Supplemental Material

through Heterologous Expression of a Whole-Genome Bacterial Artificial Chromosome Library in Streptomyces spp. Min Xu¹, Yemin Wang¹, Zhilong Zhao², Guixi Gao¹, Sheng-Xiong Huang³, Qianjin Kang¹, Xinyi He¹, Shuangjun Lin¹, Xiuhua Pang², Zixin Deng¹, and Meifeng Tao^{1*} ^{*}To whom correspondence should be addressed. Email: tao meifeng@situ.edu.cn **CONTENTS** BAC vector pHL921......2 Figure S1. Figure S2. Determination of the size of BAC inserts by PFGE......4 Figure S3. Figure S4. Figure S5. Figure S6. Figure S7. Detection and identification of CDAs produced by S. lividans SBT18/8D1......7 Figure S8. Disrupted cda BGC in S. lividans SBT5 and S. lividans SBT18......8 Detection of 8D1-1 and 8D1-2 in S. rochei Sal35 and S. coelicolor M1152/8D1 Figure S9. Figure S10. Chiral analysis of Hpg6 (a) and 3-OH-Asn9 (b) in 8D1-1......9 Figure S11. Comparison of gene organization in *cda2* and *cda* pathways from *S. rochei* Table S1. ¹³C (201 MHz) and ¹H NMR (800 MHz) data of 8D1-1 in deuterated DMSO

Title: Functional Genome Mining for Metabolites Encoded by Large Gene Clusters

Table S3. Predicted functions of the proteins encoded by the cryptic BGC in contig 4.....17



Figure S1. BAC vector pHL921. *repE*, *parA*, and *parB* provide stable replication at unit number in *E. coli. aacC4*, apramycin resistance. *aadA*, streptomycin/spectinomycin resistance. *oriT*, origin of transfer; *attP*, attachment site of phage Φ C31; *int*, Φ C31 integrase. The two *Bam*HI sites are for the cloning of large fragment genomic DNA partially digested with *Sau*3AI. The unique *SmI* site (ATTT^AAAT) is very rare in high G+C *Streptomyces* DNA and was used to linearize BAC constructs for evaluation of their sizes using pulsed field gel electrophoresis (PFGE). DNA sequence: GenBank KP823602.



Figure S2. Streptothricins produced by *S. rochei* Sal35. (a), Extracted ion chromatography of streptothricins produced by *S. rochei* Sal35. ST-F, E, D, C and minute amounts of ST-B were detected in *S. rochei* Sal35 ferment. Note different scale on y-axis. (b), Mass spectra of streptothricins.



Figure S3. Determination of the size of BAC inserts by PFGE. *Smi*l digested BACs were evaluated by PFGE, an average insert size of ~100 kb (= 113 kb - 13 kb vector) and insertion rate of 95% (19/20) could be estimated from the PFGE gel. NEB MidRange IPFG Marker was used here and the PFGE condition was: switch time: 1-25 s, voltage: 6 V/cm; included angle, 120°; 14 °C in 0.5 x TBE for 20 h.



Figure S4. Output of the antimicrobial screening in LEXAS. Antibiosis observed after overlay with *S. aureus*, *B. mycoides*, *M. smegmatis* mc² 155, and *S. sake* and incubated for 16-24 h. BAC clones bearing BGCs producing bioactive secondary metabolites could be directly visualized and screened by LEXAS.



Figure S5. Contigs grouped by *Pvu*II restriction digestion. (a), *Pvu*II restriction digestion of 2B8, 8E6, and 8H1. 8H1 shared no similar bands with other positive BACs was named contig1, 2B8 and 8E6 formed contig2 and shared a ~77 kb overlap. (b), The relative arrangement of 2B8 and 8E6 on *S. rochei* sal35 chromosome. (c), *Pvu*II restriction digestion of 2F3, 6E1, 6F11, and 8A11. These 4 BAC clones shared a ~68 kb overlap and formed contig4. (d), The relative arrangement of contig4 BAC clones on *S. rochei* Sal35 chromosome. The black arrows on the chromosome indicated the boundaries of contig 2 (in b) or contig 4 (in d). Restriction digestion and contig mapping of contig3 were exhibited in the main text (Figure 4). DNA markers used here were 1 kb DNA ladder (Dongsheng Biotech, M1) and 1 kb plus ladder (Invitrogen, M2).



Figure S6. Heterologous expression and identification of borrelidin. (a), PCR verification of *bor* pathway in 8H1. Specific 0.7 kb, 1.5 kb, and 1.2 kb PCR products targeting *borE*, *borA2*, and *borJ* were amplified and sequenced which indicated the presence of *bor* pathway in 8H1. (b) Comparative metabolic profiling of *S. coelicolor* M1152/8H1 and *S. coelicolor* M1152/pHL921. The borrelidin peak was noted and growth inhibition zone around 8H1 in LEXAS was also shown. Molecular mass determined for the borrelidin peak by HRESI-QTOF, and m/z of 490.3179 ([M+H]⁺, calculated 490.3169, error 2.0 ppm) for borrelidin was shown as insert.



Figure S7. Detection and identification of CDAs produced by S. lividans SBT18/8D1. (a),

Detection of CDAs from extract of *S. lividans* SBT18/8D1. The minor peak in the vector control was not CDA. *S. coelicolor* YF11/pAfsRS_{cla} propagated in SV2 medium (1) producing CDAs was set as control. (b), Mass spectra of 8D1-1, 8D1-2, CDA3a/b and CDA4a/b detected in the *S. lividans* SBT18/8D1 extract. (c), Chemical structures of CDA3a/b and CDA4a/b (2).



Figure S8. Disrupted *cda* BGC in *S. lividans* SBT5 and *S. lividans* SBT18. *cda* pathway was disrupted in *S. lividans* SBT5/SBT18 by deleting a 7,116 bp region covering *cdaPS3* and a downstream hydrolase gene, and two pleiotropic activators *afsR_{cla}* and *afsS_{cla}* were introduced at the deletion site (3). Dashed arrows, deleted genes in *S. lividans* SBT5/SBT18. Dashed boxes, A-T-C-A-T-TE domains of CdaPS3 that were deleted in *S. lividans* SBT5/SBT18.



Figure S9. Detection of 8D1-1 and 8D1-2 in *S. rochei* Sal35 and *S. coelicolor* M1152/8D1 by HPLC. *S. rochei* Sal35 was fermented in SFM, GYM, YBP, and R3 media and *S. coelicolor* M1152/8D1 was fermented in R3 medium. 8D1-1 and 8D1-2 were not detected

in the *S. rochei* Sal35 extracts. The minor peak in *S. rochei* Sal35 SFM extract was not 8D1-2.



Figure S10. Chiral analysis of Hpg6 (a) and 3-OH-Asn9 (b) in 8D1-1. 8D1-1 acid hydrolysate was derivatized with FDAA and analyzed by LC-MS, and compared to derivatized standards (4). As the terminal amide bond of 3-OH-Asn would also be hydrolyzed during the acid hydrolyzation in 6N HCl to form 3-OH-Asp and 3-OH-D-Asp was not commercially available, so 3-OH-L-Asp was used as the standard to determine the configuration of 3-OH-Asn(1). Hpg6 and 3-OH-Asn9 were determined both in D-configuration. Chemical structures of FDAA-L/D-Hpg and FDAA-3-OH-L/D-Asp and the mass fragmentation of FDAA-L/D-Hpg were also shown.



Figure S11. Comparison of gene organization in *cda2* and *cda* pathways from *S. rochei* Sal35 and *S. coelicolor* A3(2). The two pathways share identical gene organization except for a 7 genes region absent in *cda2* pathway in *S. rochei* Sal35 and underlined with dashed line which is not required for CDA biosynthesis. DNA sequence identity is shown between the two gene clusters.

	\checkmark					
CDA2PSI_E1 CDA2PSI_E2 CDA2PSI_E CDAPSI_E1 CDAPSI_E2 CDAPSI_E2 CDAPSI_E STFAA_E STFAA_E VpSA_E VpSA_E VpSA_E VpSB_E1 VpSB_E2 AcmE_E BacA_E BacA_E E	AYGPTERTEM GOMEDI PHSPVELTEV MARVAGI ASGPVEATET MGWEAGI ASGPVEATET MGWEAGI -GPAPETEW MARIAGI ASGPVEATET MGWEAGI VEGEULTET OKNEFAN VEGEULTET OKNEFAN VEGEULTET OKNEFAN CEGUTETT OKNEFAN -GEVENTEV MALLGE -GEVENTEV MALLGE -GEVENTEV MALLGE AVGEVPETEV MALLGE VTGETELET OKREFAN VGGVPETEV OKSEFA	20 20 20 20 20 20 20 20 20 20	D FORDSTARRA VJORHDAL D GORLAGALOS VJOHDAL D GORLAGALOS VJOHDAL D ERDVTALGA VJORHDAL D RERVYAJOR VJOHDAL D RERVYAJOR VJOHDAL D RERVYAJOR IJOHDAL D RERVYAJOR IJOHDAL D EGLWROAFGO IVEHHDAL D EGLWROAFGO IVEHHDAL D PERVEKTIOA LIEHHDAL T EKTVAAGFA VJOAHDMI D FERVEKTIOA LIEHHDAL S VYAJAAGES VJOHTHDI G EKRIVTAVOA IJOHHDAL D ERVEKVFKK IJEGHDAL A ERVYEKVFKK IJEGHDAL	R WTIEDPEP R RVAG WTIEDPEP R RVAG WTIEDPEP R RVAG WTIEDPEP R RVAD GSTERAP R YYREG GAIKQINA R VYREG GLIVOVYK R RVPDE GRIUVOVYK R RVPDE GRIUVOVYK R RVPDE GRIUVEG R RVPDE GRIUVER R RVPDE GRIUVEG R RVPDE GRIUVER R RVPDE GRUVE R RVPDE GRUVE R RVPE GRUP R RVPE GRUP R RVPE GUIVE	B0 90 - SVTPBCL VRFDAVEL - SUTPBCL VRFDAVEL - STDASCL TRAAGDV - GLTDERFRF YSVDLKN - GLESKVSF FUNLY - GLESKVSF FUNLY - SUTGSVT RVAL - SAMDAGLVT RVAL - GTUSATUVE RVAL - GUSATUVE RVAL - GUSATUVE	100 IIIII E SAVRSAVIEQ IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
CDA2PSI_E1 CDA2PSI_E2 CDA2PSI_E CDAPSI_E1 CDAPSI_E2 CDAPSI_E2 CDAPSI_E SrfAB_E SrfAB_E VpSA_E VpSA_E VpSB_E1 VpSB_E2 AcnB_E BacA_E BacA_E BacB_E	RAAARSRIAE ADGRID RAAARSRIAE ADGRID RAAARDRIAE ABGRID NVAERDRIDE ADGRID CAARDRIAE ADGRID GAAERDRISE ADGRID GAAERDRISE ADGRID SARLQNSINL EVGPLVK SARLQNSINL EVGPLVK RNKLQSSIDL EBGPLVH REDAVSKIDE SAGVIAR VRIACSSIDL SAGVIAR VRIACSSIDL SAGVIAR VRIACSSIDL SAGVIAR VRIACSSIDL SAGVIAR VRIACSSIDL SAGVIAR VRIACSSID SAGVIAR VRIACSSID RAGVIA UDAQARHA ERVID TGIQKKSSI RKGKIVH ATNIQKDISI SEGKIK	130 140 1 AVW LDR.GAARDG AVL VDR.GARRAG AVL VDR.GARRAG AVL VDR.GARRAG AVL VDR.GARRAG AVL VDR.GADRAGG AVL VDR.GADRAGG MUL VDR.GPDRPG LAL FHTQNGD MUL FARKDGD AGL FQAEDGD AGL FQAEDGD AVW VDA.GPDRIG AVW VDA.GPDREG AVW VDA.GPDREG AVW VDA.GPDREG AVW VDA.GPDREG AVW VDA.GPDREG AVW FDA.GPR.PR.PG	50 L VLVANHLVVD GVTWRTIU VLVANHLVVD GVTWRTIV L VLVANHLVVD GVTWRTIV L LLVANHLVVD GVTWRTIV L ILVANHLVVD GVTWRTIV L ILVANHLVVD GVSWRTIP L MAATHLVVD GVSWRTIP L IVAAHLVVD GVSWRTIP L IVAAHLVVD GVSWRTIP L IVAAHLVVD GVSWRTIP L IVANHLVVD GVSWRTIP L IVAY HLVVD VVOWHHLSVD GVSWRTIP L VVANHLAVVD GVSWRTIP L IATHLVVD GVSWRTIP L IATHLVVD GVSWRTIP	170 180 170 180 180	G DEWRENALL TÜLAGOPR G TEWRENALL TÜLAGOPR G TEWRENALS SDLAGTAV G TEWRENALS SDLAGTAV G TEWRENALS SDLAGTAV G TEWRENALS SDLAGTAV T DEFKOMALS RDLAGTAV T DEFKOMALS RELATE T SFROMALS QEMAVSAE V TEYRONALL GEMAVSAE V TEYRONALL TECALSE V SFROMALC TECALSE V SFROMALC TECALSE V SFROMALL AAFARPA T DEVGEFAROV KKVAQSRR T DEVGEFAROV KKVAQSRR	TE EDLEMRSVU VE ADITHREHT TE ADTOHILDAV TE EDLEMRSVU VE ADITHREHT TE ADTOHILDAV TE EDLEMRSVU ADITHREHT TE ADTOHILDAV TE ADITHREHT TE ADITHREHT TE ADITHREHT ADITHREHT VE ADITHREHT TE ADITHREHT VE ADITHREHT TA ADITHREHT TA ADALWITEVI LL KEDEVARAL
CDA2PSI_E1 CDA2PSI_E2 CDA2PSI_E CDAPSI_E2 CDAPSI_E2 CDAPSI_E2 PheE SrfAA_E SrfAA_E SrfAA_E VpsA_E VpsB_E1 VpsB_E2 AccmB_E BacA_E BacB_E	230 GDTPRAL HVDPAR AGAGAL HVDPAR GDTPHTL RLDPAR GDTPHTL RLDPAR GDTPHTL RLDPAR TGAGTL RLDPAR TGAGTL RLDPAR TGAGTL RLDPAR TGAGTL RLDPAR GDCPA-AC RLDPAR QUARANLPK DRESCDQ AEQUSP-LEK DCETCG GQCPAA-TR ELTSCA GGDPAP-AG DS DGAKRRLG ELDPMC TPVDPFLA FLDPAR EDVPFIPA EKLER -KGKMAFLPK HRQAH	240 250 2 THA TAGEITAELD ADTTEALLT TYQE EAGRVCIDLL SPITEALLT THG RSGOVILDL ADTTEALLT THA RSGOVILDL ADTTEALLT THA RSGOVILDL ADTTEALLT THA RSGOVILDL PDITDALLT KQK NIRYVCMELT IETEKLLK KQK HIRTLISALT VOIENLIR RMK HIKTIEFSUT ADDIKHLIT ANG EVRSKIEVIN FITSALVG G QAHSWSTRIS GTEVITUA TVS TAGRRSWIVP REHAGTAE TVG RARSLIVSLP PEVVSPLIT TFF HSATLSING PDVAKLIR NYE NSRIERISG SQTETEQLK	GO 270 W VPGVCRATEN CVLLSTFA V VPGVFASEN DULLTAFG W VPGVFASEN DVLLTAFA W VPGVFASEN DVLLTAFA W VPGVFASEN DVLLTAFA W VPGVFRASEN DVLLTAFA V VPGVFRASEN DVLLTAFA V VPGVFRASEN DVLLTAFA V VPGVFRASEN DVLLTAFA V VPGVFRASEN DVLLTAFA V VPGVFGSEN DVLLTAFA Q 19GAFACGF DVLLAGAA V VPGVFGSEN DVLLAGAA V VPGVFGSEN DVLLAGAA N AFKANNEN DVLLTALA N AFKANNEN DVLLTALA	280 290 III VAOARHGRAE DPSREVVVI FA VADARRGRAE DPSREVVVI FA VADARRGRAE DPDREVVVI FA VASARRAG GAENVTO TA VASARRAG GAENVTO TA VASARRAG GAENVTO TA VASARRAG GAENVTO TA VASARRAG GAENVTO TA VASARRAG GAENVTO TA VASARGORAGIUTO A VASARGORAGIUTO GA VTRAG	BSHGRHDEAV PGAELSRI I BSHGRHDIV PGAELSRI I BSHGRHDIV PGAELSRI I BSHGRHDIV PGAELSRI I BSHGRHDIV PGAELSRI I BSHGRHDI PUNISRI I BSHGR-BDII PUNISRI I BSHGR-BDII PUNISRI I BSHGR-DIV GILBRI V BSHGR-BEIV GOLBRI V BSHGR-DIV BIN BSHGR-DIV GULBRI M BSHGR-DDIV QUVITRI	3330 AS WETSMYPVRI AG WETSMYPVRI AG WETSMYPVRI AG WETSMYPVRI AG WETSMYPVRI TG WETSMYPVRI UG WETSCYPVVL VG WETSMYPWVL VG WETSMYPWVL VG WETSMYPWVL VG WETSMYPWRI LG WETSMYPVRI IG WETSMYPVRI IG WETSMYPVRI IG WETSMYPVRI IG WETSMYPVRI IG WETSMYPVRI
CDA2PSI_E1 CDA2PSI_E2 CDA2PSI_E CDAPSI_E1 CDAPSI_E1 CDAPSI_E2 PheE SrfAA SrfAA SrfAA SrfAA SrfAA SrfAA SrfAA SafAE VpsA_E VpsA_E VpsA_E VpsA_E VpsA_E AcmE BacA_E BacA_E BacA_E	340 APPA	350 360 3 ALIR T-IRAVKEOI RAVPGDEIG ALQ DGLKOVKEOI RAVPGDEIG ALQ DGLKOVKEOI RAVPGDEIG ALR A-LKAVKEOI RTVPGDEIG ALR A-LKAVKEOI RTVPGDEIG ALR A-LKAVKEOI RTVPGDEIG ALG RALKVKEOI RTVPGDEIG DLS YOIKOIKENI RRIENKGIG DLS YOIKOIKENI RRIENKGIG ALG RLLKTVKEOI RAVPGDEGIG ALG RLLKTVKEOI RAVPGDEGIG ALG RLLKTVKEOI RAVPGDEGIG ALG RLLKTVKEOI RAVPGDEGIG ALG PALKKVKEOI RAVPGDEGIG ALG PALKKVKEOI RAVPGDEGIG ALG PALKKVKEOI RAVPGDEGIG ALG PALKTVKEOI RAVP	380 Y GLLRHIN-BR Y GLRHIN-BR Y	390 PE FORSTAGE PE PDGENYLGE PE PDGENYLGE RURDER RURDER PE PDGENYLGE RURDER PE PDGENYLGE RURDER PE PDGENYLGE RURDER RURDER PE PDGENYLGE PE PDGENYLGE PE STRYLGE PE STR	10 420	430 440 V ACI DGAMPLA GV AASR BAAPMA GV AGAB DTPLA DV ACI DGAMPLA DV ACI DGAMPLA GL AASR TAPMA GL AASR TAPMA GL AASR TAPMA GL AASR TAPMA GL AGE DTPLA GV ACH DTPLA GE LGAG DAVLA AG GAAP STPLR GG LGAG DAVLA AG GAAP STPLA GG LGAG DAVLA GG GGL VALA GG GG GGL VALA GG GGL VALA GG GG GGL VALA GG GGL VALA GG GGL VALA GG GGL VALA GG GG GGL VALA GG GGL VALA GGL GGL GGL GGL GGL GGL GGL GGL GGL GGL
CDA2PSI E1 CDA2PSI E2 CDAPSI E1 CDAPSI E1 CDAPSI E1 CDAPSI E TyCB_E STFAA_E VpSB_E VpSA_E VpSB_E1 VpSB_E1 VpSB_E2 AcmE_E BacB_E	450 HAVDISAVH EGADGER HAVDISAVVH EGADGER GLEVNSMVR DTPDGED CAIDMINAIS MGR SVEDDSVVA GR	460 470 4 IRA RWYSRTELE TEGTORLAD IRA RWYSRTELE TEGTORLAD IRA RWYSRTALE PEDTORLAD IRA WYSRTALE PEDTORLAD IRA WYARALVP DHDARLAE IRA WYARALVP DHDARLAE IRA WYARALVP DHDARLAE IHA WYARALVP DHDARLAE IHA SYSTALE PEDTORLAD IRA WYARALVP DHDARLAE IHA SYSTRALE PEDTORLAD IFA SYSTRALE PEDTORLAD IFA SYSTRALE PENTORLAD ITA SYSTRALE PENTORLAD ITA SYSTRALE PENTORLAD ITA MENGORY ITA MENGORY ITA MENGORY ITA MENGORY ITA MARNILD ITA WWAPDLWP ITA MENGORY ITA MENGORY ITA MENGORY ITA MENGORY ITA MENGORY ITA MENGORY ITA SYSTRELE ITA WENGORY ITA SYSTRE	80 190 A MFRMERRUVE EROPGAG H MFRAEALVE HAGRESTG G MFRAEALVE HAGRESTG G MFRAEALVE HAGRESTG G MFRAEALVE GADRAGTG G MFRAEALVE GADRAGTG S YKQHLAITE HCVGKEDT S YKAHLAITE HCVGKEDT S YKAHLAITE HCLSQDGT FKENEMERE HADREDTG G MLAVIAGUAR GREDEAAG A MLANTGLAV HAGDEAAG D MFQATGLAV HADREDTG S FKENEKAUD HCVDREST S FKENEKAUD HCVDREST	500 510 GL TPSDIAHPTI GODIEH GL TPSDVIDSI SOSIEF GL TPSDVIDSI SOSIEF GL TPSDVIDSI SOSIEF GL TPSDVIGSI SOSIEF GL TPSDVIGSI SOSIEF GL TPSDVIGSI SOSIEF FYSDFFFL SLEWMDI EY TPSDLGDEI SMELONI EY TPSDLGDEI SMELONI GL TSSDFALL HISDFALVAL GODIEAL GH TSSDFALVAL GODIEAL GH TPSDFFLIC GY TPSDFUTCI SCHDEVEU GY TPSDFU		

Figure S12. Homology alignment of epimerization domains. The point mutation of conserved Val139 to Met139 was noted with the red asterisk in Cda2PS1 E1 domain. This point mutation along with other point mutations should be responsible for the inactivation of Cda2PS1 E1 domain which resulted in the release of the linear CDAs from the Cda2 NRPS

assembly line. PheE was selected from gramicidin S biosynthesis pathway, TycB_E, SrfAA_E and SrfAB_E, VpsA_E, VpsB_E1, and VpsB_E2, AcmB_E, BacA_E, and BacB_E were from tyrocidine, surfactin, vancomycin, actinomycin, and bacitracin biosynthesis pathway. The conserved amino acids in epimerization domains were labelled by black arrows (5).

Table S1. ¹³C (201 MHz) and ¹H NMR (800 MHz) data of 8D1-1 in deuterated DMSO (DMSO-*d6*).



2' 124.4 7.13 (s)

Continued

Position	δ _C	δ _H (Mult, <i>J_{H-H}</i>)	Position	δ _C	δ _H (Mult, <i>J_{H-H}</i>)
Hpg7			Glu11		
2'	128.7	7.16 (d, 8.3)	1	171.3	
3'	115.4	6.68 (d, 8.3)	2	52.3	4.28 (br)
4'	157.2		3	27.6	1.69 (m), 1.95 (m)
5'	115.4	6.68 (d, 8.3)	4	30.3	2.19 (m), 2.23 (m)
6'	128.7	7.16 (d, 8.3)	5	174.7	
α-amine		8.21 (br)	α-amine		7.61 (br)
Asp8			Trp12		
1	171.8		1	173.8	
2	50.3	4.57 (m)	2	53.5	4.45 (m)
3	37.3	2.48 (m), 2.61 (m)	3	27.6	3.04 (m),
					3.18 (dd, 14.6,
					5.3)
4	172.4		α-amine		8.19 (br)
α-amine		8.48 (br)	1'amine		10.80 (s)
Gly9			2'	124.2	7.18 (s)
1	169.3		3'	110.1	
2	42.9	3.72 (brd, 12),	4'	118.6	7.52 (d, 7.8)
		3.85 (brd, 12)			
α-amine		8.31 (br)	5'	118.8	6.99 (t, 7.8)
Han10			6'	121.4	7.06 (t, 7.8)
1	168.9		7'	111.9	7.33 (d, 7.8)
2	56.6	4.62 (m)	8'	136.5	
3	72.5	4.11 (d, 4)	9'	127.6	
4	174				
α-amine		8.08 (br)			



Table S2. ¹³C (201 MHz) and ¹H NMR (800 MHz) data of 8D1-2 in DMSO-*d*6.

Position	δ _C	δ _H (Mult, <i>J_{H-H}</i>)	Position	δ_{C}	δ _H (Mult, <i>J_{H-H}</i>)
Hex1			Trp4		
1	170.9		1'-amine		10.75 (br)
2	75.2	4.23 (br)	2'	124.4	7.15 (s)
3	65.4	4.27 (brd, 10.4)	3'	110.2	
4	33.8	1.6 (m), 1.69 (m)	4'	118.9	7.60 (d, 7.8)
5	19.8	1.28 (m), 1.47 (m)	5'	118.6	6.97 (t, 7.8)
6	13.7	0.83 (t, 7.3)	6'	121.3	7.05 (t, 7.8)
Ser2			7'	111.7	7.30 (d, 7.8)
1	170.3		8'	136.5	
2	54.7	4.46 (m)	9'	127.6	
3	62.1	3.55 (m), 3.69 (m)	Asp5		
α-amine		7.91 (brd, 7.6)	1	171.2	
Thr3			2	49.8	4.63 (m)
1	170.3		3	36.6	2.41 (m), 2.68 (m)
2	58.8	4.15 (brdd, 2.8, 8.0)	4	172.2	
3	66.8	3.95 (m)	α-amine		8.36 (d, 7.2)
4	19.9	0.86 (d, 6)	Asp6		
α-amine		7.91 (brd, 7.6)	1	170.4	
Trp4			2	50.1	4.62 (m)
1	171.9		3	36.6	2.55 (m), 2.65 (m)
2	54.2	4.5 (m)	4	172.2	
3	28.2	2.9 (m), 3.15 (m)	α-amine		8.22 (br)
α-amine		7.95 (d, 7.3)			

Position	δ_{C}	δ _H (Mult, <i>J_{H-H}</i>)	Position	δ_{C}	δ _H (Mult, <i>J_{H-H}</i>)
Hpg7			Glu11		
1	170.3		1	171.5	
2	56.1	5.36 (d, 7.3)	2	52.2	4.3 (m)
3 (1')	128.9		3	27.8	1.69 (m), 1.94 (m
2'	128.6	7.14 (d, 8.0)	4	30.3	2.19 (m), 2.24 (m
3'	115.4	6.67 (d, 8.0)	5	174.7	
4'	157.2		α-amine		7.65 (br)
5'	115.4	6.67 (d, 8.0)	Trp12		
6'	128.6	7.14 (d, 8.0)	1	173.7	
α-amine		8.17 (br)	2	53.5	4.45 (m)
Asp8			3	27.5	3.04 (m), 3.18 (m)
1	171.3		α-amine		8.22 (br)
2	50.1	4.60 (m)	1'-amine		10.79(br)
3	36.8	2.46 (m), 2.65 (m)	2'	124.2	7.17 (s)
4	172.2		3'	110.1	
α-amine		8.51 (br)	4'	118.6	7.52 (d, 7.8)
Gly9			5'	118.9	6.99 (t, 7.8)
1	169.2		6'	121.4	7.06 (t, 7.8)
2	42.8	3.74 (brdd, 4.5, 16),	7'	111.9	7.34 (d, 7.8)
		3.83 (brdd, 4.5, 16)			
α-amine		8.22 (br)	8'	136.5	
Han10			9'	127.6	
1	168.8				
2	56.3	4.63 (m)			
3	72.4	4.11 (brd, 3.7)			
4	174				
α-amine		7.98 (br)			

orf1 orf2 orf3	orf4 orf5 orf6	Sorf7orf9 orf11 orf12 orf13 orf14 orf16 orf17	orf18 orf19 orf20	orf21 orf23orf24
		orf8 orf10 orf15		orf22 orf25 1 kb
Regula	tor 📃	Biosynthesis LanC	Transporter	Others
orf	Product	Homolog (source),	Identity/	Conserved domain,
	size (aa)	accession no.	similarity (%)	accession no.
orf1	224	DNA binding response regulator [<i>Streptomyces</i> sp. NRRL F-4835], WP_030976652.1	99/99	CitB, COG2197
orf2	394	Two-component sensor histidine kinase [<i>Streptomyces</i> sp. NRRL F-5650], WP_051851823.1	98/98	HisKA_3, pfam07730
orf3	206	hypothetical protein [<i>Streptomyces</i> sp NRRL F-4835], WP_051890672.1	o. 99/99	-
orf4	370	3-oxoacyl-ACP synthase [<i>Streptomyces</i> sp. NRRL F-4835], WP_030976658.1	100/100	KAS_I_II, cd00834
orf5	474	acyl CoA ligase [<i>Streptomyces aizunensis</i>], AAX98201.1	44/59	CaiC, COG0318
orf6	235	Beta-ketoacyl-ACP reductase [<i>Streptomyces olivochromogenes</i>], KUN43576.1	64/75	FabG, PRK05557
orf7	255	enoyl-[acyl-carrier-protein] reductase [<i>Streptomyces</i> sp. NRRL F-4835], WP_030976664.1	100/100	Fabl, COG0623
orf8	29	hypothetical protein	-	-
orf9	55	hypothetical protein	-	-
orf10	29	hypothetical protein	-	-
orf11	858	serine/threonine protein kinase [<i>Nocardia</i> sp. NRRL S-836], KOV80059.1	49/62	LanC_Ser/Thrkinase, cd04791
orf12	234	GNAT family N-acetyltransferase [<i>Streptomyces</i> sp. NRRL F-4835], WP_051890673.1	100/100	RimL, COG1670
orf13	488	beta-ketoacyl-[acyl-carrier-protein] synthase II [<i>Streptomyces</i> sp. NRRL F-5650], WP_051851828.1	88/100	FabF, TIGR03150
orf14	316	hypothetical protein [<i>Streptomyces</i>], WP_030976672.1	100/100	-
orf15	82	acyl carrier protein [<i>Streptomyces</i>], WP_030976674.1	100/100	AcpP, PRK00982

Table S3. Predicted functions of the proteins encoded by the cryptic BGC in contig 4.

Continued				
orf16	528	transporter, major facilitator family protein [<i>Streptomyces rimosus</i>], WP 003981566.1	53/64	MFS_1, pfam07690
orf17	302	- 3-oxoacyl-ACP reductase [<i>Streptomyces</i>], WP_030976662.1	100/100	BKR_SDR_c, cd05333
orf18	575	acetolactate synthase large subunit [<i>Streptomyces</i> sp. NRRL F-5650], WP_031036547.1	99/99	Acolac_lg, TIGR00118
orf19	413	aldehyde dehydrogenase [<i>Streptomyces</i> sp. NRRL F-5650], WP_051851819.1	99/99	AdhE, COG1012
orf20	424	hypothetical protein [<i>Streptomyces</i> sp. NRRL F-4835], WP_051890678.1	98/98	2A0121, TIGR00900
orf21	430	transcriptional regulator [<i>Streptomyces</i> sp. NRRL F-4835], WP_030976691.1	99/99	HTH_XRE, cd00093
orf22	228	4'-phosphopantetheinyl transferase [<i>Streptomyces</i>], WP_003986505.1	46/58	ACPS, pfam01648
orf23	248	two-component system response regulator [<i>Streptomyces neyagawaensis</i>], WP_055537867.1	41/52	CitB, COG2197
orf24	401	two-component sensor histidine kinase [<i>Streptomyces</i> sp. NRRL F-4835], WP_030976698.1	100/100	COG4585, COG4585
orf25	231	DNA-binding response regulator [<i>Streptomyces</i> sp. NRRL F-4835], WP_051890679.1	100/100	CitB, COG2197

Functions of each protein were annotated by BlastP (<u>http://blast.ncbi.nlm.nih.gov/Blast.cgi</u>). *orf*4, *orf*5, *orf*6, *orf*7, *orf*12, *orf*13, *orf*15, *orf*17, *orf*18, *orf*19, and *orf*22 were PKS related genes, and *orf*11 was a LanC like Ser/Thr kinase gene. We could not tell the structure of the secondary metabolites encoded by this BGC from its sequence. Genes organization of the BGC was also shown upon Table S5.

References

- Kempter C, Kaiser D, Haag S, Nicholson G, Gnau V, Walk T, Gierling KH, Decker H, Zähner H, Jung G. 1997. CDA: Calcium - Dependent Peptide Antibiotics from *Streptomyces coelicolor* A3 (2) Containing Unusual Residues. Angewandte Chemie International Edition in English 36:498-501.
- Hojati Z, Milne C, Harvey B, Gordon L, Borg M, Flett F, Wilkinson B, Sidebottom PJ, Rudd BA, Hayes MA, Smith CP, Micklefield J. 2002. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. Chem Biol 9:1175-1187.
- Bai T, Yu Y, Xu Z, Tao M. 2014. Construction of *Streptomyces lividans* SBT5 as an efficient heterologous expression host. Journal of Huazhong Agricultural University 33:1-6.
- 4. Bhushan R, Bruckner H. 2004. Marfey's reagent for chiral amino acid analysis: a review. Amino Acids 27:231-247.
- Stachelhaus T, Walsh CT. 2000. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. Biochemistry 39:5775-5787.