Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination

Behnam Enghiad1,2,4, Chunshuai Huang1,4, Fang Guo2, Guangde Jiang1, Bin Wang1, S. Kasra Tabatabaei1, Teresa A. Martin2 & Huimin Zhao1,2,3 ✉

Direct cloning represents the most efficient strategy to access the vast number of uncharacterized natural product biosynthetic gene clusters (BGCs) for the discovery of novel bioactive compounds. However, due to their large size, repetitive nature, or high GC-content, large-scale cloning of these BGCs remains an overwhelming challenge. Here, we report a scalable direct cloning method named Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination (CAPTURE) which consists of Cas12a digestion, a DNA assembly approach termed T4 polymerase exo + fill-in DNA assembly, and Cre-lox in vivo DNA circularization. We apply this method to clone 47 BGCs ranging from 10 to 113 kb from both Actinomycetes and Bacilli with ~100% efficiency. Heterologous expression of cloned BGCs leads to the discovery of 15 previously uncharacterized natural products including six cyclic head-to-tail heterodimers with a unique 5/6/6/6/5 pentacyclic carbon skeleton, designated as bipentaromycins A–F. Four of the bipentaromycins show strong antimicrobial activity to both Gram-positive and Gram-negative bacteria such as methicillin-resistant Staphylococcus aureus, vancomycin-resistant Enterococcus faecium, and bioweapon Bacillus anthracis. Due to its robustness and efficiency, our direct cloning method coupled with heterologous expression provides an effective strategy for large-scale discovery of novel natural products.

1 Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. 2 Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. 3 Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. 4 These authors contributed equally: Behnam Enghiad, Chunshuai Huang. ✉ Email: zhao5@illinois.edu
Natural products or small molecules produced by living organisms have played an important role in human medicine and agriculture for centuries. Recent analysis of all FDA approved small molecule drugs between 1981 and 2019 has revealed that ~34% of these compounds are either natural products or natural products derivatives and a significant number of these compounds are produced by microorganisms\(^1\). With recent advances in genome sequencing and bioinformatics tools, microbial natural product biosynthetic gene clusters (BGCs) are uncovered at an unprecedented rate. However, only a relatively small fraction of these BGCs have been associated with known natural products\(^2\)-\(^4\). As a result, microorganisms potentially possess a huge reservoir of currently uncharacterized novel natural products\(^5\)-\(^7\).

To date, a wide variety of approaches have been developed to investigate these uncharacterized BGCs\(^6\)-\(^7\), which can be grouped into native host based\(^8\)-\(^9\) and heterologous host based\(^10\)-\(^11\) strategies. Due to lack of developed genetic tools, the native host based strategies are not generally applicable to most microorganisms. In contrast, since BGCs can be cloned from purified genomic DNA, the heterogeneous host based strategies can be applied to BGCs from all microorganisms including those yet to be cultivated in the laboratory. Traditionally, microbial BGCs have been cloned by construction of genomic DNA libraries and subsequent screening to find the correct clone\(^12\)-\(^13\). Although this random cloning strategy has successfully been used for cloning microbial BGCs, it is not suitable for large-scale discovery of novel natural products due to its untargeted nature and extensive screening requirement. Recently, with the availability of public genome sequences, a variety of methods for direct cloning of microbial BGCs have been developed\(^14\). In vivo homologous recombination strategies including site-specific recombination (SSR) in <i>Saccharomyces cerevisiae</i>\(^10\)-\(^15\) and linear-linear homologous recombination (LLHR) in <i>Escherichia coli</i>\(^16\)-\(^17\) take advantage of the high recombination efficiency of the cloning host to clone genomic fragments into plasmids. Unfortunately, many microbial BGCs have repetitive DNA sequences that can negatively interfere with successful recombination events and acquisition of correct clones using these strategies\(^15\). Alternatively, microbial BGCs can also be cloned using in vitro strategies. Cas9-assisted targeting of chromosomal segments (CATCH) method\(^18\) uses a combination of in vitro Cas9 digestion of genomic DNA in agarose gel-plugs and isothermal Gibson assembly\(^19\) for cloning genomic DNA fragments into linearized DNA vectors. Although this approach can potentially be used to clone BGCs with repetitive DNA sequences, due to the low efficiency of Gibson assembly in assembling DNA molecules with high GC-content\(^20\), this approach is not suitable for cloning large BGCs (>50 kb) from high GC-content organisms including Actinobacteria, one of the most gifted producers of natural products\(^21\).

Here, we report an inexpensive, rapid, robust, and highly efficient direct cloning method named Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination (CAPTURE), which allows direct cloning of large genomic fragments (up to 113 kb) including those with high GC-content and sequence repeats into functional vectors. We characterize the efficiency and robustness of CAPTURE by cloning 47 natural product BGCs from both Actinomycetes and Bacilli. We also demonstrate the utility of CAPTURE in large-scale discovery of natural products by direct cloning and heterologous expression of uncharacterized microbial BGCs, which has led to the discovery of 15 previously uncharacterized natural products.

**Results**

**Design of the CAPTURE method.** The workflow of CAPTURE is shown in Fig. 1a. In the first step, purified microbial genomic DNA is digested by Cas12a enzyme to release the BGC fragment. In the second step, two DNA receivers are amplified by polymerase chain reaction (PCR). The DNA receivers each carry a loxP site at their ends. To limit the possibility of creating false positive colonies, none of the DNA receivers carry both resistance marker and <i>E. coli</i> origin of replication. Other than these two essential elements, DNA receivers can also carry elements required for heterologous expression. This strategy would allow direct cloning of BGCs into vectors ready for heterologous expression. The DNA receivers are amplified from a set of universal receiver plasmids which can be designed for a specific heterologous host. Examples of universal receiver plasmids designed for <i>Streptomyces</i> and <i>Bacillus subtilis</i> heterologous hosts are shown in Supplementary Fig. 1. After PCR, the amplified DNA receivers are mixed with the digested genomic DNA in a T4 polymerase exo + fill-in DNA assembly reaction to join the three fragments and create a linear DNA product. In the third and final step, DNA assembly products are transformed into <i>E. coli</i> cells harboring a helper plasmid expressing Cre recombinase and phage lambda Red Gam protein for in vivo Cre-lox circularization. The resulting colonies are then checked for correct clones.

**Development of the CAPTURE method by creating an efficient in vivo DNA circularization system based on site-specific recombination.** The creation of a direct cloning method based on in vivo site-specific recombination was based on the fact that due to the creation of by-products such as DNA concatemers (e.g., Supplementary Fig. 2a), DNA circularization is the main limiting step in the assembly of large DNA molecules in vitro\(^22\). To potentially overcome this challenge, we sought to utilize a DNA circularization system in which instead of circular DNA, <i>E. coli</i> cells are transformed with linear DNA molecules which are able to circularize in vivo. By eliminating the formation of DNA by-products, this strategy can in turn increase the frequency of DNA circularization (Supplementary Fig. 2b).

In vivo circularization of linear DNA molecules can occur with the help of homologous recombination or site-specific recombinase enzymes\(^23\). Unlike homologous recombination which promotes recombination between all DNA sequences with high degrees of homology, site-specific recombination only occurs between defined DNA sequences which are recognized by the site-specific recombinase enzyme\(^24\). As a result, for DNA circularization, site-specific recombination offers an advantage over homologous recombination because of its ability in circularization of DNA molecules containing sequence repeats. Previously, a few systems based on site-specific recombinase proteins were developed for circularization of linear DNA molecules in <i>E. coli</i>\(^25\)-\(^26\). However, these systems either have low circularization efficiency or lack transient expression of recombinase and require RecBC deficient strains of <i>E. coli</i>. Therefore, we decided to create a general transient <i>E. coli</i> circularization system by placing the required elements for DNA circularization on a plasmid (a.k.a. helper plasmid). This circularization helper plasmid (Fig. 1b) carries three essential elements: (1) a site-specific recombinase to carry out DNA circularization; (2) phage lambda Red gam gene for inhibition of linear DNA degradation by <i>E. coli</i> RecBCD complex\(^27\); (3) pSC101 temperature-sensitive origin of replication to enable curing of <i>E. coli</i> cells from the helper plasmid after circularization\(^28\). We evaluated two frequently used site-specific recombinases, Flp and Cre, by using two test plasmids carrying either two FRT or loxP sites flanking a lacZ gene (Supplementary Figs. 3 and 4a). In the event of successful recombination, lacZ gene would be removed and colonies would show white color in blue-white screening. Thus the ratio of white colonies to the total number of
acquired colonies can be used to calculate recombination frequency. As shown in Fig. 1c, when Flp recombinase expression was fully induced, we observed only ~0.3% recombination frequency for the test plasmid. Contrarily, even leaky expression of Cre recombinase from L-arabinose inducible promoter was enough to promote recombination at ~78% frequency and when the expression was fully induced, Cre recombinase was able to carry out recombination at close to 100% frequency. We next assessed the efficiency of Flp and Cre helper plasmids in circularization of linear DNA molecules. This time, the test plasmids were amplified by PCR and the linear PCR products (Supplementary Fig. 4b) were transformed into *E. coli* cells harboring Flp (pBE11) or Cre (pBE12 and pBE14) helper plasmids. As shown in Fig. 1d, Cre helper plasmids were able to exhibit $10^8$–$10^9$ c.f.u/µg transformation efficiency for a 3 kb linear DNA molecule, which is essentially the same efficiency as transforming a circular DNA molecule with the same size (Supplementary Fig. 5).

Following creation of an efficient in vivo DNA circularization system in *E. coli*, we sought to directly compare the efficiency of DNA circularization in vitro and in vivo for large DNA molecules. For this purpose, two plasmid DNA molecules with sizes of 50 and 73 kb each carrying two *loxP* sites, were constructed (Supplementary Fig. 6). These plasmids were then linearized by restriction enzymes. For in vitro DNA circularization, the linearized DNA was ligated to itself using T4 DNA ligase and transformed into *E. coli* cells. For in vivo DNA circularization, the linearized DNA was transformed directly into *E. coli* cells and transformed into *E. coli* harboring pBE11 (Flp), pBE12 (Cre), pBE14 (Cre and recA1) helper plasmids. Both pBE12 and pBE14 *E. coli* cells exhibited transformation efficiencies similar to circular DNA. A comparison of in vitro versus in vivo circularization for two large (50 kb, 73 kb) linear DNA molecules. In vivo circularization showed ~33-fold and 150-fold higher frequency than in vitro circularization for 50 kb and 73 kb molecules, respectively. Circularization frequencies were calculated based on the number of colonies acquired for each circularization experiment in comparison to the number of colonies acquired after transforming the original circular DNA (see Methods for full description). Each experiment was performed in three biological replicates and data are presented as mean values ± standard deviation (SD). Source data are provided as a Source Data file.
cells harboring helper plasmid pBE14. As shown in Fig. 1e, for the 50 kb linear DNA, the frequency of in vivo DNA circularization was ~33-fold higher than that of in vitro circularization and as the DNA size increased to 73 kb, this ratio grew to ~150-fold. These results indicate that due to creation of DNA by-products (i.e., DNA concatemers created in the ligation reaction), in vitro DNA circularization for long DNA molecules occurs at low frequencies. However, by eliminating formation of by-products, the circularization frequency can be significantly increased using an efficient in vivo DNA circularization system.

**Development of the CAPTURE method by characterizing various genomic DNA digestion/DNA assembly combinations.**

To determine which genomic DNA digestion/DNA assembly combination will work most effectively with in vivo circularization for targeted cloning from microbial genomic DNA, we evaluated several systems. For genomic DNA digestion, we tested restriction enzymes and Cas12a. Unlike Cas9, Cas12a is able to create staggered cuts on DNA molecules and the resulting cohesive ends can potentially be used for DNA assembly applications. For DNA assembly, we examined three strategies: DNA ligation, isothermal Gibson assembly, and our newly developed DNA assembly approach termed T4 DNA polymerase exonuclease treatment (Fig. 2a). We determined the cloning efficiencies for four of the digestion and DNA assembly combinations (restriction enzymes/Gibson, FnCas12a/ligation, FnCas12a/Gibson, and FnCas12a/T4 exo + fill-in) using four high GC-content (72–77% range) BGCs from Actinomycetes with sizes ranging from 21-87 kb as model BGCs (Fig. 2b and Supplementary Table 1). For each cloning experiment, we selected at least seven colonies and analyzed the purified plasmids for each colony by restriction digestion. The cloning efficiencies were calculated as the ratio of correct colonies to the total number of checked colonies. As shown in Fig. 2b, the FnCas12a/T4 exo + fill-in strategy exhibited close to 100% cloning efficiency for all four target BGCs. RE restriction enzymes. For each cloning experiment, at least seven colonies were selected and the purified plasmids from each colony were analyzed by restriction digestion. The cloning efficiencies were calculated as the ratio of correct colonies to the total number of checked colonies. As shown in Fig. 2b, the FnCas12a/T4 exo + fill-in strategy exhibited close to 100% cloning efficiency for all four BGCs. Restriction enzymes/Gibson exhibited an average of ~63% cloning efficiency for 21, 38, and 62 kb target BGCs. However, we were unable to clone the 87 kb target BGC using this approach. FnCas12a/ligation exhibited close to 100% cloning efficiency for the 21 kb target BGC and ~57% efficiency for the 62 kb target BGC but only ~5% cloning efficiency for the 38 kb target BGC. We were also unable to clone the 87 kb target BGC using FnCas12a/ligation combination. The low efficiency of cloning for some target BGCs using FnCas12a/ligation may come from the fact that Cas12a enzymes do not generate defined sticky ends and also possess non-specific trans-cleavage activity on dsDNA. Consequently, ligation of sticky ends generated by Cas12a enzymes is not highly reliable. For FnCas12a/Gibson combination, we observed close to 100% cloning efficiency for the 21 kb target BGC and ~67% efficiency for the 38 kb target BGC. However, for cloning the larger target BGCs, this approach showed very low efficiencies (~10% for the 62 kb target BGC and ~5% for the 87 kb target BGC). Taken together, compared to the other three combinations, FnCas12a/T4 exo + fill-in, i.e., the CAPTURE method, exhibited the highest reliability and cloning efficiency for the four model BGCs.

**Large-scale cloning of uncharacterized BGCs and discovery of novel natural products.** To determine the overall robustness and cloning efficiency of CAPTURE, we used this method to clone a total of 43 uncharacterized natural product BGCs from 14 Streptomyces and three Bacillus species (Supplementary Figs. 7–13, Fig. 2c, and Table 1). The selected BGCs ranged from 10 to 113 kb in size. Out of the 43 selected target BGCs, we were able to...
from the 43 BGCs resulted in identification of seven BGCs with positive HPLC peaks (Fig. 3, Supplementary Figs. 14–18). To date, we have been able to identify the corresponding natural products for five of these seven BGCs. BGC #1 from Streptomyces sp. NRRL F-6131 encoding a putative type II polyketide synthases (PKSs) pathway was heterologously expressed in the model host S. avermitilis SUK17. HPLC analysis of the metabolite profile led to the discovery of six cyclic head-to-tail heterodimers with a unique 5/6/6/5/6 pentacyclic carbon skeleton, designated as dipentaromycins A–F (1–6, Fig. 3a). The chemical structures of these compounds were elucidated by high-resolution electrospray ionization mass spectrometry (HRE-SIMS) and nuclear magnetic resonance (NMR) (Supplementary Figs. 19–60, Supplementary Tables 2–4). One of these compounds was named as dipentaromycin E (5) and its molecular formula was established as C_{42}H_{49}O_{15} by the negative HRE-SIMS (m/z 777.1619 [M−H]−, calc for 777.1614, Supplementary Fig. 47), indicating 31 degrees of unsaturation. Detailed analyses of the 1H, 13C, and 2D NMR spectroscopic data revealed that 5 was a heterodimer and allowed the establishment and connectivity of units A (rings A/B/C/D/E) and B (rings A′/B′/C′/D′/′E′) via C-2′−C-8′ and C-8′−C-′2. This assignment was further validated by X-ray crystallographic analysis with Flack parameter 0.0075 (CCDC no. 1993850; Fig. 3a, Supplementary Table 5). We determined the antimicrobial activity of compounds 3–6 using a series of Gram-positive and Gram-negative bacteria including six ESKAPE pathogens commonly associated with antimicrobial resistance and bioweapon Bacillus anthracis as well as five commonly used fungal species (Supplementary Table 6). Interestingly, these compounds showed potent antibacterial activity towards both Gram-positive bacteria (e.g., Bacillus anthracis str. Sterne with minimal inhibitory concentration (MIC) of 4 µg/mL, methicillin-resistant Staphylococcus aureus USA300 with MIC of 32 µg/mL, vancomycin-resistant Enterococcus faecium U503 with MIC of 32 µg/mL) and Gram-negative bacteria (e.g., Pseudomonas fluorescens Pf-5 with MIC of 16–32 µg/mL). These four compounds showed different MIC values toward some of the bacteria (e.g., for Staphylococcus epidermidis 15 × 154, 3 and 4 had a MIC value of 4 µg/mL while 5 and 6 had a MIC value of 32 µg/mL). Notably, none of these compounds showed any antifungal activity. Metabolic profiling analyses of the other positive BGCs led to the isolation of three angucyclines (7, 8, and 10) with post-PKS tailoring enzyme-modified angular tetracyclic structural skeletons and an unexpected 2,3-tricyclic derivative (9), two anthraquinone or naphthoquinone derivatives (11 and 12) containing an acetylcysteine moiety, two known citreodiols and a oxidative derivative (10), and, to the best of our knowledge, novel molecules allenmycin A and B (16 and 17) (Fig. 3b,c, Supplementary Figs. 61–106, Supplementary Tables 7–10).

To gain insight into the potential reasons for lack of compound production from 36 of the 43 uncharacterized BGCs, we randomly selected 10 BGCs for transcriptional analysis in the heterologous host. For each of the 10 BGCs, we chose three targets corresponding to predicted core biosynthetic genes and performed RT-PCR analysis after 24 and 48 h of cell growth (Supplementary Figs. 107–111). For 6 of the 10 analyzed BGCs, we observed either very low or no detectable levels of RNA transcription after 30 cycles of RT-PCR for one or more analyzed biosynthetic genes. These results indicate that at least for 60% of the analyzed BGCs, lack or very low levels of transcription for the main biosynthetic genes plays an important role in lack of detectable compound production. As a result, strategies to increase levels of transcription such as insertion of strong promoters in front of core biosynthetic genes, overexpression of pathway-specific activators, or deletion of pathway-specific repressors can potentially lead to successfully clone all BGCs without any failure. The CAPTURE method exhibited close to 100% cloning efficiency for all target BGCs tried including 11 target BGCs between 50 and 99 kb and three target BGCs larger than 100 kb (Table 1). These results indicate that the CAPTURE method is an extremely robust and highly efficient direct cloning approach for cloning BGCs ranging from 10 to 113 kb in size regardless of their GC-content.

Next, we heterologously expressed all 43 cloned uncharacterized BGCs in either Streptomyces avermitilis, Streptomyces lividans, or Bacillus subtilis. HPLC analysis of crude extracts

Table 1 Summary of the results from the 43 uncharacterized BGCs cloned by CAPTURE.

| Target BGC | Size | Predicted BGC type | Cloning efficiency | HPLC |
|-----------|------|--------------------|-------------------|------|
| Streptomyces noursei ATCC 11455 | C29 | 26 kb Lanthipeptide | 100% | No new peak |
| C30 | 76 kb NRPS | 100% | No new peak |
| Streptomyces inopinus NRRL 8089 | C5 | 13 kb Siderophore | 100% | No new peak |
| Streptomyces lavendulae NRRL B-24963 | C19 | 22 kb Lasso peptide | 100% | No new peak |
| Streptomyces griseorubescens ATCC 14511 | C6 | 59 kb NRPS, type I PKS | 100% | 1 new peak |
| C7 | 113 kb type I PKS, NRPS | 100% | 4 new peaks (colonies) |
| C11 | 24 kb Lasso peptide | 100% | No new peak |
| C15 | 11 kb Siderophore | 100% | No new peak |
| C20 | 100 kb NRPS, type I PKS, Lanthipeptide | 86% | 12 new peaks (colonies) |
| C31 | 20 kb Lasso peptide | 100% | No new peak |
| C39 | 103 kb NRPS, type III PKS | 100% | 10 new peaks (colonies) |
| C41 | 45 kb type I PKS | 86% | No new peak |
| Streptomyces griseoflavus NRRL B-1830 | C28 | 65 kb NRPS-like, type I PKS | 100% | No new peak |
| Streptomyces sp. NRRL B-11253 | C7 | 55 kb Terpene-arylpolypeptide | 100% | No new peak |
| Streptomyces sp. NRRL F-5635 | C10 | 72 kb type II PKS | 100% | No new peak |
| Streptomyces sp. NRRL F-6131 | C4 | 40 kb type II PKS | 100% | 6 new peaks |
| C5 | 50 kb type II PKS | 100% | No new peak |
| C8 | 73 kb type I PKS-NRPS | 100% | No new peak |
| C9 | 51 kb NRPS | 100% | No new peak |
| C11 | 46 kb type II PKS-NRPS | 100% | No new peak |
| C14 | 29 kb Lanthipeptide | 100% | No new peak |
| C33 | 36 kb type I PKS-NRPS | 100% | No new peak |
| Streptomyces alni NRRL B-24661 | C10 | 50 kb TIPKS, hglE-KS | 57% | 1 new peak |
| C24 | 48 kb NRPS-like | 100% | No new peak |
| Streptomyces alni NRRL B-24661 | C7 | 59 kb type II PKS | 100% | No new peak |
| C11 | 46 kb type II PKS | 100% | 4 new peaks |
| C28 | 65 kb NRPS-like, type II PKS | 100% | No new peak |
| Streptomyces bicolor NRRL B-3897 | C6 | 47 kb CDPS | 100% | No new peak |
| C15 | 67 kb type II PKS | 100% | 3 new peaks |
| Streptomyces cyanosporus NRRL B-2570 | C2 | 69 kb type II PKS | 100% | 2 new peaks |
| C24 | 43 kb NRPS-like | 100% | No new peak |
| Streptomyces griseoluscs NRRL B-5429 | C8 | 27 kb type I PKS | 100% | No new peak |
| C11 | 46 kb type I PKS | 100% | No new peak |
| C20 | 28 kb type I PKS | 100% | No new peak |
| C21 | 35 kb type I PKS | 71% | No new peak |
| C25 | 28 kb type I PKS | 100% | 2 new peaks |
| C38 | 25 kb type I PKS | 100% | 3 new peaks |
| Streptomyces peucetius WC-3868 | C19 | 45 kb type I PKS | 100% | No new peak |
| Bacinilis meguetan DSM 319 | C2 | 26 kb Phosphonate | 100% | No new peak |
| Bacinilis meguetan QM B-1551 | C1 | 19 kb Ladderane | 100% | No new peak |
| C2 | 19 kb Lanthipeptide | 100% | No new peak |
| Bacillus thuringiensis YST-0202 | C14 | 14 kb LAP | 100% | No new peak |
| C2 | 14 kb Arylpolyene | 100% | No new peak |
the production of associated compounds from these BGCs in a heterologous host. For the remaining 40% of the BGCs, while low levels of transcription from other essential biosynthetic genes might still play a role in lack of compound production, a variety of previously described heterologous expression limiting factors such as absence of required precursors or other missing host factors can also explain lack of detectable compound production in the heterologous host.

**Discussion**

By combining Cas12a digestion, T4 polymerase exo + fill-in DNA assembly, and in vivo Cre-lox DNA circularization, we were able to create an extremely robust and highly efficient method named CAPTURE for direct cloning of natural product BGCs. CAPTURE uses Cre-lox site-specific recombination instead of homologous recombination for the creation of circular DNA molecules which allows cloning targets with repetitive DNA
sequences. Moreover, the use of T4 polymerase exo + fill-in DNA assembly allows efficient assembly of DNA molecules with high GC-content. As a result, using CAPTURE, BGCs ranging from 10-113 kb regardless of their GC-content or repetitive DNA sequence can be directly cloned into vectors ready for heterologous expression in as short as 3-4 days. Compared to other direct cloning methods (see Supplementary Table 1), CAPTURE also offers high-levels of robustness and efficiencies. The high success rate of CAPTURE combined with short time-frame and low cost, makes this method suitable for large-scale cloning of microbial BGCs and discovery of novel natural products.

With the increasing number of sequenced microbial genomes, more and more uncharacterized natural product BGCs are being uncovered. However, identification and structure elucidation of the compounds associated with these BGCs has remained a major challenge. Direct cloning and heterologous expression represent the most efficient strategy for study of these uncharacterized BGCs and discovery of their associated compounds. As demonstrated in this study, direct cloning and heterologous expression of intact uncharacterized BGCs results in the production of associated compounds for ~16% of the BGCs. It was estimated that the genus of Streptomyces alone has the capacity to produce as many as 100,000 natural products with only a small fraction of them discovered so far. However even with ~16% success rate for heterologous expression, by utilizing the large-scale cloning capabilities of CAPTURE, a significant number of natural products can be discovered at an accelerated rate. Discovery of bipenamycins with strong antibacterial activity toward meticillin-resistant Staphylococcus aureus, vancomycin-resistant Enterococcus faecium, and bioweapon Bacillus anthracis only provides an example of the multitude of uncharacterized bioactive compounds present in microorganisms. Therefore, we envision CAPTURE to become the method of choice for direct cloning of uncharacterized BGCs and large-scale discovery of novel natural products with unique bioactivities for use in drug development and agricultural applications.

Methods

Bacterial strains and reagents. E. coli strain NEB10β (New England Biolabs, MA) was used for all cloning experiments. E. coli strains WM4026 and FD25867/pUZ8002 were used as conjugation donors for Streptomyces heterologous hosts. S. avermitilis SUKA17 and S. lividans TK24 were used for heterologous expression of natural product BGCs cloned from Actinomycete strains. B. Subtilis JH642 was used for heterologous expression of natural product BGCs cloned from Bacillus strains. E. coli conjugation donor strain WM6026 was a gift from William Metcalf (University of Illinois at Urbana-Champaign). S. avermitilis SUKA17 was a gift from Haruo Ikeda (Kitasato University, Japan). S. griseorhizogenes ATCC 14511 was ordered directly from ATCC. All remaining Actinomycete strains were obtained from the Agricultural Research Service (NRRl) culture collection, Peoria, IL. All Bacillus strains were obtained from Bacillus Genetic Stock Center (Columbus, OH). Restriction enzymes, NEBuffers, T4 DNA polymerase, E. coli DNA ligase, dNTPs, NAD⁺, NEBuilder HiFi DNA assembly master mix, and Q5 DNA polymerase were purchased from New England Biolabs (Ipswich, MA). All DNA oligonucleotides were ordered from Integrated DNA Technologies (Coraliale, IA).

Genomic DNA isolation. Actinomycete strains were grown at 30 °C in 100 mL of modified MYG medium (10 g/L malt extract broth, 4 g/L yeast extract, 4 g/L glucose) using 250 mL baffle flasks containing glass beads. Bacillus strains were grown at 30 °C in 100 mL of LB medium using 250 mL baffle flasks. The cultures were allowed to reach late exponential—stationary growth phase and then centrifuged at 3220 × g for 15 min. After removing the supernatant, the cells were fully resuspended in 12 mL of cell resuspension buffer (50 mM Tris-HCl pH 8.0, 2.5 mM EDTA) and transferred into 50 mL conical tubes. A combination of cell lysis enzymes (lysosyme 30 mg, achromopeptidase 20 mg (if necessary), mutanolysin 300 U (if necessary)), and 0.6 mg RNAse A) was added to the sample and gently mixed. The mixture was incubated at 37 °C for 1-2 h to allow cell lysis and then 6 mg of proteinase K was added, gently mixed and the sample was continued to be incubated at 37 °C for 1 more hour. Following this incubation, 1.2 mL of 10% sodium dodecyl sulphate (SDS) was added, gently mixed and the sample was incubated at 50 °C for 1-2 h until the solution became clear. Recovery of genomic DNA from cell lysis was performed by phenol-chloroform extraction. A total of 15 mL of phenol-chloroform-isomyl alcohol (25:24:1, v/v, pH 8.0) was added to the sample. The mixture was placed on a rocker to allow gentle mixing without creating emulsion. Once the aqueous phase became completely white, the sample was centrifuged at 22,000 × g for 45 min at room temperature. Without disturbing the precipitated proteins, the aqueous phase was transferred into a new 50 mL conical tube and 10 mL of chloroform was added. The sample was mixed as described above and centrifuged at 22,000 × g for 10 min. Without disturbing the precipitated proteins, the aqueous phase was aliquoted into 1.7 mL centrifuge tubes and the DNA was recovered using isopropanol and sodium acetate precipitation. The DNA pellets were washed 3 times with 70% ethanol (v/v) and air-dried. The DNA was rehydrated in 10 mM Tris-HCl pH 8.0 and incubated at 50 °C with occasional tube inversion until fully dissolved and stored at -80 °C.

Expression and purification of FnCas12a enzyme. The FnCas12a expression plasmid pET28-FnCas12a-TEV was obtained from Jin Wang (Shanghai Institute of Plant Physiology & Ecology, Chinese Academy of Sciences, China) and transformed into E. coli KRX (Promega, WI) according to manufacturer’s protocol. The strain was cultivated overnight at 37 °C in 5 mL LB medium supplemented with 50 µg/mL kanamycin. Following overnight incubation, 2 mL of the overnight culture was added to 400 mL Terrific Broth containing 50 µg/mL kanamycin and incubated at 37 °C until the OD₆₀₀ of 1.2-1.5 was reached. The culture was then cold shocked by incubation in an ice bath for 15 min and protein expression was induced by addition of isopropl β-D-1-thiogalactopyranoside (IPTG) and L-ramhnose to final concentrations of 1 mM and 0.1% (w/v), respectively. Expression was continued by incubation at 37 °C for 16-20 h. The culture was harvested at 30 °C for 10 min and after removing the supernatant, the cell paste was stored at -80 °C until purification. For the purification step, frozen cells were thawed at room temperature and then fully resuspended in 30 mL of binding buffer (20 mM Tris-HCl pH 8.0, 1 M NaCl, 20 mM imidazole). The resuspended cells were lysed by sonication on ice for 10 min pulse time (30% amplitude, 5 s on, 5 s off). The solution was centrifuged 3 times at 20,000 × g at 4 °C and the supernatant was used for purification using fast protein liquid chromatography (FPLC) and 1 mL Histrap HP column (GE Healthcare, IL). After equilibration with 5 column volume (CV) of binding buffer, the column was loaded with the supernatant and washed with 10 CV of washing buffer. The column was then eluted using a gradient from 20 to 500 mM imidazole into 15, 1 mL fractions, and fractions containing protein of the expected size were pooled and concentrated using Amicon Ultra 15 50 kDa filtration units (MilliporeSigma, MA). The purified proteins were diluted to ~10 μM concentration using storage buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 15% (v/v) glycerol) and the aliquots were stored at -80 °C.

DNA design for FnCas12a guide RNA and DNA receivers. FnCas12a guide RNA molecules were transcribed from dsDNA templates with a size of 60 bp generated by hybridization of two ssDNA oligonucleotides. The dsDNA templates carry sequences for T7 promoter, FnCas12a gRNA scaffold, and a BGC specific spacer. For all guide RNA molecules, the spacer length of 18 nt was used and the spacer GC-content was kept in 40–70% range. In the selection of FnCas12a targets, to direct the generated BGC fragment ends, a 39 bp sequence upstream of the PAM sequence was kept inside the BGC fragment. The chosen gRNAs used for each BGC were checked by Basic Local Alignment Search Tool (BLAST) to avoid occurrence of undesired cleavage in the middle of the BGC fragment.

DNA receivers were amplified using Q5 DNA polymerase following manufacturer’s protocol using 1 ng of plasmid DNA template. For DNA receivers used in cloning target BGCs of larger than 90 kb, to avoid occurrence of false positive colonies, after PCR, 1 μL of DpnI restriction enzyme was added to the mixture and incubated at 37 °C for 45 min. The PCR products were run on 1% (w/v) agarose gel for 30 min at 120 V and purified by Zymoclean Gel DNA Recovery Kit (Zymo Research, CA) following manufacturer’s protocol. For T4 DNA polymerase exo + fill-in assembly, each DNA receiver carried 39 bp of homology to the BGC fragment. This 39 bp homology included 24 bp sequence downstream of the PAM sequence (including the 18 bp of spacer), 3 bp of the TTV (V=A, C, or G) as a spacer sequence, and 12 bp of the GC-content for the 39 bp sequence was kept in 40–70% range while avoiding 4 or more consecutive G nucleotides. Examples of ssDNA templates and primers used for receiver amplification can be found in Supplementary Data 1.

Guide RNA transcription and genomic DNA digestion. For each guide RNA, 1 μL of each 100 μM forward and reverse template ssDNA oligonucleotides were mixed in a 20 μL reaction containing 2 μL of 10× NEBuffer 3.1. The mixture was incubated at 98 °C for 3 min and the temperature was then slowly lowered with the rate of 0.1 °C/s until 10 °C was reached. A total of 2 μL of the hybridized mixture was used as template for in vitro RNA transcription using HiScribe T7 quick high yield RNA Synthesis kit (New England Biolabs, MA) following manufacturer’s protocol for transcription of short RNA molecules. After RNAse treatment, the described above and centrifuged at 22,000 × g for 10 min. Without disturbing the precipitated proteins, the aqueous phase was aliquoted into 1.7 mL centrifuge tubes and the DNA was recovered using isopropanol and sodium acetate precipitation. The DNA pellets were washed 3 times with 70% ethanol (v/v) and air-dried. The DNA was rehydrated in 10 mM Tris-HCl pH 8.0 and incubated at 50 °C with occasional tube inversion until fully dissolved and stored at -80 °C.
**Preparation of electrocompetent \( E. coli \) cells for Cre-lox in vivo recombinacion**

\( E. coli \) NEB10β cells harboring pBE14 circularization helper plasmid were grown overnight at 30 °C in modified SOB medium (20 mL Bacto tryptone, 5 g/L Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl) supplemented with 8 μg/mL tetracycline hydrochloride. Following overnight incubation, 100 μL of the overnight culture was added to 10 mL of modified SOB medium supplemented with 8 μg/mL tetracycline hydrochloride and incubated at 30 °C. After 1 h of incubation (OD\(_{600}\) of ~0.2), 100 μL of 1 M L-arabinose was added to the culture to induce cassette excision. The cells were grown at 30 °C until the OD\(_{600}\) of 0.45-0.55 was reached (~1.5 h after induction). The culture was then centrifuged at 3,220 × g for 7 min at room temperature and the supernatant was immediately removed. The cells were gently resuspended in 1 mL of room temperature 10% (v/v) glycerol and centrifuged at 20,000 × g for 30 s. After washing, without disturbing the cell pellet, the supernatant was gently removed and this washing step was repeated 2 more times. The cells were then resuspended in a final volume of 70 μL 10% (v/v) glycerol using wide-bore pipette tips and transformed with DNA using Gene Pulser XCell Electroporation system (Bio-Rad, CA). For cloning BGCs, electroporation was performed by gentle mixing of 3

**In vitro assembly of digested genomic DNA and DNA receivers**

Gibbon iso-thermal assembly of DNA molecules was performed by adding 3-3.75 μg of digested DNA to 100 ng of DNA receiver amplified from plasmid pBE44, 35 ng of DNA receiver amplified from plasmid pBE45, and 7.5 μL of 2x NEBuilder HiFi DNA assembly mastermix in a 15 μL reaction. After very gentle mixing using wide-bore pipette tips, the mixture was incubated at 50 °C for 90 min and stored at 10 °C until transformation.

For cloning target BGCs of larger than 50 kb, to increase DNA transformation efficiency, instead of modified SOB medium containing 5 g/L yeast extract, the overnight culture was added to modified SOB medium containing 1 g/L yeast extract. The induction and harvesting steps were carried out at the same OD\(_{600}\) as mentioned above.

For cloning target BGCs of larger than 50 kb, to increase DNA transformation efficiency, instead of modified SOB medium containing 5 g/L yeast extract, the overnight culture was added to modified SOB medium containing 1 g/L yeast extract. The induction and harvesting steps were carried out at the same OD\(_{600}\) as mentioned above.

**Analysis of cloned BGCs**

Upon electroporation, 1 mL of LB medium supplemented with 5 mM MgCl\(_2\) (ref. 57) was added to the cuvette and the suspended cells were transferred into 14 mL round-bottom Falcon tubes. The culture incubated at 37 °C for 25 min, with 500 μL of LB agar plates containing appropriate antibiotics (50 μg/mL ampicillin or 100 μg/mL carbenicillin) with blue/white screening and incubated at 37 °C until colonies appeared. For target BGCs of larger than 90 kb, the cells were plated on LB agar plates containing 100 μg/mL ampicillin supplemented with 0.2% (w/v) glucose. For analysis of correct clones from plasmids harboring the 15A origin of replication, at least 7 single white colonies were picked and grown at 37 °C in LB medium supplemented with appropriate antibiotics. The plasmid DNA was purified from the cultures using Qiagen Spin Miniprep Kit (Qiagen, Germany) following manufacturer’s protocol. The purified plasmids were then digested by appropriate restriction enzymes and the samples were analyzed by gel electrophoresis. A full DNA map for the cloned BGC plasmid was created based on the BGC and receiver DNA sequence and the resulting restriction digestion patterns were compared to the correct digestion pattern predicted by the SnapGene software. For analysis of correct clones from plasmids harboring the BAC origin of replication, at least 7 single white colonies were picked and grown at 37 °C in LB medium supplemented with 0.1% (w/v) glucose and appropriate antibiotics and grown at 37 °C. The BACmid DNA was purified from the cultures using Plasmid Midi kit (Qiagen, Germany) following manufacturer’s protocol. After Plasmid-Safe DNase (Lucigen, WI) treatment of purified DNA in Cutsmap buffer (New England Biolabs, MA), the DNA was digested by addition of appropriate restriction enzymes to the same mixture and the samples were analyzed by agarose gel electrophoresis.

**Heterologous expression, product extraction, and HPLC analysis**

The correct recombinant plasmids containing BGCs from Actinomycetes were transformed into \( E. coli \) WM6026 (supplemented with 2.6-diaminopimelic acid at the final concentration 40 μg/mL for growth) or \( E. coli \) ET100D7/PUB0802. The plasmids were then transferred into S. avermitilis SUKA17 and S. lividans TK24 hetero-oligomers by intergeneric conjugation procedures68. Following conjugation, exconjugants were randomly selected and inoculated into liquid MMY medium (Malt extract 10 g/L, Yeast extract 4 g/L, Maltose 4 g/L, pH 7.2-7.4) and incubated at 28 °C for 4-5 days. BGC plasmids cloned from Bacillus strains were introduced into \( B. subtilis \) HB642 + sip hetero-oligomer by natural competence transformation. After transformation, recombinant strains carrying correct integration at the amyE loci were cultivated in liquid LB medium supplemented with 100 μg/mL spectinomycin for 3 days at 30 °C. After growth, the fermentation cultures from both \( S. streptomyces \) and \( B. subtilis \) heterologous hosts were harvested and centrifuged. The mycelia and supernatant solution were then separated and the mycelia were extracted with equal volume of acetone. After removal of the acetone by evaporation, the extract of mycelia and the supernatant then were extracted with equal volume of ethyl acetate. The ethyl acetate fraction was evaporated to dryness under reduced pressure. The crude extract was subjected to normal phase silica gel (200-400 mesh) column chromatography eluted with a gradient solvent of hexane/ethyl acetate (from 1000:0 to 1:000, v/v). Subfractions were further successively purified by Sephadex LH-20 column chromatography and reversed-phase semi-preparative high performance liquid chromatography.
performance liquid chromatography (HPLC) to yield compounds. The semi-preparative HPLC was carried out on an Agilent HPLC using the following solvent systems: A (water, supplemented with 0.1% trifluoroacetic acid) and B (acetonitrile supplemented with 0.1% trifluoroacetic acid).

RT-PCR analysis of uncharacterized BGCs in the heterologous host. For analysis of BGCs in S. avermitilis heterologous host, each strain was first inoculated into liquid MTM medium and incubated at 28 °C for 48 h. 500 µL of the initial culture was then transferred to 10 mL of fresh liquid MTM medium and incubated at 28 °C. Samples for RT-PCR analysis were taken after 24 h and 48 h of growth and the cell pellets were stored at -80 °C until RNA extraction. For analysis of BGCs in B. subtilis heterologous host, each strain was first inoculated into liquid LB medium supplemented with 100 µg/mL spectinomycin and incubated at 30 °C for 24 h. 500 µL of the initial culture was then added to 10 mL of fresh LB medium supplemented with 100 µg/mL spectinomycin and incubated at 30 °C. Samples for RT-PCR analysis were taken after 24 h and 48 h of growth and the cell pellets were stored at -80 °C until RNA extraction. For RNA extraction, 100 µL of frozen cell pellets were first grinded using a mortar and pestle cooled with liquid nitrogen. Total RNA was then purified from the ground cells using PureLink RNA mini kit (ThermoFisher Scientific, MA) following manufacturer’s protocol. 4 µg of total RNA was first incubated with edDNase enzyme (ThermoFisher Scientific) to remove any contaminating genomic DNA and reverse transcription was then performed by SuperScript IV VILO mastermix (ThermoFisher Scientific). The resulting cDNA product was used as template for RT-PCR analysis of different biosynthetic genes using Q5 DNA polymerase. For each biosynthetic gene, a 100-bp PCR product was used as template for RT-PCR analysis of different 1H, 13C, and HSQC NMR spectroscopic data (Supplementary Figs. 48-52) data suggested that this was a heterodimer and allowed the establishment and confirmation of the structures of compound 2 (δH 1.75, 5.00 H, C 35.0). Finally, the structure of compound 5 was determined to be C24H23NO8S by HRESIMS (m/z 562.1025, Supplementary Fig. 95). The connectivity of the two units via C-2 was established. In S. avermitilis and B. subtilis respectively.

In vitro vs. in vivo circularization for long linear DNA molecules. To perform the in vitro vs. in vivo circularization, 10 µg of the 50 and 73 kb plasmids (Supplementary Fig. 6) were first linearized by 120 U of restriction enzymes in 300 µL reactions by incubation at restriction enzyme’s optimal temperature for at least 6 h. Following digestion, the reaction mixture was treated with 3 µL of 20 mg/mL of phenol:chloroform extraction in 5PRIME PLG light phase-lock gel tubes as described previously. The DNA was recovered using phase-lock gel tubes and the DNA pellets were washed 2 times with ethanol and sodium acetate precipitation and the DNA pellets were washed 2 times with ethanol. DNA was recovered using phase-lock gel tubes and the DNA pellets were washed 2 times with ethanol and sodium acetate precipitation and the DNA pellets were washed 2 times with ethanol.

The molecular formula of compound 7 was determined to be C36H48O31 by HRESIMS (m/z 877.1619 [M+H]+, Supplementary Fig. 61). 1H NMR and 13C NMR spectroscopic data (Supplementary Figs. 62 and 63, Supplementary Table 7) of 7 were similar to those of fujianmycin A44, except an additional hydroxyl group attached to C-6 in 7. This assignment was further supported by HMBC correlations (Supplementary Fig. 60) from H-6/5′-H to C-6a, C-12b, and from 6-NH2 to C-6a. The NOESY correlation (Supplementary Fig. 67) between Me-13 and H-4 indicated that they were in the same orientation. Biogenetically, considering that the coupling constant value (JH13-H14 9.3 Hz) between H-3 and H-4 was almost identical to that of fujianmycin A (JH13-H14 9.5 Hz), the absolute configurations of the chiral centers of 7 were assigned as R and 4. Consequently, the structure of 7 was determined and designated as 6-hydroxyfujianmycin A1.

6-hydroxyfujianmycin A1 (8) has the same molecular formula of C36H48O31 as that of 7, as determined by HRESIMS (m/z 877.1719 [M+H]+, Supplementary Fig. 68). The NMR spectroscopic data (Supplementary Figs. 69–74, Supplementary Table 7) of 8 were strikingly similar to those of 7. The coupling constant value (JH13-H14 2.9 Hz) between H-3 and H-4 indicated it was cis-configuration. Moreover, the chemical shift changes of those carbons surrounding C-4 suggested that 8 was a stereoisomer of 7 at C-4 position. Therefore, the absolute configurations of 8 were tentatively deduced as 3R and 4S. Consequently, the structure of 7 was determined and the compound was named as 6-hydroxyfujianmycin A1.

Structure elucidation. NMR data were analyzed by MestReNova software. HRESIMS data were analyzed by Masslynx software. X-ray single crystals were visualized by Olex2. Bipentamycin structures were drawn using ChemBioDraw.

Compound 5 was obtained as a red crystal. The molecular formula of 5 was established as C19H20O5 by the negative HRESIMS (m/z 337.0717 [M−H]-, Supplementary Fig. 51). The molecular formula of the compound was determined as C19H20O5 by the negative HRESIMS (m/z 337.0717 [M−H]-, Supplementary Fig. 51). The molecular formula of the compound was determined as C19H20O5 by the negative HRESIMS (m/z 337.0717 [M−H]-, Supplementary Fig. 51).
Supplementary Table 9) were similar to those of griseusin B44, except for the presence of an N-acetylcysteine moiety at C-4 and a ketone carbonyl group at C-17 in 12. These assignments were confirmed by HMBC correlations from H-4 to C-17, from H-1 to C-4, as well as from H-16/H-18 to C-17 (δ 205.0). Finally, 12 was named N-acetylcysteigriseusin.

Compound 13 was assigned a molecular formula of C_{19}H_{19}O_{11} based on the basis of HR/EI-MS (m/z 395.1022 [M + H]+), calculated for 395.1016, Supplementary Fig. 101. The 1D and 2D NMR spectroscopic data (Supplementary Figs. 102-106, Supplementary Table 10) of 13 were highly similar to those of citreodiol. The only difference was that a hydroxyl group presented at C-7 (δ 73.1) in citreodiol was found to be changed to a keto group (δ 218.3) in 13, which was further supported by HMBC correlations from H-2 (δ 2.22) and 6-Me (δ 1.42) to C-7. Ultimately, compound 13 was designated as 2,6-dimethyl-6-hydroxy-7-oxo-2,4-octadienoic acid methyl ester.

Physiochemical properties for compounds 1-13. Bipentaromycin A (1). Red powder; 1H and 13C NMR: see Supplementary Table 2, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 377.1043; found, 377.1145.

Bipentaromycin A (1). Red powder; 1H and 13C NMR: see Supplementary Table 2, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 377.1043; found, 377.1145.

Bipentaromycin B (2). Red powder; 1H and 13C NMR: see Supplementary Table 2, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 377.1043; found, 377.1145.

Bipentaromycin C (3). Red powder; 1H and 13C NMR: see Supplementary Table 3, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 749.1301; found, 749.1320.

Bipentaromycin D (4). Red powder; 1H and 13C NMR: see Supplementary Table 3, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 749.1301; found, 749.1320.

Bipentaromycin E (5). Red crystal; 1H and 13C NMR: see Supplementary Table 4, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 777.1614; found, 777.1619.

Bipentaromycin F (6). Red powder; 1H and 13C NMR: see Supplementary Table 4, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 777.1614; found, 777.1645.

6-Hydroxyfujianmycin A (7). Yellow powder; 1H and 13C NMR: see Supplementary Table 7, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 337.0718; found, 337.0719.

6-Hydroxyfujianmycin A (7). Yellow powder; 1H and 13C NMR: see Supplementary Table 7, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 337.0718; found, 337.0719.

2.3-secopredagogue (8). Yellow powder; 1H and 13C NMR: see Supplementary Table 8, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 327.1238; found, 327.1239.

Pregilvocarcin W (10). Yellow powder; 1H and 13C NMR: see Supplementary Table 8, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 318.0772; found, 318.0771.

N-Acetylcysteinansaoquinone (11). Yellow powder; 1H and 13C NMR: see Supplementary Table 9, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 484.1072; found, 484.1093.

N-Acetylcysteigriseusin (12). Yellow powder; 1H and 13C NMR: see Supplementary Table 9, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 484.1072; found, 484.1093.

2,6-Dimethyl-6-hydroxy-7-oxo-2,4-octadienoic acid methyl ester (13). White powder; 1H and 13C NMR: see Supplementary Table 10, HRMS (m/z): [M + H]+ calcd. for C_{19}H_{13}O_{11}, 195.1016; found, 195.1027.

Antimicrobial assays. The antimicrobial activities of bipentaromycins were measured against five fungi and 17 bacteria including nine pathogenic strains using broth microdilution method44. These indicator strains were: Bacillus subtilis ATCC 14948, Escherichia coli ATCC 25922, Staphylococcus aureus ATCC 25923, Salmonella typhimurium ATCC 14028, Enterococcus faecalis ATCC 29212, and Mycoplasma pulmonis ATCC 41674. Fungi and bacteria were cultured in Brain Heart Infusion (BHI) at 37 °C. Overnight cultures were diluted into fresh culture medium and re-inoculated in 96-well microtitre plates, which were supplemented with compounds ranging from 128 to 0.25 μg/mL. The Minimum Inhibitory Concentration (MIC) value was the concentration that resulted in no visible growth after 16–18 h cultivation. All experiments were tested in triplicates. Ampicillin (Amp) was used as a positive control for antibacterial assays.

Data availability Source data are provided with this paper. Deposition number of crystallographic data for 5 is CCDC 1993850. The data supporting the findings of this study are available within the article and its Supplementary Information files and all other data are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

References
1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1961 to 09/2019. J. Nat. Prod. 83, 770–803 (2020).
2. Bentley, S. D. et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417, 141–147 (2002).
3. Ikeda, H. et al. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat. Biotechnol. 21, 526–531 (2003).
4. Dejong, C. A. et al. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat. Chem. Biol. 12, 1007–1014 (2016).
5. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat. Prod. Rep. 26, 1362–1384 (2009).
6. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol. 13, 509–523 (2015).
7. Zhang, M. M., Wang, Y., Ang, E. L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. Nat. Prod. Rep. 33, 963–987 (2016).
8. Wang, B., Guo, F., Dong, S.-H. & Zhao, H. Activation of silent biosynthetic gene clusters using transcription factor decoys. Nat. Chem. Biol. 15, 111–114 (2019).
9. Zhang, M. M. et al. CRISPR–Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. Nat. Chem. Biol. 13, 607–609 (2017).
10. Yamanaka, K. et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl Acad. Sci. USA 111, 1957–1962 (2014).
11. Cleverger, K. D. et al. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat. Chem. Biol. 13, 895–901 (2017).
12. Zhang, J. J., Tang, X. & Moore, B. S. Genetic platforms for heterologous expression of microbial natural products. Nat. Prod. Rep. 36, 1313–1332 (2019).
13. Deng, Q., Zhou, L., Luo, M., Deng, Z. & Zhao, C. Heterologous expression of avermectin biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers. Synth. Syst. Biotechnol. 2, 59–64 (2017).
14. Wang, B., Tang, S., Li, W. & Lu, Y. in Nature Communications. https://doi.org/10.1038/s41467-021-21275-4 (2021).
21. Barka, E. A. et al. Taxonomy, physiology, and natural products of Actinobacteria. Microbiol. Mol. Biol. Rev. 80, 1–43 (2016).

22. Dugaczyk, A., Boyer, H. W. & Goodman, H. M. Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. J. Mol. Biol. 96, 171–184 (1975).

23. Łobocka, M. B. et al. Genome of bacteriophage P1. J. Bacteriol. 186, 7032–7068 (2004).

24. Grindley, N. D., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. Annu. Rev. Biochem. 75, 567–605 (2006).

25. Guo, Y.-Y. et al. A strategy for enhanced circular DNA construction efficiency based on DNA cyclicization after microbial transformation. Microb. Cell Fact. 14, 18 (2015).

26. Sauer, B. & Henderson, N. The cyclization of linear DNA in Escherichia coli by site-specific recombination. Gene 70, 331–341 (1988).

27. Karu, A., Sakaki, Y., Echols, H. & Linn, S. The γ protein specified by bacteriophage λ. Structure and inhibitory activity for the recBC enzyme of Escherichia coli. J. Biol. Chem. 250, 7377–7387 (1975).

28. Hashimoto-Gotoh, T., Franklin, F., Nordheim, A. & Timmis, K. Specific-purpose plasmid cloning vectors l. Low copy number, temperature-sensitive, mobilization-defective pSC101-derived containing vectors. Gene 16, 227–235 (1981).

29. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015).

30. Leti, C. et al. The CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) method for efficient editing of large DNA constructs in vitro. Nucleic Acids Res. 45, e74 (2017).

31. Li, S.-Y., Zhao, G.-P. & Wang, J. C-Brick: a new standard for assembly of biological parts using Cpf1. ACS Synth. Biol. 5, 1383–1388 (2016).

32. Fuchs, R. T. et al. Cas12a trans-cleavage can be modulated in vitro and is active on ssDNA, dsDNA, and RNA. Preprint at bioRxiv https://doi.org/10.1101/600890 (2019).

33. Wang, R., Guo, F., Huang, C. & Zhao, H. Unraveling the iterative type I polyketide synthases hidden in Streptomyces. Proc. Natl Acad. Sci. USA 117, 8449–8454 (2020).

34. Olano, C. et al. Activation and identification of five clusters for secondary metabolites in Streptomyces albus 11074. Microb. Biotechnol. 7, 242–256 (2014).

35. Laureti, L. et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces anbofaciens. Proc. Natl Acad. Sci. USA 108, 6258–6263 (2011).

36. Bunet, R. et al. Characterization and manipulation of the pathway-specific late regulator AlpW reveals Streptomyces anbofaciens as a new producer of kinamycins. J. Bacteriol. 193, 1142–1153 (2011).

37. Guillón, S. et al. Isolation, characterization, and heterologous expression of the biosynthesis gene cluster for the antitumor anthracycline stefimycin. Appl. Environ. Microbiol. 72, 4172–4183 (2006).

38. Gomez-Escribanro, J. P. & Bibb, M. J. Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. J. Ind. Microbiol. Biotechnol. 41, 425–431 (2014).

39. Wateve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus Streptomyces? Arch. Microbiol. 176, 386–390 (2001).

40. Sheng, Y., Mancino, V. & Birren, B. Transformation of Escherichia coli with large DNA molecules by electroporation. Nucleic Acids Res. 23, 1990–1996 (1995).

41. Novakova, J., Izsakova, A., Grivalsky, T., Ottmann, C. & Farkashovo, M. Improved method for high-efficiency electrotransformation of Escherichia coli with the large BAC plasmids. Folia Microbiol. 59, 53–61 (2014).

42. Blin, K. et al. antiSMASH 5.0: updates to the secondary metabolism genome mining pipeline. Nucleic Acids Res. 47, W81–W87 (2019).

43. Kieser, T., Ribb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical Streptomyces Genetics. (John Innes Foundation, 2000).

44. Vanga, D. G. & Kaliappan, K. P. Total synthesis of rubiginone A2, C2, and fujiamycin A. RSC Adv. 4, 12716–12722 (2014).

45. Yang, C., Huang, C., Zhang, W., Zhu, Y. & Zhang, C. Heterologous expression of flavostatin gene cluster leads to a bioactive heterodimer. Org. Lett. 17, 5324–5327 (2015).

46. Liu, T., Fischer, C., Beninga, C. & Rohr, J. Oxidative rearrangement processes in the biosynthesis of gilvocarcin V. J. Am. Chem. Soc. 126, 12262–12263 (2004).

47. Abdelfattah, M. S. Mansoquinone: isolation and structure elucidation of new antibacterial aromatic polyketides from terrestrial Streptomyces sp. Egs. Nat. Prod. Res. 23, 212–218 (2009).

48. Igarashi, M. et al. 4′-Deacetyl(-)-griseusins A and B, new naphthoquinone antibiotics from an actinomycete. J. Antibiot. 48, 1502–1505 (1995).

49. Tietz, J. s. al. A new genome-mining tool refines the lasso peptide biosynthetic landscape. Nat. Chem. Biol. 13, 470–478 (2017).

Acknowledgements
We thank William Metcalf for helpful discussions and sharing plasmids and microbial strains. We are grateful to Furuong Sun (School of Chemical Sciences Mass Spectrometry Laboratory) for HRESIMS data analysis, Xuqiong Gao (the Carl R. Woese Institute for Genomic Biology) for NMR measurement, and Toby J. Woods (School of Chemical Sciences Materials Chemistry Laboratory) for X-ray diffraction analysis. We thank Wenyun Zhang (South China Sea Institute of Oceanology, Chinese Academy of Sciences) and Haiyan Tian (Jinan University) for insightful discussion of structure elucidation. Some of this data was collected in the IGB Core on a 600 MHz NMR funded by NIH grant number S10-RR028833. This work was supported by grant GM077596 and AI144967 from the National Institutes of Health.

Author contributions
B.E. and H.Z. conceived and designed the study. B.E., C.H. and H.Z. wrote the manuscript. B.E. performed the method development experiments. B.E., C.H., F.G., G.J., B.W. and S.K.T. performed the uncharacterized BGCS cloning experiments. C.H. F.G. and G.J. performed the heterologous expression and HPLC analysis experiments. C.H. performed structure elucidation of compounds and pigmentamycins antimicrobial activity assays. T.M. helped with heterologous expression experiments.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21275-4.

Correspondence and requests for materials should be addressed to H.Z.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021