Electronic Supplementary Information

Biosynthesis of the uridine-derived nucleoside antibiotic A-94964: identification and characterization of the biosynthetic gene cluster provide insight into the biosynthetic pathway

Taro Shiraishi, Makoto Nishiyama, and Tomohisa Kuzuyama

Biotechnology Research Center and Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo, 113-8567, Japan

Address correspondence to: Tomohisa Kuzuyama. Fax, +81-3-5841-8030; E-mail, utkuz@mail.ecc.u-tokyo.ac.jp

Contents

Materials and methods

Bacterial strains, plasmids, and culture conditions. *Escherichia coli* DH5α was used for routine cloning, *E. coli* XL1BLUE MRF' was used for the construction of a cosmid library, *E. coli* BW25141/pKD46[1] was used for the gene deletion in a cosmid by λRed recombination system, *E. coli* DH5α/pTH18cs::cre[2] was used for the removal of the deletion cassette by Cre/*loxP* recombination system. *Streptomyces* sp. SANK 60404 is a gift from Daiichi Sankyo (Tokyo, Japan), which had previously been isolated from a soil sample collected in Okinawa, Japan.[3] A pOJ446[4] vector was used to construct the cosmid library of *Streptomyces* sp. SANK 60404. *Streptomyces albus* G153 was used as a heterologous host for A-94964 production. *Streptomyces* sp. SANK 60404 or each *S. albus* transformant was inoculated into 10 mL TSB medium (30 g L^{-1} tryptic soya broth) and incubated with shaking (300 rpm) at 30°C for 2 day. Two mL of the preculture was inoculated into 100 mL of the A-94964 production medium and incubated with rotating (180 rpm) at 27°C for 7 day.[3] A-94964 standard is also a gift from Daiichi Sankyo.

Genomic DNA isolation, DNA sequencing and de novo assembly.[5] *Streptomyces* sp. SANK 60404 mycelium was inoculated in TSB medium. After 2 days at 30°C, genomic DNA was isolated by phenol chloroform extraction.[3] The isolated genomic DNA was then subjected to Illumina DNA sequencing. A 10-kb mate pair library sequencing was performed with an Illumina Genome Analyzer IIx. The Edena de novo short-reads assembler (Genomic Research Laboratory, Geneva, Switzerland) was used for de novo assembly.[6] An assembly of the sequence reads yielded 1,565 contigs with 7,706,959 total base pairs.

Cosmid library construction, screening, and sequencing. Total DNA from *Streptomyces* sp. SANK 60404 was prepared and partially digested with Sau3AI. DNA fragments larger than 20 kb were ligated with a BamHI- and phosphatase-treated pOJ446, packaged with a LAMBDA INN packaging kit (Nippon Gene, Saitama, Japan), and introduced into *E. coli* XL1Blue MRF′ cells according to the manufacturer's instructions. The resulting cosmid library of SANK 60404 was screened by colony hybridization with a DNA fragment containing the *orf17* as a probe (ECL Direct; GE Healthcare Japan, Tokyo, Japan). A positive cosmid cosT26 was sequenced by the shotgun method (Genotech, Inc., Daejeon, Korea) and annotated with the FramePlot 4.0beta (http://nocardia.nih.go.jp/fp4/) and protein BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Introduction of the cosmid library into *S. albus* G153 cell was performed with polyethylene glycol-mediated protoplast transformation. Protoplasts were prepared using the standard protocol.[4]

Gene deletion by λRed recombination system. The schematic scheme is represented for gene

deletion by λRed recombination system (Fig. S1).[7] Using λRed recombination system, each target gene was deleted by replacement with the *aph* gene that confers kanamycin-resistance. The plasmid pKU479 was used as a template for amplification of the *aph* gene inserted between two *mut-loxP* sequences. Each *mut-loxP* sequences-containing *aph* gene cassette with approximately 40-nt homology arms corresponding to the flanking regions of a target gene was amplified by PCR with a set of primers listed in Table S2. The amplified gene cassettes were individually introduced into *E. coli* BW25141/pKD46/cosT26 or *E. coli* BW25141/pKD46/cosT49. Kanamycin-resistant clones were then selected and each cosmid containing the *mut-loxP* sequences-containing *aph* gene cassette was prepared from the selected clones. The prepared cosmid was introduced into *E. coli* DH5α/pTH18cs::cre to remove the *aph* gene cassette by Cre/*loxP* recombination system. Kanamycin-sensitive clones were then selected and each cosmid, where the target gene is deleted, was prepared from the selected clones. The prepared cosmid was introduced into *S. albus*, and the resultant transformant was used for the following metabolite analysis.

Gene complementation of *anb10* in *S. albus*::cosT49Δ*anb10*. The *anb10* gene was amplified by PCR from cosT49, using the primers listed in Table S2. The primers were designed to amplify the *anb10* gene from the upstream region that includes its ribosomal binding site. A DNA fragment containing the *anb10* gene was cloned downstream of the *rpsJ* promoter in pKU1021[7] to give pKU1021*anb10*. *S. albus*::cosT49Δ*anb10* was transformed with pKU1021*anb10* according to a previously reported protocol.[4]

Analysis of metabolites. After cultivation of *Streptomyces* sp. SANK 60404 or each *S. albus* transformant, two-times volume of acetone was added to the culture broth. After the extraction by acetone, the acetone was evaporated in vacuo, the remaining residue was dissolved in 50% acetonitrile. The resultant solution was analyzed on an HPLC system (Jasco, Tokyo, Japan) equipped with a Capcell Pak C18 UG120 column (4.6 φ × 250 mm; Shiseido, Tokyo, Japan) under the following condition: mobile phase 50% acetonitrile + 50% acetonitrile with 5.0 mM triethylamine phosphate (pH 3.0) at flow rate of 1.0 mL/min. The resultant solution was also analyzed on a high-resolution Triple TOF 5600 MS instrument (SCIEX, Tokyo, Japan) equipped with a UFLC Nexera system (Shimadzu, Kyoto, Japan). In the MS and MS/MS analysis, a Capcell Pak C18 IF2 column (2.0 φ × 50 mm; Shiseido, Tokyo, Japan) or an ACQUITY UPLC BEH Amide column (2.1 φ× 50 mm; Waters, Tokyo) was used. LC condition was as follows: (i) for CAPCELL PAK C18 IF column, mobile phase A, 10% acetonitrile + 10 mM ammonium formate (pH 3.0); mobile phase B, 90% acetonitrile + 10 mM ammonium formate (pH 3.0); 2–98% B over 5 min, 98% B for 5 min, and then 2% A for 5 min, at a flow rate of 0.4 ml/min; (ii) for ACQUITY UPLC BEH Amide column, mobile phase A, 50%

acetonitrile + 10 mM ammonium acetate (pH 9.0); mobile phase B, 95% acetonitrile + 10 mM ammonium acetate (pH 9.0); 100–0% B over 5 min, 0% B for 2.5 min, and then 100% B for 2.5 min, at a flow rate of 0.4 ml/min. MS and MS/MS analyses were simultaneously performed using electrospray ionization in positive mode.

Table S1. Proposed functions of each ORF encoded in cosT49. Accession number: LC431526.

ORFs	aa	Proposed function	BLAST hit protein [Origin]	%Id/Si	E-value	Gene Bank ID	CPZ[8]	TUN[9]
Orf1	212	unknown	hypothetical protein [*Streptomyces* sp. SANK 60405]	64/75	1.00E-71	BAJ05881	Cpz8	
Orf2	272	PAPS 3'-phosphatase	PAPS 3'-phosphatase [*Streptomyces* sp. SN-1061M]	65/72	2.00E-76	ADC96653	Cpz7	
Orf3	492	sulfotransferase	aryl sulfotransferase [*Streptomyces* sp. SANK 60405]	56/72	7.00E-154	BAJ05878	Cpz4	
Orf4	343	type III PKS	putative type III polyketide synthase [*Streptomyces* sp. SANK 60405]	67/77	5.00E-126	ADC96652	Cpz6	
Anb1 (Orf5)	319	dehydrogenase	hypothetical protein [*Streptomyces xylophagus*]	62/77	3.00E-146	WP_043663702		
Anb2 (Orf6)	384	monooxgenase	LLM class flavin-dependent oxidoreductase [*Streptomyces alboviridis*]	78/87	0	WP_032755504		
Anb3 (Orf7)	205	unknown	hypothetical protein [*Streptomyces leeuwenhoekii*]	69/83	4.00E-104	WP_029381177		
Anb4 (Orf8)	304	ABC transporter ATP-binding subuni	ABC transporter ATP-binding protein [*Streptomyces* sp. WM6368]	79/90	4.00E-179	WP_053701923		TunI
Anb5 (Orf9)	256	ABC transporter permease subunit	ABC transporter permease [*Streptomyces leeuwenhoekii*]	71/85	2.00E-124	WP_047121428		TunJ
Anb6 (Orf10)	262	type I PKS (ACP)	hypothetical protein [*Streptomyces alboviridis*]	48/59	4.00E-68	WP_032755492		
Anb7 (Orf11)	261	type I PKS (ER)	SDR family oxidoreductase [*Streptomyces alboviridis*]	69/85	1.00E-128	WP_032755490		
Anb8 (Orf12)	1378	type I PKS (KS, AT)	type I polyketide synthase [*Streptomyces leeuwenhoekii*]	66/76	0	WP_029381182		
Anb9 (Orf13)	830	type I PKS (DH, KR)	SDR family NAD(P)-dependent oxidoreductase [*Streptomyces leeuwenhc*	63/76	0	WP_047121429		
Anb10 (Orf14)	515	oxidoreductase	gfo/Idh/MocA family oxidoreductase [*Streptomyces leeuwenhoekii*]	67/76	0	WP_049976676		
Anb11 (Orf15)	739	aminoglycoside phosphotransferase	hypothetical protein [*Streptomyces leeuwenhoekii*]	59/71	1.00E-161	WP_107408999		
Anb12 (Orf16)	282	glycine transferase	hypothetical protein [Streptomyces leeuwenhoekii]	64/74	9.00E-116	WP_078648000		
Anb13 (Orf17)	222	deacetylase	PIG-L family deacetylase [*Streptomyces leeuwenhoekii*]	79/90	3.00E-131	WP_029381188		TunE
Anb14 (Orf18)	307	methyltransferase	putative sugar O-methyltransferase [*Micromonospora rosaria*]	49/58	1.00E-81	WP_067370002		
Anb15 (Orf19)	507	glycosyl transferase	glycosyl transferase [*Halopolyspora algeriensis*]	45/59	1.00E-119	WP_114454111		
Anb16 (Orf20)	181	FMN reductase	FMN reductase (NADPH) [*Streptomyces* sp. WMMB 714]	59/72	2.00E-56	WP_045864492		
Orf21	128	unknown	glyoxalase [*Streptomyces* sp. CB03238]	92/95	2.00E-80	WP_084902204		
Orf22	281	unknown	hypothetical protein [*Streptomyces* sp. CB03234]	90/93	3.00E-157	WP_073759106		

Table S2. Oligonucleotides used in this study. Start codon and stop codon are underlined except for anb10comp_fw and anb10comp_rv.

Oligonucleotide	Sequence	Description
orf9p_fw	5'-gggcgggatgatcaccgggccgaggaa-3'	screening of cosmid library
orf9p_rv	5'-gggtacgcgtgcactgcctgaccacc-3'	screening of cosmid library
orf17p_fw	5'-gggtcggtcgacgtcggactgatgcagc-3'	screening of cosmid library
orf17p_rv	5'-gggtcgtcgacgtcggactgatgcagc-3'	screening of cosmid library
dorf5_fw	5'-AGCCGATGCACATACCTGTCGATCACCTGAGAAATCTCATGCCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
dorf5_rv	5'-GAGTGCCACCGTGCCTCCCGTTCACGCGTGTTGTGGTCTCACCGGGTACCGAGCGAACGCGTT-3'	gene deletion
dorf22_fw	5'-CCCGTGACGGTCACTGGTGCGAACAGCGGGTCAGAGCCTCACCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
dorf22_rv	5'-TGTCCGGCAGGCACGTCTACAGAGAACGAGGGTCCCGCATGCCGGGTACCGAGCGAACGCGTT-3'	gene deletion
dorf21_fw	5'-CGCTCGCCGCCCCCGTCGCCTCGGCGTAGGGTCCGCGTATGCCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
dorf21_rv	5'-GACCGTCACGGGGGTCGGCCACTGGTGCGAACGGTGTGTCACCGGGTACCGAGCGAACGCGTT-3'	gene deletion
dorf20_fw	5'-GTACCCAGGCGGGCGCCGAACCCAAGGAGTTCCCACCATGCCAGTGAATTCGAGCGACTCGAGTA-3'	gene deletion
dorf20_rv	5'-GACCGTCACGGGGGTCGGCCACTGGTGCGAACGGTGTGTCACCGGGTACCGAGCGAACGCGTT-3'	gene deletion
dorf19_fw	5'-ACGCTCGTACCGCGGGACGGGCCACCCACGCGAAGAGCATGCCAGTGAATTCGAGCGACTCGAG-3'	gene deletion
dorf19_rv	5'-GGTTCGGCGCCCGCCTGGGTACGTGGTCGGCGCCGAGGTCACCGGGTACCGAGCGAACGCGT-3'	gene deletion
danb8-9_fw	5'-GTTCACGAATGACATTCGAACCGGTTGCCATAGTCGGCGTGCCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
danb8-9_rv	5'-AACGCGTTCGCTCGGTACCCGGTGACGTACACCGACCAGCACACCTACCACCGCGGCCGTTCC-3'	gene deletion
danb3_fw	5'-GGACAGCAGACCGACGTACTACCGCGAGACGGGGGATCATGCCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
danb3_rv	5'-AACGCGTTCGCTCGGTACCCGGTGAGCCGGAAGAGCGGGCCGGGACCGAGCGTGACACAGGAG-3'	gene deletion
danb11_fw	5'-CCGGCCGACGACCTGCTGGACGCGTCGCTGTGAGCGCGGTGCCAGTGAATTCGAGCGACTCGAGT-3'	gene deletion
danb11_rv	5'-ATCCGCCTGGCGCGTCGGTCGCACGGTCATCGATGGCGTGCCCGGGTACCGAGCGAACGCGTT-3'	gene deletion
danb10_fw	5'-CACACCTACCACCGCGGCCGTTCCAAGGGGGAGCCGAAGTGCCAGTGAATTCGAGCGACTCGAG-3'	gene deletion
danb10_rv	5'-GACCTCGGTGCGGGCCTCGGTGAGTCCGGTCACCGCGCTCACCGGGTACCGAGCGAACGCGT-3'	gene deletion
anb10comp_fw	5'-GGtctagaCAAGGGGGAGCCGAAGTGAAGAACGTCCTCGTGATCGG-3'(XbaI site underlined)	gene complementation
anb10comp_rv	5'-GGaagcttTCACAGCGACGCGTCCAGCAGGTCGTCGGCCGGT-3'(HindIII site underlined)	gene complementation

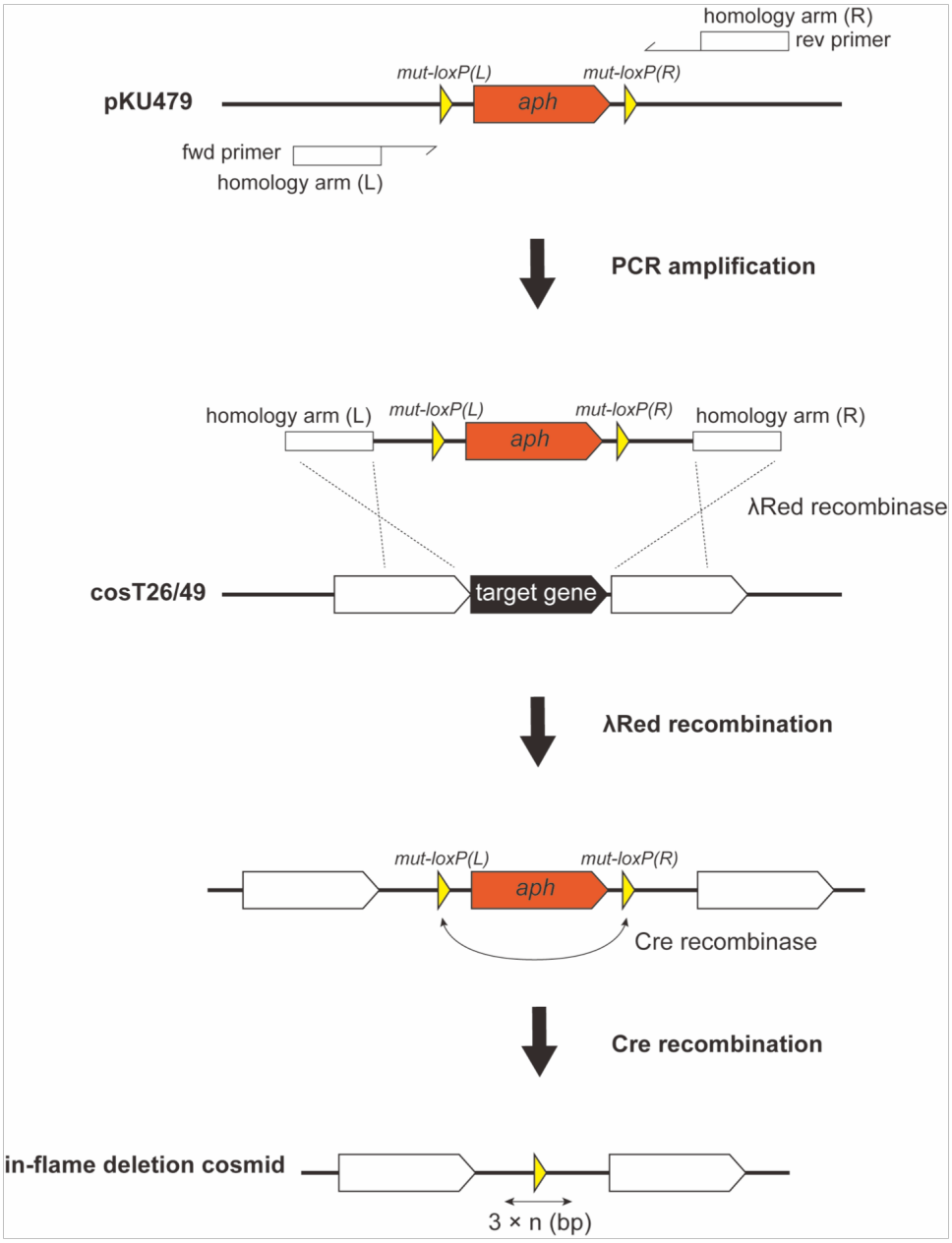

Figure S1. Schematic representation for gene deletion by λRed and Cre/*loxP* recombination systems.

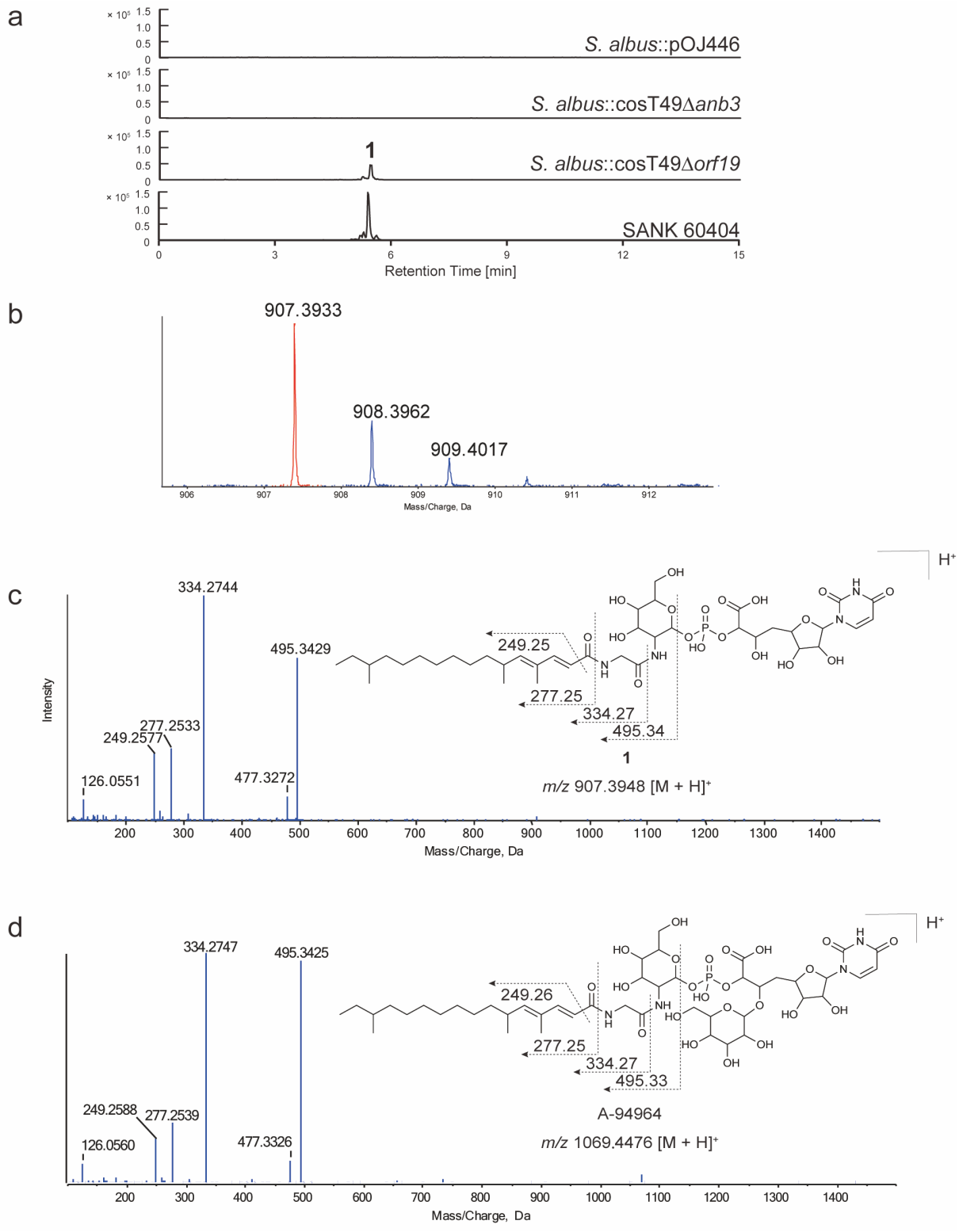

Figure S2. Metabolite analysis of *S. albus*::cosT49Δ*orf19*. a. Extracted ion count chromatograms for a possible intermediate of A-94964 that lacks one sugar moiety (**1**) (*m/z* 907.3948 $[M+H]^+$). b. MS spectrum of **1**. c. MS/MS spectrum of **1**. The predicted fragmentation patterns of **1** are also shown. d. MS/MS spectrum of A-94964. The predicted fragmentation patterns of A-94964 are also shown.

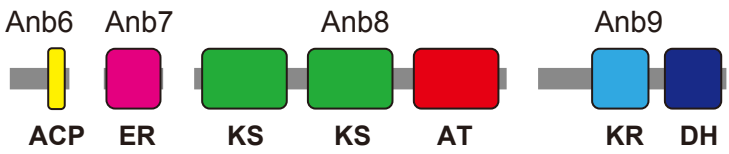

Figure S3. Domain architecture of Anb6-9. ACP, acyl carrier protein; ER, enoyl reductase; KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase.

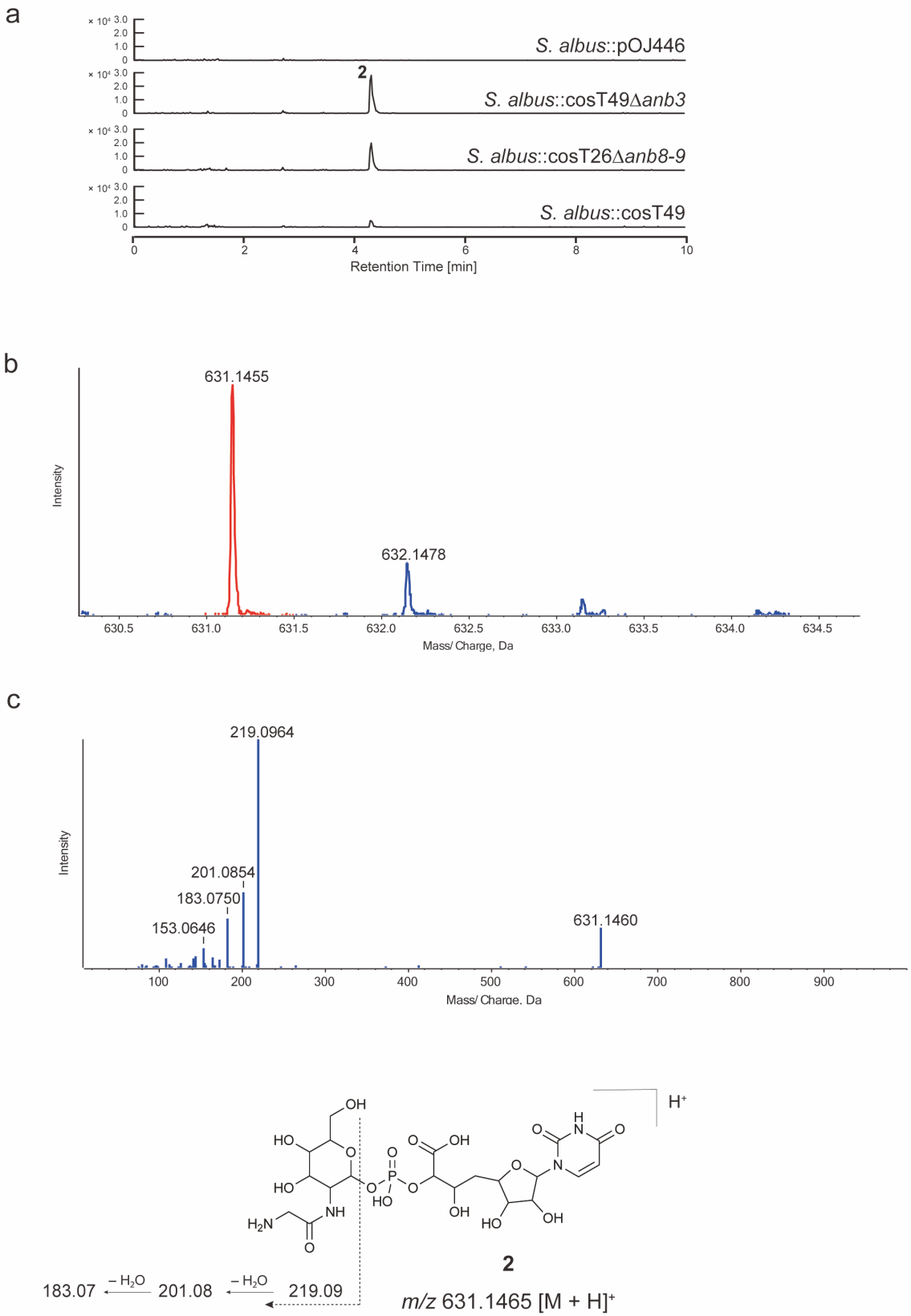

Figure S4. Metabolite analysis of *S. albus*::cosT26Δ*anb3* and *S. albus*::cosT49Δ*anb8-9*. a. Extracted ion count chromatograms for **2** (*m/z* 631.1465 $[M+H]^+$). b. MS spectrum of **2**. c. MS/MS spectrum of **2**. The predicted fragmentation patterns of **2** are also shown.

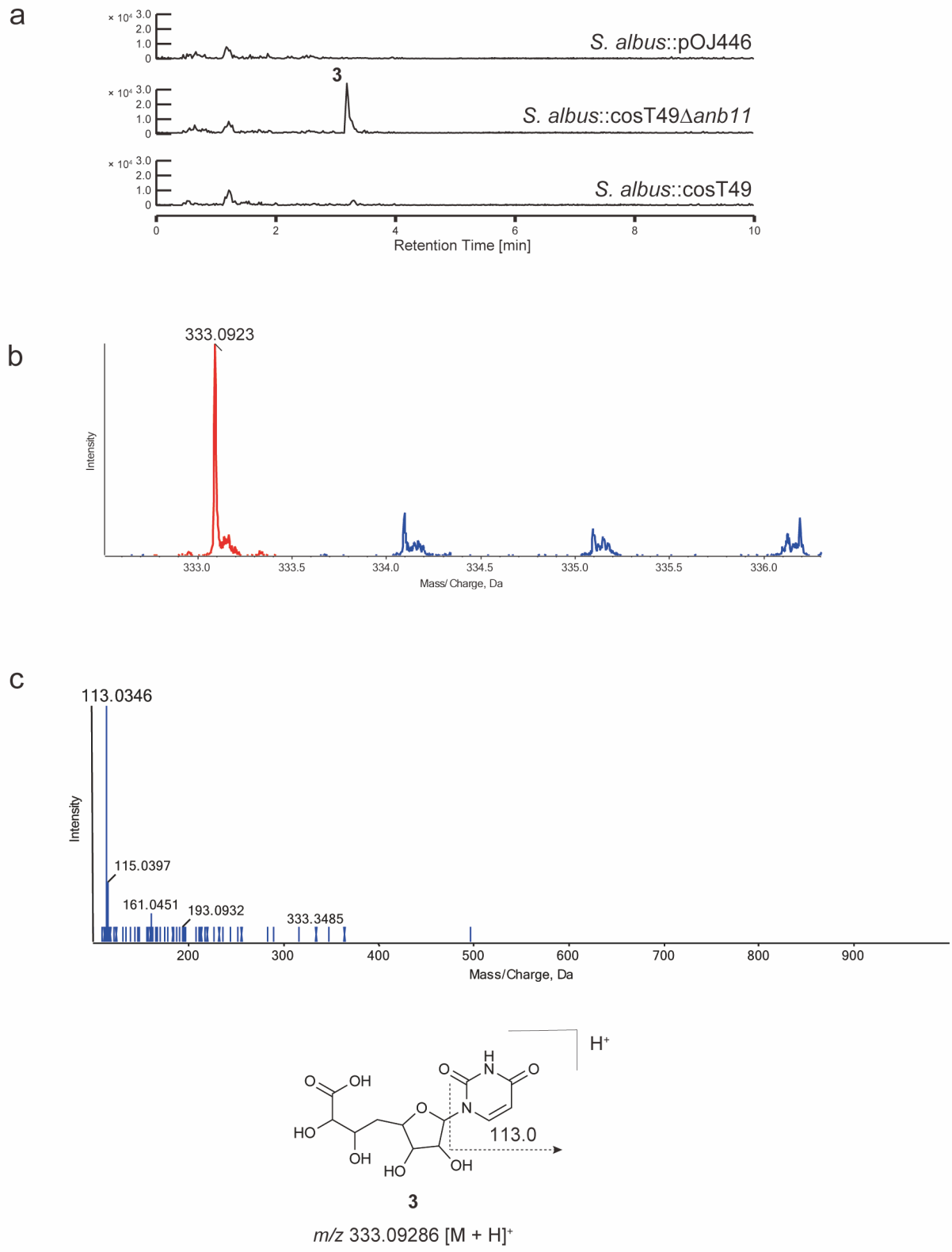

Figure S5. Metabolite analysis of *S. albus*::cosT49Δ*anb11*. a. Extracted ion count chromatograms for **3** (*m/z* 333.09286 $[M+H]^+$). b. MS spectrum of **3**. c. MS/MS spectrum of **3**. The predicted fragmentation patterns of **2** are also shown.

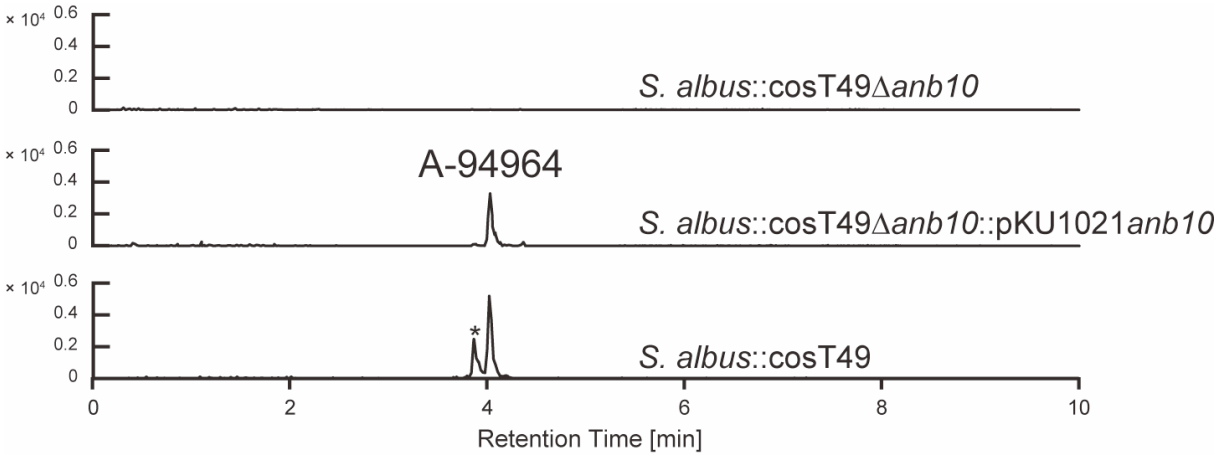

Figure S6. Metabolite analysis of *S. albus*::cosT49Δ*anb10*::pKU1021*anb10*. Extracted ion count chromatograms for A-94964 (*m/z* 1069. 4476 $[M+H]^+$) obtained via LC-MS analysis of the culture extracts from each transformant. The metabolite marked with asterisk (*) corresponds to an isomer of A-94964.

Reference

1 K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sci.*, 2000, **97**, 6640–6645.

2 N. Suzuki, Y. Tsuge, M. Inui and H. Yukawa, *Appl. Microbiol. Biotechnol.*, 2005, **67**, 225–233.

3 R. Murakami, Y. Fujita, M. Kizuka, T. Kagawa, Y. Muramatsu, S. Miyakoshi, T. Takatsu and M. Inukai, *J. Antibiot*, 2008, **61**, 537–544.

4 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, 2000.

5 A. Meguro, T. Tomita, M. Nishiyama and T. Kuzuyama, *ChemBioChem*, 2013, **14**, 316–321.

6 D. Hernandez, P. François, L. Farinelli, M. Østerås and J. Schrenzel, *Genome Res.*, 2008, **18**, 802–809.

7 M. Komatsu, K. Komatsu, H. Koiwai, Y. Yamada, I. Kozone, M. Izumikawa, J. Hashimoto, M. Takagi, S. Omura, K. Shin-ya, D. E. Cane and H. Ikeda, *ACS Synth. Biol.*, 2013, **2**, 384–396.

8 L. Kaysser, L. Lutsch, S. Siebenberg, E. Wemakor, B. Kammerer and B. Gust, *J. Biol. Chem.*, 2009, **284**, 14987–14996.

9 F. J. Wyszynski, A. R. Hesketh, M. J. Bibb and B. G. Davis, *Chem. Sci.*, 2010, **1**, 581–589.