Supplementary information

Engineered CRISPR-Cas9 for *Streptomyces* sp. genome editing to improve specialized metabolite production

Duck Gyun Kim[1], Boncheol Gu[1], Yujin Cha[1], Jeonghan Ha[1], Yongjae Lee[2], Gahyeon Kim[2], Byung-Kwan Cho[2,3,4], & Min-Kyu Oh[1]*

[1]Department of Chemical & Biological Engineering, Korea University, Seoul 02841, Republic of Korea

[2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[3]KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[4]Graduate School of Engineering Biology, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

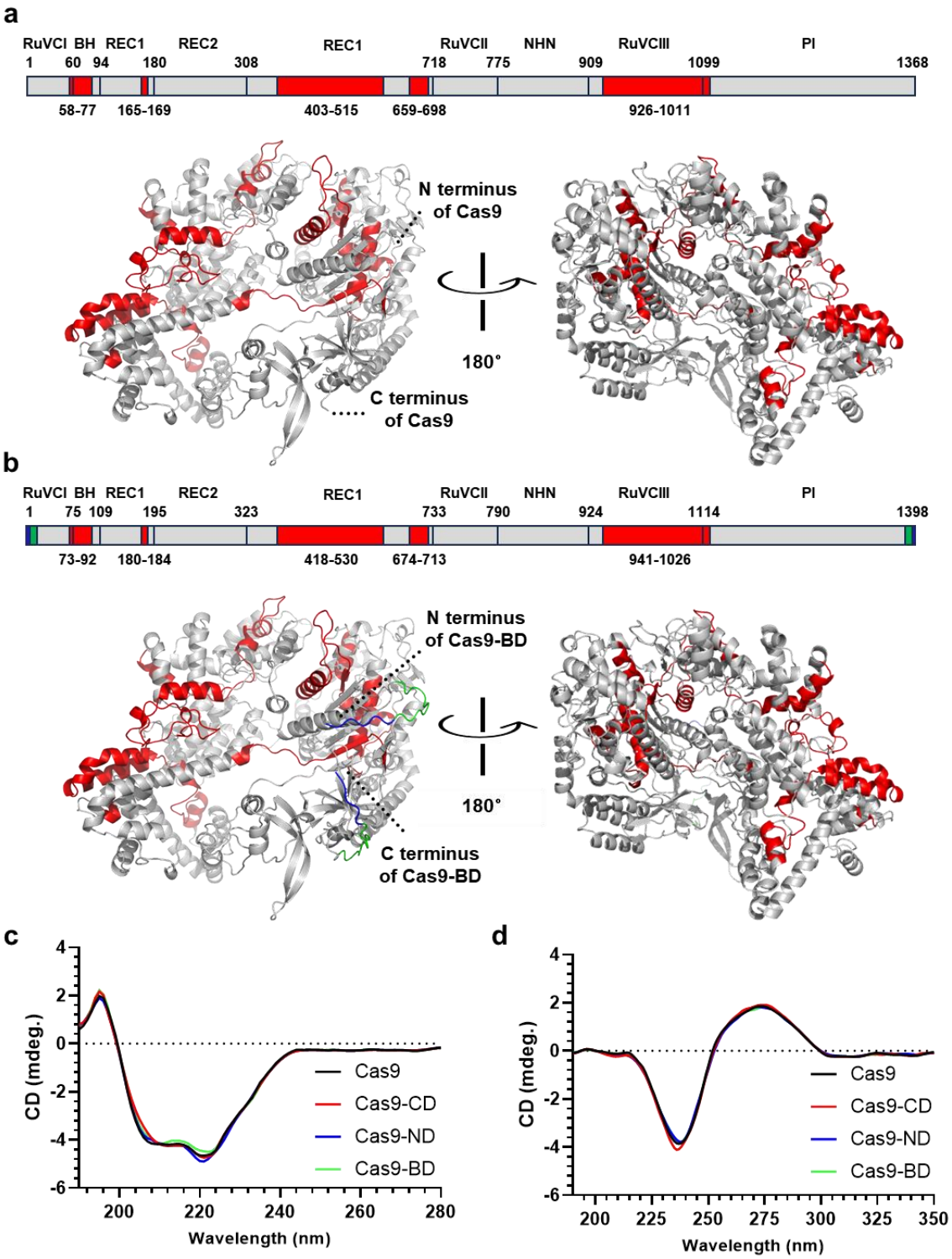
***Correspondence**:

Min-Kyu Oh,

Department of Chemical and Biological Engineering, Korea University, Anam-Ro 145 Seongbuk-Gu, Seoul 02841, Korea.

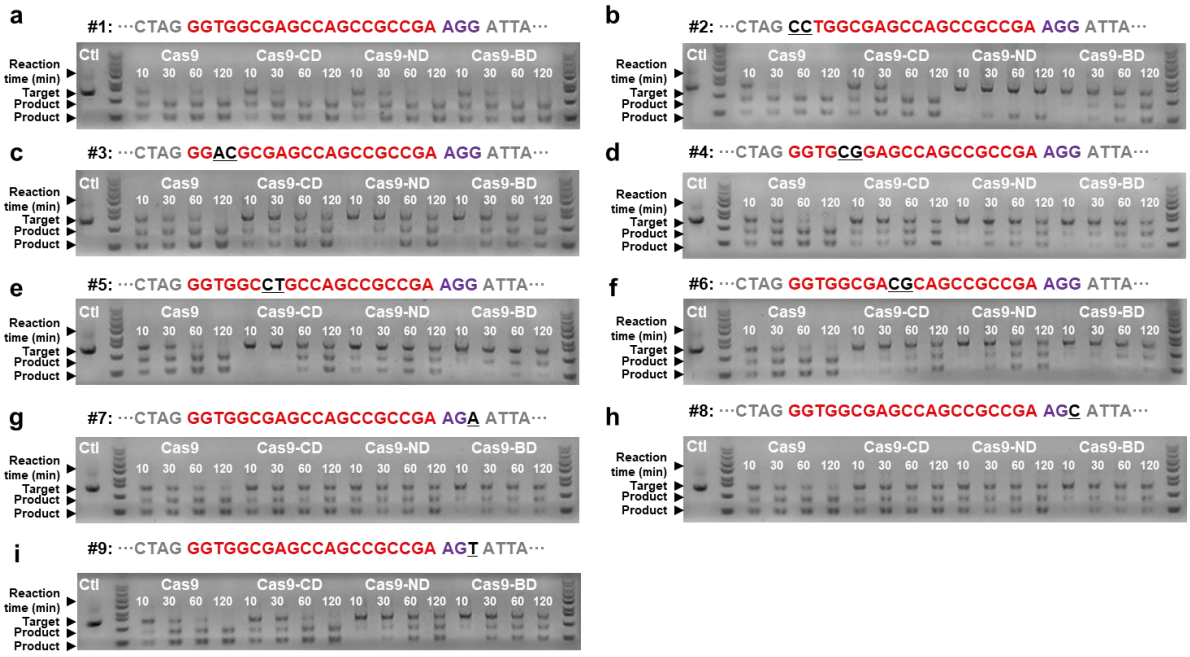
E-mail: mkoh@korea.ac.kr

Tel: +82-2-3290-3308

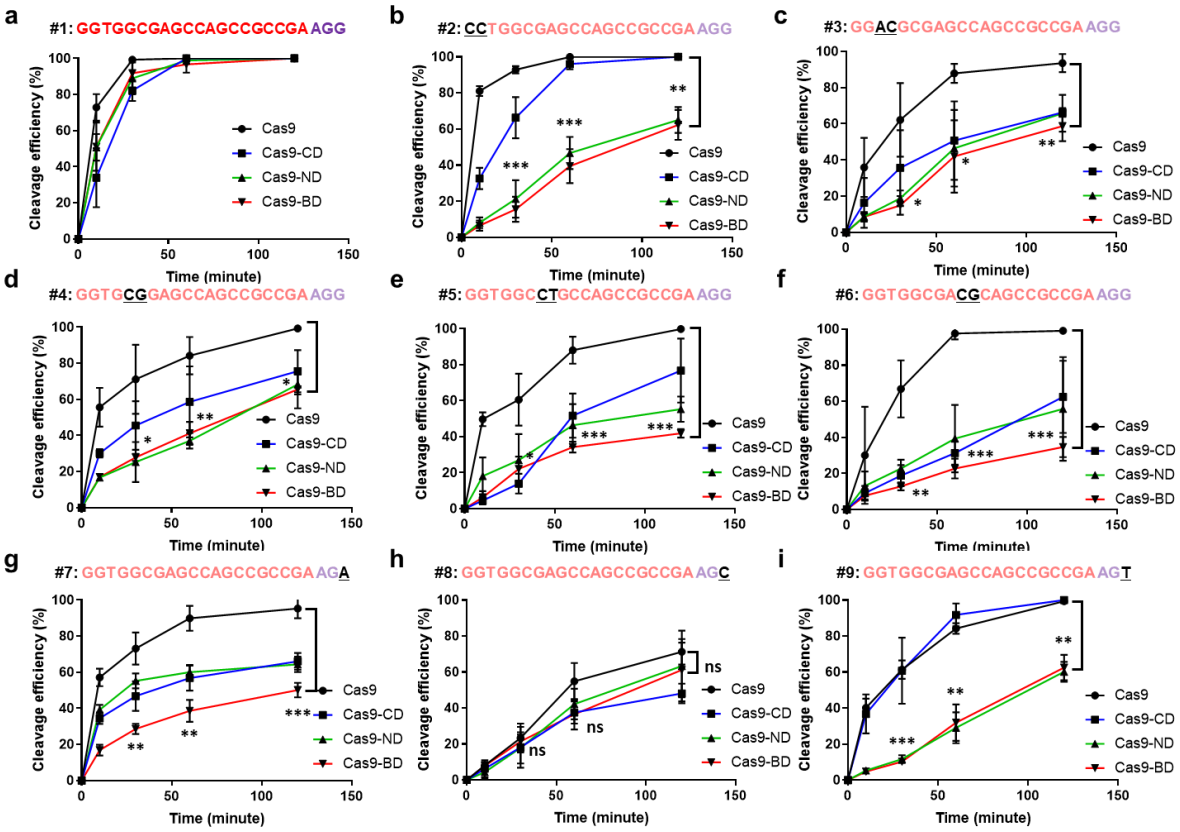


Supplementary Fig. 1. The structural analysis of Cas9 originated from *Streptococcus pyogenes* and modified Cas9s. a. Organizations and structures of Cas9 from *S. pyogenes* and **b.** modified Cas9-BD predicted by using AlphaFold3 (https://www.alphafoldserver.com)[1]. The predicted structures were visualized using PyMOL[2]. The SpCas9 was colored gray and regions in active site containing residues for recognition of binding between target DNA and sgRNA[3] was red. The polyaspartate was colored blue and glycine-serin linker was green. **c.** The circular dichroism (CD) spectroscopy analysis of purified proteins including Cas9 and modified Cas9s, and **d.** protein-sgRNA complexes. Cas9, wild-type Cas9

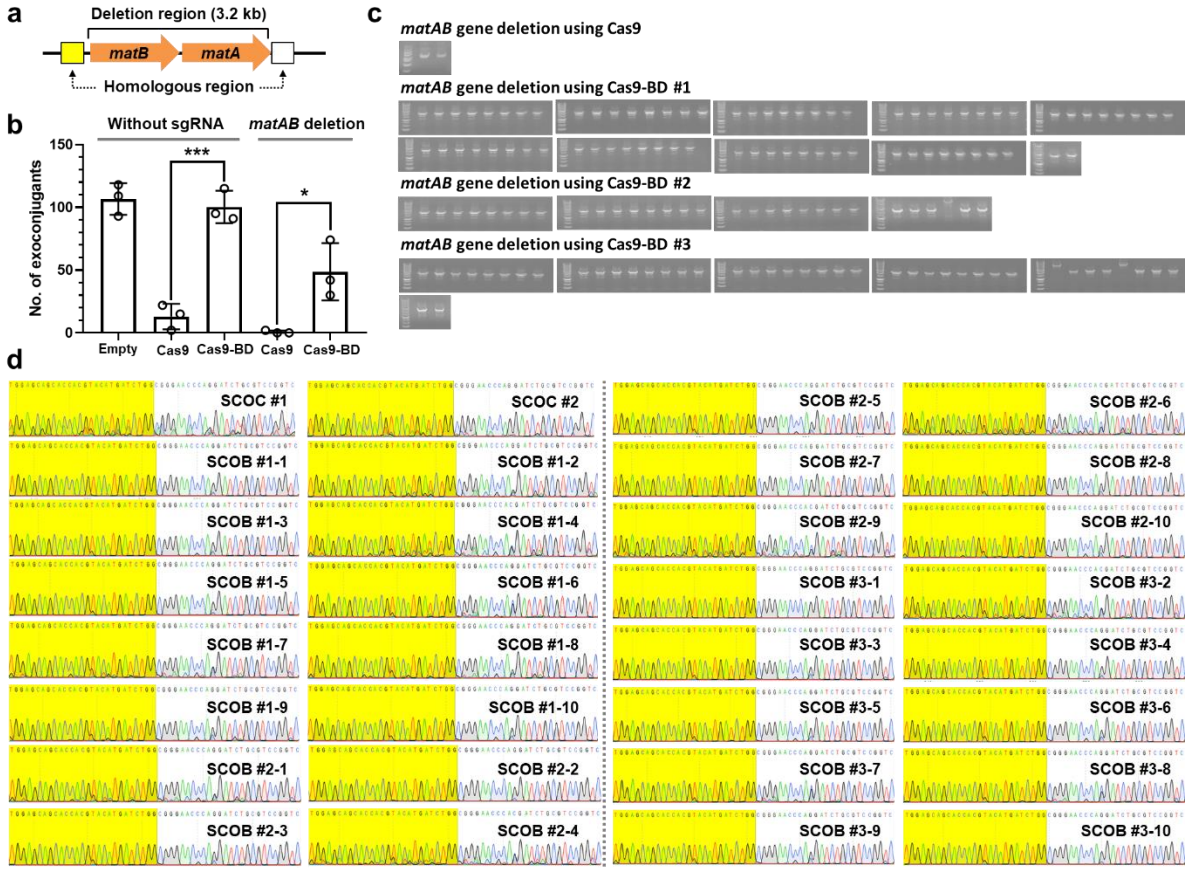
originated from *S. pyogenes*; Cas9-CD, Cas9 linked with polyaspartate at C terminus; Cas9-ND, Cas9 linked with polyaspartate at N terminus; Cas9-BD, Cas9 linked with poly-aspartate at both of N and C termini. Source data are provided as a Source Data file.



Supplementary Fig. 2. The visualized products of *in vitro* cleavage reaction using linearized DNA and various Cas9s. The time-series results of *in vitro* reaction using various Cas9-sgRNA complexes with various DNAs, including one with on-target sequence (**a**), off-target sequence (**b, c, d, e, f**), and non-PAM sequence (**g, h, i**). The DNA sequences used in the experiment are presented above each figure with underlined mismatch sequences in case of off-target and non-PAM subsrtates. Ctl, DNA after reaction without any Cas9; Cas9, Cas9-CD, Cas9-ND, and Cas9-BD, DNA after reaction with the corresponding protein-sgRNA complexes; 10, 30, 60, 120, the reaction time (min). Source data are provided as a Source Data file.



Supplementary Fig. 3. The calculated cleavage efficiencies of various Cas9s with each substrate. The results calculated via visualized products (Supplementary Fig. 2) of *the in vitro* reaction using various Cas9s with various substrates, including on-target sequences (**a**), off-target sequences (**b, c, d, e, f**), and non-PAM sequences (**g, h, i**). The sequences of the substrates are presented above each figure. Reactions were performed in triplicates (n = 3). The products were visualized by a transilluminator and the intensities of DNA fragment were measured using scanned image. The mean of the intensity ratios was plotted with error bar representing standard deviation. P values were determined by unpaired two-tailed Student's t-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Source data are provided as a Source Data file.



e

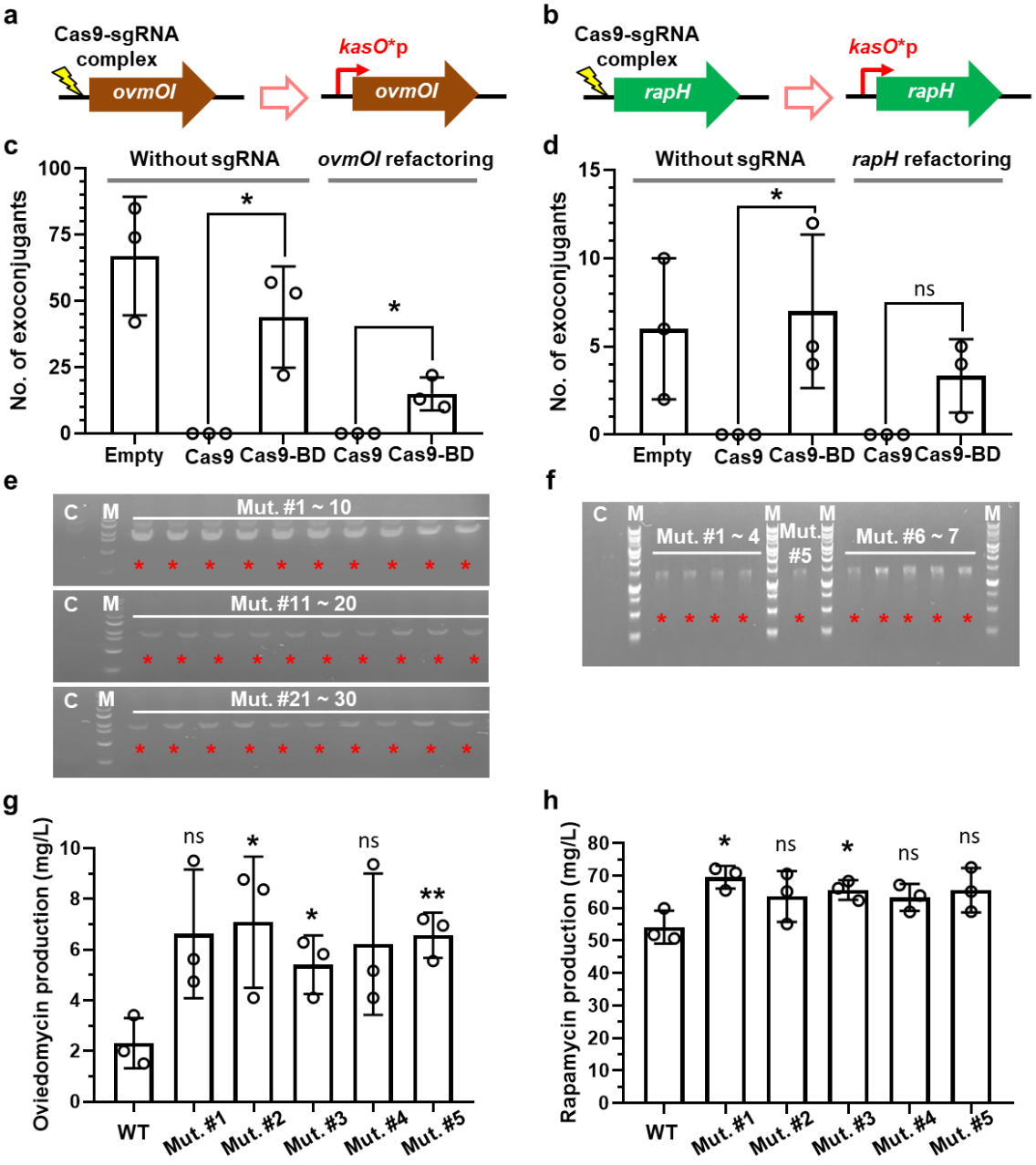
Strain	Cas9 #1	Cas9 #2	Cas9-BD #1	Cas9-BD #2
Raw reads	10,347,834	13,441,601	16,383,244	15,692,416
Mapped reads	10,032,879	13,086,429	15,926,895	15,167,112
Uniquely mapped reads	9,841,404	12,835,690	15,617,107	14,876,032
Average read length	51	51	51	51
Coverage	57.91	75.53	91.89	87.53

f

Type	Position	Reference	Mutation	Frequency (%)				Gene (Protein)	Change
				Cas9 #1	Cas9 #2	Cas9-BD #1	Cas9-BD #2		
SNV	2,411,510	G	A	0	93.75	0	0	SCO2241 (Glutamate synthetase)	Missense (G363D)
SNV	3,259,197	G	A	0	93.75	0	0	SCO2990 (Hypothetical protein)	Missense (A181G)
SNV	4,979,273	G	T	100	97.96	0	0	SCO4632 (ATP/GTP binding protein)	Missense (Q740H)
SNV	6,249,097	T	A	0	0	94.44	98.33	SCO5802 (ATP-dependent helicase)	Missense (W431R)

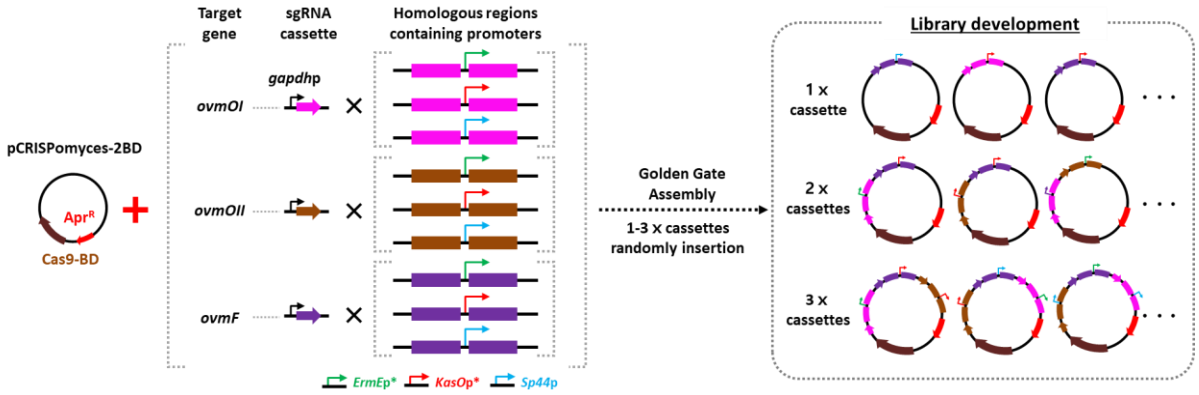
Supplementary Fig. 4. The *matAB* gene deletion in *Streptomyces coelicolor* M1146 strain using Cas9 and Cas9-BD. a. The edited region and homologous sequence regions (yellow and white boxes) for repair after *matAB* gene deletion in the *S. coelicolor* genome. **b.** The number of exconjugants for each conjugation experiment. Empty, conjugation with pCRISPomyces-2 plasmid lacking gene coded Cas9 protein; Cas9, conjugation with pCRISPomyces-2 plasmid; Cas9-BD, conjugation with pCRISPomyces-2BD plasmid. Conjugations were performed in triplicates individually (n = 3). The mean was plotted with error bar representing standard deviation. P values were determined by unpaired two-tailed Student's t-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). **c.** PCR amplification and **d.** sequence analysis of exconjugants to confirm editing efficiency. **e.** Summary of whole genome sequencing performed with two exconjugants each after the editing experiments with Cas9 and Cas9-BD. **f.** The off-target

edited sites revealed by whole genome sequencing. Source data are provided as a Source Data file. SNV, single nucleotide variation.



Supplementary Fig. 5. BGC refactoring to improve production of secondary metabolites in *S. antibioticus* NRRL 3238 and *S. rapamycinicus* NRRL 5491. The scheme of experiments that exchanged promoters of *ovmOI* (**a**) and *rapH* genes (**b**) to *kasO** promoter using Cas9 or Cas9-BD with sgRNA. The numbers of exconjugants for *S. antibioticus* NRRL 3238 (**c**) and *S. rapamycinicus* NRRL 5491 (**d**) obtained with conjugation experiments. Empty, conjugation with pCRISPomyces-2 plasmid; All conjugations were performed in triplicate (*n* = 3). The refactoring of *ovmOI* (**e**) and *rapH* promoters (**f**) confirmed by results of PCR amplification. C, results of PCR amplification using genomic DNA of wild type *S. antibioticus* NRRL 3238 or *S. rapamycinicus* NRRL 5491 as template; M, DNA ladder marker; Mut. #n, results of PCR amplification using genomic DNA of exconjugants of refactoring experiments. Productions of oviedomycin (**g**) and rapamycin (**h**) of five exconjugants with flask cultivation. Cultivations were performed in triplicates separately (*n* = 3). The mean was plotted with error bar

representing standard deviation. *P* values were determined by unpaired two-tailed Student's *t*-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). Source data are provided as a Source Data file.

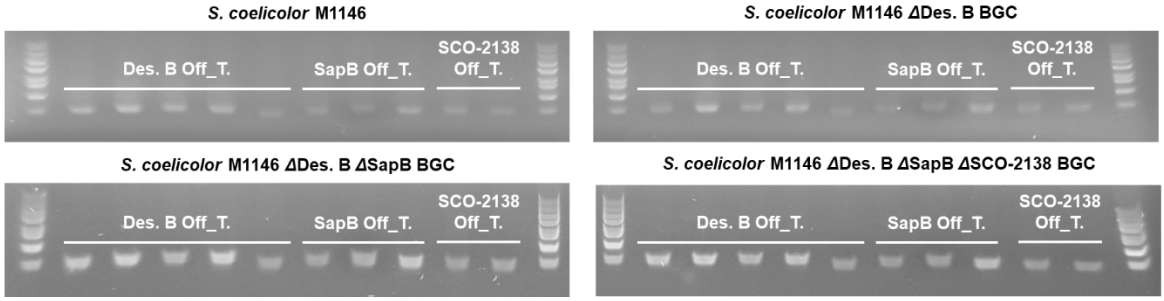


Supplementary Fig. 6. The scheme of plasmid library development for combinatorial replacement of multiple promoters in the oviedomycin BGC. The increased gene expression in the BGC was intended by replacing 1~3 promoters combinatorially by one conjugation experiment using Cas9-BD.

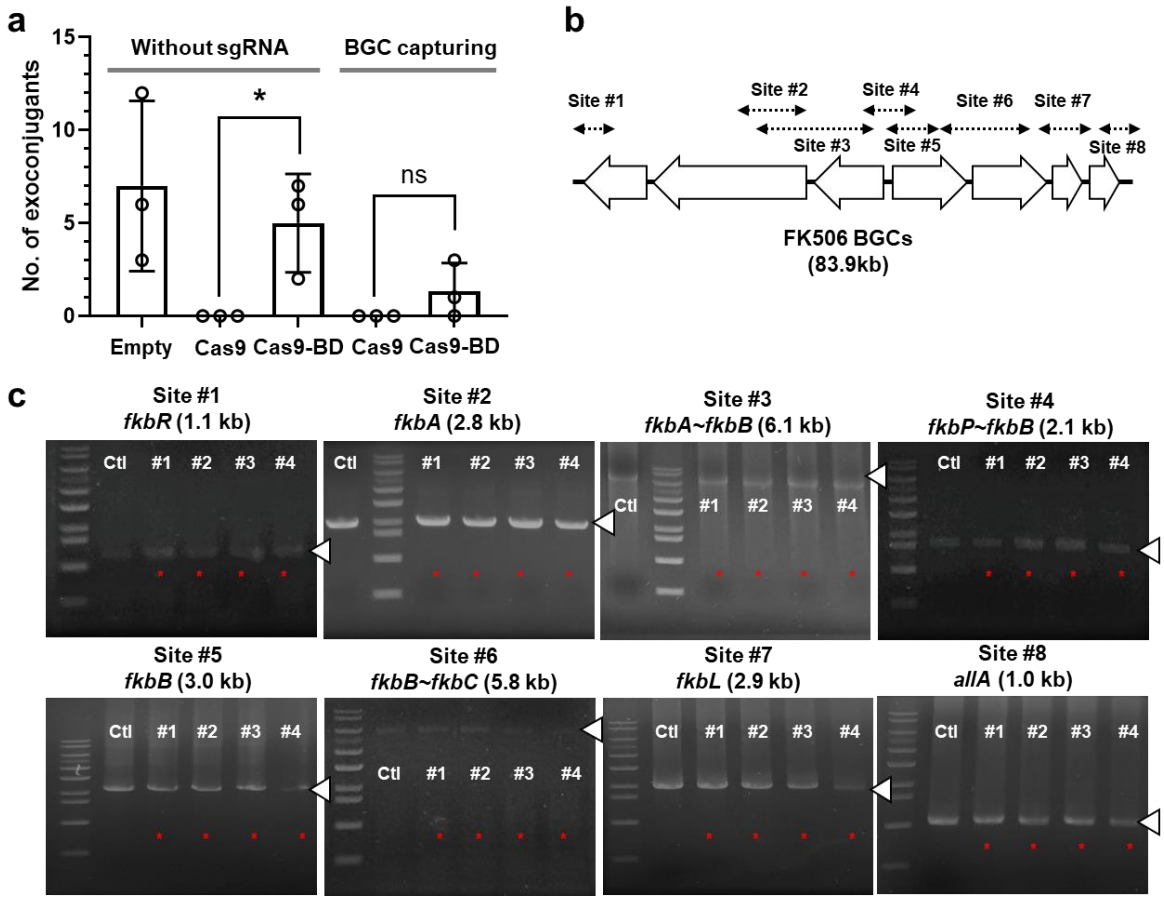
a

Site	Sequence (5` to 3`)	Location	Gene (Protein)
Des. B On-T. site	GCGGGATCGGTGACGTCCGC CGG	3,036,850 – 3,036,872	SCO2782 (Pyridoxal-dependent decarboxylase)
Des. B Off-T. site #1	TTGGTGACGGTGACGTCCGC GGT	1,260,337 – 1,260,359	SCO1187 (Cellulase B)
Des. B Off-T. site #2	GCGACGGCGGTGACGTCCGC CGT	3,146,344 – 3,146,366	SCO2892 (Hypothetical protein)
Des. B Off-T. site #3	GTGCAGCCGGTGACGTCCGC CGC	3,496,107 – 3,496,129	SCO3189 (Hypothetical protein)
Des. B Off-T. site #4	TCGGTGGCGGTGACGTCCGC GGC	7,160,014 – 7,160,036	SCO6611 (Hypothetical protein)
Des. B Off-T. site #5	ACGTGCGCGGTGACGTCCGC GGA	7,348,877- 7,348,899	SCO6764 (Squalene-hopene cyclase)
SapB On-T. site	AGGACTTCCAGGGCGCGCCC CGG	7,249,514 – 7,249,536	SCO6683 (ABC transporter ATP-binding protein)
SapB Off-T. site #1	CCGATGTCCAGGGCGCGCCC GGC	1,008,042 – 1,008,064	SCO0958 (Hypothetical protein)
SapB Off-T. site #2	GCGGCGCCCAGGGCGCGCCC AGT	2,019,049 – 2,019,071	SCO1885 (Oxidoreductase)
SapB Off-T. site #3	AGCACGTCCAGGGCGCGCCC GGG	5,497,248 – 5,497,270	SCO5146 (Methyltransferase)
SCO-2138 On-T. site	GACGGCTCCGACGTCGACCA GGG	2,626,970 – 2,626,992	SCO2446 (Peptidase)
SCO-2138 Off-T. site #1	GTCCAGGCCGACGTCGACCA GGA	2,095,876 – 2,095,898	SCO1958 (Excinuclease ABC subunit A)
SCO-2138 Off-T. site #2	GAGGCCGCCGACGTCGACCA GGC	5,134,826 – 5,134,848	SCO4792 (DNA-binding response regulator)

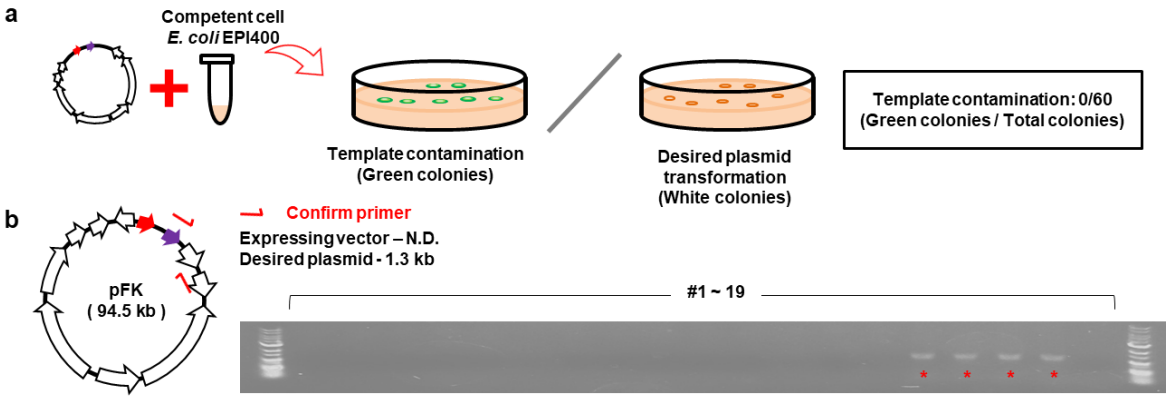
b



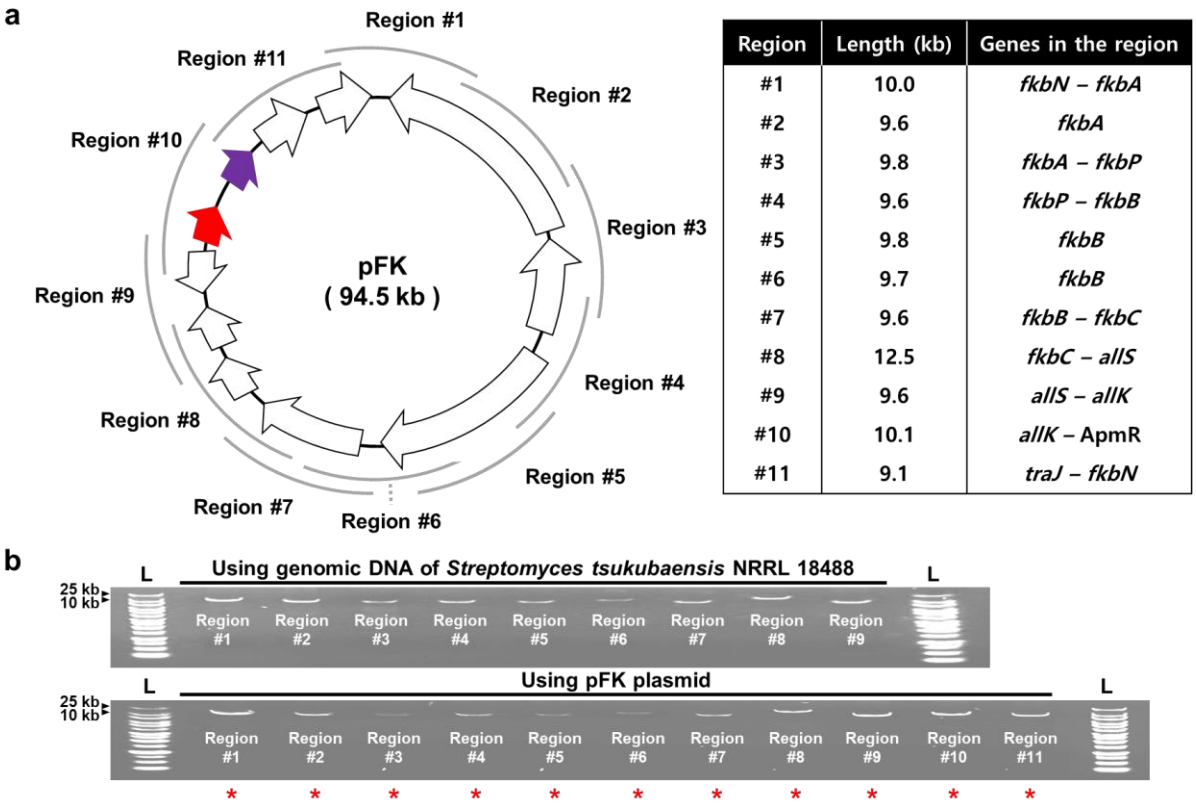
Supplementary Fig. 7. Prediction and confirmation of off-target sites during multiple BGC deletion experiment in *S. coelicolor* M1146. a. Predicted off-target sites based on sgRNA for each BGC deletion. Off-target sites including putative PAM sequence were predicted by using BLAST searches with NCBI database and RGEN OFFinder Tool[4]. **b.** Confirmation by PCR amplification of the off-target sites with a few exconjugants generated by multiple deletion using Cas9-BD. Sequencing of the PCR products were also used for confirmation. Source data are provided as a Source Data file.



Supplementary Fig. 8. In vivo capturing efficiency of FK506 BGC using Cas9-BD with *S. tukubaensis* NRRL 18488. a. Numbers of exconjugants after conjugation**.** Conjugations were performed in triplicates individually (n = 3). The mean was plotted with error bar representing standard error. P values were determined by two-tailed unpaired Student's t-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). **b.** Location of confirmation PCR products in FK506 BGC. Eight sites in the BGC were amplified using PCR. **c.** The PCR results of the eight sites using four purified capturing plasmids from in vivo capturing experiments. Ctl, PCR product using genomic DNA of *S. tsukubaensis* NRRL 18488 as a template. Source data are provided as a Source Data file.

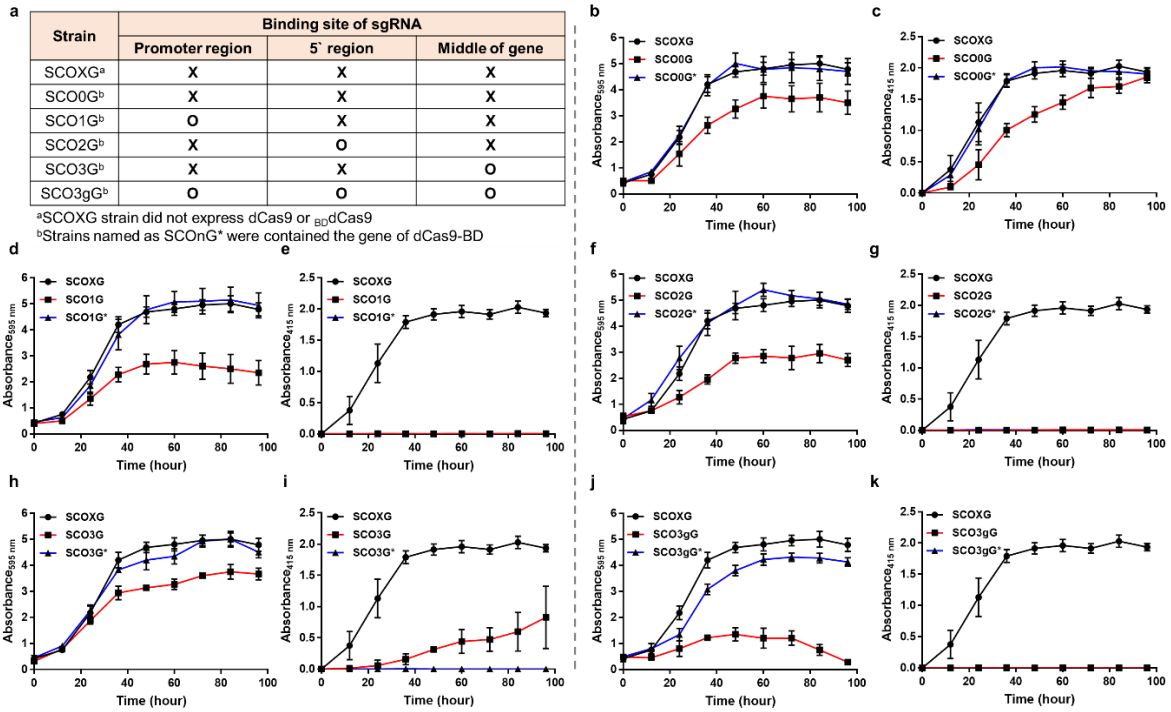


Supplementary Fig. 9. **Subcloning of FK506 BGC into the expression vector. a** After ligation of FK506 BGC with linearized expressing vector, plasmids were transformed into *E. coli* EPI 400 strain. The strain that harbored the desired plasmid was selected by white color. **b** The developed plasmid, pFK, was confirmed by PCR amplification, which showed the band at 1.3 kb. N. D., not detected. Source data are provided as a Source Data file.



Region	Length (kb)	Genes in the region
#1	10.0	*fkbN – fkbA*
#2	9.6	*fkbA*
#3	9.8	*fkbA – fkbP*
#4	9.6	*fkbP – fkbB*
#5	9.8	*fkbB*
#6	9.7	*fkbB*
#7	9.6	*fkbB – fkbC*
#8	12.5	*fkbC – allS*
#9	9.6	*allS – allK*
#10	10.1	*allK* – ApmR
#11	9.1	*traJ – fkbN*

Supplementary Fig. 10. **Verification of FK506 BGC in the pFK plasmid. a** To confirm the integrity of the FK506 BGC within the pFK plasmid, PCR amplification and sequence analysis were performed on 11 regions covering the entire plasmid. The lengths and contained genes of each region were listed. **b** The results of the PCR amplification are presented. Sequence analysis of each PCR product was performed and the results were presented in Supplementary Data. Source data are provided as a Source Data file.



Supplementary Fig. 11. Culture results of *S. coelicolor* M1146 expressed dCas9 or Cas9-BD with various types of sgRNAs targeting *gusA*. Strains used in this experiment (**a**). the three sgRNAs targeted different regions of gusA were designed and expressed in each strain . Cell growth (**b, d, f, h, j**) and expression of the *gusA* expression levels (**c, e, g, h, k**) measured via phenylamine staining and GUS assay. Results of the strains expressing wild-type dCas9 and dCas9-BD without sgRNA (**b, c**), with sgRNA binding *gusA* promoter region (**d, e**), sgRNA binding 5` region of *gusA* gene (**f, g**), sgRNA binding the middle of *gusA* gene (**h, i**), and all three sgRNAs (**j, k**). Flask cultivations were performed in triplicates (*n* = 3). The mean was plotted with error bar representing standard deviation. Source data are provided as a Source Data file.

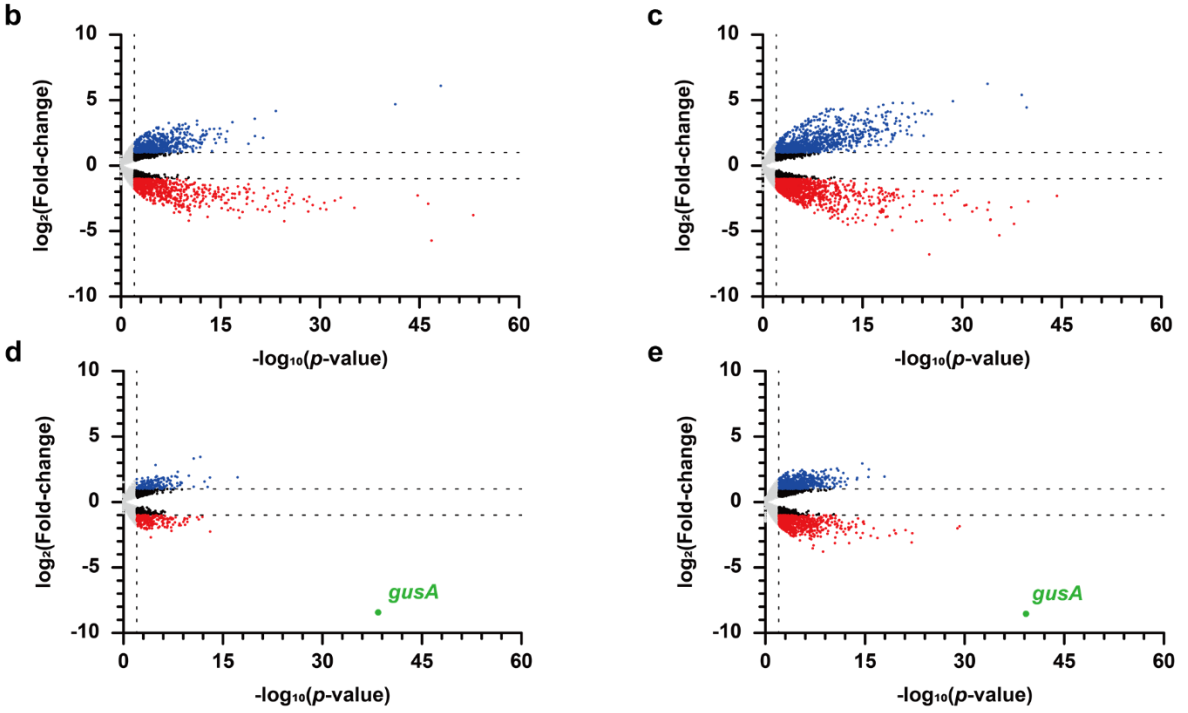
a

Gene	SCOXG	SCO0G	SCO0G*	SCO3G	SCO3G*
gusA	O	O	O	O	O
dCas9		O		O	
dCas9-BD			O		O
sgRNAs				O	O

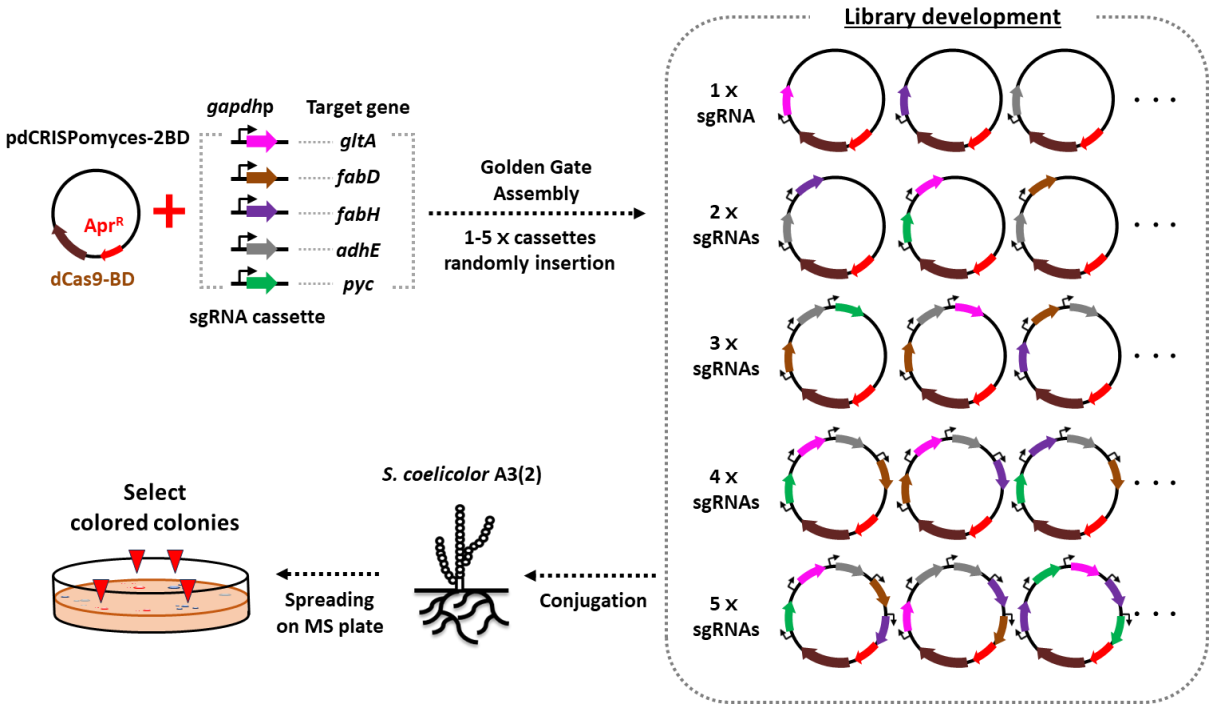
Strain	SCOXG		
Replicates	#1	#2	#3
Raw reads	48,856,558	55,776,192	48,833,434
Uniquely mapped reads	43,431,459	46,362,384	43,020,393
Average read length	50.68	50.67	50.67
Coverage	185.93	198.80	184.11

Strain	SCO0G			SCO0G*		
Replicates	#1	#2	#3	#1	#2	#3
Raw reads	45,914,956	53,883,490	59,519,944	50,342,058	48,496,000	46,161,984
Uniquely mapped reads	40,608,587	50,935,011	46,248,354	46,725,854	44,820,282	42,432,210
Average read length	50.65	50.67	50.65	50.66	50.66	50.66
Coverage	232.57	295.04	264.03	259.94	249.15	238.34

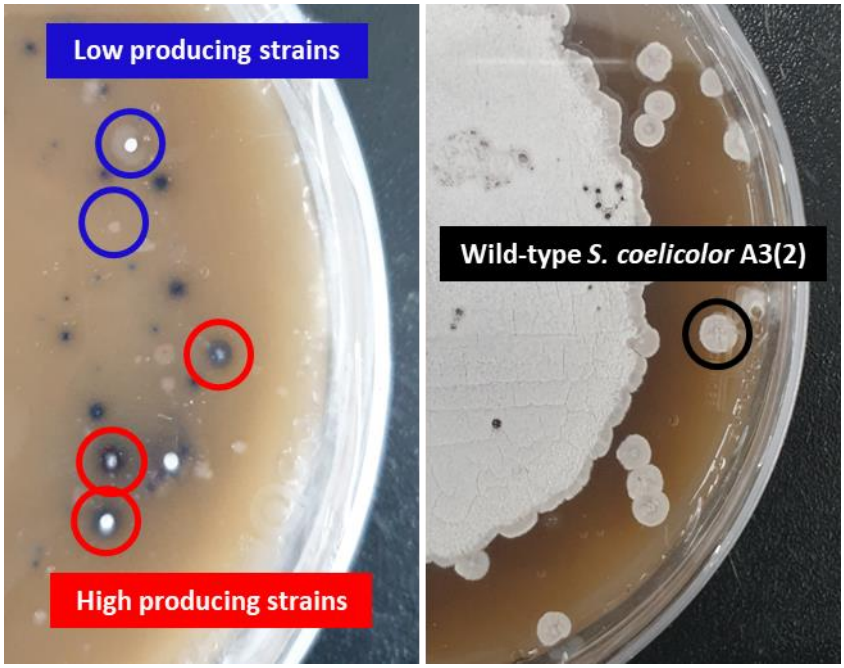
Strain	SCO3G			SCO3G*		
Replicates	#1	#2	#3	#1	#2	#3
Raw reads	45,526,090	45,582,836	46,330,178	44,470,950	54,044,806	48,631,508
Uniquely mapped reads	40,029,809	43,233,611	44,128,131	36,273,537	48,947,163	43,066,342
Average read length	50.65	50.65	50.67	50.66	50.65	50.65
Coverage	227.87	249.63	254.09	204.94	274.50	235.35



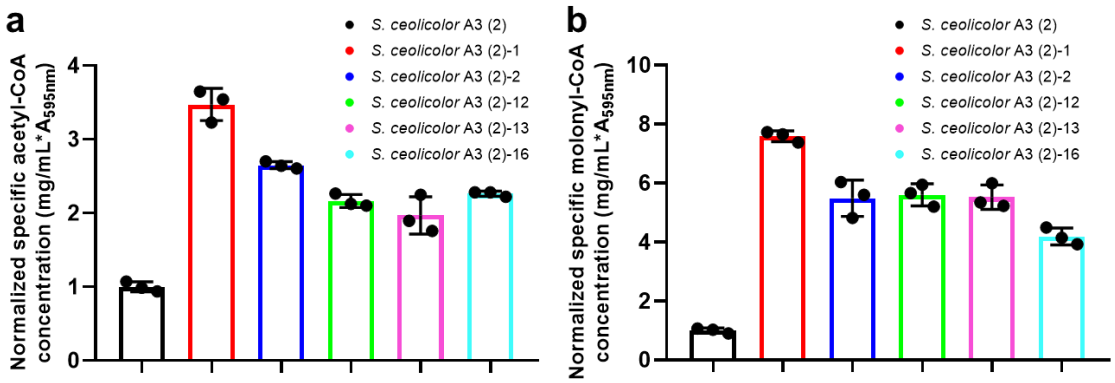
Supplementary Fig. 12. Transcriptional analysis results to compare effect of dCas9 and dCas9-BD with sgRNAs targeting *gusA* gene. Five different strains used for transcriptional analysis and the quality of data in the RNA-seq experiment (**a).** After acquiring transcript levels, the differential expression genes (DEGs) were analyzed. Genes expressed in each SCO0G, SCO0G*, SCO3G, and SCO3G* were compared with those in SCOXG strain (**b, c**, **d**, **e**). Red and blue dots indicated down- or up-regulated genes more than 2-fold with significance level (P<0.01). Black dots indicated unchanged genes, while gray dots presented unsignificant results (*P* > 0.01). The differential expression level of *gusA* gene was marked with green dot. Source data are provided as a Source Data file. ns, not significant.



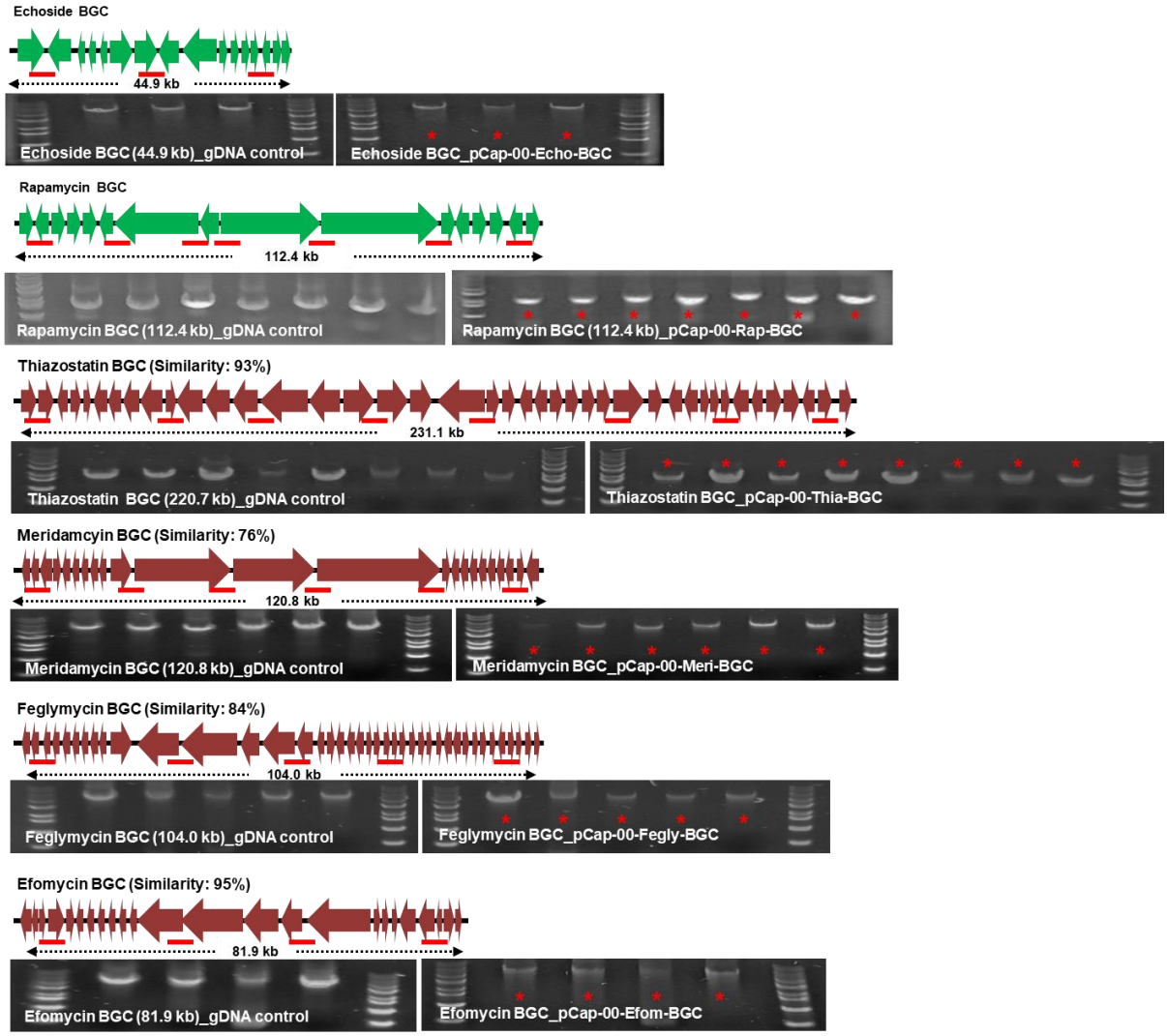
Supplementary Fig. 13. Development of plasmid library for inhibiting multiple genes by dCas9_BD. The scheme about development of sgRNA plasmid library targeting five metabolic genes for inhibiting multiple genes combinatorially using dCas9-BD.



Supplementary Fig. 14. Selection of *S. coelicolor* A3(2) with improved secondary metabolites productions with dCas9-BD. After transformation of library containing dCas9-BD and multiple sgRNAs, the rapidly colored *S. coelicolor* A3(2) strains (red circle) representing increased produced secondary metabolites, actinorhodin, undecylprodigiosin, and coelimycin P2, were selected. In contrast, some colonies (Blue circled) showed lower color than wild-type strain (Black circled) after conjugation.



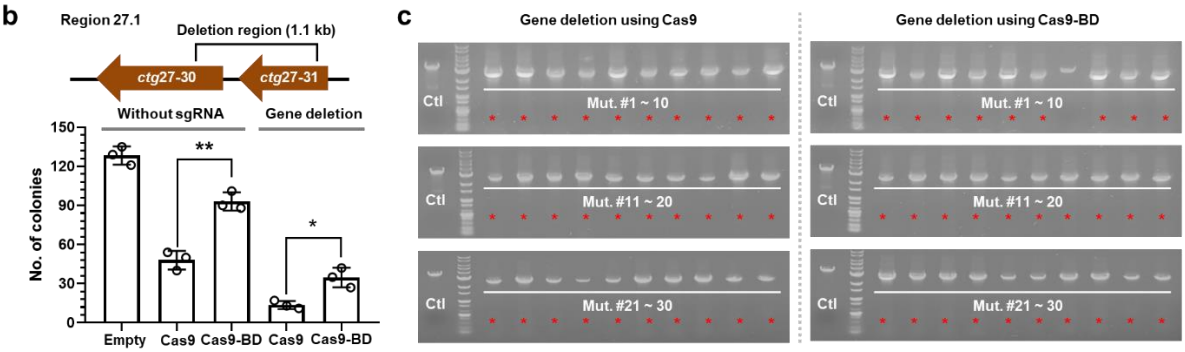
Supplementary Fig. 15. Measurement of acetyl-CoA and malonyl-CoA concentration in the engineered strains by Cas9-BD for higher secondary metabolite production. The concentrations of acetyl-CoA (**a**) and malonyl-CoA (**b**) in wild-type and engineered strains. Flask cultivations were performed in triplicates (*n* = 3). The mean was plotted with error bar representing standard deviation. Source data are provided as a Source Data file.



Supplementary Fig. 16. Captured BGCs fom *Streptomyces rapamycinicus* NRRL 5491 via developed *in vivo* capturing method. Six BGCs of *S. rapamycinicus* NRRL 5491 were captured by *in vivo* capturing method. The BGCs of echoside and rapamaycin, were confirmed in genomic DNA *S. rapamycinicus* NRRL 5491 as well as in the captured plasmids. Their production of metabolites, were also confirmed (green). The other four BGCs were mined by antiSMASH[5] and the similarities with confirmed other BGCs were presented with percentages. The pCap-00 harboring BGCs were confirmed by PCR on various loci in BGCs and amplified sites were pointed with red lines. Source data are provided as a Source Data file.

a

Strain	GC content of genome (%)	Template plasmid	Promoter	Number of exoconjugants	
				With Cas9	With Cas9-BD
Arthrobacter hankyoungi I2-34	68.2%	pCRISPomyces-2	p*rpsL*	48.0±5.89	93.0±5.72
Methylorubrum extorquens AM1	68.5%	pCM66	p*lac*	23.7±5.25	523±88.4
			p*mxaF*	0.0±0.0	1.0±0.8
Streptomyces argenteolus supsp. *toyonakensis*	72.2%	pCRISPomyces-2	p*rpsL*	25	307
Micromonospora antibiotica MMS20-R2-23	72.5%	pCRISPomyces-2	p*rpsL*	0	118
Micromonospora humidisoli MMS20-R2-29	72.5%	pCRISPomyces-2	p*rpsL*	0	1



Supplementary Fig. 17. Application of Cas9 and Cas9-BD with various types of strains with high GC content genomes. a. Summary of transformation experiments expressing Cas9 and Cas9-BD to various strains with high-GC contents. The GC contents of *A. hankyoungi* I2-34[6], *M. extorquens* AM1[7], *S. argenteolus* supsp. *toyonakensis*[8], *M. antibiotica* MMS20-R2-23[9], and *M. humidisoli* MMS-20-R2-29[9], and number of exconjugants were presented. **b.** Number of exconjugants of *A. hankyoungi* I2-34 after deletion of ctg27-30 region with Cas9 and Cas9-BD plasmids. Empty, conjugation with pCRISPomyces-2 plasmid lacking gene coded Cas9 protein; Cas9, conjugation with pCRISPomyces-2 plasmid; Cas9-BD, conjugation with pCRISPomyces-2BD plasmid. Conjugations were performed in triplicates individually (n = 3). The mean was plotted with error bar representing standard error. P values were determined by two-tailed unpaired Student's t-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). **c.** Gene editing confirmed by PCR amplification using different exconjugants. Confirmation with sequence analysis were also used. Source data are provided as a Source Data file.

Reference

1. Abramson J., *et al*., Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature,* **630**, 493-500 (2024).

2. Yuan S., Chan HS., Hu Z., Using PyMOL as a platform for computational drug design, *Wires. Comput. Mol. Sci.,* **7**, 2, e1298 (2017).

3. Nishimasu H*, et al.* Crystal structure of Cas9 in complex with sgRNA and target DNA. *Cell* **156**, 935-949 (2014).

4. Sangsu Bae, Jeongbin Park, Jin-Soo Kim, Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases, *Bioinformatics,* **10**, 1473-1475 (2014).

5. Blin, Kai*, et al.* antiSMASH 6.0: improving cluster detecteion and comparison capabilities. *Nucleic Acid Res.* **49**, W29-W35 (2021).

6. Siddiqi M. Z*., et al.*, *Arthrobacter hankoungi* sp. nov., Isolated From Wet Land, *Current Microbiology*, **80**, 3, 92 (2023).

7. Le H. T. Q*., et al.*, Development of Methylorubrum extorquens AM1 as a promising platform strain for enhanced violacein production from co-utilization of methoanl and acetate, *Metabolic engineering*, **72**, 150-160 (2022).

8. Schoch C. L., *et al.*, NCBI Taxonomy: a comprehensive update on curation, resources and tools, *Database*, **2020**, baaa062 (2020).

9. Lee D. H., *et al.*, *Micromonospora antibiotica* sp. nov. and *Micromonospora humidisoli* sp. nov., two new actinobacterial species exhibiting antimicrobial potential, *Evolutionary Microbiology*, **72**, 10, 00552 (2022).