Bacterial natural product discovery by heterologous expression

Adjo E. Kadjo,[a,b] Alessandra S. Eustáquio[a,b*]

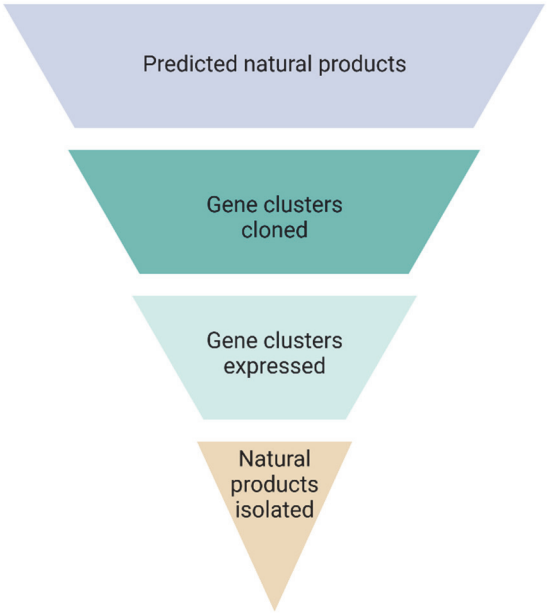
[a]Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, 60607, USA

[b]Center for Biomolecular Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA

To whom correspondence should be addressed: ase@uic.edu

Keywords: Heterologous expression, bacteria, metabolites, natural products, synthetic biology.

Graphical Abstract



One-Sentence Summary: At least 63 new families of bacterial natural products were discovered using heterologous expression in the last five 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 figure.

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 realize bioinformatics predictions into products. However, challenges remain such as gene cluster prioritization, cloning of the complete gene cluster, appropriate expression, product identification, and isolation of products in practical yields. Here we reviewed the literature from the past five 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 identified. 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.

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 and Cragg 2020). Although the structural complexity of NPs allows for effective target interactions, it also prevents easy synthetic access. Moreover, with increased resistance to known molecules, it is necessary to discover new compounds.

Advances in DNA sequencing and bioinformatics have revealed the untapped NP potential of microorganisms (Bachmann, Van Lanen, and Baltz 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), in part because many NPs are not produced in quantities that facilitate discovery and development (Baltz 2017). The exploration of orphan biosynthetic gene clusters (BGCs) offers an avenue for discovery.

Strategies that have been used to activate gene expression and access the biosynthetic potential of microorganisms 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). While genetic engineering of native producers can also give information about NP biosynthesis and ecological roles, tools for genetic manipulation must be developed for each strain of interest. Finally, native producer centric methods cannot be used for uncultured bacteria.

Heterologous expression of uncharacterized biosynthetic gene clusters (BGCs) in an established host offers great potential for NP discovery. 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 (**Figure 1**).

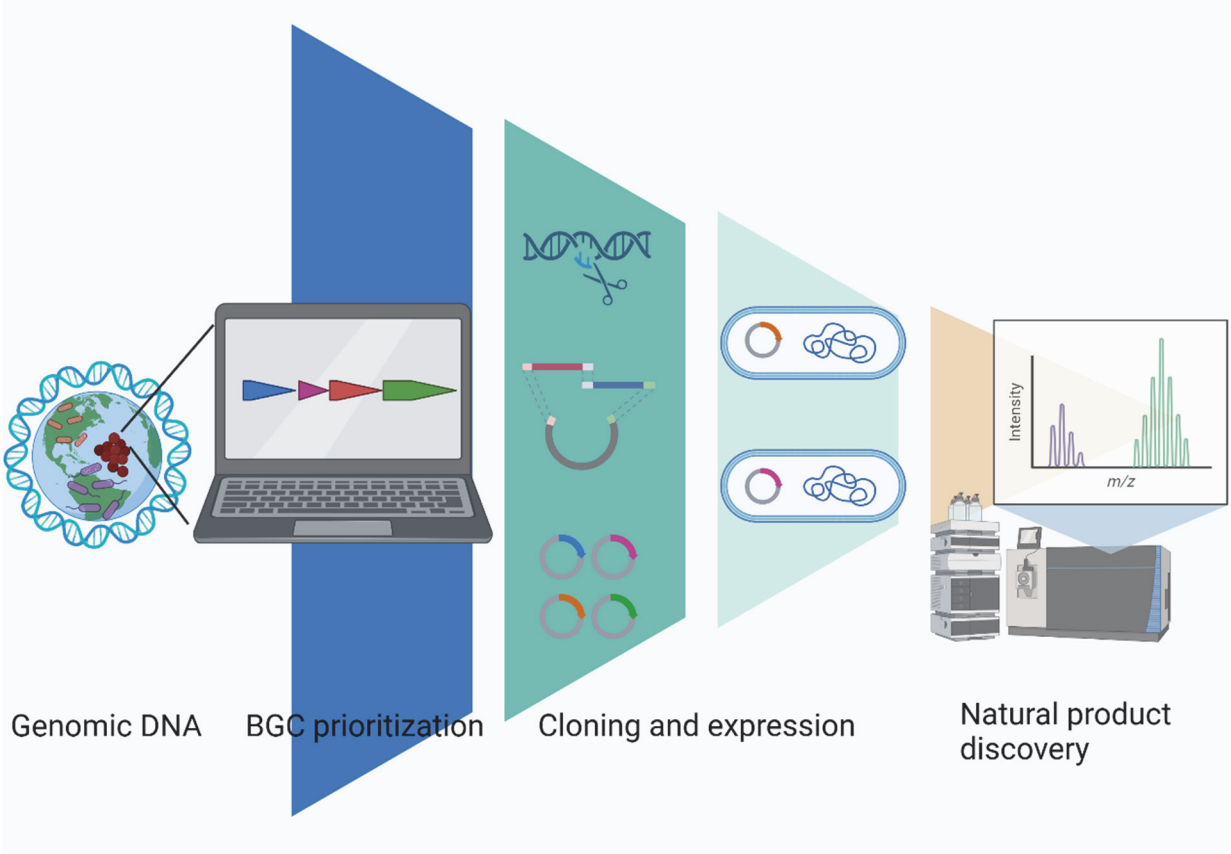


Figure 1. Pipeline for natural product discovery by heterologous expression. Genomic DNA is sequenced and analyzed with bioinformatic tools. Biosynthetic gene clusters (BGCs) are predicted and prioritized. The selected BGCs are cloned and transferred into a suitable host for expression. Metabolomics 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 the figure.

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 to June 2023 using search terms “heterologous expression and bacteria”, and “biosynthetic gene cluster” or “discovery” or “genome mining”. 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, below we discuss the rationale for BGC prioritization, cloning methods, biosynthetic class, source taxa, and host choice (**Figures 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 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 (Kautsar et al. 2019) 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 here (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 (**Figure 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 sedimis* (Han et al. 2022), and by the discovery of 31 other compounds listed in **SI Table S1** (Hashimoto et al. 2021; Hao et al. 2019; Zhang et al. 2021; Lasch et al. 2021; Enghiad et al. 2021; Shi et al. 2019; Yuet et al. 2020; De Rond, Asay, and Moore 2021; Wang et al. 2023; Liu et al. 2021; Lasch, Stierhof, et al. 2020; Li et al. 2020; Myronovskyi et al. 2018; Gao et al. 2023; Cheng et al. 2023; Koos and Link 2019; Kaweewan, Nakagawa, and Kodani 2021; S. H. Liu et al. 2019; Shuai et al. 2020; Lasch, Gummerlich, et al. 2020; Paulus et al. 2022; Bothwell et al. 2021; Gummerlich et al. 2020; Alberti et al. 2019; Vermeulen et al. 2022; Zhang et al. 2021; Unno et al. 2020).

Eighteen studies (36%) prioritized the expression of a NP class or subclass. For instance, Libis *et al.* targeted adenylation and ketosynthase domains in a 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 (Carson et al. 2023; Cheung-Lee et al. 2020; Cheung-Lee, Cao, and Link 2019; Gomez-Escribano et al. 2019; Cao et al. 2021), lanthipeptides (Arias-Orozco et al. 2021; Kaweewan et al. 2023; Singh et al. 2019; Thetsana et al. 2022), and thiopeptides (Santos-Aberturas et al. 2019). For further examples, see **SI Table S1** (Shi et al. 2021; J. Liu et al. 2019; Nguyen et al. 2022; Bosch et al. 2020).

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 genes encoding adenylation domains similar to known calcium dependent antibiotics (Wu et al. 2019; Hover et al. 2018). Likewise, the BGCs encoding glycopeptides GP1416 and GP6738 were selected for expression because of their sequence similarities to known glycopeptide antibiotic BGCs (Xu et al. 2020). Finally, one study used activity-guided prioritization of expressed DNA fragments. 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. 2020).

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

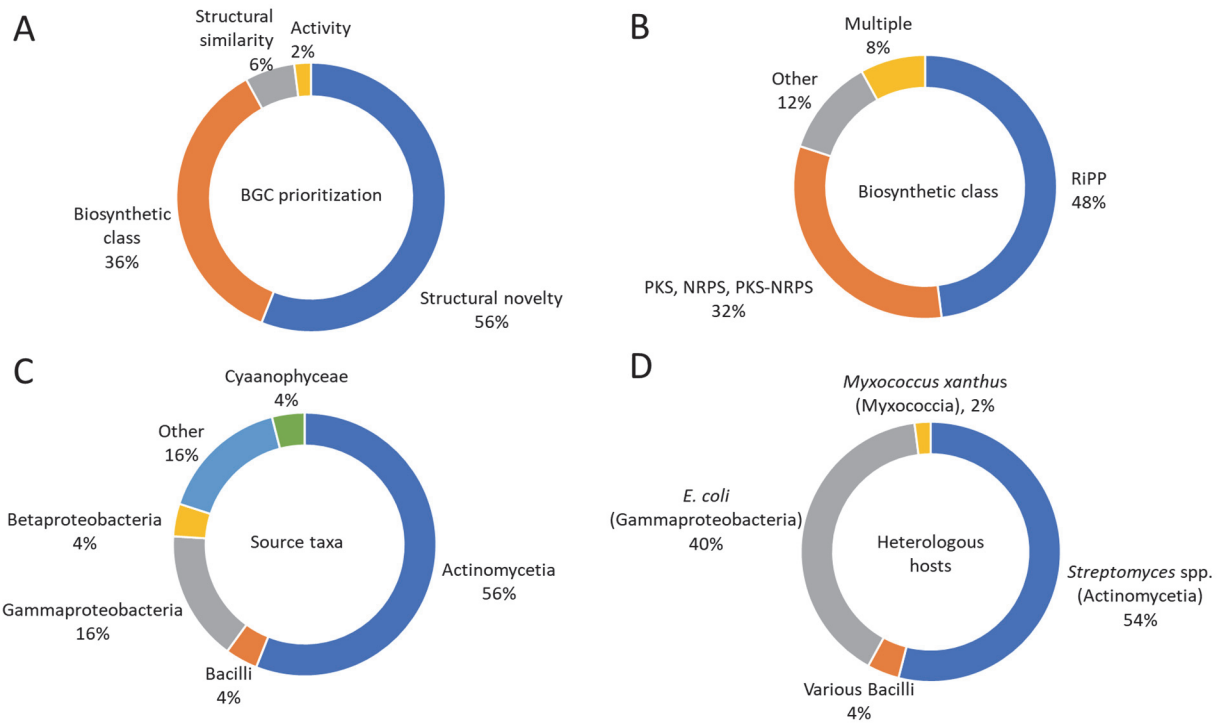


Figure 2. Discovery of bacterial natural products by heterologous expression in the last five years. The 50 studies reviewed here are broken down by (**A**) Rationale for BGC prioritization. (**B**) Biosynthetic class of prioritized BGCs. (**C**) Source of BGCs. (**D**) Heterologous hosts used.

Cloning methods

Three general methods were used to clone BGCs of interest for heterologous expression: DNA synthesis, direct cloning or assembly methods, and generation of random libraries (**Figure 3**). Information about library generation and direct cloning and assembly methods can be found here (Huo et al. 2019; 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 polymerase chain reaction (PCR) was used in 87% of the studies expressing only RiPPs (Han et al. 2022; Zhang et al. 2021; Wang et al. 2023; Koos and Link 2019; Kaweewan, Nakagawa, and Kodani 2021; Bothwell et al. 2021; Vermeulen et al. 2022; Unno et al. 2020; Ayikpoe et al. 2022; Nguyen et al. 2022; Bösch et al. 2020; Mevaere et al. 2018; Carson et al. 2023; Cheung-Lee et al. 2020; Cheung-Lee, Cao, and Link 2019; Gomez-Escribano et al. 2019; Cao et al. 2021; Arias-Orozco et al. 2021; Kaweewan et al. 2023; Singh et al. 2019; Thetsana et al. 2022). Cas12a-assisted precise targeted cloning using in vivo DNA circularization (CAPTURE), transformation associated recombination (TAR) cloning, and random library generation were used in the remaining studies (Ren et al. 2023; Santos-Aberturas et al. 2019; M. Xu et al. 2020) (**Figure 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, 80% of the studies aiming to express only PKS or NRPS used random library generation often in combination with assembly methods (**Figure 3**) (Libis et al. 2022; Hashimoto et al. 2021; Lasch et al. 2021; Shi et al. 2019; Yuet et al. 2020; Liu et al. 2021; Lasch, Stierhof, et al. 2020; Gao et al. 2023; S. H. Liu et al. 2019; Lasch, Gummerlich, et al. 2020; Paulus et al. 2022; Wu et al. 2019; Hover et al. 2018; Myronovskyi et al. 2018; Cheng et al. 2023; Libis et al. 2022). 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 chromosome of the hosts (Paulus et al. 2022; Hover et al. 2018; Wu et al. 2019). Moreover, recombination methods like Red/ET can be used to introduce a selective marker (Lasch et al. 2021; Lasch, Stierhof, et al. 2020; Shuai et al. 2020) 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).

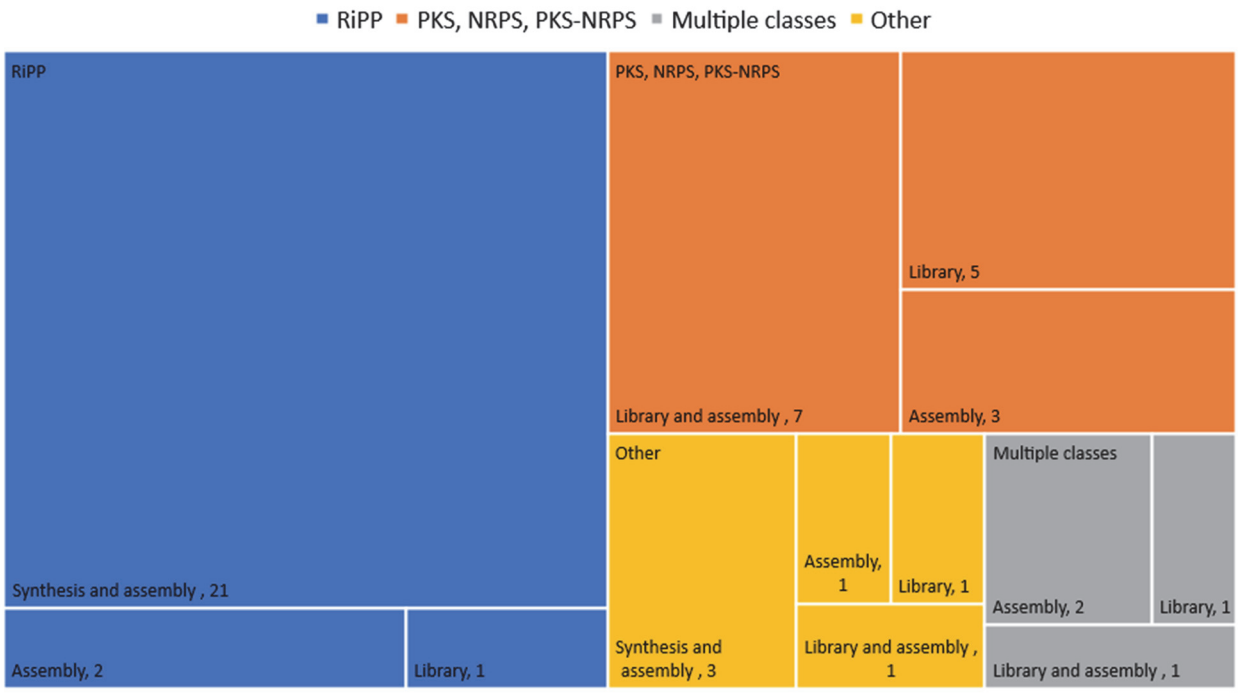


Figure 3: **Cloning techniques used for different BGC classes**.

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 (**Figure 2D**) matching the source of the BGCs in most cases (**Figure 4A**). Phylogeny relatedness was mentioned by Hao *et al*. and six other studies as the reason for host choice (Han et al. 2022; Hao et al. 2019; X. Wang et al. 2023; Shi et al. 2021; Mevaere et al. 2018; Carson et al. 2023; Xu et al. 2020). 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 (Wang et al. 2019), although exceptions exist (Zhang et al. 2017). Even though not specifically

mentioned as reason, strains related to the source organism were used as host for expression in 28 other studies (Libis et al. 2022; Hashimoto et al. 2021; Lasch et al. 2021; Enghiad et al. 2021; Shi et al. 2019; Liu et al. 2021; Lasch, Stierhof, et al. 2020; Li et al. 2020; Myronovskyi et al. 2018; Gao et al. 2023; Cheng et al. 2023; Kaweewan, Nakagawa, and Kodani 2021; S. H. Liu et al. 2019; Shuai et al. 2020; Lasch, Gummerlich, et al. 2020; Paulus et al. 2022; Gummerlich et al. 2020; Alberti et al. 2019; Vermeulen et al. 2022; Zhang et al. 2021; Unno et al. 2020; Liu et al. 2019; Ren et al. 2023; Nguyen et al. 2022; Gomez-Escribano et al. 2019; Arias-Orozco et al. 2021; Thetsana et al. 2022; Santos-Aberturas et al. 2019). 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 *E. coli* was used as host (**Figure 2D** and **Figure 4A**). Twenty studies (40%) used *E. coli* as host (**Figure 2D**) (Han et al. 2022; Zhang et al. 2021; Yuet et al. 2020; De Rond, Asay, and Moore 2021; Wang et al. 2023; Koos and Link 2019; Kaweewan, Nakagawa, and Kodani 2021; Bothwell et al. 2021; Vermeulen et al. 2022; Unno et al. 2020; Ayikpoe et al. 2022; Nguyen et al. 2022; Bösch et al. 2020; Carson et al. 2023; Cheung-Lee et al. 2020; Cheung-Lee, Cao, and Link 2019; Cao et al. 2021; Kaweewan et al. 2023; Singh et al. 2019; Thetsana et al. 2022). All but two (Yuet et al. 2020; De Rond, Asay, and Moore 2021) of these 20 studies expressed RiPP BGCs (**Figure 4B**). Codon optimization was performed during the cloning step in three of the studies (Yuet et al. 2020; Ayikpoe et al. 2022; Carson et al. 2023). Moreover, low yields were observed after *E. coli* expression of lanthipeptides from *Pedobacter lusitanus* NL19 and *Thermosporothrix hazakensis*. Co-expression of the BGC with the tRNA-Glu and glutamyl-tRNA from the source organism was used to improve the yield of encoded products (Bothwell et al. 2021; Kaweewan et al. 2023).

In contrast, 98% of studies exploring PKS, NRPS and other classes utilized *Streptomyces* hosts (**Figure 4B**) matching the source of the BGCs in most cases. Other hosts explored in the covered literature include *Myxococcus xanthus*, *Lactococcus lactis*, and *Streptococcus mutans*.

M. xanthus was used to express a refactored BGC from the closely related *Sorangiineae* 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).

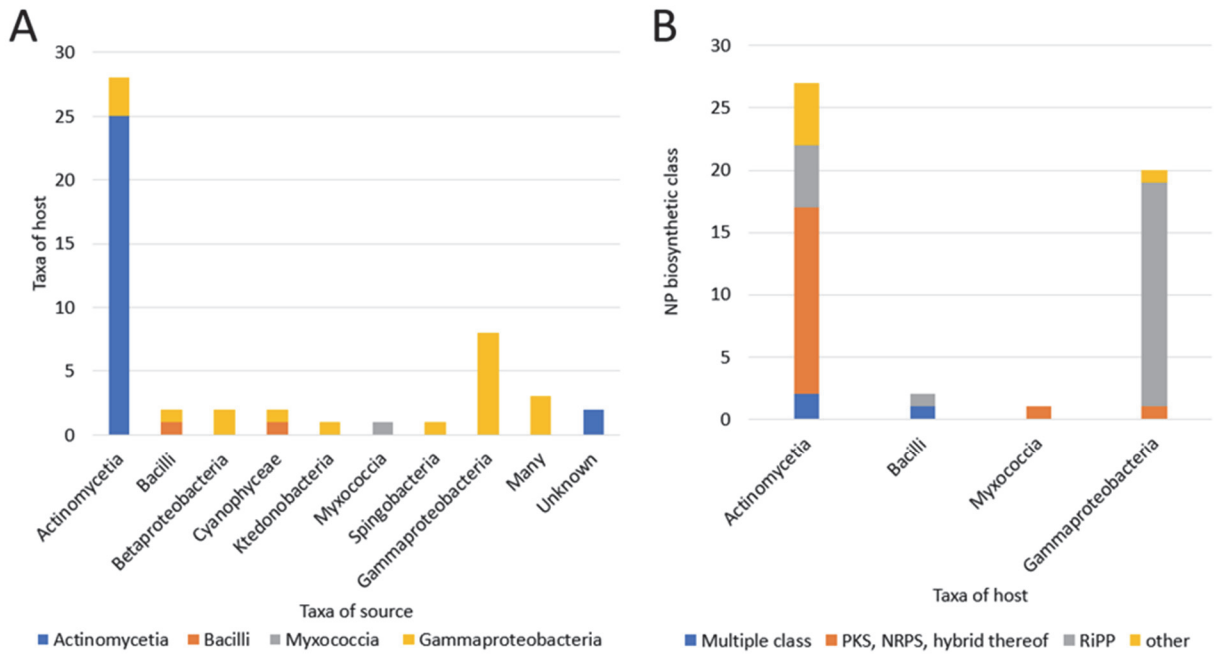


Figure 4. **Metrics regarding source taxa, host taxa, and biosynthetic class.** (**A**) Relationship between the taxa of the BGC source and the taxa of the heterologous host. (**B**) Relationship between the biosynthetic class of the expressed BGC and the taxa of the host used.

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. From the surveyed literature in the last five 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. Fortuitously, four large-scale studies from which success rates can be derived were included in the 50 studies reviewed here (**Table 1**) (Enghiad et al. 2021; Libis et al. 2022; Gummerlich et al. 2020; Ayikpoe 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 of *Saccharothrix espanaensis* with low similarity to known compounds. Of the 31 BGCs predicted by antiSMASH, six were excluded 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* ΔYA6 strains with 11% expression success rate. Of the four detected products, two were produced in enough 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 production. Five out of the seven were produced in large enough 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 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 production. 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 lassopeptides 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).

289 **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 *Saccharothrix espanaensis*	25	17 (68%)	Multiple	100	Random library	*S. lividans* DYA *S. albus* J1074	4 (11%)	2	(Gummerlich et al. 2020)
14 *Streptomyces spp.* 3 *Bacillus spp.*	43	43 (100%)	Multiple	10-113	CAPTURE	*S. avermitilis* SUKA17 *S. lividans* TK24 *B. subtilis* JH642	7 (16%)	5	(Enghiad et al. 2021)
100 *Streptomyces spp.*	Orphan PKS, NRPS, PKS-NRPS	58 (72%)	PKS, NRPS	140	Random library	*S. albus* J1074 *S. lividans* RedStrep 1.7	15 (24%)	3	(Libis et al. 2022)
1 Bacteroidota 10 Pseudomonadota 3 Cyanobacteriota 5 Actinomycetota 8 Bacillota.	96	83 (86%)	RiPPs	<18	Golden Gate assembly of synthetic genes	*E. coli* BL21 (DE3)	27 (32%)	3	(Ayikpoe et al. 2022)

290

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

296 ## Conclusions and remaining challenges

297 From January 2018 to June 2023, at least 50 studies were published that used heterologous
298 expression and combined led to the discovery of 63 new families of natural products (**Figure 2**).
299 Most of the studies (56%) prioritized BGCs based on structural novelty, followed by biosynthetic
300 class (36%). The main biosynthetic classes prioritized were RiPPs (48%) and PKS, NRPS, and
301 hybrids thereof (32%). Actinomycetia genomes were mined most frequently (56%) followed by
302 Gammaproteobacteria (16%). Cloning methods used depended on BGC size which is correlated
303 with biosynthetic class (**Figure 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, direct cloning and assembly were the methods of choice.

The main hosts (**Figure 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 (**Figure 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 (**Figure 4B**).

Heterologous expression is made possible by bioinformatics tools for genome mining, biosynthetic knowledge base, improved cloning techniques, host development, and improved isolation, and detection of the desired NP (Huo et al. 2019; Liu et al. 2020; Avalon et al. 2022; Caesar et al. 2021) (**Figure 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 four factors that can lead to failure and addressed those factors with their design. Genes were codon optimized for the host of choice, the synthetic genes were placed under a promoter known to work well in the host, toxicity was avoided by producing inactive precursors that were then converted to the final products in vitro, and purification was facilitated by inserting a His-tag. These improvements 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 clever 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, Liu, and Tang 2020).

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

With a host that is phylogenetically related to the source DNA, codon optimization 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 (**Figure 4A**). By expressing nine BGCs in 25 hosts, Wang *et al*. (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. Further, 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*. (Wang et al. 2019) reached 67% success rate. Other yet unknown or unexplored factors may help improve the success rate further in the future.

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

Alberti, Fabrizio, Daniel J. Leng, Ina Wilkening, Lijiang Song, Manuela Tosin, and Christophe Corre. 2019. “Triggering the Expression of a Silent Gene Cluster from Genetically Intractable Bacteria Results in Scleric Acid Discovery.” *Chemical Science* 10 (2): 453–63. https://doi.org/10.1039/C8SC03814G.

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.

Arias-Orozco, Patricia, Maartje Inklaar, Judith Lanooij, Rubén Cebrián, and Oscar P. Kuipers. 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): 2579–91. https://doi.org/10.1021/acssynbio.1c00224.

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, Richard S., Chengyou Shi, Alexander J. Battiste, Sara M. Eslami, Sangeetha Ramesh, Max A. Simon, Ian R. Bothwell, et al. 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, Brian O, Steven G Van Lanen, and Richard H Baltz. 2014. “Microbial Genome Mining for Accelerated Natural Products Discovery: Is a Renaissance in the Making?” *Journal of Industrial Microbiology and Biotechnology* 41 (2): 175–84. https://doi.org/10.1007/s10295-013-1389-9.

Baltz, Richard H. 2017. “Gifted Microbes for Genome Mining and Natural Product Discovery.” *Journal of Industrial Microbiology and Biotechnology* 44 (4–5): 573–88. 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, Kai, Thomas Wolf, Marc G. Chevrette, Xiaowen Lu, Christopher J. Schwalen, Satria A. Kautsar, Hernando G. Suarez Duran, et al. 2017. “AntiSMASH 4.0—Improvements in Chemistry Prediction and Gene Cluster Boundary Identification.” *Nucleic Acids Research* 45 (W1): W36–41. https://doi.org/10.1093/nar/gkx319.

Bothwell, Ian R., Tânia Caetano, Raymond Sarksian, Sónia Mendo, and Wilfred A. Van Der Donk. 2021. “Structural Analysis of Class I Lanthipeptides from *Pedobacter Lusitanus* NL19 Reveals an Unusual Ring Pattern.” *ACS Chemical Biology* 16 (6): 1019–29. https://doi.org/10.1021/acschembio.1c00106.

Bösch Nina M., Borsa Mariana, Greczmiel Ute, Morinaka Brandon I., Gugger Muriel, Oxenius Annette, Vagstad Anna L., Piel Jörn. 2020. Landornamides: Antiviral Ornithine-Containing Ribosomal Peptides Discovered through Genome Mining *Angewandte Chemie (International ed. in English)*, *59*, 11763. https://doi.org/10.1002/anie.201916321.

Caesar, Lindsay K., Rana Montaser, Nancy P. Keller, and Neil L. Kelleher. 2021. “Metabolomics and Genomics in Natural Products Research: Complementary Tools for Targeting New Chemical Entities.” *Natural Product Reports* 38 (11): 2041–65. https://doi.org/10.1039/D1NP00036E.

Cao, Li, Moshe Beiser, Joseph D. Koos, Margarita Orlova, Hader E. Elashal, Hendrik V. Schröder, and A. James Link. 2021. “Cellulonodin-2 and Lihuanodin: Lasso Peptides with an

Aspartimide Post-Translational Modification." *Journal of the American Chemical Society* 143 (30): 11690–702. https://doi.org/10.1021/jacs.1c05017.

Carson, Drew V., Monica Patiño, Hader E. Elashal, Alexis Jaramillo Cartagena, Yi Zhang, Megan E. Whitley, Larry So, et al. 2023. "Cloacaenodin, an Antimicrobial Lasso Peptide with Activity against *Enterobacter*." *ACS Infectious Diseases* 9 (1): 111–21. https://doi.org/10.1021/acsinfecdis.2c00446.

Cheng, Ziqian, Qingbo Zhang, Jing Peng, Xiaoyang Zhao, Liang Ma, Changsheng Zhang, and Yiguang Zhu. 2023. "Genomics-Driven Discovery of Benzoxazole Alkaloids from the Marine-Derived Micromonospora Sp. SCSIO 07395." *Molecules* 28 (2): 821. https://doi.org/10.3390/molecules28020821.

Cheung-Lee, Wai Ling, Li Cao, and A. James Link. 2019. "Pandonodin: A Proteobacterial Lasso Peptide with an Exceptionally Long C-Terminal Tail." *ACS Chemical Biology* 14 (12): 2783–92. https://doi.org/10.1021/acschembio.9b00676.

Cheung-Lee, Wai Ling, Madison E. Parry, Chuhan Zong, Alexis Jaramillo Cartagena, Seth A. Darst, Nancy D. Connell, Riccardo Russo, and A. James Link. 2020. "Discovery of Ubonodin, an Antimicrobial Lasso Peptide Active against Members of the *Burkholderia Cepacia* Complex." *ChemBioChem* 21 (9): 1335–40. https://doi.org/10.1002/cbic.201900707.

Covington, Brett C., Fei Xu, and Mohammad R. Seyedsayamdost. 2021. "A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters." *Annual Review of Biochemistry* 90 (1): 763–88. https://doi.org/10.1146/annurev-biochem-081420-102432.

De Rond, Tristan, Julia E. Asay, and Bradley S. Moore. 2021. "Co-Occurrence of Enzyme Domains Guides the Discovery of an Oxazolone Synthetase." *Nature Chemical Biology* 17 (7): 794–99. https://doi.org/10.1038/s41589-021-00808-4.

Enghiad, Behnam, Chunshuai Huang, Fang Guo, Guangde Jiang, Bin Wang, S. Kasra Tabatabaei, Teresa A. Martin, and Huimin Zhao. 2021a. "Cas12a-Assisted Precise

459 Targeted Cloning Using in Vivo Cre-Lox Recombination." *Nature Communications* 12 (1):
460 1171. https://doi.org/10.1038/s41467-021-21275-4.
461 Gao, Yunsheng, Christine Walt, Chantal D. Bader, and Rolf Müller. 2023. "Genome-Guided
462 Discovery of the Myxobacterial Thiolactone-Containing Sorangibactins." *ACS Chemical*
463 *Biology* 18 (4): 924–32. https://doi.org/10.1021/acschembio.3c00063.
464 Gavriilidou, Athina, Satria A. Kautsar, Nestor Zaburannyi, Daniel Krug, Rolf Müller, Marnix H.
465 Medema, and Nadine Ziemert. 2022. "Compendium of Specialized Metabolite
466 Biosynthetic Diversity Encoded in Bacterial Genomes." *Nature Microbiology* 7 (5): 726–
467 35. https://doi.org/10.1038/s41564-022-01110-2.
468 Gomez-Escribano, Juan Pablo, Jean Franco Castro, Valeria Razmilic, Scott A. Jarmusch,
469 Gerhard Saalbach, Rainer Ebel, Marcel Jaspars, Barbara Andrews, Juan A. Asenjo, and
470 Mervyn J. Bibb. 2019. "Heterologous Expression of a Cryptic Gene Cluster from
471 Streptomyces Leeuwenhoekii C34 ^{T} Yields a Novel Lasso Peptide, Leepeptin." Edited by
472 Andrew J. McBain. *Applied and Environmental Microbiology* 85 (23): e01752-19.
473 https://doi.org/10.1128/AEM.01752-19.
474 Gummerlich, Nils, Yuriy Rebets, Constanze Paulus, Josef Zapp, and Andriy Luzhetskyy. 2020.
475 "Targeted Genome Mining—From Compound Discovery to Biosynthetic Pathway
476 Elucidation." *Microorganisms* 8 (12): 2034.
477 https://doi.org/10.3390/microorganisms8122034.
478 Han, Yu, Xiaotong Wang, Youming Zhang, and Liujie Huo. 2022. "Discovery and Characterization
479 of Marinsedin, a New Class II Lanthipeptide Derived from Marine Bacterium *Marinicella*
480 *Sediminis* F2 ^{T}." *ACS Chemical Biology* 17 (4): 785–90.
481 https://doi.org/10.1021/acschembio.2c00021.
482 Hao, Tingting, Zhoujie Xie, Min Wang, Liwei Liu, Yuwei Zhang, Weicang Wang, Zhao Zhang, et
483 al. 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, Takuya, Junko Hashimoto, Noritaka Kagaya, Takehiro Nishimura, Hikaru Suenaga, Makoto Nishiyama, Tomohisa Kuzuyama, and Kazuo Shin-ya. 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): 354–58. https://doi.org/10.1038/s41429-021-00410-9.

Hegemann, Julian D, Joy Birkelbach, Sebastian Walesch, and Rolf Müller. 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, Bradley M., Seong-Hwan Kim, Micah Katz, Zachary Charlop-Powers, Jeremy G. Owen, Melinda A. Ternei, Jeffrey Maniko, et al. 2018. "Culture-Independent Discovery of the Malacidins as Calcium-Dependent Antibiotics with Activity against Multidrug-Resistant Gram-Positive Pathogens." *Nature Microbiology* 3 (4): 415–22. https://doi.org/10.1038/s41564-018-0110-1.

Huo, Liujie, Joachim J. Hug, Chengzhang Fu, Xiaoying Bian, Youming Zhang, and Rolf Müller. 2019. "Heterologous Expression of Bacterial Natural Product Biosynthetic Pathways." *Natural Product Reports* 36 (10): 1412–36. https://doi.org/10.1039/C8NP00091C.

Jiang, Yang, Syam Sundar Neti, Ian Sitarik, Priya Pradhan, Philip To, Yingzi Xia, Stephen D. Fried, Squire J. Booker, and Edward P. O'Brien. 2023. "How Synonymous Mutations Alter Enzyme Structure and Function over Long Timescales." *Nature Chemistry* 15 (3): 308–18. https://doi.org/10.1038/s41557-022-01091-z.

Kautsar, Satria A, Kai Blin, Simon Shaw, Jorge C Navarro-Muñoz, Barbara R Terlouw, Justin J J van der Hooft, Jeffrey A van Santen, et al. 2019. "MIBiG 2.0: A Repository for Biosynthetic

509 Gene Clusters of Known Function." *Nucleic Acids Research*, October, gkz882.
510 https://doi.org/10.1093/nar/gkz882.
511 Kaweewan, Issara, Shinta Ijichi, Hiroyuki Nakagawa, and Shinya Kodani. 2023. "Heterologous
512 Production of New Lanthipeptides Hazakensins A and B Using a Cryptic Gene Cluster of
513 the Thermophilic Bacterium Thermosporothrix Hazakensis." *World Journal of Microbiology*
514 *and Biotechnology* 39 (1): 30. https://doi.org/10.1007/s11274-022-03463-6.
515 Kaweewan, Issara, Hiroyuki Nakagawa, and Shinya Kodani. 2021. "Heterologous Expression of
516 a Cryptic Gene Cluster from Marinomonas Fungiae Affords a Novel Tricyclic Peptide
517 Marinomonasin." *Applied Microbiology and Biotechnology* 105 (19): 7241–50.
518 https://doi.org/10.1007/s00253-021-11545-y.
519 Kent, Ross, Samantha Halliwell, Kate Young, Neil Swainston, and Neil Dixon. 2018. "Rationalizing
520 Context-Dependent Performance of Dynamic RNA Regulatory Devices." *ACS Synthetic*
521 *Biology* 7 (7): 1660–68. https://doi.org/10.1021/acssynbio.8b00041.
522 Kloosterman, Alexander M., Kyle E. Shelton, Gilles P. Van Wezel, Marnix H. Medema, and
523 Douglas A. Mitchell. 2020. "RRE-Finder: A Genome-Mining Tool for Class-Independent
524 RiPP Discovery." Edited by Marcelino Gutierrez. *MSystems* 5 (5): e00267-20.
525 https://doi.org/10.1128/mSystems.00267-20.
526 Koos, Joseph D., and A. James Link. 2019. "Heterologous and *in Vitro* Reconstitution of
527 Fuscanodin, a Lasso Peptide from *Thermobifida Fusca*." *Journal of the American*
528 *Chemical Society* 141 (2): 928–35. https://doi.org/10.1021/jacs.8b10724.
529 Lasch, Constanze, Nils Gummerlich, Maksym Myronovskyi, Anja Palusczak, Josef Zapp, and
530 Andriy Luzhetskyy. 2020. "Loseolamycins: A Group of New Bioactive Alkylresorcinols
531 Produced after Heterologous Expression of a Type III PKS from Micromonospora
532 Endolithica." *Molecules* 25 (20): 4594. https://doi.org/10.3390/molecules25204594.
533 Lasch, Constanze, Marc Stierhof, Marta Rodríguez Estévez, Maksym Myronovskyi, Josef Zapp,
534 and Andriy Luzhetskyy. 2020. "Dudomycins: New Secondary Metabolites Produced after

Heterologous Expression of an Nrps Cluster from Streptomyces Albus Ssp. Chlorinus Nrrl B-24108." *Microorganisms* 8 (11): 1800. https://doi.org/10.3390/microorganisms8111800.

Lasch C, Stierhof M, Estévez MR, 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, Zhong, Yuanyuan Jiang, Xingwang Zhang, Yimin Chang, Shuai Li, Xiaomin Zhang, Shanmin Zheng, et al. 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): 5846–51. https://doi.org/10.1021/acscatal.0c01575.

Libis, Vincent, Logan W. MacIntyre, Rabia Mehmood, Liliana Guerrero, Melinda A. Ternei, Niv Antonovsky, Ján Burian, Zongqiang Wang, and Sean F. Brady. 2022. "Multiplexed Mobilization and Expression of Biosynthetic Gene Clusters." *Nature Communications* 13 (1): 5256. https://doi.org/10.1038/s41467-022-32858-0.

Liu, Jingyi, Xia Wu, Mingdong Yao, Wenhai Xiao, and Jian Zha. 2020. "Chassis Engineering for Microbial Production of Chemicals: From Natural Microbes to Synthetic Organisms." *Current Opinion in Biotechnology* 66 (December): 105–12. https://doi.org/10.1016/j.copbio.2020.06.013.

Liu, Jing, Xie, Xiulan, Li, Shu-Ming. 2019. Guanitrypmycin Biosynthetic Pathways Imply Cytochrome P450 Mediated Regio- and Stereospecific Guaninyl-Transfer Reactions. *Angewandte Chemie (International ed. in English)*, *58*, 11534. https://doi.org/10.1002/anie.201906891.

Liu, Shuang He, Wen Wang, Kai Biao Wang, Bo Zhang, Wei Li, Jing Shi, Rui Hua Jiao, Ren Xiang Tan, and Hui Ming Ge. 2019. "Heterologous Expression of a Cryptic Giant Type I PKS Gene Cluster Leads to the Production of Ansaseomycin." *Organic Letters* 21 (10): 3785–88. https://doi.org/10.1021/acs.orglett.9b01237.

Liu, Yang, Haibo Zhou, Qiyao Shen, Guangzhi Dai, Fu Yan, Xiaoju Li, Xiangmei Ren, et al. 2021. "Discovery of Polycyclic Macrolide Shuangdaolides by Heterologous Expression of a Cryptic *Trans* -AT PKS Gene Cluster." *Organic Letters* 23 (17): 6967–71. https://doi.org/10.1021/acs.orglett.1c02589.

Mevaere, Jimmy, Christophe Goulard, Olha Schneider, Olga N. Sekurova, Haiyan Ma, Séverine Zirah, Carlos Afonso, Sylvie Rebuffat, Sergey B. Zotchev, and Yanyan Li. 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.

Myronovskyi, Maksym, Birgit Rosenkränzer, Suvd Nadmid, Petar Pujic, Philippe Normand, and Andriy Luzhetskyy. 2018. "Generation of a Cluster-Free Streptomyces Albus Chassis Strains for Improved Heterologous Expression of Secondary Metabolite Clusters." *Metabolic Engineering* 49 (September): 316–24. https://doi.org/10.1016/j.ymben.2018.09.004.

Newman, David J., and Gordon M. Cragg. 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, Nguyet A., Ying Cong, Rachel C. Hurrell, Natalie Arias, Neha Garg, Aaron W. Puri, Eric W. Schmidt, and Vinayak Agarwal. 2022. "A Silent Biosynthetic Gene Cluster from a Methanotrophic Bacterium Potentiates Discovery of a Substrate Promiscuous Proteusin Cyclodehydratase." *ACS Chemical Biology* 17 (6): 1577–85. https://doi.org/10.1021/acschembio.2c00251.

Paulus, Constanze, Maksym Myronovskyi, Josef Zapp, Marta Rodríguez Estévez, Maria Lopatniuk, Birgit Rosenkränzer, Anja Palusczak, and Andriy Luzhetskyy. 2022. "Miramides A–D: Identification of Detoxin-like Depsipeptides after Heterologous Expression of a

585 Hybrid NRPS-PKS Gene Cluster from Streptomyces Mirabilis Lu17588." *Microorganisms*
586 10 (9): 1752. https://doi.org/10.3390/microorganisms10091752.
587 Ren, Hengqian, Shravan R. Dommaraju, Chunshuai Huang, Haiyang Cui, Yuwei Pan, Marko
588 Nesic, Lingyang Zhu, David Sarlah, Douglas A. Mitchell, and Huimin Zhao. 2023.
589 "Genome Mining Unveils a Class of Ribosomal Peptides with Two Amino Termini." *Nature*
590 *Communications* 14 (1): 1624. https://doi.org/10.1038/s41467-023-37287-1.
591 Santos-Aberturas, Javier, Govind Chandra, Luca Frattaruolo, Rodney Lacret, Thu H Pham,
592 Natalia M Vior, Tom H Eyles, and Andrew W Truman. 2019. "Uncovering the Unexplored
593 Diversity of Thioamidated Ribosomal Peptides in Actinobacteria Using the RiPPER
594 Genome Mining Tool." *Nucleic Acids Research* 47 (9): 4624–37.
595 https://doi.org/10.1093/nar/gkz192.
596 Shi, Jing, Xiang Xu, Pei Yi Liu, Yi Ling Hu, Bo Zhang, Rui Hua Jiao, Ghader Bashiri, Ren Xiang
597 Tan, and Hui Ming Ge. 2021. "Discovery and Biosynthesis of Guanipiperazine from a
598 NRPS-like Pathway." *Chemical Science* 12 (8): 2925–30.
599 https://doi.org/10.1039/D0SC06135B.
600 Shi, Jing, Ying Jie Zeng, Bo Zhang, Fen Li Shao, Yan Chi Chen, Xiang Xu, Yang Sun, Qiang Xu,
601 Ren Xiang Tan, and Hui Ming Ge. 2019. "Comparative Genome Mining and Heterologous
602 Expression of an Orphan NRPS Gene Cluster Direct the Production of Ashimides."
603 *Chemical Science* 10 (10): 3042–48. https://doi.org/10.1039/C8SC05670F.
604 Shuai, Hui, Maksym Myronovskyi, Suvd Nadmid, and Andriy Luzhetskyy. 2020. "Identification of
605 a Biosynthetic Gene Cluster Responsible for the Production of a New Pyrrolopyrimidine
606 Natural Product—Huimycin." *Biomolecules* 10 (7): 1074.
607 https://doi.org/10.3390/biom10071074.
608 Singh, M., Chaudhary, S., & Sareen, D. (2020). Roseocin, a novel two-component lantibiotic from
609 an actinomycete. *Molecular microbiology*, *113*(2), 326–337.
610 https://doi.org/10.1111/mmi.14419.

611 Thetsana, Chanaphat, Shinta Ijichi, Issara Kaweewan, Hiroyuki Nakagawa, and Shinya Kodani.
612 2022. “Heterologous Expression of a Cryptic Gene Cluster from a Marine Proteobacterium
613 Thalassomonas Actiniarum Affords New Lanthipeptides Thalassomonasins A and B.”
614 *Journal of Applied Microbiology* 132 (5): 3629–39. https://doi.org/10.1111/jam.15491.
615 Tietz, Jonathan I, Christopher J Schwalen, Parth S Patel, Tucker Maxson, Patricia M Blair, Hua-
616 Chia Tai, Uzma I Zakai, and Douglas A Mitchell. 2017. “A New Genome-Mining Tool
617 Redefines the Lasso Peptide Biosynthetic Landscape.” *Nature Chemical Biology* 13 (5):
618 470–78. https://doi.org/10.1038/nchembio.2319.
619 Unno, Kohta, Issara Kaweewan, Hiroyuki Nakagawa, and Shinya Kodani. 2020. “Heterologous
620 Expression of a Cryptic Gene Cluster from Grimontia Marina Affords a Novel Tricyclic
621 Peptide Grimoviridin.” *Applied Microbiology and Biotechnology* 104 (12): 5293–5302.
622 https://doi.org/10.1007/s00253-020-10605-z.
623 Vermeulen, Ross, Anton Du Preez Van Staden, Leonardo Joaquim Van Zyl, Leon M. T. Dicks, and
624 Marla Trindade. 2022. “Unusual Class I Lanthipeptides from the Marine Bacteria
625 *Thalassomonas Viridans*.” *ACS Synthetic Biology* 11 (11): 3608–16.
626 https://doi.org/10.1021/acssynbio.2c00480.
627 Wang, Gaoyan, Zhiying Zhao, Jing Ke, Yvonne Engel, Yi-Ming Shi, David Robinson, Kerem
628 Bingol, et al. 2019. “CRAGE Enables Rapid Activation of Biosynthetic Gene Clusters in
629 Undomesticated Bacteria.” *Nature Microbiology* 4 (12): 2498–2510.
630 https://doi.org/10.1038/s41564-019-0573-8.
631 Wang, Wenfang, Guosong Zheng, and Yinhua Lu. 2021. “Recent Advances in Strategies for the
632 Cloning of Natural Product Biosynthetic Gene Clusters.” *Frontiers in Bioengineering and*
633 *Biotechnology* 9 (July): 692797. https://doi.org/10.3389/fbioe.2021.692797.
634 Wang, Xiaotong, Zongjie Wang, Zhiqi Dong, Yihai Yan, Youming Zhang, and Liujie Huo. 2023.
635 “Deciphering the Biosynthesis of Novel Class I Lanthipeptides from Marine

Pseudoalteromonas Reveals a Dehydratase PsfB with Dethiolation Activity." *ACS Chemical Biology* 18 (5): 1218–27. https://doi.org/10.1021/acschembio.3c00135.

Wu, Changsheng, Zhuo Shang, Christophe Lemetre, Melinda A. Ternei, and Sean F. Brady. 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): 3910–19. https://doi.org/10.1021/jacs.8b12087.

Xu, Min, Wenliang Wang, Nicholas Waglechner, Elizabeth J. Culp, Allison K. Guitor, and Gerard D. Wright. 2020. "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. (2020). 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, Yan, Nicholas Liu, and Yi Tang. 2020. "Recent Developments in Self-Resistance Gene Directed Natural Product Discovery." *Natural Product Reports* 37 (7): 879–92. https://doi.org/10.1039/C9NP00050J.

Yuet, Kai P., Corey W. Liu, Stephen R. Lynch, James Kuo, Wesley Michaels, Robert B. Lee, Abigail E. McShane, Brian L. Zhong, Curt R. Fischer, and Chaitan Khosla. 2020. "Complete Reconstitution and Deorphanization of the 3 MDa Nocardiosis-Associated

660 Polyketide Synthase." *Journal of the American Chemical Society* 142 (13): 5952–57.
661 https://doi.org/10.1021/jacs.0c00904.
662 Zhang, Jia Jia, Xiaoyu Tang, Michelle Zhang, Darlene Nguyen, and Bradley S. Moore. 2017.
663 "Broad-Host-Range Expression Reveals Native and Host Regulatory Elements That
664 Influence Heterologous Antibiotic Production in Gram-Negative Bacteria." Edited by Julian
665 E. Davies. *MBio* 8 (5): e01291-17. https://doi.org/10.1128/mBio.01291-17.
666 Zhang, Yuqiao, Tingting Yao, Yuechen Jiang, Huayue Li, Weicheng Yuan, and Wenli Li. 2021.
667 "Deciphering a Cyclodipeptide Synthase Pathway Encoding Prenylated Indole Alkaloids
668 in Streptomyces Leeuwenhoekii." Edited by Haruyuki Atomi. *Applied and Environmental*
669 *Microbiology* 87 (11): e02525-20. https://doi.org/10.1128/AEM.02525-20.
670 Zhang, Y., Hong, Z., Zhou, L., Zhang, Z., Tang, T., Guo, E., Zheng, J., Wang, C., Dai, L., Si, T., &
671 Wang, H. 2022. Biosynthesis of Gut-Microbiota-Derived Lantibiotics Reveals a Subgroup
672 of S8 Family Proteases for Class III Leader Removal. *Angewandte Chemie (International*
673 *ed. in English)*, *61*(6), e202114414. https://doi.org/10.1002/anie.202114414.