



Published in final edited form as:

*J Nat Prod.* 2017 April 28; 80(4): 1200–1204. doi:10.1021/acs.jnatprod.6b01121.

## A PCR-independent Method of Transformation Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete

Charles B. Larson<sup>†,‡</sup>, Max Crüsemann<sup>†,⊥</sup>, and Bradley S. Moore<sup>†,§,\*</sup>

<sup>†</sup>Center for Marine Biotechnology & Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA

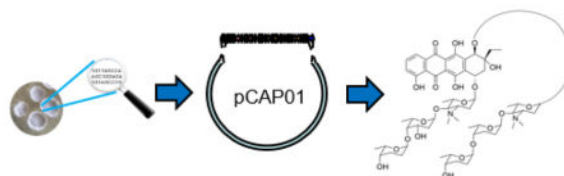
<sup>‡</sup>Biomedical Sciences Graduate Program, University of California at San Diego, La Jolla, CA, USA

<sup>§</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA

### Abstract

The transformation-associated recombination cloning methodology facilitates the genomic capture and heterologous expression of natural product biosynthetic gene clusters (BGCs). We have streamlined this procedure by introduction of synthetic DNA gene blocks for the efficient capture of BGCs. We show the successful capture and expression of the aromatic polyketide antitumor agent cosmomycin from streptomycete bacteria and the discovery of new cosmomycin analogs by mass spectral molecular networking.

### Graphical Abstract



With the advent of rapid and inexpensive genome sequencing methods, the process of natural product discovery is advancing to incorporate biosynthetic logic.<sup>1</sup> Traditional discovery efforts have focused on the collection, culturing, and subsequent analysis of organismal extracts to detect pharmacological activities harbored in bacteria, plants, or animals in a “grind and find” approach.<sup>2</sup> However, this manner of investigation can often overlook compounds produced in minuscule amounts or in unique environmental conditions

\*Corresponding author: bsmoore@ucsd.edu.

<sup>⊥</sup>present address: Institute of Pharmaceutical Biology, University of Bonn, 53115 Bonn, Germany

#### Supplemental Information

The Supporting Information is available free of charge on the ACS Publications website at DOI:

HRMS and MS/MS data for compounds **1–3**, **A** and **B**; <sup>1</sup>H and 2D NMR of **1** and **A**; analysis of plasmid pCAP01-COS, molecular network of organisms containing the cosmomycin BGC.

that are difficult to replicate in the laboratory. Some of the most well-studied bacteria have been shown to produce only a subset of natural product chemicals even though their genomes frequently encode the biosynthetic machinery to synthesize many additional compounds of often unknown structure and function.<sup>3</sup> Recent advances in synthetic biology have allowed investigators to avoid this traditional discovery approach of culturing and extracting wild type strains by directly manipulating the genetic sequence that encodes components of biosynthetic pathways.<sup>4</sup> Using this genetic information, entire pathways can be targeted within the native producer or removed in their entirety and expressed in a heterologous host.<sup>5</sup> Utilizing these techniques, many pathways of interest have been elucidated that are intractable to traditional natural product discovery approaches.<sup>6</sup>

In order to clone large biosynthetic gene clusters (BGCs) that are regularly 40–60 kilobases in size, transformation-associated recombination (TAR) has emerged as a powerful method to selectively capture and incorporate genes encoding complete biosynthetic pathways into multi-host plasmids. This method was originally described by Kouprina and Larionov for cloning selective genomic loci from human DNA,<sup>7</sup> and later adapted by Brady to capture BGCs from environmental DNA (eDNA) libraries, including the BGC producing the MRSA-active antibiotic tetracycline A.<sup>8</sup> Our laboratory developed TAR vectors and methods to capture and express an assortment of BGCs directly from gDNA from taxonomically diverse microbes.<sup>6,9,10</sup> In addition to direct cloning from libraries or genomes, TAR has been applied as a DNA assembly technique<sup>11</sup> for the creation of a fully synthetic genome and has facilitated the refactoring of BGCs to support natural product production.<sup>12</sup>

Despite its successes, challenges are often experienced during the experimentally involved TAR cloning process. Vector assembly in particular is a tedious, multistep process that relies heavily on PCR. Complications can arise during this assembly process, especially with high GC or repetitive sequences that are common among actinomycete bacteria known to harbor large numbers of BGCs. Herein we report a modified approach employing a fully synthetic “capture arm,” thus eliminating the need for traditional PCR amplification during the vector assembly process. This new methodology allows for a significant decrease in the duration of the cloning process and opens the door for higher-throughput applications. To demonstrate the feasibility of this optimized procedure, we set out to capture and heterologously express the cosmomycin BGC from a marine streptomycete bacterium.

We previously reported the utilization of TAR-based direct cloning and expression of BGCs with the taromycin series of cyclic lipopeptide antibiotics.<sup>6</sup> Briefly, our original protocol with the *Saccharomyces cerevisiae*/*Escherichia coli* shuttle-actinobacterial chromosome integrative capture vector pCAP01 required a pair of one kilobase (kb) regions, homologous to the two flanking regions of the BGC of interest, to be amplified from genomic DNA (Figure 1A). These fragments were then assembled, inserted into the pCAP01 capture vector, and linearized prior to use in the capture/transformation steps. The time consuming process of the multiple PCR reactions and subsequent ligations necessary to obtain the assembled capture vector demanded development of a more efficient protocol.

We sought to streamline the original protocol by replacing all vector assembly steps by a simple digestion/ligation reaction. To do so, we designed a synthetic DNA fragment

containing shortened capture arms and the appropriate restriction sites to facilitate both its insertion into pCAP01 and plasmid linearization (Figure 1B). We selected an orphan 54 kb BGC encoding a type II PKS from *Streptomyces* sp. CNT-302 (Figure 2) as a target pathway. The type II PKS shares 94% and 90% identity with the cosmomycin<sup>13,14</sup> KS and CLF subunits, respectively. In addition, the pathway contains 32 genes encoding putative tailoring enzymes, and three genes encoding glycosyltransferases indicative of a highly glycosylated end product. Proposed functions and homologies for each gene product in the pathway are shown in Supplementary Information Table S2. Initially investigated for its extensive glycosylation biochemistry involving three separate glycosyltransferases, we analyzed CNT-302 extracts by glycogenomic mass spectral analysis<sup>15</sup> to no avail. Although this BGC appeared non-functional in the native host under the growth conditions tested, we explored its heterologous expression using the modified TAR capturing method.

We employed capture arms of 360 base pairs (bps), about a third of the length previously used,<sup>6</sup> although capture arms as small as 60 bps have been shown to be effective.<sup>16</sup> The size of the capture arms was determined by the maximum size of commercial oligonucleotide synthesis at the time, and included three restriction sites flanking the two homology capture arms for ligation and linearization. After insertion of the 750 bp synthetic double stranded DNA into pCAP01 to form plasmid pCAP01-COM, we transformed yeast spheroplasts with digested *S. sp* CNT-302 gDNA and linear pCAP01-COM on selective media. Yeast clones were screened by PCR and subsequent analysis via restriction digestion, confirmed the successful capture of the type II PKS BGC in plasmid pCAP01-COS in three of 200 transformants screened (Supporting Information Figures S1 and S2). After the isolation of pCAP01-COS, we confirmed the vector by restriction digest and transformed it into an *E. coli* ET12567 shuttle, which integrated the vector into the genome of *S. coelicolor* M512<sup>17</sup> by conjugation. To confirm the stable integration of pCAP01-COS, several regions of the plasmid were PCR amplified and Sanger sequenced (Figures S3 and S4). In order to determine if any glycosylated natural products were produced by this type II PKS BGC, R5 media was inoculated with heterologous host into which pCAP01-COS was integrated (M512/pCAP01-COS). After 10 days, the culture media was extracted with EtOAc and subjected to mass spectral (MS) molecular networking.<sup>18</sup>

Spectral networks, originally developed for proteomics, have recently been adopted as a tool for general MS data analysis.<sup>19</sup> The foundation of a molecular network is the comparison of MS/MS spectra for molecular ions captured by mass spectrometry. Molecular families are then created based on the similarity of the analyzed spectra. Analysis of the network, which included extracts of M512 as a control, M512/pCAP01-COS, and CNT-302 immediately revealed a five molecule family in *S. coelicolor* M512/pCAP01-COS absent in the 'native' producer, including the previously described antitumor molecules cosmomycin C (**1**) (HR-ESI-MS  $m/z$  1173.5964, calcd.  $m/z$  for  $C_{60}H_{88}N_2O_{21}$  ( $[M+H]^+$ ) 1173.5952), cosmomycin D (**2**) (HR-ESI-MS  $m/z$  1189.5889, calcd.  $m/z$  for  $C_{60}H_{88}N_2O_{22}$  ( $[M+H]^+$ ), 1189.5901), and deoxy-cosmomycin C (**3**) (HR-ESI-MS  $m/z$  1157.6007, calcd.  $m/z$  for  $C_{60}H_{88}N_2O_{20}$  ( $[M+H]^+$ ) 1157.6003) (Figure 3).<sup>20</sup> These three molecules were identified by mass fragmentation patterns and isotope distributions in comparison to those previously reported (Figures S5–8).<sup>21</sup> Additionally, two new cosmomycin analogs were identified in *S.*

*coelicolor* M512/pCAP01-COS, which appear to be missing methyl and/or hydroxyl groups on the inner amino sugar rhodosamine or aglycone when compared to cosmomycin C (**1**). These include compound **A** (HR-ESI-MS  $m/z$  1159.5793 (calcd.  $m/z$  for  $C_{59}H_{86}N_2O_{21}$  ( $[M+H]^+$ ) 1159.5796) and compound **B** (HR-ESI-MS  $m/z$  1143.5846, calcd.  $m/z$  for  $C_{59}H_{86}N_2O_{20}$  ( $[M+H]^+$ ) 1143.5847) which have masses consistent with desmethyl-cosmomycin C and desmethyl, deoxy-cosmomycin C respectively (Figures S9–11).

The two major cosmomycin species present in culture, **1** and **A**, were further purified and analyzed by  $^1H$  and HSQC NMR (Figures S12–17). Because of the small amount of **B** produced in culture, further structural elucidation was not pursued. The chemical shifts observed for both **1** and **A** closely matched literature values for cosmomycin analogs A447 C and D, obelmycin C, and cosmomycin D (Table S1). While **1** could be unequivocally established as cosmomycin C, spectroscopic analysis of the closely related **A** suggested that it was a new analogue. The C-14 triplet methyl peak at 1.09 ppm in **1** is conspicuously absent in the 1D NMR and HSQC spectra of compound **A** (Figures S14–16). Additionally, we observed a new peak in the HSQC spectrum of **A** at 1.47 ppm correlating to a carbon with a chemical shift of 26.8 ppm, matching expected values of the terminal methyl at C-13. Thus, these data suggest that the canonical propionate starter unit of the cosmomycin type II PKS is replaced with acetate in the heterologous system, thereby shortening the carbon chain by one carbon in **A**.

After the identification of the cosmomycin series in *S. coelicolor* M512/pCAP01-COS, genome analysis revealed a homologous BGC in *S. sp.* CNS-615 (Figure 2). Following analysis of culture extracts from *S. sp.* CNS-615 and subsequent molecular networking against M512/pCAP01-COS and *S. sp.* CNT-302, *S. sp.* CNS-615 was similarly identified as a producer of **1–3** but not of any of the acetate-primed desmethyl species **A** or **B** (Supplemental Figure S18).

Identification of the cosmomycin series in culture extracts of *S. coelicolor* M512/pCAP01-COS demonstrates that the type II PKS BGC captured via TAR is sufficient for the production of the cosmomycin natural products. These results confirm the findings by Padilla and coworkers, who demonstrated the necessity of two glycosyltransferases in the biosynthesis of cosmomycin D from *Streptomyces olindensis* which share 95% and 94% identity with the glycosyltransferases captured within the pathway from CNT-302.<sup>13</sup>

During the course of this study, the genome sequence of *S. olindensis* was published,<sup>14</sup> which included a type II PKS BGC with an average amino acid identity of 90% to the *cos* cluster captured in this study (Figure 1, Table S2). Knowledge of the final product has allowed us to confirm complete absence of production by the native “producer” *Streptomyces sp.* CNT-302 under a range of media conditions (Figure S19), yet expression was achieved in *S. coelicolor* M512/pCAP01-COS without the need for genetic manipulation of the pathway. This observation illustrates the potential for using TAR capturing and heterologous expression to unleash otherwise silent gene clusters, while also demonstrating the synergy of combining heterologous expression and molecular networking to connect genes to molecules and reveal new members of molecular families.

The TAR cloning protocol has been successful at capturing BGCs directly from genomic<sup>6</sup> and metagenomic<sup>8</sup> DNA samples for the efficient characterization of natural products and the processes involved in their assembly. However, the staggering number of available bacterial genomes and orphan biosynthetic clusters therein suggests the need for a high-throughput methodology to investigate novel biosynthetic chemistry. The lengthy TAR protocol requiring individual optimization and troubleshooting is unsuitable for high throughput technologies. Here, we present a streamlined version of the protocol, avoiding PCR assembly of each custom capture vector, instead utilizing synthetic DNA inserted with a simple ligation. Although the synthetic capture arms used in this study cost approximately \$1/bp, at the time of publication the prices has fallen to \$0.15/bp, and the time savings compared to PCR-based assembly of capture vectors is estimated to be one week for a single investigator. We have demonstrated that this new method is as effective as the original approach, showing similar efficiency of transformation in addition to effectively capturing and expressing the cosmomycin biosynthetic pathway. Since this work was performed, the pCAP system has undergone further improvements such as the shortening of the synthetic insert length to 144 bp by including 60 bp capture arms, the addition of regions homologous to the vector backbone allowing for one-step assembly, and the insertion of a 5-fluoroorotic acid-mediated negative selection mechanism which dramatically increases efficiency from ~1–2% to over 50%.<sup>22</sup> These recent advancements allow us to entertain the idea of a high-throughput TAR protocol, which could potentially automate a large number of capture experiments through rapid, PCR independent assembly of capture vectors and subsequent parallel transformations with prepared gDNA. Furthermore, whole pathway capture allows for the simple manipulation of biosynthetic machinery. Through manipulating biosynthetic genes, novel analogues and compounds can be engineered and produced from known clusters.<sup>23</sup> In the case of the cosmomycin series, new glycosylation patterns are envisioned through glycorandomization engineering processes.<sup>24</sup> Considering the promising antitumor activity of the known cosmomycin series and the importance of the carbohydrate moiety for bioactivity,<sup>25</sup> this approach could yield powerful analogs with even greater efficacy. With this engineering approach to modifying biosynthesis, combined with the powerful TAR pathway cloning, there is a large area of chemical space waiting to be explored.

## Experimental Methods

### General Experimental Procedures

NMR spectra were collected at 298K using a Bruker Avance III 600 MHz spectrometer fitted with a 5mm TCI cryoprobe. The HSQC spectra were acquired using 12.5% non-uniform sampling and reconstructed using SMILE and NMRPipe. Spectra were recorded in CDCl<sub>3</sub> at 25 °C, and chemical shifts are given on the  $\delta$  scale referenced to residual chloroform ( $\delta$ H 7.26) (Supplemental Figures S12–17, Table S1). For mass spectrometry analysis, the Agilent 1290 liquid chromatography system coupled to an Agilent 6530 Q-TOF was used. Liquid chromatography fractionation was carried out on both the Agilent Prepstar and the Agilent 1260 Infinity LC systems. Electroporations were carried out on an Eppendorf Electroporator 2510.

## TAR Cloning Procedure

The *cos* BGC was identified in the genome of *Streptomyces* sp. CNT-302 (GenBank accession ARIM00000000.1) using the antiSMASH software suite,<sup>26</sup> and homologous capture arms were designed with 360 base pairs of homology to the boundaries of the cluster. The capture arms were synthesized as a DNA geneblock (Integrated DNA Technologies) with a BamHI restriction site between them. Capture arms were inserted into the SpeI and XhoI restriction sites in the pCAP01 (AddGene accession number 59981) backbone using digestion and ligation to form plasmid pCAP01-COM. Successful assembly was confirmed by restriction analysis.

1 µg of pCAP01-COM was linearized using the BamHI restriction site included in the synthetic insert sequence, and the linear vector was gel purified using a Qiagen Gel Extraction Kit following the standard manufacturer protocol. From this point onward, a slightly modified version of the original TAR protocol<sup>10</sup> was followed and can be found deposited with the pCAP01 sequence on the Addgene database. Briefly, yeast spheroplasts (200 µL) were transformed with linear vector (200 ng) and gDNA (1 µg) from CNT-302 previously digested with ClaI. Transformed spheroplasts were plated on selective agar and incubated for 4 days at 30 °C. Colonies were PCR screened for regions of the target biosynthetic cluster, and plasmid from positive hits (1.5% of colonies analyzed) were isolated using the Zymoprep Yeast Plasmid Miniprep I kit (Zymo Research) and transformed into *E. coli* Top10 (Life Technologies) by electroporation (Figures S1 and S2). The correct capture of the biosynthetic cluster into pCAP01-COM was confirmed by restriction analysis, and the resulting plasmid was named pCAP01-COS.

The plasmid pCAP01-COS was transformed into *S. coelicolor* M512 using tri-parental conjugation<sup>27</sup> with *E. coli* ET12567 cells containing either pUB307 conjugation plasmid or the pCAP01-COS vector containing the captured cluster. Three sequential rounds of selection were performed on selective media (MS agar + nalidixic acid (100 µg/mL) + kanamycin (50 µg/mL) + chloramphenicol (35 µg/mL)) to ensure the plasmid was integrated and maintained. Integration was confirmed by PCR amplification and Sanger sequencing (Figures S3 and S4).

## Heterologous Expression of the Pathway and Identification of the Products

Spores of *S. coelicolor* M512/pCAP01-COS were used to streak plates of R5 production media. After one week, plugs of the agar plate (1 cm in diameter) were removed, washed with 5 mL H<sub>2</sub>O, and extracted with 5 mL EtOAc. During the course of this study, a second marine streptomycete (*Streptomyces* sp. CNS-615, GenBank accession AQPE00000000.1) was found to contain the same target type II PKS BGC. In addition to M512 and M512/pCAP01-COS, the two marine *Streptomyces* sp. strains containing the 54 kb pathway (CNT-302 and CNS-615) were also grown and extracted in the same manner. Production of secondary metabolites was confirmed by HPLC-HR-ESI-MSMS analysis, carried out on an Agilent 1290 liquid chromatography system coupled to an Agilent 6530 Q-TOF (200 2000m/z, 20 keV) (Figures S5–11). Prepared samples were dissolved in MeOH and injected on a C18 column (Phenomenex Luna 5µm C18(2) 100A 100x4.6 mm) with an initial mobile phase of 10% MeCN at a flow rate of 0.7 mL/min. Over the course of 30 min, the MeCN



concentration was linearly increased to 100%. The collected data were subjected to the GNPS molecular networking workflow<sup>19</sup> and analyzed as described previously (Cosine Score = 0.7).<sup>18</sup> Samples were then grouped and visualized using Cytoscape (Figures S18 and S19).

### Purification of Compounds 1 and A

Liquid R5 media was inoculated with *S. coelicolor* M512/pCAP01-COS and grown for 10 days. One liter of culture supernatant was extracted three times with 500 mL EtOAc and subsequently dried. The extract was then loaded on a preparative HPLC column (Agilent Pursuit XR5 5 $\mu$ m C18, 100 x 21.2 mm) and subjected to a standard reversed-phase gradient from 5% to 100% MeCN in 100 mM ammonium acetate buffer over 60 min at a flow rate of 15 mL/min. All subsequent HPLC purifications utilized the same 100 mM ammonium acetate buffer. Individual fractions collected from the preparative column containing cosmomycin family members (eluted at ~40% MeCN) were then subjected to another round of HPLC purification (Phenomenex Luna 5 $\mu$ m C8 100A, 250 x 10 mm). Cosmomycin family members were eluted on an isocratic 40% MeCN concentration and individual peaks were collected, dried, and subjected to a final round of C18 purification (Phenomenex Kinetex 5 $\mu$ m XB-C18 100A, 250 x 4.6 mm) using the same conditions as the previous C8 column purification. The major individual peaks were collected and dried on a lyophilizer for NMR analysis. Two liters of extracted culture supernatant yielded 2.5 mg of **1** and 1.5 mg of **A**.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

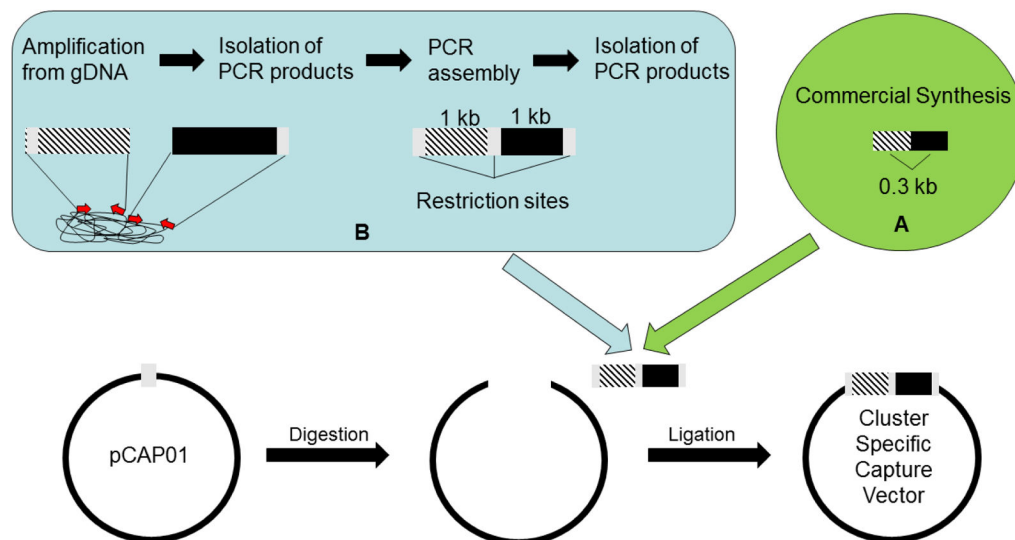
We thank P. Jensen and W. Fenical (UC San Diego) for providing *Streptomyces* strains CNT-302 and CNS-615, M. Bibb (John Innes Centre, UK) for providing *S. coelicolor* M512, Y. Su at the UCSD Molecular Mass Spectrometry Facility for mass spectrometry assistance, B. Duggan for NMR spectroscopy assistance, and E. O'Neill, X. Tang, S. McKinnie, Y. Kudo, and J. Li (UC San Diego) for valuable discussions. This work was supported by U.S. National Institutes of Health grant R01-GM085770 and a DFG postdoctoral fellowship CR464-1 to M.C. The genome sequences were provided by the Jensen lab and obtained by the U.S. Department of Energy Joint Genome Institute with support by the Office of Science, U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### References

1. Harvey AL, Edrada-Ebel R, Quinn RJ. Nat Rev Drug Discov. 2015; 14:111–129. [PubMed: 25614221]
2. Cragg GM, Newman DJ. Biochim Biophys Acta. 2013; 1830:3670–3695. [PubMed: 23428572]
3. Nett M, Ikeda H, Moore BS. Nat Prod Rep. 2009; 26:1362–1384. [PubMed: 19844637]
4. Medema, MH., Breitling, R., Takano, E. Synthetic Biology in Streptomyces. Vol. 497. Bacteria Academic Press; 2011. p. 485-502.
5. Ongley SE, Bian X, Neilan BA, Muller R. Nat Prod Rep. 2013; 30:1121–38. [PubMed: 23832108]
6. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Proc Natl Acad Sci U S A. 2014; 111:1957–1962. [PubMed: 24449899]
7. Kouprina N, Larionov V. Nat Protoc. 2008; 3:371–7. [PubMed: 18323808]
8. Kallifidas D, Kang HS, Brady SF. J Am Chem Soc. 2012; 134:19552–19555. [PubMed: 23157252]

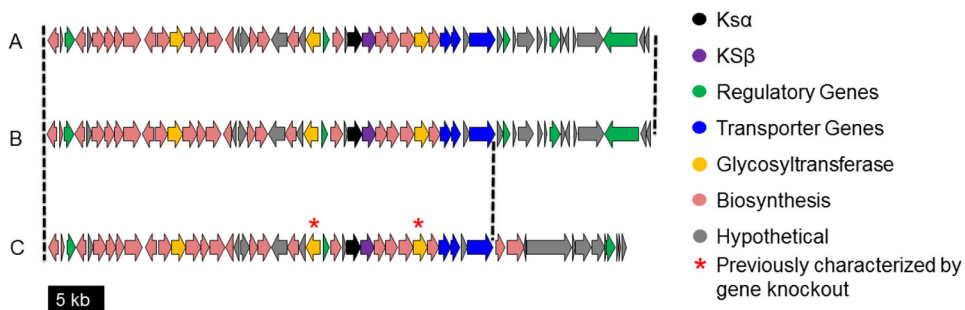
9. Ross AC, Gulland LE, Dorrestein PC, Moore BS. *ACS Synth Biol.* 2014; 4:414–420. [PubMed: 25140825]
10. Li Y, Li Z, Yamanaka K, Xu Y, Zhang W, Vlamakis H, Kolter R, Moore BS, Qian PY. *Sci Rep.* 2015; 5:9383. [PubMed: 25807046]
11. Chao R, Yuan Y, Zhao H. *FEMS Yeast Res.* 2014; 15:1–9.
12. Luo Y, Li BZ, Liu D, Zhang L, Chen Y, Jia B, Zeng BX, Zhao H, Yuan YJ. *Chem Soc Rev.* 2015; 44:5265–5290. [PubMed: 25960127]
13. Garrido LM, Lombo F, Baig I, Nur EAM, Furlan RL, Borda CC, Brana A, Mendez C, Salas JA, Rohr J, Padilla G. *Appl Microbiol Biotechnol.* 2006; 73:122–131. [PubMed: 16810496]
14. Rojas JD, Starcevic A, Baranasic D, Ferreira-Torres MA, Contreras CA, Garrido LM, Araujo WL, de Souza RF, Zucko J, Hranueli D, Long PF, Cullum J, Padilla G. *Genome Announc.* 2014; 2:1–2.
15. Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, Moore BS. *Proc Natl Acad Sci U S A.* 2013; 110:E4407–4416. [PubMed: 24191063]
16. Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, Stubbs L, Kouprina N, Larionov V. *Nucleic Acids Res.* 2001; 29:e32. [PubMed: 11239009]
17. Floriano B, Bibb M. *Mol Microbiol.* 1996; 21:385–96. [PubMed: 8858592]
18. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC. *J Nat Prod.* 2013; 76:1686–1699. [PubMed: 24025162]
19. Wang, et al. *Nature Biotechnology.* 2016; 34:828–837.
20. Hirayama K, Akashi S, Ando T, Horino I, Etoh Y, Morioka H, Shibai H, Murai A. *Biol Mass Spectrom.* 1987; 14:305–312.
21. Kelso C, Rojas JD, Furlan RL, Padilla G, Beck JL. *Eur J Mass Spectrom.* 2009; 15:73–81.
22. Tang X, Li J, Millan-Aguinaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. *ACS Chem Biol.* 2015; 10:2841–2849. [PubMed: 26458099]
23. Metsä-Ketelä M, Palmu K, Kunnari T, Ylihonko K, Mäntsälä P. *Antimicrob Agents Chemother.* 2003; 47:1291–1296. [PubMed: 12654660]
24. Langenhan JM, Griffith BR, Thorson JS. *J Nat Prod.* 2005; 68:1696–1711. [PubMed: 16309329]
25. Furlan RL, Watt SJ, Garrido LM, Amarante-Mendes GP, Nur-e-alam M, Rohr J, Braña A, Mendez C, Salas JA, Sheil MM, Beck JL, Padilla G. *J Antibiot(Tokyo).* 2004; 57:647–654. [PubMed: 15638325]
26. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Muller R, Wohlleben W, Breitling R, Takano E, Medema MH. *Nucleic Acids Res.* 2015; 43:W237–243. [PubMed: 25948579]
27. Kieser, T., Bibb, MJ., Buttner, MJ., Chater, KF., Hopwood, DA. *Practical Streptomyces Genetics.* Vol. 1. John Innes Foundation; Norwich: 2000. p. 250-251.





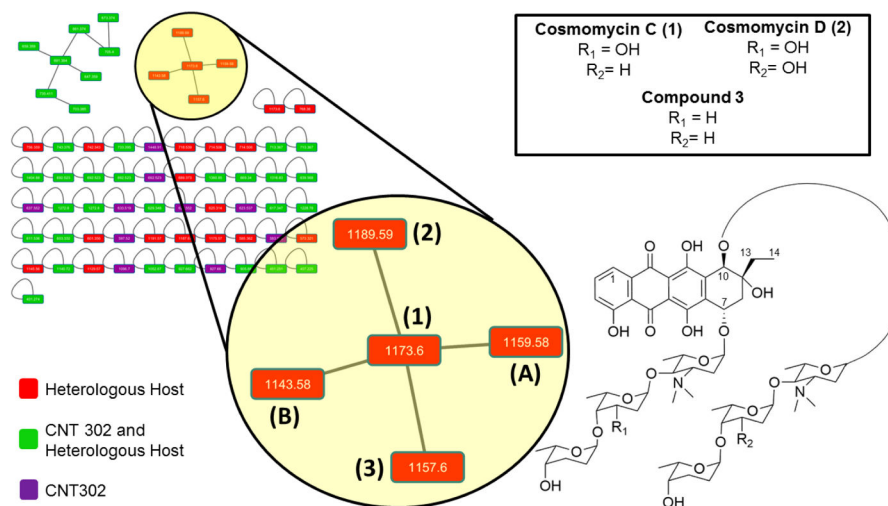
**Figure 1. Capture Vector Assembly**

(A) Cluster specific capture arms, one kilobase in size, are amplified from genomic DNA and subsequently assembled using PCR. (B) Cluster specific capture arms (360 bp) are synthesized commercially as a single 750 bp dsDNA gene block including restriction sites for insertion and linearization.



**Figure 2. Gene Map of Biosynthetic Gene Clusters Associated with the Biosynthesis of Cosmomycins C and D**

The cosmomycin BGCs from (A) *Streptomyces* sp. CNS-615, (B) *Streptomyces* sp. CNT-302, and (C) *Streptomyces olindensis* are >90% similar on the amino acid sequence level. A gene by gene comparison can be found in Table S2. Homology of the *S.* sp. CNT-302 and the *S. olindensis* BGCs is restricted to the first 40 kb of the captured sequence (denoted by dotted lines), suggesting the last 14 kb of genomic material captured in this study is not necessary for cosmomycin production. Genes marked with a '\*' have been interrogated by gene knockout experiments in previous studies.



**Figure 3. Cosmomycin Cluster Identified by Molecular Networking**

Nodes matching the masses for cosmomyacin C (**1**) and D (**2**) network with nodes matching values of cosmomyacin derivatives lacking methyl groups and/or oxygen. These include deoxy-cosmomycin C (**3**), 13-desmethyl-cosmomycin C (**A**), and desmethyl, deoxy-cosmomycin C (**B**).