Simple cloning of large natural product biosynthetic gene cluster by CRISPR/Cas12a-mediated fast direct capturing strategy

Mindong Liang[1, †], Leshi Liu[1, †], Weishan Wang[2, †], Xiaoqian Zeng[1, †], Jiakun Liu[1, 3], Loganathan Karthik[1], Guoliang Zhu[1], Linquan Bai[4], Chengjian Hou[1], Xiangyin Chen[1], Liming Ouyang[1], Xueting Liu[1], Bin Hu[5], Xuekui Xia[6], Yaojun Tong[1, 7], Chunbo Lou[3], Gil Alterovitz[8], Gao-Yi Tan[1, *], Li-Xin Zhang[1, *]

[1]State Key Laboratory of Bioreactor Engineering (SKLBE), and School of Biotechnology, East China University of Science and Technology (ECUST), Shanghai 200237, China.
[2]State Key Laboratory of Microbial Resources and CAS Key Laboratory of Pathogenic Microbiology and Immunology, Chinese Academy of Sciences (CAS), Beijing 100101, China.
[3]Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen 518055, China.
[4]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University (SJTU), Shanghai 200240, China.
[5]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China.
[6]Key Biosensor Laboratory of Shandong Province, Biology Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250013, China.
[7]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 21 2800 Kgs. Lyngby, Denmark.
[8]Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115, USA.

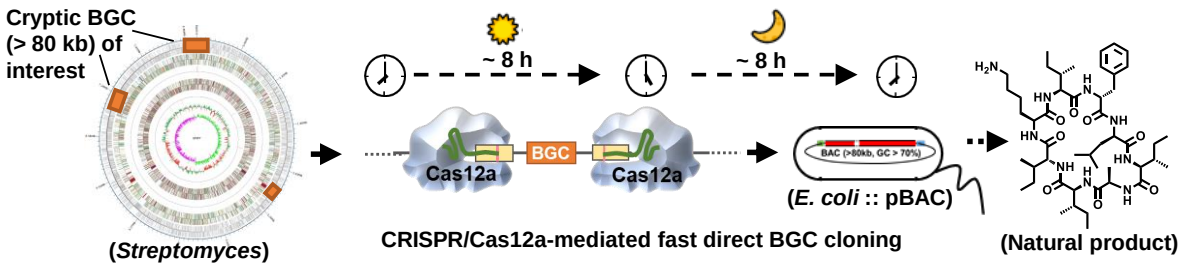
*To whom correspondence should be addressed. Tel/Fax: +86 21 64253020; Email: tangy@ecust.edu.cn. Correspondence may also be addressed to Tel/Fax: +86 21 64252575, Email: lxzhang@ecust.edu.cn

[†]These authors contributed equally to the paper as first authors. The authors wish it to be known that, in their opinion, the first 4 authors should be regarded as joint First Authors.

ABSTRACT

Directly cloning of biosynthetic gene clusters (BGCs) from even unculturable microbial genomes revolutionized nature products-based drug discovery. However, it is still very challenging to efficiently cloning, for example, the large (e.g. > 80kb) BGCs, especially for samples with high GC content in Streptomyces. In this study, by combining the advantages of CRISPR/Cas12a cleavage and bacterial artificial chromosome (BAC) library construction, we developed a simple, fast yet efficient in vitro platform for direct cloning of large BGCs based on CRISPR/Cas12a, named CAT-FISHING (CRISPR/Cas12a-mediated fast direct biosynthetic gene cluster cloning). It was demonstrated by the efficient direct cloning of large DNA fragments from bacterial artificial chromosomes or high GC (>70%) Streptomyces genomic DNA. Moreover, surugamides, encoded by a captured 87-kb gene cluster, was expressed and identified in a cluster-free Streptomyces chassis. These results indicate that CAT-FISHING is now poised to revolutionize bioactive small molecules (BSMs) drug discovery and lead a renaissance of interest in microorganisms as a source of BSMs for drug development.

TABLE OF CONTENTS



SIGNIFICANCE STATEMENT

Natural products (NPs) are one of the most important resources for drug leads. One bottleneck of NPs-based drug discovery is the inefficient cloning approach for BGCs. To address it, we established a simple, fast and efficient BGC directed cloning method CAT-FISHING by combining the advantages of CRISPR/Cas12a (e.g. high specificity) and bacterial artificial chromosome (BAC) library (e.g. large DNA fragment and high GC content). As demonstrations, a series of DNA fragments ranging from 49 kb to 139 kb were successfully cloned. After further optimization, our method was able to efficiently clone and express an 87-kb long, GC-rich (76%) surugamides BGC in a Streptomyces chassis with reduced time-cost. CAT-FISHING presented in this study would much facilitate the process of NPs discovery.

INTRODUCTION

Microorganisms, especially *Streptomyces*, remain unrivalled in their ability to produce bioactive small molecules (BSMs), some of which reached the market without any chemical modifications required, a testimony to the remarkable potential of *Streptomyces* to produce novel drugs. The expedition of *Streptomyces* genomes deciphered a large unexploited pool of novel biosynthetic gene clusters (BGCs), responsible for new but silent BSMs (1,2). However, the cloning of BGCs in *Streptomyces* is often very difficult because of the high GC and large size of those BGCs. It was found that 92% (1760/1910) of the characterized BGCs are smaller than 85kb, and 40% (756/1910) with > 70% GC content. In *Streptomyces*, 84% (534/634) BGCs have a GC content over 70% (Supplementary Figure S1). To date, various processes have been developed for BGC cloning, such as genomic library (*i.e.*, cosmid, fosmid and BAC) construction, recombination-based or RecET/Redαβ-based cloning, and Gibson assembly *etc* (Table 1). Additionally, the emergence of the CRISPR/Cas9 technique has enabled several new DNA cloning methods such as ExoCET, and CATCH *etc* (3-5). However, it is relatively cumbersome and time-consuming of using the aforementioned methods. Simple, fast and efficient strategy for large BGCs, especially with high GC content, is in urgent need to make natural BSMs accessible and affordable.

CRISPR/Cas12a is a single RNA-guided (crRNA) endonuclease of a Class II CRISPR/Cas system (19). Unlike Cas9 proteins, Cas12a recognizes a T-rich protospacer-adjacent motif (PAM) instead of a G-rich PMA and generates dsDNA breaks with staggered ends instead of blunt ends. Besides the

genome editing applications (19,20), CRISPR/Cas12a has been widely used in nucleic acid-based diagnostic applications (21-23), small molecule detection (24,25) *etc*. Moreover, it worth noting that CRISPR/Cas12a possesses obvious superiority in DNA assembly with regard to its programable endonuclease activity and the DNA sticky ends of 4- or 5-nt overhangs (19). Based on these features, Li *et al.* developed a CRISPR/Cas12a-based DNA assembly standard, namely C-Brick (26). Subsequently, a DNA assembly method (namely, CCTL) was reported (27). These features suggested the probability of being able to directly clone of large BGCs by using CRISPR/Cas12a.

Bacterial artificial chromosome (BAC) library construction is a classical method for cloning large DNA fragments, but it is time consuming and labour intensive, as well as technically demanding (6,28). However, compared to other PCR-based or recombination-based cloning methods that have recently emerged (Table 1), the BAC library is indiscriminate toward insertion in a DNA sequence. This makes BAC library construction suitable for application to high GC content DNA samples. In this study, by combining the DNA cleavage activity of CRISPR/Cas12a with the unique features of a BAC library construction, we have developed a CRISPR/Cas12a-derived method, designated CAT-FISHING (CRISPR/Cas12a-mediated fast direct biosynthetic gene cluster cloning). As a proof of concept, DNA fragments (or BGCs) of > 80kb have been fast cloned from different high GC content DNA samples (*e.g.* bacterial artificial chromosome plasmid, 66% GC; or *Streptomyces* genomic DNA, 73% GC) by CAT-FISHING. Furthermore, the captured 87-kb surugamides BGC has also been successfully expressed in the *S. albus* J1074-derived cluster-free chassis strain (29).

MATERIAL AND METHODS

Strains, plasmids and media. The strains and plasmids used in this work are present in Supplementary Table S2. *Escherichia coli* and its derivatives were cultivated on Luria-Bertani (LB) 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 (SFM) 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 Bioscience, San Jose, CA, USA). In the fermentation experiments, seeds were grown in TSB (trypticase soy broth, Oxoid Ltd), and R4 medium was used for subsequent fermentations (in a 250-mL Erlenmeyer flask, 30°C and 200 rpm).

Capture plasmid construction. The primers for capture plasmid construction are listed in Supplementary Table S3. 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) corresponding to the flanking regions of the target DNA fragment (or BGC) in to pBAC2015 (12). Assembly of multiple DNA fragments was carried out by Gibson assembly or with the EZmax one-step seamless cloning kit (Tolo Biotechnology, Shanghai). Plasmid DNA was isolated from *E. coli* by using alkaline lysis (30).

Genomic DNA isolation. For genomic DNA isolation, *S. albus* J1074 was cultured in Oxoid TSB (30 g/L) supplemented with glycine (5 g/L). According to *Practical Streptomyces Genetics* (30), after cultivation at 200 rpm and 30°C for 24 ~ 36 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 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 mould (100-μl holes). The blocks were removed from the mould 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), then three times with TE for 30 min. After removing all of the liquid, the agarose plugs could be used for CRISPR/Cas12 digestion, but could also be stored for up to 1 month at 4°C in 70% ethanol.

crRNA preparation. The oligonucleotides used as templates for crRNA transcription are given in Supplementary Table S4. According to our previous study (24), 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 & ConcentratorTM-5 kit (Zymo Research), and subsequently quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). RNase-free materials (Axygen Scientific, Union City, CA, USA) and conditions were applied during the entire experimental process.

CRISPR/Cas12a-based DNA restriction and ligation. The Cas12a (LbCas12a) protein used in this study was overexpressed in pET28a and then purified by fast protein liquid chromatography (FPLC; AKTA Explorer 100, GE Healthcare). In the CRISPR/Cas12a cutting system, NEBuffer™ 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/mL BSA, pH 7.9) was adopted as the reaction buffer. For pBAC-ZL or capture plasmid cleavage, plasmid DNA was incubated with Cas12a protein and the corresponding crRNA pairs at 37°C for 1 h. After the reaction, the linearized capture plasmids or DNA fragments of pBAC-ZL were prepared using isopropanol and ethanol (12). 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 analysed 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 sec 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 inactive Cas12a protein, the LMP agarose gel was hydrolysed 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 T4 DNA ligase (NEB).

Electro-transformation of *E. coli*. Following 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 electro-transformation. The

high efficiency electro-transformation of *E. coli* cells was accomplished according to a previous study (31). The following electro-transformations were performed in 2-mm cuvettes using the Bio-Rad GenePulser Xcell™ system (electroporator 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, $MgCl_2$ 10 mM, glucose 20 mM) was added and the mixture was transferred into a 15-mL Falcom™ tube. After shaking at 200 rpm for 1 h at 37°C, the strains were collected and spread on selective LB 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 S5.

Expression of BGCs in *Streptomyces* and LC-MS analysis. The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette amplified from pSET152 has been introduced into BAC plasmids by Red/ET recombination (32). The resulting plasmid was introduced into *S. albus* Del14 by triparental conjugation according to a previous study (6). The transconjugants were screened and verified by PCR using the primers listed in Supplementary Table S5. After fermentation, the production of target natural product was qualitatively analysed using a high-resolution Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA).

RESULTS

Design and workflow of CAT-FISHING

By combining the DNA cleavage activity of CRISPR/Cas12 with the unique features of a BAC library construction, an *in vitro* platform (designated CAT-FISHING) for large BGC cloning has been developed. The flow chart of CAT-FISHING is presented in Figure 1, and includes three steps. The first step is the capture plasmid construction and CRISPR/Cas12a- based plasmid digestion. In this step, two homolog arms (each arm containing at least one PAM site) that flank the target GBC were selected and amplified by PCR. Then the BAC plasmid backbone containing the two homolog arms and selection marker (*e.g.,* antibiotic resistance gene, counter selection gene or *lacZ*), and the designated capture plasmid was constructed via the DNA assembly method. Under the guidance of crRNAs, the two selected PAM motif regions on the left and right homolog arms were simultaneously digested by CRISPR/Cas12a, resulting in the linear capture plasmid. The second step is genomic DNA isolation and CRISPR/Cas12-based genome digestion. According to the BAC library construction protocol, genomic DNA plugs from the target strain were prepared. And the genomic DNA was digested by the CRISPR/Cas12a system guided by the two designed crRNAs that were previously used in step one. The last step is ligation and transformation. The resulting linear capture plasmid and the digested genomic DNA from steps I and II, respectively, were mixed and ligated by T4 DNA ligase. Then ligation products were introduced into *E. coli* by electroporation. Target BGCs could then be obtained from transformants by PCR-based screening.

Evaluation of CAT-FISHING cloning efficiency

The principle underlying CAT-FISHING is the cohesive end ligation of two linearized DNA fragments by T4 DNA ligase. However, different from the widely used restriction endonuclease-based DNA cloning methods, here the cohesive ends were generated by paired crRNA-guided CRISPR/Cas12a cleavage. This study therefore evaluated and compared the cloning efficiencies achieved by applying two different kinds of cohesive ends that were individually generated by NEB restriction endonuclease and CRSIPR/Cas12a. As shown in Supplementary Figure S2, plasmid pGY2020 derived from the pCC2-FOS Fosmid vector (Epicentre) was constructed, and this plasmid contains two PAM sites (PAM1 and PAM2) as well as two NEB restriction endonucleases (EcoRI and HindIII). And the specific DNA fragment (ampicillin resistant gene Amp^R) was cloned into pGY2020 by the CRISPR/Cas12a or NEB restriction enzymes-based method. The clone number of the CRISPR/Cas12a-based method was 34% ($P > 0.05$) lower than that of the NEB restriction enzymes-based method. However, there was no significant difference in the true positive rate between these two cloning methods (Figure 2A-2B and Supplementary Figure S3-S4). These results indicate that CAT-FISHING should be able to clone target DNA fragments with a relatively high efficiency.

Cloning of a target DNA fragment from a BAC plasmid

In order to further demonstrate CAT-FISHING in a simplified system, a 137-kb BAC plasmid pBAC-ZL was used to evaluate its cloning performance on a large DNA fragment. As shown in Figure 2C, a 50-kb fragment and an 80-kb fragment could be obtained by using the corresponding crRNAs-guided CRISPR/Cas12a cleavage. Under the guidance of the corresponding crRNA pairs, the BAC plasmid was digested by CRISPR/Cas12a, and 50-kb and 80-kb target bands were observed on the agarose gel after PFGE (Supplementary Figure S5). By using the corresponding capture plasmid, two target DNA fragments were also successfully cloned from the BAC plasmid, as shown in Figure 2D-2E and Supplementary Figure S6-S7. For 50-kb DNA fragment, of more than 100 transformants about 95% were the right clones. For the 80-kb DNA fragment, the number of transformants and the true positive rate were both lower, and about 50% of the transformants were the right clones. These results indicated that CAT-FISHING could achieve high cloning rates of the 50-kb and 80-kb DNA fragments from the BAC plasmid. It also needs to be noted that, for the 80-kb DNA fragment, the cloning difficulty was obviously greater.

Direct cloning of target BGCs from Streptomyces genomic DNA

In order to directly clone large BGCs from genomic DNA, a very simple procedure for fast cloning has been developed, as shown in Figure 3A. After genomic DNA isolation and subsequent CRISPR/Cas12a digestion, the resulting sample containing a mixture of genomic DNA could be directly used for subsequent ligation and transformation without prior DNA fragment isolation by PFGE and purification. In this study, a 49-kb paulomycin gene cluster (33), an 87-kb surugamides gene cluster (34) as well as a 139-kb candicidin gene cluster (35) were selected from the chromosome of *S. albus* J1074 to demonstrate this method. As shown in Table 2, the 49-kb paulomycin gene cluster (GC content 71%) and the 87-kb surugamides gene cluster (GC content 76%) were successfully captured by CAT-FISHING, as confirmed by PCR and restriction mapping

(Figure 3B-3C). Additionally, the 139-kb candicidin gene cluster (GC content 75%) was also captured with CAT-FISHING, albeit with a much lower efficiency. These results indicate that CAT-FISHING is a simple and fast method for directly cloning large BGCs from high GC genomic DNA samples (Supplementary Table S1).

Expression of the target BGC in a cluster-free *Streptomyces* chassis strain

To thoroughly check the sequence and functional integrity of these BGCs that were captured by CAT-FISHING, as well as to prove access to genome mining through this route, a captured 87-kb surugamides gene cluster was expressed in a cluster-free *Streptomyces* chassis strain. The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette was introduced into the target plasmid by Red/ET recombination. By applying ET12567/pUC307-mediated triparental conjugation, the resulting plasmid pBAC2015-87kb-J1074-int was integrated into the chromosome of *S. albus* Del14. During the subsequent fermentation study, surugamide A (RT = 4.53 min, m/z = 912.6293) and other components (*i.e.,* surugamides D, G, H and I) were produced in *S. albus* Del14-87kb. LC−MS/MS analysis further confirmed the production of surugamide A (Figure 4).

DISCUSSION

Cloning and manipulation of large DNA fragments is a fundamental and important platform technology for mining results BSMs from rich microbial genome sources. In this study, CAT-FISHING combined the advantages of CRISPR/Cas12a cleavage and BAC library construction. Instead of using restriction enzymes (*e.g.* HindIII, EcoRI, or BamHI) for the random fragmentation of the genomic DNA (8), CAT-FISHING utilizes CRISPR/Cas12a together with specific crRNA pairs to precisely cut the genomic DNA (Figure 1). This theoretically makes CAT-FISHING a more simple and direct method for cloning of large BGCs with high GC contents.

There are two concerns about CRISPR/Cas12a-based DNA cleavage, off-target (non-specific cleavage) and inaccuracy of the cleavage sites (26,36). Both could reduce the cloning efficiency of CAT-FISHING. Lei *et al.* reported that the cleavage specificity of CRISPR/Cas12a would enhance with a shorter crRNA spacer (*i.e.*, 17 ~ 19 nt) (27). In this study, we used crRNA with an 18-nt spacer has been applied to evaluate the cloning efficiency of CAT-FISHING. As shown in Figure 2A-2B, the number of transformants obtained by CAT-FISHING were fewer than the control, while no significant difference ($P > 0.05$) was observed. On the other hand, non-specific cleavage by CRISPR/Cas12a could be minimized by decreasing the Cas12a concentration and shortening the cleavage time. As a result, when a purified 137-kb BAC plasmid was used to demonstrate CAT-FISHING, the plasmid was almost completely digested and no non-specific cleavage products appeared on the agarose gel (Figure 2C and Supplementary Figure S5). To some extent, these results verify the relatively higher efficiency of cloning 50-kb and 80-kb DNA fragments from BAC plasmid (Figure 2D-2E). In addition, due to its high efficiency, CAT-FISHING also could be used as an efficient *in vitro* large DNA fragment editing system.

PFGE is a powerful and essential tool for isolating large DNA fragments. During BAC library construction, however, PFGE is time consuming, as it often takes 16 ~ 24 hours to separate specific size DNA fragments (8). Moreover, the following operational steps, such as DNA elution and purification, could drastically decrease the DNA integrity/amount as well as the subsequent ligation or transformation efficiency. In this study, after many preliminary tests, we found that, following agarose digestion, the resulting mixture from a CRISPR/Cas12a-treated genomic DNA plug could be directly used for ligation with the vector and subsequent electro-transformation (Figure 3A). Without the need for PFGE or the preparation of purified high molecular weight genomic DNA fragments, and compared to previously reported large DNA fragment cloning approaches (*e.g.* ExoCET, CATCH, TAR *etc.*), the cloning process in CAT-FISHING has been greatly simplified (3,5,10,18). In this study, by applying CAT-FISHING, we found the ratio of right clones that contain a 49-kb target BGC was 4~5%, and that for an 87-kb target BGC was 2~4% (Figure 3B-3C and Table 2). Probably due to the complexity of the un-purified DNA mixture sample and high activity of T4 DNA ligase, many short DNA fragments or incomplete pieces of target BGCs were inserted into capture plasmids. Therefore, it is reasonable to predict that, if necessary, through DNA fragment isolation and purification, the cloning performance toward a 139-kb BGC should also be dramatically improved. However, compared to the BAC library in which often only a few right strains could be screened out of thousands of clones (i.e., 1/1000 ~ 1/2000) (6,37), CAT-FISHING is a simpler method with a greater efficiency for cloning target BGCs.

Streptomyces are the source of a majority of antibiotic classes in current clinical and agricultural use (38-41). *S. albus* J 1074 is one of the most widely used *Streptomyces* chassis for genome mining (42). In this study, *S. albus* Del14, which is a *S. albus* J 1074-derived cluster-free chassis strain (29), has been used to demonstrate the sequence and functional integrity of the BGCs obtained by CAT-FISHING. As shown in Figure 4, the 87-kb target BGC was successfully expressed, and the corresponding NP surugamides (inhibitors of cathepsin B) could be confirmed by LC-MS. During genome mining, BGC cloning and expression is the most important starting point for the next step of bioactivity analysis and structure identification of target BSMs. The current results present a case study for NP production in *Streptomyces* by applying CAT-FISHING. Lastly, as an *in vitro* manipulation platform, not limited to actinomyces, CAT-FISHING could easily be extended to genome mining in fungi and other microbial resources (43).

Based on the results described above, we concluded that, by combining the advantages of CRISPR/Cas12a cleavage and BAC library construction, CAT-FISHING offers a simple, fast but efficient direct cloning strategy for targeting large BGCs from high GC content genomic DNA. And, in addition to genome editing, DNA assembly, nucleic acid and small molecule detection *etc*., this study also expanded the application of CRISPR/Cas12 to direct cloning of large DNA fragments *in vitro*. This innovation of a fundamental platform technology for use in genome mining through application of the CRISPR/Cas12 system would facilitate the discovery of novel BSMs from microbial sources.

ACKNOWLEDGEMENT

The authors would like to thank Assoc. Prof. Wei Ma (School of Life Sciences and Biotechnology, Shanghai Jiao Tong University) for his helpful suggestion in pulsed-field gel electrophoresis.

FUNDING

This work was supported by the National Natural Science Foundation of China [31720103901, 31870040, 21877038], the “111” Project of China [B18022], the Fundamental Research Funds for the Central Universities [22221818014], the Open Project Funding of the State Key Laboratory of Bioreactor Engineering, and the Shandong Taishan Scholar Award to L.-X. Z.

REFERENCES

1. 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.
2. Culp, E.J., Yim, G., Waglechner, 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.
3. Jiang, W. and Zhu, T.F.J.N.P. (2016) Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc.* **11**, 960-975.
4. 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
5. Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A.F. *et al.* (2017) ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.*, **46**, e28-e28.
6. Tan, G.Y., Deng, K., Liu, X., Tao, H., Chang, Y., Chen, J., Chen, K., Sheng, Z., Deng, Z. and Liu, T.J.A.S.B. (2017) Heterologous biosynthesis of spinosad: an omics-guided large polyketide synthase gene cluster reconstitution in *Streptomyces*. *ACS Synth. Biol.*, **6**, 995-1005.
7. Xu, M., Wang, Y., Zhao, Z., Gao, G., Huang, S.X., Kang, Q., He, X., Lin, S., Pang, X. and Deng, Z. (2016) Functional genome mining for metabolites encoded by large gene clusters using heterologous expression of a whole genomic BAC library in Streptomyces. *Appl. Environ. Microbiol.*, **82**, 5795-5805.
8. Luo, M. and Wing, R.A. (2003) In Grotewold, E. (ed.), *Plant Functional Genomics*. Humana Press, Totowa, NJ, pp. 3-19.
9. Yu, Y., Bai, L., Minagawa, K., Jian, X., Li, L., Li, J., Chen, S., Cao, E., Mahmud, T., Floss, H.G. *et al.* (2005) Gene cluster responsible for validamycin biosynthesis in *Streptomyces hygroscopicus* jinggangensis 5008. *Appl.Environ. Microbiol.*, **71**, 5066-5076.

10. Kazuya, Y., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Victor, N., Dorrestein, P.C. and Moore, B.S. (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1957-1962.
11. Noskov, V.N., Karas, B.J., Young, L., Chuang, R.-Y., Gibson, D.G., Lin, Y.-C., Stam, J., Yonemoto, I.T., Suzuki, Y., Andrews-Pfannkoch, C. *et al.* (2012) Assembly of large, high G+C bacterial dna fragments in yeast. *ACS Synth. Biol.*, **1**, 267-273.
12. Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A.F., Fu, J. *et al.* (2016) RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.*, **11**, 1175-1190.
13. Bian, X., Huang, F., Wang, H., Klefisch, T., Müller, R. and Zhang, Y. (2014) Heterologous production of glidobactins/luminmycins in escherichia coli nissle containing the glidobactin biosynthetic gene cluster from *Burkholderia* DSM7029. *ChemBioChem*, **15**, 2221-2224.
14. Li, L., Zhao, Y., Ruan, L., Yang, S., Ge, M., Jiang, W. and Lu, Y.J. (2015) A stepwise increase in pristinamycin II biosynthesis by *Streptomyces pristinaespiralis* through combinatorial metabolic engineering. *Metab. Eng.*, **29**, 12-25.
15. Song, C., Luan, J., Cui, Q., Duan, Q., Li, Z., Gao, Y., Li, R., Li, A., Shen, Y., Li, Y. *et al.* (2019) Enhanced heterologous spinosad production from a 79-kb synthetic multioperon assembly. *ACS Synth. Biol.*, **8**, 137-147.
16. Li, L., Jiang, W. and Lu, Y. (2018) In Jensen, M. K. and Keasling, J. D. (eds.), *Synthetic Metabolic Pathways: Methods and Protocols*. Springer New York, New York, NY, pp. 203-209.
17. Casini, A., MacDonald, J.T., Jonghe, J.D., Christodoulou, G., Freemont, P.S., Baldwin, G.S. and Ellis, T. (2013) One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res.*, **42**, e7-e7.
18. 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. Commu.*, **6**, 8101.
19. Bernd, Z., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Patrick, E., Volz, S.E., Julia, J., John, V.D.O. and Aviv, R.J.C. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759-771.
20. 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. Commu.*, **8**, 15179.
21. 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.
22. 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.
23. Li, S.-Y., Cheng, Q.-X., Liu, J.-K., Nie, X.-Q., Zhao, G.-P. and Wang, J. (2018) CRISPR-Cas12a has both *cis*- and *trans*-cleavage activities on single-stranded DNA. *Cell Res.*, **28**, 491-493.

24. 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. Commu.*, **10**, 3672.

25. Xiong, Y., Zhang, J., Yang, Z., Mou, Q., Ma, Y., Xiong, Y. and Lu, Y. (2020) Functional DNA Regulated CRISPR-Cas12a sensors for point-of-care diagnostics of non-nucleic-acid targets. *J. Am. Chem. Soc.*, **142**, 207-213.

26. 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.

27. 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-e74.

28. Farrar, K. and Donnison, I.S. (2007) Construction and screening of BAC libraries made from Brachypodium genomic DNA. *Nat. Protoc.*, **2**, 1661-1674.

29. 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.

30. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) *Practical streptomyces genetics*. John Innes Foundation Norwich.

31. 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.

32. Zhang, Y., Muyrers, J.P.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.*, **18**, 1314-1317.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. Tan, G.Y. and Liu, T. (2016) rational synthetic pathway refactoring of natural products biosynthesis in actinobacteria. *Metab. Eng.*, **39**, 228.

39. Gao, Q., Tan, G.-Y., Xia, X. and Zhang, L. (2017) Learn from microbial intelligence for avermectins overproduction. *Curr. Opin. Biotechnol.*, **48**, 251-257.

40. Tan, G.-Y., Peng, Y., Lu, C., Bai, L. and Zhong, J.-J. (2015) Engineering validamycin production by tandem deletion of γ-butyrolactone receptor genes in *Streptomyces hygroscopicus* 5008. *Metab. Eng.*, **28**, 74-81.

41. 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.

42. Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V. and Luzhetskyy, A. (2014) Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics*, **15**, 97.

43. Ziemert, N., Alanjary, M. and Weber, T. (2016) The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.*, **33**, 988-1005.

TABLE AND FIGURES LEGENDS

Table 1. Previous approaches that have been used for BGC cloning *in vitro*.

Approach	Brief. description	ULCC	Limitations	Cloned BGC	Ref.
Library construction					
Cosmid	i) Genomic DNA isolation; ii) Partial digestion or DNA fragmentation; iii) Ligation and transformation; iv) Target single clone screening.	~ 40kb	Time consuming; Labour intensive.	Validamycin, Spinosad, Borrelidin, etc.	(6-9)
Fosmid		~ 40kb			
BAC		> 90kb			
Yeast recombination					
TAR	i) Genomic DNA isolation; ii) Transformation and recombination in yeast; iii) Plasmid isolation and *E. coli* transformation.	67kb	Mis-priming or mis-annealing for high GC content or repeated sequence (e.g. Polyketide synthase)	Taromycin, etc.	(10)
YA	i) DNA fragment preparation by PCR; ii) PCR products transformation and recombination in yeast; iii) Plasmid isolation and *E. coli* transformation.	> 200kb	Mis-priming or mis-annealing for high GC content or repeated sequence; Mutation caused by PCR	N/A	(11)
RecET/Redαβ					
RecET	i) Genomic DNA isolation; ii) DNA digestion by specific restriction enzymes; iii) Transformation and application of RecET.	~ 50kb	Restriction enzymes limitation; Mis-priming or mis-annealing for high GC content or repeated sequence (e.g. Polyketide synthase)	Glidobactins, etc.	(12-14)
ExoCET	i) Genomic DNA isolation; ii) DNA digestion by restriction enzymes or CRISPR/Cas9; iii) Transformation and application of RecET.	106kb	Mis-priming or mis-annealing for high GC content or repeated sequence (e.g. Polyketide synthase)	Salinomycin, Spinosad.	(5,15)
Gibson assembly					
GA	i) DNA fragment preparation by PCR; ii) Gibson assembly; iii) Transformation.	72kb	Mutation caused by PCR; Mismatched linker pairings for high GC content sequence	Pristinamycin	(14,16,17)
CRISPR/Cas9					
CATCH	i) Genomic DNA isolation; ii) DNA digestion by CRISPR/Cas9; iii) Gibson assembly and transformation.	~ 100kb	Mismatched linker pairings for high GC content sequence	N/A	(3,18)
Cas9, λ packaging	i) Genomic DNA isolation; ii) DNA digestion by CRISPR/Cas9; iii) *in vitro* λ packaging and ligation; iv) Transformation	~ 40kb	Difficult for large gene clusters cloning, e.g. > 50kb.	Sisomicin	(4)

BAC: bacterial artificial chromosome; ULCC: upper limit of cloning capacity; BGC: biosynthetic gene cluster; NP: natural product; CATCH: Cas9-assisted targeting of chromosome segments; TAR: transformation-associated recombination; YA: yeast assembly; GA: Gibson assembly; ExoCET: exonuclease 466 combined with RecET recombination.

Table 2. The cloning efficiencies of high-GC target BGCs from genomic DNA.

Items		Target BGCs		
		Paulomycin	Surugamides	Candicidin
Source strain		*S. albus* J1074	*S. albus* J1074	*S. albus* J1074
Cloning vector		pBAC2015	pBAC2015	pBAC2015
Genome size of source strain		6.8 Mb	6.8 Mb	6.8 Mb
GC content of genome		73%	73%	73%
Type of target NP		Quinone glycoside	Polyketide	Polyketide
Size of target BGC		49kb	87kb	139kb
GC content of target BGCs		71%	76%	75%
White clone, c.f.u (/mL)		659 ± 28 *	508 ± 56 *	330 ± 47*
Corrected/checked (by PCR)	Repeat 1	4/96	3/96	1/192
	Repeat 2	4/96	2/96	0/96
	Repeat 3	5/96	4/96	0/96
Percentage of right clone (verified by restriction enzyme digestion from 5 PCR-verified clones)		100%	100%	N/A

* (Optional) Commercially available ElectroMAXTM DH10BTM cells (Thermo Fisher Scientific, Invitrogen™) could be used to increase the number of transformants. All experiments were performed in triplicate.

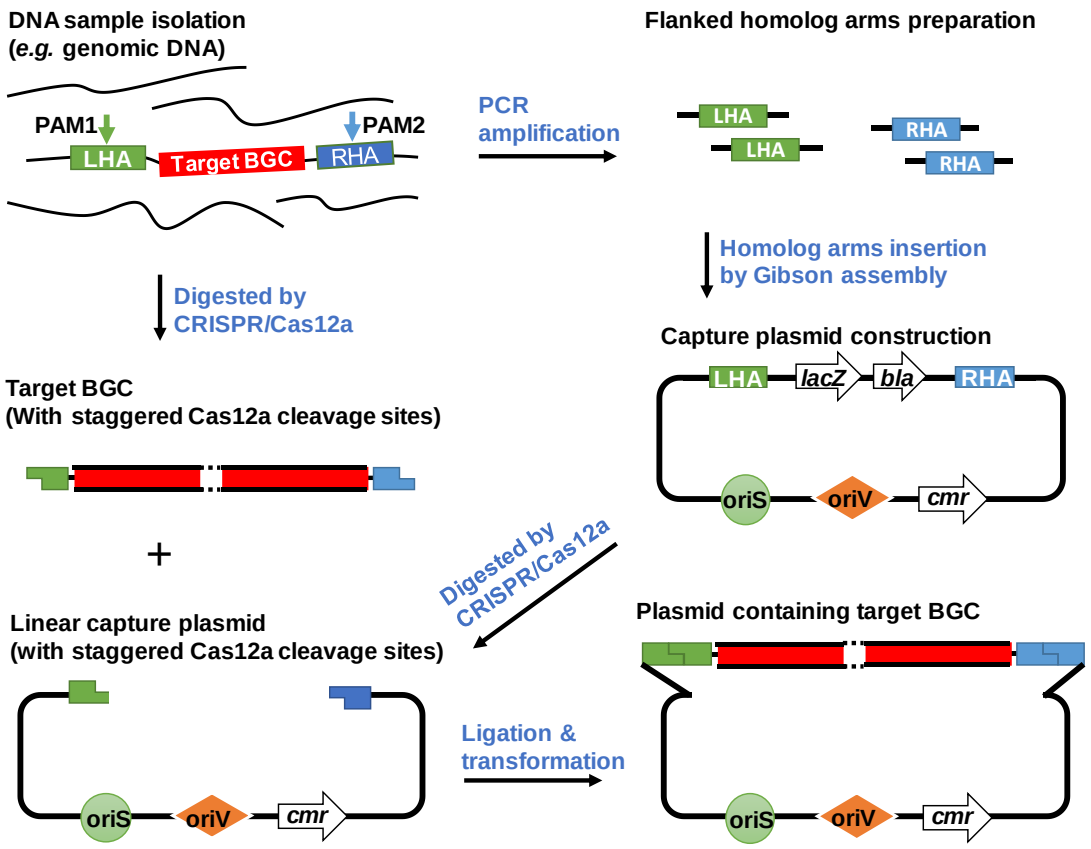


Figure 1. Schematic diagram of CAT-FISHING. LHA: left homology arm, RHA: right homology arm, BGC: biosynthetic gene cluster.

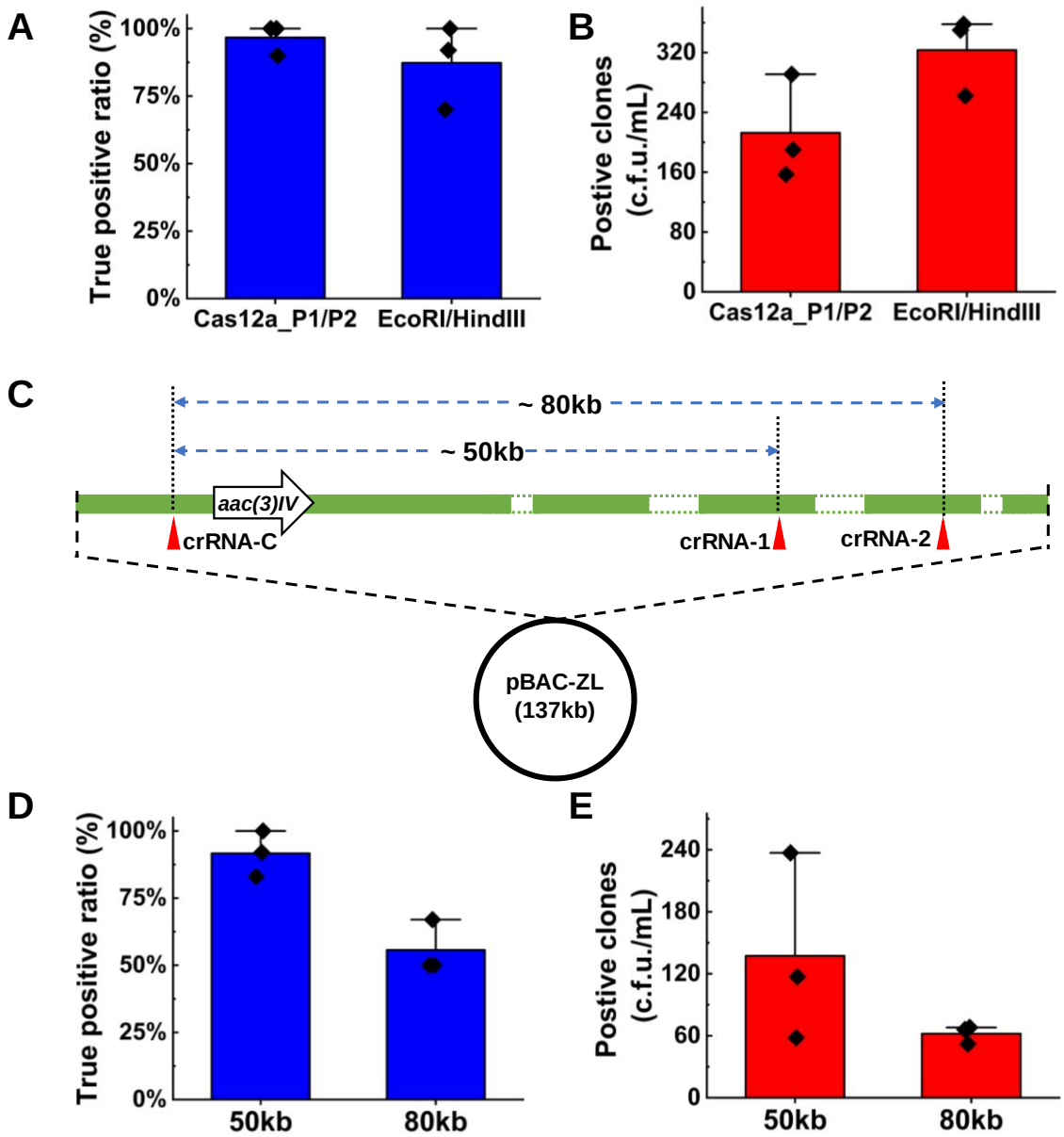


Figure 2. Cloning of DNA fragments with various of lengths by CAT-FISHING. (A, B) Comparison of the clone numbers and positive rates of the CAT-FISHING and NEB restriction enzymes-based methods. (C) Two different target segments with different lengths (50 kb and 80 kb) in the BAC plasmid. (D, E) Determination of the clone numbers and positive rates for the DNA fragments of different lengths.

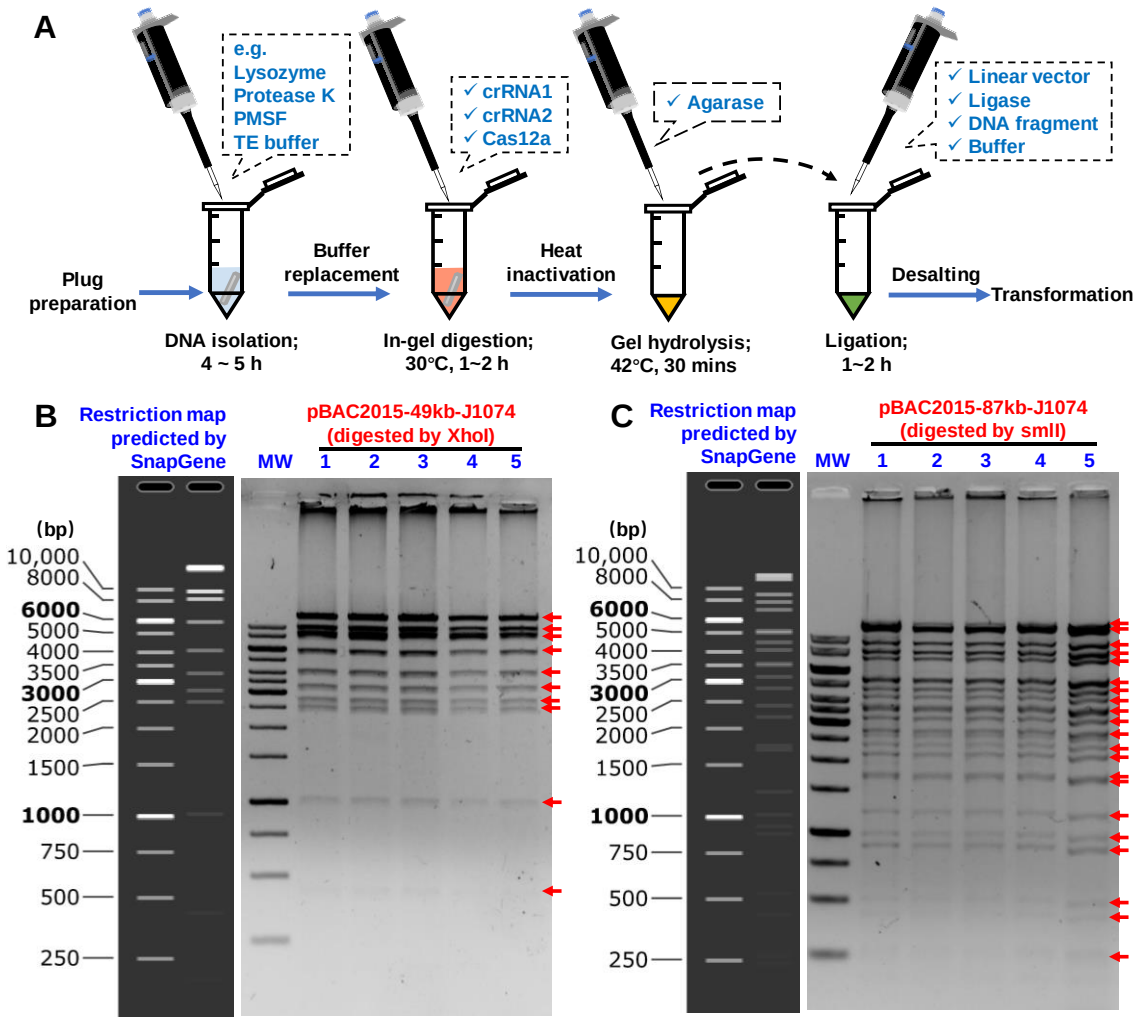


Figure 3. Direct cloning of large BGCs by CAT-FISHING. (A) A workflow of the target BGC capturing by CAT FISHING. (B, C). Validation of five randomly selected positive clones containing a paulomycin or surugamides gene cluster by restriction enzymes digestion. XhoI and SmlI have, respectively, been used for paulomycin and surugamides gene cluster restriction. Bands are indicated by arrows.

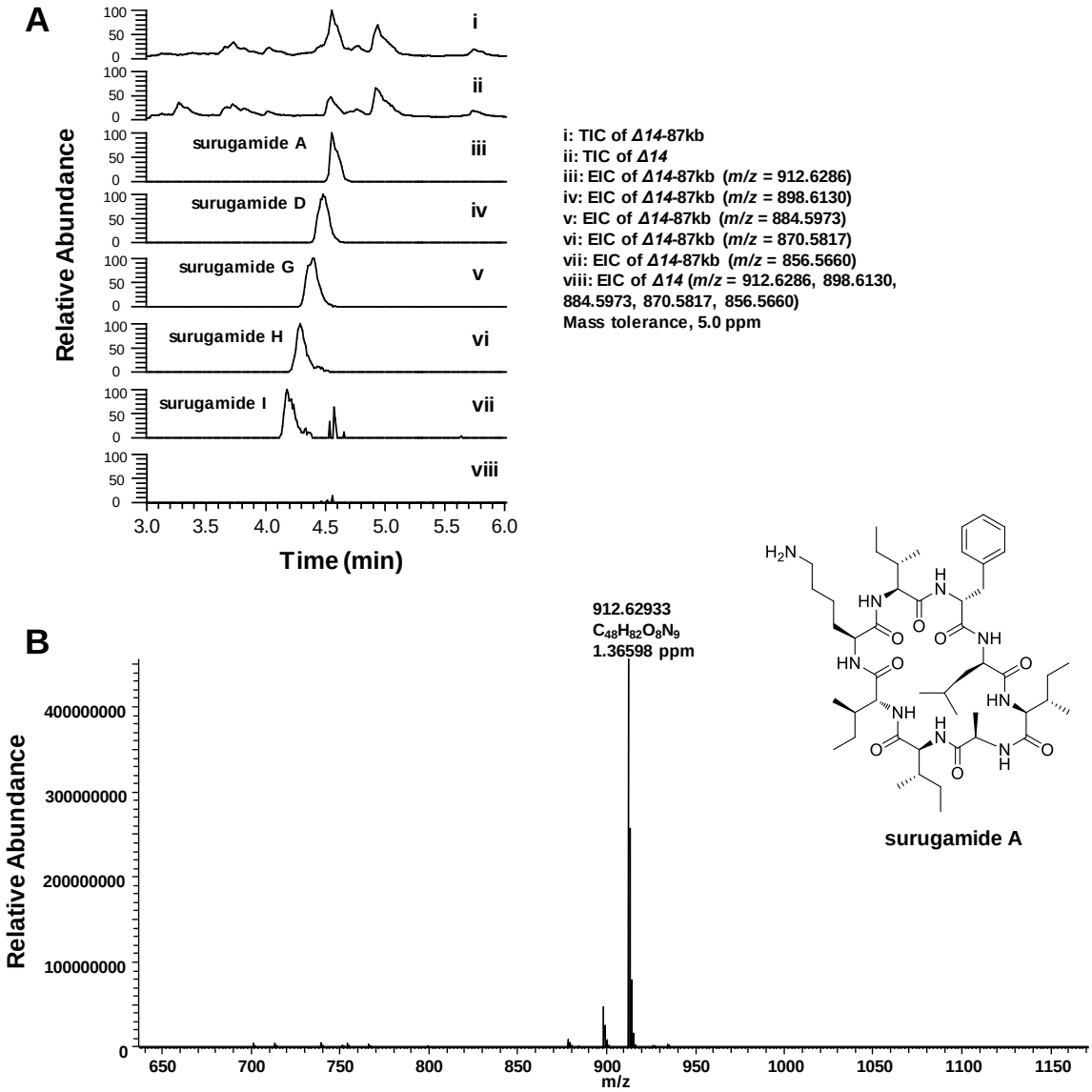


Figure 4. Expression of pBAC2015-87kb-J1074-int in S. albus Del 14. (A) Detection of surugamide components by LC-MS in S. albus Del-87kb. (B) High-resolution mass spectrum of surugamide A (m/z = 912.62933 Da; RT = 4.53 min).

Supplementary Information for

Simple cloning of large natural product biosynthetic gene cluster by CRISPR/Cas12a-mediated fast direct capturing strategy

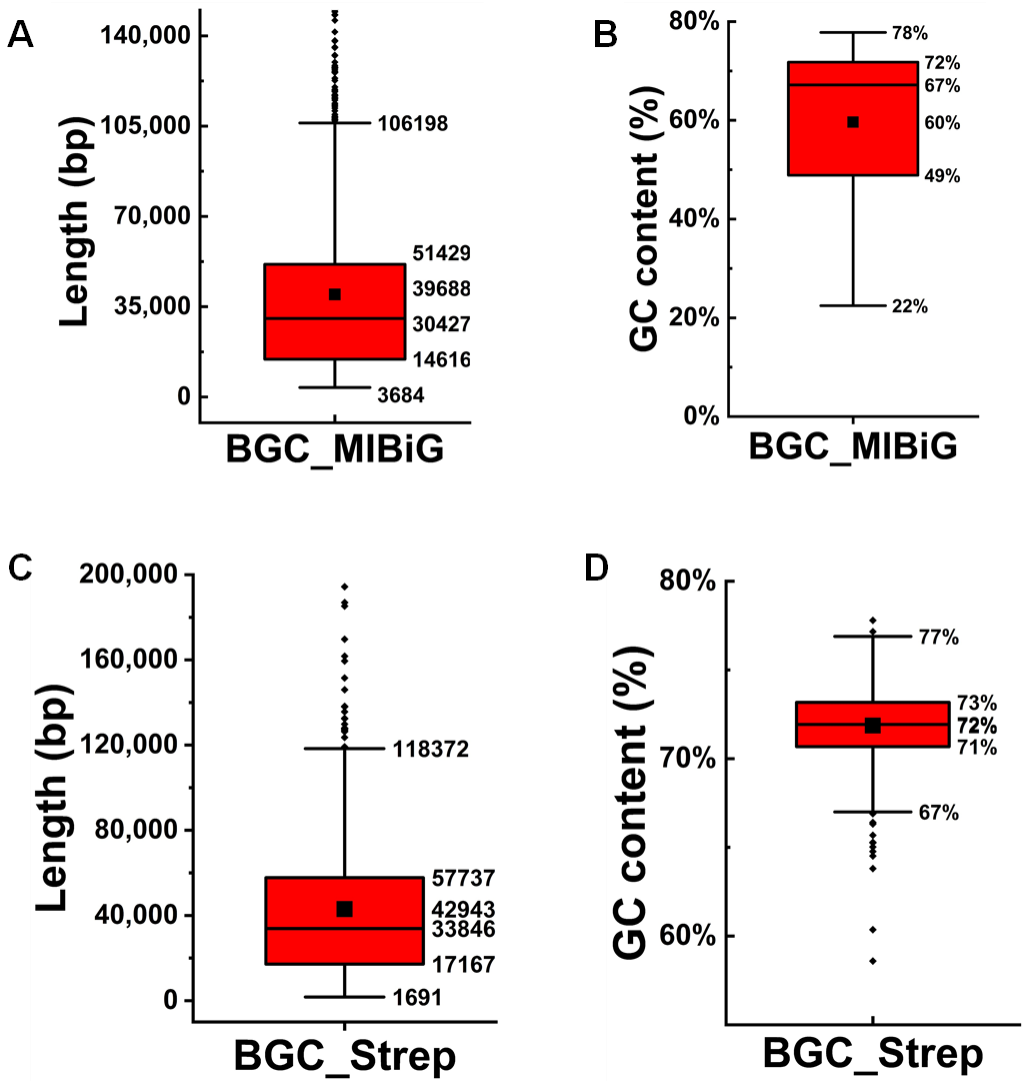
Mindong Liang[1, †], Leshi Liu[1, †], Weishan Wang[2, †], Xiaoqian Zeng[1, †], Jiakun Liu[1, 3], Loganathan Karthik[1], Guoliang Zhu[1], Linquan Bai[4], Chengjian Hou[1], Xiangyin Chen[1], Liming Ouyang[1], Xueting Liu[1], Bin Hu[5], Xuekui Xia[6], Yaojun Tong[1, 7], Chunbo Lou[3], Gil Alterovitz[8], Gao-Yi Tan[1, *], Li-Xin Zhang[1, *]

[1]State Key Laboratory of Bioreactor Engineering (SKLBE), and School of Biotechnology, East China University of Science and Technology (ECUST), Shanghai 200237, China.
[2]State Key Laboratory of Microbial Resources and CAS Key Laboratory of Pathogenic Microbiology and Immunology, Chinese Academy of Sciences (CAS), Beijing 100101, China.
[3]Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen 518055, China.
[4]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University (SJTU), Shanghai 200240, China.
[5]Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China.
[6]Key Biosensor Laboratory of Shandong Province, Biology Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250013, China.
[7]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 21 2800 Kgs. Lyngby, Denmark.
[8]Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115, USA.

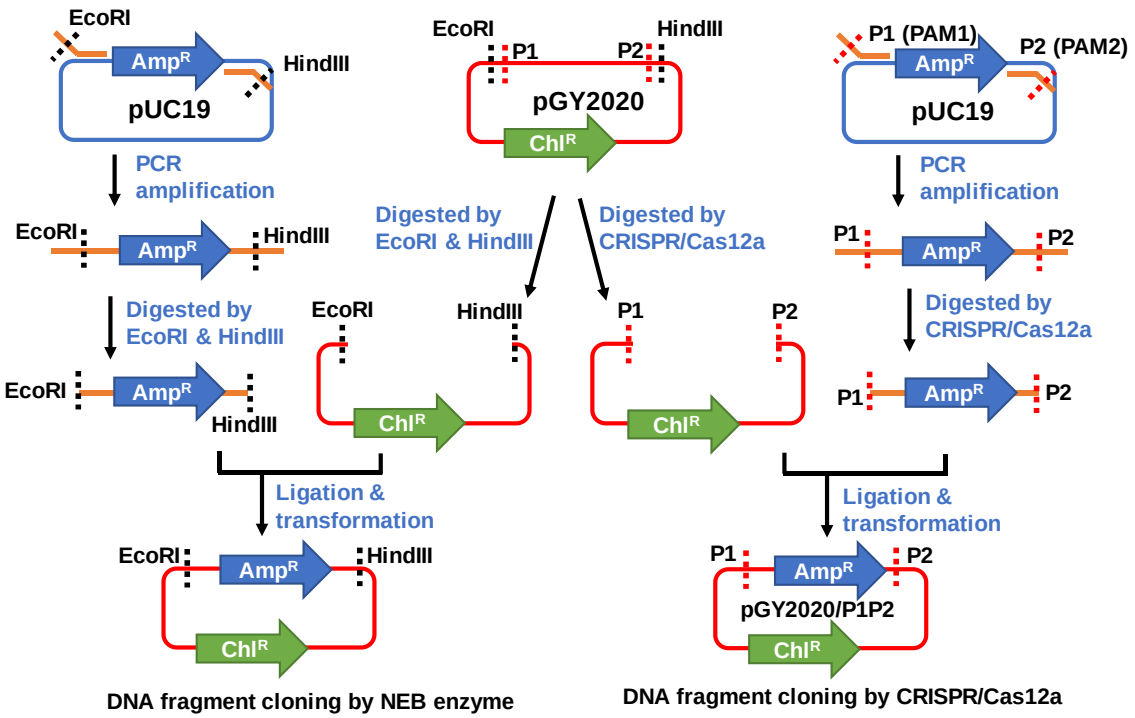
*To whom correspondence should be addressed. Tel/Fax: +86 21 64253020; Email: tangy@ecust.edu.cn. Correspondence may also be addressed to Tel/Fax: +86 21 64252575, Email: lxzhang@ecust.edu.cn
†These authors contributed equally to the paper as first authors. The authors wish it to be known that, in their opinion, the first 4 authors should be regarded as joint First Authors.

This file includes:

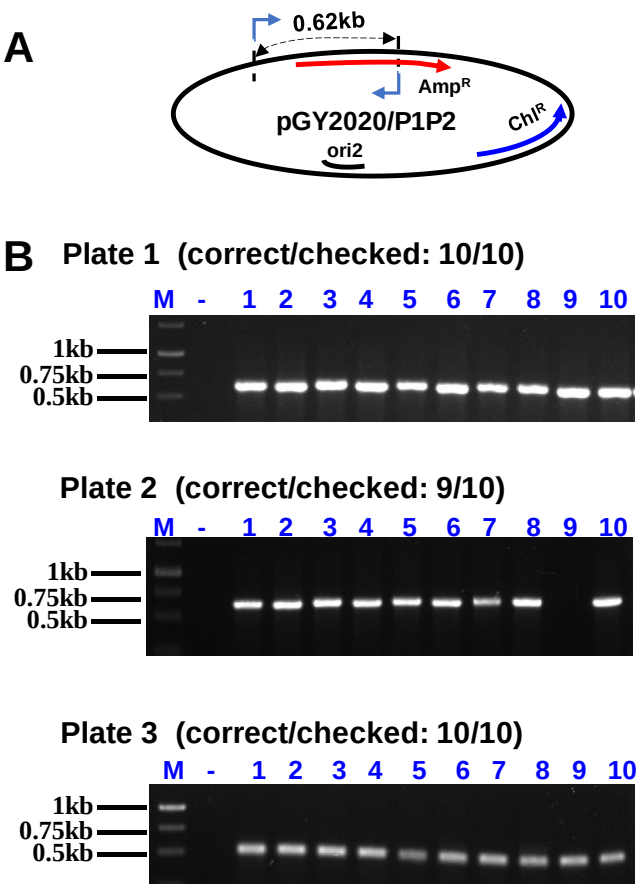
Supplementary Figures S1 to S7
Supplementary Tables S1 to S5
SI References



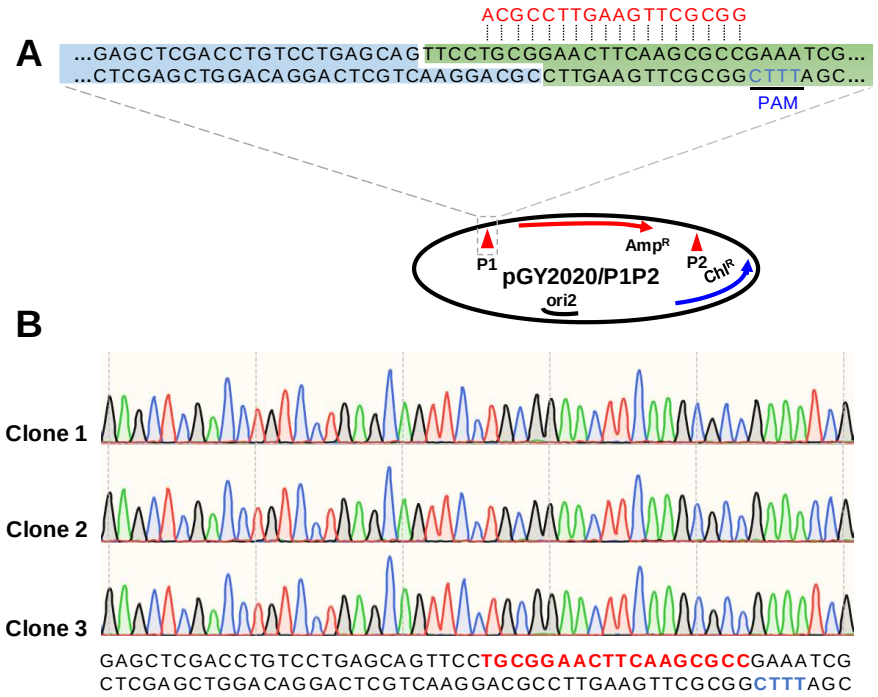
Supplementary Figure S1. Statistical analysis of GC content and length of characterized BGCs in the MIBiG database. A-B. Distribution of BGCs length and GC content in characterized gene clusters identified in the MIBiG database. C-D. Distribution of BGCs length and GC content in characterized gene clusters identified in Streptomyces. 1910 characterized BGCs were download from MIBiG database (https://mibig.secondarymetabolites.org/stats)



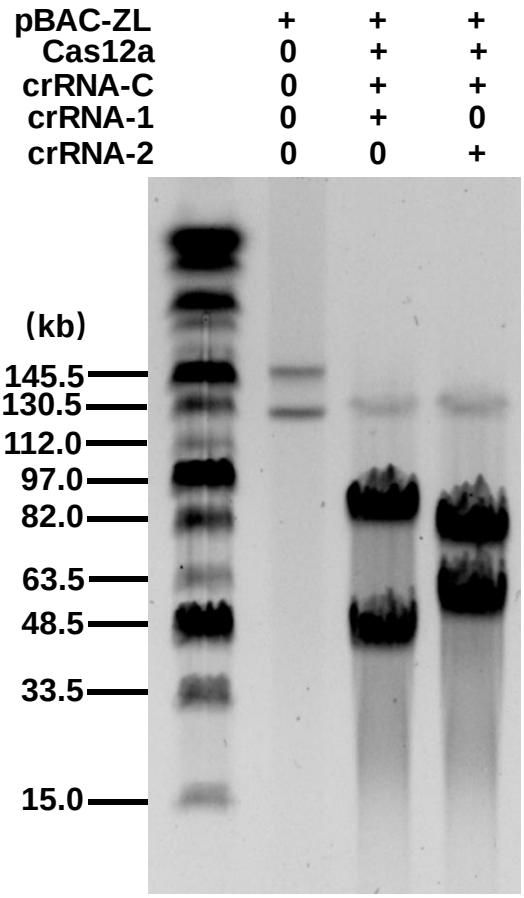
Supplementary Figure S2. Workflow of DNA fragment cloning by the CRISPR/Cas12a or NEB restriction enzymes-based method.



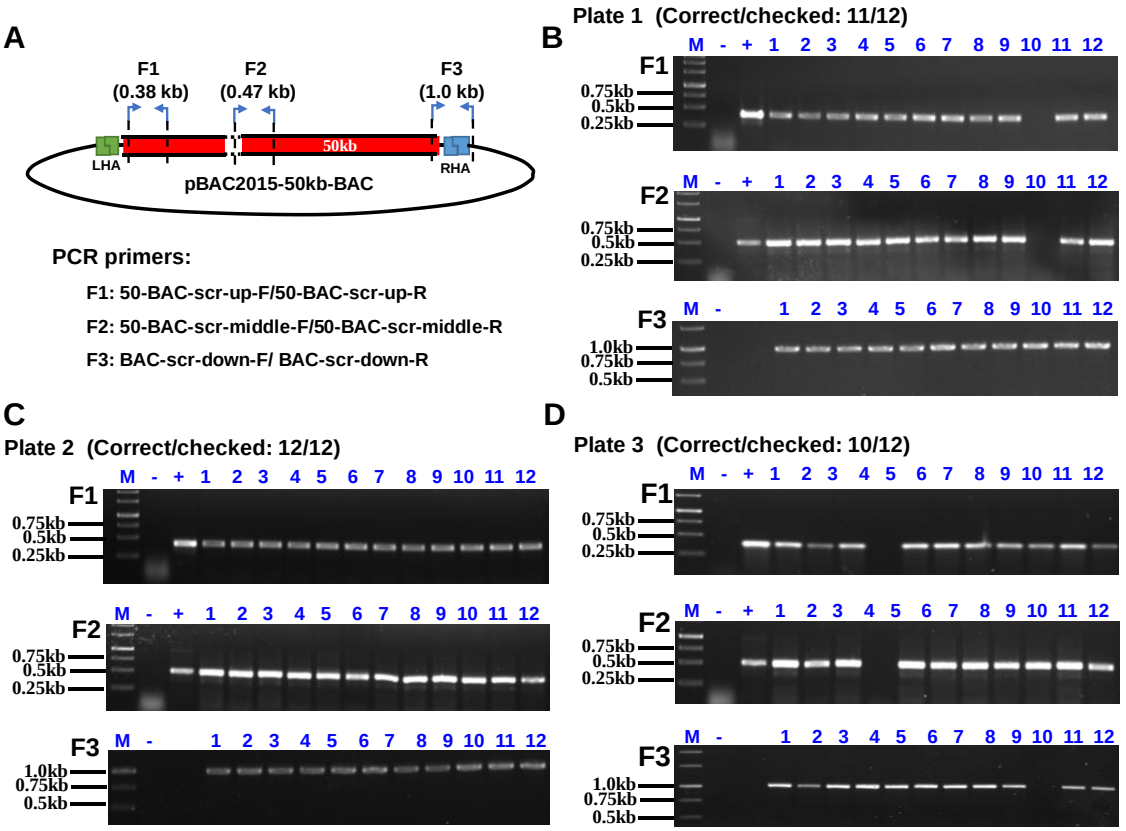
Supplementary Figure S3. Screening of right clones by PCR amplification. A. Schematic diagram of PCR screening of right clones with primers Amp-Cas12a-scr-F/R. B. PCR amplification of ten randomly selected clones. All experiments were performed in triplicate. "-" represented blank control, genomic DNA of E. coli DH10B was used as PCR template.



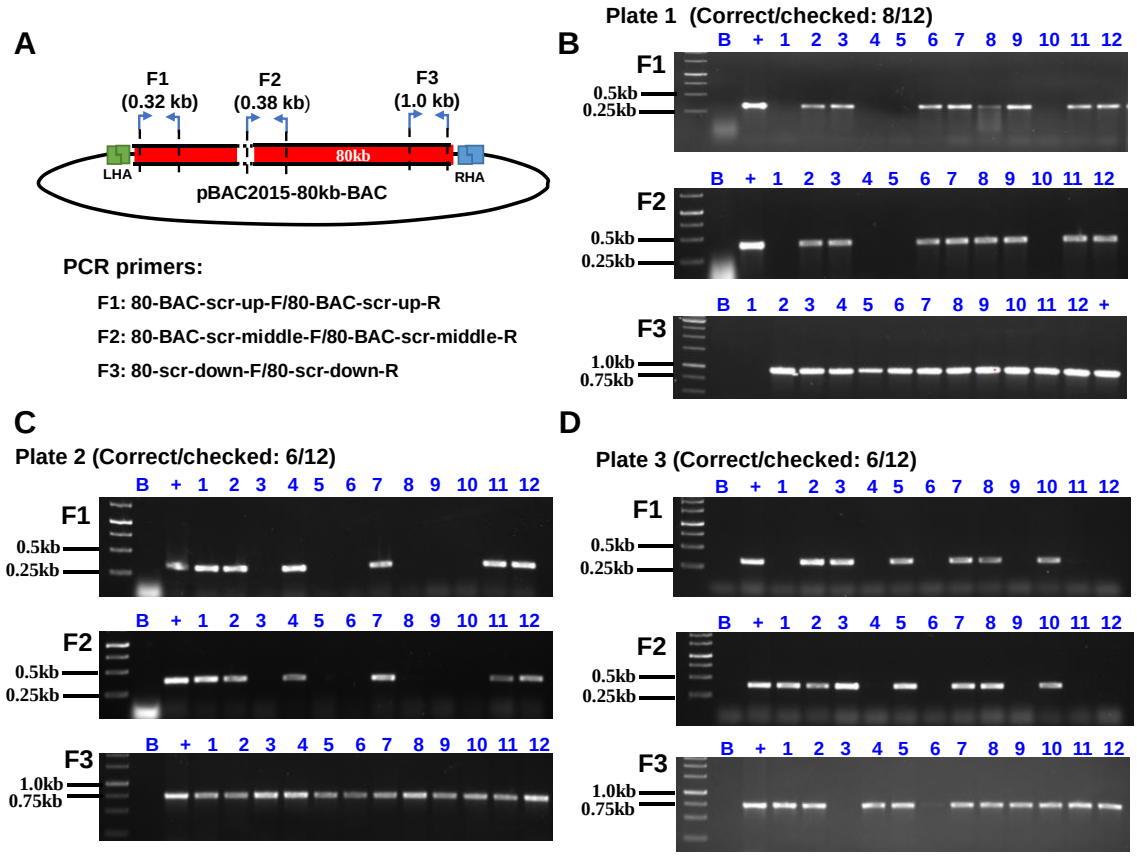
Supplementary Figure S4. Verification of Cas12a-mediated cohesive end ligation by DNA sequencing. A. Schematic representation of crRNA design and cohesive end ligation. 18-nt spacer crRNAs were employed, Cas12a mainly cleavage after the 14th base, generating 8-nt cohesive ends (1). B. Conformation of cohesive end ligation by junction sequencing. Three clones were randomly selected for sequencing.



Supplementary Figure S5. CRISPR/Cas12a-digested BAC plasmid (pBAC-ZL) analysed by PFGE. PFGE was performed in 0.5% agarose at 6 V/cm with a 1 ~ 25 sec switching pulse time for 16 ~ 18 h in 0.5 × TBE buffer.



Supplementary Figure S6. Schematic representation of crRNA design and cohesive end ligation. F1, F2 and F3 are the PCR products that were amplified using 50-BAC-scr-up-F/R, 50-BAC-scr-middle-F/R and BAC-scr-down-F/R, respectively. Primer sequences are given in Supplementary Table 5. B, C and D. PCR screening of 12 randomly selected clones containing pBAC2015-50kb-BAC in three independent experiments.



Supplementary Figure S7. Screening of right clones that containing pBAC2015-80kb-BAC by PCR amplification. A. Schematic representation of crRNA design and cohesive end ligation. F1, F2 and F3 are the PCR products that were amplified using 80-BAC-scr-up-F/R, 80-BAC-scr-middle-F/R and 80-scr-down-F/R, respectively. Primer sequences are given in Supplementary Table 5. B, C and D. PCR screening of 12 randomly selected clones containing pBAC2015-80kb-BAC in three independent experiments.

Supplementary Table S1. Comparison of different protocols of large DNA fragment cloning.

Method	Procedure	Timing	Cloning performance	References
CATCH*	sgRNA template preparation and in vitro transcription Agarose plug preparation and in-gel cell lysis In-gel Ca9 digestion Preparation of cloning vectors Ligation and electrotransformation Validation of positive clones	6 h 2 d 3 h 4 h 2 h 1 d In total: 3 ~ 4 d	Strain: *Bacillus. subtilis* GC% of genome: 45% Target fragment: unknown PKS DNA fragment length: 78 kb GC% of fragment: 46% Cloning efficiency: 12% ULCC: 150 kb (GC content: 51%)	(2,3)
RecET* (ExoCET)	Preparation of the E. coli strain Preparation of linear vectors for direct cloning Preparation of cassettes and vectors for engineering Restriction digestion of genomic to release the target gene cluster Preparation of the overnight E. coli culture Preparation of electro-competent E. coli cells Electroporation of genomic and the linear vector into E. coli cells Setting up cell cultures for restriction analysis Screening of correct recombinants by restriction analysis	2 d 1 ~ 7 d 1 d 5 – 6 h 12 h (overnight) 3 – 3.5 h 1.5 h 12 h (overnight) 7–9 h In total: > one week	Strain: *S. albus* DSM4139 GC% of genome: 72.6% Target BGC: salinomycin BGC length: 106 kb GC% of BGC: 74% Cloning efficiency: 4 ~ 8 % ULCC: 106 kb (GC content: 74%)	(4,5)
CAT-FISHING	crRNA template preparation and in vitro transcription Genomic DNA isolation and in-gel Cas12a digestion Plasmid construction and Cas12a digestion Ligation and transformation Validation of positive clones	~ 6 h ~ 1 d ~ 1 d 2 h 1 d In total: 3 ~ 4 d	Strain: *S. albus* J1074 GC% of genome: 73.3% Target BGC: surugamide BGC length: 87 kb GC% of BGC: 76% Cloning efficiency: ~ 4% ULCC: 139 kb (GC content: 75%)	This study

ULCC: upper limit of cloning capacity; BGC: biosynthetic gene cluster
* The description of procedures adapted from Nature Protocol(3,4).

Supplementary Table S2. The strains and plasmids used in this study.

Plasmid or strains	Features	Source/Ref.
Plasmid		
pUC19	High copy number cloning vector that conveys the Amp resistance	NEB
pCC2FOS	Fosmid vector, used for fosmid library construction.	EpicentreBio
pET28a	IPTG-inducible expression vector, used for recombination protein expression	Novagen
pSET152	Site-specific integrating vector; Containing the *ΦC31* attachment site and *oriT*, and can be transferred by conjugation into *Streptomycetes* from *E. coli*	(6)
pBAC2015	Copy-control BAC vector without repetitive sequences.	(4)
pUB307	RPI-derived self-transmissible plasmid, providing *in trans* the function for the mobilization of the *oriT*-containing BAC plasmid	(7)
pSC101-BAD-αβγA-tet	Redαβγ expression plasmid for linear plus circular homologous recombination	(8)
pBAC-ZL	BAC plasmid pIndigoBAC536-S containing 137kb DNA fragment	This study
pGY2020	Amp resistance gene in pCC2FOS	This study
pBAC2015-C50	pBAC2015 derived plasmid, used for 50kb DNA fragment (from pBAC-ZL) cloning	This study
pBAC2015-C80	pBAC2015 derived plasmid, used for 80kb DNA fragment (from pBAC-ZL) cloning	This study
pBAC2015-CS49	pBAC2015 derived plasmid, used for 49kb paulomycin biosynthetic gene cluster cloning	This study
pBAC2015-CS87	pBAC2015 derived plasmid, used for 87kb paulomycin biosynthetic gene cluster cloning	This study
pBAC2015-50kb-BAC	pBAC2015-C50 containing the 50kb DNA fragment (from pBAC-ZL)	This study
pBAC2015-80kb-BAC	pBAC2015-C80 containing the 80kb DNA fragment (from pBAC-ZL)	This study
pBAC2015-49kb-J1074	pBAC2015-CS49 containing the whole paulomycin biosynthetic gene cluster (49kb)	This study
pBAC2015-87kb-J1074	pBAC2015-CS87 containing the whole surugamides biosynthetic gene cluster (87kb)	This study
pBAC2015-49kb-J1074-int	The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette from pSET152 was introduced into the pBAC2015-49kb-J1074 by replacement of chloramphenicol-resistant (cmr) gene	This study
pBAC2015-87kb-J1074-int	The *aac(3)IV-oriT-attP(ΦC31)-int(ΦC31)* cassette from pSET152 was introduced into the pBAC2015-87kb-J1074 by replacement of chloramphenicol-resistant (cmr) gene	This study

Supplementary Table S2. The strains and plasmids used in this study (continued).

Plasmid or strains	Features	Source/Ref.
E. coli		
EPI300™-T1R	F– *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) (StrR) φ80d*lac*Z ΔM15 Δ*lacX*74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G *trf*A *ton*A *dhfr*	EpicentreBio
DH10B	F– *mcrA* ∆(*mrr-hsd*RMS-*mcrBC*) φ80d*lac*Z ∆M15 ∆lacX74 *deo*R *rec*A1 *end*A1 *ara*D139 ∆(*ara*, *leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G	Gibco BRL
BL21(DE3)	F– *ompT hsdSB*(rB- mB-) *gal dcm* (DE3)	TIANGEN, Shangai
ET12567	F–*dam13*::Tn*9 dcm6 hsdM hsdR recF143 zjj201*::Tn*10 galK2 galT22ara14 lacY1 xyl5 leuB6 thi1 tonA31 rpsL136 hisG4 tsx78 mtli glnV44*	(9)
ET12567(pUB307)	ET12567 containing plasmid pUB307	This study
Streptomyces		
S. albus J1074	Wild-type strain, heterologous host	(10)
S. albus Del14	Derivative of J1074 with deletion of 15 gene cluster	(11)
S. albus Del14-49kb	BAC plasmid pBAC2015-49kb-M integrated into the genome of Del14	This study
S. albus Del14-87kb	BAC plasmid pBAC2015-87kb-M integrated into the genome of Del14	This study

Supplementary Table S3. The sequence of PCR primers for plasmid construction.

Name	Sequence (5' → 3')
pGY2020 construction	
pCC2-hom-arm-up-F	TACAACGACACCTAGACCACGGTGGAGCTGCGCAACCGGC
pCC2-hom-arm-up-R	ACAGACAAGCTGTGACCGTCTCCTGGGGCGTGACACCACC
pCC2-hom-arm-dn-F	GCGCGGGGAGAGGCGGTTTGGACATTGCACTCCACCGCTG
pCC2-hom-arm-dn-R	AGGAAACAGCCTAGGAACACGGAAGTTTTTTCAGGCATCG
pBAC2015-C50 construction	
50kb-BAC-arm-up-F	TTATCTATGCTCGGGGGATGCCGCGTGGTACC
50kb-BAC-arm-up-R	ACAGACAAGCTGTGACCGTCGGCGGCATCCCGATCAGCGC
BAC-arm-dn-F	GCGCGGGGAGAGGCGGTTTGGACATTGCACTCCACCGCTG
BAC-arm-dn-R	TAGAGAGGATACCGGAGATCCTTTGATCTTTTC
BAC-lacZ-F	GACGGTCACAGCTTGTCTGTAAGC
BAC-lacZ-R	CAAACCGCCTCTCCCCGCGCGTTGG
pBAC2015-C80 construction	
80kb-BAC-arm-up-F	TTATCTATGCTCGGGGGTGGAGCTGCGCAACCG
80kb-BAC-arm-up-R	ACAGACAAGCTGTGACCGTCTCCTGGGGCGTGACACCACC
BAC-lacZ-F	GACGGTCACAGCTTGTCTGTAAGC
BAC-lacZ-R	CAAACCGCCTCTCCCCGCGCGTTGG
BAC-arm-dn-F	GCGCGGGGAGAGGCGGTTTGGACATTGCACTCCACCGCTG
BAC-arm-dn-R	TAGAGAGGATACCGGAGATCCTTTGATCTTTTC
pBAC2015-CS49 construction	
50kb-up-hom-arm-F	CACTCACTCACCCCGGTCACATCGTTATCTATGCTCGGCGATCGCGCTGGAGTCCTTCG
50kb-up-hom-arm-R	TTTGCGTATTGGGCAATTCTCATGTTTGACCGCTTATCTGGATCTGCCCTTTCCACTC
50kb-lacZ-bla-F	ATAAGCGGTCAAACATGAGA
50kb-lacZ-bla-R	AATGAAGTTTTAAATCAATC
50kb-dn-hom-arm-F	TTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTAATCCTGTCGACCGTCAAG
50kb-dn-hom-arm-R	AGGAAATTTATCTTGATCATATAAATAGAGAGGATACCGTGACCAAGCGGTACGACTTC
pBAC2015-CS87 construction	
80kb-up-hom-arm-F	CATCGTTATCTATGCTCGGGGATGTCGACGTCCAGGGTG
80kb-up-hom-arm-R	CGGTTTGCGTATTGGGCAATTGATTCGCCGGCGTTCTG
80kb-lacZ-F	ATTGCCCAATACGCAAAC
80kb-lacZ-R	CTAAGAAACCATTATTATC
80kb-dn-hom-arm-F	TGATAATAATGGTTTCTTAGGGAAGCGGTCTCCTGAAGC
80kb-dn-hom-arm-R	ATAAATAGAGAGGATACCGGTTTTCCCCCGTTGATGAGTGG

Supplementary Table S4. The sequence of crRNA for CRISPR/Cas12a cleavage.

Name	Sequence (5' → 3')	Cleavage
T7-promoter-F	GAAATTAATACGACTCACTATAGGG	
50kb-BAC-up-crRNA-R	GTGCAGACCCGGATTCCGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	50kb BAC
80kb-BAC-dn-rRNA-R	CGATCAACGGCACTGTTGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	50kb BAC
80kb-BAC-up-crRNA-R	TGCGGAACTTCAAGCGCCATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	80kb BAC
80kb-BAC-dn-crRNA-R	CGATCAACGGCACTGTTGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	80kb BAC
50kb-up-crRNA-R	GTACGCGGGCAGCGTGAGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	49kb BGC
50kb-dn-crRNA-R	GGACAGAGATTTCGCCAAATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	49kb BGC
80-up-crRNA-R	CCTGGCCGCGCCCGCCCGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	87kb BGC
80kb-dn-crRNA-R	CACAGTTTCGGTCCTCGGATCTACAACAGTAGAAATTCCCTATAGTGAGTCGTATTAAT	87kb BGC

Supplementary Table S5. The sequence of PCR primers for screening, verification and modification

Name	Sequence (5' → 3')	Note
Amp-Cas12a-scr-F	GGGAACCGGAGCTGAATGAA	pGY2020-P1_P2
Amp-Cas12a-scr-R	CTGACGCTCAGTGGAACGAA	pGY2020-P1_P2
50-BAC-scr-up-F	GGCAGTTTCATCGTGGCGTA	50kb BAC-Ch-U
50-BAC-scr-up-R	GCGGGACTCACATGGGTTTT	50kb BAC-Ch-U
50-BAC-scr-mid-F	CAATCCGATGACACGGCACA	50kb BAC-Ch-M
50-BAC-scr-mid-R	CCGTGGTTGTCGCTGTACTC	50kb BAC-Ch-M
BAC-scr-down-F	TGCCATCAACTCGGCAAGAT	50kb BAC-Ch-D
BAC-scr-down-R	TCTTTCGCGAAGGCTTGAGT	50kb BAC-Ch-D
80-BAC-scr-up-F	TGGTGTGTTGTCGTTGTCGG	80kb BAC-Ch-U
80-BAC-scr-up-R	GGAGATCTGGGCGAACTCCT	80kb BAC-Ch-U
80-BAC-scr-mid-F	GGCAGTTTCATCGTGGCGTA	80kb BAC-Ch-M
80-BAC-scr-mid-R	GCGGGACTCACATGGGTTTT	80kb BAC-Ch-M
BAC-scr-down-F	TGCCATCAACTCGGCAAGAT	80kb BAC-Ch-D
BAC-scr-down-R	TCTTTCGCGAAGGCTTGAGT	80kb BAC-Ch-D
50-scr-up-F	TCCATTCCGTGCCATGCG	49kb BGC-Ch-U
50-scr-up-R	GCGACGAGAGAGGATGTG	49kb BGC-Ch-U
50-scr-middle-F	ACAAGGCTCCTGACAGG	49kb BGC-Ch-M
50-scr-middle-R	GAGGTGGTGCACCTGG	49kb BGC-Ch-M
50-scr-down-F	GTGTATCGCGCCGCTG	49kb BGC-Ch-D
50-scr-down-R	CTATGCTCCAGACATC	49kb BGC-Ch-D
80-scr-up-F	GCAGCGAACTGCCTGGT	87kb BGC-Ch-U
80-scr-up-R	CCGATGAGGTCGTTCAC	87kb BGC-Ch-U
80-scr-middle-F	ATGCCCGTCAACTGCTCCTG	87kb BGC-Ch-M
80-scr-middle-R	TGATTTCCCGACCGTTT	87kb BGC-Ch-M
80-scr-down-F	GCCGGTTCAGGCGCGCT	87kb BGC-Ch-D
80-scr-down-R	AGAAGGGCAAGTTGTGC	87kb BGC-Ch-D
Redαβ-scr-F	ACGTATGTGGTGTGACCGGA	49kb BGC-Red
Redαβ-scr-R	CATTGCACTCCACCGCTGAT	49kb BGC-Red
Redαβ-scr-F	ACGTATGTGGTGTGACCGGA	87kb BGC- Red
Redαβ-scr-R	CATTGCACTCCACCGCTGAT	87kb BGC- Red

SI References

1. 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-e74.
2. 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. Commu.*, **6**, 8101.
3. Jiang, W. and Zhu, T.F.J.N.P. (2016) Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc.,* **11**, 960-975.
4. Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A.F., Fu, J. *et al.* (2016) RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.*, **11**, 1175.
5. Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A.F. *et al.* (2017) ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.*, **46**, 2697-2697.
6. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Nagaraja Rao, R. and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to Streptomyces spp. *Gene*, **116**, 43-49.
7. 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 Microb. Lett.*, **155**, 223-229.
8. Zhang, Y., Muyrers, J.P.P., Testa, G. and Stewart, A.F. (2000) DNA cloning by homologous recombination in Escherichia coli. *Nat. Biotechnol.*, **18**, 1314-1317.
9. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, **111**, 61-68.
10. Chater, K.F. and Wilde, L.C. (1980) Streptomyces albus G Mutants Defective in the SalGI Restriction-Modification System. *J. Gen. Microbiol.*, **116**, 323-334.
11. 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.