Genetic platforms for heterologous expression of microbial natural products

Journal:	*Natural Product Reports*
Manuscript ID	NP-REV-04-2019-000025.R1
Article Type:	Review Article
Date Submitted by the Author:	03-Jun-2019
Complete List of Authors:	Zhang, Jia Jia; University of California San Diego, Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography Tang, Xiaoyu; University of California San Diego, Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography Moore, Bradley; University of California San Diego, Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography; Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego

Genetic platforms for heterologous expression of microbial natural products

Jia Jia Zhang,[a*] Xiaoyu Tang,[a] and Bradley S. Moore[a,b*]

[a]*Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA.*
[b]*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA.*
**Email: jzhang@microbechembio.org; bsmoore@ucsd.edu*

Natural products are of paramount importance in human medicine. Not only are most antibacterial and anticancer drugs derived directly from or inspired by natural products, many other branches of medicine, such as immunology, neurology, and cardiology, have similarly benefited from natural product-based drugs. Typically, the genetic material required to synthesize a microbial specialized product is arranged in a multigene biosynthetic gene cluster (BGC), which codes for proteins associated with molecule construction, regulation, and transport. The ability to connect natural product compounds to BGCs and vice versa, along with ever-increasing knowledge of biosynthetic machineries, has spawned the field of genomics-guided natural product genome mining for the rational discovery of new chemical entities. One significant challenge in the field of natural product genome mining is how to rapidly link orphan biosynthetic genes to their associated chemical products. This review highlights state-of-the-art genetic platforms to identify, interrogate, and engineer BGCs from diverse microbial sources, which can be broken into three stages: (1) cloning and isolation of genomic loci, (2) heterologous expression in a host organism, and (3) genetic manipulation of cloned pathways. In the future, we envision natural product genome mining will be rapidly accelerated by *de novo* DNA synthesis and refactoring of whole biosynthetic pathways in combination with systematic heterologous expression methodologies.

1. Introduction

Natural products, or specialized small molecules produced by living organisms, have long fascinated chemists due to their complex three-dimensional structures, which make them challenging to produce using synthetic organic chemistry.[1] Perhaps more importantly, structural complexity and multidimensionality confer many natural products with potent and specific biological activities, making them privileged scaffolds in the quest to develop new medicines.[2] Natural products have played an invaluable role in deepening our understanding of various cellular processes and even facilitated the discovery of important and evolutionarily conserved macromolecules, such as the protein kinase mTOR, or mechanistic target of rapamycin, which is blocked by the *Streptomyces* natural product rapamycin.[3] Rapamycin has become a life-saving immunosuppressant drug,

and the discovery of mTOR spawned a vibrant field of study uncovering its central role in physiology, metabolism, aging, and common diseases such as cancer and epilepsy.[4]

The ability of all living organisms to biosynthesize endogenous, specialized small molecules is genetically encoded. Making connections between isolated small molecules and the genes responsible for their construction has been particularly productive in microorganisms, which generally cluster elements involved in natural product biosynthesis along their genomes in “biosynthetic gene clusters” (BGCs).

Clustering of specialized genetic elements carries an additional benefit: the ability to clone and transfer whole BGCs to heterologous host organisms for expression and characterization. Heterologous expression of BGCs for characterization or identification of new chemical entities is advantageous for several reasons. As more and more genome sequences from diverse microbial sources become available, heterologous expression circumvents the need to develop new genetic tools to interrogate pathways from each new genus or species of interest. Furthermore, it enables characterization of BGCs from microbes that have yet to be cultured such as those identified from obligate symbionts or environmental DNA (eDNA). Successful heterologous reconstitution of a BGC allows rapid delineation of all essential specialized genes involved in the production of a microbial bio-chemical. Finally, genetic platforms for interrogation of BGCs that are developed can be optimized and universally applied, both for heterologous expression as well as rapid genetic manipulation of cloned pathways to perform biosynthetic investigations or BGC refactoring.

In this article, we review genetic platforms that have been established for heterologous expression of microbial natural products. Integrated platforms include three stages: 1) cloning and isolation of selective genomic loci containing BGCs, 2) expression in a heterologous host organism, and 3) genetic manipulation of cloned pathways for interrogation or activation (Fig. 1). Following this workflow, we highlight advancements in cloning methods, heterologous hosts, and genetic manipulation of large microbial BGCs used for successful heterologous reconstitution of various natural products (Fig. 2). Lastly, we highlight recent studies that utilize BGC refactoring, DNA synthesis, and *de novo* pathway design to access new chemical entities.

2. Cloning of microbial BGCs

Due to the large size, repetitive nature, and high GC-content of many microbial BGCs, cloning has remained a challenging step in heterologous reconstitution of natural product pathways. While advancements in genome sequencing, bioinformatics, and molecular biology techniques have changed the landscape of BGC cloning over the years, many different types of cloning methods, including library-based methods, assembly methods, and direct cloning methods, have been developed and continue to be used successfully today. In this section, we highlight these techniques and their associated advantages and limitations (Table 1).

2.1 Library-based methods

Early efforts to clone microbial BGCs relied heavily on library-based methods, which involve the generation of a clone library of random genomic DNA (gDNA) fragments in *Escherichia coli.* Library generation is particularly useful when complete genome sequence information is lacking, or when it is advantageous to catalogue the complete genetic material from an organism or metagenomic sample. Furthermore, library generation breaks the chromosome into smaller chunks that can be more easily sequenced and assembled than whole chromosomes and has therefore been a vital component of numerous genome sequencing efforts. Many BGCs have been cloned from cultured organisms into cosmid[5-9] and fosmid[10-14] libraries, which hold inserts of approximately 40 kb in size and are packaged and delivered to *E. coli* by bacteriophages. While these libraries hold similar sized inserts, fosmids exist at low or single copy number in *E. coli* cloning hosts and are therefore considered more stable than cosmids, particularly for highly repetitive DNA. Brady and colleagues have pioneered the generation of cosmid libraries for discovery and characterization of BGCs from soil metagenomic DNA,[15-21] thus accessing pathways from organisms that have yet to be cultured. Cosmid libraries from environmental DNA (eDNA) can be maintained and continually re-screened for new categories of specialized biosynthetic genes.[16, 19-21] Cloning of metagenomic DNA is particularly challenging due to the low enrichment and potentially low quality of genetic material from an individual microbe within a heterogeneous population, but library cloning has proven to be a successful approach for accessing BGCs from eDNA. However, BGCs

cloned using these methods are often split across multiple library clones and thus need to be stitched together and trimmed before subsequent use.[8, 10, 19, 21] Analysis of 540 full BGCs deposited in the MIBiG database[22] as of August 2018 revealed that these pathways range in size from 204 bp (representing a small ribosomally encoded peptide) to 148,229 bp, with an average of approximately 36 kb. Although this is below the 40 kb threshold, it is likely that an average-sized BGC will be split across multiple fosmid or cosmid clones.

Library vectors that hold larger inserts include those based on bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs). Unlike cosmid and fosmid libraries, BAC and PAC libraries are prepared by direct transfer of DNA to *E. coli* via electroporation. High molecular weight DNA can be isolated using special techniques to promote uniform insert sizes above 100 kb, which are stably maintained within low copy number vector backbones. Several recent studies have utilized BAC library clones to characterize large BGCs (>55 kb) encoding assembly line biosynthetic pathways, which include modular polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters.[23-27] The quinolidomicin A_1 (**1**) BGC, which encodes a PKS and spans over 200 kb in size, was very recently cloned in a BAC library and represents the largest BGC cloned and heterologously expressed to-date.[28] However, uniform, high molecular weight DNA can be difficult to isolate, and low copy number plasmids can also be more challenging to work with. Because of these technical difficulties, research groups are increasingly outsourcing large-insert library generation to companies. Several reports of large BGC cloning have leveraged PAC libraries generated by the Canadian company Bio S&T using vector pESAC13, which can be readily transferred to various *Streptomyces* heterologous hosts by conjugation.[29-33]

Recently, a group of researchers studying fungal secondary metabolism retrofitted a BAC vector backbone with an autonomously replicating sequence from *Aspergillus* (AMA1) to generate a self-replicating fungal artificial chromosome (FAC) that can be used for library generation.[34] FAC libraries generated using this vector can be directly transferred to the host *Aspergillus nidulans* for heterologous expression and mass spectrometry screening for identification of new chemical entities. Libraries were constructed and screened from the gDNA of *A. terrus*, *A. aculeatus*, and *A. wentii* and resulted in the identification of 15 unique mass signals produced by cryptic biosynthetic machinery, a few of which have been characterized in greater detail, including benzomalvin A/D (**2**) and acu-dioxomorpholines A (**3a**) and B (**3b**).[35-37]

Library-based cloning methods have contributed greatly to our understanding of microbial BGCs and continue to be used routinely and successfully. However, library cloning is an untargeted approach in which most library clones do not contain genomic regions of interest and thus must be extensively screened. With the proliferation of publicly available genome sequence information, more targeted approaches, such as assembly and direct cloning methods, are being utilized with greater frequency.

2.2 Assembly methods

Methods that rely on *in vitro* assembly of BGCs from smaller fragments represent an attractive alternative to library-based approaches, as small fragments of DNA can be generated quickly and cheaply using polymerase chain reaction (PCR) and are easy to work with. In 2009, Gibson and co-workers reported the *in vitro* assembly of very large DNA molecules leveraging the concerted action of three enzymes – a 5' exonuclease, a DNA polymerase, and a DNA ligase – in an isothermal, single-reaction method.[38] Gibson assembly kits can be purchased commercially and are used ubiquitously in molecular biology. Within the field of natural products research, Gibson assembly has been successfully used to reconstruct small BGCs, usually around 10 kb in length.[39-43] Direct pathway cloning, or DiPaC, which relies on Gibson assembly, has been successfully used to clone small but also larger (20 to 55 kb) BGCs in a stepwise, combinatorial fashion for heterologous production of compounds such as hapalosin (**4**) and sodorifen (**5**).[44-46] Alternatively, a method called twin-primer assembly, or TPA, is designed for assembly of PCR-amplified fragments without the use of enzymes.[47] TPA relies of annealing of complementary single-stranded overhangs designed into PCR primer sequences and has been used to assemble a 31 kb plasmid at approximately 50% fidelity.[47]

Phage recombinases have also been used to assemble BGCs. Also referred to as integrases, these enzymes catalyze recombination across relatively short recognition sequences, often referred to as *attP* and *attB* sites, to generate new *attL* and *attR* sequence junctions. In this way, DNA fragments can be stitched together in a

specific order and orientation into a self-replicating construct and selected for using selectable markers. Serine integrase recombinational assembly, or SIRA, utilizes purified bacteriophage integrases ΦC31 and Bxb1 as well as their recombination directionality factors (RDFs) to assemble multiple DNA fragments into functional plasmids.[48, 49] Site-specific recombination-based tandem assembly, or SSRTA, is a similar approach that leverages other serine integrases, including ΦBT1, TG1, and ΦRv1 in addition to Bxb1.[50, 51] However, when directly compared, Bxb1 displayed the highest *in vitro* assembly efficiency.[52] Homing endonucleases, which recognize long asymmetric sequences that are therefore rarely present in natural DNA, have also been used for the iterative assembly of standardized DNA parts through a cut-and-paste mechanism similar to standard restriction-digestion and ligation cloning in a method called iBrick.[53] Recently, programmable DNA-guided restriction enzymes have been developed leveraging the Argonaute enzyme from *Pyrococcus furiosus*, which enables the cleavage of virtually any DNA sequence and generation of defined sticky ends for facile DNA assembly.[54] These artificial restriction enzymes, or AREs, are easily programmable and will likely find broad utility in biological research.

In vivo DNA assembly methods have been widely used to reconstruct large microbial BGCs. DNA assembly leveraging high rates of *in vivo* homologous recombination in yeast has been shown to be accurate for assembly of up to 10 fragments of DNA, which is particularly useful for large and high GC-content gene clusters that can only be amplified in 2-5 kb pieces.[55-59] Several named methods, including DNA assembler,[60-67] overlap extension PCR-yeast homologous recombination (ExRec),[68] and yeast recombinational cloning-enabled pathway transfer and expression tool (yTREX),[69] all rely on homologous recombination in yeast for assembly of multiple fragments of DNA into functional plasmids for heterologous expression of natural product BGCs. Although not as naturally recombinant as yeast, *E. coli* strains equipped with Red/ET recombineering machinery have also been used for *in vivo* assembly of BGCs in a method called artificial gene operon assembly system, or AGOS.[70]

While assembly methods generally require the least technical expertise and have the greatest potential for automation, they are often not feasible or cumbersome for large BGCs and present the greatest risk for introducing mutations into cloned pathways. Direct cloning methods, in which whole BGCs are directly targeted for cloning, represent the most elegant approach to obtaining BGCs for heterologous expression and are becoming increasingly user friendly.

2.3 Direct cloning methods

Yeast homologous recombination, in addition to catalyzing multi-part DNA assembly, has been extensively leveraged for direct cloning of whole BGCs with pre-defined boundaries. Transformation-associated recombination (TAR) cloning in *Saccharomyces cerevisiae* can be used for the selective isolation of any genomic fragment into a circular yeast artificial chromosome (YAC), which can propagate, segregate, and be selected for in yeast.[71] YAC clones arise from homologous recombination between gDNA fragments, which can be prepared by random shearing or enzymatic digestion, and a linearized TAR cloning vector containing two targeting hooks with sequences identical to those flanking the genomic loci of interest. Following successful cloning in yeast, constructs are shuttled to *E. coli* for detailed characterization and verification by PCR, sequencing, and/or restriction digestion. For the purposes of cloning BGCs for heterologous expression, TAR cloning vectors can also include elements for transfer to and maintenance in a heterologous host organism. Moore and colleagues developed first-generation TAR cloning vectors, pCAP01[72] and pCAPB02,[73] which are yeast-*E. coli*-*Streptomyces* and yeast-*E. coli*-*Bacillus* shuttle vectors, respectively, and can be assembled into a cluster-specific capture vector by addition of two long homology arms of ~1 kb each. Many bacterial BGCs have been directly cloned and interrogated using these first-generation TAR vectors.[73-82] However, pCAP vector backbones possess yeast origins of replication, which are essential for cloning bacterial DNA but also result in high rates of plasmid recircularization via non-homologous end joining (NHEJ) to greatly increase the number of empty vectors that must be screened against. By adapting a previously established method utilizing counterselection to select against plasmid recircularization via NHEJ,[83] two second-generation

vectors, pCAP03[84] and pCAP05,[55] were developed and used to clone and express bacterial BGCs.[85] These vectors also employ much shorter homology arms of 50 bp each, which simplifies the procedure for preparing cluster-specific capture vectors in addition to significantly improving the efficiency of the TAR cloning experiment by introducing a mechanism for counterselection.[84] Furthermore, pCAP05 is a broad-host-range expression vector for Gram-negative host organisms, thus expanding the host range of the pCAP vector series to include organisms such as *Pseudomonas* and *Agrobacterium*.[86]

TAR cloning is most efficient when homologous sequences are located as closely as possible to DNA ends,[87] perhaps because homologous recombination is a mechanism to repair DNA double-strand breaks (DSBs) that arise during mitosis. Thus, the success of TAR cloning is also greatly enhanced if restriction sites can be identified just beyond the boundaries of the BGC of interest and the targeting hooks are designed as closely as possible to these sites. Unfortunately, this may not be feasible for many BGCs due to a lack of available restriction sites associated with commercially available restriction enzymes that do not also cut within the BGC. Recently, Lee, Larionov, and Kouprina reported a method combining CRISPR/Cas9-mediated *in vitro* digestion of DNA with TAR cloning, which resulted in a dramatic increase in the fraction of positive clones.[88] Although this method has not yet been reported for TAR cloning of a microbial BGC for heterologous expression, the use of CRISPR/Cas9 for *in vitro* digestion of gDNA is generally applicable for a number of direct cloning methods and has already been used in combination with RecE catalyzed linear-linear homologous recombination (LLHR) as well as Gibson assembly (detailed below).

Reconstitution of Rac prophage enzymes RecE and RecT in *E. coli* enables *in vivo* homologous recombination of two linear DNA fragments in a method called LLHR,[89] which is highly analogous to TAR cloning in yeast. While arguably not as robust, LLHR is more attractive than TAR because it uses *E. coli* as a cloning host, which makes it faster and also eliminates a step compared to the TAR cloning process. LLHR has been pioneered as a method for BGC cloning by Zhang, Stewart, Müller and colleagues and has been applied for investigation of BGCs from various microbial sources.[89-93] Exonuclease combined with RecET recombination, or ExoCET, is based on the principle that direct cloning efficiencies can be improved by pre-annealing of linear vectors and target DNA using the exonuclease T4 DNA polymerase before delivery to *E. coli* cells.[94, 95] Furthermore, LLHR operates on the same principle as homologous recombination in yeast, wherein precise digestion of gDNA prior to cloning using either commercially available restriction enzymes or CRISPR/Cas9 programmed restriction digestion greatly enhances cloning efficiencies.[94]

Finally, various *in vitro* methods have been used for direct cloning of BGCs. Single-strand overlapping annealing, or SSOA, is an *in vitro* approach that, like ExoCET, also utilizes restriction digestion coupled with exonuclease treatment, but does not leverage any *in vivo* recombination machinery.[96] Direct cloning of large pathways from complex mixtures of gDNA using Gibson assembly has also been reported.[97-99] In work by Leadlay and colleagues on expression of the anticancer compound conglobatin (**6**), a precise 41 kb fragment of DNA was generated by gDNA digestion using restriction enzymes XhoI and EcoRI, and digested DNA was carefully purified by gel electrophoresis to remove fragments less than 20 kb in size.[97] Alternatively, CRISPR/Cas9 can be used for programmed DNA digestion in agarose gel plugs prior to Gibson assembly, as has been reported for Cas9 assisted targeting of chromosome segments, or CATCH, which has been successfully used for targeting up to 100 kb in a single step.[98, 99] Thus, like *in vivo* methods, *in vitro* direct cloning approaches greatly benefit from careful preparation of DNA through pre-treatment and purification.

If natural restriction sites are not present at the boundaries of a BGC of interest, researchers have also relied on genetic manipulation to introduce unique sites for restriction enzymes or homing endonucleases so that BGCs can be precisely targeted for digestion and self-ligation; in various reports, this process has been referred to as plasmid recovery,[100] plasmid rescue,[101] or iCATCH.[102] Editing of chromosomal sequences upstream and downstream of BGCs has also been used to introduce ΦBT1 or Cre-lox recombination sites to excise genomic loci via *in vitro* or *in vivo* recombination.[103-106] However, an important drawback to relying on genetic manipulation is that it is a labor-intensive approach that not all natural product producing organisms are amenable to, which may be the reason why cloning and heterologous expression were pursued for BGC characterization in the first place.

While still not trivial, cloning of large BGCs has become increasingly accessible, and many viable approaches have been developed and used to isolate DNA from complex genetic backgrounds. Following successful cloning, BGCs must be transferred to a heterologous host that can stably maintain and express the exogenous DNA. Furthermore, successful heterologous reconstitution also requires that the host is equipped with all biosynthetic building blocks and non-clustered enzymatic machinery or accessory factors essential for natural product production. In the next section, we highlight various host organisms that have been developed, optimized, and used for heterologous expression of natural product BGCs.

3. Hosts for heterologous expression

There are several obvious features of a good heterologous host organism: it grows fast, is genetically manipulatable, and is easy to work with in the laboratory. However, there are many additional traits that organisms must carry, either naturally or through engineering, to enable heterologous reconstitution of natural product BGCs. Some genera of host organisms have been developed and used more extensively than others, particularly *Streptomyces*, as this genus has been a naturally rich source of antibiotics and other small molecule natural products over the years. There is an underlying assumption, pervasive throughout the natural products research community, that BGCs are best expressed in organisms most closely related to their original source. Although this assumption is theoretically sound, it has not been rigorously proven, in part because it becomes completely irrelevant once a suitable host organism has been identified. Although it remains essentially impossible to predict host compatibility, educated guesses can be made, particularly as we understand more about biosynthetic mechanism and regulation. For all BGCs, there will be an assortment of additional elements that need to be “borrowed” from the host organism, ranging from biosynthetic precursors such as fatty acids, amino acids, and acyl-CoAs, to enzymes that post translationally modify biosynthetic enzymes such as phosphopantetheinyl transferases (PPTases), to regulatory genes that control BGC expression. These host elements must be compatible with exogenous BGCs or independently supplied from alternative sources. It has been shown that BGCs often move by horizontal gene transfer; thus, they are frequently not conserved throughout a taxonomic group.[107, 108] Conversely, this means that similar BGCs can naturally function in organisms that are not closely related phylogenetically. Regardless of whether a more closely related host is truly “superior”, it is ideal to have many different types of heterologous hosts available, particularly as we identify new BGCs from diverse microbial sources. Many studies have shown this empirically through successful expression of BGCs in some hosts but not others.[15, 58, 109-112]

Heterologous hosts can be engineered for enhanced expression of microbial BGCs in a background devoid of competing or contaminating pathways. General strategies for host optimization include genome minimization through deletion of native BGCs, nucleases, and proteases, as well as introduction of chromosomal integration elements, PPTases and other post translational modifying enzymes, or genes involved in precursor biosynthesis. In this section we review microbial heterologous host systems that have been developed, optimized, and used to successfully reconstitute natural product BGCs. Various bacterial and fungal genera are discussed (Table 2).

2.1 Actinobacteria

To date, actinobacterial hosts of the genus *Streptomyces* have been most widely used for BGC expression, primarily because many bioactive natural products have been isolated from *Streptomyces* and therefore the greatest attention has been paid to studying their secondary metabolism and developing optimized *Streptomyces* hosts. *Streptomyces* species that have been used as heterologous hosts include *S. coelicolor*,[113] *S. avermitilis*,[114, 115] *S. lividans*,[116] *S. albus*,[117, 118] *S. venezuelea*,[119] *S. ambofaciens* and *S. roseosporus*,[120] *S. flavogriseus*,[109, 121] *S. chattanoogensis*,[122] and *S. chartreusis*.[123] Additionally, fast-growing and moderately thermophilic strains of *Streptomyces* have been isolated and tested for heterologous expression, although these organisms have not been characterized at the strain level.[124] *Streptomyces* hosts have been optimized to various extents using several general strategies. Genome minimization through curing of self-replicating plasmids and deletion of non-essential genes, including native BGCs, has been widely applied. Deletion of endogenous BGCs has the two-fold effect of removing sinks for biosynthetic precursors and simultaneously

simplifying the host's chemical profile.[113-115, 117, 122] Additionally, empirically identified mutations in *rpoB* and *rpsL* (which encode the RNA polymerase β-subunit and ribosomal protein S12, respectively) of *S. coelicolor* and *S. lividans* pleiotropically enhance secondary metabolite production and have been introduced.[113] Remodeling of global regulatory circuits through introduction of new positive regulators or up- and down-regulation of native positive and negative regulators, respectively, has also been leveraged,[116, 118] in addition to expression of codon-optimized efflux pumps to reduce toxicity and aid in natural product purification.[116] Please note that the *Streptomyces* hosts and references listed in this section are not comprehensive, as in depth review of optimized *Streptomyces* hosts is covered elsewhere in this issue.

Long and often repetitive natural product BGCs must be stably maintained in the host, either as an autonomously replicating sequence or integrated into the host genome. Replicative expression plasmids such as pOJ446, which is an *E. coli-Streptomyces* shuttle vector based on the *S. coelicolor* A3(2) plasmid SCP2*, have been widely used.[125-127] However, chromosomal integration of BGCs is more stable and, if irreversible, circumvents the need to use antibiotics for plasmid maintenance. Many BGCs have been integrated into *Streptomyces* genomes leveraging recombinases from actinophages, most notably ΦC31.[128-130] The ΦC31 *attB* attachment site is naturally present in the genome of many *Streptomyces* species and enables site-specific integration of large BGCs into the host chromosome. Other actinophages, including TG1, SV1, ΦBT1, R4, ΦHau, and ΦJoe, have been identified, characterized, and in some cases used for BGC integration.[131-133] Having access to multiple attachment sites is helpful for performing complementation experiments to confirm that gene deletion mutants do not carry polar effects. Furthermore, BGCs can be split and integrated into multiple genomic loci to simplify cloning and assist in BGC engineering.[134, 135] The precise site of genomic integration could be important in a heterologous expression experiment, as it has recently been shown that gene expression can vary up to eight-fold depending on the position of chromosomal integration in *S. albus*.[136] Thus, having access to multiple attachment sites leveraging different actinophages can be a valuable asset for heterologous expression of BGCs in *Streptomyces*.

ΦC31 attachment sites are also naturally present in the genomes of several other Actinobacteria and have been leveraged for BGC expression in the rare actinomycete *Nonomuraea* sp. strain ATCC 39727[6, 137] and the nocardioform actinomycete *Amycolatopsis japonicum*.[138] Alternatively, the marine actinomycete *Salinispora tropica* CNB-440, which possesses several pseudo ΦC31 attachment sites, was engineered to introduce an authentic ΦC31 attachment site for integration of BGCs into the salinosporamide (*sal*) biosynthetic locus, simultaneously abolishing salinosporamide production to free up biosynthetic precursors.[139] The resulting strain, *S. tropica* CNB-4401, represents the first marine actinomycete heterologous host and is readily compatible with expression vectors containing the ΦC31 integrase. Finally, *Saccharopolyspora erythraea*, the original producer of the antibiotic erythromycin, was used as a host for heterologous expression of the spinosad BGC, encoding production of spinosyns A (**7a**) and D (**7b**), from *Saccharopolyspora spinosa* via chromosomal integration by double-crossover homologous recombination.[140]

2.2 Firmicutes

The rod-shaped, low G+C content Firmicute *Bacillus subtilis* has many desirable qualities of a heterologous host organism. It produces the cyclic lipopeptide surfactin and thus harbors the promiscuous PPTase Sfp, which has been used to activate carrier proteins from diverse microbial sources.[141] Thus, *B. subtilis* is naturally equipped with an important element for expression of assembly line BGCs. Furthermore, it has the capacity for natural genetic competence and homologous recombination, making it easy to transform and genetically manipulate, and it grows relatively fast, particularly in comparison to *Streptomyces*. Despite these attractive features, *Bacillus* has been used relatively infrequently for heterologous expression of BGCs, perhaps due to a natural reluctance to test non-*Bacillus* pathways in a *Bacillus* host organism. Consequently, there has not been the same level of investment in development and optimization of *Bacillus* heterologous hosts. *B. subtilis* and *B. amyloliquefaciens* have been used to express a number of BGCs, primarily those encoding peptide products originating from other *Bacilli*.[73, 82, 142-147] In most cases, gene clusters have been integrated into the *amyE* locus of *B. subtilis* hosts via double-crossover homologous recombination. Interestingly, Li *et al.* attempted to express the surfactin BGC in *B. subtilis* ROM77, in which the native surfactin locus is disrupted, from a self-replicating plasmid but encountered insurmountable plasmid instability, prompting the authors to build an integratable heterologous expression plasmid instead.[73] Beyond *Bacillus*, *Lactococcus lactis* is another

Firmicute that has been leveraged for heterologous expression of small bacteriocin BGCs responsible for production of lactococcin Z and plantaricyclin A, which are produced by other *Lactococci* and *Lactobacilli*, respectively.[148, 149]

2.3 Proteobacteria

Although the majority of isolated natural products have been sourced from Gram-positive Actinobacteria, Gram-negative organisms such as Proteobacteria and Cyanobacteria are increasingly being recognized as capable and underexplored producers of bioactive small molecules.[150] Gram-negative Proteobacteria include many dangerous pathogens such as *Salmonella*, *Vibrio*, and *Yersinia*, which are difficult to treat using standard antibiotics. Conversely, Proteobacteria are also important commensals found within the microbiomes of many higher organisms, including humans. Thus, heterologous expression of proteobacterial BGCs represents an important tool for studying secondary metabolism of microorganisms that greatly influence human health and disease.

The disparate GC-content between microorganisms has often been cited as a reason for host-BGC incompatibility, although robust evidence supporting this claim is lacking. More likely, host factors that must be borrowed for heterologous production, whether regulatory or biosynthetic, may be missing or incompatible between BGC and host. Currently, it is not clear whether Gram-positive hosts can robustly express Gram-negative BGCs or vice versa, as this has not been rigorously tested or reported. Regardless, many Gram-negative bacteria are easier to work with than *Streptomyces*, and thus expansion of heterologous expression platforms to include more proteobacterial hosts represents a valuable endeavor, particularly for those species that grow fast, are easy to manipulate, and are more naturally suited to natural product production than *E. coli*.

Of course, the most obvious proteobacterial expression host is the laboratory workhorse *E. coli*, which belongs to the class of γ-proteobacteria. *E. coli* is very amenable to laboratory manipulation and, under ideal conditions, has a doubling time of only 20 minutes. Unfortunately, commonly used strains of *E. coli* are not naturally equipped to support expression of many natural product BGCs for several reasons; in some cases, specific pitfalls have been addressed and overcome. The luminmycin A (**8a**) BGC from the γ-proteobacterium *Photorhabdus luminescens* and the glidobactin A (**8b**) BGC from the β-proteobacterium *Burkholderia* DSM7029, both encoding hybrid NRPS/PKS pathways, have been expressed in the probiotic strain *E. coli* Nissle 1917 under the control of a tetracycline inducible promoter.[151, 152] Transcriptional regulation appears to be a common challenge that must be overcome for BGC expression in laboratory strains of *E. coli*;[153] for this reason, *E. coli* BL21(DE3), equipped with the T7 RNA polymerase (RNAP), has been used to express pathways retrofitted with the orthogonal and inducible T7 promoter. BGCs tested include those originating from other Proteobacteria,[58, 74, 154] but importantly also include NRPS and PKS pathways from *Streptomyces*.[155, 156] T7 is used extensively, particularly for *E. coli* protein expression, but is limited for BGC expression as it can only reliably promote transcripts up to 20 kb in length.[153] In BL21(DE3), T7 is controlled by arabinose induction, which is amplified in an engineered strain, BT2, through genetic disruption of the endogenous arabinose catabolism pathway.[157]

Post-translational activation of biosynthetic enzymes represents another common challenge for heterologous expression of BGCs in *E. coli*. As the native PPTases of *E. coli* are often not able to effectively activate acyl or peptidyl carrier proteins involved in natural product biosynthesis,[158] *E. coli* BAP1, which harbors a genomically integrated copy of the Sfp PPTase from *B. subtilis*,[157] and GB05-MtaA, which harbors the MtaA PPTase from *Stigmatella aurantiaca*,[89] were generated. Use of these strains has proven essential for heterologous expression of several assembly line biosynthetic pathways.[44, 159-163] BAP1 is derived from BL21(DE3) and, in addition to Sfp, harbors a re-engineered *prp* operon to enhance supply of the polyketide precursor propionyl-CoA.[157] Additional genetic manipulations have been made in BAP1 to further enhance precursor supply for production of the antibiotic erythromycin A (**9**), a glycosylated polyketide macrolide. These manipulations include deletion of *ygfH*, resulting in generation of *E. coli* TB3 to further support propionyl-CoA supply,[164] and introduction of *pccAB*, resulting in generation of *E. coli* BTRAP for conversion of propionyl-CoA to methylmalonyl-CoA.[157] Finally, *E. coli* LF01 was generated from TB3 via sequential deletion of genes from pathways capable of siphoning carbon from deoxysugar biosynthesis to improve erythromycin glycosylation.[165] Although these manipulations were made specifically for enhanced heterologous erythromycin production, they

demonstrate the potential versatility of *E. coli* to be engineered for optimal heterologous expression of any natural product BGC given a comprehensive understanding of biosynthetic processes.

Despite challenges associated with heterologous expression of assembly line biosynthetic pathways, laboratory strains of *E. coli* are generally well suited for production of ribosomally synthesized and post-translationally modified peptides, or RiPPs. RiPPs represent an important and growing class of natural products;[166] RiPP BGCs are also relatively small, making them tractable for heterologous expression and engineering. Recently, van der Donk, Tavassoli, and colleagues reported the generation of a vast library of 10^6 lanthipeptide analogs in *E. coli,* which was screened for inhibitory activity against a key protein-protein interaction involved in HIV infection.[167]

Because most heterologous expression experiments in *E. coli* rely on self-replicating plasmids as opposed to chromosomal integration, strains of *E. coli* have also been engineered for enhanced plasmid stability, for example *E. coli* BT3 and BTRA, which are also derivatives of BAP1.[157] Tools for chromosomal integration of large DNA fragments in *E. coli* have also been developed, including a Cre-lox based system called recombinase-assisted genome engineering, or RAGE.[168, 169] In theory, RAGE can be applied to a range of organisms to enable chromosomal integration of large BGCs for heterologous expression.

Beyond *E. coli*, the γ-proteobacterium *Pseudomonas* has been used extensively for BGC heterologous expression. *Pseudomonas* grows fast, is genetically manipulatable, and species such as *P. putida* are generally recognized as safe (GRAS). Furthermore, *Pseudomonas* appears to be more naturally suited to BGC expression than *E. coli*, as most *Pseudomonas* used for heterologous expression to-date have not been purposefully modified.[40, 58, 110, 111, 170-176] *P. putida* KT2440 harbors only a single PPTase, but *in vivo* testing revealed that this PPTase was capable of effectively activating carrier proteins from *S. coelicolor* and several Myxobacteria.[177] *P. putida* KT2440 has been metabolically engineered for methylmalonyl-CoA biosynthesis to enable heterologous production of PKS pathways that require this extender unit.[178] Furthermore, *P. putida* KT2440 has been modified more extensively through a series of genomic deletions to enhance its ability to serve as a chassis for gene expression, resulting in the generation of *P. putida* EM383, which displays superior growth properties and improved genetic stability.[179] Despite this enhanced stability, expression of large BGCs in *P. putida* from self-replicating plasmids is still not as reliable as chromosomal integration, and therefore robust methods for site-specific integration of large BGCs need to be further developed and optimized for this host. It is worth reiterating that, because *P. putida* grows fast and can be transformed quickly and easily using electroporation, testing of genetic variants using this heterologous host can be performed in a much shorter time frame than for even the fastest growing *Streptomyces*. Furthermore, as *P. putida* has demonstrated ability to activate *S. coelicolor* carrier proteins, it may be worthwhile to test *Streptomyces* BGCs in *Pseudomonas* heterologous hosts.

Outside of γ-proteobacteria, δ-proteobacteria of the order myxobacteria have also been leveraged as heterologous hosts. Myxobacteria are incredibly interesting bacteria that exhibit social behaviors, moving and feeding in multicellular groups called swarms. Furthermore, analogous to *Streptomyces* and *Aspergillus*, myxobacteria undergo complex morphological and developmental changes during their life cycle, forming spore-producing structures, or fruiting bodies, under conditions of starvation.[180] Myxobacteria have large genomes (9-10 Mbp) relative to other bacteria and are prolific producers of natural products. Unfortunately, many species of myxobacteria do not make ideal expression hosts, as are difficult to work with and require extensive microbiological experience to culture and manipulate. Nevertheless, a few model myxobacteria have been successfully leveraged for expression of BGCs primarily from other myxobacteria, although the oxytetracycline (**10**) BGC from *Streptomyces rimosus* was also successfully reconstituted in a myxobacterial host.[181] *Myxococcus xanthus* DK1622 has been the most widely used myxobacterial host, and a strain of *M. xanthus* has been generated in which the native myxochromide A BGC has been inactivated in an attempt to improve heterologous production.[173, 182] Single- or double- crossover homologous recombination has been used for site-specific chromosomal integration of BGCs;[181, 183] however, this can be difficult, especially for large constructs. Transposition has also been used in both *M. xanthus*[111, 171] and *Corallococcus macrosporus* GT-2,[184] a moderately thermophilic, faster growing myxobacteria. While more efficient than homologous recombination, integration via transposition, which is stochastic, can change host transcriptomes, leading to varying levels of heterologous compound production and preventing reproducibility between experiments.[173, 185] This also makes it impossible to compare results between experiments, for example if BGCs are genetically manipulated and re-transferred to the heterologous host. Thus, Mx9 and Mx8 integrases, identified from

bacteriophages, have also been successfully applied for site-specific BGC integration.[100, 112, 186] It is noteworthy that the pyxidicycline A (**11**) BGC from *Pyxidicoccus fallax* An d48, which was expressed in both *M. xanthus* and *Stigmatella aurantiaca* DW4/3-1, displayed distinct, host-specific production profiles, where different analogs were favored in the hosts and native producer.[186] This has also been observed for other BGCs and may indicate key differences in precursor availability, biosynthesis, or regulation across host organisms.[139]

Finally, various α- and β-proteobacterial hosts have been used for heterologous expression, albeit much more rarely. An eDNA library prepared using a broad-host-range cosmid vector was screened in the β-proteobacterium *Ralstonia metallidurans*, resulting in the identification of terpene and type III PKS products that could not be produced in *E. coli*.[187] Subsequently, broad-host-range eDNA clones were screened in a more diverse range of proteobacterial hosts, including the α-proteobacteria *Agrobacterium tumefaciens* and *Caulobacter vibrioides* and the β-proteobacterium *Burkholderia graminis*, in addition to *R. metallidurans*, *P. putida*, and *E. coli*.[110] eDNA expressed across these six hosts showed minimal overlap in host compatibility.[110] A similar broad-host-range approach was taken for expression of the violacein (**12**) BGC from the γ-proteobacterium *Pseudoalteromonas luteoviolacea* 2ta16 and demonstrated that although the α-proteobacterium *A. tumefaciens* was the least phylogenetically related host organism tested, it produced the greatest yield of violacein due to differences in BGC regulation.[58] Finally, *Rhodobacter capsulatus* was successfully engineered for expression of various pigment BGCs,[172] and *Burkholderida* DSM 7029 has been used for heterologous production of epothilone A (**13**) and vioprolide B (**14**), which both originate from myxobacteria.[173, 188]

2.4 Cyanobacteria

Cyanobacteria are photosynthetic bacteria, found in freshwater and marine ecosystems, that have historically been a rich source of bioactive natural products and toxins.[189] The cyanobacterial phylum is diverse, but unfortunately most Cyanobacteria grow slowly and are difficult to genetically manipulate. Despite these challenges, recent advances have been made toward developing genetic tools and cyanobacterial hosts for heterologous expression of intransigent BGCs. A broad-host-range vector was constructed and tested in an array of “model” cyanobacterial hosts, including *Synechococcus elongatus* PCC7942, *Synechocystis* sp. PCC6803, *Anabaena* sp. PCC7120, *Leptolyngbya* sp. BL0902, and *Nostoc punctiforme* ATCC29133.[190] More recently, *S. elongatus* PCC7942 was used for heterologous expression of polybrominated diphenyl ethers, through integration of PBDE biosynthetic genes, identified in sponge-derived metagenomic cyanobacterial DNA sequences, via homologous recombination into neutral sites in the *S. elongatus* genome.[191] Gene expression was driven by a synthetic promoter-riboswitch, providing precise control over protein production. *S. elongatus* PCC7942 has also been optimized specifically for polyketide production via introduction of modules that aid in PKS extender unit production, regulation (via the T7 RNAP), and post-translational modification (via the Sfp PPTase).[192] *S.* sp. PCC6803 was successfully engineered for photosynthetic overproduction of mycosporine-like amino acids (MAA) through promoter refactoring and co-expression of the *Anabaena* sp. PCC7120 PPTase (APPT).[193] APPT has been characterized and displays broad substrate scope and good catalytic efficiency against a range of cyanobacterial and *Streptomyces* carrier proteins.[194] Finally, *Anabaena* sp. PCC7120 has been further explored for heterologous expression of BGCs using replicative plasmids, resulting in successful heterologous production of lyngbyatoxin A (**15**) from *Moorea producens* and MAA from *Nostoc flagelliforme*.[195, 196]

2.5 Fungi

Filamentous fungi, like filamentous bacteria, produce many bioactive natural products and harbor scores of uncharacterized NRPS, PKS, and miscellaneous BGCs. However, many important differences underly bacterial and fungal transcription, translation, and basic cell biology. Therefore, it seems reasonable that identifying viable fungal hosts would be very important for heterologous expression of fungal BGCs. Fungal hosts carry more biological complexity than bacteria but in most cases are easy to cultivate, genetically manipulatable, and much more tractable than higher eukaryotes.

The baker's yeast *Sacchromyces cerevisiae* has been successfully used for expression of many biosynthetic genes and gene clusters from filamentous fungi.[197-207] Like *E. coli*, *S. cerevisiae* is genetically tractable, grows quickly, and has been a longstanding laboratory workhorse and model organism. Also like *E. coli*, it is not a naturally gifted natural product producer. For this reason, *S. cerevisiae* has been outfitted with the NpgA PPTase from *Aspergillus nidulans* specifically for expression of NRPS and PKS genes.[208] *S. cerevisiae* has also undergone extensive metabolic engineering to increase PKS precursor supply, specifically acetyl-CoA.[197, 198] The majority of heterologous expression experiments in *S. cerevisiae* have involved only a single or few individual genes expressed from replicative vectors,[199-204] although entire pathways involving more than three or four genes have also been reconstituted in their entirety.[205-207] Most BGCs expressed in *S. cerevisiae* to date originate from other fungi, but it has also been used to reconstitute the bacterial NRPS BpsA from *Streptomyces lavendulae*.[209] Recently, HEx (Heterologous EXpression) was established as a high-throughput platform for expression of fungal BGCs in *S. cerevisiae*.[210] Using this platform, 41 orphan BGCs from diverse fungal species were interrogated, resulting in 22 detectable compounds.[210] Finally, new methods for BGC integration into yeast chromosomes are being developed, for example a method called delta integration CRISPR-Cas, or Di-CRISPR.[211]

Despite many successes, there are significant differences between intron processing in filamentous fungi and yeast.[212] Although intron sequences can be effectively removed by cloning from cDNA, this requires RNA isolation and also makes it difficult to clone and express multi-gene pathways. Thus, filamentous fungal hosts have also been developed. *Aspergillus nidulans* is a model organism that can be genetically manipulated and is highly recombinant; it has been used extensively for expression of BGCs from *Aspergilli* and other ascomycete fungi.[34-37, 213-216] *A. nidulans* has been engineered through deletion of endogenous BGCs, and characterized promoters can be leveraged in this host.[217] *Aspergillus oryzae* has been successfully used for heterologous reconstitution of the antibiotic pleuromutilin (**16**), a diterpene produced by the basidiomycete *Clitopilus passeckerianus*.[218] Finally, *Fusarium heterosporum* ATCC 74349, an ascomycete that produces high levels of the natural product equisetin,[219] has been leveraged to express BGCs from *Aspergillus flavus*, *Aspergillus terreus*, *Hapsidospora irregularis*, and an uncharacterized fern endophyte isolated from Papua New Guinea.[220-222] Furthermore, fusions of fungal PKS/NRPS genes were tested in this host strain to determine guidelines for engineering hybrid PKS/NRPS pathways.[223]

Heterologous expression of microbial BGCs, enabled by the diverse range of hosts reviewed in this section, has greatly expanded our knowledge of natural product biosynthesis, bioactivity, and regulation. While it is worthwhile to develop new and diverse host organisms, current platforms are not designed to be broadly applicable, as dual cloning and expression vectors outfitted with host-specific elements make it difficult to test expression of cloned BGCs in different hosts, particularly across bacterial and fungal phyla. More universal systems would greatly streamline the ability to test BGC expression across multiple hosts. One potential approach involves introduction of modular expression elements (ideally for site-specific chromosomal integration) after cloning, enabling vectors to be easily retrofitted for expression testing across different hosts (Fig. 3A). Alternatively, diverse hosts can be engineered to carry the same BGC integration site, such that a single vector system is broadly compatible with many different expression hosts (Fig. 3B). This could be achieved, for example, leveraging a system like recombinase-assisted genome engineering (RAGE) to introduce the same attachment site in diverse hosts organisms.[169]

4. Genetic manipulation of cloned BGCs

Cloned constructs can be quickly edited leveraging *E. coli* recombineering tools such as PCR targeting and λ-Red recombination.[224, 225] Therefore, cloning of BGCs also enables rapid genetic manipulation of biosynthetic pathways through gene knock-in or knock-out experiments. Edited constructs can then be re-introduced to heterologous hosts to test for expression. In this manner, BGCs can be engineered for activation or enhanced production of chemical products,[226] generation of new analogs,[227] or interrogation of biosynthetic mechanism.[228, 229] This process can be iterated to quickly connect biosynthetic genotype and chemotype. Standard *E. coli* recombineering leveraging antibiotic resistance cassettes makes most gene deletion and insertion experiments routine, although care should be taken to avoid polar effects in bacteria. Alternatively, yeast can also be used, leveraging natural recombination and various auxotrophic markers or resistance

cassettes, for deletion or insertion.[72] This is particularly attractive for large, highly repetitive BGCs cloned into high copy number *E. coli* vectors, since recombineering strains of *E. coli* are inherently less genetically stable and could catalyze unwanted plasmid recombination resulting in BGC rearrangements or deletions. However, yeast manipulations take longer and are only possible with vectors containing yeast origins of replication and selection markers.

Selection markers flanked by FRT (flippase recognition target) sequences can be removed using Flp (flippase) recombinase-mediated excision, leaving behind a small scar sequence. This adds an additional step, and clones must be carefully screened since there is no mechanism to select for excision. Recently, marker-less deletions were successfully made *in vitro* using CRISPR-Cas9 in a method termed ICE (*in vitro* CRISPR/Cas9 mediated editing).[230] While this report simplifies the process of generating marker-less deletion mutants, it remains significantly challenging to make subtle marker-less changes, for example point mutations, within large DNA constructs, as these types of manipulations would usually be performed using PCR mutagenesis. One potential application for improved marker-less mutation methods is for editing assembly line BGCs, as these generally involve large, multigene pathways. Recently, we developed a method for facile generation of point mutations in large cloned constructs, which paves the way for easy editing of large BGCs in the future.[231]

5. Conclusions and future perspectives

With advances in DNA sequencing technology and bioinformatics tools, researchers have identified huge untapped microbial biosynthetic potential *in silico*, and genomics-based approaches have become routine for natural product discovery. The diverse strategies summarized in this review are advancing our ability to leverage microbial genome sequence information for the purpose of discovering new metabolites and understanding biosynthetic logic.

Over the past few decades, hundreds of natural product BGCs have been cloned and expressed in various heterologous hosts, the vast majority associated with known natural products (estimate based on analysis of the MIBiG database).[22] Thus, the next frontier lies in eliciting expression of so-called “silent” or “cryptic” BGCs to identify their chemical products and understand their biological activity. One promising approach has been the swapping of native promoters of “silent” BGCs with well-characterized constitutive or inducible promoters. This so-called “promoter refactoring’ strategy is particularly compatible with heterologous expression platforms, in which undetermined native transcriptional regulation systems can be replaced with well-characterized and controllable regulatory elements. While this approach has garnered significant interest and enthusiasm,[57, 232-236] current cloning and editing methods make it impractical to perform large-scale BGC refactoring, requiring substantial investments in time, energy, and money. In the future, we envision that DNA synthesis will play an outsized role in obtaining BGCs for functional characterization. Although currently costly and impractical for large, repetitive, and high GC-content BGCs, a small number of studies have already used *de novo* DNA synthesis to access and refactor BGCs less than 10 kb in size.[237-239] With increasing demand for synthetic DNA driven by the field of synthetic biology,[240] technical advances will likely spur decreased cost and enhanced capabilities of large-scale *de novo* DNA synthesis.[241] Conceivably, this will shift the natural product genome mining paradigm from an empirical “clone-edit-test” workflow to a streamlined and hypothesis-driven “design-build-test” workflow (Fig. 4). *De novo* DNA synthesis will enable experimental characterization of numerous BGCs identified from metagenomic sequence datasets, which represent the largest genomic reservoir for small molecule discovery and currently remains largely untapped.[242] Furthermore, establishing various synthetic transcriptional regulatory elements, including promoter sequences, ribosome binding sites, and terminator sequences, will expand the toolbox for BGC re-design and even enable generation of artificial biosynthetic pathways.[243-249] Finally, multi-omics techniques, including genomics, transcriptomics, proteomics, and metabolomics, can be integrated within heterologous expression platforms to diagnose production bottlenecks in a systematic manner.[250]

One hurdle BGC re-design cannot overcome is missing or incompatible “host factors” or biosynthetic elements essential for heterologous reconstitution of natural product production, especially when working with unconventional biosynthetic pathways where we may not know that a missing factor exists at all. As mentioned previously, current host selection and testing is limited and largely based on the assumption that the ideal heterologous host is most closely phylogenetically related to the native producing organism. However, this hypothesis has not been rigorously tested and there is limited empirical evidence to support this notion,

particularly for “silent” or “cryptic” BGCs. Expanding the spectrum of heterologous hosts that can be tested for BGC expression in a streamlined manner will likely enhance our ability to characterize novel BGCs. Beyond phylogeny, other biological factors such as down-scale cultivability, growth speed, transformation efficiency, and genetic amenability should also be considered in terms of the development of new heterologous hosts with potential for high-throughput screening campaigns. Finally, hosts extending into vastly distant biological domains, including archaea[251] and eukaryotes such as *Chlamydomonas reinhardtii*,[252] a green microalga, and *Nicotiana benthamiana*,[253] a wild relative of tobacco, will hopefully experience greater development and use as we continue to explore natural product biosynthesis across the tree of life.

Briefly, although this review focuses on heterologous expression-based genome mining strategies, we would also like to direct readers to a small number of studies that have pioneered the quest for mining natural products from native producers, including genetic and chemical induction-based approaches.[254-257] Furthermore, cell-free natural product biosynthesis, also referred to as enzymatic total synthesis or *in vitro* biosynthesis, has become a powerful approach to truly test our understanding of biosynthetic mechanism.[258-263] Cell-free methods also hold potential for obtaining scalable quantities of valuable bio-chemicals.[264, 265]

Over the past few decades, we have witnessed a renaissance within the field of natural products research, reinvigorated by the development of genomics-guided identification of microbial BGCs. The cloning methods, heterologous hosts, and genetic editing tools cited in this review will continue to be developed and utilized for BGC characterization. We remain in the infancy of investigating natural product biosynthesis and biological activity, with many challenges still to be addressed and even greater discoveries yet to be made.

6. Conflicts of interest

There are no conflicts to declare.

7. Acknowledgements

We acknowledge the generosity of the National Institutes of Health (N.I.H.) grants R01-GM085770 and R01-AI117712 for supporting our work on developing genetic platforms to mine microbial natural products. J.J.Z. was supported by a graduate fellowship from the National Science Foundation (N.S.F.) and N.I.H. grant F31-AI129299.

8. References

1. K. C. Nicolaou and S. Rigol, *J. Antibiot. (Tokyo)*, 2018, **71**, 153-184.
2. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629-661.
3. J. Heitman, N. R. Movva and M. N. Hall, *Science*, 1991, **253**, 905-909.
4. D. M. Sabatini, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 11818-11825.
5. K. Flinspach, L. Westrich, L. Kaysser, S. Siebenberg, J. P. Gomez-Escribano, M. Bibb, B. Gust and L. Heide, *Biopolymers*, 2010, **93**, 823-832.
6. L. C. Foulston and M. J. Bibb, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13461-13466.
7. L. Kaysser, X. Tang, E. Wemakor, K. Sedding, S. Hennig, S. Siebenberg and B. Gust, *Chembiochem*, 2011, **12**, 477-487.
8. K. Yamanaka, K. S. Ryan, T. A. Gulder, C. C. Hughes and B. S. Moore, *J. Am. Chem. Soc.*, 2012, **134**, 12434-12437.
9. I. G. U. Pait, S. Kitani, F. W. Roslan, D. Ulanova, M. Arai, H. Ikeda and T. Nihira, *J. Ind. Microbiol. Biotechnol.*, 2018, **45**, 77-87.
10. L. Kaysser, P. Bernhardt, S. J. Nam, S. Loesgen, J. G. Ruby, P. Skewes-Cox, P. R. Jensen, W. Fenical and B. S. Moore, *J. Am. Chem. Soc.*, 2012, **134**, 11988-11991.
11. S. M. Mantovani and B. S. Moore, *J. Am. Chem. Soc.*, 2013, **135**, 18032-18035.
12. M. Schorn, J. Zettler, J. P. Noel, P. C. Dorrestein, B. S. Moore and L. Kaysser, *ACS Chem. Biol.*, 2014, **9**, 301-309.

13. F. Leipoldt, J. Santos-Aberturas, D. P. Stegmann, F. Wolf, A. Kulik, R. Lacret, D. Popadic, D. Keinhorster, N. Kirchner, P. Bekiesch, H. Gross, A. W. Truman and L. Kaysser, *Nat Commun*, 2017, **8**, 1965.
14. F. Wolf, F. Leipoldt, A. Kulik, D. Wibberg, J. Kalinowski and L. Kaysser, *Chembiochem*, 2018, **19**, 1189-1195.
15. M. D. McMahon, C. Guan, J. Handelsman and M. G. Thomas, *Appl. Environ. Microbiol.*, 2012, **78**, 3622-3629.
16. J. G. Owen, Z. Charlop-Powers, A. G. Smith, M. A. Ternei, P. Y. Calle, B. V. Reddy, D. Montiel and S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 4221-4226.
17. H. A. Iqbal, L. Low-Beinart, J. U. Obiajulu and S. F. Brady, *J. Am. Chem. Soc.*, 2016, **138**, 9341-9344.
18. J. K. Bitok, C. Lemetre, M. A. Ternei and S. F. Brady, *FEMS Microbiol. Lett.*, 2017, **364**, fnx155.
19. B. M. Hover, S. H. Kim, M. Katz, Z. Charlop-Powers, J. G. Owen, M. A. Ternei, J. Maniko, A. B. Estrela, H. Molina, S. Park, D. S. Perlin and S. F. Brady, *Nat Microbiol*, 2018, **3**, 415-422.
20. J. Peek, M. Lilic, D. Montiel, A. Milshteyn, I. Woodworth, J. B. Biggins, M. A. Ternei, P. Y. Calle, M. Danziger, T. Warrier, K. Saito, N. Braffman, A. Fay, M. S. Glickman, S. A. Darst, E. A. Campbell and S. F. Brady, *Nat Commun*, 2018, **9**, 4147.
21. Z. Feng, J. H. Kim and S. F. Brady, *J. Am. Chem. Soc.*, 2010, **132**, 11902-11903.
22. M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Dusterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kotter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H. W. Nutzmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gomez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y. Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S. Doyle, A. J. Driessen, M. Ehling-Schulz, K. D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Hofte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H. J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Mendez, M. Metsa-Ketela, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Muller, B. A. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J. L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Sussmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano and F. O. Glockner, *Nat. Chem. Biol.*, 2015, **11**, 625-631.
23. Q. Tu, J. Herrmann, S. Hu, R. Raju, X. Bian, Y. Zhang and R. Muller, *Sci. Rep.*, 2016, **6**, 21066.
24. M. Xu, Y. Wang, Z. Zhao, G. Gao, S. X. Huang, Q. Kang, X. He, S. Lin, X. Pang, Z. Deng and M. Tao, *Appl. Environ. Microbiol.*, 2016, **82**, 5795-5805.
25. Q. Deng, L. Zhou, M. Luo, Z. Deng and C. Zhao, *Synth Syst Biotechnol*, 2017, **2**, 59-64.
26. M. Crusemann, R. Reher, I. Schamari, A. O. Brachmann, T. Ohbayashi, M. Kuschak, D. Malfacini, A. Seidinger, M. Pinto-Carbo, R. Richarz, T. Reuter, S. Kehraus, A. Hallab, M. Attwood, H. B. Schioth, P. Mergaert, Y. Kikuchi, T. F. Schaberle, E. Kostenis, D. Wenzel, C. E. Muller, J. Piel, A. Carlier, L. Eberl and G. M. Konig, *Angew. Chem. Int. Ed. Engl.*, 2018, **57**, 836-840.
27. H. Liu, H. Jiang, B. Haltli, K. Kulowski, E. Muszynska, X. Feng, M. Summers, M. Young, E. Graziani, F. Koehn, G. T. Carter and M. He, *J. Nat. Prod.*, 2009, **72**, 389-395.
28. T. Hashimoto, J. Hashimoto, I. Kozone, K. Amagai, T. Kawahara, S. Takahashi, H. Ikeda and K. Shin-Ya, *Org Lett*, 2018, **20**, 7996-7999.
29. A. C. Jones, B. Gust, A. Kulik, L. Heide, M. J. Buttner and M. J. Bibb, *PLoS One*, 2013, **8**, e69319.
30. J. F. Castro, V. Razmilic, J. P. Gomez-Escribano, B. Andrews, J. A. Asenjo and M. J. Bibb, *Appl. Environ. Microbiol.*, 2015, **81**, 5820-5831.
31. Z. Qin, J. T. Munnoch, R. Devine, N. A. Holmes, R. F. Seipke, K. A. Wilkinson, B. Wilkinson and M. I. Hutchings, *Chem. Sci.*, 2017, **8**, 3218-3227.
32. J. Tu, S. Li, J. Chen, Y. Song, S. Fu, J. Ju and Q. Li, *Microb Cell Fact*, 2018, **17**, 28.

33. B. Kepplinger, S. Morton-Laing, K. H. Seistrup, E. C. L. Marrs, A. P. Hopkins, J. D. Perry, H. Strahl, M. J. Hall, J. Errington and N. E. E. Allenby, *ACS Chem. Biol.*, 2018, **13**, 207-214.
34. J. W. Bok, R. Ye, K. D. Clevenger, D. Mead, M. Wagner, A. Krerowicz, J. C. Albright, A. W. Goering, P. M. Thomas, N. L. Kelleher, N. P. Keller and C. C. Wu, *BMC Genomics*, 2015, **16**, 343.
35. K. D. Clevenger, J. W. Bok, R. Ye, G. P. Miley, M. H. Verdan, T. Velk, C. Chen, K. Yang, M. T. Robey, P. Gao, M. Lamprecht, P. M. Thomas, M. N. Islam, J. M. Palmer, C. C. Wu, N. P. Keller and N. L. Kelleher, *Nat. Chem. Biol.*, 2017, **13**, 895-901.
36. M. T. Robey, R. Ye, J. W. Bok, K. D. Clevenger, M. N. Islam, C. Chen, R. Gupta, M. Swyers, E. Wu, P. Gao, P. M. Thomas, C. C. Wu, N. P. Keller and N. L. Kelleher, *ACS Chem. Biol.*, 2018, **13**, 1142-1147.
37. K. D. Clevenger, R. Ye, J. W. Bok, P. M. Thomas, M. N. Islam, G. P. Miley, M. T. Robey, C. Chen, K. Yang, M. Swyers, E. Wu, P. Gao, C. C. Wu, N. P. Keller and N. L. Kelleher, *Biochemistry*, 2018, **57**, 3237-3243.
38. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd and H. O. Smith, *Nat Methods*, 2009, **6**, 343-345.
39. L. Linares-Otoya, V. Linares-Otoya, L. Armas-Mantilla, C. Blanco-Olano, M. Crusemann, M. L. Ganoza-Yupanqui, J. Campos-Florian, G. M. Konig and T. F. Schaberle, *Microbiology*, 2017, **163**, 1409-1414.
40. N. M. Vior, R. Lacret, G. Chandra, S. Dorai-Raj, M. Trick and A. W. Truman, *Appl. Environ. Microbiol.*, 2018, **84**.
41. J. Mevaere, C. Goulard, O. Schneider, O. N. Sekurova, H. Ma, S. Zirah, C. Afonso, S. Rebuffat, S. B. Zotchev and Y. Li, *Sci. Rep.*, 2018, **8**, 8232.
42. X. Cai, V. L. Challinor, L. Zhao, D. Reimer, H. Adihou, P. Grun, M. Kaiser and H. B. Bode, *Org Lett*, 2017, **19**, 806-809.
43. M. E. Yurkovich, R. Jenkins, Y. Sun, M. Tosin and P. F. Leadlay, *Chem. Commun. (Camb.)*, 2017, **53**, 2182-2185.
44. C. Greunke, E. R. Duell, P. M. D'Agostino, A. Glockle, K. Lamm and T. A. M. Gulder, *Metab Eng*, 2018, **47**, 334-345.
45. P. M. D'Agostino and T. A. M. Gulder, *ACS Synth Biol*, 2018, **7**, 1702-1708.
46. E. R. Duell, P. M. D'Agostino, N. Shapiro, T. Woyke, T. M. Fuchs and T. A. M. Gulder, *Microb Cell Fact*, 2019, **18**, 32.
47. J. Liang, Z. Liu, X. Z. Low, E. L. Ang and H. Zhao, *Nucleic Acids Res.*, 2017, **45**, e94.
48. S. D. Colloms, C. A. Merrick, F. J. Olorunniji, W. M. Stark, M. C. Smith, A. Osbourn, J. D. Keasling and S. J. Rosser, *Nucleic Acids Res.*, 2014, **42**, e23.
49. F. J. Olorunniji, C. Merrick, S. J. Rosser, M. C. M. Smith, W. M. Stark and S. D. Colloms, *Methods Mol. Biol.*, 2017, **1642**, 303-323.
50. L. Zhang, X. Ou, G. Zhao and X. Ding, *J. Bacteriol.*, 2008, **190**, 6392-6397.
51. L. Zhang, G. Zhao and X. Ding, *Sci. Rep.*, 2011, **1**, 141.
52. X. Wang, B. Tang, Y. Ye, Y. Mao, X. Lei, G. Zhao and X. Ding, *Acta Biochim Biophys Sin (Shanghai)*, 2017, **49**, 44-50.
53. J. K. Liu, W. H. Chen, S. X. Ren, G. P. Zhao and J. Wang, *PLoS One*, 2014, **9**, e110852.
54. B. Enghiad and H. Zhao, *ACS Synth Biol*, 2017, **6**, 752-757.
55. J. H. Kim, Z. Feng, J. D. Bauer, D. Kallifidas, P. Y. Calle and S. F. Brady, *Biopolymers*, 2010, **93**, 833-844.
56. O. Bilyk, O. N. Sekurova, S. B. Zotchev and A. Luzhetskyy, *PLoS One*, 2016, **11**, e0158682.
57. K. D. Bauman, J. Li, K. Murata, S. M. Mantovani, S. Dahesh, V. Nizet, H. Luhavaya and B. S. Moore, *Cell Chem Biol*, 2019, **26**, 724-736.
58. J. J. Zhang, X. Tang, M. Zhang, D. Nguyen and B. S. Moore, *MBio*, 2017, **8**, e01291.
59. Y. Shi, Z. Jiang, X. Li, L. Zuo, X. Lei, L. Yu, L. Wu, J. Jiang and B. Hong, *Acta Pharm Sin B*, 2018, **8**, 283-294.
60. Z. Shao, H. Zhao and H. Zhao, *Nucleic Acids Res.*, 2009, **37**, e16.
61. Z. Shao, Y. Luo and H. Zhao, *Mol. Biosyst.*, 2011, **7**, 1056-1059.
62. Z. Shao, Y. Luo and H. Zhao, *Methods Mol. Biol.*, 2012, **898**, 251-262.
63. Z. Shao and H. Zhao, *Methods Enzymol.*, 2012, **517**, 203-224.
64. Z. Shao, G. Rao, C. Li, Z. Abil, Y. Luo and H. Zhao, *ACS Synth Biol*, 2013, **2**, 662-669.
65. Z. Shao and H. Zhao, *Methods Mol. Biol.*, 2013, **1073**, 85-106.

66. Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb and H. Zhao, *Nat Commun*, 2013, **4**, 2894.
67. Z. Shao and H. Zhao, *Curr. Protoc. Chem. Biol.*, 2014, **6**, 65-100.
68. O. Schimming, F. Fleischhacker, F. I. Nollmann and H. B. Bode, *Chembiochem*, 2014, **15**, 1290-1294.
69. A. Domrose, R. Weihmann, S. Thies, K. E. Jaeger, T. Drepper and A. Loeschcke, *Synth Syst Biotechnol*, 2017, **2**, 310-319.
70. P. Basitta, L. Westrich, M. Rosch, A. Kulik, B. Gust and A. K. Apel, *ACS Synth Biol*, 2017, **6**, 817-825.
71. N. Kouprina and V. Larionov, *Nat. Protoc.*, 2008, **3**, 371-377.
72. K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein and B. S. Moore, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 1957-1962.
73. Y. Li, Z. Li, K. Yamanaka, Y. Xu, W. Zhang, H. Vlamakis, R. Kolter, B. S. Moore and P. Y. Qian, *Sci. Rep.*, 2015, **5**, 9383.
74. A. C. Ross, L. E. Gulland, P. C. Dorrestein and B. S. Moore, *ACS Synth Biol*, 2015, **4**, 414-420.
75. B. Bonet, R. Teufel, M. Crusemann, N. Ziemert and B. S. Moore, *J. Nat. Prod.*, 2015, **78**, 539-542.
76. C. Cano-Prieto, R. Garcia-Salcedo, M. Sanchez-Hidalgo, A. F. Brana, H. P. Fiedler, C. Mendez, J. A. Salas and C. Olano, *Chembiochem*, 2015, **16**, 1461-1473.
77. L. Ray, K. Yamanaka and B. S. Moore, *Angew. Chem. Int. Ed. Engl.*, 2016, **55**, 364-367.
78. Z. R. Li, J. Li, J. P. Gu, J. Y. Lai, B. M. Duggan, W. P. Zhang, Z. L. Li, Y. X. Li, R. B. Tong, Y. Xu, D. H. Lin, B. S. Moore and P. Y. Qian, *Nat. Chem. Biol.*, 2016, **12**, 773-775.
79. C. B. Larson, M. Crusemann and B. S. Moore, *J. Nat. Prod.*, 2017, **80**, 1200-1204.
80. N. Wu, H. Huang, T. Min and H. Hu, *Acta Biochim Biophys Sin (Shanghai)*, 2017, **49**, 1129-1134.
81. R. Garcia-Salcedo, R. Alvarez-Alvarez, C. Olano, L. Canedo, A. F. Brana, C. Mendez, F. de la Calle and J. A. Salas, *Mar. Drugs*, 2018, **16**, 259.
82. Y. Hu, F. Nan, S. W. Maina, J. Guo, S. Wu and Z. Xin, *J. Biotechnol.*, 2018, **288**, 1-8.
83. V. N. Noskov, N. Kouprina, S. H. Leem, I. Ouspenski, J. C. Barrett and V. Larionov, *BMC Genomics*, 2003, **4**, 16.
84. X. Tang, J. Li, N. Millan-Aguinaga, J. J. Zhang, E. C. O'Neill, J. A. Ugalde, P. R. Jensen, S. M. Mantovani and B. S. Moore, *ACS Chem. Biol.*, 2015, **10**, 2841-2849.
85. F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin and C. Corre, *Chem. Sci.*, 2019, **10**, 453-463.
86. J. J. Zhang, K. Yamanaka, X. Tang and B. S. Moore, *Methods Enzymol.*, 2019, **621**, 87-110.
87. N. Kouprina and V. Larionov, *Nat Rev Genet*, 2006, **7**, 805-812.
88. N. C. Lee, V. Larionov and N. Kouprina, *Nucleic Acids Res.*, 2015, **43**, e55.
89. J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Muller, A. F. Stewart and Y. Zhang, *Nat. Biotechnol.*, 2012, **30**, 440-446.
90. X. Xu, H. Zhou, Y. Liu, X. Liu, J. Fu, A. Li, Y. Z. Li, Y. Shen, X. Bian and Y. Zhang, *J. Nat. Prod.*, 2018, **81**, 1060-1064.
91. F. Huang, J. Tang, L. He, X. Ding, S. Huang, Y. Zhang, Y. Sun and L. Xia, *Microb Cell Fact*, 2018, **17**, 31.
92. T. Thongkongkaew, W. Ding, E. Bratovanov, E. Oueis, A. A. M. A. Garci, N. Zaburannyi, K. Harmrolfs, Y. Zhang, K. Scherlach, R. Muller and C. Hertweck, *ACS Chem. Biol.*, 2018, **13**, 1370-1379.
93. Y. Tang, S. Frewert, K. Harmrolfs, J. Herrmann, L. Karmann, U. Kazmaier, L. Xia, Y. Zhang and R. Muller, *J. Biotechnol.*, 2015, **194**, 112-114.
94. H. Wang, Z. Li, R. Jia, J. Yin, A. Li, L. Xia, Y. Yin, R. Muller, J. Fu, A. F. Stewart and Y. Zhang, *Nucleic Acids Res.*, 2018, **46**, e28.
95. C. Song, J. Luan, Q. Cui, Q. Duan, Z. Li, Y. Gao, R. Li, A. Li, Y. M. Shen, Y. Z. Li, A. F. Stewart, Y. Zhang, J. Fu and H. Wang, *ACS Synth Biol*, 2018, **8**, 137-147.
96. R. Y. Wang, Z. Y. Shi, J. C. Chen and G. Q. Chen, *ACS Synth Biol*, 2012, **1**, 291-295.
97. Y. Zhou, A. C. Murphy, M. Samborskyy, P. Prediger, L. C. Dias and P. F. Leadlay, *Chem. Biol.*, 2015, **22**, 745-754.
98. W. Jiang, X. Zhao, T. Gabrieli, C. Lou, Y. Ebenstein and T. F. Zhu, *Nat Commun*, 2015, **6**, 8101.
99. W. Tang, Z. Guo, Z. Cao, M. Wang, P. Li, X. Meng, X. Zhao, Z. Xie, W. Wang, A. Zhou, C. Lou and Y. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 2818-2823.
100. C. Osswald, N. Zaburannyi, C. Burgard, T. Hoffmann, S. C. Wenzel and R. Muller, *J. Biotechnol.*, 2014, **191**, 54-63.
101. H. J. Nah, M. W. Woo, S. S. Choi and E. S. Kim, *Microb Cell Fact*, 2015, **14**, 140.

102. J. Wang, A. Lu, J. Liu, W. Huang, J. Wang, Z. Cai and G. Zhao, *Acta Biochim Biophys Sin (Shanghai)*, 2019, **51**, 97-103.
103. R. X. Dai, B. Zhang, G. P. Zhao and X. M. Ding, *Eng Life Sci*, 2015, **15**, 655-659.
104. Y. Yu, B. Tang, R. Dai, B. Zhang, L. Chen, H. Yang, G. Zhao and X. Ding, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 2621-2633.
105. D. Du, L. Wang, Y. Tian, H. Liu, H. Tan and G. Niu, *Sci. Rep.*, 2015, **5**, 8740.
106. H. N. Nguyen, K. I. Ishidoh, H. Kinoshita and T. Nihira, *J. Microbiol. Methods*, 2018, **150**, 47-54.
107. N. Ziemert, A. Lechner, M. Wietz, N. Millan-Aguinaga, K. L. Chavarria and P. R. Jensen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E1130-1139.
108. A. C. Letzel, J. Li, G. C. A. Amos, N. Millan-Aguinaga, J. Ginigini, U. R. Abdelmohsen, S. P. Gaudencio, N. Ziemert, B. S. Moore and P. R. Jensen, *Environ. Microbiol.*, 2017, **19**, 3660-3673.
109. Y. Martinez-Burgo, R. Alvarez-Alvarez, R. Perez-Redondo and P. Liras, *J. Biotechnol.*, 2014, **186**, 21-29.
110. J. W. Craig, F. Y. Chang, J. H. Kim, S. C. Obiajulu and S. F. Brady, *Appl. Environ. Microbiol.*, 2010, **76**, 1633-1641.
111. J. Fu, S. C. Wenzel, O. Perlova, J. Wang, F. Gross, Z. Tang, Y. Yin, A. F. Stewart, R. Muller and Y. Zhang, *Nucleic Acids Res.*, 2008, **36**, e113.
112. S. C. Wenzel, H. Hoffmann, J. Zhang, L. Debussche, S. Haag-Richter, M. Kurz, F. Nardi, P. Lukat, I. Kochems, H. Tietgen, D. Schummer, J. P. Nicolas, L. Calvet, V. Czepczor, P. Vrignaud, A. Muhlenweg, S. Pelzer, R. Muller and M. Bronstrup, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 15560-15564.
113. J. P. Gomez-Escribano and M. J. Bibb, *Microb Biotechnol*, 2011, **4**, 207-215.
114. M. Komatsu, T. Uchiyama, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2646-2651.
115. M. Komatsu, K. Komatsu, H. Koiwai, Y. Yamada, I. Kozone, M. Izumikawa, J. Hashimoto, M. Takagi, S. Omura, K. Shin-ya, D. E. Cane and H. Ikeda, *ACS Synth Biol*, 2013, **2**, 384-396.
116. Q. Peng, G. Gao, J. Lu, Q. Long, X. Chen, F. Zhang, M. Xu, K. Liu, Y. Wang, Z. Deng, Z. Li and M. Tao, *Front. Microbiol.*, 2018, **9**, 3042.
117. M. Myronovskyi, B. Rosenkranzer, S. Nadmid, P. Pujic, P. Normand and A. Luzhetskyy, *Metab Eng*, 2018, **49**, 316-324.
118. D. Kallifidas, G. Jiang, Y. Ding and H. Luesch, *Microb Cell Fact*, 2018, **17**, 25.
119. S. M. Salem, P. Kancharla, G. Florova, S. Gupta, W. Lu and K. A. Reynolds, *J. Am. Chem. Soc.*, 2014, **136**, 4565-4574.
120. D. C. Alexander, J. Rock, X. He, P. Brian, V. Miao and R. H. Baltz, *Appl. Environ. Microbiol.*, 2010, **76**, 6877-6887.
121. R. Alvarez-Alvarez, Y. Martinez-Burgo, R. Perez-Redondo, A. F. Brana, J. F. Martin and P. Liras, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 9451-9463.
122. Q. T. Bu, P. Yu, J. Wang, Z. Y. Li, X. A. Chen, X. M. Mao and Y. Q. Li, *Microb Cell Fact*, 2019, **18**, 16.
123. B. Zhang, K. B. Wang, W. Wang, S. F. Bi, Y. N. Mei, X. Z. Deng, R. H. Jiao, R. X. Tan and H. M. Ge, *Org Lett*, 2018, **20**, 2967-2971.
124. W. Chen and Z. Qin, *BMC Microbiol.*, 2011, **11**, 243.
125. J. M. Winter, M. C. Moffitt, E. Zazopoulos, J. B. McAlpine, P. C. Dorrestein and B. S. Moore, *J. Biol. Chem.*, 2007, **282**, 16362-16368.
126. G. Zhang, H. Zhang, S. Li, J. Xiao, G. Zhang, Y. Zhu, S. Niu, J. Ju and C. Zhang, *Appl. Environ. Microbiol.*, 2012, **78**, 2393-2401.
127. F. Kudo, T. Tsunoda, M. Takashima and T. Eguchi, *Chembiochem*, 2016, **17**, 2143-2148.
128. H. M. Thorpe and M. C. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5505-5510.
129. B. Gust, G. Chandra, D. Jakimowicz, T. Yuqing, C. J. Bruton and K. F. Chater, *Adv. Appl. Microbiol.*, 2004, **54**, 107-128.
130. A. S. Eustaquio, B. Gust, U. Galm, S. M. Li, K. F. Chater and L. Heide, *Appl. Environ. Microbiol.*, 2005, **71**, 2452-2459.
131. M. C. Smith, R. W. Hendrix, R. Dedrick, K. Mitchell, C. C. Ko, D. Russell, E. Bell, M. Gregory, M. J. Bibb, F. Pethick, D. Jacobs-Sera, P. Herron, M. J. Buttner and G. F. Hatfull, *J. Bacteriol.*, 2013, **195**, 4924-4935.
132. B. Fayed, E. Younger, G. Taylor and M. C. Smith, *BMC Biotechnol.*, 2014, **14**, 51.
133. P. C. Fogg, J. A. Haley, W. M. Stark and M. C. Smith, *Appl. Environ. Microbiol.*, 2017, **83**, e02767.

134. B. Fayed, D. A. Ashford, A. M. Hashem, M. A. Amin, O. N. El Gazayerly, M. A. Gregory and M. C. Smith, *Appl. Environ. Microbiol.*, 2015, **81**, 8402-8413.
135. L. Li, K. Wei, X. Liu, Y. Wu, G. Zheng, S. Chen, W. Jiang and Y. Lu, *Metab Eng*, 2019, **52**, 153-167.
136. B. Bilyk, L. Horbal and A. Luzhetskyy, *Microb Cell Fact*, 2017, **16**, 5.
137. E. J. Sherwood, A. R. Hesketh and M. J. Bibb, *J. Bacteriol.*, 2013, **195**, 2309-2321.
138. P. N. Schwarz, A. Buchmann, L. Roller, A. Kulik, H. Gross, W. Wohlleben and E. Stegmann, *Biotechnol J*, 2018, **13**, 1700527.
139. J. J. Zhang, B. S. Moore and X. Tang, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 8437-8446.
140. J. Huang, Z. Yu, M. H. Li, J. D. Wang, H. Bai, J. Zhou and Y. G. Zheng, *Appl. Environ. Microbiol.*, 2016, **82**, 5603-5611.
141. J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Nat. Prod. Rep.*, 2014, **31**, 61-108.
142. A. M. Herzner, J. Dischinger, C. Szekat, M. Josten, S. Schmitz, A. Yakeleba, R. Reinartz, A. Jansen, H. G. Sahl, J. Piel and G. Bierbaum, *PLoS One*, 2011, **6**, e22389.
143. S. K. Choi, S. Y. Park, R. Kim, S. B. Kim, C. H. Lee, J. F. Kim and S. H. Park, *J. Bacteriol.*, 2009, **191**, 3350-3358.
144. S. A. Borisova, B. T. Circello, J. K. Zhang, W. A. van der Donk and W. W. Metcalf, *Chem. Biol.*, 2010, **17**, 28-37.
145. S. Zobel, J. Kumpfmuller, R. D. Sussmuth and T. Schweder, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 681-691.
146. J. Kumpfmuller, K. Methling, L. Fang, B. A. Pfeifer, M. Lalk and T. Schweder, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 1209-1220.
147. Q. Liu, Q. Shen, X. Bian, H. Chen, J. Fu, H. Wang, P. Lei, Z. Guo, W. Chen, D. Li and Y. Zhang, *Sci. Rep.*, 2016, **6**, 34623.
148. G. M. Daba, N. Ishibashi, T. Zendo and K. Sonomoto, *J. Appl. Microbiol.*, 2017, **123**, 1124-1132.
149. J. Borrero, E. Kelly, P. M. O'Connor, P. Kelleher, C. Scully, P. D. Cotter, J. Mahony and D. van Sinderen, *Appl. Environ. Microbiol.*, 2018, **84**, e01801.
150. J. Masschelein, M. Jenner and G. L. Challis, *Nat. Prod. Rep.*, 2017, **34**, 712-783.
151. X. Bian, A. Plaza, Y. Zhang and R. Muller, *J. Nat. Prod.*, 2012, **75**, 1652-1655.
152. X. Bian, F. Huang, H. Wang, T. Klefisch, R. Muller and Y. Zhang, *Chembiochem*, 2014, **15**, 2221-2224.
153. D. C. Stevens, T. P. Hari and C. N. Boddy, *Nat. Prod. Rep.*, 2013, **30**, 1391-1411.
154. V. Agarwal, A. A. El Gamal, K. Yamanaka, D. Poth, R. D. Kersten, M. Schorn, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2014, **10**, 640-647.
155. K. Watanabe, K. Hotta, A. P. Praseuth, K. Koketsu, A. Migita, C. N. Boddy, C. C. Wang, H. Oguri and H. Oikawa, *Nat. Chem. Biol.*, 2006, **2**, 423-428.
156. Y. Wang and B. A. Pfeifer, *Metab Eng*, 2008, **10**, 33-38.
157. H. Zhang, L. Fang, M. S. Osburne and B. A. Pfeifer, *Methods Mol. Biol.*, 2016, **1401**, 121-134.
158. R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923-936.
159. W. Zhang, Y. Li and Y. Tang, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20683-20688.
160. H. Zhang, Y. Wang, J. Wu, K. Skalina and B. A. Pfeifer, *Chem. Biol.*, 2010, **17**, 1232-1240.
161. D. C. Stevens, K. R. Conway, N. Pearce, L. R. Villegas-Penaranda, A. G. Garza and C. N. Boddy, *PLoS One*, 2013, **8**, e64858.
162. X. Bian, F. Huang, F. A. Stewart, L. Xia, Y. Zhang and R. Muller, *Chembiochem*, 2012, **13**, 1946-1952.
163. X. Cai, S. Nowak, F. Wesche, I. Bischoff, M. Kaiser, R. Furst and H. B. Bode, *Nat. Chem.*, 2017, **9**, 379-386.
164. H. Zhang, B. A. Boghigian and B. A. Pfeifer, *Biotechnol. Bioeng.*, 2010, **105**, 567-573.
165. L. Fang, M. Guell, G. M. Church and B. A. Pfeifer, *Biotechnol. Prog.*, 2018, **34**, 271-276.
166. P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein,

R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108-160.
167. X. Yang, K. R. Lennard, C. He, M. C. Walker, A. T. Ball, C. Doigneaux, A. Tavassoli and W. A. van der Donk, *Nat. Chem. Biol.*, 2018, **14**, 375-380.
168. C. N. Santos, D. D. Regitsky and Y. Yoshikuni, *Nat Commun*, 2013, **4**, 2503.
169. C. N. Santos and Y. Yoshikuni, *Nat. Protoc.*, 2014, **9**, 1320-1336.
170. S. C. Wenzel, F. Gross, Y. Zhang, J. Fu, A. F. Stewart and R. Muller, *Chem. Biol.*, 2005, **12**, 349-356.
171. Y. Chai, S. Shan, K. J. Weissman, S. Hu, Y. Zhang and R. Muller, *Chem. Biol.*, 2012, **19**, 361-371.
172. A. Loeschcke, A. Markert, S. Wilhelm, A. Wirtz, F. Rosenau, K. E. Jaeger and T. Drepper, *ACS Synth Biol*, 2013, **2**, 22-33.
173. F. Yan, D. Auerbach, Y. Chai, L. Keller, Q. Tu, S. Huttel, A. Glemser, H. A. Grab, T. Bach, Y. Zhang and R. Muller, *Angew. Chem. Int. Ed. Engl.*, 2018, **57**, 8754-8759.
174. C. Ramel, M. Tobler, M. Meyer, L. Bigler, M. O. Ebert, B. Schellenberg and R. Dudler, *BMC Biochem.*, 2009, **10**, 26.
175. A. Dudnik, L. Bigler and R. Dudler, *Microbiol. Res.*, 2013, **168**, 73-76.
176. A. M. Kretsch, G. L. Morgan, J. Tyrrell, E. Mevers, I. Vallet-Gely and B. Li, *Org Lett*, 2018, **20**, 4791-4795.
177. F. Gross, D. Gottschalk and R. Muller, *Appl. Microbiol. Biotechnol.*, 2005, **68**, 66-74.
178. F. Gross, M. W. Ring, O. Perlova, J. Fu, S. Schneider, K. Gerth, S. Kuhlmann, A. F. Stewart, Y. Zhang and R. Muller, *Chem. Biol.*, 2006, **13**, 1253-1264.
179. E. Martinez-Garcia, P. I. Nikel, T. Aparicio and V. de Lorenzo, *Microb Cell Fact*, 2014, **13**, 159.
180. J. Munoz-Dorado, F. J. Marcos-Torres, E. Garcia-Bravo, A. Moraleda-Munoz and J. Perez, *Front. Microbiol.*, 2016, **7**, 781.
181. D. C. Stevens, M. R. Henry, K. A. Murphy and C. N. Boddy, *Appl. Environ. Microbiol.*, 2010, **76**, 2681-2683.
182. H. Sucipto, D. Pogorevc, E. Luxenburger, S. C. Wenzel and R. Muller, *Metab Eng*, 2017, **44**, 160-170.
183. O. Perlova, J. Fu, S. Kuhlmann, D. Krug, A. F. Stewart, Y. Zhang and R. Muller, *Appl. Environ. Microbiol.*, 2006, **72**, 7485-7494.
184. O. Perlova, K. Gerth, S. Kuhlmann, Y. Zhang and R. Muller, *Microb Cell Fact*, 2009, **8**, 1.
185. L. P. Zhu, X. J. Yue, K. Han, Z. F. Li, L. S. Zheng, X. N. Yi, H. L. Wang, Y. M. Zhang and Y. Z. Li, *Microb Cell Fact*, 2015, **14**, 105.
186. F. Panter, D. Krug, S. Baumann and R. Muller, *Chem. Sci.*, 2018, **9**, 4898-4908.
187. J. W. Craig, F. Y. Chang and S. F. Brady, *ACS Chem. Biol.*, 2009, **4**, 23-28.
188. X. Bian, B. Tang, Y. Yu, Q. Tu, F. Gross, H. Wang, A. Li, J. Fu, Y. Shen, Y. Z. Li, A. F. Stewart, G. Zhao, X. Ding, R. Muller and Y. Zhang, *ACS Chem. Biol.*, 2017, **12**, 1805-1812.
189. J. K. Nunnery, E. Mevers and W. H. Gerwick, *Curr. Opin. Biotechnol.*, 2010, **21**, 787-793.
190. A. Taton, F. Unglaub, N. E. Wright, W. Y. Zeng, J. Paz-Yepes, B. Brahamsha, B. Palenik, T. C. Peterson, F. Haerizadeh, S. S. Golden and J. W. Golden, *Nucleic Acids Res.*, 2014, **42**, e136.
191. V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2017, **13**, 537-543.
192. J. Roulet, A. Taton, J. W. Golden, A. Arabolaza, M. D. Burkart and H. Gramajo, *Metab Eng*, 2018, **49**, 94-104.
193. G. Yang, M. A. Cozad, D. A. Holland, Y. Zhang, H. Luesch and Y. Ding, *ACS Synth Biol*, 2018, **7**, 664-671.
194. G. Yang, Y. Zhang, N. K. Lee, M. A. Cozad, S. E. Kearney, H. Luesch and Y. Ding, *Sci. Rep.*, 2017, **7**, 11888.
195. P. Videau, K. N. Wells, A. J. Singh, W. H. Gerwick and B. Philmus, *ACS Synth Biol*, 2016, **5**, 978-988.
196. J. L. Shang, Z. C. Zhang, X. Y. Yin, M. Chen, F. H. Hao, K. Wang, J. L. Feng, H. F. Xu, Y. C. Yin, H. R. Tang and B. S. Qiu, *Environ. Microbiol.*, 2018, **20**, 200-213.
197. J. Cardenas and N. A. Da Silva, *Metab Eng*, 2014, **25**, 194-203.
198. J. Cardenas and N. A. Da Silva, *Metab Eng*, 2016, **36**, 80-89.
199. H. Zhou, K. Qiao, Z. Gao, J. C. Vederas and Y. Tang, *J. Biol. Chem.*, 2010, **285**, 41412-41421.
200. Y. Xu, T. Zhou, S. Zhang, P. Espinosa-Artiles, L. Wang, W. Zhang, M. Lin, A. A. Gunatilaka, J. Zhan and I. Molnar, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 12354-12359.

201. A. O. Zabala, Y. H. Chooi, M. S. Choi, H. C. Lin and Y. Tang, *ACS Chem. Biol.*, 2014, **9**, 1576-1586.
202. M. Wang, M. Beissner and H. Zhao, *Chem. Biol.*, 2014, **21**, 257-263.
203. X. M. Mao, Z. J. Zhan, M. N. Grayson, M. C. Tang, W. Xu, Y. Q. Li, W. B. Yin, H. C. Lin, Y. H. Chooi, K. N. Houk and Y. Tang, *J. Am. Chem. Soc.*, 2015, **137**, 11904-11907.
204. Y. H. Chooi, C. Krill, R. A. Barrow, S. Chen, R. Trengove, R. P. Oliver and P. S. Solomon, *Appl. Environ. Microbiol.*, 2015, **81**, 177-186.
205. Y. Li, Y. H. Chooi, Y. Sheng, J. S. Valentine and Y. Tang, *J. Am. Chem. Soc.*, 2011, **133**, 15773-15785.
206. W. Xu, Y. H. Chooi, J. W. Choi, S. Li, J. C. Vederas, N. A. Da Silva and Y. Tang, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 6472-6475.
207. C. M. Bond and Y. Tang, *Metab Eng*, 2019, **51**, 1-8.
208. K. K. Lee, N. A. Da Silva and J. T. Kealey, *Anal. Biochem.*, 2009, **394**, 75-80.
209. M. Wehrs, J. P. Prahl, J. Moon, Y. Li, D. Tanjore, J. D. Keasling, T. Pray and A. Mukhopadhyay, *Microb Cell Fact*, 2018, **17**, 193.
210. C. J. B. Harvey, M. Tang, U. Schlecht, J. Horecka, C. R. Fischer, H. C. Lin, J. Li, B. Naughton, J. Cherry, M. Miranda, Y. F. Li, A. M. Chu, J. R. Hennessy, G. A. Vandova, D. Inglis, R. S. Aiyar, L. M. Steinmetz, R. W. Davis, M. H. Medema, E. Sattely, C. Khosla, R. P. St Onge, Y. Tang and M. E. Hillenmeyer, *Sci Adv*, 2018, **4**, eaar5459.
211. S. Shi, Y. Liang, E. L. Ang and H. Zhao, *Methods Mol. Biol.*, 2019, **1927**, 73-91.
212. D. C. Anyaogu and U. H. Mortensen, *Front. Microbiol.*, 2015, **6**, 77.
213. J. W. van Dijk, C. J. Guo and C. C. Wang, *Org Lett*, 2016, **18**, 6236-6239.
214. T. S. Lin, Y. M. Chiang and C. C. Wang, *Org Lett*, 2016, **18**, 1366-1369.
215. W. B. Yin, Y. H. Chooi, A. R. Smith, R. A. Cacho, Y. Hu, T. C. White and Y. Tang, *ACS Synth Biol*, 2013, **2**, 629-634.
216. N. Liu, Y. S. Hung, S. S. Gao, L. Hang, Y. Zou, Y. H. Chooi and Y. Tang, *Org Lett*, 2017, **19**, 3560-3563.
217. Y. M. Chiang, C. E. Oakley, M. Ahuja, R. Entwistle, A. Schultz, S. L. Chang, C. T. Sung, C. C. Wang and B. R. Oakley, *J. Am. Chem. Soc.*, 2013, **135**, 7720-7731.
218. F. Alberti, K. Khairudin, E. R. Venegas, J. A. Davies, P. M. Hayes, C. L. Willis, A. M. Bailey and G. D. Foster, *Nat Commun*, 2017, **8**, 1831.
219. J. W. Sims, J. P. Fillmore, D. D. Warner and E. W. Schmidt, *Chem. Commun. (Camb.)*, 2005, **2**, 186-188.
220. T. B. Kakule, R. C. Jadulco, M. Koch, J. E. Janso, L. R. Barrows and E. W. Schmidt, *ACS Synth Biol*, 2015, **4**, 625-633.
221. R. C. Jadulco, M. Koch, T. B. Kakule, E. W. Schmidt, A. Orendt, H. He, J. E. Janso, G. T. Carter, E. C. Larson, C. Pond, T. K. Matainaho and L. R. Barrows, *J. Nat. Prod.*, 2014, **77**, 2537-2544.
222. T. B. Kakule, S. Zhang, J. Zhan and E. W. Schmidt, *Org Lett*, 2015, **17**, 2295-2297.
223. T. B. Kakule, Z. Lin and E. W. Schmidt, *J. Am. Chem. Soc.*, 2014, **136**, 17882-17890.
224. K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci. U S A*, 2000, **97**, 6640-6645.
225. B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1541-1546.
226. D. Montiel, H. S. Kang, F. Y. Chang, Z. Charlop-Powers and S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 8953-8958.
227. T. H. Eyles, N. M. Vior and A. W. Truman, *ACS Synth Biol*, 2018, **7**, 1211-1218.
228. X. Tang, K. Eitel, L. Kaysser, A. Kulik, S. Grond and B. Gust, *Nat. Chem. Biol.*, 2013, **9**, 610-615.
229. X. Tang, J. Li and B. S. Moore, *Chembiochem*, 2017, **18**, 1072-1076.
230. Y. Liu, W. Tao, S. Wen, Z. Li, A. Yang, Z. Deng and Y. Sun, *MBio*, 2015, **6**, e01714-01715.
231. J. J. Zhang, X. Tang, T. Huan, A. C. Ross and B. S. Moore, *bioRxiv*, 2019, DOI: 10.1101/560987, 560987.
232. M. Myronovskyi and A. Luzhetskyy, *Nat. Prod. Rep.*, 2016, **33**, 1006-1019.
233. B. Baral, A. Akhgari and M. Metsa-Ketela, *Synth Syst Biotechnol*, 2018, **3**, 163-178.
234. H. Ren, S. Biswas, S. Ho, W. A. van der Donk and H. Zhao, *ACS Chem. Biol.*, 2018, **13**, 2966-2972.
235. F. Yan, C. Burgard, A. Popoff, N. Zaburannyi, G. Zipf, J. Maier, H. S. Bernauer, S. C. Wenzel and R. Muller, *Chem. Sci.*, 2018, **9**, 7510-7519.
236. P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509-523.

237. C. J. Guo, F. Y. Chang, T. P. Wyche, K. M. Backus, T. M. Acker, M. Funabashi, M. Taketani, M. S. Donia, S. Nayfach, K. S. Pollard, C. S. Craik, B. F. Cravatt, J. Clardy, C. A. Voigt and M. A. Fischbach, *Cell*, 2017, **168**, 517-526.
238. K. Gemperlein, M. Hoffmann, L. Huo, P. Pilak, L. Petzke, R. Muller and S. C. Wenzel, *Metab Eng*, 2017, **44**, 213-222.
239. R. C. Coates, B. P. Bowen, E. Oberortner, L. Thomashow, M. Hadjithomas, Z. Zhao, J. Ke, L. Silva, K. Louie, G. Wang, D. Robinson, A. Tarver, M. Hamilton, A. Lubbe, M. Feltcher, J. L. Dangl, A. Pati, D. Weller, T. R. Northen, J. F. Cheng, N. J. Mouncey, S. Deutsch and Y. Yoshikuni, *J. Ind. Microbiol. Biotechnol.*, 2018, **45**, 567-577.
240. G. Wright, *Nature*, 2014, **509**, S13.
241. S. Kosuri and G. M. Church, *Nat Methods*, 2014, **11**, 499-507.
242. C. Lok, *Nature*, 2015, **522**, 270-273.
243. T. H. Segall-Shapiro, E. D. Sontag and C. A. Voigt, *Nat. Biotechnol.*, 2018, **36**, 352-358.
244. M. J. Smanski, S. Bhatia, D. Zhao, Y. Park, B. A. W. L, G. Giannoukos, D. Ciulla, M. Busby, J. Calderon, R. Nicol, D. B. Gordon, D. Densmore and C. A. Voigt, *Nat. Biotechnol.*, 2014, **32**, 1241-1249.
245. W. R. Birmingham, C. A. Starbird, T. D. Panosian, D. P. Nannemann, T. M. Iverson and B. O. Bachmann, *Nat. Chem. Biol.*, 2014, **10**, 392-399.
246. S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante and C. D. Smolke, *Science*, 2015, **349**, 1095-1100.
247. X. Luo, M. A. Reiter, L. d'Espaux, J. Wong, C. M. Denby, A. Lechner, Y. Zhang, A. T. Grzybowski, S. Harth, W. Lin, H. Lee, C. Yu, J. Shin, K. Deng, V. T. Benites, G. Wang, E. E. K. Baidoo, Y. Chen, I. Dev, C. J. Petzold and J. D. Keasling, *Nature*, 2019, **567**, 123-126.
248. C. Osswald, G. Zipf, G. Schmidt, J. Maier, H. S. Bernauer, R. Muller and S. C. Wenzel, *ACS Synth Biol*, 2014, **3**, 759-772.
249. M. Winn, D. Francis and J. Micklefield, *Angew. Chem. Int. Ed. Engl.*, 2018, **57**, 6830-6833.
250. G. Y. Tan, K. Deng, X. Liu, H. Tao, Y. Chang, J. Chen, K. Chen, Z. Sheng, Z. Deng and T. Liu, *ACS Synth Biol*, 2017, **6**, 995-1005.
251. J. C. Charlesworth and B. P. Burns, *Archaea*, 2015, **2015**, 282035.
252. L. Doron, N. Segal and M. Shapira, *Front Plant Sci*, 2016, **7**, 505.
253. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell and A. Osbourn, *Metab Eng*, 2017, **42**, 185-193.
254. B. Wang, F. Guo, S. H. Dong and H. Zhao, *Nat. Chem. Biol.*, 2019, **15**, 111-114.
255. M. M. Zhang, F. T. Wong, Y. Wang, S. Luo, Y. H. Lim, E. Heng, W. L. Yeo, R. E. Cobb, B. Enghiad, E. L. Ang and H. Zhao, *Nat. Chem. Biol.*, 2017, **13**, 607-609.
256. M. R. Seyedsayamdost, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 7266-7271.
257. F. Xu, Y. Wu, C. Zhang, K. M. Davis, K. Moon, L. B. Bushin and M. R. Seyedsayamdost, *Nat. Chem. Biol.*, 2019, **15**, 161-168.
258. Q. Cheng, L. Xiang, M. Izumikawa, D. Meluzzi and B. S. Moore, *Nat. Chem. Biol.*, 2007, **3**, 557-558.
259. J. A. Kalaitzis, Q. Cheng, P. M. Thomas, N. L. Kelleher and B. S. Moore, *J. Nat. Prod.*, 2009, **72**, 469-472.
260. V. Agarwal and B. S. Moore, *ACS Chem. Biol.*, 2014, **9**, 1980-1984.
261. R. Teufel, L. Kaysser, M. T. Villaume, S. Diethelm, M. K. Carbullido, P. S. Baran and B. S. Moore, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 11019-11022.
262. A. W. Goering, J. Li, R. A. McClure, R. J. Thomson, M. C. Jewett and N. L. Kelleher, *Acs Synthetic Biology*, 2017, **6**, 39-44.
263. S. M. K. McKinnie, Z. D. Miles, P. A. Jordan, T. Awakawa, H. P. Pepper, L. A. M. Murray, J. H. George and B. S. Moore, *J. Am. Chem. Soc.*, 2018, **140**, 17840-17845.
264. H. Luhavaya, R. Sigrist, J. R. Chekan, S. M. K. McKinnie and B. S. Moore, *Angew. Chem. Int. Ed. Engl.*, 2019, DOI: 10.1002/anie.201901571.
265. J. R. Chekan, S. M. K. McKinnie, M. L. Moore, S. G. Poplawski, T. P. Michael and B. S. Moore, *Angew. Chem. Int. Ed. Engl.*, 2019, DOI: 10.1002/anie.201902910.

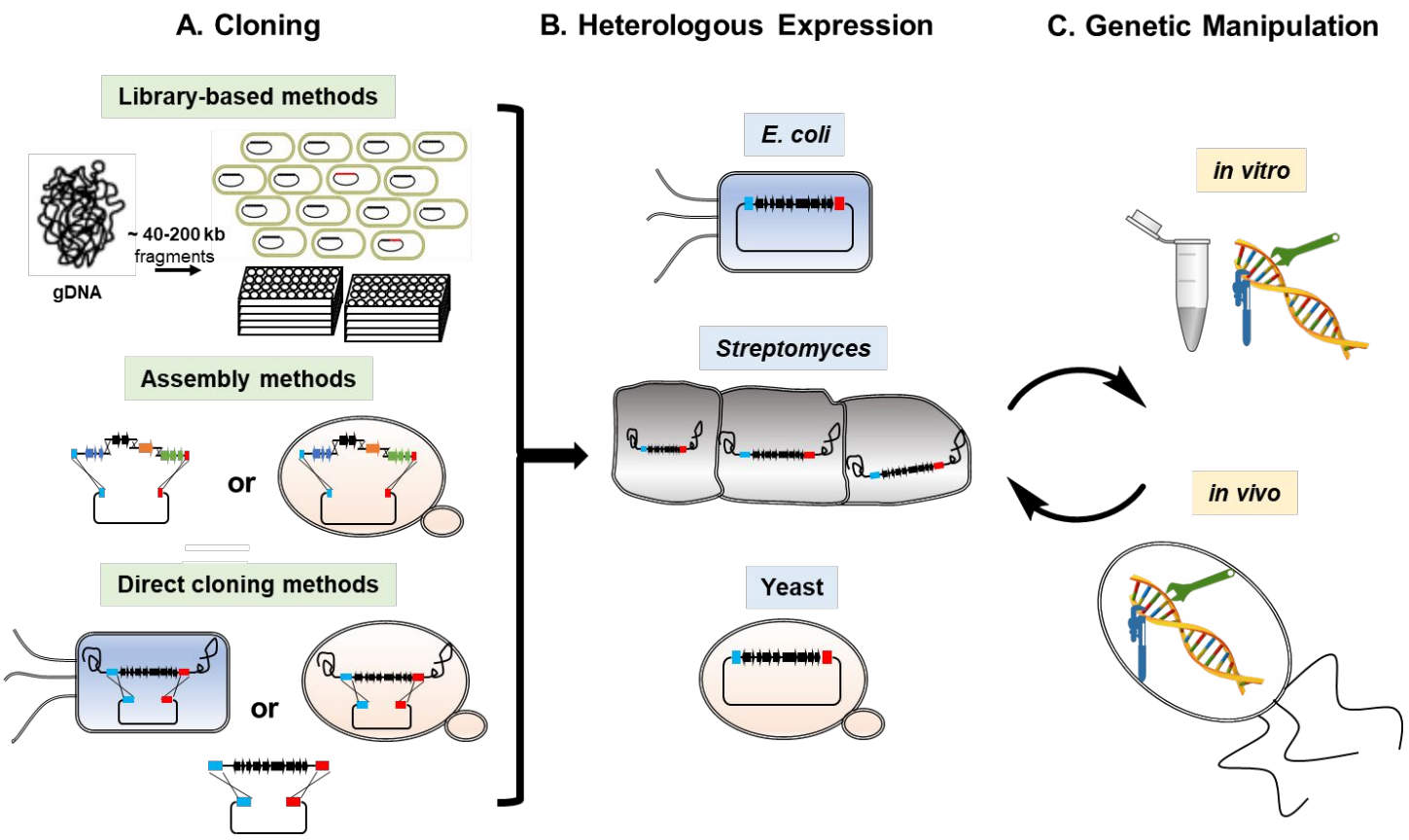

Fig. 1 General workflow for (A) cloning, (B) heterologous expression, and (C) genetic manipulation of microbial BGCs.

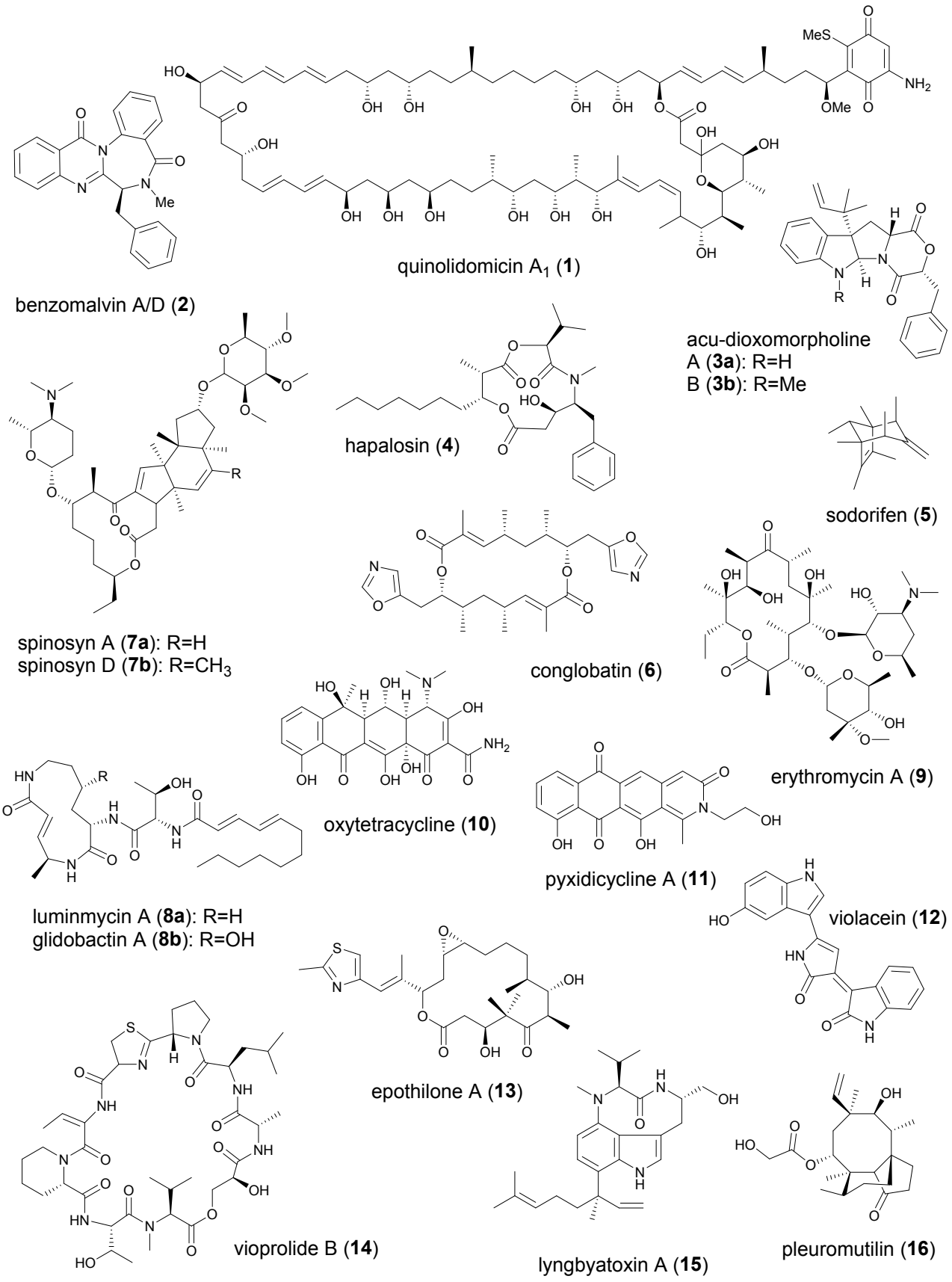

Fig. 2 Chemical structures of natural products mentioned in this article.

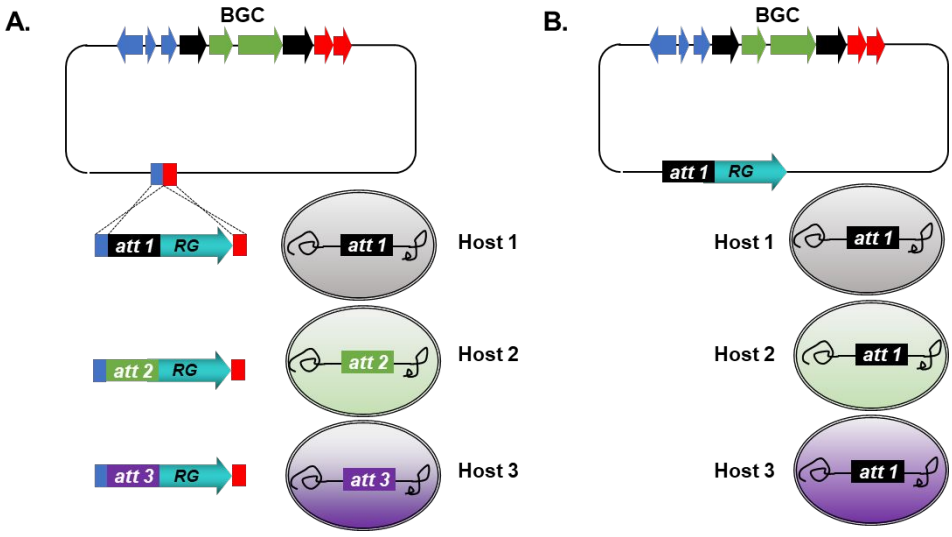

Fig. 3 Streamlined design of universal BGC expression platforms. (A) Modular expression vector system, leveraging knock-in of modular expression elements for BGC integration across many hosts. (B) Single vector system, compatible across several engineered hosts. *RG*: resistance gene; *att*: attachment site.

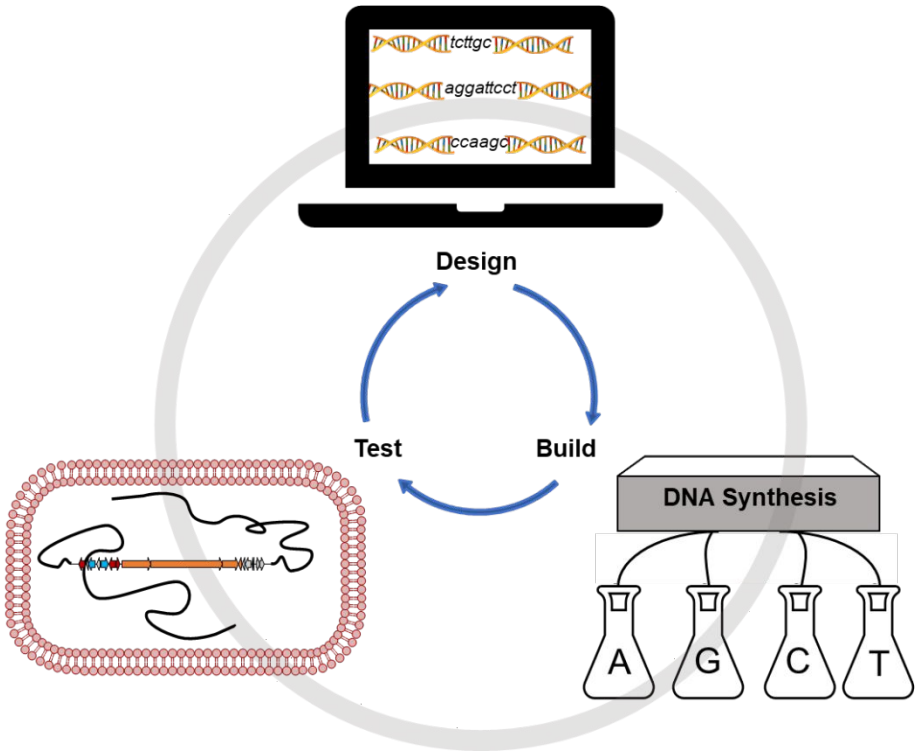

Fig. 4 "Design, build, test" workflow for future natural product genome mining.

Table 1 Summary of BGC cloning methods and their associated advantages and disadvantages.

		Examples	Advantages	Disadvantages	References
Library-based	Cosmid/fosmid	-	Sequence-independent, whole genome cloning	Untargeted; BGCs may be split across multiple library clones	5-14
	eDNA	-	Access BGCs from uncultured microbes	Unknown producer and final product; laborious screening necessary	15-21
	BAC/PAC	LEXAS	Large insert stability	Technically challenging	23-33
	FAC	-	Library can be readily screened in *A. nidulans*	Requires extensive screening, some false positives	34-37
Assembly	*in vitro*	Gibson, DiPAC, TPA, SIRA, SSRTA, AREs	Technically easier; rapid and potential to be automated	Impractical or cumbersome for large BGCs	38-54
	in vivo	DNA assembler, ExRec, yTREX, AGOS	Can assemble many DNA fragments (10+)	Difficult to troubleshoot; risk of mutation	55-70
Direct cloning	TAR	-	Robust direct cloning of whole BGCs	Can be technically challenging; must use yeast	72-82, 84-86
	LLHR	ExoCET	Utilizes *E. coli* as a cloning host	Technically challenging	88-95
	in vitro	SSOA, Gibson, CATCH, plasmid rescue/recovery	Streamlined *in vitro* approach	Requires careful preparation and/or manipulation of gDNA	96-106

eDNA, environmental DNA; BAC, bacterial artificial chromosome; PAC, P1 artificial chromosome; FAC, fungal artificial chromosome; TAR, transformation-associated recombination; LLHR, linear-linear homologous recombination; LEXAS, library expression analysis system; DiPAC, direct pathway cloning; TPA, twin-primer assembly; SIRA, serine integrase recombinational assembly; SSRTA, site-specific recombination-based tandem assembly; ARE, artificial restriction enzyme; ExRec, overlap extension PCR-yeast homologous recombination; yTREX, yeast recombinational cloning-enabled pathway transfer and expression tool; AGOS, artificial gene operon assembly system; ExoCET, exonuclease combined with RecET recombination; SSOA, single-strand overlapping annealing; CATCH, Cas9 assisted targeting of chromosome segments.

Table 2 Summary of heterologous hosts used for BGC expression.

	Phylum	Genus	Details	References
Bacteria				
Gram-positive				
	Actinobacteria			
		Streptomyces	Most widely used host, extensively optimized; site-specific BGC integration catalyzed by actinophage recombinases	109, 113-136
		Nonomuraea	All compatible with ΦC31 integrase;	6, 137
		Amycolatopsis	engineered strain *S. tropica* CNB-4401	138
		Salinispora	represents first marine actinomycete host	139
		Saccharopolyspora	*S. erythraea*, native erythromycin producer, engineered for spinosad production	140
	Firmicutes			
		Bacillus	Harbors Sfp PPTase, easily transformed and integrated with BGCs but underutilized	73, 82, 142-147
		Lactococcus	Leveraged for expression of small bacteriocin BGCs	148-149
Gram-negative				
	Proteobacteria			
	γ	*Escherichia*	Strains include Nissle 1917, BL21(DE3), BAP1, GB05-MtaA; also optimized for precursor flux and genetic stability	44, 58, 74, 89, 151-167
		Pseudomonas	Talented host; potential for expression of myxobacterial and *Streptomyces* BGCs	40, 58, 110-111, 170-178
	δ	*Myxococcus*	*M. xanthus, S. aurantiaca* leveraged for expression of many myxobacterial BGCs using transposition and phage integration;	99, 111-112, 171, 173, 181-183, 185-186
		Stigmatella	can be difficult to work with and grow slowly	186
		Corallococcus	*C. macrosporus* is moderately thermophilic	184
	α	*Agrobacterium*	Leveraged for expression of violacein,	58, 110
		Caulobacter	zeaxanthin, and prodigiosin pigments,	110
		Rhodobacter	various eDNA library clones	172
	β	*Ralstonia*	Used for expression of carotenoid, type III	110, 187
		Burkholderia	PKS products from eDNA; also epothilones and vioprolides	110, 173, 188
	Cyanobacteria			
		Synechococcus	*S. elongatus* PCC7942, *S.* sp. PCC6803	190-193
		Anabaena	*Anabaena* sp. PCC7120 used for expression of cyanobacterial pathways	190, 195-196
Fungi				
		Saccharomyces	HEx platform developed for high-throughput functional evaluation of fungal BGCs	197-211
		Aspergillus	More naturally suited for expression of	34-37, 213-218
		Fusarium	BGCs from filamentous fungi	220-223

Author Biographies

Jia Jia Zhang studied biomedical engineering at Harvard College, earning her bachelor's degree in 2013. She completed her PhD in Marine Chemical Biology in 2019 at the Scripps Institution of Oceanography (University of California San Diego), where she was an NSF graduate research fellow and NIH predoctoral fellow in the laboratory of Prof. Bradley S. Moore. Her PhD work focused on development of genetic tools for heterologous expression and genetic manipulation of natural product biosynthetic gene clusters.

Xiaoyu Tang is a research scientist at the biotech company Ginkgo Bioworks in Boston. He received his Ph.D. in Pharmaceutical Sciences from the Eberhard Karls Universität Tübingen (Germany) in August 2013, where he discovered, for the first time, the genuine sulfate donors for an arylsulfate sulfotransferase. Before moving to Boston, he was a postdoc at the Scripps Institution of Oceanography and J. Craig Venter Institute in the laboratories of Prof. Bradley S. Moore and Prof. Anna Edlund. His research at Scripps and JCVI focused on development of new genome mining strategies and genetics tools for discovery of genetically encoded small molecules from marine bacteria and the human microbiota, as well as study of their biosynthesis and biological roles.

Bradley Moore is Professor of Marine Chemical Biology at the Scripps Institution of Oceanography and Professor of Pharmaceutical Chemistry at the Skaggs School of Pharmacy and Pharmaceutical Sciences at UC San Diego. He received his B.S. in chemistry from the University of Hawaii, his PhD from the University of Washington, and was a postdoc at the University of Zurich. His research focuses on the molecular and genomic basis of natural product biosynthesis and the application of new genetic tools and biocatalysts to produce bioactive molecules.