. [[CrossRef](#)] [[PubMed](#)]
12. Xu, Y.; Zhao, F. Single-cell metagenomics: Challenges and applications. *Protein Cell* **2018**, *9*, 501–510. [[CrossRef](#)] [[PubMed](#)]
13. Van Dorst, J.M.; Hince, G.; Snape, I.; Ferrari, B.C. Novel culturing techniques select for heterotrophs and hydrocarbon degraders in a subantarctic soil. *Sci. Rep.* **2016**, *6*, 1–13. [[CrossRef](#)] [[PubMed](#)]
14. Nayfach, S.; Shi, Z.J.; Seshadri, R.; Pollard, K.S.; Kyrpides, N.C. New insights from uncultivated genomes of the global human gut microbiome. *Nature* **2019**, *568*, 505–510. [[CrossRef](#)]
15. Scott, T.A.; Piel, J. The hidden enzymology of bacterial natural product biosynthesis. *Nat. Rev. Chem.* **2019**, *3*, 404–425. [[CrossRef](#)]
16. Garrity, G.M. A new genomics-driven taxonomy of Bacteria and Archaea: Are we there yet? *J. Clin. Microbiol.* **2016**, *54*, 1956–1963. [[CrossRef](#)]
17. Krishnani, K.K.; Kathiravan, V.; Natarajan, M.; Kailasam, M.; Pillai, S.M. Diversity of sulfur-oxidizing bacteria in greenwater system of coastal aquaculture. *Appl. Biochem. Biotechnol.* **2010**, *162*, 1225–1237. [[CrossRef](#)]
18. Hedlund, B.P.; Dodsworth, J.A.; Murugapiran, S.K.; Rinke, C.; Woyke, T. Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter". *Extremophiles* **2014**, *18*, 865–875. [[CrossRef](#)]
19. Madhavan, A.; Sindhu, R.; Parameswaran, B.; Sukumaran, R.K.; Pandey, A. Metagenome analysis: A powerful tool for enzyme bioprospecting. *Appl. Biochem. Biotechnol.* **2017**, *183*, 636–651. [[CrossRef](#)]
20. Wei, Y.; Zhang, L.; Zhou, Z.; Yan, X. Diversity of gene clusters for polyketide and nonribosomal peptide biosynthesis revealed by metagenomic analysis of the yellow sea sediment. *Front. Microbiol.* **2018**, *9*, 295. [[CrossRef](#)]
21. Handelsman, J.; Rondon, M.R.; Brady, S.F.; Clardy, J.; Goodman, R.M. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem. Biol.* **1998**, *5*, R245–R249. [[CrossRef](#)]
22. Venter, J.C.; Remington, K.; Heidelberg, J.F.; Halpern, A.L.; Rusch, D.; Eisen, J.A.; Wu, D.; Paulsen, I.; Nelson, K.E.; Nelson, W. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **2004**, *304*, 66–74. [[CrossRef](#)] [[PubMed](#)]
23. Kimura, N. *Novel Biological Resources Screened from Uncultured Bacteria by a Metagenomic Method*; Elsevier Inc.: Amsterdam, The Netherlands, 2018; ISBN 9780128134030.
24. Lam, K.N.; Cheng, J.; Engel, K.; Neufeld, J.D.; Charles, T.C. Current and future resources for functional metagenomics. *Front. Microbiol.* **2015**, *6*, 1196. [[CrossRef](#)] [[PubMed](#)]
25. Escobar-Zepeda, A.; Vera-Ponce de León, A.; Sanchez-Flores, A. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Front. Genet.* **2015**, *6*, 348. [[CrossRef](#)] [[PubMed](#)]

26. Stepanauskas, R. Single cell genomics: An individual look at microbes. *Curr. Opin. Microbiol.* **2012**, *15*, 613–620. [[CrossRef](#)]
27. Blainey, P.C. The future is now: Single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* **2013**, *37*, 407–427. [[CrossRef](#)]
28. Blainey, P.C.; Quake, S.R. Dissecting genomic diversity, one cell at a time. *Nat. Methods* **2014**, *11*, 19–21. [[CrossRef](#)]
29. Nováková, J.; Farkašovský, M. Bioprospecting microbial metagenome for natural products. *Biologia* **2013**, *68*, 1079–1086. [[CrossRef](#)]
30. Dias, R.; Silva, L.C.F.; Eller, M.R.; Oliveira, V.M.; De Paula, S.O.; Silva, C.C. Metagenomics: Library construction and screening methods. *V Metagenom. Methods Appl. Perspect.* **2014**, *5*, 28–34.
31. Benbelgacem, F.F.; Salleh, H.M.; Noorbach, I.A. Construction of Metagenomic DNA Libraries and Enrichment Strategies. In *Multifaceted Protocol in Biotechnology*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 23–42.
32. Craig, J.W.; Chang, F.-Y.; Kim, J.H.; Obiajulu, S.C.; Brady, S.F. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* **2010**, *76*, 1633–1641. [[CrossRef](#)]
33. Iqbal, H.A.; Low-Beinart, L.; Obiajulu, J.U.; Brady, S.F. Natural Product Discovery through Improved Functional Metagenomics in *Streptomyces*. *J. Am. Chem. Soc.* **2016**, *138*, 9341–9344. [[CrossRef](#)] [[PubMed](#)]
34. Loganathachetti, D.S.; Muthuraman, S. Biomedical potential of natural products derived through metagenomic approaches. *RSC Adv.* **2015**, *5*, 101200–101213. [[CrossRef](#)]
35. Gaida, S.M.; Sandoval, N.R.; Nicolaou, S.A.; Chen, Y.; Venkataramanan, K.P.; Papoutsakis, E.T. Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nat. Commun.* **2015**, *6*, 7045. [[CrossRef](#)]
36. Martinez, A.; Kolvek, S.J.; Yip, C.L.T.; Hopke, J.; Brown, K.A.; MacNeil, I.A.; Osburne, M.S. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* **2004**, *70*, 2452–2463. [[CrossRef](#)]
37. Craig, J.W.; Chang, F.-Y.; Brady, S.F. Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem. Biol.* **2009**, *4*, 23–28. [[CrossRef](#)]
38. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669–685. [[CrossRef](#)]
39. Mirete, S.; Morgante, V.; González-Pastor, J.E. Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.* **2016**, *38*, 143–149. [[CrossRef](#)]
40. Feng, Z.; Chakraborty, D.; Dewell, S.B.; Reddy, B.V.B.; Brady, S.F. Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J. Am. Chem. Soc.* **2012**, *134*, 2981–2987. [[CrossRef](#)]
41. Lim, H.K.; Chung, E.J.; Kim, J.-C.; Choi, G.J.; Jang, K.S.; Chung, Y.R.; Cho, K.Y.; Lee, S.-W. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* **2005**, *71*, 7768–7777. [[CrossRef](#)]
42. Wang, G.-Y.-S.; Graziani, E.; Waters, B.; Pan, W.; Li, X.; McDermott, J.; Meurer, G.; Saxena, G.; Andersen, R.J.; Davies, J. Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2000**, *2*, 2401–2404. [[CrossRef](#)]
43. Gillespie, D.E.; Brady, S.F.; Bettermann, A.D.; Cianciotto, N.P.; Liles, M.R.; Rondon, M.R.; Clardy, J.; Goodman, R.M.; Handelsman, J. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* **2002**, *68*, 4301–4306. [[CrossRef](#)]
44. Blunt, J.W.; Carroll, A.R.; Copp, B.R.; Davis, R.A.; Keyzers, R.A.; Prinsep, M.R. Marine natural products. *Nat. Prod. Rep.* **2018**, *35*, 8–53. [[CrossRef](#)] [[PubMed](#)]
45. Brady, S.F.; Chao, C.J.; Clardy, J. Long-chain N-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* **2004**, *70*, 6865–6870. [[CrossRef](#)]
46. Brady, S.F.; Clardy, J. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J. Nat. Prod.* **2004**, *67*, 1283–1286. [[CrossRef](#)] [[PubMed](#)]
47. Williamson, L.L.; Borlee, B.R.; Schloss, P.D.; Guan, C.; Allen, H.K.; Handelsman, J. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl. Environ. Microbiol.* **2005**, *71*, 6335–6344. [[CrossRef](#)]
48. Uchiyama, T.; Abe, T.; Ikemura, T.; Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat. Biotechnol.* **2005**, *23*, 88–93. [[CrossRef](#)]
49. Uchiyama, T.; Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl. Environ. Microbiol.* **2010**, *76*, 7029–7035. [[CrossRef](#)]
50. Podar, M.; Abulencia, C.B.; Walcher, M.; Hutchison, D.; Zengler, K.; Garcia, J.A.; Holland, T.; Cotton, D.; Hauser, L.; Keller, M. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* **2007**, *73*, 3205–3214. [[CrossRef](#)]
51. Nasuno, E.; Kimura, N.; Fujita, M.J.; Nakatsu, C.H.; Kamagata, Y.; Hanada, S. Phylogenetically novel LuxI/LuxR-type quorum sensing systems isolated using a metagenomic approach. *Appl. Environ. Microbiol.* **2012**, *78*, 8067–8074. [[CrossRef](#)]
52. Guan, C.; Ju, J.; Borlee, B.R.; Williamson, L.L.; Shen, B.; Raffa, K.F.; Handelsman, J. Signal mimics derived from a metagenomic analysis of the gypsy moth gut microbiota. *Appl. Environ. Microbiol.* **2007**, *73*, 3669–3676. [[CrossRef](#)]
53. Meier, M.J.; Paterson, E.S.; Lambert, I.B. Metagenomic analysis of an aromatic hydrocarbon contaminated soil using substrate-induced gene expression. *Appl. Environ. Microbiol.* **2015**. [[CrossRef](#)]

54. Ufarté, L.; Potocki-Veronese, G.; Laville, E. Discovery of new protein families and functions: New challenges in functional metagenomics for biotechnologies and microbial ecology. *Front. Microbiol.* **2015**, *6*, 563. [[CrossRef](#)]
55. Colin, P.-Y.; Kintsjes, B.; Gielen, F.; Miton, C.M.; Fischer, G.; Mohamed, M.F.; Hyvönen, M.; Morgavi, D.P.; Janssen, D.B.; Hollfelder, F. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* **2015**, *6*, 10008. [[CrossRef](#)] [[PubMed](#)]
56. Scanlon, T.C.; Dostal, S.M.; Griswold, K.E. A high-throughput screen for antibiotic drug discovery. *Biotechnol. Bioeng.* **2014**, *111*, 232–243. [[CrossRef](#)] [[PubMed](#)]
57. Warnecke, F.; Hugenholtz, P. Building on basic metagenomics with complementary technologies. *Genome Biol.* **2007**, *8*, 231. [[CrossRef](#)] [[PubMed](#)]
58. Hosokawa, M.; Hoshino, Y.; Nishikawa, Y.; Hirose, T.; Yoon, D.H.; Mori, T.; Sekiguchi, T.; Shoji, S.; Takeyama, H. Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosens. Bioelectron.* **2015**, *67*, 379–385. [[CrossRef](#)] [[PubMed](#)]
59. Jiang, C.-J.; Chen, G.; Huang, J.; Huang, Q.; Jin, K.; Shen, P.-H.; Li, J.-F.; Wu, B. A novel β -glucosidase with lipolytic activity from a soil metagenome. *Folia Microbiol.* **2011**, *56*, 563–570. [[CrossRef](#)]
60. Jiménez, D.J.; Montana, J.S.; Alvarez, D.; Baena, S. A novel cold active esterase derived from Colombian high Andean forest soil metagenome. *World J. Microbiol. Biotechnol.* **2012**, *28*, 361–370. [[CrossRef](#)]
61. Bunterngsook, B.; Kanokratana, P.; Thongaram, T.; Tanapongpipat, S.; Uengwetwanit, T.; Rachdawong, S.; Vichitsoonthonkul, T.; Eurwilachitr, L. Identification and characterization of lipolytic enzymes from a peat-swamp forest soil metagenome. *Biosci. Biotechnol. Biochem.* **2010**, *74*, 1848–1854. [[CrossRef](#)]
62. Brady, S.F.; Clardy, J. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew. Chem. Int. Ed.* **2005**, *44*, 7063–7065.
63. Long, P.F.; Dunlap, W.C.; Battershill, C.N.; Jaspars, M. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *ChemBioChem* **2005**, *6*, 1760–1765. [[CrossRef](#)]
64. Chung, E.J.; Lim, H.K.; Kim, J.-C.; Choi, G.J.; Park, E.J.; Lee, M.H.; Chung, Y.R.; Lee, S.-W. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Appl. Environ. Microbiol.* **2008**, *74*, 723–730. [[CrossRef](#)] [[PubMed](#)]
65. Curtis, T.P.; Sloan, W.T.; Scannell, J.W. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10494–10499. [[CrossRef](#)]
66. Lambalot, R.H.; Gehring, A.M.; Flugel, R.S.; Zuber, P.; LaCelle, M.; Marahiel, M.A.; Reid, R.; Khosla, C.; Walsh, C.T. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **1996**, *3*, 923–936. [[CrossRef](#)]
67. Pfeifer, B.A.; Admiraal, S.J.; Gramajo, H.; Cane, D.E.; Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **2001**, *291*, 1790–1792. [[CrossRef](#)]
68. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Brady, S.F. Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3757–3762. [[CrossRef](#)] [[PubMed](#)]
69. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Guimarães, D.O.; de Frias, U.A.; Pupo, M.T.; Seepe, P.; Feng, Z.; Brady, S.F. Global biogeographic sampling of bacterial secondary metabolism. *Elife* **2015**, *4*, e05048. [[CrossRef](#)] [[PubMed](#)]
70. Reddy, B.V.B.; Kallifidas, D.; Kim, J.H.; Charlop-Powers, Z.; Feng, Z.; Brady, S.F. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* **2012**, *78*, 3744–3752. [[CrossRef](#)]
71. Reddy, B.V.B.; Milshteyn, A.; Charlop-powers, Z.; Brady, S.F. Resource eSNaPD: A Versatile, Web-Based Bioinformatics Platform for Surveying and Mining Natural Product Biosynthetic Diversity from Metagenomes. *Chem. Biol.* **2014**, *21*, 1023–1033. [[CrossRef](#)]
72. Wu, C.; Shang, Z.; Lemetre, C.; Ternei, M.A.; Brady, S.F. Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria. *J. Am. Chem. Soc.* **2019**, *141*, 3910–3919. [[CrossRef](#)]
73. Stein, J.L.; Marsh, T.L.; Wu, K.Y.; Shizuya, H.; DeLong, E.F. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **1996**, *178*, 591–599. [[CrossRef](#)]
74. Weinstock, G.M. Genomic approaches to studying the human microbiota. *Nature* **2012**, *489*, 250–256. [[CrossRef](#)] [[PubMed](#)]
75. Loman, N.J.; Misra, R.V.; Dallman, T.J.; Constantinidou, C.; Gharbia, S.E.; Wain, J.; Pallen, M.J. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **2012**, *30*, 434–439. [[CrossRef](#)] [[PubMed](#)]
76. Hildebrand, M.; Waggoner, L.E.; Liu, H.; Sudek, S.; Allen, S.; Anderson, C.; Sherman, D.H.; Haygood, M. bryA: An unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. *Chem. Biol.* **2004**, *11*, 1543–1552. [[CrossRef](#)] [[PubMed](#)]
77. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14002–14007. [[CrossRef](#)] [[PubMed](#)]
78. Kačar, D.; Schleissner, C.; Cañedo, L.M.; Rodríguez, P.; de la Calle, F.; Galán, B.; García, J.L. Genome of *Labrenzia* sp. PHM005 reveals a complete and active trans-AT PKS gene cluster for the biosynthesis of labrenzin. *Front. Microbiol.* **2019**, *10*, 2561. [[CrossRef](#)] [[PubMed](#)]
79. Piel, J.; Hui, D.; Wen, G.; Butzke, D.; Platzer, M.; Fusetani, N.; Matsunaga, S. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16222–16227. [[CrossRef](#)]
80. Ziemert, N.; Podell, S.; Penn, K.; Badger, J.H.; Allen, E.; Jensen, P.R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **2012**, *7*, e34064. [[CrossRef](#)]

81. Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Charlop-Powers, Z.; Calle, P.Y.; Kim, J.H.; Brady, S.F. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11797–11802. [[CrossRef](#)]
82. Owen, J.G.; Charlop-Powers, Z.; Smith, A.G.; Ternei, M.A.; Calle, P.Y.; Reddy, B.V.B.; Montiel, D.; Brady, S.F. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4221–4226. [[CrossRef](#)]
83. Kang, H.; Brady, S.F. Arimetamycin A: Improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes. *Angew. Chem. Int. Ed.* **2013**, *52*, 11063–11067. [[CrossRef](#)]
84. Chang, F.-Y.; Ternei, M.A.; Calle, P.Y.; Brady, S.F. Targeted metagenomics: Finding rare tryptophan dimer natural products in the environment. *J. Am. Chem. Soc.* **2015**, *137*, 6044–6052. [[CrossRef](#)] [[PubMed](#)]
85. Zotchev, S.B. Marine actinomycetes as an emerging resource for the drug development pipelines. *J. Biotechnol.* **2012**, *158*, 168–175. [[CrossRef](#)] [[PubMed](#)]
86. Niu, G. Genomics-driven natural product discovery in actinomycetes. *Trends Biotechnol.* **2018**, *36*, 238–241. [[CrossRef](#)]
87. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257. [[CrossRef](#)]
88. Hover, B.M.; Kim, S.-H.; Katz, M.; Charlop-Powers, Z.; Owen, J.G.; Ternei, M.A.; Maniko, J.; Estrela, A.B.; Molina, H.; Park, S. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat. Microbiol.* **2018**, *3*, 415–422. [[CrossRef](#)] [[PubMed](#)]
89. Wrighton, K.H. Discovering antibiotics through soil metagenomics. *Nat. Rev. Drug Discov.* **2018**, *17*, 241. [[CrossRef](#)]
90. Nayfach, S.; Pollard, K.S. Toward accurate and quantitative comparative metagenomics. *Cell* **2016**, *166*, 1103–1116. [[CrossRef](#)]
91. Abbasi, M.N.; Fu, J.; Bian, X.; Wang, H.; Zhang, Y.; Li, A. Recombineering for Genetic Engineering of Natural Product Biosynthetic Pathways. *Trends Biotechnol.* **2020**, *38*, 715–728. [[CrossRef](#)]
92. Mardanov, A.V.; Kadnikov, V.V.; Ravin, N.V. Metagenomics: A paradigm shift in microbiology. In *Metagenomics*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 1–13.
93. Ottesen, E.A.; Hong, J.W.; Quake, S.R.; Leadbetter, J.R. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **2006**, *314*, 1464–1467. [[CrossRef](#)]
94. Binga, E.K.; Lasken, R.S.; Neufeld, J.D. Something from (almost) nothing: The impact of multiple displacement amplification on microbial ecology. *ISME J.* **2008**, *2*, 233–241. [[CrossRef](#)] [[PubMed](#)]
95. Hochmuth, T.; Niederkrüger, H.; Gernert, C.; Siegl, A.; Taudien, S.; Platzer, M.; Crews, P.; Hentschel, U.; Piel, J. Linking chemical and microbial diversity in marine sponges: Possible role for Poribacteria as producers of methyl-branched fatty acids. *ChemBioChem* **2010**, *11*, 2572–2578. [[CrossRef](#)] [[PubMed](#)]
96. Mori, T.; Cahn, J.K.B.; Wilson, M.C.; Meoded, R.A.; Wiebach, V.; Martinez, A.F.C.; Helfrich, E.J.N.; Albersmeier, A.; Wibberg, D.; Dätwyler, S. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated Enttheonella sponge symbionts. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1718–1723. [[CrossRef](#)] [[PubMed](#)]
97. Siegl, A.; Hentschel, U. PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ. Microbiol. Rep.* **2010**, *2*, 507–513. [[CrossRef](#)]
98. Siegl, A.; Kamke, J.; Hochmuth, T.; Piel, J.; Richter, M.; Liang, C.; Dandekar, T.; Hentschel, U. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* **2011**, *5*, 61–70. [[CrossRef](#)]
99. Lu, S.; Zong, C.; Fan, W.; Yang, M.; Li, J.; Chapman, A.R.; Zhu, P.; Hu, X.; Xu, L.; Yan, L. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **2012**, *338*, 1627–1630. [[CrossRef](#)]
100. Sha, Y.; Sha, Y.; Ji, Z.; Ding, L.; Zhang, Q.; Ouyang, H.; Lin, S.; Wang, X.; Shao, L.; Shi, C. Comprehensive Genome Profiling of Single Sperm Cells by Multiple Annealing and Looping-Based Amplification Cycles and Next-Generation Sequencing from Carriers of Robertsonian Translocation. *Ann. Hum. Genet.* **2017**, *81*, 91–97. [[CrossRef](#)]
101. Lasken, R.S. Single-cell genomic sequencing using multiple displacement amplification. *Curr. Opin. Microbiol.* **2007**, *10*, 510–516. [[CrossRef](#)]
102. Piel, J.; Cahn, J. Opening up the Single-Cell Toolbox for Microbial Natural Products Research. *Angew. Chem. Int. Ed.* **2019**. [[CrossRef](#)]
103. Fieseler, L.; Horn, M.; Wagner, M.; Hentschel, U. Discovery of the novel candidate phylum “Poribacteria” in marine sponges. *Appl. Environ. Microbiol.* **2004**, *70*, 3724–3732. [[CrossRef](#)]
104. Woyke, T.; Tighe, D.; Mavromatis, K.; Clum, A.; Copeland, A.; Schackwitz, W.; Lapidus, A.; Wu, D.; McCutcheon, J.P.; McDonald, B.R. One bacterial cell, one complete genome. *PLoS ONE* **2010**, *5*, e10314. [[CrossRef](#)]
105. Rinke, C.; Schwientek, P.; Sczyrba, A.; Ivanova, N.N.; Anderson, I.J.; Cheng, J.-F.; Darling, A.; Malfatti, S.; Swan, B.K.; Gies, E.A. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **2013**, *499*, 431–437. [[CrossRef](#)] [[PubMed](#)]
106. Rinke, C. Single-cell genomics of microbial dark matter. In *Microbiome Analysis*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 99–111.
107. Wang, L.; Ravichandran, V.; Yin, Y.; Yin, J.; Zhang, Y. Natural products from mammalian gut microbiota. *Trends Biotechnol.* **2019**, *37*, 492–504. [[CrossRef](#)] [[PubMed](#)]
108. Lepage, P.; Leclerc, M.C.; Joossens, M.; Mondot, S.; Blottière, H.M.; Raes, J.; Ehrlich, D.; Doré, J. A metagenomic insight into our gut’s microbiome. *Gut* **2013**, *62*, 146–158. [[CrossRef](#)] [[PubMed](#)]

109. Sudarikov, K.; Tyakht, A.; Alexeev, D. Methods for the metagenomic data visualization and analysis. *Curr. Issues Mol. Biol.* **2017**, *24*, 37–58. [[CrossRef](#)] [[PubMed](#)]
110. Thompson, L.R.; Sanders, J.G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K.J.; Prill, R.J.; Tripathi, A.; Gibbons, S.M.; Ackermann, G. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **2017**, *551*, 457–463. [[CrossRef](#)]
111. Bitok, J.K.; Lemetre, C.; Ternei, M.A.; Brady, S.F. Identification of biosynthetic gene clusters from metagenomic libraries using PPTase complementation in a *Streptomyces* host. *FEMS Microbiol. Lett.* **2017**, *364*, 1–8. [[CrossRef](#)]
112. Santana-Pereira, A.L.R.; Sandoval-Powers, M.; Monsma, S.; Zhou, J.; Santos, S.R.; Mead, D.A.; Liles, M.R. Discovery of novel biosynthetic gene cluster diversity from a soil metagenomic library. *Front. Microbiol.* **2020**, *11*, 585398. [[CrossRef](#)]
113. Ghurye, J.S.; Cepeda-Espinoza, V.; Pop, M. Metagenomic Assembly: Overview, Challenges and Applications. *Yale J. Biol. Med.* **2016**, *89*, 353–362.
114. Ling, L.L.; Schneider, T.; Peoples, A.J.; Spoering, A.L.; Engels, I.; Conlon, B.P.; Mueller, A.; Schäberle, T.F.; Hughes, D.E.; Epstein, S. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517*, 455–459. [[CrossRef](#)]
115. Bhushan, A.; Egli, P.J.; Peters, E.E.; Freeman, M.F.; Piel, J. Genome mining-and synthetic biology-enabled production of hypermodified peptides. *Nat. Chem.* **2019**, *11*, 931–939. [[CrossRef](#)]