Published in final edited form as:

Curr Opin Biotechnol. 2017 December ; 48: 21–27. doi:10.1016/j.copbio.2017.02.008.

Breaking the Silence: New Strategies for Discovering Novel Natural Products

Hengqian Ren^{1,†}, Bin Wang^{2,†}, and Huimin Zhao^{1,2,3,*}

¹Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

²Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

³Departments of Chemistry and Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

Natural products have been a prolific source of antibacterial and anticancer drugs for decades. One of the major challenges in natural product discovery is that the vast majority of natural product biosynthetic gene clusters (BGCs) have not been characterized, partially due to the fact that they are either transcriptionally silent or expressed at very low levels under standard laboratory conditions. Here we describe the strategies developed in recent years (mostly between 2014–2016) for activating silent BGCs. These strategies can be broadly divided into two categories: approaches in native hosts and approaches in heterologous hosts. In addition, we briefly discuss recent advances in developing new computational tools for identification and characterization of BGCs and high-throughput methods for detection of natural products.

Keywords

natural products; silent pathways; pathway refactoring; synthetic biology

Introduction

Natural products are a prolific source for drug discovery. Over the past 30 years, a large fraction of anticancer drugs and anti-infectives came from either natural products or derivatives of natural products [1]. Moreover, due to an increase in drug resistance, there is an ever-pressing need for new drugs [2,3]. However, since the Golden Age of natural product discovery in the 1950s–1960s, traditional methods have only yielded known natural products unfortunately, indicating a need for new methods that can discover novel natural

*To whom correspondence should be addressed. Phone: (217) 333-2631. Fax: (217) 333-5052. zhao5@illinois.edu.

†These authors contributed equally.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

products. Thanks to advances in DNA sequencing technologies and bioinformatics, researchers have become increasingly aware of the vast untapped biosynthetic capacity of microorganisms. Next generation sequencing can generate genome sequences in a fast and inexpensive way, while computational tools such as antiSMASH [4] can identify biosynthetic gene clusters (BGCs) from genome sequences in batches. Currently millions of putative BGCs have been identified, which have potential to produce a vast number of novel natural products [5^{**}]. Moreover, in recent years, researchers have identified numerous BGCs from human microbiota. Due to potential microbiota-host interactions, these BGCs might be more likely to produce compounds related to human health and disease [6].

Despite the existence of numerous putative BGCs, only a tiny fraction of them have been characterized to date, and one of the key reasons is that in most cases BGCs are expressed poorly or not at all in their native hosts due to tight regulation under conventional laboratory culture conditions. Because of their tight regulation, most BGCs will not be activated unless special environmental cues exist. These BGCs are commonly known as “cryptic pathways.” In order to activate those cryptic pathways, many innovative approaches have been developed and comprehensively reviewed elsewhere [5^{**},7].

A typical workflow for a modern natural product discovery strategy begins with identification of silent BGCs using bioinformatics tools, followed by activation of target BGCs using synthetic biology strategies and structural characterization of the corresponding natural products (Figure 1). In this review, we will provide a brief update on the approaches for activating cryptic pathways that were developed in the last two years. In addition, we will discuss recent progress in the development of computational tools and product detection methods for genome mining.

Computational tools for identification of BGCs

Based on the implementation strategies, computational tools for identification of BGCs mainly fall into two groups. The first group is rule-based tools, such as ClusterScan, NP.searcher, NaPDoS, SEARCHPKS, and SMURF for mining PKS/NRPS, BAGEL for RiPPs, and antiSMASH [8]. These tools use predefined rulesets to identify signature genes that are associated with biosynthetic pathways. As a result they can predict known classes of BGCs with high confidence and positive rates but cannot readily identify BGCs that use completely different enzymatic mechanisms. Notable among them is antiSMASH, which in its 3rd version can detect 44 different classes of BGCs, and upon integration with ClusterFinder which uses a rule-independent algorithm further provides alternative options for automatic genome-scale predictions of BGCs of both known and novel classes [4,9]. The second group is rule-independent tools, such as ClusterFinder that predicts candidate BGCs via calculating the frequencies of Pfam domains that are commonly found in BGCs, thus those genomic regions rich in such Pfam domains hold more possibilities as candidates even in the total absence of signature genes [9]. By contrast, EvoMining takes advantage of phylogenomic analysis based evolutionary logic to identify functionally diverged paralogs of primary metabolic enzymes, and presumes that the additional copies are repurposed for secondary metabolite biosynthesis [10]. These rule-independent tools can identify new BGC classes; but at this stage it is important to reduce false-positive rate such that new rules can

emerge. A newly established website called Secondary Metabolite Bioinformatics Portal (SMBP) provides information on all aspects of secondary metabolite bioinformatics, enabling quick access to those computational tools and databases [8].

Activation of silent BGCs in native hosts

Actinomycetes, especially the genus *Streptomyces*, harbor dozens of BGCs per genome but control their expressions strictly [11]. Therefore, it remains an overwhelming challenge to activate these BGCs and identify the associated natural products if they are transcriptionally silent. Two main strategies have been developed to activate silent BGCs, either to unlock the suppression of BGC gene expression in the native hosts, or to directly bypass the regulation system by refactoring and reconstructing controlling elements in BGCs in heterologous hosts.

Genetic manipulation in native hosts takes advantages of the intact elements in BGCs, *e.g.*, precursor supply and end-product export, but it needs to break through the regulatory system first. Overexpressing positive regulator genes and deleting negative regulator genes are two main approaches used to activate transcription. For example, a new biaryl polyketide venemycin and two known ansatrienins in *Streptomyces venezuelae* and *Streptomyces* sp. XZQH13 were discovered when the large ATP-binding regulators of the LuxR family (LAL) transcriptional activator genes, *vemR* and *astGI* were constitutively expressed, respectively [12,13]. Moreover, two novel chattamycins were discovered when overexpressing an atypical response regulator (ARR) family regulator gene, *chaI* [14]; and overexpression of a LuxR family activator gene *namI* led to production of three new naphthalenic octaketide ansamycins [15]. For the jadomycin and gaudimycin biosynthetic pathways, however, no significant effects were observed when ARR family activator genes were overexpressed [16]. This approach is proved to be pathway-specific, thus the efficacy varies case by case. However, due to its simplicity, this approach may be further developed for high throughput activation of silent BGCs on a strain level. In addition, overexpression of the homologs of the target activator genes may provide an alternative strategy as indicated in the activation of a mureidomycin BGC in *S. roseosporus* [17].

Another approach to activate cryptic BGCs is to artificially knock-in a strong promoter upstream of the target BGC to initiate the expression of the biosynthetic machinery, and thus synthesis of the corresponding natural product (Figure 2). Traditional promoter knock-in strategies based on homologous double-crossover recombination are often time-consuming and labor-intensive especially when applied to streptomycetes [18,19]. Recently we developed a highly efficient clustered, regularly interspaced, short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system for multiplex genome editing in streptomycetes, either targeting at precise gene deletion [20] or promoter knock-in [21]. By applying this one-step strategy to five *Streptomyces* species, multiple silent BGCs belonging to different classes were activated, among which a novel pentangular polyketide from *S. viridochromogenes* was characterized [21]. Considering the common multi-operon architecture in BGCs, this strategy may be used to iteratively knock-in promoters and activate the silent full-length BGCs, while different combinations of activated operons (by promoter insertion) can also provide insights into the biosynthetic mechanism (Figure 2). A

system called CRISPRi was recently applied to *S. coelicolor*, which used a catalytically dead Cas9 (dCas9) to interfere with gene expression in a sequence-specific manner [22], and thus can be potentially applied to repress transcription of negative/positive regulatory genes for discovery of natural products from silent BGCs. Other tools such as the *in vitro* DNA editing system [23] and the transformation-associated recombination (TAR) method [24*] were also established based on the versatile CRISPR/Cas9 system, which further advances technological development in natural product discovery. Successful application of the CRISPR/Cas9 system in fungi has also been reported [25] and reviewed previously [26].

Other promising approaches to activate silent BGCs, such as one-strain-many-compounds (OSMAC) [27], co-culturing [28], ribosome engineering [29], and the use of elicitors [30,31], are all applicable to a wide range of hosts and may be scalable. However, the resulting host physiological disturbances are global, which make comparative metabolic profiling sometimes very difficult. Another promising approach named reporter-guided mutant selection (RGMS) employs genome-scale mutagenesis to create genetic diversity, and then directly selects desired target-activated mutants after double-reporter screening [16*]. This approach possesses a pathway-specific trait to either activate silent BGCs or improve production of weakly expressed BGCs. However, it also has built-in limitations, *e.g.*, requiring hosts to be kanamycin and hygromycin sensitive, as well as having no background yellow color.

Activation of silent BGCs in heterologous hosts

Emerging synthetic biology tools have enabled researchers to activate silent BGCs in heterologous hosts. The tight regulation of silent BGCs can be relieved by either direct cloning or refactoring using well-characterized promoters. Heterologous hosts possess significant growth advantages over most native hosts and therefore make activation of BGCs from unculturable organisms possible.

One of the most straightforward approaches to activate silent BGCs in a heterologous host is direct cloning. A major obstacle in this approach is the difficulty of handling large DNA fragments. Recently researchers have developed multiple synthetic biology tools that leverage homologous recombination (HR) to specifically capture the gene cluster of interest. Transformation-associated recombination (TAR) in *Saccharomyces cerevisiae* has been utilized for cloning large genomic fragments [32]. Random genomic DNA fragments are co-transformed with an excess of the TAR cloning vector into *S. cerevisiae*. Since the cloning vector shares homologous regions to the target DNA sequences at both ends, once the cell gets both the cloning vector and the genomic fragment with the target DNA region, recombination will occur in high efficiency and result in the formation of a circular yeast artificial chromosome (YAC) molecule with the target sequence. This approach was successfully used to clone DNA fragments up to 250 kb from complex genomes [33]. The entire procedure takes approximately two weeks and generates positive clones with 1%–5% yield. After further refinement for direct capture of gene clusters and discovery of taromycin A in *Streptomyces coelicolor* [34**], Moore and coworkers successfully extended this method to the heterologous expression of BGCs in *Bacillus subtilis* and *Escherichia coli*

(Figure 3) [35,36]. Moreover, by coupling with a target-directed genome mining process, a distinct group of thiotetronic acid natural products was discovered [37].

In addition to *S. cerevisiae*, *E. coli* can also be used for direct cloning of BGCs when assisted by phage recombination systems, such as the lambda Red system and the ET recombination system (Figure 3) [38,39]. By using the full length RecET in *E. coli*, Fu and coworkers discovered increased HR efficiency between two linear DNA molecules [40]. They then cloned all 10 megasynthetase gene clusters from *Photorhabdus luminescens* and identified novel metabolites luminmycin A and luminmide A/B. Similar to TAR, the application of RecET has afforded the heterologous production of disorazol and salinomycin in *Myxococcus xanthus* and *S. coelicolor* respectively [41,42]. In addition, a cryptic NRPS gene cluster from *Paenibacillus larvae* was cloned and activated by heterologous expression in *E. coli*, which produced a novel compound, sevadicin [43]. Recently this method was optimized by coupling with the lambda Red system for convenient modifications of the plasmid [44].

One of the major problems associated with the abovementioned direct cloning methods is the low yield of positive clones, which is caused by the nonspecific targeting of random genomic fragments. Based on the fact that the CRISPR/Cas9 system can cleave at almost any site, Jiang and coworkers developed a Cas9-assisted targeting of chromosome segments (CATCH) method for direct cloning of BGCs (Figure 3) [45,46*]. The target BGC is cleaved from intact genomic DNA by Cas9 *in vitro* and then ligated with a cloning vector through Gibson assembly. The whole procedure takes approximately eight hours over several days and yields positive clones varying from 20% (for 100 kb inserts) to 60% (for 50 kb inserts).

There are other direct cloning methods that rely on site-specific recombination reactions. For example, Hu and coworkers developed a Cre/loxP based method [47]. The loxP sites are first integrated into the boundary regions of the target BGC with elements that are necessary for plasmid replication. Then the Cre recombinase is expressed *in vivo* and the whole region flanked by loxP is circularized as a plasmid. The resulting plasmid is isolated and transformed into *E. coli*. Besides using recombination *in vivo*, the circularized BGC can also be generated by recombination catalyzed by phiBT1 integrase *in vitro* in a very similar manner [48].

In addition to direct cloning, refactoring is another strategy to activate silent BGCs in heterologous hosts. Refactoring usually involves replacement of original regulatory elements and sometimes can be coupled with direct cloning. For example, the taromycin gene cluster was not functional immediately after direct cloning, but was activated after knockout of the negative regulator encoded in the gene cluster [34**]. The more commonly used refactoring strategy is to replace the native promoters by the well-characterized constitutive promoters. One way to accomplish this is through HR in *S. cerevisiae* [49*]. The target BGC is first cloned into a shuttle vector and then co-transformed with promoter cassettes into *S. cerevisiae*. Since each promoter cassette has the homologous sequences to their corresponding insertion sites in the target BGC as well as different selection markers, the promoters will be inserted into the designated sites by HR and positive clones can be selected in the appropriate medium. By using this method, an indolotryptoline BGC was

activated, which led to the discovery of lazirimides A and B. This strategy was further optimized by using CRISPR/Cas9, which generates DNA double-strand breaks at the promoter insertion sites to enable one-step insertion of multiple promoters with a single selection marker or no marker and therefore increases the HR efficiency [24]. Another strategy to refactor BGCs is through DNA assembly, which builds up the whole construct including the biosynthetic genes, promoters and other elements from linear DNA fragments. Previously we developed a plug-and-play refactoring workflow based on DNA assembly and successfully activated a silent polycyclic tetramate macrolactam BGC [50,51]. For more details about this strategy, readers are referred to recent reviews [52,53].

High-throughput product detection

Besides activation of silent BGCs, another challenge for discovery of natural products through genome mining is how to detect the product of the corresponding BGC, specifically in a high-throughput manner. Due to advances in the understanding of biosynthetic mechanisms of natural products and mass spectrometry, many computational tools have been developed to connect the BGC to its product based on MS/MS data. For example, an automated genomes-to-natural products platform (GNP) was recently developed to predict, combinatorialize and identify the product of modular gene clusters (PKS and NRPS) from LC-MS/MS of crude extracts [54]. In order to match the polyketides and nonribosomal peptides to their corresponding BGCs, they further developed generalized retro-biosynthetic assembly prediction engine (GRAPE) and global alignment for natural products cheminformatics (GARLIC) [55]. Besides processing the BGCs and MS/MS spectra individually, a new strategy was developed to group BGCs and MS/MS spectrums into gene cluster families (GCFs) and molecular families (MFs) by similarity [56]. The relationships between GCFs and MFs can be utilized to aid in genome mining of unsequenced organisms. A plethora of MS/MS-based computational tools such as NRPquest and RiPPquest have been developed, which were reviewed elsewhere [57,58]. Although most of these computational tools are limited to the analysis of peptide natural products, it can be expected that such MS/MS based approaches will be extended to other natural products in the future. Recently, the Global Natural Products Social Molecular Networking (GNPS) was set up, which encourages researchers around the world to share their MS data online and can therefore further accelerate the identification of novel natural products from MS data [59]. Besides MS based methods, improvements in NMR probes, experiments, data mining and databases have also made NMR based high-throughput detection of natural products possible, which was reviewed elsewhere [60,61].

Conclusions and future perspectives

Multiple new approaches for activating silent BGCs were developed in the last two years. These new approaches were accomplished with varying degrees of success, which provide room for further refinement and improvement. Approaches in the native hosts can avoid precursor availability issues or protein folding related issues, but are limited to culturable microorganisms. Most genomics driven approaches require the native hosts to be genetically trackable. In contrast, greater DNA manipulation freedom can be obtained when the target silent BGC is activated in a heterologous host, but currently the choice of the appropriate

host organism still largely depends on previous studies, which may or may not afford the most appropriate host. Therefore, in addition to discovering more novel natural products, more knowledge about secondary metabolism is also needed in order to help researchers optimize the existing approaches and develop more general and scalable approaches.

Another major challenge in natural product discovery is to increase the throughput. Currently the characterization of BGCs cannot keep pace with the rate of BGC identification. Most of the approaches for activating silent BGCs described in this review require extensive genetic manipulations. However, with emerging DNA manipulation tools and advances in automated synthetic biology, it appears that this issue could be solved in the near future. In addition, many MS-based methods have been developed to date. Although most of those methods are limited to peptide natural products, they provide a new avenue for high-throughput MS based product detection. We envision a new natural product discovery paradigm in which the BGC identification algorithm will be coupled with high-throughput pathway activation and product detection methods. Such paradigm will lead to discovery of numerous novel natural products with important biological activities.

Acknowledgments

We gratefully acknowledge financial support from the National Institutes of Health (GM077596) and the Visiting Investigator Programme of Agency for Science, Technology and Research of Singapore.

Abbreviations

antiSMASH	antibiotics & Secondary Metabolite Analysis SHell
BGC	Biosynthetic Gene Cluster
CATCH	Cas9-Assisted Targeting of CHromosome Segments
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
GARLIC	Global Alignment for Natural Products Cheminformatics
GNP	Genomes-to-Natural Products Platform
GNPS	Global Natural Products Social Molecular Networking
GRAPE	Generalized Retro-Biosynthetic Assembly Prediction Engine
HR	Homologous Recombination
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NRPS	Non-ribosomal Peptide Synthetase
PKS	Polyketide Synthase
TAR	Transformation-Associated Recombination
YAC	Yeast Artificial Chromosome

LAL	Large ATP-binding regulators of the LuxR family
ARR	Atypical Response Regulator
OSMAC	One-Strain-Many-Compounds
RGMS	Reporter-Guided Mutant Selection

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

*of special interest

**of outstanding interest

1. Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod.* 2016; 79:629–661. [PubMed: 26852623]
2. Clatworthy AE, Pierson E, Hung DT. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol.* 2007; 3:541–548. [PubMed: 17710100]
3. World Health Organization. Antimicrobial resistance: global report on surveillance. 2014
4. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015; 43:237–243. [PubMed: 25435544]
- 5**. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol.* 2015; 13:509–523. This is a comprehensive review on various approaches for activating silent BGCs. [PubMed: 26119570]
6. Donia MS, Fischbach MA. Small molecules from the human microbiota. *Science.* 2015; 349:1254766. [PubMed: 26206939]
7. Zarins-Tutt JS, Barberi TT, Gao H, Mearns-Spragg A, Zhang L, Newman DJ, Goss RJM. Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *J Nat Prod.* 2016; 33:54–72.
8. Weber T, Kim HU. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol.* 2016; 1:69–79. [PubMed: 29062930]
9. Cimermanic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell.* 2014; 158:412–421. [PubMed: 25036635]
10. Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-Gómez F. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. *Genome Biol Evol.* 2016; 8:1906–1916. [PubMed: 27289100]
11. Liu G, Chater KF, Chandra G, Niu G, Tan H. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev.* 2013; 77:112–143. [PubMed: 23471619]
12. Thanapipatsiri A, Gomez-Escribano JP, Song L, Bibb M, Al-Bassam M, Chandra G, Thamchaipenet A, Challis G, Bibb M. Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. *ChemBioChem.* 2016; 17:2189–2198. [PubMed: 27605017]
13. Xie C, Deng JJ, Wang HX. Identification of AstG1, a LAL family regulator that positively controls ansatrienins production in *Streptomyces sp* XZQH13. *Curr Microbiol.* 2015; 70:859–864. [PubMed: 25784540]
14. Zhou ZX, Xu QQ, Bu QT, Guo YY, Liu SP, Liu Y, Du YL, Li YQ. Genome mining-directed activation of a silent angucycline biosynthetic gene cluster in *Streptomyces chattanoogensis*. *ChemBioChem.* 2015; 16:496–502. [PubMed: 25511454]

15. Li SR, Li YY, Lu CH, Zhang JL, Zhu J, Wang HX, Shen YM. Activating a cryptic ansamycin biosynthetic gene cluster to produce three new naphthalenic octaketide ansamycins with n-pentyl and n-butyl side chains. *Org Lett*. 2015; 17:3706–3709. [PubMed: 26167742]
- 16•. Guo F, Xiang SH, Li LY, Wang B, Rajasarkka J, Grondahl-Yli-Hannuksela K, Ai GM, Metsä-Ketela M, Yang KG. Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab Eng*. 2015; 28:134–142. This work applied the reporter method for activation of silent gene clusters and used a double-reporter system, xylE-neo, to lower the false positive rate dramatically. In principle, the target cluster activated mutants can be directly selected after the double-reporter screening. [PubMed: 25554073]
17. Jiang LJ, Wang L, Zhang JH, Liu H, Hong B, Tan HR, Niu GQ. Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in *Streptomyces roseosporus* NRRL 15998. *Sci Rep*. 2015; 5:14111. doi: 10.11038/srep14111 [PubMed: 26370924]
18. Ahuja M, Chiang Y-M, Chang S-L, Praseuth MB, Entwistle R, Sanchez JF, Lo H-C, Yeh H-H, Oakley BR, Wang CCC. Illuminating the diversity of aromatic polyketide synthases in *Aspergillus nidulans*. *J Am Chem Soc*. 2012; 134:8212–8221. [PubMed: 22510154]
19. Bok JW, Soukup AA, Chadwick E, Chiang YM, Wang CCC, Keller NP. VeA and MvlA repression of the cryptic orsellinic acid gene cluster in *Aspergillus nidulans* involves histone 3 acetylation. *Mol Microbiol*. 2013; 89:963–974. [PubMed: 23841751]
20. Cobb RE, Wang YJ, Zhao HM. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol*. 2015; 4:723–728. [PubMed: 25458909]
- 21•. Zhang M, Wong F, Wang Y, Luo S, Lim Y, Heng E, Yeo W, Cobb R, Enghiad B, Ang E, et al. An efficient CRISPR-Cas9 mediated knock-in strategy for activation of silent *Streptomyces* biosynthetic gene clusters for natural product discovery. *Nat Chem Biol*. revision under review. This work applied the CRISPR-Cas9 system based promoter knockin strategy to activate more than ten silent BGCs in five different *Streptomyces* species, demonstrating a great potential for high throughput discovery of natural products in the native hosts.
22. Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol*. 2015; 4:1020–1029. [PubMed: 25806970]
23. Liu Y, Tao W, Wen S, Li Z, Yang A, Deng Z, Sun Y. In vitro CRISPR/Cas9 system for efficient targeted DNA editing. *Mbio*. 2015; 6:e01714–01715. [PubMed: 26556277]
- 24•. Kang HS, Charlop-Powers Z, Brady SF. Multiplexed CRISPR/Cas9- and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth Biol*. 2016; 5:1002–1010. This work combined CRISPR/Cas9 and TAR to enable efficient large gene cluster assembly and multiplexed promoter engineering in a single round. It also provides a mCRISTAR python package and Webapp for transcriptional activation of complex natural product BGCs. [PubMed: 27197732]
25. Weber J, Valiante V, Nødvig CS, Mattern DJ, Slotkowski RA, Mortensen UH, Brakhage AA. Functional reconstitution of a fungal natural product gene cluster by advanced genome editing. *ACS Synth Biol*. 2016; doi: 10.1021/acssynbio.1026b00203
26. Billingsley JM, DeNicola AB, Tang Y. Technology development for natural product biosynthesis in *Saccharomyces cerevisiae*. *Curr Opin Biotechnol*. 2016; 42:74–83. [PubMed: 26994377]
27. Bode HB, Bethe B, Höfs R, Zeeck A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem*. 2002; 3:619–627. [PubMed: 12324995]
28. Netzker T, Fischer J, Weber J, Mattern DJ, König CC, Valiante V, Schroeckh V, Brakhage AA. Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front Microbiol*. 2015; 6:299. [PubMed: 25941517]
29. Ochi K, Tanaka Y, Tojo S. Activating the expression of bacterial cryptic genes by rpoB mutations in RNA polymerase or by rare earth elements. *J Ind Microbiol Biotechnol*. 2014; 41:403–414. [PubMed: 24127067]
30. Okada BK, Seyedsayamdost MR. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol Rev*. 2016; doi: 10.1093/femsre/fuw1035
31. Seyedsayamdost MR. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci U S A*. 2014; 111:7266–7271. [PubMed: 24808135]

32. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc Natl Acad Sci U S A*. 1996; 93:491–496. [PubMed: 8552668]
33. Kouprina N, Larionov V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat Protoc*. 2008; 3:371–377. [PubMed: 18323808]
- 34•. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A*. 2014; 111:1957–1962. This is the first application of TAR in natural product discovery. [PubMed: 24449899]
35. Li YX, Li ZR, Yamanaka K, Xu Y, Zhang WP, Vlamakis H, Kolter R, Moore BS, Qian PY. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium *Bacillus subtilis*. *Sci Rep*. 2015; 5:9383. [PubMed: 25807046]
36. Ross AC, Gulland LES, Dorrestein PC, Moore BS. Targeted capture and heterologous expression of the *Pseudoalteromonas alterochromide* gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. *ACS Synth Biol*. 2015; 4:414–420. [PubMed: 25140825]
37. Tang X, Li J, Millán-Aguinaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem Biol*. 2015; 10:2841–2849. [PubMed: 26458099]
38. Murphy KC. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol*. 1998; 180:2063–2071. [PubMed: 9555887]
39. Zhang Y, Buchholz F, Muirers JP, Stewart AF. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet*. 1998; 20:123–128. [PubMed: 9771703]
40. Fu J, Bian XY, Hu SB, Wang HL, Huang F, Seibert PM, Plaza A, Xia LQ, Muller R, Stewart AF, et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol*. 2012; 30:440–446. [PubMed: 22544021]
41. Tu Q, Herrmann J, Hu S, Raju R, Bian X, Zhang Y, Müller R. Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Sci Rep*. 2016; 6:21066. [PubMed: 26875499]
42. Yin J, Hoffmann M, Bian X, Tu Q, Yan F, Xia L. Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci Rep*. 2015; 5:15081. [PubMed: 26459865]
43. Tang Y, Frewert S, Harmrolfs K, Herrmann J, Karmann L, Kazmaier U, Xia L, Zhang Y, Müller R. Heterologous expression of an orphan NRPS gene cluster from *Paenibacillus larvae* in *Escherichia coli* revealed production of sevadicin. *J Biotechnol*. 2015; 194:112–114. [PubMed: 25529345]
44. Wang H, Li Z, Jia R, Hou Y, Yin J, Bian X, Li A, Müller R, Stewart AF, Fu J, et al. RecET direct cloning and Red $\alpha\beta$ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat Protoc*. 2016; 11:1175–1190. [PubMed: 27254463]
45. Jiang W, Zhu TF. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc*. 2016; 11:960–975. [PubMed: 27101517]
- 46•. Jiang WJ, Zhao XJ, Gabrieli T, Lou CB, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of Chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun*. 2015; 6:6101. This paper describes a direct cloning method based on the in vitro cleavage of a target BGC from genomic DNA using CRISPR/Cas9. [PubMed: 25592476]
47. Hu S, Liu Z, Zhang X, Zhang G, Xie Y, Ding X, Mo X, Stewart AF, Fu J, Zhang Y, et al. “Cre/loxP plus BAC”: a strategy for direct cloning of large DNA fragment and its applications in *Photorhabdus luminescens* and *Agrobacterium tumefaciens*. *Sci Rep*. 2016; 6:29087. [PubMed: 27364376]
48. Dai RX, Zhang B, Zhao GP, Ding XM. Site-specific recombination for cloning of large DNA fragments *in vitro*. *Eng Life Sci*. 2015; 15:655–659.
- 49•. Montiel D, Kang HS, Chang FY, Charlop-Powers Z, Brady SF. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc Natl Acad Sci U S A*. 2015; 112:8953–8958. This paper describes a HR-based method for replacing promoters of a silent BGC on the shuttle vector. [PubMed: 26150486]

50. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun.* 2013; 4:2894. [PubMed: 24305602]
51. Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth Biol.* 2013; 2:662–669. [PubMed: 23968564]
52. Chao R, Yuan Y, Zhao H. Recent advances in DNA assembly technologies. *FEMS Yeast Res.* 2015; 15:1–9.
53. Zhang MM, Wang Y, Ang EL, Zhao H. Engineering microbial hosts for production of bacterial natural products. *Nat Prod Rep.* 2016; 33:963–987. [PubMed: 27072804]
54. Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, Zechel DL, Ma B, Magarvey NA. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat Commun.* 2015; 6:8421. The authors of this paper established a genomes-to-natural products (GNP) platform for predicting, combinatorializing and identifying the products of modular gene clusters (PKS and NRPS) from LC-MS/MS data of crude extracts. [PubMed: 26412281]
55. Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster AL, Magarvey NA. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat Chem Biol.* 2016; 12:1007–1014. [PubMed: 27694801]
56. Nguyen DD, Wu C-H, Moree WJ, Lamsa A, Medema MH, Zhao X, Gavilan RG, Aparicio M, Atencio L, Jackson C, et al. MS/MS networking guided analysis of molecule and gene cluster families. *Proc Natl Acad Sci U S A.* 2013; 110:2611–2620. [PubMed: 23359698]
57. Henke MT, Kelleher NL. Modern mass spectrometry for synthetic biology and structure-based discovery of natural products. *Nat Prod Rep.* 2016; 33:942–950. [PubMed: 27376415]
58. Mohimani H, Pevzner PA. Dereplication, sequencing and identification of peptidic natural products: from genome mining to peptidogenomics to spectral networks. *J Nat Prod.* 2015; 33:73–86.
59. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapon CA, Luzzatto-Knaan T, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016; 34:828–837. [PubMed: 27504778]
60. Gökay O, Albert K. From single to multiple microcoil flow probe NMR and related capillary techniques: A review. *Anal Bioanal Chem.* 2012; 402:647–669. [PubMed: 21969176]
61. Halabalaki M, Vougiannopoulou K, Mikros E, Skaltsounis AL. Recent advances and new strategies in the NMR-based identification of natural products. *Curr Opin Biotechnol.* 2014; 25:1–7. [PubMed: 24484874]

Highlights

- Most BGCs are silent due to their tight regulation under conventional laboratory culture conditions.
- New methods have been developed to activate silent BGCs in both native hosts and heterologous hosts.
- Computational tools have been developed for high-throughput identification of BGCs and natural products.

Figure 1.

Overview of the workflow for a modern natural product discovery strategy. Silent BGCs are first identified from genomic sequences by bioinformatics tools and then activated by different synthetic biology strategies. Products are characterized by high-throughput methods and the data obtained can further guide the next round genome mining.

Figure 2.

Workflow of the promoter knock-in strategy. Golden Gate and Gibson assembly are used to insert sgRNA (fusion of the crRNA and tracrRNA) and homology arms to a vector harboring a codon-optimized *SpCas9* gene to yield knock-in plasmids which are then transformed into the native host. The codon-optimized *SpCas9* cuts specific DNA sequence guided by sgRNA, and then inserts strong promoters into the chromosome via homologous recombination. Serial insertion of strong promoters upstream of each operon in a silent BGC can eventually result in a fully engineered gene cluster. Different combinations of engineered operons can also produce intermediates which provide important information on the biosynthetic mechanism. Fermentation followed by comparative analysis of metabolites by LC-MS or NMR enables the characterization of the corresponding natural products.

Figure 3.

Scheme for direct cloning *in vivo* (left) and *in vitro* (right). In the TAR based *in vivo* method, the genomic DNA containing the target BGC is first fragmented. The resulting fragments and the vector with homology arms to the target BGC are then co-transformed to either *S. cerevisiae* or *E. coli* (using the Red system or ET recombination system). The target BGC and the vector can form a circular DNA molecule through homologous recombination. In comparison, the *in vitro* method (CATCH) uses CRISPR/Cas9 to precisely cleave the target BGC. The resulting target fragment is purified and cloned into the vector through Gibson assembly. Both *in vivo* and *in vitro* methods require the screening of correct constructs.