

# Bacterial natural product discovery by heterologous expression

Adjo E. Kadjo<sup>1,2</sup> and Alessandra S. Eustáquio <sup>1,2</sup>

<sup>1</sup>Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA

<sup>2</sup>Center for Biomolecular Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA

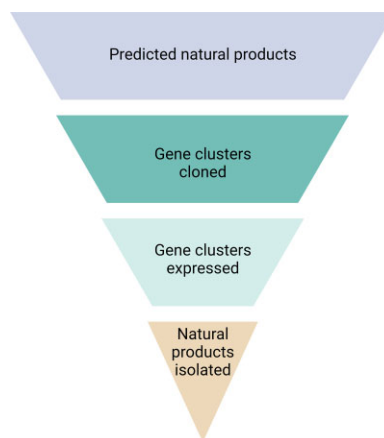
Correspondence should be addressed to: Alessandra S. Eustáquio at [ase@uic.edu](mailto:ase@uic.edu)

**Abstract:** Natural products have found important applications in the pharmaceutical and agricultural sectors. In bacteria, the genes that encode the biosynthesis of natural products are often colocalized in the genome, forming biosynthetic gene clusters. It has been predicted that only 3% of natural products encoded in bacterial genomes have been discovered thus far, in part because gene clusters may be poorly expressed under laboratory conditions. Heterologous expression can help convert bioinformatics predictions into products. However, challenges remain, such as gene cluster prioritization, cloning of the complete gene cluster, high level expression, product identification, and isolation of products in practical yields. Here we reviewed the literature from the past 5 years (January 2018 to June 2023) to identify studies that discovered natural products by heterologous expression. From the 50 studies identified, we present analyses of the rationale for gene cluster prioritization, cloning methods, biosynthetic class, source taxa, and host choice. Combined, the 50 studies led to the discovery of 63 new families of natural products, supporting heterologous expression as a promising way to access novel chemistry. However, the success rate of natural product detection varied from 11% to 32% based on four large-scale studies that were part of the reviewed literature. The low success rate makes it apparent that much remains to be improved. The potential reasons for failure and points to be considered to improve the chances of success are discussed.

**One-Sentence Summary:** At least 63 new families of bacterial natural products were discovered using heterologous expression in the last 5 years, supporting heterologous expression as a promising way to access novel chemistry; however, the success rate is low (11–32%) making it apparent that much remains to be improved—we discuss the potential reasons for failure and points to be considered to improve the chances of success. BioRender was used to generate the graphical abstract figure.

**Keywords:** Heterologous expression, Bacteria, Metabolites, Natural products, Synthetic biology

## Graphical abstract



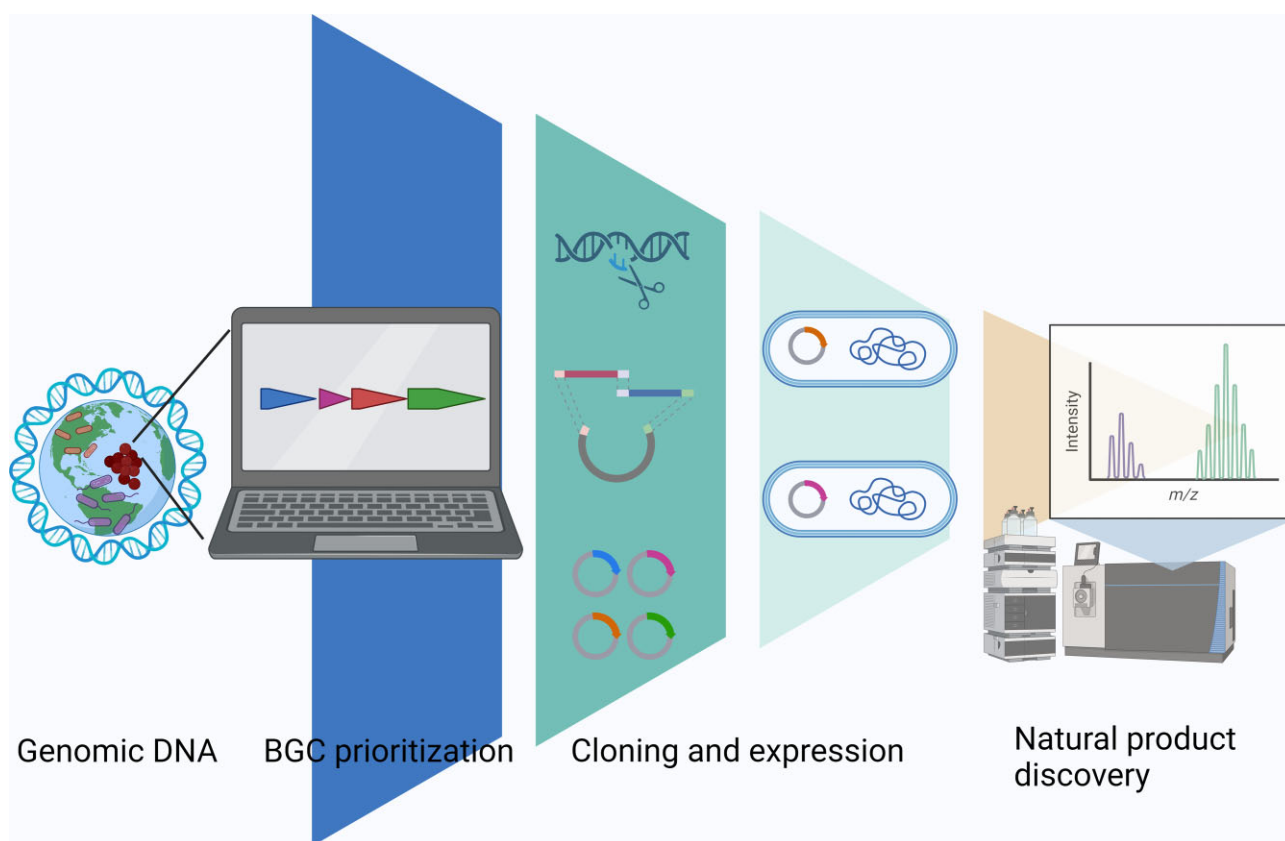
## Introduction

Natural products play important roles in the agricultural and pharmaceutical sectors. For instance, most small molecule drugs approved by the US Food and Drug Administration are either natural product (NP), NP derivatives, or synthetic compounds with NP

pharmacophores (Newman & Cragg, 2020). To counteract drug resistance, it is necessary to discover new compounds. Additionally, it has been estimated that 85% of the human, disease-associated proteome lack an associated therapeutic (Neklesa et al., 2017), implying a large therapeutic gap that could be at least partially filled by NP discovery.

Received: September 5, 2023. Accepted: December 4, 2023.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society of Industrial Microbiology and Biotechnology. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



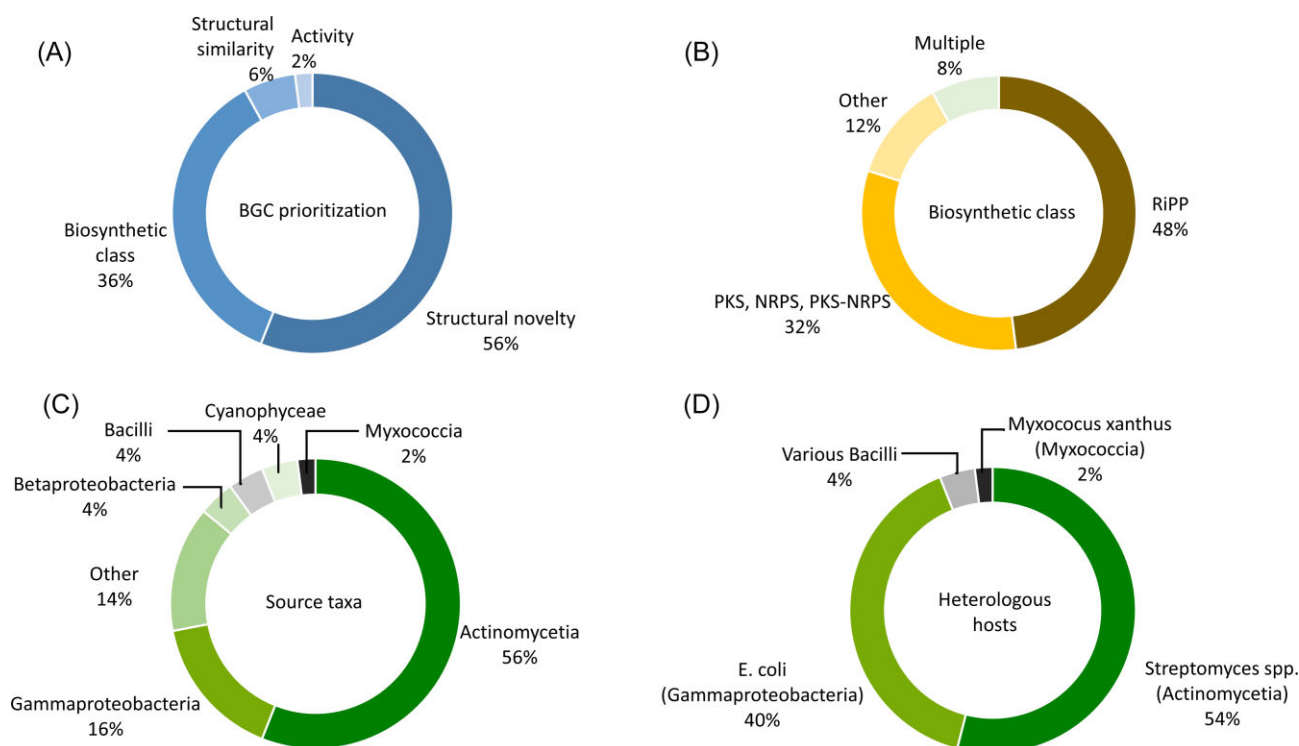
**Fig. 1.** Pipeline for natural product discovery by heterologous expression. Genomic DNA is sequenced and analyzed with bioinformatic tools. Biosynthetic gene clusters (BGCs) are predicted and prioritized. The selected BGCs are cloned and transferred into a suitable host for expression. Metabolomics tools are used for natural product detection, and if enough quantities are produced, the natural product is isolated and characterized. BioRender was used to generate this figure.

Advances in DNA sequencing and bioinformatics have revealed the untapped NP biosynthesis potential of microorganisms (Bachmann et al., 2014; Gavriilidou et al., 2022). A recent study predicted that only 3% of NPs encoded in bacterial genomes have been discovered (Gavriilidou et al., 2022). The discovery of new NPs is hampered by two main factors. First, most microorganisms from which NPs can be discovered remain uncultured. Second, the conditions used in the laboratory to study culturable microorganisms may not be appropriate for production in amounts that enable discovery and development (Baltz, 2017).

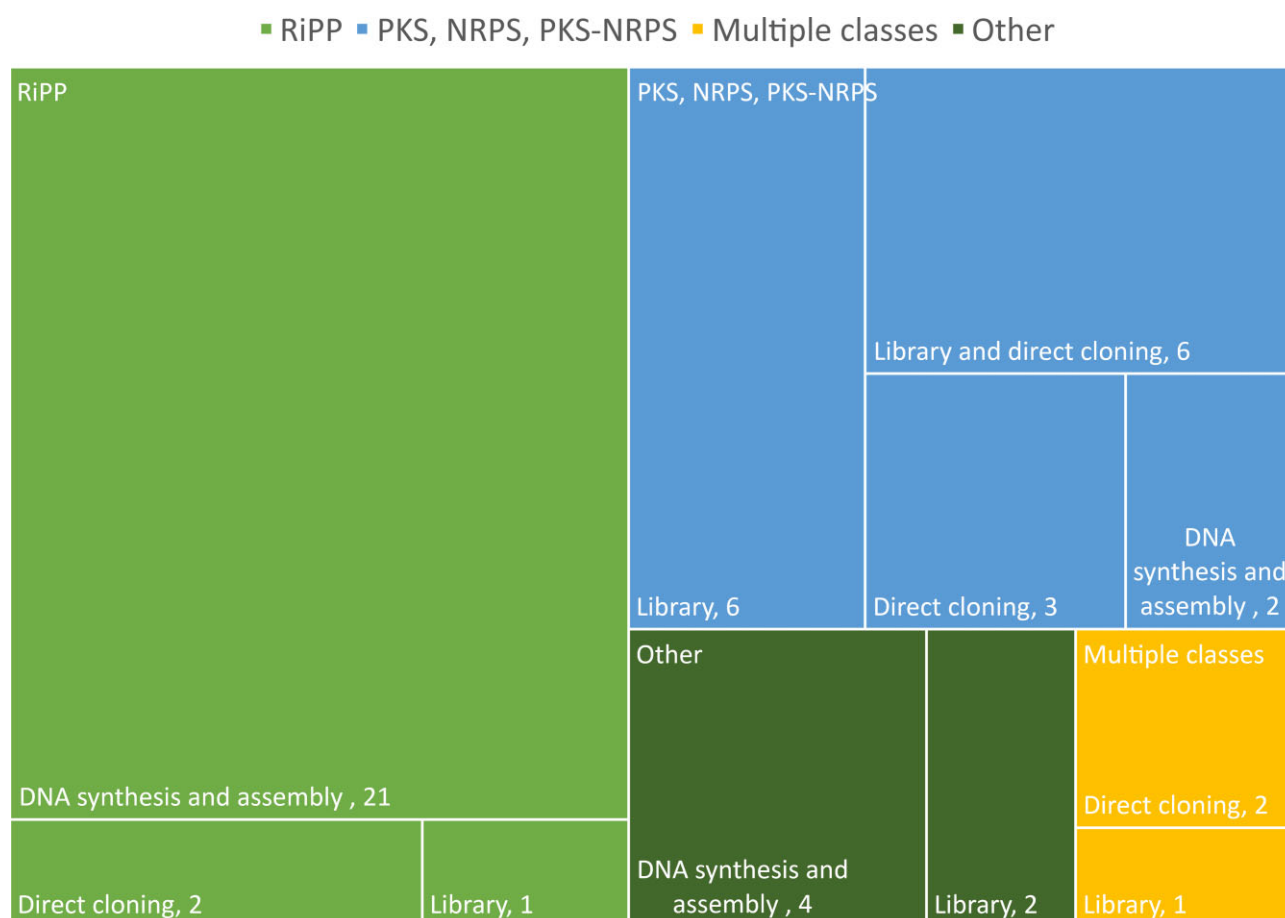
In bacteria, the genes that encode the biosynthesis of a NP are often colocalized in the genome, forming biosynthetic gene clusters (BGCs). Most BGCs have not been connected to a NP and are thus termed orphan. The exploration of orphan BGCs in the native producers or through heterologous expression offers an avenue for discovery. Strategies that have been used to activate gene expression and access the biosynthetic potential of microorganisms in the native producers include variation of the culture conditions, addition of elicitors to the culture media, co-cultivation to replicate the environmental conditions that promote NP production, and genetic approaches as reviewed here (Covington et al., 2021). Genetic approaches can be targeted to a specific NP of interest. Examples include deletion of pathway-specific negative regulators, overexpression of positive regulators, or promoter exchange (Covington et al., 2021). A drawback of genetic engineering of native producers is that tools for genetic manipulation must be developed for each strain of interest. Moreover, native producer centric methods cannot be used for uncultured bacteria.

Alternatively, heterologous expression of orphan BGCs in an established host strain offers great potential for NP discovery (Fig. 1). BGCs that are not well expressed under laboratory conditions can be refactored for activation and NPs from uncultured bacteria or metagenomes can be explored as well. However, some major challenges are BGC prioritization, cloning of the complete BGC, appropriate expression of the BGC in the selected host, and the identification and isolation of products in practical yields.

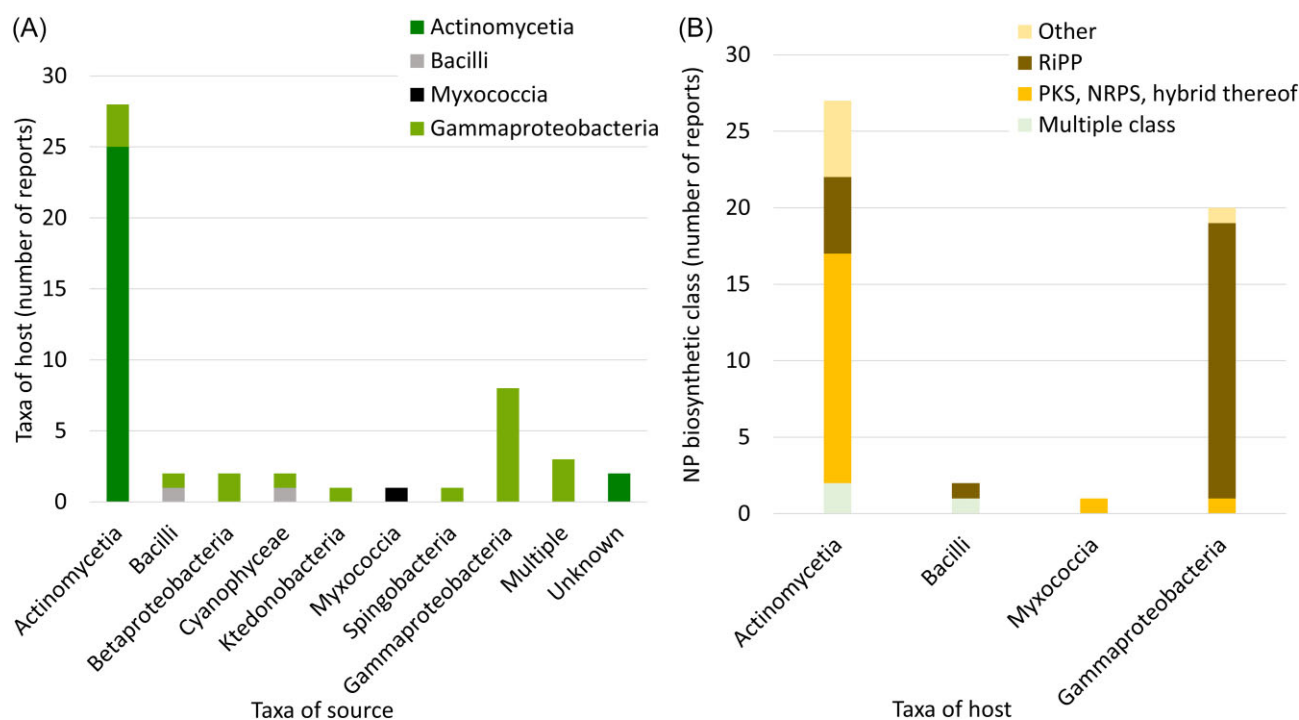
To provide insight on what has worked and potential causes of heterologous expression failure, we searched PubMed and Web of Science for articles published between January 2018 and June 2023 using search terms 'natural product' AND 'heterologous expression' AND 'bacteria'. We also searched with the terms 'biosynthetic gene cluster' AND 'heterologous expression' AND 'bacteria' AND 'discovery'. Additionally, we used the search terms 'genome mining' AND 'heterologous expression' AND 'bacteria'. Because our aim was to focus the review on heterologous expression for NP discovery, studies were excluded if they only reported rediscovery of known NPs, close congeners of known NPs, or NPs previously detected in native producers. Despite careful analysis, we expect we may have missed relevant articles that were not identified using the search terms above and apologize to researchers whose work we inadvertently omitted. Based on the 50 identified articles (Supplemental Table S1), below we discuss the rationale for BGC prioritization, cloning methods, biosynthetic class, source taxa, and host choice (Figs 2–4). We then summarize and discuss large scale studies that have allowed the determination of success rates (Table 1). We conclude with a discussion of remaining



**Fig. 2.** Discovery of bacterial natural products by heterologous expression in the last 5 years. The 50 studies reviewed here are broken down by (A) Rationale for BGC prioritization. (B) Biosynthetic class of prioritized BGCs. (C) Source of BGCs. (D) Heterologous hosts used.



**Fig. 3.** Cloning techniques used for different BGC classes. DNA synthesis refers to DNA synthesis or PCR. Assembly methods include restriction enzymes mediated methods and recombination-based methods. Random libraries were generated using cosmid or BAC vectors.



**Fig. 4.** Metrics regarding source taxa, host taxa, and biosynthetic class. **(A)** Relationship between the taxa of the BGC source (x-axis) and the taxa of the heterologous host (y-axis). Studies that tested more than one source taxa were labeled as multiple. Host taxa (class) are color-coded as indicated. **(B)** Relationship between the taxa of the host used (x-axis) and the biosynthetic class of the expressed BGC (y-axis). NP biosynthetic classes are color coded as indicated.

challenges. Our goal is to obtain insights on approaches used with the hope of informing researchers whose goal is to find natural products from orphan BGCs.

## BGC Prioritization, Biosynthetic Class, and Host Taxa

The first step in genome mining is BGC selection, which includes the rationale for the selected BGC and the identification of the genes to be cloned. To this end, databases containing information about known BGCs such as MIBiG (Terlouw et al., 2023) are very helpful in the race for novel chemistry. However, they remain incomplete because most known NPs have not yet been connected to their BGCs.

Compared to activation of orphan BGCs in their native host, identification of the complete set of genes necessary for NP biosynthesis in a heterologous host is more challenging because (a) genes may be necessary that are not part of the cluster, and (b) cluster boundaries may not be accurately predicted. Detailed information about the principles underlining current bioinformatic tools for BGC prediction can be found in reference (Z. Xu et al., 2023).

For the studies reviewed here, the rationale for BGC selection was based either on (a) predicted structural novelty, (b) biosynthetic class, (c) structure similarity to known antibiotic class, or (d) biological activity of library clone (Fig. 2A). The most frequent approach (56%) was to prioritize novelty by expressing unusual BGCs found in rare and/or understudied bacteria, exemplified by the discovery of the lanthipeptide marinsedin from *Marinicella sedimim* (Han et al., 2022), and by the discovery of 31 other compounds listed in Supplemental Table S1 (Alberti et al., 2019; Bothwell et al., 2021; Cheng et al., 2023; De Rond et al., 2021; Enghiad et al.,

2021; Gao et al., 2023; Gummerlich et al., 2020; Hao et al., 2019; Hashimoto et al., 2021; Kaweewan et al., 2021; Lasch et al., 2021; Lasch, Gummerlich, et al., 2020, 2020; Li et al., 2020; S. H. Liu et al., 2019; Y. Liu et al., 2021; Myronovskiy et al., 2018; Paulus et al., 2022; Shi et al., 2019; Shuai et al., 2020; Unno et al., 2020; Vermeulen et al., 2022; X. Wang et al., 2023; Yuet et al., 2020; Y. Zhang et al., 2021; Y. Zhang et al., 2022).

Eighteen studies (36%) prioritized the expression of a NP class or subclass. For instance, Libis et al. targeted adenylation and ketosynthase domains in a genomic library generated from 100 *Streptomyces* strains to identify nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) BGCs (Libis et al., 2022). Ayikpoe et al. used the bioinformatic tool RODEO (Tietz et al., 2017) to mine phylogenetic diverse bacteria to discover ribosomally synthesized and post translationally modified peptides (RiPPs) (Ayikpoe et al., 2022). Ren et al. used RODEO and RRE-Finder (Kloosterman et al., 2020) to uncover a new RiPP subclass (Ren et al., 2023). Several other studies focused on RiPP subclasses such as lasso peptides (Cao et al., 2021; Carson et al., 2023; Cheung-Lee et al., 2019; Cheung-Lee et al., 2020; Gomez-Escribano et al., 2019; Mevaere et al., 2018), lanthipeptides (Arias-Orozco et al., 2021; Kaweewan et al., 2023; Singh et al., 2020; Thetsana et al., 2022), and thiopeptides (Santos-Aberturas et al., 2019). For further examples, see Supplemental Table S1 (Bösch et al., 2020; J. Liu et al., 2019; Nguyen et al., 2022; Shi et al., 2021).

The third prioritization criterion was based on structure similarity to known antibiotic classes (6%). For example, cadasides and malacidins calcium dependent antibiotics were discovered by searching for NRPS adenylation domains similar to those encoding known calcium dependent antibiotics (Hover et al., 2018; Wu et al., 2019). Likewise, the BGCs encoding glycopeptides GP1416 and GP6738 were selected for expression because of their sequence similarities to known glycopeptide antibiotic BGCs

**Table 1.** Summary of large scale, heterologous expression studies to discover natural products.

| BGC source  | No. of BGCs selected for cloning | No. of BGCs cloned (success rate) | Biosynthetic class | BGC/insert size (kb) | Cloning method                          | Host(s) used  | No. of BGCs expressed (success rate) | No. of NP families isolated | Ref.                      |
|---|----------------------------------|-----------------------------------|--------------------|----------------------|---|---|--------------------------------------|-----------------------------|---------------------------|
| 1 <i>Saccharothrix espanaensis</i>  | 25                               | 17 (68%)                          | Multiple           | 100                  | Random library                          | <i>S. lividans</i> DYA<br><i>S. albus</i> J1074                                     | 4 (11%)                              | 2                           | (Gummerlich et al., 2020) |
| 14 <i>Streptomyces</i> spp.<br>3 <i>Bacillus</i> spp.                                       | 43                               | 43 (100%)                         | Multiple           | 10–113               | CAPTURE                                 | <i>S. avermitilis</i> SUKA17<br><i>S. lividans</i> TK24<br><i>B. subtilis</i> JH642 | 7 (16%)                              | 5                           | (Enghiad et al., 2021)    |
| 100 <i>Streptomyces</i> spp.  | Orphan PKS, NRPS, PKS-NRPS       | 58 (72%)                          | PKS, NRPS          | 140                  | Random library                          | <i>S. albus</i> J1074<br><i>S. lividans</i> RedStrep 1.7                            | 15 (24%)                             | 3                           | (Libis et al., 2022)      |
| 1 Bacteroidota<br>10 Pseudomonadota<br>3 Cyanobacteriota<br>5 Actinomycetota<br>8 Bacillota | 96                               | 83 (86%)                          | RiPPs              | <18                  | Golden Gate assembly of synthetic genes | <i>E. coli</i> BL21 (DE3)   | 27 (32%)                             | 3                           | (Ayikpoe et al., 2022)    |

(M. Xu et al., 2020b). Finally, one study used activity-guided prioritization of library clones. A bacterial artificial chromosome (BAC) library of a *Streptomyces rochei* strain was generated, expressed in *Streptomyces lividans* and active clones were prioritized leading to the discovery of a lanthipeptide (M. Xu et al., 2020a).

In summary, over the last 5 years, at least 63 NP families were discovered and characterized using heterologous expression and reported in 50 studies. A family is here defined as structurally similar NPs isolated after the heterologous expression of a singular BGC. Most of the studies reviewed here focused on RiPPs (48%), followed by PKSs, NRPSs, and hybrid PKS-NRPSs (32%), other biosynthetic classes that included terpenoids, alkaloids, and oxazolones (12%), and multiple classes (8%) (Fig. 2B). Most studies (56%) focused on actinomycetes (Fig. 2C). While actinomycetes are indeed gifted (Gavriliidou et al., 2022), mining of diverse phyla is expected to lead to further chemical diversity (Hegemann et al., 2023).

## Cloning Methods

Three general methods were used to clone BGCs of interest for heterologous expression: DNA chemical synthesis or polymerase chain reaction (PCR) followed by assembly, direct cloning from genomic DNA, and generation of random libraries (Fig. 3). Various assembly methods were used to construct plasmids for expression, to perform promoter engineering, or to obtain complete BGCs from library clones. Information about library generation, direct cloning, and assembly methods can be found in references (Huo et al., 2019; W. Wang et al., 2021). Briefly, random libraries are collections of genomic DNA generated by fragmentation of the DNA followed by cloning of the DNA fragments in BAC or cosmid vectors. Direct cloning involves cloning the BGC of interest directly from genomic DNA but not using PCR. Instead, methods such as transformation associated recombination (TAR) cloning or Cas12a-assisted precise targeted cloning using *in vivo* DNA circularization (CAPTURE) are used. Assembly methods include restriction enzyme mediated methods such as Golden Gate assembly, and recombination-based methods such as TAR cloning in yeast, Red/ET-mediated recombineering in bacteria, and *in vitro* Gibson assembly (Huo et al., 2019; W. Wang et al., 2021).

DNA synthesis is advantageous for simplicity, if mutations are to be introduced, or if codon optimization is required. However, synthesis is practical only for small BGCs such as RiPPs since the maximum size that DNA synthesis suppliers currently offer is 1.8 kb for fragments and 5 kb for clonal DNA. Accordingly, DNA synthesis or PCR to amplify the BGC from genomic DNA was used in 87% of the studies expressing only RiPPs (Arias-Orozco et al., 2021; Ayikpoe et al., 2022; Bösch et al., 2020; Bothwell et al., 2021; Cao et al., 2021; Carson et al., 2023; Cheung-Lee et al., 2019; Cheung-Lee et al., 2020; Gomez-Escribano et al., 2019; Kaweewan et al., 2021, 2023; Koos & Link, 2019; Mevaere et al., 2018; Nguyen et al., 2022; Singh et al., 2020; Thetsana et al., 2022; Unno et al., 2020; Vermeulen et al., 2022; X. Wang et al., 2023; Y. Zhang et al., 2022). TAR cloning, CAPTURE, and random library generation were used in the remaining studies (Ren et al., 2023; Santos-Aberturas et al., 2019; M. Xu et al., 2020a) (Fig. 3).

In contrast, random libraries and direct cloning techniques are often used for cloning of PKS, NRPS, and PKS-NRPS hybrids in line with the larger size of the BGCs compared to RiPPs. In fact, 88% of the studies aiming to express only PKS or NRPS used random library generation and or direct cloning, often in combination with assembly methods (Fig. 3) (Alberti et al., 2019; Gao et al., 2023; Hashimoto et al., 2021; Hover et al., 2018; Lasch et al., 2021; Lasch,

Gummerlich, et al., 2020, 2020; Libis et al., 2022; S. H. Liu et al., 2019; Y. Liu et al., 2021; Myronovskyi et al., 2018; Paulus et al., 2022; Shi et al., 2019; Wu et al., 2019; M. Xu et al., 2020b). For example, the libraries generated for the discovery of cadasides, malacindins, and miramides produced cosmids with portions of the desired BGCs. The overlapping pieces of the BGCs were then assembled into a BAC by TAR cloning and integrated into the chromosomes of the hosts (Hover et al., 2018; Paulus et al., 2022; Wu et al., 2019). Moreover, recombination methods like Red/ET can be used to introduce a selective marker (Lasch et al., 2021) or insert integration machinery into the vector (Gummerlich et al., 2020). Studies using random libraries used hosts related to the source organism to ensure the presence of regulatory elements necessary for BGC expression since the BGCs were often not refactored, except for two studies that engineered promoters (Gao et al., 2023; Lasch, Gummerlich, et al., 2020).

In summary, while DNA synthesis is convenient for small BGCs, large BGCs are still technically challenging to obtain through synthesis. Random libraries can accommodate large pieces of DNA; however, much effort must be invested in finding clones that cover the whole BGC of interest. For example, Hashimoto et al. screened 1 536 BAC clones to identify the one leading to the production of the polyketide JBIR-159 (Hashimoto et al., 2021). Hover et al. and Libis et al. generated libraries of 25 000 and 60 000 clones, respectively (Hover et al., 2018; Libis et al., 2022). To facilitate analysis of such libraries, Hover et al. used barcoded primers and profiled the obtained clones with the bioinformatic platform eSNAPD (environmental Surveyor of Natural Product Diversity) (Hover et al., 2018). Libis et al. also used barcoded primers and analyzed their multigenome library with the targeted sequencing workflow CONKAT-seq (co-occurrence network analysis of targeted sequences) that uses co-occurrence patterns in libraries to identify chromosomally clustered domains thus increasing the probability of assembling complete BGCs from complex samples (Libis et al., 2019, 2022). In contrast, direct cloning methods are convenient as they allow targeted cloning of a BGC of interest but some, such as TAR cloning, have the limitation of not being suitable for BGCs that contain repetitive DNA sequences due to unwanted recombination that can result in BGC deletions. Cas-assisted excision from genomic DNA followed by *in vitro* assembly with a vector has been recently used to circumvent the drawbacks of TAR cloning (Enghiad et al., 2021; Montaser & Kelleher, 2020; J.-W. Wang et al., 2015).

## Host Choice

After cloning the BGC of interest, transferring it into a suitable host is necessary for optimal yields. The heterologous host must be genetically tractable and ideally, easy to culture. *Streptomyces* spp. (Actinomycetia) were used as hosts in 54% of the studies (Fig. 2D) matching the source of the BGCs in most cases (Fig. 4A). Indeed, phylogeny relatedness was mentioned by Hao et al. and four other studies as the reason for host choice (Hao et al., 2019; Mevaere et al., 2018; Shi et al., 2021; X. Wang et al., 2023; M. Xu et al., 2020b). Phylogeny as a choice is backed up by a recent study showing that the yield of heterologously expressed NPs is often higher when the host is closely related to the native strain (G. Wang et al., 2019), although exceptions exist (J. J. Zhang et al., 2017). Even though not specifically mentioned as reason, strains related to the source organism were used as host for expression in 29 other studies (Alberti et al., 2019; Carson et al., 2023; Cheng et al., 2023; Enghiad et al., 2021; Gao et al., 2023; Gomez-Escribano et al., 2019; Gummerlich et al., 2020; Han et al., 2022; Hashimoto

et al., 2021; Kaweewan et al., 2021; Lasch et al., 2021; Lasch, Gummerlich, et al., 2020, 2020; Li et al., 2020; Libis et al., 2022; J. Liu et al., 2019; S. H. Liu et al., 2019; Y. Liu et al., 2021; Myronovskiy et al., 2018; Nguyen et al., 2022; Paulus et al., 2022; Ren et al., 2023; Santos-Aberturas et al., 2019; Shi et al., 2019; Shuai et al., 2020; Thetsana et al., 2022; Unno et al., 2020; Vermeulen et al., 2022; M. Xu et al., 2020a; Y. Zhang et al., 2021). In total, 35 studies (70%) used hosts that fall in the same class as the source BGC strain.

When phylogeny was not taken into consideration, the model organism *Escherichia coli* was used as host (Fig. 2D and Fig. 4A). Twenty studies (40%) used *E. coli* as host (Fig. 2D) (Ayikpoe et al., 2022; Bösch et al., 2020; Bothwell et al., 2021; Cao et al., 2021; Carson et al., 2023; Cheung-Lee et al., 2019; Cheung-Lee et al., 2020; De Rond et al., 2021; Han et al., 2022; Kaweewan et al., 2021, 2023; Koos & Link, 2019; Nguyen et al., 2022; Singh et al., 2020; Thetsana et al., 2022; Unno et al., 2020; Vermeulen et al., 2022; X. Wang et al., 2023; Yuet et al., 2020; Y. Zhang et al., 2022). All but two (De Rond et al., 2021; Yuet et al., 2020) of these 20 studies expressed RiPP BGCs (Fig. 4B). Codon optimization was performed during the cloning step in three of the studies (Ayikpoe et al., 2022; Carson et al., 2023; Yuet et al., 2020). Moreover, to improve yields of heterologously produced lanthipeptides in *E. coli*, co-expression with tRNA-Glu and glutamyl-tRNA synthetase from the source organisms was used (Bothwell et al., 2021; Kaweewan et al., 2023).

In contrast, 98% of studies exploring PKS, NRPS, and other classes utilized *Streptomyces* hosts (Fig. 4B) matching the source of the BGCs in most cases. Other hosts explored in the covered literature included *Myxococcus xanthus*, *Lactococcus lactis*, and *Streptococcus mutans*. *Myxococcus xanthus* was used to express a refactored BGC from the closely related *Sorangineae* sp. strain MSr11367 (Gao et al., 2023). Arias-Orozco et al. chose *L. lactis* as a host to facilitate NP purification given the accumulation of peptides in inclusion bodies (Arias-Orozco et al., 2021). Hao et al. developed *S. mutans* UA159 as a host and used it to discover mutanocyclin from human oral bacteria. The natural competence system of *S. mutans* facilitates BGC transfer (Hao et al., 2019).

In summary, *Streptomyces* spp. (Actinomycetia) and *E. coli* (Gammaproteobacteria) were the most frequently used hosts (54% and 40% of studies, respectively). The class of the host matched the source of the expressed BGC in 70% of studies with PKS, NRPS, and hybrid PKS-NRPS BGCs, which were from Actinomycetia and expressed in *Streptomyces* strains, whereas RiPPs from various sources were primarily expressed in *E. coli*.

## Success Rate of Natural Product Discovery by Heterologous Expression

With the increased sophistication of bioinformatic tools and cloning techniques, heterologous expression has become more attractive for NP discovery. From the surveyed literature in the last 5 years, we identified 50 studies reporting NP discovery by heterologous expression. Because only successful attempts are usually reported, it is impossible to predict success rate based on small scale studies. Fortunately, four large-scale studies from which success rates can be derived were included in the 50 studies reviewed here (Table 1) (Ayikpoe et al., 2022; Enghiad et al., 2021; Gummerlich et al., 2020; Libis et al., 2022). We summarize these four studies below in chronological order of publication.

Gummerlich et al. used a random library approach to express BGCs from *Saccharothrix espanaensis* with low similarity to known compounds. Of the 31 BGCs predicted by antiSMASH, six were ex-

cluded because they were predicted to encode known NPs or congeners of known NPs. After a BAC library generation, 15 BACs covering 17 of the remaining 25 BGCs were expressed in *S. albus* J1074 and *S. lividans*  $\Delta$ YA6 strains with 11% expression success rate. Of the four detected products, two were produced in sufficient quantity for isolation. (Gummerlich et al., 2020).

Enghiad et al. designed a Cas12a-assisted approach termed CAPTURE for direct cloning of large BGCs from genomic DNA. Briefly, the targeted BGC is digested from genomic DNA with Cas12a, assembled *in vitro* with two DNA receivers containing either an origin of replication or a selection marker, and the generated linear fragment is circularized by Cre-lox recombination *in vivo*. Using CAPTURE, Enghiad et al. cloned 43 orphan BGCs from *Streptomyces* spp. and *Bacillus* spp. ranging in size from 10 to 113 kb. (Enghiad et al., 2021). BGCs from *Streptomyces* were expressed in *Streptomyces avermitilis* and *Streptomyces lividans* whereas BGCs from *Bacillus* were introduced in *Bacillus subtilis*. HPLC peaks were observed for seven BGCs (all from *Streptomyces*) giving a 16% success rate for NP detection. Five out of the seven were produced in sufficient quantities for structural characterization. After investigating the expression of BGCs without products, the authors found that 60% of those BGCs had low to no detectable RNA in the culture condition tested. They suggested that further refactoring of the BGCs by promoter engineering, expression of positive regulators or deletion of repressors could improve the success rate (Enghiad et al., 2021).

Next, Libis et al. generated a DNA library from 100 pooled *Streptomyces* strains. The generated library with an average insert size of 140 kb was analyzed using a targeted sequence workflow to identify clones with complete BGCs. The authors estimated a 72% cloning success for complete PKS and NRPS BGCs. They then selected 58 orphan BGCs that were expressed in *S. albus* J1074 and *S. lividans* RedStrep 1.7. Fifteen out of the 58 orphan BGCs produced a differential mass spectral feature compared to the control, one of which was a known compound, giving a 24% success rate for NP detection. Three out of the 14 new NPs were isolated for structural characterization (Libis et al., 2022). The authors noted the advantage of expressing the BGCs in different hosts to improve success rate as there was only partial overlap regarding which BGC was expressed in which host (only 36% expressed in both hosts).

Finally, Ayikpoe et al. (Ayikpoe et al., 2022) attempted cloning and expression of 96 RiPP BGCs. They used Golden Gate assembly of two to nine synthetic genes into BGCs of <18 kb in size. With up to five genes, the assembly success rate was 100%, decreasing thereafter. Overall, they achieved 86% cloning success with 83 of the 96 RiPPs successfully cloned. During the cloning step, the authors refactored the targeted RiPP BGCs by promoter replacement, *E. coli* codon optimization, and incorporation of a histidine tag (except for the lasso peptides and thiopeptides) to facilitate NP isolation. After expression in *E. coli* BL21, 27 of the 83 cloned BGCs produced mass features corresponding to the modified peptide, giving a 32% success rate. The authors noted that classes of RiPPs for which modifying enzymes have not been reconstituted in *E. coli* were not expressed. Of the 30 peptides detected by mass spectrometry, six were tested for activity, and three bioactive peptides were structurally characterized (Ayikpoe et al., 2022).

Taken together, 11% to 32% of the cloned BGCs were successfully expressed. Ayikpoe et al. had the highest success rate at 32%. For the other three studies, further refactoring of the BGCs by promoter exchange could have improved the success rate as indicated by the authors. Moreover, factors such as NP toxicity, codon bias and differential regulatory network may also help explain the low success rate.

## Conclusions and Remaining Challenges

From January 2018 to June 2023, at least 50 studies were published that used heterologous expression for NP discovery. Combined, these 50 studies led to the discovery of 63 new families of natural products (Fig. 2). Most of the studies (56%) prioritized BGCs based on structural novelty, followed by biosynthetic class (36%). The main biosynthetic classes prioritized were RiPPs (48%) and PKS, NRPS, and hybrids thereof (32%). This trend reflects the bioinformatics tools available that are tailored to these classes. In fact, of the 33 tools available to predict BGCs reviewed by Xu *et al.*, nine are able to predict only RiPPs, seven are able to predict PKS, NRPS, and hybrid PKS-NRPS BGCs, and the remaining 15 tools can predict multiple types of biosynthetic pathways (Z. Xu *et al.*, 2023). Moreover, RiPPs are more amenable to heterologous expression due to the relatively smaller size of their BGCs compared to PKS and NRPS BGCs.

Actinomycetia genomes were mined most frequently (56%) followed by Gammaproteobacteria (16%). It has been shown that three-fourths of gene cluster families are unique to each bacterial phylum and that diversity drops at each taxonomic rank (Gavriliidou *et al.*, 2022). Thus, exploration of understudied taxa is expected to result in new NPs, expanding the known NP chemical space. However, new approaches will need to be developed to enable access to BGCs from understudied bacteria, such as cultivation conditions to facilitate isolation of specific taxa and a collection of versatile host strains to enable the expression of BGCs from varied sources.

Cloning methods used in the covered literature depended on BGC size, which is correlated with biosynthetic class (Fig. 3). RiPP BGCs tend to be smaller and DNA synthesis or PCR was used most frequently (87%), whereas for large PKS and NRPS BGCs random libraries and direct cloning were the methods of choice. Newer cloning methods are being developed, such as Cas-assisted excision from genomic DNA followed by *in vitro* assembly with a vector, which has been used to circumvent the drawbacks of homologous recombination-based direct cloning (Enghiad *et al.*, 2021; Montaser & Kelleher, 2020; J.-W. Wang *et al.*, 2015).

The main hosts (Fig. 2D) used were *Streptomyces* spp. (54%) and *E. coli* (40%). Most studies (70%) used a host that falls in the same class as the source BGC strain. When phylogeny was not taken into consideration, *E. coli* was the host of choice (Fig. 4A). There was a correlation between host choice and biosynthetic class. For example, 90% of the studies expressing RiPPs from various sources used *E. coli* as host, whereas 98% of the studies targeting PKS, NRPS, and other classes used *Streptomyces* spp. matching the source of the BGCs in most cases (Fig. 4B).

A common question with heterologous expression is whether the heterologously identified compounds are the ones that would have been detected in the native producers. This question cannot always be answered because the biosynthetic gene cluster may be unexpressed in the native host. In fact, eight studies used heterologous expression because they were unable to find the NP associated to their BGC of interest in the native host (Bösch *et al.*, 2020; Bothwell *et al.*, 2021; Cheng *et al.*, 2023; Han *et al.*, 2022; Li *et al.*, 2020; J. Liu *et al.*, 2019; Shi *et al.*, 2019; X. Wang *et al.*, 2023). Only two other of the covered papers were indeed able to identify the heterologous products in the native producers (Gao *et al.*, 2023; J. Liu *et al.*, 2019). Conversely, fralnimycin, isolated after expression of an aminocyclitol-like BGC from *Frankia alni* in *S. albus* Del14 likely derives from the incorporation of a precursor from the host. Fralnimycin is biosynthesized via esterification of salicylic acid and tryptophol. However, no genes responsible for the

biosynthesis of tryptophol were found in the heterologously expressed fosmid (Myronovskiy *et al.*, 2018).

Heterologous expression is made possible by bioinformatics tools for genome mining, biosynthetic knowledge base, improved cloning techniques, host development, improved isolation, and detection of the desired NP (Avalon *et al.*, 2022; Caesar *et al.*, 2021; Huo *et al.*, 2019; J. Liu *et al.*, 2020) (Fig. 1). The studies covered here support heterologous expression as a promising way to access novel chemistry. In particular, the recently reported large-scale studies are significant and instrumental in revealing current limitations for the field to address. The low success rate of large-scale studies makes it apparent that much remains to be improved to allow us to seamlessly go from DNA to natural products using heterologous expression.

The reasons for failure to be considered are numerous. Ayikpoe *et al.* considered three factors that can lead to failure and addressed those factors with their design, i.e., gene expression, product toxicity, and product purification, which likely explain their higher success rate compared to the other large-scale studies (Table 1). First, gene expression was addressed with codon optimization for the host of choice, and the synthetic genes were placed under a promoter known to work well in the host. Second, toxicity was avoided by producing inactive precursors that were then converted to the final products *in vitro*. Third, purification was facilitated by inserting a His-tag. Their design may explain the higher success rate compared to the other studies. Yet, biosynthetic class also plays a role as some of the strategies used are applicable to RiPPs but not to other classes. For example, the strategy to circumvent potential toxicity was to exclude the protease gene because products containing the leader peptide are expected to be inactive. *In vitro* enzymatic cleavage of the leader peptides was then used to obtain the final products. This strategy also allowed the insertion of a His-Tag at the N-terminus of the precursor peptide, which can then be removed during leader peptide cleavage. This approach is very elegant and works well for RiPPs (with lasso peptides and thiopeptides as exceptions) but not for other biosynthetic classes. Including resistance genes is an alternative strategy that is applicable to all biosynthetic classes, if resistance genes are associated with the BGC. Incidentally, resistance genes can also be used as a prioritization rationale to identify antibiotics (Yan *et al.*, 2020).

We next discuss four overall factors that can affect success, some of which are easier to address than others, that is, gene expression, host choice, incomplete BGCs, and sequence errors.

Insufficient gene expression is a major hurdle. Gene expression can be improved with refactoring using well-characterized promoters and codon optimization, as done by Ayikpoe *et al.* However, refactoring alone is not sufficient as evidenced by the 32% success rate. Because different bacteria have different codon preferences, it is assumed that optimizing the sequence by using codons that are common in the host, improves expression. However, codon optimization is not always successful (Jenkins *et al.*, 2023). Moreover, while protein yield may be improved with codon optimization, protein function may be reduced (Arsin *et al.*, 2021). It has been shown that many genes contain rare codons interspaced among common codons and that rare codons are important to slow down translation and allow proper protein folding (Y. Liu, 2020). Thus, *codon harmonization*—reproducing the order of common and rare codons for the host of interest—can lead to improved enzyme activity (Angov *et al.*, 2008; Arsin *et al.*, 2021; Y. Liu, 2020; Mignon *et al.*, 2018). A new open source tool for codon optimization/harmonization became available during revision of this article (Schmidt *et al.*, 2023). We also speculate that any



recoding (even codon harmonization) may lead to detrimental, context-dependent transcription and translation effects (Biziaev et al., 2022; Jiang et al., 2023; Kent et al., 2018) that cannot yet be predicted and avoided.

With a host that is phylogenetically related to the source DNA, codon optimization/harmonization can be avoided, and regulatory elements are expected to be better recognized by the host. In fact, 70% of the studies selected phylogenetically related hosts (Fig. 4A). By expressing nine BGCs in 25 hosts, Wang et al. (G. Wang et al., 2019) demonstrated that success rate in terms of yield and number of congeners detected tends to improve with increasing relatedness (defined as 16S rRNA sequence identity) between source and host. Yet, some versatile host strains were identified that had higher success rates (produced more natural products), outperforming others despite lower relatedness. Thus, host selection is an important part of heterologous expression pipelines. It seems that the highest success rates would be achieved when using many host strains rather than only one. Because increasing the number of hosts increases complexity and costs, it is then advantageous to select and develop several versatile strains to be added to the heterologous expression toolbox. Versatile strains may also help address other potential reasons for failure, such as missing precursors and incorrect folding of proteins.

Addressing incomplete BGCs remains a challenge because of split clusters and because determination of the boundaries requires experimental validation. To help define the DNA region to be cloned, algorithms have been developed for automatic prediction of cluster boundaries (Blin et al., 2017) and boundaries can be estimated by comparing BGCs from multiple strains (Adamek et al., 2017). For rare and unusual BGCs, estimating the boundaries becomes more difficult. In this respect, being generous regarding the number of surrounding genes to be included can pay off. At the same time, increasing the size of the predicted BGC also makes cloning more difficult. Finally, for synthetic DNA, any errors in the original sequence data will carry over into synthetic constructs; thus, the quality of the source genome sequence is important.

Given all the potential reasons for failure, some of which are difficult to address such as incomplete BGCs, aiming for 100% heterologous expression success rate for unknown BGCs appears unrealistic as of 2023. Yet, the studies reviewed here point to a combination of approaches that should be used to improve the chances of success, including promoter replacement to ensure transcription, addressing product toxicity, and testing various, versatile hosts. We expect these combined approaches may help double the current best success rate of 32%. In fact, by testing different hosts, Wang et al. (G. Wang et al., 2019) reached 67% success rate. Other yet unknown or unexplored factors may help improve the success rate further in the future.

## Supplementary Material

Supplementary material is available online at *JIMB* ([www.academic.oup.com/jimb](http://www.academic.oup.com/jimb)).

## Funding

This work was supported by the Directorate of Biological Sciences of the National Science Foundation (MCB-2237551 to A. S. E.) and by the National Institute of General Medical Sciences (GM129344 to A.S.E.), National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- Adamek, M., Spohn, M., Stegmann, E., & Ziemert, N. (2017). Mining bacterial genomes for secondary metabolite gene clusters. *Methods in Molecular Biology* (Clifton, N.J.), 1520, 23–47. [https://doi.org/10.1007/978-1-4939-6634-9\\_2](https://doi.org/10.1007/978-1-4939-6634-9_2)
- Alberti, F., Leng, D. J., Wilkening, I., Song, L., Tosin, M., & Corre, C. (2019). Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery. *Chemical Science*, 10(2), 453–463. <https://doi.org/10.1039/C8SC03814G>
- Angov, E., Hillier, C. J., Kincaid, R. L., & Lyon, J. A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE*, 3(5), e2189. <https://doi.org/10.1371/journal.pone.0002189>
- Arias-Orozco, P., Inklaar, M., Lanooij, J., Cebrián, R., & Kuipers, O. P. (2021). Functional expression and characterization of the highly promiscuous lanthipeptide synthetase SyncM, enabling the production of lanthipeptides with a broad range of ring topologies. *ACS Synthetic Biology*, 10(10), 2579–2591. <https://doi.org/10.1021/acssynbio.1c00224>
- Arsin, H., Jasilionis, A., Dahle, H., Sandaa, R.-A., Stokke, R., Nordberg Karlsson, E., & Steen, I. H. (2021). Exploring codon adjustment strategies towards *Escherichia coli*-based production of viral proteins encoded by HTH1, a novel prophage of the marine bacterium *hypnocyclicus thermotrophus*. *Viruses*, 13(7), 1215. <https://doi.org/10.3390/v13071215>
- Avalon, N. E., Murray, A. E., & Baker, B. J. (2022). Integrated metabolomic-genomic workflows accelerate microbial natural product discovery. *Analytical chemistry*, 94(35), 11959–11966. <https://doi.org/10.1021/acs.analchem.2c02245>
- Ayikpoe, R. S., Shi, C., Battiste, A. J., Eslami, S. M., Ramesh, S., Simon, M. A., Bothwell, I. R., Lee, H., Rice, A. J., Ren, H., Tian, Q., Harris, L. A., Sarksian, R., Zhu, L., Frerk, A. M., Precord, T. W., Van Der Donk, W. A., Mitchell, D. A., & Zhao, H. (2022). A scalable platform to discover antimicrobials of ribosomal origin. *Nature Communications*, 13(1), 6135. <https://doi.org/10.1038/s41467-022-33890-w>
- Bachmann, B. O., Van Lanen, S. G., & Baltz, R. H. (2014). Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *Journal of Industrial Microbiology and Biotechnology*, 41(2), 175–184. <https://doi.org/10.1007/s10295-013-1389-9>
- Baltz, R. H. (2017). Gifted microbes for genome mining and natural product discovery. *Journal of Industrial Microbiology and Biotechnology*, 44(4–5), 573–588. <https://doi.org/10.1007/s10295-016-1815-x>
- Biziaev, N., Sokolova, E., Yanvarev, D. V., Toropygin, I. Y., Shuvalov, A., Egorova, T., & Alkalaeva, E. (2022). Recognition of 3' nucleotide context and stop codon readthrough are determined during mRNA translation elongation. *The Journal of biological chemistry*, 298(7), 102133. <https://doi.org/10.1016/j.jbc.2022.102133>
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, M. H. (2017). antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, 45(W1), W36–W41. <https://doi.org/10.1093/nar/gkx319>

- Bösch, N. M., Borsa, M., Greczmiel, U., Morinaka, B. I., Gugger, M., Oxenius, A., Vagstad, A. L., & Piel, J. (2020). Landornamides: Antiviral ornithine-containing ribosomal peptides discovered through genome mining. *Angewandte Chemie (International ed. in English)*, 59(29), 11763–11768. <https://doi.org/10.1002/anie.201916321>
- Bothwell, I. R., Caetano, T., Sarkisian, R., Mendo, S., & Van Der Donk, W. A. (2021). Structural analysis of class I lanthipeptides from *Pedobacter lusitanus* NL19 reveals an unusual ring pattern. *ACS Chemical Biology*, 16(6), 1019–1029. <https://doi.org/10.1021/acscchembio.1c00106>
- Caesar, L. K., Montaser, R., Keller, N. P., & Kelleher, N. L. (2021). Metabolomics and genomics in natural products research: Complementary tools for targeting new chemical entities. *Natural Product Reports*, 38(11), 2041–2065. <https://doi.org/10.1039/D1NP00036E>
- Cao, L., Beiser, M., Koos, J. D., Orlova, M., Elashal, H. E., Schröder, H. V., & Link, A. J. (2021). Cellulonodin-2 and Lihuanodin: Lasso peptides with an aspartimide post-translational modification. *Journal of the American Chemical Society*, 143(30), 11690–11702. <https://doi.org/10.1021/jacs.1c05017>
- Carson, D. V., Patiño, M., Elashal, H. E., Cartagena, A. J., Zhang, Y., Whitley, M. E., So, L., Kayser-Browne, A. K., Earl, A. M., Bhattacharyya, R. P., & Link, A. J. (2023). Cloacaenodin, an antimicrobial lasso peptide with activity against *Enterobacter*. *ACS Infectious Diseases*, 9(1), 111–121. <https://doi.org/10.1021/acsinfectdis.2c00446>
- Cheng, Z., Zhang, Q., Peng, J., Zhao, X., Ma, L., Zhang, C., & Zhu, Y. (2023). Genomics-driven discovery of benzoxazole alkaloids from the marine-derived micromonospora sp. SCSIO 07395. *Molecules (Basel, Switzerland)*, 28(2), 821. <https://doi.org/10.3390/molecules28020821>
- Cheung-Lee, W. L., Cao, L., & Link, A. J. (2019). Pandonodin: A proteobacterial lasso peptide with an exceptionally long C-terminal tail. *ACS Chemical Biology*, 14(12), 2783–2792. <https://doi.org/10.1021/acscchembio.9b00676>
- Cheung-Lee, W. L., Parry, M. E., Zong, C., Cartagena, A. J., Darst, S. A., Connell, N. D., Russo, R., & Link, A. J. (2020). Discovery of Ubonodin, an antimicrobial lasso peptide active against members of the *Burkholderia cepacia* Complex. *Chembiochem*, 21(9), 1335–1340. <https://doi.org/10.1002/cbic.201900707>
- Covington, B. C., Xu, F., & Seyedsayadost, M. R. (2021). A natural product Chemist's guide to unlocking silent biosynthetic gene clusters. *Annual Review of Biochemistry*, 90(1), 763–788. <https://doi.org/10.1146/annurev-biochem-081420-102432>
- De Rond, T., Asay, J. E., & Moore, B. S. (2021). Co-occurrence of enzyme domains guides the discovery of an oxazolone synthetase. *Nature Chemical Biology*, 17(7), 794–799. <https://doi.org/10.1038/s41589-021-00808-4>
- Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S. K., Martin, T. A., & Zhao, H. (2021). Cas12a-assisted precise targeted cloning using in vivo cre-lox recombination. *Nature Communications*, 12(1), 1171. <https://doi.org/10.1038/s41467-021-21275-4>
- Gao, Y., Walt, C., Bader, C. D., & Müller, R. (2023). Genome-guided discovery of the myxobacterial thiolactone-containing sorangibactins. *ACS Chemical Biology*, 18(4), 924–932. <https://doi.org/10.1021/acscchembio.3c00063>
- Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., & Ziemert, N. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nature Microbiology*, 7(5), 726–735. <https://doi.org/10.1038/s41564-022-01110-2>
- Gomez-Escribano, J. P., Castro, J. F., Razmilic, V., Jarmusch, S. A., Saalbach, G., Ebel, R., Jaspars, M., Andrews, B., Asenjo, J. A., & Bibb, M. J. (2019). Heterologous expression of a cryptic gene cluster from *Streptomyces leeuwenhoekii* C34<sup>T</sup> yields a novel lasso peptide, leepeptin. *Applied and Environmental Microbiology*, 85(23), e01752–19. <https://doi.org/10.1128/AEM.01752-19>
- Gummerlich, N., Rebets, Y., Paulus, C., Zapp, J., & Luzhetskyy, A. (2020). Targeted genome mining—From compound discovery to biosynthetic pathway elucidation. *Microorganisms*, 8(12), 2034. <https://doi.org/10.3390/microorganisms8122034>
- Han, Y., Wang, X., Zhang, Y., & Huo, L. (2022). Discovery and characterization of Marinisedin, a new class II lanthipeptide derived from marine bacterium *Marinicella sediminis* F2<sup>T</sup>. *ACS Chemical Biology*, 17(4), 785–790. <https://doi.org/10.1021/acscchembio.2c00021>
- Hao, T., Xie, Z., Wang, M., Liu, L., Zhang, Y., Wang, W., Zhang, Z., Zhao, X., Li, P., Guo, Z., Gao, S., Lou, C., Zhang, G., Merritt, J., Horsman, G. P., & Chen, Y. (2019). An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters. *Nature Communications*, 10(1), 3665. <https://doi.org/10.1038/s41467-019-11673-0>
- Hashimoto, T., Hashimoto, J., Kagaya, N., Nishimura, T., Suenaga, H., Nishiyama, M., Kuzuyama, T., & Shin-ya, K. (2021). A novel oxazole-containing tetraene compound, JBIR-159, produced by heterologous expression of the cryptic trans-AT type polyketide synthase biosynthetic gene cluster. *The Journal of Antibiotics*, 74(5), 354–358. <https://doi.org/10.1038/s41429-021-00410-9>
- Hegemann, J. D., Birkelbach, J., Walesch, S., & Müller, R. (2023). Current developments in antibiotic discovery: Global microbial diversity as a source for evolutionary optimized anti-bacterials. *EMBO Reports*, 24(1), e56184. <https://doi.org/10.15252/embr.202256184>
- Hover, B. M., Kim, S.-H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., Maniko, J., Estrela, A. B., Molina, H., Park, S., Perlin, D. S., & Brady, S. F. (2018). Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant gram-positive pathogens. *Nature Microbiology*, 3(4), 415–422. <https://doi.org/10.1038/s41564-018-0110-1>
- Huo, L., Hug, J. J., Fu, C., Bian, X., Zhang, Y., & Müller, R. (2019). Heterologous expression of bacterial natural product biosynthetic pathways. *Natural Product Reports*, 36(10), 1412–1436. <https://doi.org/10.1039/C8NP00091C>
- Jenkins, M. C., Parker, C., O'Brien, C., Campos, P., Tucker, M., & Miska, K. (2023). Effects of codon optimization on expression in *Escherichia coli* of protein-coding DNA sequences from the protozoan *Eimeria*. *Journal of Microbiological Methods*, 211, 106750. <https://doi.org/10.1016/j.mimet.2023.106750>
- Jiang, Y., Neti, S. S., Sitarik, I., Pradhan, P., To, P., Xia, Y., Fried, S. D., Booker, S. J., & O'Brien, E. P. (2023). How synonymous mutations alter enzyme structure and function over long timescales. *Nature Chemistry*, 15(3), 308–318. <https://doi.org/10.1038/s41557-022-01091-z>
- Kaweewan, I., Ijichi, S., Nakagawa, H., & Kodani, S. (2023). Heterologous production of new lanthipeptides hazakensins A and B using a cryptic gene cluster of the thermophilic bacterium *Thermosporothrix hazakensis*. *World Journal of Microbiology and Biotechnology*, 39(1), 30. <https://doi.org/10.1007/s11274-022-03463-6>
- Kaweewan, I., Nakagawa, H., & Kodani, S. (2021). Heterologous expression of a cryptic gene cluster from *Marinomonas fungiae* affords a novel tricyclic peptide marinomonasin. *Applied Microbiology and Biotechnology*, 105(19), 7241–7250. <https://doi.org/10.1007/s00253-021-11545-y>
- Kent, R., Halliwell, S., Young, K., Swainston, N., & Dixon, N. (2018). Rationalizing context-dependent performance of dynamic RNA regulatory devices. *ACS Synthetic Biology*, 7(7), 1660–1668. <https://doi.org/10.1021/acssynbio.8b00041>

- Kloosterman, A. M., Shelton, K. E., Van Wezel, G. P., Medema, M. H., & Mitchell, D. A. (2020). RRE-finder: A genome-mining tool for class-independent RiPP discovery. *mSystems*, 5(5), e00267–20. <https://doi.org/10.1128/mSystems.00267-20>
- Koos, J. D. & Link, A. J. (2019). Heterologous and *in vitro* reconstitution of Fuscanodin, a lasso peptide from *thermobifida fusca*. *Journal of the American Chemical Society*, 141(2), 928–935. <https://doi.org/10.1021/jacs.8b10724>
- Lasch, C., Gummerlich, N., Myronovskyi, M., Paluszczak, A., Zapp, J., & Luzhetskyy, A. (2020). Loseolamycins: A group of new bioactive alkylresorcinols produced after heterologous expression of a type III PKS from *micromonospora endolithica*. *Molecules (Basel, Switzerland)*, 25(20), 4594. <https://doi.org/10.3390/molecules25204594>
- Lasch, C., Stierhof, M., Estévez, M. R., Myronovskyi, M., Zapp, J., & Luzhetskyy, A. (2020). Dudomycins: New secondary metabolites produced after heterologous expression of an Nrps cluster from *streptomyces albus* ssp. *Chlorinus* nr1 B-24108. *Microorganisms*, 8(11), 1800. <https://doi.org/10.3390/microorganisms8111800>
- Lasch, C., Stierhof, M., Estévez, M. R., Myronovskyi, M., Zapp, J., & Luzhetskyy, A. (2021). Bonsecamin: A new cyclic pentapeptide discovered through heterologous expression of a cryptic gene cluster. *Microorganisms*, 9(8), 1640. <https://doi.org/10.3390/microorganisms9081640>
- Li, Z., Jiang, Y., Zhang, X., Chang, Y., Li, S., Zhang, X., Zheng, S., Geng, C., Men, P., Ma, L., Yang, Y., Gao, Z., Tang, Y.-J., & Li, S. (2020). Fragrant venezuelaenes A and B with A 5–5–6–7 tetracyclic skeleton: Discovery, biosynthesis, and mechanisms of Central catalysts. *ACS Catalysis*, 10(10), 5846–5851. <https://doi.org/10.1021/acscatal.0c01575>
- Libis, V., Antonovsky, N., Zhang, M., Shang, Z., Montiel, D., Maniko, J., Ternei, M. A., Calle, P. Y., Lemetre, C., Owen, J. G., & Brady, S. F. (2019). Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences. *Nature Communications*, 10(1), 3848. <https://doi.org/10.1038/s41467-019-11658-z>
- Libis, V., MacIntyre, L. W., Mehmood, R., Guerrero, L., Ternei, M. A., Antonovsky, N., Burian, J., Wang, Z., & Brady, S. F. (2022). Multi-plexed mobilization and expression of biosynthetic gene clusters. *Nature Communications*, 13(1), 5256. <https://doi.org/10.1038/s41467-022-32858-0>
- Liu, J., Wu, X., Yao, M., Xiao, W., & Zha, J. (2020). Chassis engineering for microbial production of chemicals: From natural microbes to synthetic organisms. *Current Opinion in Biotechnology*, 66, 105–112. <https://doi.org/10.1016/j.copbio.2020.06.013>
- Liu, J., Xie, X., & Li, S. M. (2019). Guanitrypmycin biosynthetic pathways imply cytochrome P450 mediated regio- and stereospecific guaninyl-transfer reactions. *Angewandte Chemie (International ed. in English)*, 58(33), 11534–11540. <https://doi.org/10.1002/anie.201906891>
- Liu, S. H., Wang, W., Wang, K. B., Zhang, B., Li, W., Shi, J., Jiao, R. H., Tan, R. X., & Ge, H. M. (2019). Heterologous expression of a cryptic giant type I PKS gene cluster leads to the production of ansaseomycin. *Organic Letters*, 21(10), 3785–3788. <https://doi.org/10.1021/acs.orglett.9b01237>
- Liu, Y. (2020). A code within the genetic code: Codon usage regulates co-translational protein folding. *Cell Communication and Signaling*, 18(1), 145. <https://doi.org/10.1186/s12964-020-00642-6>
- Liu, Y., Zhou, H., Shen, Q., Dai, G., Yan, F., Li, X., Ren, X., Sun, Q., Tang, Y.-J., Zhang, Y., & Bian, X. (2021). Discovery of polycyclic macrolide shuangdaolides by heterologous expression of a cryptic *trans*-AT PKS gene cluster. *Organic Letters*, 23(17), 6967–6971. <https://doi.org/10.1021/acs.orglett.1c02589>
- Meuvaere, J., Goulard, C., Schneider, O., Sekurova, O. N., Ma, H., Zirah, S., Afonso, C., Rebuffat, S., Zotchev, S. B., & Li, Y. (2018). An orthogonal system for heterologous expression of actinobacterial lasso peptides in *Streptomyces* hosts. *Scientific Reports*, 8(1), 8232. <https://doi.org/10.1038/s41598-018-26620-0>
- Mignon, C., Mariano, N., Stadthagen, G., Lugari, A., Lagoutte, P., Donnat, S., Chenavas, S., Perot, C., Sodoyer, R., & Werle, B. (2018). Codon harmonization – going beyond the speed limit for protein expression. *FEBS Letters*, 592(9), 1554–1564. <https://doi.org/10.1002/1873-3468.13046>
- Montaser, R. & Kelleher, N. L. (2020). Discovery of the biosynthetic machinery for stravidins, biotin antimetabolites. *ACS Chemical Biology*, 15(5), 1134–1140. <https://doi.org/10.1021/acscchembio.9b00890>
- Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., & Luzhetskyy, A. (2018). Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metabolic Engineering*, 49, 316–324. <https://doi.org/10.1016/j.ymben.2018.09.004>
- Neklesa, T. K., Winkler, J. D., & Crews, C. M. (2017). Targeted protein degradation by PROTACs. *Pharmacology & Therapeutics*, 174, 138–144. <https://doi.org/10.1016/j.pharmthera.2017.02.027>
- Newman, D. J. & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83(3), 770–803. <https://doi.org/10.1021/acs.jnatprod.9b01285>
- Nguyen, N. A., Cong, Y., Hurrell, R. C., Arias, N., Garg, N., Puri, A. W., Schmidt, E. W., & Agarwal, V. (2022). A silent biosynthetic gene cluster from a methanotrophic bacterium potentiates discovery of a substrate promiscuous proteusins cyclodehydratase. *ACS Chemical Biology*, 17(6), 1577–1585. <https://doi.org/10.1021/acscchembio.2c00251>
- Paulus, C., Myronovskyi, M., Zapp, J., Rodríguez Estévez, M., Lopatniuk, M., Rosenkränzer, B., Paluszczak, A., & Luzhetskyy, A. (2022). Miramides A–D: Identification of detoxin-like depsipeptides after heterologous expression of a hybrid NRPS-PKS gene cluster from *Streptomyces mirabilis* Lu17588. *Microorganisms*, 10(9), 1752. <https://doi.org/10.3390/microorganisms10091752>
- Ren, H., Dommaraju, S. R., Huang, C., Cui, H., Pan, Y., Nesic, M., Zhu, L., Sarlah, D., Mitchell, D. A., & Zhao, H. (2023). Genome mining unveils a class of ribosomal peptides with two amino termini. *Nature Communications*, 14(1), 1624. <https://doi.org/10.1038/s41467-023-37287-1>
- Santos-Aberturas, J., Chandra, G., Frattaruolo, L., Lacret, R., Pham, T. H., Vior, N. M., Eyles, T. H., & Truman, A. W. (2019). Uncovering the unexplored diversity of thioamidated ribosomal peptides in actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Research*, 47(9), 4624–4637. <https://doi.org/10.1093/nar/gkz192>
- Schmidt, M., Lee, N., Zhan, C., Roberts, J. B., Nava, A. A., Keiser, L. S., Vilchez, A. A., Chen, Y., Petzold, C. J., Haushalter, R. W., Blank, L. M., & Keasling, J. D. (2023). Maximizing heterologous expression of engineered type I polyketide synthases: Investigating codon optimization strategies. *ACS Synthetic Biology*, 12(11), 3366–3380. <https://doi.org/10.1021/acssynbio.3c00367>
- Shi, J., Xu, X., Liu, P. Y., Hu, Y. L., Zhang, B., Jiao, R. H., Bashiri, G., Tan, R. X., & Ge, H. M. (2021). Discovery and biosynthesis of guanipiperazine from a NRPS-like pathway. *Chemical Science*, 12(8), 2925–2930. <https://doi.org/10.1039/D0SC06135B>
- Shi, J., Zeng, Y. J., Zhang, B., Shao, F. L., Chen, Y. C., Xu, X., Sun, Y., Xu, Q., Tan, R. X., & Ge, H. M. (2019). Comparative genome mining and heterologous expression of an orphan NRPS gene cluster direct the production of ashimides. *Chemical Science*, 10(10), 3042–3048. <https://doi.org/10.1039/C8SC05670F>

- Shuai, H., Myronovskyi, M., Nadmid, S., & Luzhetskyy, A. (2020). Identification of a biosynthetic gene cluster responsible for the production of a new pyrrolopyrimidine natural product—Huimycin. *Biomolecules*, 10(7), 1074. <https://doi.org/10.3390/biom10071074>
- Singh, M., Chaudhary, S., & Sareen, D. (2020). Roseocin, a novel two-component lantibiotic from an actinomycete. *Molecular microbiology*, 113(2), 326–337. <https://doi.org/10.1111/mmi.14419>
- Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J. J., Reitz, Z. L., van Santen, J. A., Selem-Mojica, N., Tørring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S. A. A., Augustijn, H. E., ... Medema, M. H. (2023). MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*, 51(D1), D603–D610. <https://doi.org/10.1093/nar/gkac1049>
- Thetsana, C., Ijichi, S., Kaweewan, I., Nakagawa, H., & Kodani, S. (2022). Heterologous expression of a cryptic gene cluster from a marine proteobacterium *Thalassomonas actiniarum* affords new lanthipeptides *thalassomonasins A and B*. *Journal of Applied Microbiology*, 132(5), 3629–3639. <https://doi.org/10.1111/jam.15491>
- Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H.-C., Zakai, U. I., & Mitchell, D. A. (2017). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology*, 13(5), 470–478. <https://doi.org/10.1038/nchembio.2319>
- Unno, K., Kaweewan, I., Nakagawa, H., & Kodani, S. (2020). Heterologous expression of a cryptic gene cluster from *Grimontia marina* affords a novel tricyclic peptide *grimoviridin*. *Applied Microbiology and Biotechnology*, 104(12), 5293–5302. <https://doi.org/10.1007/s00253-020-10605-z>
- Vermeulen, R., Van Staden, A. D. P., Van Zyl, L. J., Dicks, L. M. T., & Trindade, M. (2022). Unusual class I lanthipeptides from the marine bacteria *Thalassomonas viridans*. *ACS Synthetic Biology*, 11(11), 3608–3616. <https://doi.org/10.1021/acssynbio.2c00480>
- Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y.-M., Robinson, D., Bingol, K., Zhang, Z., Bowen, B., Louie, K., Wang, B., Evans, R., Miyamoto, Y., Cheng, K., Kosina, S., De Raad, M., Silva, L., Luhrs, A., Lubbe, A., ... Yoshikuni, Y. (2019). CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nature Microbiology*, 4(12), 2498–2510. <https://doi.org/10.1038/s41564-019-0573-8>
- Wang, J.-W., Wang, A., Li, K., Wang, B., Jin, S., Reiser, M., & Lockey, R. F. (2015). CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. *Biotechniques*, 58(4), 161–170. <https://doi.org/10.2144/000114261>
- Wang, W., Zheng, G., & Lu, Y. (2021). Recent advances in strategies for the cloning of natural product biosynthetic gene clusters. *Frontiers in Bioengineering and Biotechnology*, 9, 692797. <https://doi.org/10.3389/fbioe.2021.692797>
- Wang, X., Wang, Z., Dong, Z., Yan, Y., Zhang, Y., & Huo, L. (2023). Deciphering the biosynthesis of novel class I lanthipeptides from marine *Pseudoalteromonas* reveals a dehydratase PsfB with dethiolation activity. *ACS Chemical Biology*, 18(5), 1218–1227. <https://doi.org/10.1021/acscchembio.3c00135>
- Wu, C., Shang, Z., Lemetre, C., Ternei, M. A., & Brady, S. F. (2019). Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria. *Journal of the American Chemical Society*, 141(9), 3910–3919. <https://doi.org/10.1021/jacs.8b12087>
- Xu, M., Wang, W., Waglechner, N., Culp, E. J., Guitor, A. K., & Wright, G. D. (2020b). GPAHex-A synthetic biology platform for type IV-V glycopeptide antibiotic production and discovery. *Nature Communications*, 11(1), 5232. <https://doi.org/10.1038/s41467-020-19138-5>
- Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S. X., Lin, S., Deng, Z., & Tao, M. (2020a). Functional genome mining reveals a class V lanthipeptide containing a d-amino acid introduced by an  $F_{420} H_2$ -dependent reductase. *Angewandte Chemie (International ed. in English)*, 59(41), 18029–18035. <https://doi.org/10.1002/anie.202008035>
- Xu, Z., Park, T. J., & Cao, H. (2023). Advances in mining and expressing microbial biosynthetic gene clusters. *Critical reviews in microbiology*, 49(1), 18–37. <https://doi.org/10.1080/1040841X.2022.2036099>
- Yan, Y., Liu, N., & Tang, Y. (2020). Recent developments in self-resistance gene directed natural product discovery. *Natural Product Reports*, 37(7), 879–892. <https://doi.org/10.1039/C9NP00050J>
- Yuet, K. P., Liu, C. W., Lynch, S. R., Kuo, J., Michaels, W., Lee, R. B., McShane, A. E., Zhong, B. L., Fischer, C. R., & Khosla, C. (2020). Complete reconstitution and deorphanization of the 3 MDa nocardiosis-associated polyketide synthase. *Journal of the American Chemical Society*, 142(13), 5952–5957. <https://doi.org/10.1021/jacs.0c00904>
- Zhang, J. J., Tang, X., Zhang, M., Nguyen, D., & Moore, B. S. (2017). Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in gram-negative bacteria. *mBio*, 8(5), e01291–17. <https://doi.org/10.1128/mBio.01291-17>
- Zhang, Y., Hong, Z., Zhou, L., Zhang, Z., Tang, T., Guo, E., Zheng, J., Wang, C., Dai, L., Si, T., & Wang, H. (2022). Biosynthesis of gut-microbiota-derived lantibiotics reveals a subgroup of S8 Family proteases for class III leader removal. *Angewandte Chemie (International ed. in English)*, 61(6), e202114414. <https://doi.org/10.1002/anie.202114414>
- Zhang, Y., Yao, T., Jiang, Y., Li, H., Yuan, W., & Li, W. (2021). Deciphering a cyclodipeptide synthase pathway encoding prenylated indole alkaloids in *Streptomyces leeuwenhoekii*. *Applied and Environmental Microbiology*, 87(11), e02525–20. <https://doi.org/10.1128/AEM.02525-20>