

Strain Prioritization and Genome Mining for Eneidyne Natural Products

Xiaohui Yan,^a Huiming Ge,^a Tingting Huang,^a Hindra,^a Dong Yang,^a Qihui Teng,^a Ivana Crnovčić,^a Xiuling Li,^b Jeffrey D. Rudolf,^a Jeremy R. Lohman,^a Yannick Gansemans,^c Xiangcheng Zhu,^{d,e} Yong Huang,^d Li-Xing Zhao,^f Yi Jiang,^f Filip Van Nieuwerburgh,^c Christoph Rader,^{b,g} Yanwen Duan,^{d,e} Ben Shen^{a,g,h}

Department of Chemistry, the Scripps Research Institute, Jupiter, Florida, USA^a; Department of Cancer Biology, the Scripps Research Institute, Jupiter, Florida, USA^b; Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium^c; Xiangya International Academy of Translational Medicine, Central South University, Changsha, Hunan, China^d; Hunan Engineering Research Center of Combinatorial Biosynthesis and Natural Product Drug Discovery, Changsha, Hunan, China^e; Yunnan Institute of Microbiology, Yunnan University, Kunming, Yunnan, China^f; Department of Molecular Therapeutics, the Scripps Research Institute, Jupiter, Florida, USA^g; Natural Products Library Initiative at The Scripps Research Institute, the Scripps Research Institute, Jupiter, Florida, USA^h

X.Y., H.G., T.H., and H. contributed equally to this article.

ABSTRACT The enediynes family of natural products has had a profound impact on modern chemistry, biology, and medicine, and yet only 11 enediynes have been structurally characterized to date. Here we report a genome survey of 3,400 actinomycetes, identifying 81 strains that harbor genes encoding the enediynes polyketide synthase cassettes that could be grouped into 28 distinct clades based on phylogenetic analysis. Genome sequencing of 31 representative strains confirmed that each clade harbors a distinct enediynes biosynthetic gene cluster. A genome neighborhood network allows prediction of new structural features and biosynthetic insights that could be exploited for enediynes discovery. We confirmed one clade as new C-1027 producers, with a significantly higher C-1027 titer than the original producer, and discovered a new family of enediynes natural products, the tiancimycins (TNMs), that exhibit potent cytotoxicity against a broad spectrum of cancer cell lines. Our results demonstrate the feasibility of rapid discovery of new enediynes from a large strain collection.

IMPORTANCE Recent advances in microbial genomics clearly revealed that the biosynthetic potential of soil actinomycetes to produce enediynes is underappreciated. A great challenge is to develop innovative methods to discover new enediynes and produce them in sufficient quantities for chemical, biological, and clinical investigations. This work demonstrated the feasibility of rapid discovery of new enediynes from a large strain collection. The new C-1027 producers, with a significantly higher C-1027 titer than the original producer, will impact the practical supply of this important drug lead. The TNMs, with their extremely potent cytotoxicity against various cancer cells and their rapid and complete cancer cell killing characteristics, in comparison with the payloads used in FDA-approved antibody-drug conjugates (ADCs), are poised to be exploited as payload candidates for the next generation of anticancer ADCs. Follow-up studies on the other identified hits promise the discovery of new enediynes, radically expanding the chemical space for the enediynes family.

Received 15 November 2016 Accepted 29 November 2016 Published 20 December 2016

Citation Yan X, Ge H, Huang T, Hindra, Yang D, Teng Q, Crnovčić I, Li X, Rudolf JD, Lohman JR, Gansemans Y, Zhu X, Huang Y, Zhao L-X, Jiang Y, Van Nieuwerburgh F, Rader C, Duan Y, Shen B. 2016. Strain prioritization and genome mining for enediynes natural products. *mBio* 7(6):e02104-16. doi:10.1128/mBio.02104-16.

Editor Julian E. Davies, University of British Columbia

Copyright © 2016 Yan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Yanwen Duan, ywduan66@yahoo.com, or Ben Shen, shenb@scripps.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. External solicited reviewers: Bradley Moore, University of California San Diego; Yi Tang, UCLA; David Newman, NCI.

Natural products offer unmatched chemical and structural diversity compared to any other small-molecule families. They continue to inspire novel chemistry, enzymology, and biology investigations and remain the best sources of drugs and drug leads (1). The enediynes represent one of the most fascinating families of natural products for their unprecedented molecular architecture and extraordinary biological activities. Since the neocarzinostatin (NCS) chromophore structure was first unveiled in 1985 (2), the enediynes family has grown steadily but remains very small, with only 11 structurally characterized members and 4 additional members isolated in their cycloaromatized form known to date. Classified into two subcategories according to the size of the enediynes core structures (3–7), members of the 9-membered

enediynes subcategory include NCS, C-1027, kedarcidin (KED), maduropeptin (MDP), N1999A2, the sporolides (SPO), the cyanosporasides (CYA and CYN), and the fijiolides, with the latter four isolated in cycloaromatized form (see Fig. S1A in the supplemental material). Members of the 10-membered enediynes subcategory include the calicheamicins (CAL), the esperamicins (ESP), dynemicin (DYN), namenamicin, shishijimicin, and uncialamyacin (UCM) (Fig. S1B).

The enediynes have had a profound impact on modern chemistry, biology, and medicine (3, 4, 7). All enediynes contain a unit consisting of two acetylenic groups conjugated to a double bond or an incipient double bond within the 9- or 10-membered carbocycle (Fig. S1A and B). As a consequence of this structural fea-

ture, the enediynes share a mode of action—electronic rearrangement of the enediyne carbocycle produces a transient benzenoid diradical. When positioned within the minor groove of DNA, the diradical abstracts hydrogen atoms from the deoxyribose backbone of duplex DNA; the DNA-centered radicals can then cause interstrand cross-links (ICLs) or react with molecular oxygen, leading ultimately to DNA double-strand breaks (DSBs), or both (3, 4, 7, 8). With their exquisite mode of action and their extraordinary cytotoxicity, the enediynes have been successfully translated into clinical drugs. It is remarkable that, among the 11 enediynes known to date, 2 [NCS as poly(styrene-comaleic acid)-conjugated NCS (SMANCS) and CAL as gemtuzumab ozogamicin (Mylotarg)] have been developed into marketed drugs, 1 (C-1027) is in clinical trials, and another (UCM) is in preclinical studies, representing an astonishing and remarkable ~35% success rate with the enediyne class of natural products (3–7). A great challenge is that of developing innovative methods to discover new enediynes and producing them in sufficient quantities for chemical, biological, and clinical investigations.

Here we report strain prioritization and genome mining for enediyne natural products from the actinomycetes strain collection at the Scripps Research Institute (TSRI). By surveying 3,400 strains, we identified 81 potential producers, and genome sequencing of 31 representatives revealed at least 28 distinct enediyne biosynthetic gene clusters. We constructed an enediyne genome neighborhood network (GNN) to facilitate gene cluster annotation and to predict new structural features, thereby further streamlining the discovery of new enediyne natural products. To demonstrate the feasibility of our approach in rapidly discovering new enediynes from a large strain collection, we characterized a new C-1027 producer with a significantly higher C-1027 titer than the original *Streptomyces globisporus* producer (9, 10) and discovered the tiancimycins (TNMs), new enediyne natural products that exhibit potent cytotoxicity against a broad spectrum of cancer cell lines and kill selected cancer cells more rapidly and completely than the payloads used in FDA-approved antibody (Ab)-drug conjugates (ADCs). Production of TNMs in sufficient quantities by microbial fermentation and manipulation of TNM biosynthesis for engineering new analogues were also demonstrated.

RESULTS AND DISCUSSION

Genome survey of 3,400 actinomycetes for the enediyne polyketide synthase gene cassette identifying 81 potential enediyne producers that could be grouped into 28 distinct clades. Since the cloning of the first 9-membered (C-1027) (11) and 10-membered (CAL) (12) enediyne biosynthetic gene clusters in 2002, three additional 9-membered (NCS, MDP, and KED) (13–15) and two additional 10-membered (DYN and ESP [partial]) (16, 17) enediyne gene clusters, as well as the three clusters encoding the biosynthesis of SPO (18), CYA (19), and CYN (19), have been characterized. Comparative analysis of the 10 gene clusters revealed a set of 5 genes common to all enediynes, i.e., the enediyne polyketide synthase (PKS) gene cassette consisting of *E3*, *E4*, *E5*, *E*, and *E10* (*E3/E4/E5/E/E10*) (Fig. S2A) (17, 20–22); no apparent conservation was observed beyond the enediyne PKS gene cassettes, accounting for the structural diversity characteristic for the periphery moieties of the enediynes (Fig. S1A and B). The remarkable sequence homology prompted us to select genes within the enediyne PKS cassettes as probes to survey genomes for the presence of enediyne biosynthetic machinery. This was vali-

dated recently by a virtual survey of all bacterial genomes available in public databases, revealing the rich potential of biosynthesis of enediynes by actinomycetes (Fig. S2B and C) (5, 6).

Inspired by the accuracy and specificity of the virtual screening, we adapted our recently developed high-throughput real-time PCR method (23) to survey the TSRI actinomycetes collection for the enediyne PKS gene cassettes to identify new enediyne producers (Fig. 1A). Two sets of PCR primers were designed to specifically target *E5/E* or *E/E10* (Fig. 1B). Hits identified by both sets of the primers featured the enediyne PKS gene cassettes with *E5/E/E10* clustered together, while hits identified by only one of the two sets of primers featured an enediyne PKS gene cassette with either *E5* or *E10* separate from the *E* gene (Fig. 1B). By real-time PCR in a 384-well plate format, specific PCR products were rapidly identified, in a high-throughput manner, by melting curve analysis and confirmed by gel electrophoresis and DNA sequencing (Fig. 1C). From 3,400 representative strains, 81 distinct enediyne producers were identified on the basis of the identity of the enediyne PKS gene cassettes (i.e., *E5*, *E10*, and a 1-kb internal fragment of *E*), taxonomy, and geographic locations where the strains were isolated (Fig. 1B to E). Phylogenetic analysis of the 81 new enediyne PKS cassettes, with the known enediyne PKS cassettes as controls, was carried out using the translated 1-kb internal fragment of *E*. While each of the enediyne PKS cassettes is unique, the phylogenetic tree of the 81 new enediyne PKS cassettes subjected to 90% amino acid identity cutoff collapsed into 28 distinct clades (the pairwise comparison of the known enediyne PKS cassettes revealed amino acid sequence identities ranging from 33% to 69%) (Fig. 2A). It is therefore very significant that 27 of the 28 clades are distinct from the known enediyne PKS cassettes, indicative of novel enediynes. The CB02366 group forms a clade with the C-1027 enediyne PKS cassette (11), suggesting that these hits potentially represent new C-1027 producers (Fig. 2A, section A).

Genome sequencing of 31 representative hits from the 28 clades confirmed that they all contain an enediyne gene cluster and therefore are true enediyne producers (Fig. 2). Significantly, hits from different clades yielded distinct enediyne gene clusters, while hits from the same clade afforded highly homologous gene clusters, as exemplified by TSRI0395, TSRI0261, CB02115, and CB00072 (Fig. 2A, section B) (only the cluster from CB00072 is shown; Fig. 2C). These findings support the proposal of using the enediyne PKS cassette clades to further prioritize the hits and streamline enediyne natural-product dereplication.

Discovery of *Streptomyces* sp. strain CB02366 as a new C-1027 producer with a significantly higher titer than the original producer. C-1027 was originally isolated from *Streptomyces globisporus* in 1993 (10). C-1027 has been in clinical development as an antibody-drug conjugate (ADC) for hepatoma (24, 25). We have been studying C-1027 biosynthesis in *S. globisporus* as a model for enediyne biosynthesis and engineering (7, 11, 20, 22, 26). To confirm the four hits that form a clade with the C-1027 enediyne PKS cassette as alternative C-1027 producers (Fig. 2A, section A), we sequenced the genome of *Streptomyces* sp. strain CB02366, a representative from the clade, and indeed revealed a C-1027 biosynthetic gene cluster (Fig. 3A). We subsequently isolated C-1027 from *Streptomyces* sp. strain CB02366 and generated Δ *pkcE* mutant strain SB1036 (Fig. S3; see also Table S1 in the supplemental material), which has completely lost its ability to produce C-1027, thereby confirming that the cloned gene cluster

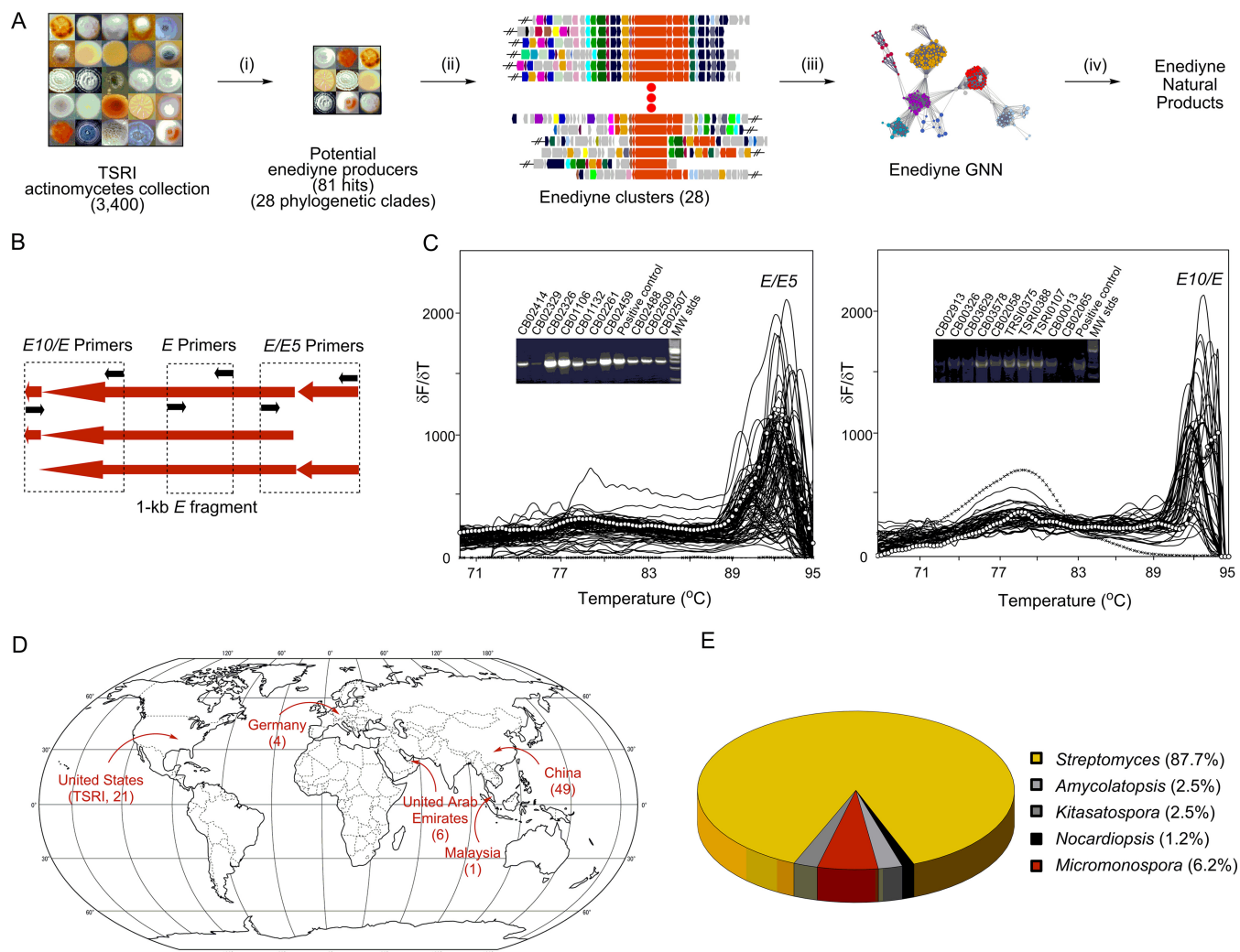


FIG 1 Strain prioritization and genome mining for eneidyne natural products. (A) A high-throughput method to survey the eneidyne biosynthetic machinery in a strain collection and prioritize the hits for new eneidyne natural-product discovery: (i) genome survey of 3,400 representative strains from the TSRI actinomycetes collection identifying 81 novel eneidyne producers; (ii) genome sequencing of 31 representative producers yielding 28 distinct eneidyne biosynthetic gene clusters; (iii) construction of an eneidyne GNN unveiling new insights for eneidyne biosynthesis and structural novelty; and (iv) fermentation optimization and production, isolation, and structural elucidation affording the new eneidyne natural products. (B) Design of PCR primers for eneidyne PKS gene cassette, targeting *E5/E* or *E10/E*, and the primers for the 1-kb internal fragment of *E*. (C) Representative melting curve analysis in real-time PCR in a 384-well plate format, as exemplified by using *E/E5* or *E10/E* primers, with each of the peaks indicating a specific product. Solid lines with open circles represent the positive controls (genomic DNAs of the C-1027 producer *S. globisporus* and the KED producer *Streptomyces* sp. strain ATCC 53650 were used as positive controls for *E/E5* and *E10/E* amplification, respectively), and dashed lines represent the negative controls with no-template DNA. Insets show PCR products of the hits that were analyzed by agarose gel electrophoresis and confirmed by DNA sequencing. (D) Geographic distribution of the 81 new eneidyne producers identified (see Table S2 for details). Numbers in parentheses are the numbers of new eneidyne producers isolated from each of the clades. (E) Taxonomic distribution of the 81 newly identified eneidyne producers in the culture collection, with approximately 88% of them belonging to the *Streptomyces* genus, ensuring that the expedient genetic tools developed in the past two decades for *Streptomyces* will be readily available to manipulate the eneidyne biosynthetic machinery in these producers for production, titer improvement, and structural diversity.

encodes the C-1027 biosynthetic machinery (Fig. 3B and C). Interestingly, CB02366, isolated from a soil sample collected in Dubai, United Arab Emirates, is classified as a *Streptomyces griseus* species (Table S2), but the original C-1027 producer, isolated from a soil sample collected in Qianjiang county, Hubei Province, China, is known as *Streptomyces globisporus* (27). And yet the two C-1027 clusters are highly homologous (with identical genetic organizations and ~90%/94% identity of DNA/amino acid sequences) (Fig. 3A and Table S3). The finding of *Streptomyces* sp. strain CB02366 as an alternative C-1027 producer is significant

because C-1027 had not been rediscovered since it was first isolated in 1993 (10). Remarkably, *Streptomyces* sp. strain CB02366 produces the C-1027 chromoprotein with a titer of ~750 mg/liter, which is minimally 10-fold higher than that produced by the original *S. globisporus* wild-type strain at ~74 mg/liter (calculated on the basis of 5.5 mg/liter of the C-1027 chromophore as determined by high-performance liquid chromatography [HPLC] analysis) (Fig. 3B) (9). Multiple producers (for either known or new eneidyne) with various growth characteristics and levels of genetic amenability also present new opportunities for yield im-

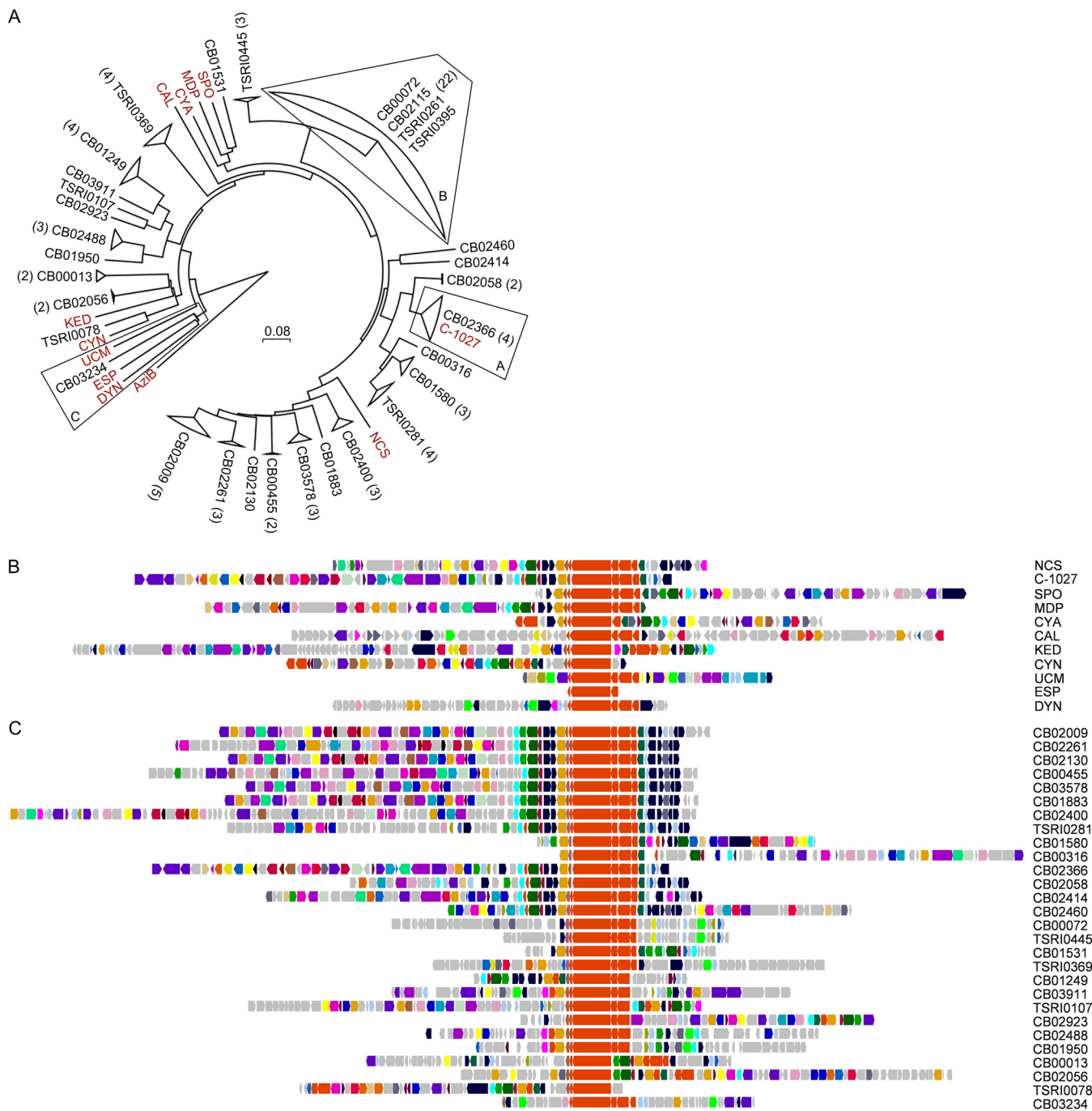


FIG 2 A genome survey of 3,400 strains from the TSRI actinomycetes collection, identifying 81 enediyne producers. (A) Phylogenetic analysis of the 81 enediyne gene clusters, based on the translated 1-kb internal fragment of the *E* genes, in comparison with the 11 known enediyne producers (red), affording 28 distinct clades upon employing a 90% amino acid identity cutoff value. Numbers in parentheses are the hits identified from each of the clades. The clades of alternative C-1027 producers (section A), multiple producers sequenced with highly homologous enediyne gene clusters (section B), and the TNM producer *Streptomyces* sp. strain CB03234 (section C) are highlighted. (B) The 10 enediyne biosynthetic gene clusters known previously and the *ucm* biosynthetic gene cluster characterized in this study (see Table S3 for detailed annotation). (C) The 28 distinct enediyne gene clusters identified upon genome sequencing of 31 representative hits from the 28 clades (see Tables S3 and S4 for detailed annotation). Genes are color-coded based on GNN annotation (see Fig. 4).

provement, structural diversity, or both by applying metabolic pathway engineering strategies and methods (1, 28). The fact that most of the new enediyne producers identified in this study are *Streptomyces* species (Fig. 1E) ensures that the expedient recombinant DNA technologies and genetic tools available can be readily adopted for these experiments.

An enediyne GNN facilitating gene cluster annotation and predicting new structural features. For the other 27 clades, genome sequencing of representative hits unveiled gene clusters that are distinct from all enediyne biosynthetic gene clusters known to date (Fig. 2), indicative of novel enediyne natural products. GNNs have recently emerged as a component of a powerful bioinformat-

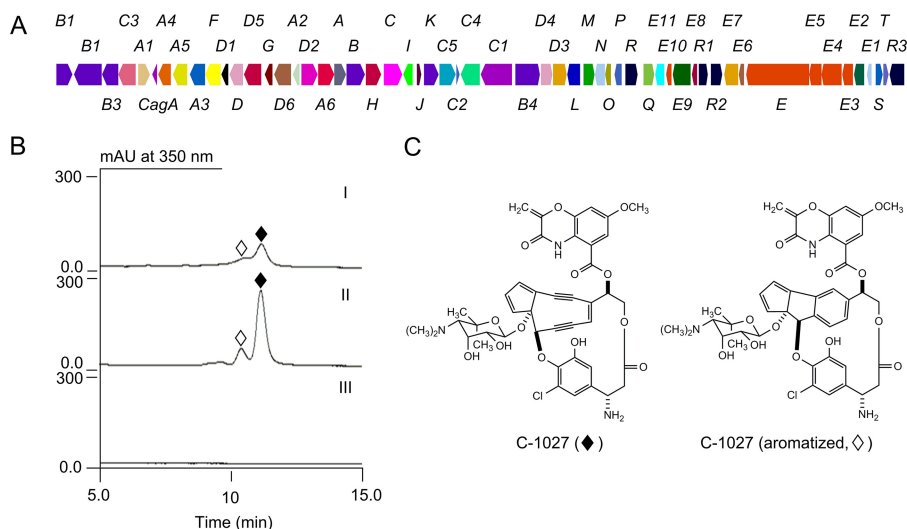


FIG 3 Characterization of *Streptomyces* sp. strain CB02366 as an alternative C-1027 producer. (A) Genetic organization of the C-1027 gene cluster from CB02366 (see Table S3). (B) HPLC analysis of fermentations of the *Streptomyces* sp. strain CB02366 wild type (II) and SB1036 (i.e., $\Delta pksE$) mutant (III), with the *S. globisporus* wild-type strain as a control (I), establishing CB02366 as a new C-1027 producer with a significantly higher C-1027 (solid diamonds) titer. Titers for C-1027 chromoprotein production were estimated on the basis of the C-1027 chromophore as determined by HPLC analysis. The *S. globisporus* wild-type strain produces ~5.5 mg/liter of the C-1027 chromophore, which translates to ~74 mg/liter of the C-1027 chromoprotein (9). mAU, milli-absorbance units. (C) The structures of C-1027 (solid diamonds) and its aromatized metabolite (open diamonds).

ics strategy to predict enzyme functions on a large scale based on their genomic context (29, 30). Given the high complexity of enediynes gene clusters (often spanning 80 kb to 120 kb of DNA and consisting of 50 to 75 open reading frames [ORFs]), the large number of functionally unassigned proteins in these gene clusters, and, most importantly, the desire to compare the new enediynes gene clusters to the known ones, we constructed an enediynes GNN to quickly and accurately analyze the new enediynes gene clusters. The enediynes GNN (Fig. 4) included all known enediynes gene clusters (Fig. 2B) and the 31 new enediynes gene clusters (i.e., the 28 distinct gene clusters and the 3 homologous gene clusters from the same clade of CB00072 [Fig. 2A, section B]) (Fig. 2C). The newly discovered enediynes biosynthetic gene clusters are diverse and rich in new chemistry (Fig. 4), featuring many enzyme families that are different from those encoded by the known enediynes gene clusters or are functionally unknown (Fig. 4C and D). Close examination of the GNN unveiled many new insights that could be exploited to predict novel structural features and guide experimental designs to discover new enediynes natural products.

The 28 distinct new gene clusters can be divided into two groups using our recently reported method to predict 9-membered or 10-membered enediynes based on the exclusive presence of *E2* or *R3* genes, respectively (5)—21 clusters expected to encode the production of 9-membered enediynes and 7 clusters expected to encode the production of 10-membered enediynes (Fig. 4B and Table S2). Among the predicted 9-membered producers (Table S2), seven contain apoprotein genes (encoding CB01883, CB02009, CB02130, CB02261, CB02366, CB02400, and CB03578) (also see Table S4), and many strains share common genes responsible for the biosynthesis of peripheral moieties, such as β -amino acids (CB00455, CB01883, CB02009, CB02130, CB02261, and CB02414), amino sugar (CB01580, CB02414, CB02460, and TSRI0369), and benzoxazolinone (CB02261, CB02130, CB02009, CB01883, CB00455, and CB02400) (Table S4). Besides these

known moieties, several predicted 9-membered enediynes gene clusters contain genes encoding the biosynthesis of unknown moieties. For example, the enediynes gene cluster from CB03578 encodes enzymes for tryptophan degradation (making those genes unique among enediynes biosynthetic gene clusters), methyltransferases, and corresponding activation and condensing enzymes, implying the presence of an anthranilic acid-like peripheral moiety (Fig. 4E and Table S4).

Among the strains harboring predicted 10-membered enediynes biosynthetic gene clusters (Table S2), one strain (TSRI0369) is proposed to produce an enediynes compound structurally similar to CAL (Table S4) and another one (CB03234) (Table S3) shares many homologous genes with the DYN biosynthetic gene cluster from *Micromonospora chersina* and the UCM gene cluster from *S. uncialis*, while the other five gene clusters (CB00072, TSRI0455, CB03911, CB02488, and CB01950) are fundamentally different from the reported 10-membered enediynes biosynthetic gene clusters (also see Table S4), implying that they might produce new families of 10-membered enediynes. This notion is highlighted by results seen with the gene cluster from CB03911, which possesses putative 10-membered enediynes resistance genes, shikimate pathway enzyme genes, and several singletons (Fig. 4F and Table S4). Chorismate metabolism has been seen in 9-membered enediynes biosynthesis (e.g., benzoxazolinone in C-1027) but was not seen in suspected 10-membered enediynes biosynthesis until now (4, 7, 11).

Tiancimycins (TNMs) from *Streptomyces* sp. strain CB03234 demonstrating our approach in rapidly discovering new enediynes. We noticed that *Streptomyces* sp. strain CB03234 is clustered together with the UCM (31), ESP (32), and DYN (33) producers upon phylogenetic analysis of their enediynes PKS cassettes (Fig. 2A, section C), suggesting that CB03234 might produce a new 10-membered enediynes. The *dyn* biosynthetic gene cluster was previously cloned from *M. chersina* (Fig. 5A) (16), but

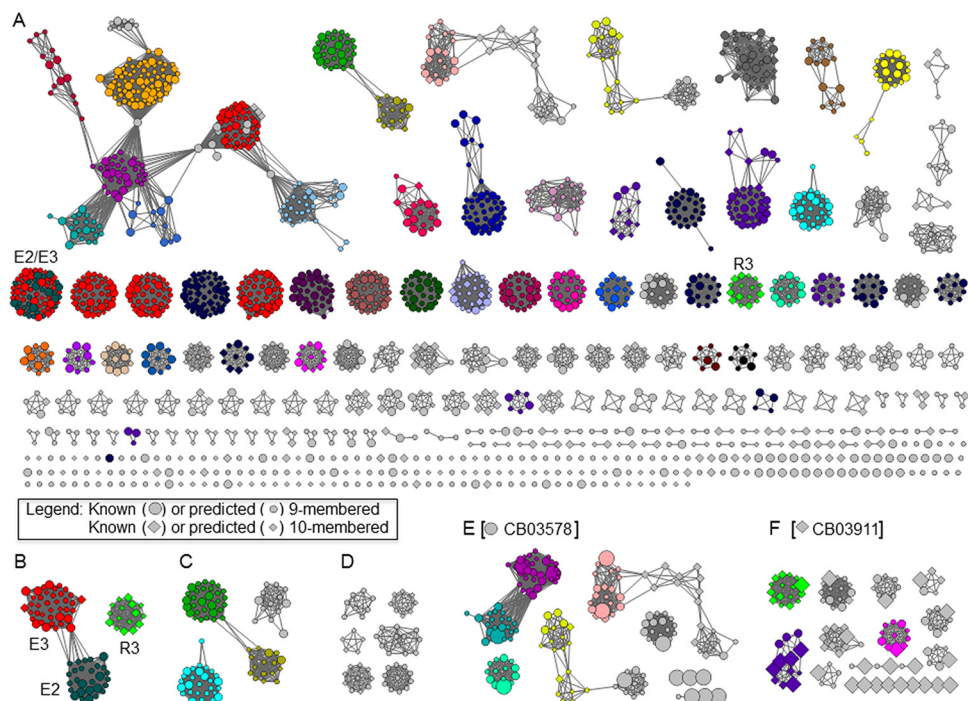


FIG 4 An enediyne GNN revealing that the enediyne biosynthetic gene clusters discovered are rich in new chemistry and diverse structures. (A) The enediyne GNN consisting of the known and new enediyne gene clusters (see Fig. 2B and C) displayed with an E value threshold of 10^{-6} . The enediyne GNN displays both the conserved nature and the diverse nature of proteins involved in enediyne biosynthesis. (B) The E2, E3, and R3 families displayed with an E value threshold of $10^{-7.5}$ to separate the E2 and E3 subfamilies for use as 9-membered versus 10-membered enediyne prediction. (C) Selected representatives of proteins of unknown function, highlighting the vast number of uncharacterized and unknown chemistries found in enediyne biosynthesis. (D) Selected representatives of proteins of unknown function found only in the new enediyne gene clusters, representing novel chemistry, enzymology, and structures found in the new enediyne producers. Big and small dots and diamonds in panels A to D denote known and predicted 9- and 10-membered enediynes, respectively. (E) A predicted 9-membered enediyne cluster from CB03578 (denoted by big dots) featuring tryptophan degradation enzymes with corresponding activation and condensing enzymes, uncharacterized protein families, and singletons. (F) A predicted 10-membered enediyne cluster from CB03911 (denoted with big diamonds), with several conserved families of proteins for regulation and resistance in 10-membered biosynthesis and shikimate pathway enzymes and singletons unique to 10-membered enediyne biosynthesis.

the gene cluster for UCM, discovered from *S. uncialis* (31), had not been cloned until now. Thus, we first identified the *ucm* gene cluster by sequencing the *S. uncialis* genome (Fig. 5A and Table S3) and inactivated the *ucmE* gene in *S. uncialis* to generate Δ *ucmE* mutant strain SB18001 (Fig. S3 and Table S1). HPLC analysis of SB18001 fermentation, with the *S. uncialis* wild type as a control, showed complete abolishment of UCM production, confirming that the cloned gene cluster encodes UCM biosynthesis (Fig. 5C). We next identified the new enediyne gene cluster (i.e., the *tnm* cluster) from *Streptomyces* sp. strain CB03234 by genome sequencing (Fig. 5A). Remarkably, while the *dyn*, *ucm*, and *tnm* clusters show little conservation in genetic organization beyond the enediyne PKS cassette (Fig. 5A), GNN analysis revealed that the three clusters share many homologous genes (Fig. 5B; also see Table S3) and may therefore encode the biosynthesis of a common enediyne core with various peripheral moieties.

To discover the new enediyne from *Streptomyces* sp. strain CB03234, we first inactivated the *tnmE* gene in *Streptomyces* sp. strain CB03234 to afford the corresponding Δ *tnmE* mutant strain SB20001 (Fig. S3 and Table S1). Comparison of the HPLC metabolite profiles between fermentations of the *Streptomyces* sp. CB03234 wild-type and SB20001 mutant strains revealed one major metabolite whose biosynthesis could be readily correlated to the *tnm* gene cluster (Fig. 5D). This metabolite was subsequently

isolated and named TNM A. The structure of TNM A was established on the basis of extensive mass spectrometry (MS) and one-dimensional (1D) and 2D nuclear magnetic resonance (NMR) analysis (Fig. 6A and B; see also Fig. S4 and Table S5), with its absolute stereochemistry established by comparison of the circular dichroism (CD) spectra of TNM A to an authentic UCM standard (Fig. 6C). TNM A featured an anthraquinone-fused 10-membered enediyne core, characteristic of both DYN and UCM. The DYN producer, isolated from a soil sample collected in Gujarat state, India, was identified as a *M. chersina* species (33). The UCM producer, extracted from the surface of a lichen specimen collected in British Columbia, Canada, was classified as a *S. uncialis* species (31). *Streptomyces* sp. strain CB03234 was isolated from a soil sample collected in Yuanjiang county, Yunnan Province, China, and has been assigned as a *Streptomyces venezuelae* species (Table S2). Pairwise comparison of the *dyn*, *ucm*, and *tnm* clusters, as exemplified by the translated products of the enediyne PKS cassette (i.e., *E10/E/E5/E4/E3*; Fig. 5A), revealed only 41% to 60% amino acid sequence identity. It is therefore fascinating that three distantly related strains, harboring gene clusters with distinct organizations, biosynthesize structurally related enediynes, indicative of an intricate evolution for these complex natural-product biosynthetic machineries.

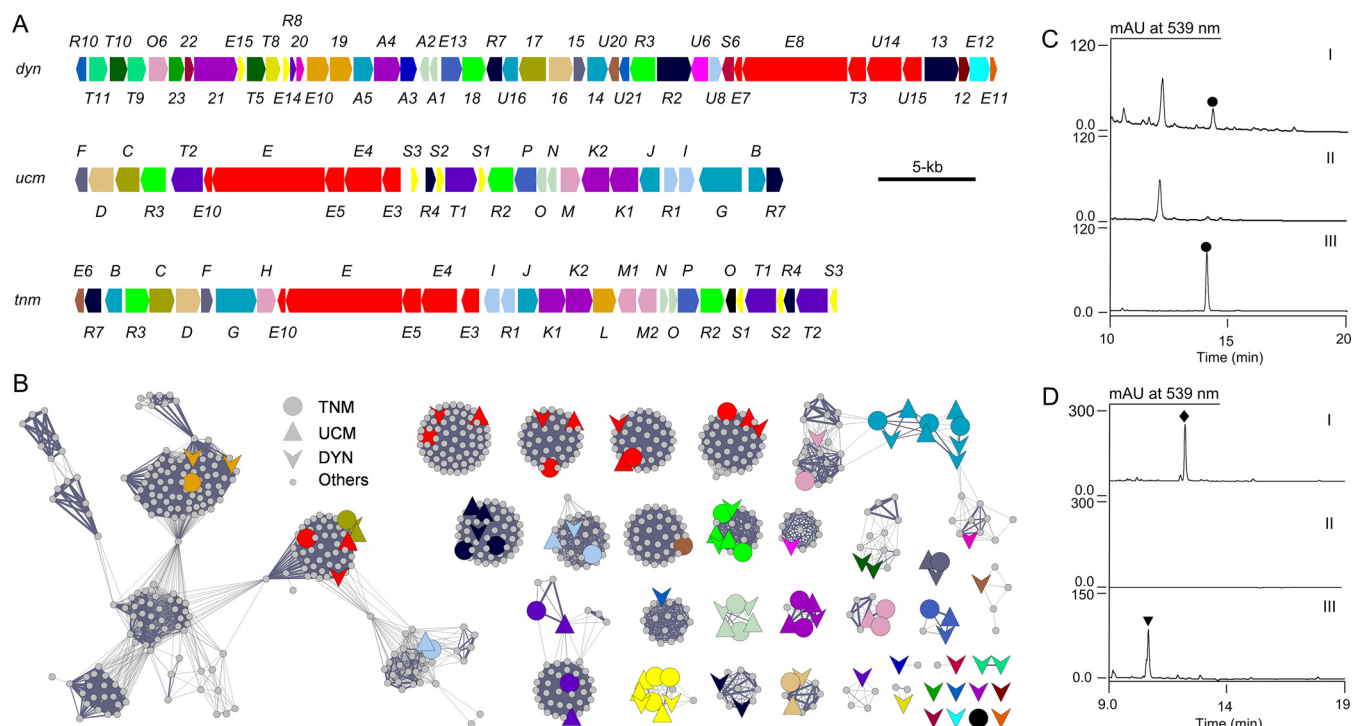


FIG 5 Characterization of *Streptomyces* sp. strain CB03234 as a TNM producer. (A) Genetic organization of the *ucm*, *tnm*, and *dyn* gene clusters (see Table S3). Genes are color-coded based on GNN annotation (see Fig. 4). (B) GNN analysis (E value of 10^{-6}) unveiling functional similarity among the *dyn*, *ucm*, and *tnm* biosynthetic gene clusters. (C) HPLC analysis of fermentations of the *S. uncialis* wild-type strain (I) and SB18001 (i.e., $\Delta ucmE$) mutant (II), in comparison with an authentic UCM standard (III), establishing the cloned gene cluster encoding UCM (solid circles) biosynthesis. (D) HPLC analysis of fermentations of the *Streptomyces* sp. strain CB03234 wild type (I) and the SB20001 (i.e., $\Delta tnmE$) (II) and SB20002 (i.e., $\Delta tnmH$) (III) mutants, establishing the cloned gene cluster encoding TNM A (solid diamond) biosynthesis and demonstrating the feasibility of manipulating TNM biosynthesis in *Streptomyces* sp. strain CB03234 as exemplified by the engineered production of TNM C (inverted triangle) from SB20002.

Manipulation of TNM biosynthesis in *Streptomyces* sp. strain CB03234 demonstrating the feasibility of analogue engineering. Microbial genome mining and metabolic pathway engineering are rapidly changing the landscape of discovery and structural diversity of natural products (1, 28–30, 34, 35). Development of an expedient genetic system for *in vivo* manipulation of the targeted biosynthetic machinery is of paramount importance to implement these emerging strategies. Thus, a critical decision in manipulating enediyne biosynthesis is the selection of the producers that are compatible with the expedient technologies and tools of recombinant DNA work in *Streptomyces* species and related organisms that have been developed in the past two decades (1, 28, 34, 36, 37). Accordingly, the TSRI actinomycetes collection is enriched with *Streptomyces* species (Fig. 1E), and this selection is vital to overcoming the current challenges of, and meeting future objectives for, enediyne discovery, biosynthesis, and engineering in their native producers.

We have developed an expedient genetic system for *Streptomyces* sp. strain CB03234. The feasibility of manipulating TNM biosynthesis in *Streptomyces* sp. strain CB03234 to generate novel analogues has been demonstrated by the engineered production of a TNM analogue. Thus, we inactivated the *tnmH* gene, encoding a putative *O*-methyltransferase (Fig. 5A and B and Table S3), to generate $\Delta tnmH$ mutant strain SB20002 (Fig. S3 and Table S1). HPLC analysis of SB20002 fermentation, with the *Streptomyces* sp. strain CB03234 wild type as a control, showed complete abolishment of TNM A production and a concomitant accumulation of a

new metabolite (Fig. 5D), which was subsequently identified as TNM C on the basis of extensive MS, CD, and 1D and 2D NMR analysis (Fig. 6; see also Fig. S4 and Table S5). TNM C featured an -OH group at C-7, as would be expected from inactivating the *tnmH* gene, but an additional side chain at C-26, revealing new insights into TNM A biosynthesis. Thus, TnmH-catalyzed *O*-methylation at C-7 most likely takes place early in TNM A biosynthesis, without which the side chain at C-26 cannot be fully processed *en route* to TNM A as has been proposed previously for UCM biosynthesis (31). Given the biosynthetic relationship among DYN, TNM, and UCM (Fig. 5A and B), manipulation of TNM biosynthesis in CB03234 therefore provides an outstanding bioengineering platform to access the DYN and UCM scaffolds, which have been difficult to access in their native producers due to either recalcitrance to genetic manipulation (for DYN) (16) or the inability to produce in submerged fermentation (for UCM) (31).

TNMs exhibiting potent cytotoxicity, with rapid and complete killing, toward a broad spectrum of cancer cell lines. The enediynes are among the most cytotoxic molecules known to date, and they are active in many tumor types. Although the natural enediynes have limited use as clinical drugs, both polymer-based delivery systems and ADCs have shown great clinical success or promise in anticancer therapy (3, 7). For example, poly(styrene-co-maleic acid)-conjugated NCS (SMANCS) has been marketed since 1994 for use against hepatoma (38). Several ADCs have been developed, including an anti-CD33 monoclonal antibody (MAb)-CAL conjugate (i.e., gemtuzumab ozogamicin) for acute myeloid

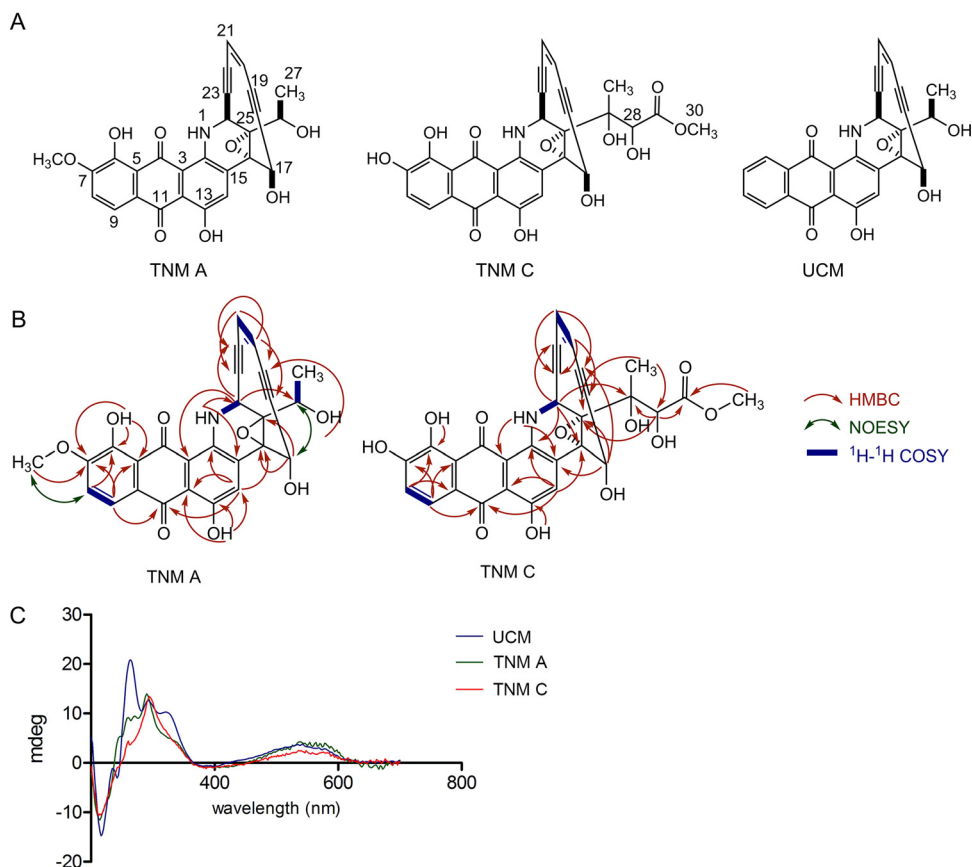


FIG 6 Structural elucidation of TNM A and C on the basis of 1D and 2D NMR and CD spectroscopic data analysis. (A) Structures of TNM A and C in comparison with that of UCM. (B) Key ^1H - ^1H correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY), and heteronuclear multiple bond correlation (HMBC) analysis of TNM A and C, supporting their structural assignments (also see Table S4). (C) CD spectra of TNM A and C in comparison with an authentic standard of UCM, supporting their absolute stereochemistry assignment.

leukemia (AML) and an anti-CD22 MAb-CAL conjugate (inotuzumab ozogamicin) for non-Hodgkin lymphoma (39, 40), as well as several MAb-C-1027 conjugates for hepatoma (24) and MAb-UCM conjugates for selected tumors (N. S. Chowdari, S. Gangwar, and B. Sufi, 22 August 2013, European patent application WO 2013122823 A1). These examples demonstrate that the enediyne can be developed into powerful drugs and that the new enediyne therefore represent outstanding payload candidates for ADCs.

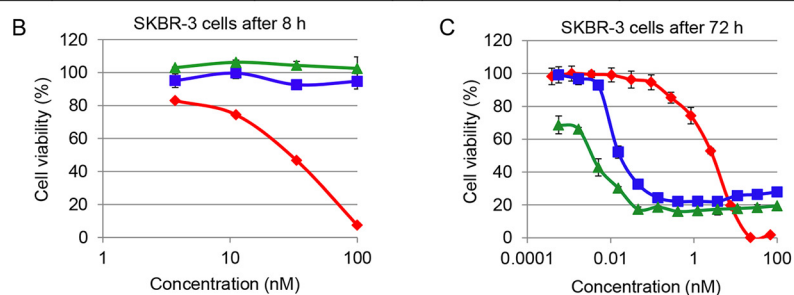
We carried out preliminary cytotoxicity evaluations of TNMs in comparison with UCM, auristatin F phenylenediamine (AFP) (41), a variant of the natural product dolostatin 10, and maytansinoid AP-3 (Fig. S1C). Auristatins and maytansinoids are used in the FDA-approved ADCs brentuximab vedotin (Adcetris) and trastuzumab emtansine (Kadcyla), respectively, as well as in many ADCs currently in clinical development (39, 40, 42, 43). The TNMs are extremely potent against a broad spectrum of cancer cell lines, with subnanomolar 50% inhibitory concentrations (IC_{50} s). For example, TNM A is more potent than UCM, particularly against breast cancer cell lines (Table 1A); UCM is currently in preclinical development as an ADC (Chowdari et al., European patent application WO 2013122823 A1) (44). Most impressively, as exemplified with the SKBR-3 breast cancer cell line, TNM A exhibited more rapid and more complete cell killing than AFP and AP-3 (Tables 1B and C), thereby minimizing the development of

potential drug resistance (45, 46), supporting the wisdom of exploiting the TNMs as payload candidates for the next generation of anticancer ADCs.

Conclusions and significance. In spite of their profound impact on modern chemistry, biology, and medicine, the enediyne family of natural products remains very small, with only 11 members structurally characterized to date. Recent advances in DNA sequencing and microbial genomics, however, clearly revealed that the biosynthetic potential of soil actinomycetes to produce enediyne is greatly underappreciated. A great challenge is that of developing innovative methods to discover new enediyne and producing them in sufficient quantities for chemical, biological, and clinical investigations. We recently reported a high-throughput real-time PCR method to prioritize strains for natural-product discovery. By adapting this method to identify strains that are highly likely to encode enediyne biosynthesis, followed by genome sequencing, bioinformatics analysis, genetic manipulation, and fermentation optimization, we have now demonstrated the feasibility of rapid discovery of new enediyne from a large strain collection. The new C-1027 producers, with a significantly higher C-1027 titer than the original producer, will impact the practical supply of the drug should it eventually be brought into clinical applications. The TNMs, with their extremely potent cytotoxicity against a broad spectrum of cancer cells and rapid and complete cell killing characteristics, in comparison with the pay-

TABLE 1 Cytotoxicity of and killing of SKBR-3 cells by TNM A^a

A	cell lines	cancer type	IC ₅₀ [nM]	
			UCM	TNM A
	MDA-MB-468	breast	0.35 ± 0.03	0.31 ± 0.03
	MDA-MB-231	breast	12 ± 0.23	2.3 ± 0.21
	SKBR-3	breast	19 ± 1.2	2.3 ± 0.19
	KPL-4	breast	5.8 ± 0.89	0.33 ± 0.02
	BT474	breast	33 ± 5.2	2.2 ± 0.35
	DYT2	breast	8.2 ± 0.13	2.3 ± 0.04
	M14	melanoma	0.41 ± 0.11	0.84 ± 0.01
	SK-MEL-5	melanoma	2.9 ± 0.29	0.33 ± 0.01
	NCI-H226	NCS lung	4.3 ± 0.17	9.2 ± 0.44
	SF-295	CNS	0.97 ± 0.10	1.1 ± 0.09
	SF-539	CNS	1.0 ± 0.33	0.91 ± 0.08



^a (A) Cytotoxicity of TNM A against selected breast and other cancer cell lines in comparison to UCM. (B) Rapid and (C) complete killing of SKBR-3 cells by TNM A (red diamonds) in comparison with AFP (green triangles) and AP-3 (blue squares) (see Fig. S1 for structures), analogues of the natural products auristatin and maytansine used as payloads in the FDA-approved ADCs brentuximab vedotin (Adcetris) and trastuzumab emtansine (Kadcyla), respectively. Each point represents the mean ± SD of results from at least three replicates, and the IC₅₀s were determined by computerized curve fitting using GraphPad Prism. ADC, antibody-drug conjugates; TNM, tiancimycin; UCM, uncialamycin.

loads used in FDA-approved ADCs and ADCs in various stages of development, are poised to be exploited as payload candidates for the next generation of anticancer ADCs. Follow-up studies on the other hits that have already been identified in this study or on application of our strategy to other strain collections promise the discovery of new enediynes, radically expanding the chemical space for the enediyne family of natural products. Our results also support strain prioritization and genome mining for the discovery of other classes of natural products from ever-growing microbial strain collections. Together with other emerging strategies and technologies, these findings will inspire continued innovations in natural-product discovery.

MATERIALS AND METHODS

General materials. Primers, plasmids, and strains used and reported in this study are summarized in Table S1 in the supplemental material.

Construction of the actinomycetes genomic DNA library. Genomic DNA (gDNA) was prepared from strains isolated from various unexplored and underexplored ecological niches (23, 47, 48). Strains were cultivated in a rich liquid medium (tryptic soy broth [TSB]) for 2 to 3 days. DNA was isolated using the salting-out protocol (36), deposited into 96-well plates, and stored at -80°C.

Real-time PCR screening of 3,400 strains for enediyne producers. Real-time PCR was performed using an Applied Biosystems 7900HT Fast real-time PCR system. Preparation of gDNA arrays and application of the real-time PCR method to survey the 3,400 actinomycetes strains for the enediyne PKS gene cassette and to identify new enediyne producers

followed the published protocol for strain prioritization with modifications (23). For further details, see Text S1 in the supplemental material.

Genome sequencing and assembly. Genome sequencing of the representative enediyne producers was performed using an Illumina MiSeq sequencer (2 × 300 paired-end sequencing) at the Next Generation Sequencing and Microarray Core Facility, TSRI. Read quality filtering was performed using a tool developed in-house. Adapter trimming and *de novo* assembly were done with CLC Genomics Workbench version 7.5.1 (CLC Bio.) using default settings. The resulting contigs were further extended and joined into a final scaffold by SSPACE version 2.0 (49) using all quality-filtered reads. The remaining gaps inside the final scaffold were partially or completely filled using the quality-filtered reads by GapFiller version 1.10 (50). The draft genome sequences of the selected hits (under BioProject [PRJNA293172](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA293172)) and *S. uncialis* DCA2648 (under BioProject [PRJNA286672](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA286672)) reported in this study have been deposited in GenBank, with their accession numbers summarized in Table S2.

Eneidiyne genome neighborhood network (GNN). Annotation of the new enediyne gene clusters and construction of the enediyne GNN were carried out as described previously (5, 6). Annotations of all the enediyne biosynthetic gene clusters reported in this study are summarized in Tables S3 and S4. The C-1027 gene cluster from *Streptomyces* sp. strain CB02366, the *ucm* gene cluster from *S. uncialis* DCA2648, and the *tnm* gene cluster from *Streptomyces* sp. strain CB03234 were deposited in GenBank under accession numbers KU597647, KT762610, and KT716443, respectively. Accession numbers for the other new enediyne gene clusters are listed in Table S2.

The enediyne GNN was constructed and included (i) the 7 known 9-membered enediyne gene clusters (encoding C-1027, NCS, MDP, KED,

SPO, CYA, and CYN), (ii) the 3 known 10-membered enediynes gene clusters (encoding CAL, DYN, and ESP), (iii) the newly characterized gene cluster for the 10-membered enediynes UCM, and (iv) the 31 new enediynes gene clusters (i.e., the 28 distinct gene clusters plus the 3 homologous gene clusters from the same clade of CB00072) discovered in this study (Fig. 2). Cytoscape v 3.0 was used for GNN generation, visualization, and analysis (51). All GNNs were displayed using the “organic” layout with edge widths corresponding to the *E* value corresponding to comparisons between proteins. A more detailed description is provided in Text S1.

Gene inactivation. Inactivation of selected genes within the cloned enediynes clusters in *Streptomyces* species was performed by gene replacement following literature procedures (36, 37). The genotypes of the resultant mutants were confirmed by PCR (Table S1) and Southern analysis (Fig. S3) as described in Text S1.

Fermentation, production, and HPLC analyses of C-1027. The *Streptomyces* sp. strain CB02366 wild-type and SB1036 (i.e., Δ *pksE*) mutant strains were cultured individually following previously reported procedures (11), with the original C-1027 producer *S. globisporus* wild-type strain as a control. The identity of C-1027 was confirmed by HPLC and high-resolution electrospray ionization–mass spectrometry HR-ESI-MS analysis (Fig. 3B) as described in Text S1. To determine C-1027 titers, HPLC analysis was calibrated with an authentic C-1027 standard (9, 11).

Fermentation, production, and HPLC analysis of UCM. Fermentation of *S. uncialis* and production of UCM were performed on solid agar medium (ISP4) following previously published procedures (31). The identity of UCM was confirmed by HPLC and HR-ESI-MS analysis with an authentic UCM standard (Fig. 5C) as described in Text S1.

Fermentation, production, and HPLC analyses of TNMA and C. The *Streptomyces* sp. strain CB03234 wild-type, SB20001 (i.e., Δ *tnmE*), and SB20002 (i.e., Δ *tnmH*) mutant strains were cultured individually in 250-ml baffled flasks containing 50 ml of TSB liquid medium. After growth at 28°C and 250 rpm for 2 days, 5 ml of seed culture was inoculated into 250-ml baffled flasks containing 50 ml of the production medium (1% soluble starch, 0.5% Pharmamedia, 0.2% CaCO₃, 0.005% CuSO₄·5H₂O, 0.0005% NaI, pH 7.0). The resulting cultures were incubated at 28°C and 250 rpm for 7 days and individually harvested. Each culture was centrifuged, the supernatant was extracted with EtOAc, and the cell pellet was extracted with CH₃COCH₃. The combined extracts were concentrated in vacuum and dissolved in CH₃OH for HPLC and HR-ESI-MS analysis (Fig. 5D) as described in Text S1. TNM titers were determined by HPLC analysis calibrated with authentic standards.

Isolation and structural elucidation of TNM A and C. For structural elucidation, TNM A (1.2 mg) was isolated from 6 liters of fermentation culture of the *Streptomyces* sp. CB03234 wild-type strain, while TNM C (0.7 mg) was isolated from 6 liters of fermentation culture of the Δ *tnmH* mutant strain SB20002. The structures of TNM A and C were established on the basis of extensive MS, CD, and NMR analysis (Fig. 6 and Table S2). For further details, see Text S1.

Cytotoxicity assay of TNMs. The IC₅₀s of TNMs against selected human cancer cell lines, including breast (MDA-MB-468, MDA-MB-231, SKBR-3, KPL-4, BT474, and DYT2), melanoma (M14 and SK-MEL-5), non-small cell lung cancer (NCI-H226), and central nervous system (SF-295 and SF-539) cells, with UCM as a control, were determined by following standard protocols described previously (47). Each point represents the mean \pm standard deviation (SD) of results from three replicates, and the IC₅₀ was determined by computerized curve fitting using GraphPad Prism (Table 1A). For comparison of the cell killing rates among TNM A, AFP, and AP-3, SK-BR-3 breast cancer cells were incubated with graded doses (ranging from 0 to 100 nM) of drugs for 8, 12, 24, and 72 h before the assays were developed as described above. Cell viability was expressed as a percentage of the level seen with untreated control cells (Tables 1B and C). Further details are provided in Text S1.

Accession number(s). Accession numbers for the C-1027 gene cluster from *Streptomyces* sp. strain CB02366, the *ucm* gene cluster from *S. un-*

cialis DCA2648, and the *tnm* gene cluster from *Streptomyces* sp. strain CB03234 are KU597647, KT762610, and KT716443, respectively. Accession numbers for the draft genomes of the selected 31 hits, as well as the UCM producer *S. uncialis* DCA2648, are summarized in Table S2.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.02104-16/-/DCSupplemental>.

Figure S1, PDF file, 0.5 MB.
Figure S2, PDF file, 0.5 MB.
Figure S3, PDF file, 0.2 MB.
Figure S4, PDF file, 0.3 MB.
Table S1, PDF file, 0.1 MB.
Table S2, PDF file, 0.4 MB.
Table S3, PDF file, 0.1 MB.
Table S4, PDF file, 0.4 MB.
Table S5, PDF file, 0.3 MB.
Text S1, PDF file, 0.3 MB.

ACKNOWLEDGMENTS

Y.D. and B.S. conceived the project and designed the experiments; X.Y., H.G., T.H., H., D.Y., Q.T., and I.C. performed experiments; X.L. and C.R. carried out the cytotoxicity assays; J.D.R. and J.R.L. contributed to bioinformatics analysis; Y.G. and F.V.N. provided genome assembly; X.Z., Y.H., L.-X.Z., Y.J., and Y.D. assisted with strain collection; B.S. wrote the manuscript with help from the other authors.

We thank Y. Li, Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Beijing, China, for the wild-type *S. globisporus* strain that produces C-1027; J. Davies, University of British Columbia, Vancouver, Canada, for the wild-type *S. uncialis* DCA2648 strain that produces UCM; K. C. Nicolaou, Rice University, Houston, TX, for an authentic sample of UCM; C. G. Nelson and T. R. Burke, Jr., NCI, Frederick, MD, for a synthetic sample of AFP; and the John Innes Centre, Norwich, United Kingdom, for providing the REDIRECT technology kit.

This work was supported in part by the Chinese Ministry of Education 111 Project B08034 (to Y.D.), National High Technology Joint Research Program of China grant 2011ZX09401-001 (to Y.D.), National High Technology Research and Development Program of China grant 2012AA02A705 (to Y.D.), and United States NIH grants CA78747 (to B.S.) and GM115575 (to B.S.). I.C. and J.D.R. are supported in part by postdoctoral fellowships from the German Research Foundation and the Arnold and Mabel Beckman Foundation, respectively.

We declare that we have no competing financial interest.

FUNDING INFORMATION

This work, including the efforts of Yanwen Duan, was funded by the Chinese Ministry of Education (111 Project B08034). This work, including the efforts of Yanwen Duan, was funded by National High Technology Joint Research Program of China (2011ZX09401-001). This work, including the efforts of Yanwen Duan, was funded by National High Technology Research and Development Program of China (2012AA02A705). This work, including the efforts of Ben Shen, was funded by HHS | National Institutes of Health (NIH) (CA78747). This work, including the efforts of Ben Shen, was funded by HHS | National Institutes of Health (NIH) (GM115575).

REFERENCES

- Shen B. 2015. A new golden age of natural product drug discovery. *Cell* 163:1297–1300. <http://dx.doi.org/10.1016/j.cell.2015.11.031>.
- Edo K, Mizugaki M, Koide Y, Seto H, Furihata K, Otake N, Ishida N. 1985. The structure of neocarzinostatin chromophore possessing a novel bicyclo[7,3,0]dodecadiyne system. *Tetrahedron Lett* 26:331–334. [http://dx.doi.org/10.1016/S0040-4039\(01\)80810-8](http://dx.doi.org/10.1016/S0040-4039(01)80810-8).
- Galm U, Hager MH, Van Lanen SG, Ju J, Thorson JS, Shen B. 2005. Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem Rev* 105:739–758. <http://dx.doi.org/10.1021/cr030117g>.
- Liang ZX. 2010. Complexity and simplicity in the biosynthesis of enediynes

- natural products. *Nat Prod Rep* 27:499–528. <http://dx.doi.org/10.1039/b908165h>.
5. Rudolf JD, Yan X, Shen B. 2016. Genome neighborhood network reveals insights into enediyne biosynthesis and facilitates prediction and prioritization for discovery. *J Ind Microbiol Biotechnol* 43:261–276. <http://dx.doi.org/10.1007/s10295-015-1671-0>.
 6. Shen B, Hindra, Yan X, Huang T, Ge H, Yang D, Teng Q, Rudolf JD, Lohman JR. 2015. Eneidyne: exploration of microbial genomics to discover new anticancer drug leads. *Bioorg Med Chem Lett* 25:9–15. <http://dx.doi.org/10.1016/j.bmcl.2014.11.019>.
 7. Van Lanen SG, Shen B. 2008. Biosynthesis of enediyne antitumor antibiotics. *Curr Top Med Chem* 8:448–459. <http://dx.doi.org/10.2174/156802608783955656>.
 8. Kennedy DR, Ju J, Shen B, Beerman TA. 2007. Designer enediynes generate DNA breaks, interstrand cross-links, or both, with concomitant changes in the regulation of DNA damage responses. *Proc Natl Acad Sci U S A* 104:17632–17637. <http://dx.doi.org/10.1073/pnas.0708274104>.
 9. Chen Y, Yin M, Horsman GP, Shen B. 2011. Improvement of the enediyne antitumor antibiotic C-1027 production by manipulating its biosynthetic pathway regulation in *Streptomyces globisporus*. *J Nat Prod* 74:420–424. <http://dx.doi.org/10.1021/np100825y>.
 10. Ken-ichiro Y, Minami Y, Azuma R, Saeki M, Otani T. 1993. Structure and cycloaromatization of a novel enediyne, C-1027 chromophore. *Tetrahedron Lett* 34:2637–2640. [http://dx.doi.org/10.1016/S0040-4039\(00\)77644-1](http://dx.doi.org/10.1016/S0040-4039(00)77644-1).
 11. Liu W, Christenson SD, Standage S, Shen B. 2002. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* 297:1170–1173. <http://dx.doi.org/10.1126/science.1072110>.
 12. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM, Thorson JS. 2002. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* 297:1173–1176. <http://dx.doi.org/10.1126/science.1072105>.
 13. Liu W, Nonaka K, Nie L, Zhang J, Christenson SD, Bae J, Van Lanen SG, Zazopoulos E, Farnet CM, Yang CF, Shen B. 2005. The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC15944 involving two iterative type I polyketide synthase. *Chem Biol* 12:293–302. <http://dx.doi.org/10.1016/j.chembiol.2004.12.013>.
 14. Van Lanen SG, Oh TJ, Liu W, Wendt-Pienkowski E, Shen B. 2007. Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. *J Am Chem Soc* 129:13082–13094. <http://dx.doi.org/10.1021/ja052375o>.
 15. Lohman JR, Huang SX, Horsman GP, Dilfer PE, Huang T, Chen Y, Wendt-Pienkowski E, Shen B. 2013. Cloning and sequencing of the kedaricin biosynthetic gene cluster from *Streptoalloteichus* sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics. *Mol Biosyst* 9:478–491. <http://dx.doi.org/10.1039/c3mb25523a>.
 16. Gao Q, Thorson JS. 2008. The biosynthetic genes encoding for the production of the dynamic enediyne core in *Micromonospora chersina* ATCC53710. *FEMS Microbiol Lett* 282:105–114. <http://dx.doi.org/10.1111/j.1574-6968.2008.01112.x>.
 17. Zazopoulos E, Huang K, Staffa A, Liu W, Bachmann BO, Nonaka K, Ahlert J, Thorson JS, Shen B, Farnet CM. 2003. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat Biotechnol* 21:187–190. <http://dx.doi.org/10.1038/nbt784>.
 18. McGlinchey RP, Nett M, Moore BS. 2008. Unraveling the biosynthesis of the sporolide cyclohexenone building block. *J Am Chem Soc* 130:2406–2407. <http://dx.doi.org/10.1021/ja104888m>.
 19. Lane AL, Nam SJ, Fukuda T, Yamanaka K, Kauffman CA, Jensen PR, Fenical W, Moore BS. 2013. Structures and comparative characterization of biosynthetic gene clusters for cyclosporin, enediyne-derived natural products from marine actinomycetes. *J Am Chem Soc* 135:4171–4174. <http://dx.doi.org/10.1021/ja311065v>.
 20. Horsman GP, Chen Y, Thorson JS, Shen B. 2010. Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes. *Proc Natl Acad Sci U S A* 107:11331–11335. <http://dx.doi.org/10.1073/pnas.1003442107>.
 21. Liu W, Ahlert J, Gao Q, Wendt-Pienkowski E, Shen B, Thorson JS. 2003. Rapid PCR amplification of minimal enediyne PKS cassette leads to a predictive familial classification model. *Proc Natl Acad Sci U S A* 100:11959–11963. <http://dx.doi.org/10.1073/pnas.2034291100>.
 22. Zhang J, Van Lanen SG, Ju J, Liu W, Dorrestein PC, Li W, Kelleher NL, Shen B. 2008. A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proc Natl Acad Sci U S A* 105:1460–1465. <http://dx.doi.org/10.1073/pnas.0711625105>.
 23. Hindra, Huang T, Yang D, Rudolf JD, Xie P, Xie G, Teng Q, Lohman JR, Zhu X, Huang Y, Zhao LX, Jiang Y, Duan Y, Shen B. 2014. Strain prioritization for natural product discovery by a high-throughput real-time PCR method. *J Nat Prod* 77:2296–2303. <http://dx.doi.org/10.1021/np5006168>.
 24. Brukner I. 2000. C-1027 Taiho Pharmaceutical Co Ltd. *Curr Opin Oncol Endocr Met Invest Drugs* 2:344–352.
 25. Li W, Li X, Huang T, Teng Q, Crnovcic I, Rader C, Shen B. 2016. Engineered production of cancer targeting peptide (CTP)-containing C-1027 in *Streptomyces globisporus* and biological evaluation. *Bioorg Med Chem* 24:3887–3892. <http://dx.doi.org/10.1016/j.bmc.2016.04.017>.
 26. Beerman TA, Gawron LS, Shin S, Shen B, McHugh MM. 2009. C-1027, a radiomimetic enediyne anticancer drug, preferentially targets hypoxic cells. *Cancer Res* 69:593–598. <http://dx.doi.org/10.1158/0008-5472.CAN-08-2753>.
 27. Zhen YS, Ming XY, Yu B, Otani T, Saito H, Yamada Y. 1989. A new macromolecular antitumor antibiotic, C-1027: III. Antitumor activity. *J Antibiot* 42:1294–1298. <http://dx.doi.org/10.7164/antibiotics.42.1294>.
 28. Baltz RH, Ikeda H, Katz L, Kim ES, van Wezel GP, Wright G. 2016. Special issue: natural product discovery and development in the genomic era. *J Ind Microbiol Biotechnol* 43:109. <http://dx.doi.org/10.1007/s10295-015-1715-5>.
 29. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158:412–421.
 30. Zhao S, Sakai A, Zhang X, Vetting MW, Kumar R, Hillerich B, San Francisco B, Solbiati J, Steves A, Brown S, Akiva E, Barber A, Seidel RD, Babbitt PC, Almo SC, Gerlt JA, Jacobson MP. 2014. Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *Elife* 3:e03275. <http://dx.doi.org/10.7554/eLife.03275>.
 31. Davies J, Wang H, Taylor T, Warabi K, Huang XH, Andersen RJ. 2005. Uncialamycin, a new enediyne antibiotic. *Org Lett* 7:5233–5236. <http://dx.doi.org/10.1021/ol052081f>.
 32. Golik J, Clardy J, Dubay G, Groenewold G, Kawaguchi H, Konishi M, Krishnan B, Ohkuma H, Saitoh K, Doyle TW. 1987. Esperamicins, a novel class of potent antitumor antibiotics. 3. Structure of esperamicins A1, A2 and A1b. *J Am Chem Soc* 109:3462–3464.
 33. Konishi M, Ohkuma H, Matsumoto K, Tsuno T, Kamei H, Miyaki T, Oki T, Kawaguchi H, VanDuyn GD, Clardy J. 1989. Dynemicin A, a novel antibiotic with the anthraquinone and 1,5-diyne-3-ene subunit. *J Antibiot (Tokyo)* 42:1449–1452. <http://dx.doi.org/10.7164/antibiotics.42.1449>.
 34. Iqbal HA, Low-Beinart L, Obiajulu JU, Brady SF. 2016. Natural product discovery through improved functional metagenomics in *Streptomyces*. *J Am Chem Soc* 138:9341–9344. <http://dx.doi.org/10.1021/jacs.6b02921>.
 35. Ju KS, Gao J, Doroghazi JR, Wang KK, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, Lee J, Cioni JP, Evans BS, Hirota R, Labeda DP, van der Donk WA, Metcalf WW. 2015. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci U S A* 112:12175–12180. <http://dx.doi.org/10.1073/pnas.1500873112>.
 36. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. *Practical Streptomyces genetics*. John Innes Foundation, Norwich, United Kingdom.
 37. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* 100:1541–1546. <http://dx.doi.org/10.1073/pnas.0337542100>.
 38. Maeda H, Edo K, Ishida N. 1997. Neocarzinostatin: the past, present, and future of an anticancer drug. Springer-Verlag, New York, NY.
 39. Chari RVJ, Miller ML, Widdison WC. 2014. Antibody-drug conjugates: an emerging concept in cancer therapy. *Angew Chem Int Ed Engl* 53:3796–3827. <http://dx.doi.org/10.1002/anie.201307628>.
 40. Sievers EL, Senter PD. 2013. Antibody-drug conjugates in cancer therapy.

- Annu Rev Med 64:15–29. <http://dx.doi.org/10.1146/annurev-med-050311-201823>.
41. Law CL, Gordon KA, Toki BE, Yamane AK, Hering MA, Cervený CG, Petroziello JM, Ryan MC, Smith L, Simon R, Sauter G, Oflazoglu E, Doronina SO, Meyer DL, Francisco JA, Carter P, Senter PD, Copland JA, Wood CG, Wahl AF. 2006. Lymphocyte activation antigen CD70 expressed by renal cell carcinoma is a potential therapeutic target for anti-CD70 antibody-drug conjugates. *Cancer Res* 66:2328–2337. <http://dx.doi.org/10.1158/0008-5472.CAN-05-2883>.
 42. Drake PM, Rabuka D. 2015. An emerging playbook for antibody-drug conjugates: lessons from the laboratory and clinic suggest a strategy for improving efficacy and safety. *Curr Opin Chem Biol* 28:174–180. <http://dx.doi.org/10.1016/j.cbpa.2015.08.005>.
 43. Polakis P. 2016. Antibody drug conjugates for cancer therapy. *Pharmacol Rev* 68:3–19. <http://dx.doi.org/10.1124/pr.114.009373>.
 44. Nicolaou KC, Wang Y, Lu M, Mandal D, Pattanayak MR, Yu R, Shah AA, Chen JS, Zhang H, Crawford JJ, Pasunoori L, Poudel YB, Chowdari NS, Pan C, Nazeer A, Gangwar S, Vite G, Pitsinos EN. 2016. Streamlined total synthesis of uncialamycin and its application to the synthesis of designed analogues for biological investigations. *J Am Chem Soc* 138:8235–8246. <http://dx.doi.org/10.1021/jacs.6b04339>.
 45. Loganzo F, Tan X, Sung M, Jin G, Myers JS, Melamud E, Wang F, Diesl V, Follettie MT, Musto S, Lam MH, Hu W, Charati MB, Khandke K, Kim KS, Cinque M, Lucas J, Graziani E, Maderna A, O'Donnell CJ, Arndt KT, Gerber HP. 2015. Tumor cells chronically treated with a trastuzumab-maytansinoid antibody-drug conjugate develop varied resistance mechanisms but respond to alternate treatment. *Mol Cancer Ther* 14:952–963. <http://dx.doi.org/10.1158/1535-7163.MCT-14-0862>.
 46. Yu SF, Zheng B, Go M, Lau J, Spencer S, Raab H, Soriano R, Jhunjhunwala S, Cohen R, Caruso M, Polakis P, Flygare J, Polson AG. 2015. A novel anti-CD22 anthracycline-based antibody-drug conjugate (ADC) that overcomes resistance to auristatin-based ADCs. *Mol Cancer Ther* 21:3298–3306. <http://dx.doi.org/10.1158/1078-0432.CCR-14-2035>.
 47. Ma M, Rateb ME, Teng Q, Yang D, Rudolf JD, Zhu X, Huang Y, Zhao LX, Jiang Y, Li X, Rader C, Duan Y, Shen B. 2015. Angucyclines and angucyclinones from *Streptomyces* sp. CB01913 featuring C-ring cleavage and expansion. *J Nat Prod* 78:2471–2480. <http://dx.doi.org/10.1021/acs.jnatprod.5b00601>.
 48. Xie P, Ma M, Rateb ME, Shaaban KA, Yu Z, Huang SX, Zhao LX, Zhu X, Yan Y, Peterson RM, Lohman JR, Yang D, Yin M, Rudolf JD, Jiang Y, Duan Y, Shen B. 2014. Biosynthetic potential-based strain prioritization for natural product discovery: a showcase for diterpenoid-producing actinomycetes. *J Nat Prod* 77:377–387. <http://dx.doi.org/10.1021/np401063s>.
 49. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
 50. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
 51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <http://dx.doi.org/10.1101/gr.1239303>.