

Activating cryptic biosynthetic gene cluster through a CRISPR–Cas12a-mediated direct cloning approach

Mindong Liang^{1,†}, Leshi Liu^{1,†}, Fei Xu^{2,†}, Xiaoqian Zeng^{1,†}, Ruijun Wang³, Jinling Yang², Weishan Wang⁴, Loganathan Karthik¹, Jiakun Liu⁵, Zhiheng Yang¹, Guoliang Zhu¹, Shuliu Wang¹, Linqun Bai⁶, Yaojun Tong⁶, Xueting Liu¹, Min Wu⁷, Li-Xin Zhang¹ and Gao-Yi Tan^{1,*}

¹State Key Laboratory of Bioreactor Engineering and School of Biotechnology, East China University of Science and Technology, Shanghai 200237, China, ²Institute of Pharmaceutical Biotechnology and Department of Gastroenterology of the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China, ³Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, College of Life Sciences, Huzhou University, Huzhou 313000, China, ⁴State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China, ⁵Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, ⁶State Key Laboratory of Microbial Metabolism and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China and ⁷College of Life Sciences, Zhejiang University, Hangzhou 310058, China

Received January 03, 2022; Revised March 05, 2022; Editorial Decision March 07, 2022; Accepted March 10, 2022

ABSTRACT

Direct cloning of biosynthetic gene clusters (BGCs) from microbial genomes facilitates natural product-based drug discovery. Here, by combining Cas12a and the advanced features of bacterial artificial chromosome library construction, we developed a fast yet efficient *in vitro* platform for directly capturing large BGCs, named CAT-FISHING (CRISPR/Cas12a-mediated fast direct biosynthetic gene cluster cloning). As demonstrations, several large BGCs from different actinomycetal genomic DNA samples were efficiently captured by CAT-FISHING, the largest of which was 145 kb with 75% GC content. Furthermore, the directly cloned, 110 kb long, cryptic polyketide encoding BGC from *Micromonospora* sp. 181 was then heterologously expressed in a *Streptomyces* chassis. It turned out to be a new macrolactam compound, marinolactam A, which showed promising anticancer activity. Our results indicate that CAT-FISHING is a powerful method for complicated BGC cloning, and we believe that it would be an important asset to the entire community of natural product-based drug discovery.

INTRODUCTION

Microorganisms, particularly actinobacteria, remain unrivalled in their ability to produce bioactive small molecules (BSMs), some of which (e.g. avermectin) (1–3) have been used in clinics without any structural modifications. This capability is a testament to the remarkable potential of actinobacteria to produce novel drugs. The sequencing of actinobacteria (e.g. *Streptomyces*) genomes has deciphered a large unexploited pool of novel biosynthetic gene clusters (BGCs), which may encode biosynthetic machinery producing new BSMs. Unfortunately, most of these BGCs are cryptic (4,5). How to efficiently clone and heterogeneously express these BGCs is the key to natural product (NP)-based drug discovery (6). However, the cloning of BGCs from actinobacteria is difficult and time consuming because of their high GC content and large size. More than 10% (196/1910) of the characterized BGCs are >80 kb, and 40% (756/1910) have >70% GC content. This percentage is much higher in BGCs from *Streptomyces*, reaching 84% (534/634) (Supplementary Scheme S1). To date, multiple technologies have been developed for BGC cloning, such as genomic library construction (i.e. cosmid or fosmid libraries), PCR-based methods, RecET/Red $\alpha\beta$ -based cloning and Gibson assembly (Supplementary Table S1). Additionally, the emergence of the CRISPR/Cas technique has enabled several new DNA cloning methods such as Cas9 combined with an *in vitro* λ packaging system (7), ExoCET (Exo nuclease combined with RecET recombina-

*To whom correspondence should be addressed. Tel: +86 21 64253020; Fax: +86 21 64253020; Email: tangy@ecust.edu.cn

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

tion) (8), CATCH (Cas9-assisted targeting of chromosomal segments) (9,10) and CAPTURE (Cas12a-assisted precise targeted cloning using *in vivo* Cre-lox recombination) (11). However, challenges may still remain regarding the relatively low efficiency and the time-consuming nature of large (e.g. >120 kb) and high-GC (e.g. >70%) BGC cloning. More efficient methods that can be used generally during different workflows are therefore needed to enable accessible production of natural BSMs.

CRISPR/Cas12a is a single RNA-guided (crRNA) endonuclease belonging to the class II CRISPR/Cas system (12). Unlike Cas9, Cas12a recognizes a T-rich protospacer adjacent motif (PAM) instead of a G-rich PAM, and it generates dsDNA breaks with staggered ends instead of blunt ends. Beyond applications in genome editing (13), CRISPR/Cas12a has been widely used in applications including nucleic acid-based diagnosis (14,15) and small-molecule (or protein) detection (16–18). Moreover, CRISPR/Cas12a is superior in DNA assembly because of its programmable endonuclease activity and the generated sticky ends with 4- or 5-nt overhangs (12). On the basis of these features, Li *et al.* have developed a Cas12a-based DNA assembly method, named C-Brick (19). Further studies have revealed that the accuracy of Cas12a cleavage is significantly improved by shortening the length of the crRNA; on this basis, a more efficient DNA assembly method, Cpf1-assisted cutting and Taq DNA ligase-assisted ligation, was developed (20). These previous studies encouraged us to develop a new and better Cas12a-based direct large BGC cloning approach that can be efficiently applied in genome mining of novel cryptic secondary metabolites.

The classical, widely used bacterial artificial chromosome (BAC) library construction technique is indiscriminate in the DNA sequence of insertion, thus making BAC library construction particularly suitable for the cloning of large DNA fragments with high GC content. However, the procedure is time consuming, labor intensive and technically demanding (21,22). In this study, by combining the DNA cleavage activity of Cas12a and the unique advantages from BAC library construction, we developed a Cas12a-derived DNA cloning method, called CAT-FISHING (CRISPR/Cas12a-mediated fast direct biosynthetic gene cluster cloning). To thoroughly evaluate the capability of CAT-FISHING, we rapidly cloned various lengths of BGCs containing DNA fragments (up to 145.7 kb) from different DNA samples with high GC content (>70%) by using CAT-FISHING. Moreover, a captured cryptic large polyketide synthase (PKS) BGC from *Micromonospora* sp. 181 was successfully expressed in the *Streptomyces albus* J1074-derived cluster-free chassis strain (23), and a novel macrolactam compound, marinolactam A, with moderate anticancer activity was identified and characterized.

MATERIALS AND METHODS

Strains, plasmids and media

The strains and plasmids used in this work are given in Supplementary Tables S2 and S3. *Escherichia coli* and its derivatives were cultivated on Luria–Bertani agar plates (tryptone 10 g/l, yeast extract 5 g/l and NaCl 10 g/l, pH

7.2). *Streptomyces* and its derivatives were cultivated on soybean flour–mannitol agar plates (soybean flour 20 g/l, mannitol 20 g/l and agar 20 g/l, pH 7.2) or ISP4 (International *Streptomyces* Project Medium 4, BD Biosciences, San Jose, CA). In the *Streptomyces* fermentation experiments, seeds grown in TSB (trypticase soy broth, Oxoid Ltd), R4 medium (0.5% glucose, 0.1% yeast extract, 0.15% proline, 0.12% valine, 0.28% TES [*N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid], 0.5% MgCl₂·6H₂O, 0.2% CaCl₂·2H₂O, 50 mg/l casamino acid, 100 mg/l K₂SO₄ and 2 ml (per liter) trace element solution, which contains 200 mg/l FeCl₃·6H₂O, 40 mg/l ZnCl₂, 10 mg/l MnCl₂·4H₂O, 10 mg/l CuCl₂·2H₂O, 10 mg/l Na₂B₄O₇·10H₂O and 10 mg/l (NH₄)₆Mo₇O₂₄·4H₂O) or VER medium (1% glucose, 1% soluble starch, 0.5% peptone, 0.2% yeast extract, 1% glycerol, 0.25% corn starch and 0.2% CaCO₃) were used for subsequent fermentations (in a 250-ml Erlenmeyer flask, 30°C and 200 rpm).

Capture plasmid construction and genomic DNA isolation

The primers for capture plasmid construction are listed in Supplementary Table S4. The capture plasmid was constructed by introducing the *lacZ* gene as well as two PCR-amplified homology arms (each arm containing at least one PAM site, ≥30 bp) corresponding to the flanking regions of the target DNA fragment (or BGC) in pBAC2015 (24). Assembly of multiple DNA fragments was carried out by the EZmax one-step seamless cloning kit (Tolo Biotechnology, Shanghai). Briefly, 100 ng of plasmid backbone DNA amplified from pBAC2015, 50 ng of each homology arm DNA amplified from the flanking region of the target DNA fragment, 50 ng of *lacZ* cassette DNA amplified from pUC57, 2 μl of buffer and 2 μl of recombinase were mixed in a 20-μl reaction system. The mixture was incubated at 37°C for 0.5 h and then directly used for transformation. Plasmid DNA was isolated from *E. coli* using the alkaline lysis protocol (25). Alternatively, if the selected homology arm contains only one PAM site, two 30-bp (4-nt PAM site, 26-nt target recognition and cleavage sequence) homology arms were also capable for BGC cloning. The linearized capture plasmid could easily be obtained by one-step PCR with the homology arm incorporated primers using pBAC2015 as a template. The resulting PCR product could be directly used in the following step of Cas12a digestion.

For genomic DNA isolation, target *Streptomyces* or *Actinobacteria* strain was cultured in Oxoid TSB (30 g/l) supplemented with glycine (5 g/l) or sucrose (10.3%). According to *Practical Streptomyces Genetics* (25), after cultivation at 200 rpm and 30°C for 24–48 h, mycelium was collected by centrifugation (4°C, 4000 × *g*, 5 min). Mycelium was resuspended in TE25S (25 mM Tris–HCl, pH 8, 25 mM EDTA, pH 8, 0.3 M sucrose) and then the supernatant was removed (4°C, 4000 × *g*, 5 min). The mycelium density was adjusted with TE25S (to the OD₆₀₀ of 1.9–2.0) and it was mixed with an equal volume of 1.0% LMP agarose (1.0% molten solution of low melting point agarose) at 50°C and then poured into holes in a plug mold (100-μl holes). The blocks were removed from the mold and incubated at 37°C for 1 h in lysozyme solution (2 mg/ml in TE25S). The lysozyme solution was removed and the blocks were incubated at 50°C

for 2 h in proteinase K solution (1 mg/ml proteinase K in NDS; NDS: 0.5 M EDTA, pH 8, 10 mM Tris-HCl, pH 8, 1% SDS). The proteinase K solution was removed and the blocks were washed once for 30 min in TE (10 mM Tris-HCl, pH 8, 1 mM EDTA, pH 8) supplemented with 0.1 mM PMSF (phenylmethanesulfonyl fluoride, serine proteinase inhibitor) and then three times with TE for 30 min. After removing all of the liquid, the agarose plugs could be used for CRISPR/Cas12a digestion, but could also be stored for up to 1 month at 4°C in 70% ethanol.

crRNA preparation, protein purification and Cas12a-based DNA restriction and ligation

The oligonucleotides used as templates for crRNA transcription are given in Supplementary Table S5. According to our previous study (16), crRNA was prepared via *in vitro* transcription. Templates for crRNA synthesis were synthesized and annealed by using Taq DNA Polymerase PCR Buffer (Thermo Fisher Scientific). A HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB) was used for crRNA *in vitro* transcription. The resulting crRNA was purified using RNA Clean & Concentrator™-5 kit (Zymo Research), and subsequently quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). RNase-free materials (Axygen Scientific, Union City, CA) and conditions were applied during the entire experimental process.

The Cas12a (LbCas12a) protein used in this study was overexpressed in pET28a in *E. coli* BL21(DE3). According to our previous study (16), for Cas12a purification, *E. coli* cells were resuspended in lysis buffer (200 mM sodium chloride, 2 mM dithiothreitol, 5% glycerol and 50 mM Tris-HCl, pH 7.4) supplemented with protease inhibitors (Roche, EDTA-free) and then lysed by French press. The supernatant was collected by centrifugation at 12 000 rpm for 40 min at 4°C. The sample was then loaded into a His-Trap FF column (GE Healthcare) and washed by washing buffer with different concentrations of imidazole. The resulting peak fraction eluted was collected followed by desalting with dialysis. After loading into a HiTrap Q HP column (GE Healthcare), the peak fraction was collected and concentrated. The concentrated solution was further purified by gel filtration chromatography using a HiLoad 16/600 Superdex 200 pg column combining with the AKTA system (FPLC, AKTA Explorer 100, GE Healthcare).

In the CRISPR/Cas12a cutting system, NEBuffer™ 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 100 µg/ml bovine serum albumin, pH 7.9) was adopted as the reaction buffer. For pBAC-ZL or capture plasmid cleavage, 2 µl of 10× NEBuffer™ 3.1 and 200 ng of plasmid DNA were incubated with 300 nM of Cas12a protein and 300 nM of corresponding crRNA pairs at 37°C for 1 h. After the reaction, the linearized capture plasmids or DNA fragments of pBAC-ZL were prepared with isopropanol and ethanol (24). The resulting linear capture plasmid or DNA fragments of pBAC-ZL could be used for the following ligation. If necessary, the large DNA fragments could be analyzed by pulsed field gel electrophoresis (PFGE) with the CHEF-DR III apparatus (Bio-Rad, Richmond, CA). PFGE was performed in 0.5% agarose at 6 V/cm with a 1–25 s switching pulse time for 16–18 h in 0.5× TBE buffer. For genomic

DNA cleavage, plugs were initially equilibrated in 1× NEBuffer™ 3.1, then transferred into a cleavage system that contained the Cas12a protein and the corresponding crRNA pair, and finally incubated at 37°C for 1–2 h. After the reaction, following heat treatment at 65°C for 10 min to inactivate Cas12a protein, the LMP agarose gel was hydrolyzed using β-agarase I (NEB) for 30 min at 42°C. Afterward, the resulting DNA mixture could be directly used for following the ligation with the corresponding linear capture vectors by DNA ligase (NEB).

Electroporation of *E. coli*

After the transfer of ligation samples into 0.1 M glucose/1% agarose gel to desalt for 1–2 h on ice, these samples could be used for electroporation. The high-efficiency electroporation of *E. coli* cells was accomplished according to a previous study (26). The following electroporations were performed in 2-mm cuvettes using the Bio-Rad GenePulser Xcell™ system (electroporation conditions: 2500 V, 200 Ω and 25 µF). Then, to the *E. coli* cells in the cuvette, 1 ml of SOC medium (tryptone 20 g/l, yeast extract 5 g/l, NaCl 0.5 g/l, KCl 2.5 mM, MgCl₂ 10 mM, glucose 20 mM) was added and the mixture was transferred into a 15-ml Falcon™ tube. After shaking at 200 rpm for 1 h at 37°C, the strains were collected and spread on selective Luria-Bertani agar plates. The plates were incubated overnight at 37°C, and the transformants were screened and verified by PCR using the primers listed in Supplementary Table S6.

Heterologous expression of BGCs in *Streptomyces* and LC-MS analysis

The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette amplified from pSET152 was introduced into BAC plasmids by Red/ET recombination (27). The resulting plasmids were introduced into *S. albus* Del14 by triparental conjugation according to a previous study (22). The transconjugants were screened and verified by PCR with the primers listed in Supplementary Table S6. R4 medium was used for *S. albus* Del-87kb fermentation, and the production of surugamide was qualitatively analyzed with a high-resolution (HR) mass spectrometer according to a previous study (28). VER medium was used for *S. albus* Del-110kb fermentation. The 96-h fermentation broth was harvested by centrifugation, and the resultant supernatant was filtered through a 0.22-µm membrane, loaded on a pre-packed C18 column (Phenomenex, 50 µm, 65 Å, 10 g) and eluted with 5%, 30% and 100% MeCN in water in a stepwise manner. The 30% fraction containing marinolactam A was dried *in vacuo*, resuspended in MeOH and further analyzed with HR-LC-MS. LC-MS was performed on a QTOF instrument (Agilent 6546 Accurate Mass QTOF LC-MS consisting of a 1290 Infinity Series HPLC system). Samples were resolved on a Luna Omega C18 column (Phenomenex, 1.6 µm, 30 mm × 2.1 mm). The mobile phase consisted of water and MeCN (+0.1% formic acid). Elution for orfamide was performed isocratically with 5% MeCN in water (1 min) followed by a gradient of 5–100% MeCN over 6 min, and then isocratic elution with 100% MeCN over 3 min at a flow rate of 0.3 ml/min.

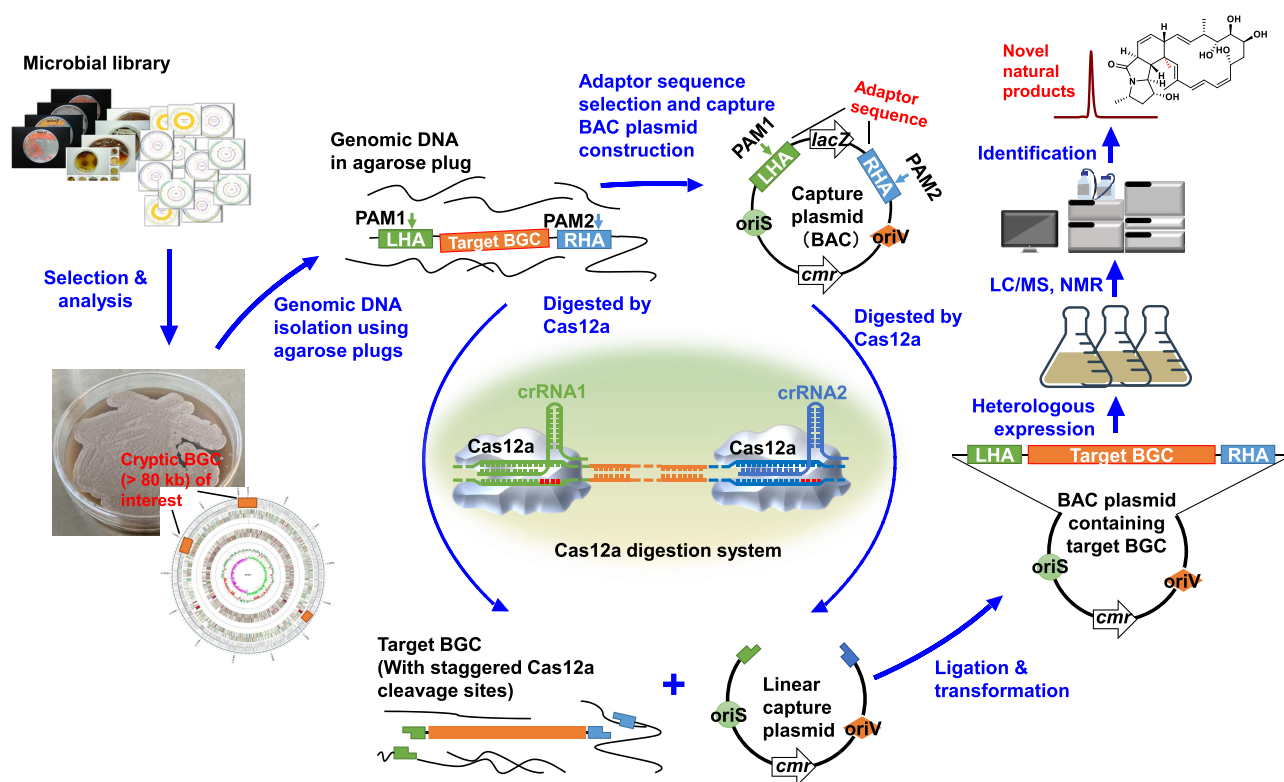


Figure 1. Schematic diagram of CAT-FISHING to capture cryptic BGCs for NP discovery. LHA: left homology arm; RHA: right homology arm; BGC: biosynthetic gene cluster; BAC: bacterial artificial chromosome; oriS and oriV are two origins of replication in the BAC plasmid; cmr: chloramphenicol resistance gene; crNA: CRISPR RNA; PAM: protospacer adjacent motif; LC/MS: liquid chromatography–tandem mass spectrometry; NMR: nuclear magnetic resonance.

Large-scale growth of *S. albus* Del14-110kb and purification of marinolactam A

Large-scale fermentation was carried out following a similar procedure to that for small-scale fermentation described earlier. Three-day seed cultures were prepared as described earlier. They were then used to inoculate several 2-l Erlenmeyer flasks containing 200 ml of VER medium with isolated mycelia to a final concentration of 0.01% (w/v). About 20 l of total culture was used for compound isolation. The flasks were incubated at 30°C/220 rpm for 7 days, at which point the desired products were purified (see later).

Marinolactam A was purified from 20 l fermentations of *S. albus* Del14-110kb. Four to six days after fermentation, the resulting supernatant was loaded on a pre-packed C18 column (Phenomenex, 50 μ m, 65 Å, 20 g) and eluted with 5%, 30% and 100% MeCN in water in a stepwise manner. The 30% fraction containing marinolactam A was dried *in vacuo*, resuspended in 30% MeCN and further purified on an Elite prep HPLC P3500 using a Sinochrom ODS-BP C18 column (Sinochrom ODS-BP, 10 μ m, 20 mm \times 250 mm) operating at a flow rate of 12 ml/min with mobile phases consisting of water and MeCN (+0.1% formic acid). Upon injection, elution was carried out isocratically with 5% MeCN in water for 2 min, followed by a gradient of 5–50% MeCN in water over 20 min, followed by a gradient of 50–100% MeCN over 5 min. Fractions were collected in 1 min intervals over the time range of 5–25 min. Peaks

containing marinolactam A, as judged by HR-HPLC–MS analysis, were pooled, dried *in vacuo*, resuspended in 30% MeCN and further purified on a semipreparative/analytical Thermo Fisher UltiMate 300 DGLC HPLC system. The sample was applied to an RP Amide-C16 column (Supelco, 5 μ m, 10 mm \times 250 mm) operating at a flow rate of 2.5 ml/min with the same mobile phases as mentioned earlier and a gradient of 10–55% MeCN in water over 30 min, followed by a gradient of 55–100% MeCN over 5 min. Peaks containing pure marinolactam A were combined and lyophilized to dryness. This procedure gave 0.8 mg of marinolactam A.

Electronic circular dichroism calculation

Conformational searches were run by employing the ‘systematic’ procedure implemented in Spartan’14 using the Merck molecular force field. All Merck molecular force field minima were reoptimized with density functional theory calculations at the B3LYP/6-31+G(d) level using the Gaussian09 program. The geometry was optimized starting from various initial conformations, with vibrational frequency calculations confirming the presence of minima. All the conformers were further optimized at the B3LYP/6-31+G(d) level to give one lowest-energy conformer for marinolactam A, which was subjected to time-dependent density functional theory calculations at the B3LYP/6-31+G(d) level in methanol with the polarizable continuum

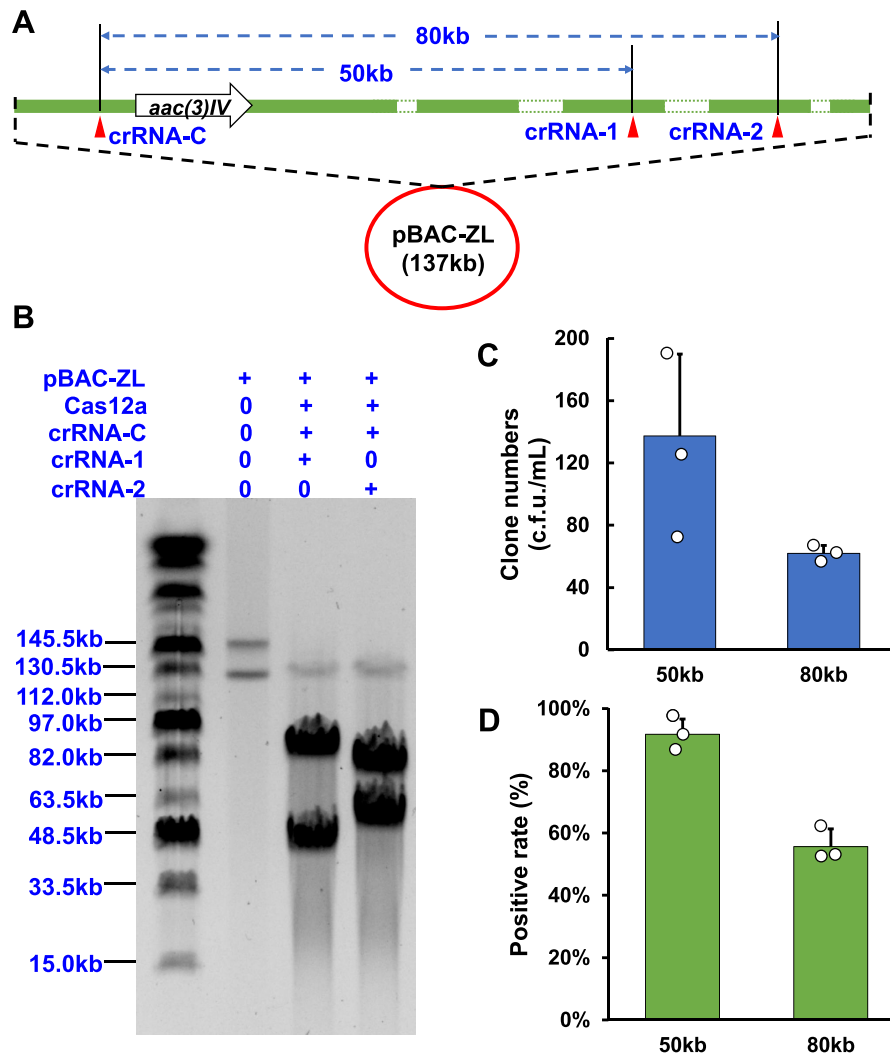


Figure 2. Direct cloning of a 50-kb and an 80-kb DNA fragment from pBAC-ZL through a Cas12a-based cloning strategy. (A) Two different target segments with different lengths (50 and 80 kb) in the BAC plasmid. (B) Analysis of Cas12a-digested BAC plasmid (pBAC-ZL) by PFGE. PFGE was performed with 0.5% agarose at 6 V/cm with a 1–25 s switching pulse time for 16–18 h in 0.5× TBE buffer. MidRange PFG Marker (NEB Biolabs) was used as a molecular size marker. (C, D) Determination of the clone numbers and positive rates for the DNA fragments of different lengths in three independent experiments/plates. The data shown are mean \pm SD.

model. Electronic circular dichroism (ECD) spectra were generated using the program SpecDis by applying a Gaussian band shape with 0.2 width for marinolactam A, from dipole length rotational strengths. The dipole velocity forms yielded negligible differences. The spectra of the conformers were combined using Boltzmann weighting, with the lowest-energy conformations accounting for 100% of the weights, respectively.

Antibiotic and cytotoxic assays

Experimental details are described in the Supplementary Data.

Data analysis

Data analysis and process were carried out using Microsoft Excel 2016. The data in figures were expressed as mean \pm standard deviation (SD).

RESULTS AND DISCUSSION

The CAT-FISHING workflow for NP discovery

The recent development of whole genome sequencing and computational biology has provided new opportunities for NP discovery from organisms with abundant cryptic BGCs, such as *Streptomyces*, *Myxobacteria*, rare actinomycetes, plant endosymbionts, and anaerobes (29). In this study, to clone and manipulate large BGCs from rich microbial genome sources, we used Cas12a together with specific crRNA pairs to precisely cut the genomic DNA, instead of using the restriction enzymes to partially digest the genomic DNA in BAC library construction. The application of Cas12a in CAT-FISHING theoretically enables cloning of large BGCs with high GC content in a simpler and more direct procedure compared to BAC library-based cloning techniques. The flowchart of using CAT-FISHING to capture novel BGCs for NP discovery is presented in Figure 1.

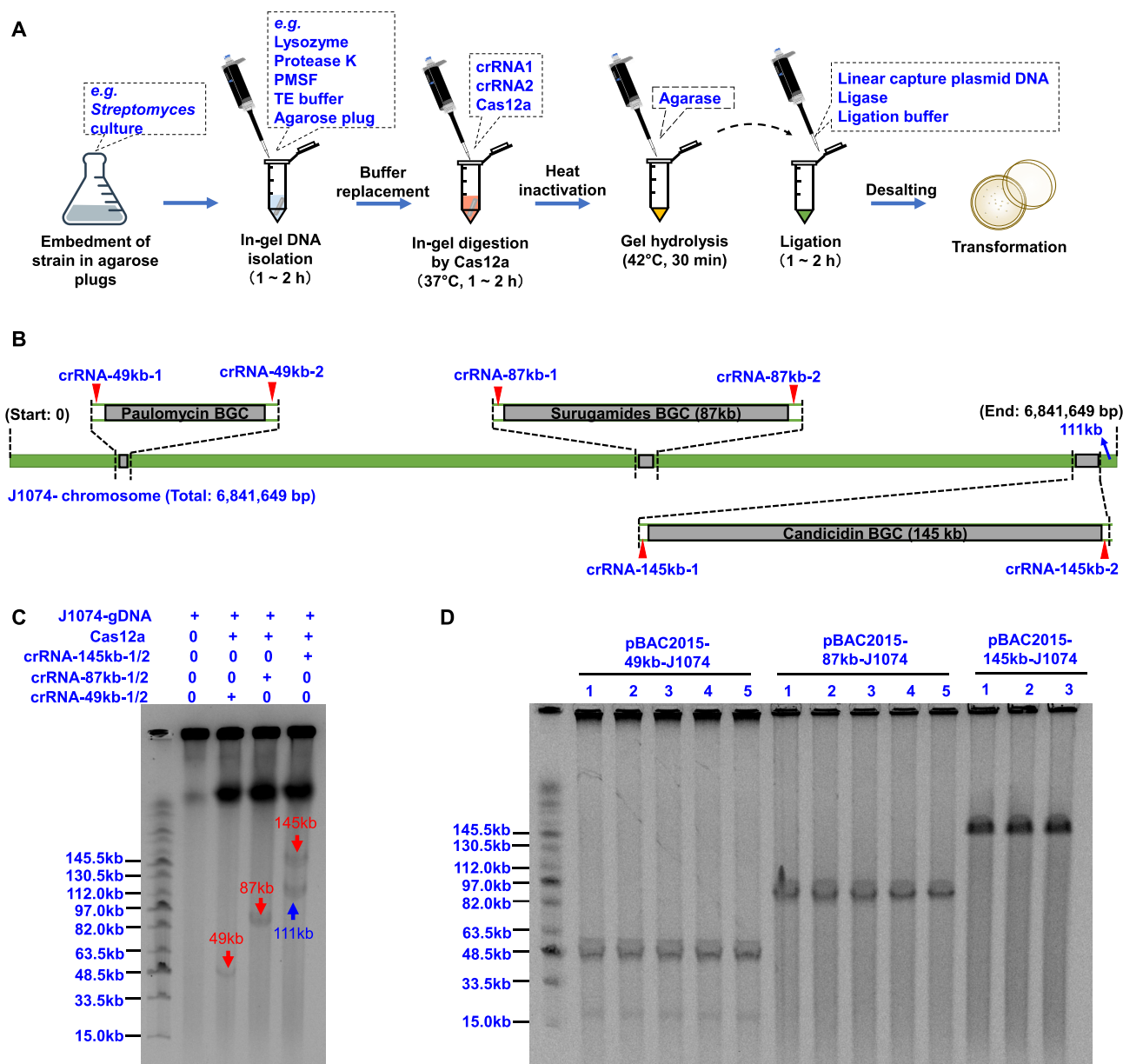


Figure 3. Direct cloning of large BGCs from genomic DNA by CAT-FISHING. **(A)** Workflow of cloning large BGCs from genomic DNA by CAT-FISHING. **(B)** Three different target BGCs of different lengths (49, 87 and 145 kb) in the chromosome DNA of *S. albus* J1074. **(C)** Analysis of CRISPR/Cas12a-digested genomic DNA of *S. albus* J1074 by PFGE. Three pairs of crRNAs were used for genomic DNA digestion to obtain the resulting target DNA fragment. Bands of target BGC are indicated by red arrows. **(D)** Validation of positive clones containing a paulomycin (pBAC2015-49kb-J1074), surugamide (pBAC2015-87kb-J1074) or candidicin (pBAC2015-145kb-J1074) gene cluster, on the basis of PFGE. PFGE was performed with 0.5% agarose at 6 V/cm with a 1–25 s switching pulse time for 16–18 h in 0.5× TBE buffer. MidRange PFG Marker (NEB Biolabs) was used as a molecular size marker.

With CAT-FISHING, a target BGC identified *in silico*, which encodes biosynthetic machineries producing bioactive NP, can easily be captured in three steps. As shown in Figure 1, the first step was to construct the capture plasmid and linearize it with Cas12a. In this step, two homology arms (each containing at least one PAM site, or ≥ 30 bp) that flank the target BGC were selected as adapter sequences and amplified by PCR. Then, the BAC plasmid containing the two adapter sequences and selection marker (e.g. antibiotic resistance gene, *lacZ* or *gfp*), designated as the capture plasmid, was constructed. Under the guidance

of crRNAs, the two selected PAM motif regions on two adapters were simultaneously digested by Cas12a, thus resulting in the linear capture plasmid. The second step was to isolate digested genomic DNA with Cas12a. According to the BAC library construction protocol (30), genomic DNA plugs from the target strain were prepared and digested by Cas12a with the two designed crRNAs previously used in step I. Third, the resultant linear capture plasmid and the digested genomic DNA from steps I and II were mixed and ligated with DNA ligase. Then, the ligation product was subsequently introduced into *E. coli* by electropora-

Table 1. The cloning efficiencies of target BGCs from different actinomycetal genomic DNA samples

Strains and target NPs	Genome features		Target BGC features			Cloning performance				
						Route 1 (simple and fast)		Route 2 (high efficiency) ^a		BGC integrity ^d
	Size	(G + C)%	Type	Length	(G + C)%	Correct/checked ^b	Efficiency	Correct/checked ^c	Efficiency	
<i>K. aureofaciens</i> DM-1 Demeclocycline	7.1 Mb	73%	PKS	41 kb	73%	11/20	~55%	NA	NA	100%
<i>Sa. spinosa</i> NHF132 Spinosad	8.9 Mb	69%	PKS	80 kb	67%	6/20	~30%	NA	NA	100%
<i>Sa. pogona</i> NRRL 30141 Butenyl-spinosyn	9.6 Mb	68%	PKS	79 kb	66%	5/20	~25%	NA	NA	100%
<i>S. avermitilis</i> ATCC 31267 Avermectin	10.5 Mb	71%	PKS	80 kb	69%	5/20	~25%	NA	NA	100%
<i>S. albus</i> J1074 Paulomycin	6.8 Mb	73%	Quinone glycoside	49 kb	71%	8/20, 5/20, 7/20	32 ± 7%	7/10	~70%	100%
				87 kb	76%	4/20, 4/20, 5/20	21 ± 3%	7/10	~70%	100%
				145 kb	75%	2/20, 1/20, 2/24	8 ± 2%	5/10	~50%	100%
<i>Micromonospora</i> sp. 181 Marinolactam A	7.6 Mb	73%	PKS	110 kb	72%	5/20	~25%	NA	NA	100%

^aRoute 2: target DNA fragment isolation and purification with PFGE (Supplementary Figure S10).

^bA total of 20 or 24 clones were randomly selected from each agar plate for PCR screening.

^cA total of 10 clones were randomly selected from agar plate for PCR screening.

^dConfirmed by digestion or PFGE of three to five PCR-validated clones.

tion. The successfully captured BGCs were then identified through PCR screening. Finally, the novel NPs encoded by the captured cryptic BGCs were obtained through heterologous expression in a suitable host, and further isolated and identified.

Evaluation of Cas12a-based DNA cloning efficiency by using a 137-kb BAC plasmid

Like that of BAC library construction, the principle underlying CAT-FISHING involves DNA ligase-mediated cohesive end ligation of a linear BAC plasmid and a large BGC containing DNA fragments (21). A previous study has reported that the cleavage specificity of Cas12a can be enhanced by using a shorter crRNA spacer (i.e. 17–19 nt) (20). In this study, an 18-nt spacer length was used. To demonstrate the Cas12a-based cloning strategy in a simplified system, we used a 137-kb BAC plasmid, pBAC-ZL (68% GC), to evaluate its cloning utility for a large DNA fragment. As shown in Figure 2A, a 50-kb fragment and an 80-kb fragment were obtained by using the corresponding crRNA-guided Cas12a cleavage. Under the guidance of the corresponding crRNA pairs, the BAC plasmid pBAC-ZL was digested with Cas12a, and 50- and 80-kb target bands were observed on an agarose gel after PFGE (Figure 2B). With the corresponding capture plasmid, two target DNA fragments were also successfully cloned from the plasmid pBAC-ZL, as shown in Figure 2C and D and Supplementary Figures S1 and S2, respectively. According to the obtained results, 94% of the 50-kb insertion-containing transformants were correct clones. Among the 80-kb insertion clones, ~50% of the transformants were correct. These results indicated that the simplified system of Cas12a-based cloning achieved high cloning rates of capture of the 50-

and 80-kb DNA fragments from the purified BAC plasmid DNA sample.

Direct cloning of targeted BGCs from different actinomycetal genomic DNA samples by CAT-FISHING

During BAC library construction, PFGE is an essential, powerful tool for separating large DNA fragments. However, PFGE is time consuming and usually requires >16 h to separate particular sizes of DNA fragments (31). To clone large BGCs from genomic DNA by CAT-FISHING directly, we developed a PFGE-free procedure (Figure 3A and Supplementary Figure S3). After genomic DNA isolation in agarose plugs and in-gel Cas12a digestion, the resultant sample containing a mixture of genomic DNA could be directly used for subsequent ligation and transformation without the PFGE-mediated isolation step. In this study, a 49-kb paulomycin BGC (GC content 71%; quinone glycoside), an 87-kb surugamide BGC (GC content 76%; nonribosomal peptide synthetase, NRPS) and a 145-kb candidin BGC (GC content 75%; PKS) from *S. albus* J1074 (28,32,33) were selected to validate this method (Table 1). After optimization of the digestion conditions, the genomic DNA of *S. albus* J1074 was digested with Cas12a, and 49-, 87- and 145-kb target bands were observed under the guidance of the corresponding crRNA pairs. Moreover, the predicted extra 111-kb DNA fragment resulting from the digestion of J1074 chromosome by Cas12a under the guidance of crRNA-145kb-2 was also observed (Figure 3B and C). Here, lysis of streptomycete mycelium within low melting point agarose plugs and subsequent in-gel Cas12a digestion was used to largely avoid mechanical shearing (e.g. pipetting and vortexing), thus resulting in a high-quality genomic DNA sample (34). The efficient and specific digestion of

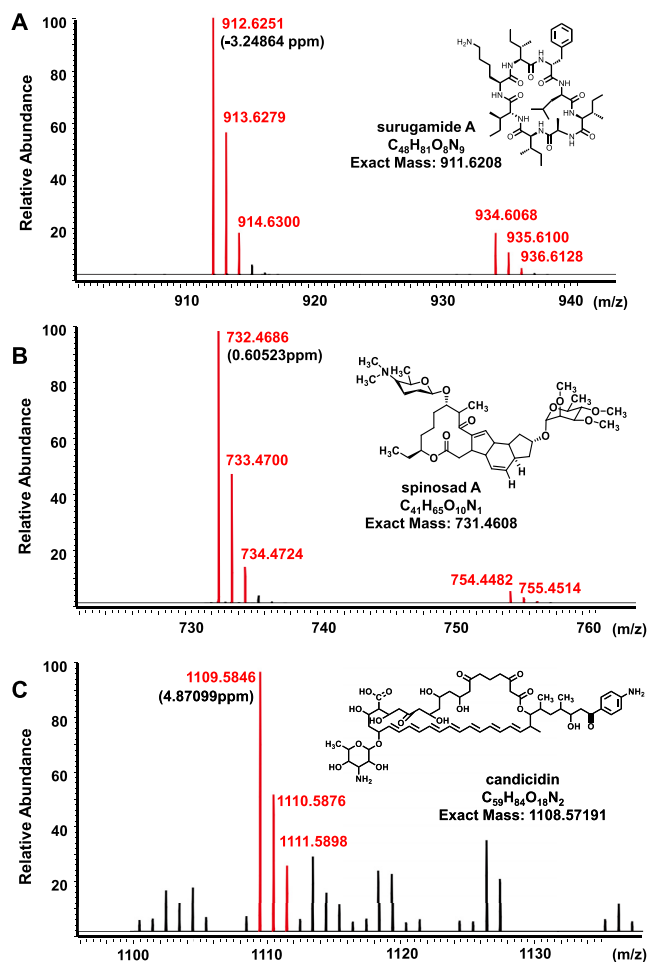


Figure 4. Heterologous biosynthesis of surugamide, spinosad and candicidin in *S. albus* Del14. Identification of surugamide, spinosad and candicidin in the fermentation broth of *S. albus* Del14-87kb (A), *S. albus* Del14-81kb (B) and *S. albus* Del14-145kb (C), respectively, by HR-MS. All the accurate mass measurements for impurities were within <5 ppm of the theoretical m/z value.

high-quality genomic DNA ensured the successful cloning of large BGCs by CAT-FISHING.

After the procedures described in Figure 3A, three different primer pairs for regions in the middle and in the left/right boundary of each target BGC were used for subsequent screening of the correct clones. We found that three BGCs from J1074 were successfully captured by CAT-FISHING, and the results were confirmed via PFGE analysis and restriction mapping (Figure 3D and Supplementary Figure S4). The ratios of correct clones containing 49-, 87- and 145-kb BGCs were ~32%, ~21% and ~8%, respectively (Table 1). When different NEB DNA ligases (e.g. T4, *E. coli* and Taq) were used, the cloning efficiency of CAT-FISHING showed significant differences. *Escherichia coli* DNA ligase had the best cloning performance. Compared with our previous results obtained with T4 DNA ligase (35), the cloning efficiency of CAT-FISHING for large BGCs was dramatically improved by using the *E. coli* DNA ligase (Supplementary Figure S5). To illustrate the general application of CAT-FISHING in other kinds of actinobac-

teria, *Kitasatospora aureofaciens* DM-1, *Saccharopolyspora spinosa* NHF132, *Sa. pogona* NRRL 30141 and *S. avermitilis* ATCC 31267 as the producer of demeclocycline (36), spinosad (37), butenyl-spinosad (38) and avermectin (39), respectively, were also selected for target BGC capturing. The ratios of correct clones containing 41-kb demeclocycline BGC, 81-kb spinosad BGC, 79-kb butenyl-spinosad BGC and 80-kb avermectin BGC were ~55%, ~30%, ~25% and ~25%, respectively (Table 1 and Supplementary Figures S6–S9). These results indicated that CAT-FISHING enables accessible capturing of target BGCs from different actinomycetal genomic DNA samples.

In contrast to the BAC library, in which only several correct clones can often be screened out of thousands of clones (i.e. 1/1000) (22,40), CAT-FISHING, by applying Cas12a, is simpler, more convenient and efficient in direct cloning large and high GC content BGCs. Only 3–4 days are needed to capture a large BGC with CAT-FISHING (Supplementary Figure S3). To further improve the cloning efficiency of CAT-FISHING, one more cloning step is added to reduce the false-positive clones. By analysis of the incorrect clones (data not shown), it was found that due to the complexity of the unpurified genomic DNA mixture sample, many short DNA fragments were inserted into capture plasmids. Therefore, through target DNA fragment isolation and purification by PFGE (Supplementary Figure S10, route 2), the cloning performance toward target BGCs has been dramatically enhanced (Table 1), which improves the accuracy rate close to that of the CAPTURE method (11). In addition, a 145-kb DNA fragment with 75% GC content was successfully captured by CAT-FISHING in this study. To our knowledge, this is the largest DNA fragment ever obtained by *in vitro* direct simple cloning from such high GC genomic DNA.

Activation of the cryptic BGC captured by CAT-FISHING in *S. albus* Del14

Off-target effects in the CRISPR/Cas system can cause unwanted mutations or deletions during DNA editing (41,42). In this study, all large BGCs captured through CAT-FISHING were carefully verified by restriction mapping, which indicated 100% BGC integrity (Table 1). To thoroughly validate the sequence and functional integrity of the BGCs captured by CAT-FISHING, as well as to demonstrate the route of BGC capture and heterologous expression, we used 81-, 87- and 145-kb BGCs, the captured gene clusters of spinosad, surugamide and candicidin, respectively, were transformed and expressed in a cluster-free *Streptomyces* chassis strain *S. albus* Del14, a widely used *Streptomyces* chassis for genome mining (23). The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette was introduced into the targeted plasmid through Red/ET recombination to generate the final expression vector (22). Through ET12567/pUC307-mediated triparental conjugation (43), the resultant plasmid was integrated into the chromosome of *S. albus* Del14, thus generating *S. albus* Del14-81kb, *S. albus* Del14-87kb and *S. albus* Del14-145kb (Supplementary Figure S11). The subsequent fermentation experiments revealed surugamide A in *S. albus* Del14-87kb, as compared with *S. albus* Del14 and *S. albus* J1074 (Sup-

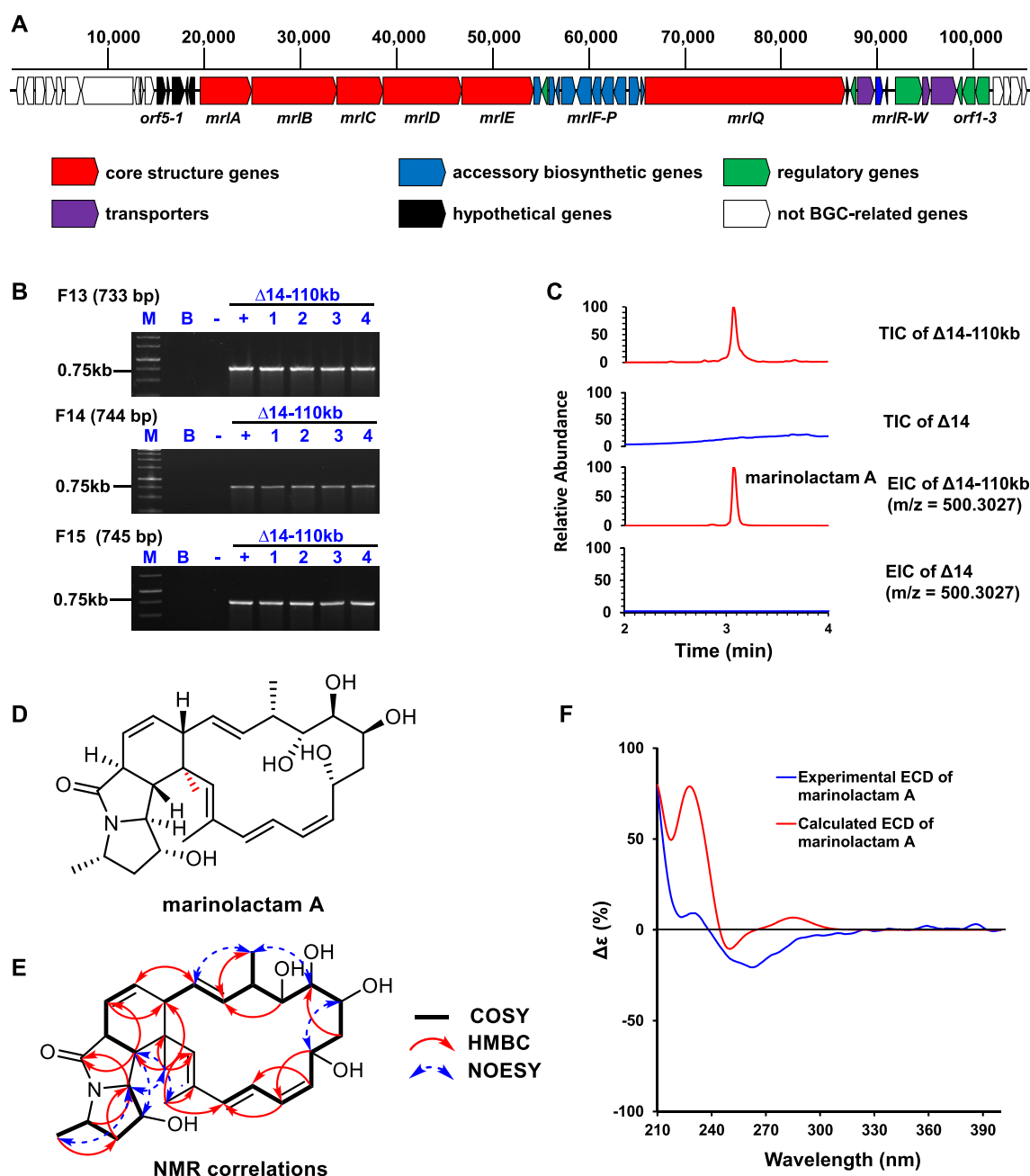


Figure 5. Cloning and heterologous expression of a 110-kb polyketide BGC from *Micromonospora* sp. 181. **(A)** The predicted 110-kb marinolactam gene cluster, determined by antiSMASH analysis. **(B)** PCR verification of *S. albus* Del14-110kb ($\Delta 14-110\text{kb}$). Five transformants were randomly selected. F13 (733 bp), F14 (744 bp) and F15 (745 bp) are the PCR products amplified with 110-scr-up-F/R, 110-scr-middle-F/R and 110-scr-down-F/R, respectively. ‘-’ represents control genomic DNA of *S. albus* Del14, used as a PCR template. ‘+’ represents positive control genomic DNA of *Micromonospora* sp. 181, used as a PCR template. **(C)** Detection of NPs by LC-MS in *S. albus* Del14-110kb ($[M + H]^+ = 500.3027$ Da). **(D)** The structure of marinolactam A. A unique quaternary carbon chiral center at the C20 position was formed; the extra C28 methylation is highlighted in red (the detailed structure information can be found in Supplementary Table S8). **(E)** NMR correlations of marinolactams. **(F)** Calculated (red) and experimental ECD (blue) spectra for marinolactam A in MeOH.

Table 2. Comparison of selected methods for large DNA fragment (>100 kb) cloning

Methods	ULCC (high GC) ^a	CE	TC	Cost	Reference
BAC library	>120 kb	~1/1000	>1 month	>\$5000 ^b	(21)
CATCH	36 kb ^c	>90%	3–4 days	NA	(10)
ExoCET	106 kb	1/24	1 week	NA	(8)
CAPTURE	113 kb	100%	3–4 days	NA	(11)
CAT-FISHING	145 kb	50%	3–4 days	~\$40 ^d	This study

CE: cloning efficiency; TC: time consuming.

^aUpper limit of cloning capacity (ULCC) for DNA sample with high GC content (>70%).

^bPrice information for BAC library construction was provided by Wuhan Eightstars Bio-Technology Co. Ltd.

^c36-kb jadomycin BGC from *Streptomyces venezuelae* (GC content: 72.3%) and 150-kb DNA fragment from *E. coli* (GC content: 50.6%) have been cloned by CATCH.

^dSee Supplementary Tables S11 and S12.

plementary Figure S12), detected by LC–HR-MS analysis (obs. [M + H]⁺ at *m/z* 912.6252 Da, obs. [M + Na]⁺ at *m/z* 934.6068 Da) (Figure 4A). Moreover, the production of spinosad A (obs. [M + H]⁺ at *m/z* 732.4686 Da, obs. [M + Na]⁺ at *m/z* 754.4482 Da) and candicidin (obs. [M + H]⁺ at *m/z* 1109.5803 Da) in *S. albus* Del14-81kb and *S. albus* Del14-145kb, respectively, could also be detected by LC–MS (Figure 4B and C). These results indicated that the CAT-FISHING captured target BGCs could be heterologously expressed in *S. albus* Del14.

Polyene macrolactams, such as BE14106, incednine and auroramycin, are a well-known class of NPs that usually exhibit significant antifungal or antiapoptotic activities (44–46). Herein, guided by antiSMASH analysis (47), we observed a 110-kb cryptic BGC (similar to micromonolactam, salinilactam and lobosamide BGCs) (48–50) in the actinomycete *Micromonospora* sp. 181 (Figure 5A and Supplementary Table S7), which has been predicted to produce polyene macrolactams (48). To demonstrate the application of CAT-FISHING in novel NP discovery, we directly captured this 110-kb polyketide encoding BGC from the 7.6-Mb chromosome DNA of *Micromonospora* sp. 181 (Table 1). As described earlier, through triparental conjugation, a 120-kb (10-kb BAC plasmid backbone and 110-kb targeted BGC) BAC plasmid was integrated into the chromosome of *S. albus* Del14, thus generating *S. albus* Del14-110kb (Figure 5B). In the following fermentation experiment, on the basis of comparison with the results for *S. albus* Del14, a new peak with (+)-HR-MS (obs. [M + H]⁺ at *m/z* 500.3027) was observed in *S. albus* Del14-110kb, thus suggesting a molecular formula of C₂₉H₄₁NO₆ (Figure 5C and D). Interestingly, this peak cannot be detected in the fermentation sample of *Micromonospora* sp. 181, which directly confirmed the efficacy of CAT-FISHING technology in activating cryptic secondary metabolites.

After subsequent isolation, a compound named marinolactam A was obtained (Figure 5D). Analysis of 1D (¹H, ¹³C)/2D (COSY, HSQC, HMBC) NMR spectra suggested that marinolactam A contained one conjugated polyene system, two isolated alkene groups, an amide group containing pyrrolizidine ring and five oxymethine protons with a planar structure similar to that of dracolactam A (49) (Supplementary Figures S13–S15). The HSQC and HMQC data showed that the structural difference was an extra C28 methylation, thus forming a unique quaternary carbon chiral center at the C20 position. Analysis of the ¹H coupling

constant and NOESY correlations revealed that the geometries of five double bonds in marinolactam A were 3*Z*, 6*E*, 14*Z*, 16*E* and 18*E*, respectively (Supplementary Table S8). Determining the stereochemistry of chiral carbon centers in macrolactam molecules is challenging. To achieve the assignments, we first deduced the relative configurations of marinolactam A by NOESY correlations, on the basis of which the absolute configurations were then assigned as 2*S*, 5*S*, 8*S*, 9*R*, 10*S*, 11*S*, 13*R*, 20*S*, 21*R*, 22*S*, 23*R* and 25*S* through analysis of the domain sequences of PKS, and validated through ECD calculations (Supplementary Table S9 and Figure 5F) (48,51).

Analysis of marinolactam BGC and the proposed biosynthetic pathway implied that the real product of the BGC should structurally resemble lobosamide and micromonolactam (48), and the octahydro-pyrroloisindole ring is formed later via a [4 + 2]-cycloaddition (Supplementary Figure S16). Unfortunately, no polyenes were detected after UV and MS analysis through whole metabolic profiles under different fermentation conditions in both *S. albus* Del14-110kb and original host *Micromonospora* sp. 181, thus indicating that the [4 + 2]-cycloaddition occurred at very high efficiency. Finally, the antibacterial activity and cytotoxicity of marinolactam A were assessed in this study. Although no antibacterial (e.g. *Staphylococcus aureus* ATCC 29213, *E. coli* K12, *Candida albicans* ATCC 90028 and *Micrococcus luteus* ATCC 10240) activity was observed in our assays, marinolactam A displayed moderate anticancer activity against HCT116 human colon cancer cells with an IC₅₀ of 18.3 μg/ml (36.6 μM), thus indicating its potentially valuable application prospects in the field of cancer treatment (Supplementary Table S10).

In summary, this study successfully coupled the accurate DNA cleavage activity of CRISPR/Cas12a with the classic BAC library construction method of cloning large DNA fragments, which gave rise to CAT-FISHING. It offers a simple, low-cost, fast and efficient strategy for targeting the largest BGC ever obtained from high-GC genomic DNA through *in vitro* direct cloning (Table 2), and diverse or cryptic BGCs captured by CAT-FISHING have been successfully expressed in a *Streptomyces* chassis. Additionally, the production and identification of the new NP marinolactam A with anticancer activity indicated that genome mining through the application of CAT-FISHING should facilitate the discovery of novel BSMs from microbial sources.

DATA AVAILABILITY

All other data are available from the corresponding author upon reasonable request. The marinolactam A BGC and deduced protein sequence have been deposited to NCBI under accession number OM891520.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Dr Wei Ma (School of Life Sciences and Biotechnology, Shanghai Jiao Tong University) for his helpful suggestion in PFGE.

Author contributions: G.-Y.T. conceived and supervised the project. G.-Y.T., M.L., F.X. and W.W. designed the experiments. M.L., L.L., X.Z., J.L., L.K., Z.Y. and S.W. performed the cloning and molecular biology experiments. G.-Y.T., M.L., F.X., L.L., X.Z. and W.W. analyzed the data. F.X., J.Y., R.W., M.W., L.L., G.Z. and X.L. performed NP isolation, LC-MS analysis and cytotoxic assays. Y.T., W.W., L.Z. and L.B. edited the manuscript. G.-Y.T. and F.X. wrote the manuscript.

FUNDING

National Key Research and Development Program of China [2020YFA0907800, 2021YFC2100600 and 2020YFA0907304]; National Natural Science Foundation of China [31720103901, 31870040 and 21877038]; National 111 Project of China [B18022]; Fundamental Research Funds for the Central Universities [22221818014]; Shanghai Science and Technology Committee [18JC1411900]; State Key Laboratory of Bioreactor Engineering. Funding for open access charge: National Key Research and Development Program of China and National Natural Science Foundation of China. *Conflict of interest statement.* None declared.

REFERENCES

- Wang, W., Li, S., Li, Z., Zhang, J., Fan, K., Tan, G., Ai, G., Lam, S.M., Shui, G., Yang, Z. *et al.* (2020) Harnessing the intracellular triacylglycerols for titer improvement of polyketides in *Streptomyces*. *Nat. Biotechnol.*, **38**, 76–83.
- Gao, Q., Tan, G.-Y., Xia, X. and Zhang, L. (2017) Learn from microbial intelligence for avermectins overproduction. *Curr. Opin. Biotechnol.*, **48**, 251–257.
- Zhuo, Y., Zhang, W., Chen, D., Gao, H., Tao, J., Liu, M., Gou, Z., Zhou, X., Ye, B.-C., Zhang, Q. *et al.* (2010) Reverse biological engineering of *hrdB* to enhance the production of avermectins in an industrial strain of *Streptomyces avermitilis*. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 11250.
- Culp, E.J., Yim, G., Wagglechner, N., Wang, W., Pawlowski, A.C. and Wright, G.D. (2019) Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat. Biotechnol.*, **37**, 1149–1154.
- D'Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W.L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R. *et al.* (2011) Antibiotic resistance is ancient. *Nature*, **477**, 457–461.
- Tong, Y. and Deng, Z. (2020) An aurora of natural products-based drug discovery is coming. *Synth. Syst. Biotechnol.*, **5**, 92–96.
- Tao, W., Chen, L., Zhao, C., Wu, J., Yan, D., Deng, Z. and Sun, Y. (2019) *In vitro* packaging mediated one-step targeted cloning of natural product pathway. *ACS Synth. Biol.*, **8**, 1991–1997.
- Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A.F. *et al.* (2018) ExoCET: exonuclease *in vitro* assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.*, **46**, e28.
- Jiang, W. and Zhu, T.F. (2016) Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc.*, **11**, 960–975.
- Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. (2015) Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.*, **6**, 8101.
- Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S.K., Martin, T.A. and Zhao, H. (2021) Cas12a-assisted precise targeted cloning using *in vivo* Cre-lox recombination. *Nat. Commun.*, **12**, 1171.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.
- Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Sun, B., Chen, B., Xu, X., Li, Y. *et al.* (2017) CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. *Nat. Commun.*, **8**, 15179.
- Li, S.-Y., Cheng, Q.-X., Wang, J.-M., Li, X.-Y., Zhang, Z.-L., Gao, S., Cao, R.-B., Zhao, G.-P. and Wang, J. (2018) CRISPR-Cas12a-assisted nucleic acid detection. *Cell Discov.*, **4**, 20.
- Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M. and Doudna, J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, **360**, 436–439.
- Liang, M., Li, Z., Wang, W., Liu, J., Liu, L., Zhu, G., Karthik, L., Wang, M., Wang, K.-F., Wang, Z. *et al.* (2019) A CRISPR-Cas12a-derived biosensing platform for the highly sensitive detection of diverse small molecules. *Nat. Commun.*, **10**, 3672.
- Zhao, X., Li, S., Liu, G., Wang, Z., Yang, Z., Zhang, Q., Liang, M., Liu, J., Li, Z., Tong, Y. *et al.* (2021) A versatile biosensing platform coupling CRISPR-Cas12a and aptamers for detection of diverse analytes. *Sci. Bull.*, **66**, 69–77.
- Zhao, X., Wang, Z., Yang, B., Li, Z., Tong, Y., Bi, Y., Li, Z., Xia, X., Chen, X., Zhang, L. *et al.* (2021) Integrating PCR-free amplification and synergistic sensing for ultrasensitive and rapid CRISPR/Cas12a-based SARS-CoV-2 antigen detection. *Synth. Syst. Biotechnol.*, **6**, 283–291.
- Li, S.-Y., Zhao, G.-P. and Wang, J. (2016) C-Brick: a new standard for assembly of biological parts using Cpf1. *ACS Synth. Biol.*, **5**, 1383–1388.
- Lei, C., Li, S.-Y., Liu, J.-K., Zheng, X., Zhao, G.-P. and Wang, J. (2017) 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.
- Farrar, K. and Donnison, I.S. (2007) Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. *Nat. Protoc.*, **2**, 1661–1674.
- Tan, G.-Y., Deng, K., Liu, X., Tao, H., Chang, Y., Chen, J., Chen, K., Sheng, Z., Deng, Z. and Liu, T. (2017) Heterologous biosynthesis of spinosad: an omics-guided large polyketide synthase gene cluster reconstitution in *Streptomyces*. *ACS Synth. Biol.*, **6**, 995–1005.
- Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P. and Luzhetskyy, A. (2018) Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.*, **49**, 316–324.
- Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A.F. and Fu, J. (2016) RecET direct cloning and Red $\alpha\beta$ recombining of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.*, **11**, 1175–1190.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) In: *Practical Streptomyces Genetics*. John Innes Foundation, Norwich, UK.
- Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.*, **16**, 6127–6145.
- Zhang, Y., Muylers, J.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.*, **18**, 1314–1317.
- Xu, F., Nazari, B., Moon, K., Bushin, L.B. and Seyedsayamdost, M.R. (2017) Discovery of a cryptic antifungal compound from

- Streptomyces albus* J1074 using high-throughput elicitor screens. *J. Am. Chem. Soc.*, **139**, 9203–9212.
29. Ziemert, N., Alanjary, M. and Weber, T. (2016) The evolution of genome mining in microbes—a review. *Nat. Prod. Rep.*, **33**, 988–1005.
 30. Luo, M. and Wing, R.A. (2003) An improved method for plant BAC library construction. In: Grotewold, E. (ed). *Plant Functional Genomics. Methods in Molecular Biology*. Vol. **236**. Humana Press, Totowa, NJ, pp. 3–19.
 31. Herschleb, J., Ananiev, G. and Schwartz, D.C. (2007) Pulsed-field gel electrophoresis. *Nat. Protoc.*, **2**, 677–684.
 32. González, A., Rodríguez, M., Braña, A.F., Méndez, C., Salas, J.A. and Olano, C. (2016) New insights into paulomycin biosynthesis pathway in *Streptomyces albus* J1074 and generation of novel derivatives by combinatorial biosynthesis. *Microb. Cell Fact.*, **15**, 56.
 33. Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y. and Deng, Z. (2003) Organizational and mutational analysis of a complete FR-008/candicidin gene cluster encoding a structurally related polyene complex. *Chem. Biol.*, **10**, 1065–1076.
 34. Zhang, M., Zhang, Y., Scheuring, C.F., Wu, C.-C., Dong, J.J. and Zhang, H.-B. (2012) Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protoc.*, **7**, 467–478.
 35. Liang, M., Liu, L., Wang, W., Zeng, X., Liu, J., Karthik, L., Zhu, G., Bai, L., Hou, C., Chen, X. *et al.* (2021) Simple cloning of large natural product biosynthetic gene clusters from *Streptomyces* by an engineered CRISPR/Cas12a system. bioRxiv doi: <https://doi.org/10.1101/2020.06.25.170191>, 12 May 2021, preprint: not peer reviewed.
 36. Wu, N., Huang, H., Min, T. and Hu, H. (2017) TAR cloning and integrated overexpression of 6-demethylchlorotetracycline biosynthetic gene cluster in *Streptomyces aureofaciens*. *Acta Biochim. Biophys. Sin.*, **49**, 1129–1134.
 37. Waldron, C., Matsushima, P., Rosteck, P.R., Broughton, M.C., Turner, J., Madduri, K., Crawford, K.P., Merlo, D.J. and Baltz, R.H. (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem. Biol.*, **8**, 487–499.
 38. Hahn, D.R., Gustafson, G., Waldron, C., Bullard, B., Jackson, J.D. and Mitchell, J. (2006) Butenyl-spinosyns, a natural example of genetic engineering of antibiotic biosynthetic genes. *J. Ind. Microbiol. Biotechnol.*, **33**, 94–104.
 39. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Ōmura, S. (1999) Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl Acad. Sci. U.S.A.*, **96**, 9509.
 40. Woo, S.-S., Jiang, J., Gill, B.S., Paterson, A.H. and Wing, R.A. (1994) Construction and characterization of bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Res.*, **22**, 4922–4931.
 41. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
 42. Tong, Y., Whitford, C.M., Robertsen, H.L., Blin, K., Jørgensen, T.S., Klitgaard, A.K., Gren, T., Jiang, X., Weber, T. and Lee, S.Y. (2019) Highly efficient DSB-free base editing for streptomycetes with CRISPR–BEST. *Proc. Natl Acad. Sci. U.S.A.*, **116**, 20366–20375.
 43. Flett, F., Mersinias, V. and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.*, **155**, 223–229.
 44. Fujita, K., Sugiyama, R., Nishimura, S., Ishikawa, N., Arai, M.A., Ishibashi, M. and Takeya, H. (2016) Stereochemical assignment and biological evaluation of BE-14106 unveils the importance of one acetate unit for the antifungal activity of polyene macrolactams. *J. Nat. Prod.*, **79**, 1877–1880.
 45. Futamura, Y., Sawa, R., Umezawa, Y., Igarashi, M., Nakamura, H., Hasegawa, K., Yamasaki, M., Tashiro, E., Takahashi, Y., Akamatsu, Y. *et al.* (2008) Discovery of incednine as a potent modulator of the anti-apoptotic function of Bcl-xL from microbial origin. *J. Am. Chem. Soc.*, **130**, 1822–1823.
 46. Lim, Y.H., Wong, F.T., Yeo, W.L., Ching, K.C., Lim, Y.W., Heng, E., Chen, S., Tsai, D.-J., Lauderdale, T.-L., Shia, K.-S. *et al.* (2018) Auroramycin: a potent antibiotic from *Streptomyces roseosporus* by CRISPR–Cas9 activation. *ChemBioChem*, **19**, 1716–1719.
 47. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H. and Weber, T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.*, **47**, W81–W87.
 48. Schulze, C.J., Donia, M.S., Siqueira-Neto, J.L., Ray, D., Raskatov, J.A., Green, R.E., McKerrow, J.H., Fischbach, M.A. and Linington, R.G. (2015) Genome-directed lead discovery: biosynthesis, structure elucidation, and biological evaluation of two families of polyene macrolactams against *Trypanosoma brucei*. *ACS Chem. Biol.*, **10**, 2373–2381.
 49. Hoshino, S., Okada, M., Awakawa, T., Asamizu, S., Onaka, H. and Abe, I. (2017) Mycolic acid containing bacterium stimulates tandem cyclization of polyene macrolactam in a lake sediment derived rare actinomycete. *Org. Lett.*, **19**, 4992–4995.
 50. Udvary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., Jensen, P.R. and Moore, B.S. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 10376.
 51. Zhao, Y., Sun, C., Huang, L., Zhang, X., Zhang, G., Che, Q., Li, D. and Zhu, T. (2021) Talarodrides A–F, nonadrides from the Antarctic sponge-derived fungus *Talaromyces* sp. HDN1820200. *J. Nat. Prod.*, **84**, 3011–3019.