Supporting Information

Kersten et al. 10.1073/pnas.1315492110

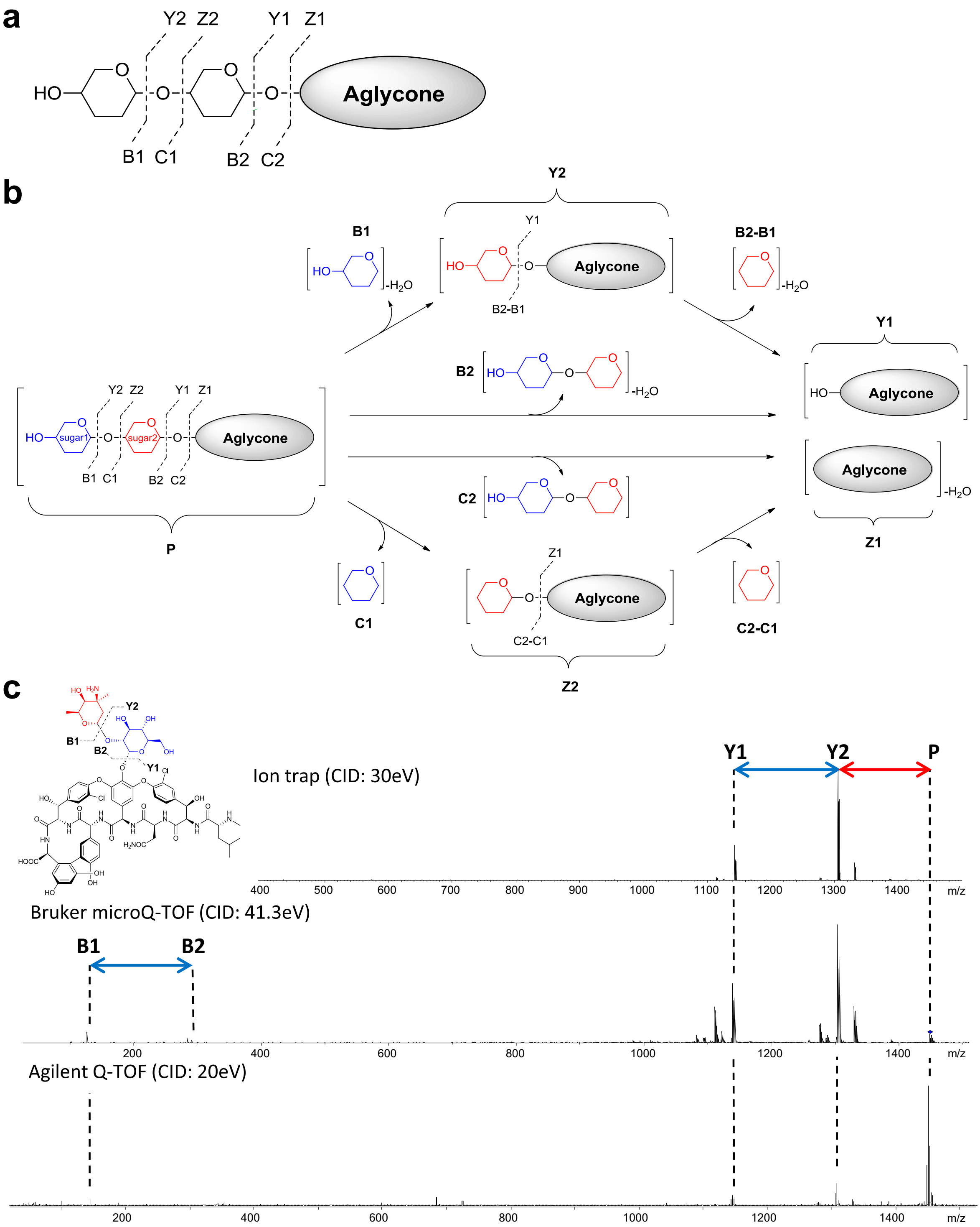
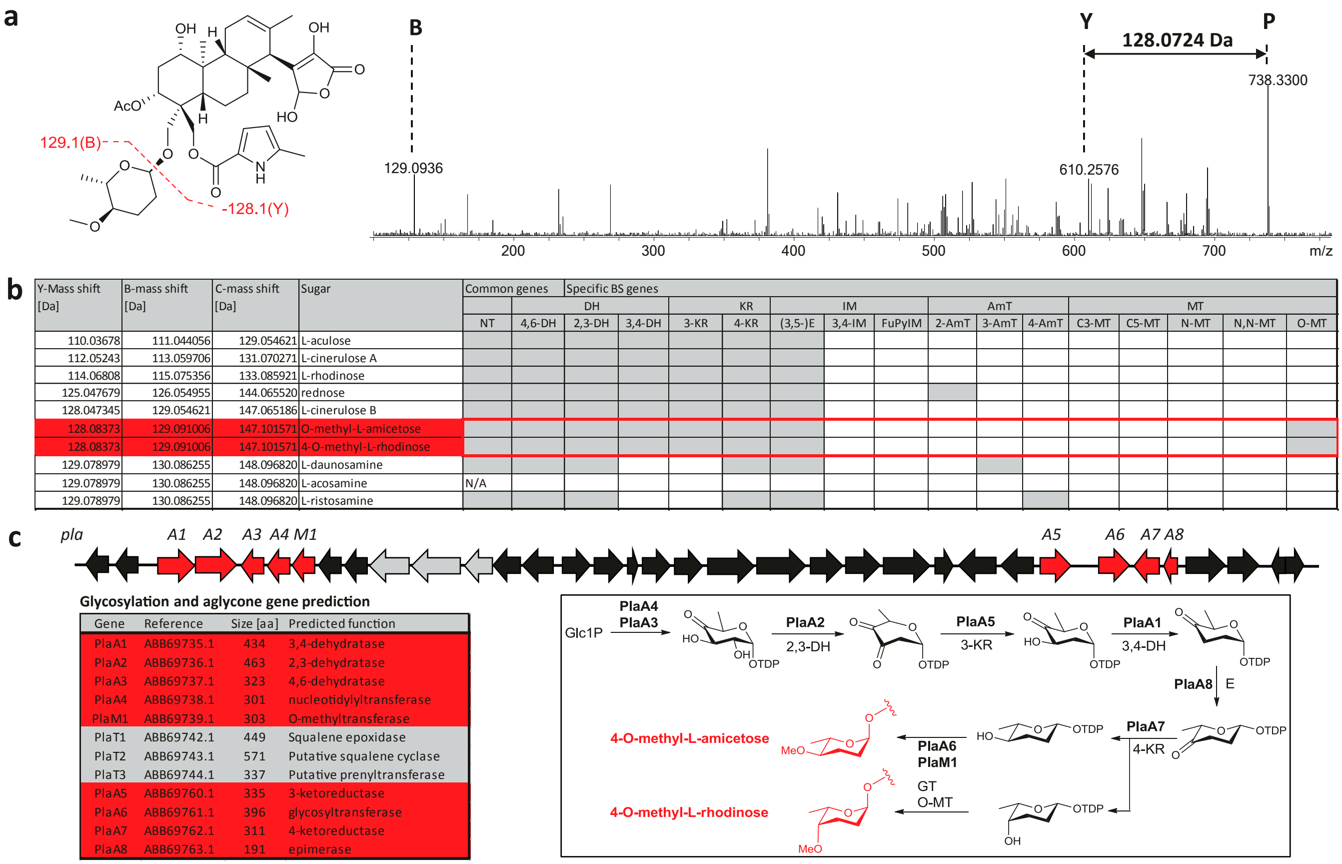


Fig. S1. Tandem mass-spectrometric analysis of glycosylated natural products (GNPs) for sugar characterization. (*A*) Tandem MS fragmentation nomenclature of O-glycosylated natural products adapted from oligosaccharide nomenclature. (*B*) Common tandem MS fragmentation (dotted line) of O-glycosidic bonds in GNPs yielding sugar mass shifts (red and blue) via the aglycone Y-ion series and sugar fragments. (*C*) Tandem MS spectra of vancomycin indicating the Y-ion sugar mass shifts and B-ions from O-glycosidic bond cleavage compared between ion trap MS and quadrupole time-of-flight (Q-TOF) MS analysis and with different collision-induced dissociation (CID) energies. Abbreviations: B, B-ion; C, C-ion; P, parent/precursor ion; Y, Y-ion; Z, Z-ion.



Y-Mass shift [Da]	B-mass shift [Da]	C-mass shift [Da]	Sugar	Common genes	Specific BS genes															
					DH			KR		IM			AmT			MT				
				NT	4,6-DH	2,3-DH	3,4-DH	3-KR	4-KR	(3,5-)E	3,4-IM	FuPyIM	2-AmT	3-AmT	4-AmT	C3-MT	C5-MT	N-MT	N,N-MT	O-MT
110.03678	111.044056	129.054621	L-aculose	▨	▨	▨	▨	▨	▨	▨										
112.05243	113.059706	131.070271	L-cinerulose A	▨	▨	▨	▨	▨	▨	▨										
114.06808	115.075356	133.085921	L-rhodinose	▨	▨	▨	▨	▨	▨	▨										
125.047679	126.054955	144.065520	rednose	▨	▨	▨	▨	▨	▨	▨			▨							
128.047345	129.054621	147.065186	L-cinerulose B	▨	▨	▨	▨	▨	▨	▨										
128.08373	129.091006	147.101571	O-methyl-L-amicetose	▨	▨	▨	▨	▨	▨	▨										▨
128.08373	129.091006	147.101571	4-O-methyl-L-rhodinose	▨	▨	▨	▨	▨	▨	▨										▨
129.078979	130.086255	148.096820	L-daunosamine	▨	▨	▨			▨	▨				▨						
129.078979	130.086255	148.096820	L-acosamine	N/A																
129.078979	130.086255	148.096820	L-ristosamine	▨	▨	▨			▨	▨					▨					

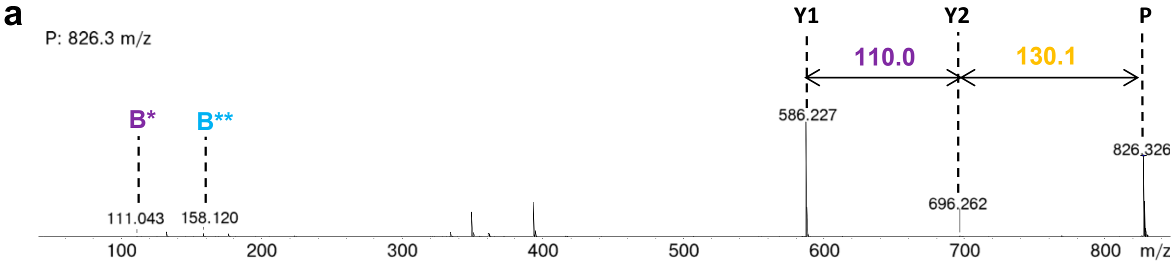
Glycosylation and aglycone gene prediction

Gene	Reference	Size [aa]	Predicted function
PlaA1	ABB69735.1	434	3,4-dehydratase
PlaA2	ABB69736.1	463	2,3-dehydratase
PlaA3	ABB69737.1	323	4,6-dehydratase
PlaA4	ABB69738.1	301	nucleotidylyltransferase
PlaM1	ABB69739.1	303	O-methyltransferase
PlaT1	ABB69742.1	449	Squalene epoxidase
PlaT2	ABB69743.1	571	Putative squalene cyclase
PlaT3	ABB69744.1	337	Putative prenyltransferase
PlaA5	ABB69760.1	335	3-ketoreductase
PlaA6	ABB69761.1	396	glycosyltransferase
PlaA7	ABB69762.1	311	4-ketoreductase
PlaA8	ABB69763.1	191	epimerase

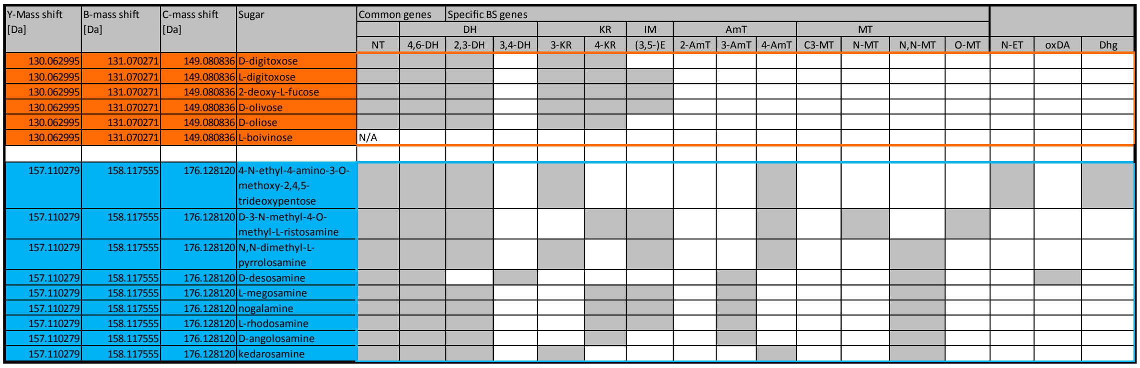
Fig. S2. Connection of glycosylated terpene phenalinolactone A from *Streptomyces* sp. Tu6071 with its gene cluster by the MS-glycogenetic code. (*A*) Tandem MS spectrum of phenalinolactone A. A putative B-ion and Y-ion mass shift of a methyltrideoxysugar was detected using the sugar mass list of the MS-glycogenetic code. (*B*) MS-glycogenetic code with observed putative sugar fragments (red) and the corresponding biosynthetic genes of these sugars (red box). (*C*) Prediction of glycosylation genes from the phenalinolactone gene cluster [GenBank accession no. DQ230532 (1)]. Predicted gene functions match the pathways of the observed MS/MS sugars, 4-*O*-methyl-L-rhodinose and 4-*O*-methyl-L-amicetose.

1. Dürr C, et al. (2006) Biosynthesis of the terpene phenalinolactone in *Streptomyces* sp. Tü6071: Analysis of the gene cluster and generation of derivatives. *Chem Biol* 13(4):365–377.

a

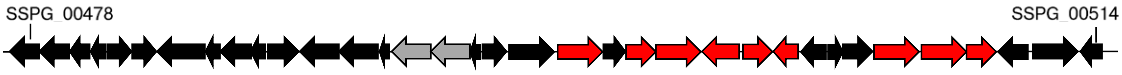


b



Y-Mass shift [Da]	B-mass shift [Da]	C-mass shift [Da]	Sugar	Common genes	Specific BS genes															
					DH			KR		IM	AmT			MT						
				NT	4,6-DH	2,3-DH	3,4-DH	3-KR	4-KR	(3,5-)E	2-AmT	3-AmT	4-AmT	C3-MT	N-MT	N,N-MT	O-MT	N-ET	oxDA	Dhg
130.062995	131.070271	149.080836	D-digitoxose	▨	▨	▨		▨	▨											
130.062995	131.070271	149.080836	L-digitoxose	▨	▨	▨		▨	▨	▨										
130.062995	131.070271	149.080836	2-deoxy-L-fucose	▨	▨	▨		▨	▨	▨										
130.062995	131.070271	149.080836	D-olivose	▨	▨	▨		▨	▨	▨										
130.062995	131.070271	149.080836	D-oliose	▨	▨	▨		▨	▨											
130.062995	131.070271	149.080836	L-boivinose	N/A																
157.110279	158.117555	176.128120	4-N-ethyl-4-amino-3-O-methoxy-2,4,5-trideoxypentose	▨	▨	▨		▨					▨					▨		▨
157.110279	158.117555	176.128120	D-3-N-methyl-4-O-methyl-L-ristosamine	▨	▨	▨			▨	▨			▨		▨		▨			
157.110279	158.117555	176.128120	N,N-dimethyl-L-pyrrolosamine	▨	▨	▨		▨		▨			▨			▨				
157.110279	158.117555	176.128120	D-desosamine	▨	▨		▨					▨				▨			▨	
157.110279	158.117555	176.128120	L-megosamine	▨	▨	▨			▨	▨		▨				▨				
157.110279	158.117555	176.128120	nogalamine	▨	▨	▨			▨	▨		▨				▨				
157.110279	158.117555	176.128120	L-rhodosamine	▨	▨	▨			▨	▨		▨				▨				
157.110279	158.117555	176.128120	D-angolosamine	▨	▨	▨			▨			▨				▨				
157.110279	158.117555	176.128120	kedarosamine	▨	▨	▨		▨					▨			▨				

c



Gene	Reference	Size [aa]	Predicted function	Closest homolog [sim/id, %]
SSBG_00478	EDY42516.1	284	daunorubicin resistance protein	BAB72060.1| AcrW [Streptomyces galilaeus] (93/89)
SSBG_00479	EDY42517.2	326	daunorubicin resistance protein DrrA	BAB72059.1| AcrV [Streptomyces galilaeus] (93/86)
SSBG_00481	EDY42519.2	206	transcriptional regulator	BAB72058.1| AclS [Streptomyces galilaeus] (80/74)
SSBG_00482	EDY42520.2	144	aglycone C-1 hydroxylase	BAB72057.1| AclR [Streptomyces galilaeus] (95/88)
SSBG_00483	EDY42521.1	269	putative hydroxylase	BAB72056.1| AclQ [Streptomyces galilaeus] (83/74)
SSBG_00484	EDY42522.1	242	N-methyltransferase	BAB72055.1| AclP [Streptomyces galilaeus] (86/80)
SSBG_00485	EDY42523.2	519	oxidoreductase (OxRed)	ABI15166 | AknOx [Streptomyces galilaeus] (corrected from AclO)
SSBG_00486	EDY42524.1	159	hypothetical protein	BAB72053.1| AclJ [Streptomyces galilaeus] (87/81)
SSBG_00487	EDY42525.1	281	transcriptional regulatory protein	BAB72052.1| AclI [Streptomyces galilaeus] (92/88)
SSBG_00488	EDY42526.1	144	CYC, cyclase	BAB72051.1| AknH [Streptomyces galilaeus] (96/91)
SSBG_00489	EDY42527.1	286	MET, methyltransferase	BAB72050.1| AknG [Streptomyces galilaeus] (90/87)
SSBG_00490	EDY42528.2	365	Propionate starter unit, AT	BAB72049.1| AknF [Streptomyces galilaeus] (86/82)
SSBG_00491	EDY42529.1	354	Propionate starter unit, KSIII	BAB72048.1| AknE2 [Streptomyces galilaeus] (88/83)
SSBG_00492	EDY42530.1	92	minPKS, ACP	BAB72047.1| AknD [Streptomyces galilaeus] (88/81)
SSBG_00493	EDY42531.1	412	minPKS, CLF	BAB72046.1| AknC [Streptomyces galilaeus] (88/81)
SSBG_00494	EDY42532.1	424	minPKS, KS	BAB72045.1| AknB [Streptomyces galilaeus] (95/89)
SSBG_00495	EDY42533.2	118	OXY, oxygenase	BAB72044.1| AknX [Streptomyces galilaeus] (77/75)
SSBG_00496	EDY42534.1	261	KR, ketoreductase	BAB72043.1| AknA [Streptomyces galilaeus] (93/87)
SSBG_00497	EDY42535.1	450	ARO, aromatase	BAB72042.1| AknE1 [Streptomyces galilaeus] (89/83)
SSBG_00498	EDY42536.1	440	glycosyltransferase (GT)	BAB72041.1| AclK [Streptomyces galilaeus] (94/88)
SSBG_00499	EDY42537.1	205	epimerase (E)	BAB72040.1| AclL [Streptomyces galilaeus] (93/88)
SSBG_00500	EDY42538.2	311	4-ketoreductase (4-KR)	BAB72039.1| AclM [Streptomyces galilaeus] (81/75)
SSBG_00501	EDY42539.2	448	2,3-dehydratase (2,3-DH)	BAB72038.1| AclN [Streptomyces galilaeus] (86/81)
SSBG_00502	EDY42540.1	369	aminotransferase (AmT)	BAB72037.1| AclZ [Streptomyces galilaeus] (91/87)
SSBG_00503	EDY42541.1	291	TDP-glucose synthase (NT)	BAB72036.1| AclY [Streptomyces galilaeus] (96/91)
SSBG_00504	EDY42542.1	239	N,N-dimethyltransferase (N,N-MT)	AAF73460.1|AF264025_11 AknX2 [Streptomyces galilaeus] (92/86)
SSBG_00505	EDY42543.1	259	CYC, cyclase	AAF73459.1|AF264025_10 AknW [Streptomyces galilaeus] (96/90)
SSBG_00507	EDY42544.1	144	putative oxidoreductase	AAF73458.1|AF264025_9 AknV [Streptomyces galilaeus] (85/81)
SSBG_00508	EDY42546.1	276	C-7 KR, ketoreductase	AAF73457.1|AF264025_8 AknU [Streptomyces galilaeus] (87/82)
SSBG_00509	EDY42547.2	445	P450-like, GTF activation	AAF73456.1|AF264025_7 AknT [Streptomyces galilaeus] (79/72)
SSBG_00510	EDY42548.2	438	glycosyltransferase (GT)	AAF73455.1|AF264025_6 AknS [Streptomyces galilaeus] (89/83)
SSBG_00511	EDY42549.1	328	4,6-dehydratase (4,6-DH)	AAF73454.1|AF264025_5 AknR [Streptomyces galilaeus] (89/84)
SSBG_00512	EDY42550.2	300	3-ketoreductase (3-KR)	AAF73453.1|AF264025_4 AknQ [Streptomyces galilaeus] (90/85)
SSBG_00513	EDY42551.2	440	3,4-dehydratase (3,4-DH)	AAF73452.1|AF264025_3 AknP [Streptomyces galilaeus] (97/92)
SSBG_00514	EDY42552.1	319	transcriptional regulatory protein	AAF73451.1|AF264025_2 AknO [Streptomyces galilaeus] (93/87)

Fig. S3. (Continued)

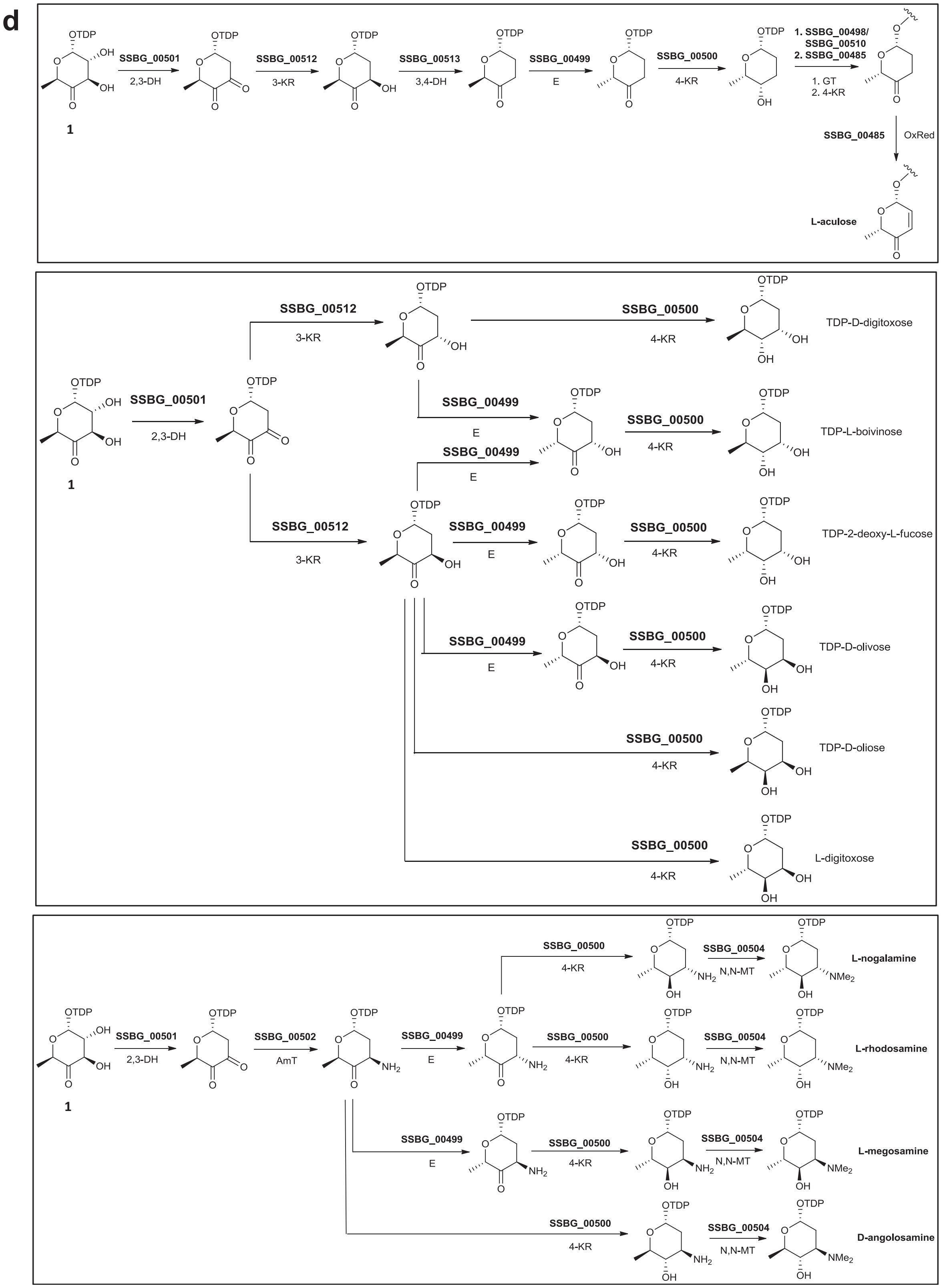


Fig. S3. Glycogenomic characterization of cinerubin B, a glycosylated anthracycline polyketide, from *Streptomyces* sp. SPB74. (*A*) Tandem MS spectrum of cinerubin B with Y-ion mass shifts (purple, orange) and B-ions (blue) corresponding to putative sugar monomers. (*B*) Characterization of candidate MS^n sugars from cinerubin B with corresponding glycosylation genes by the MS-glycogenetic code (Dataset S2). (*C*) Gene cluster analysis of cinerubin B pathway with highlighted glycosylation genes (red) and aglycone biosynthetic genes (gray). (*D*) Matching pathways of specific glycosylation genes with candidate MS/MS sugars from cinerubin B, each starting at deoxysugar biosynthetic intermediate TDP-4-keto-6-deoxy-α-D-glucose (1).

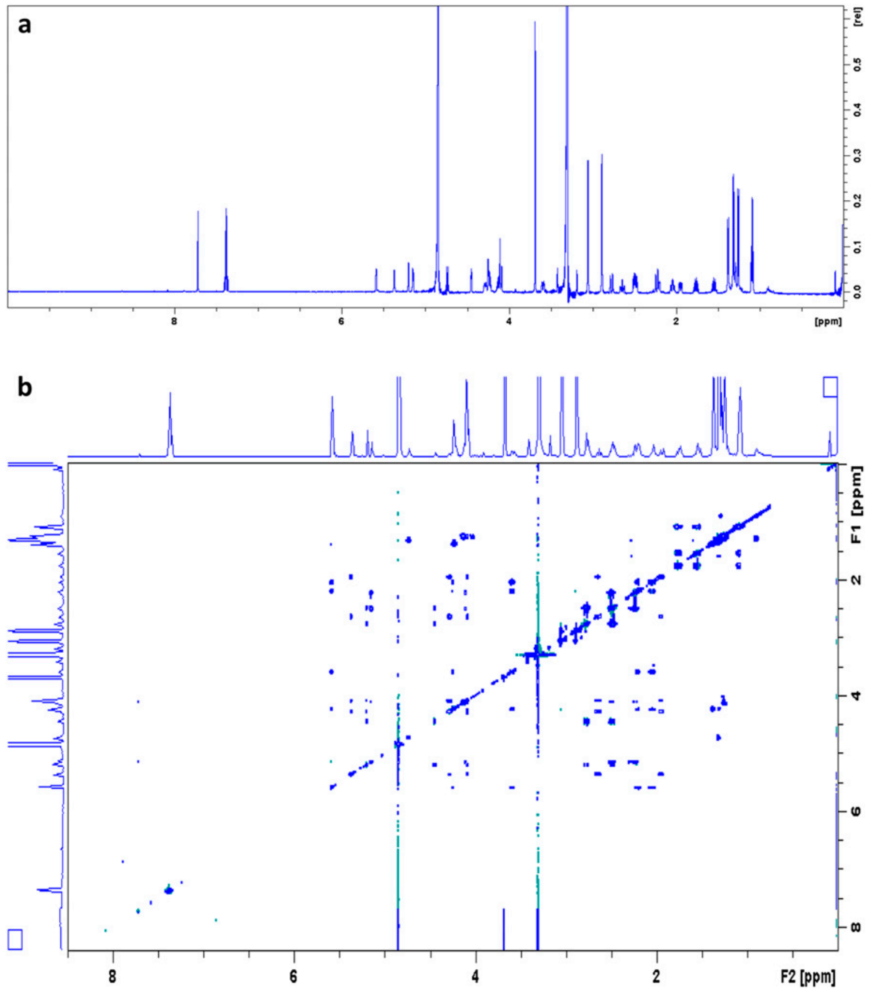


Fig. S4. (Continued)

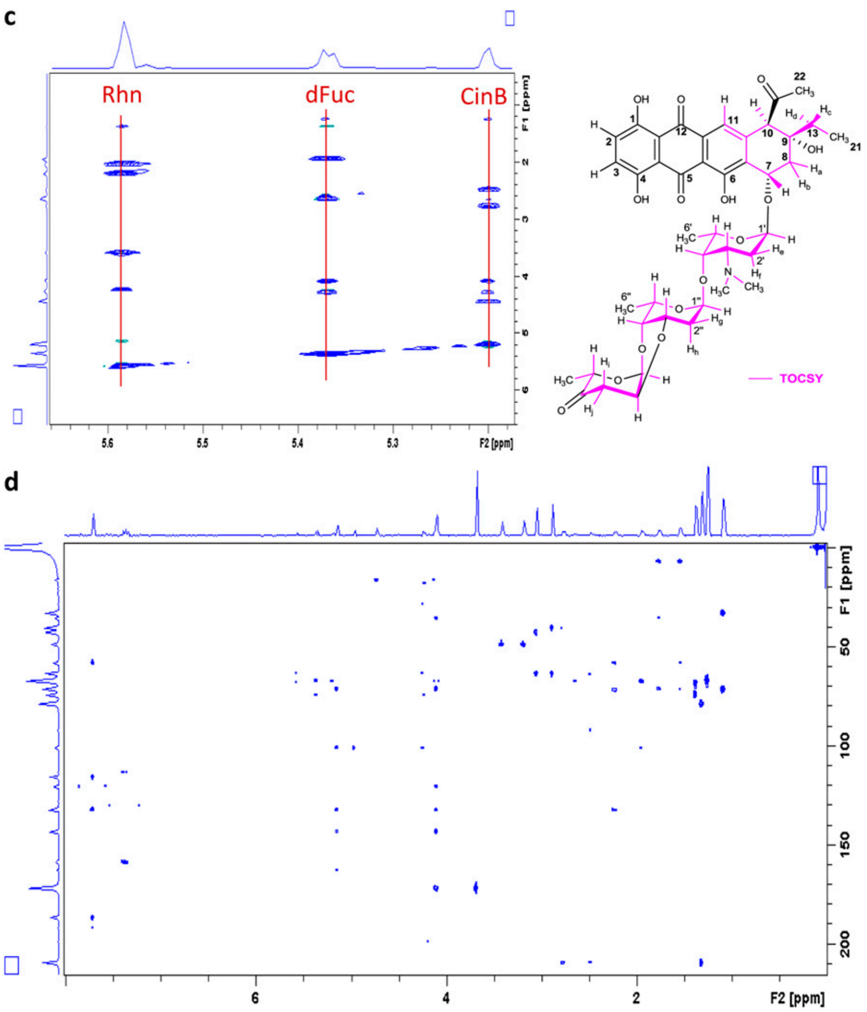


Fig. S4. (Continued)

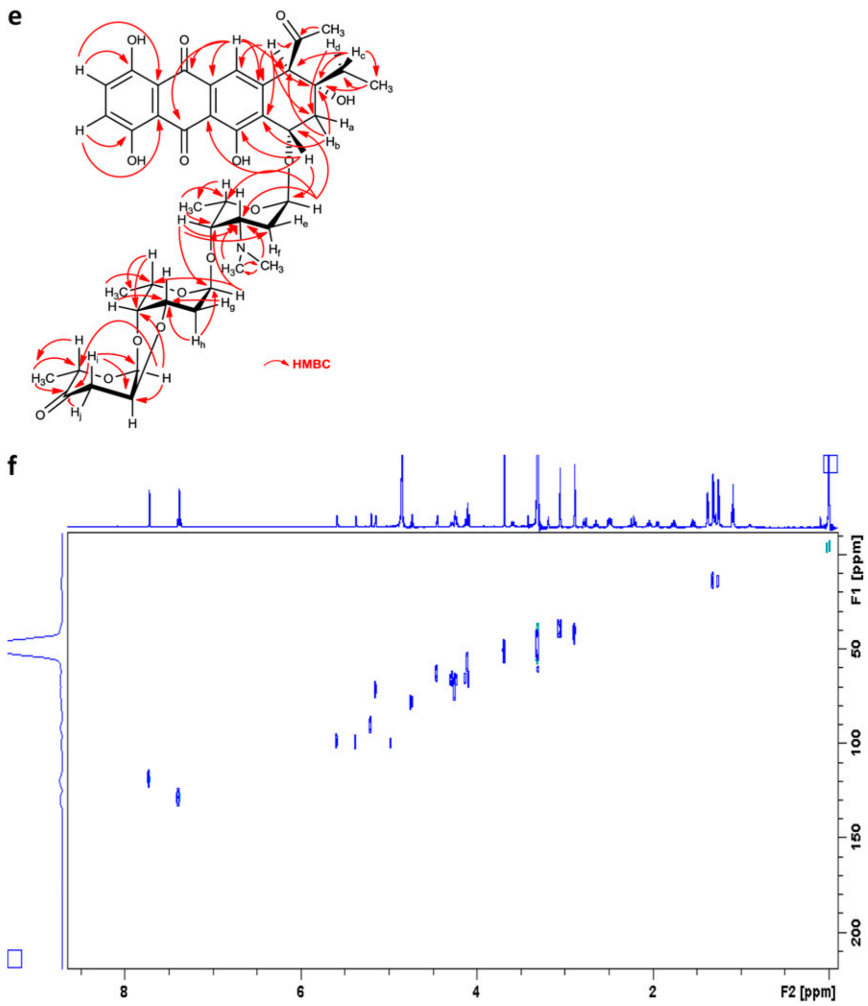


Fig. S4. (Continued)

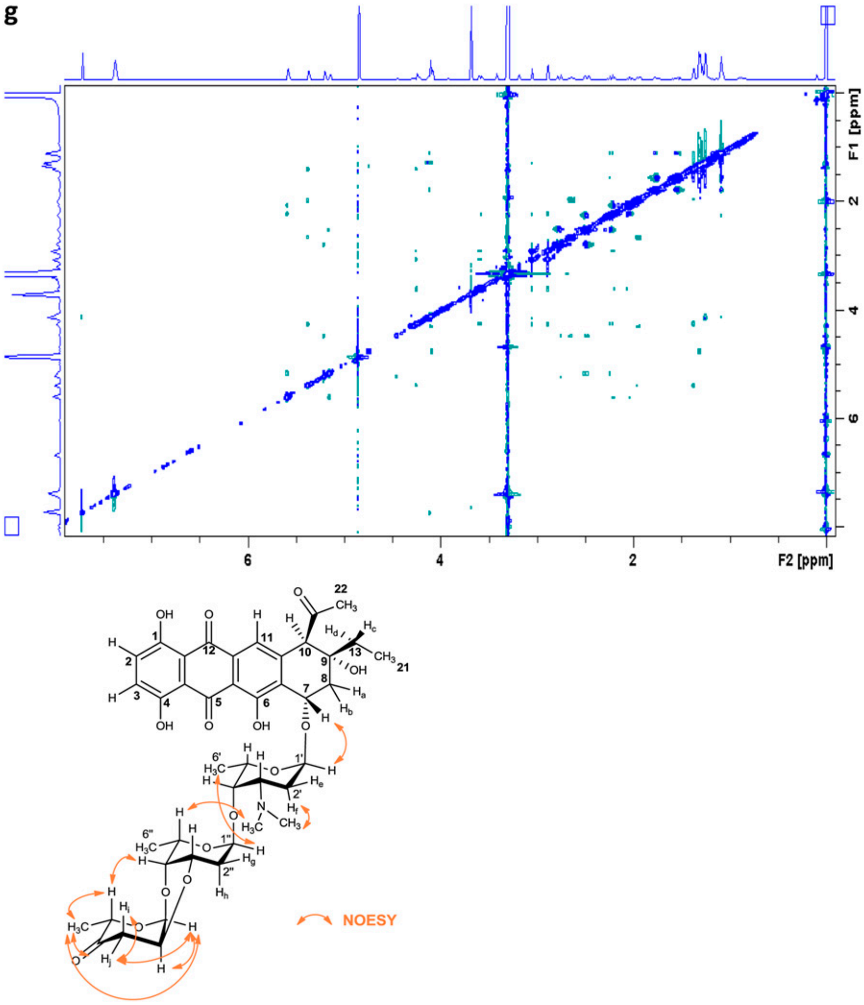
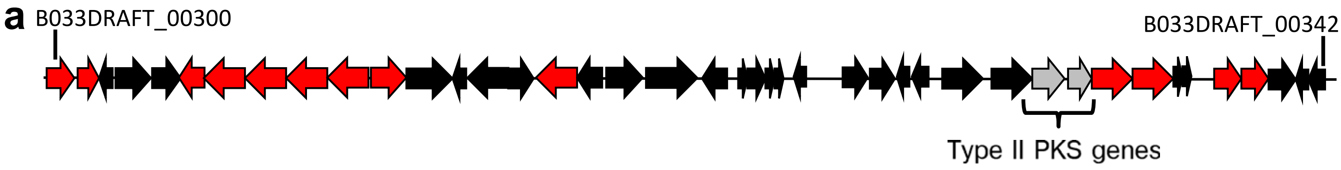


Fig. S4. NMR spectra of cinerubin B (1-hydroxyaclacinomycin A). 1D and 2D NMR analysis of cinerubin B could verify the candidate deoxysugars of the glycogenomic analysis as rhodosamine (Rhn), 2′-deoxyfucose (dFuc) and cinerulose B (CinB), which is attached via a 1‴,2‴-*O*,*O*-di-glycosidic bond to 2′-deoxyfucose. The sugar stereochemistry was assigned based on a ^{1}H-^{1}H NOESY experiment. All spectra were observed in MeOD-d4, 600 MHz, 298 K. (*A*) ^{1}H NMR spectrum. The detailed annotations are listed in Table S1. (*B* and *C*) ^{1}H-^{1}H TOCSY spectra and annotations. The spectrum was observed with a mixing time of 90 ms. *B* is a full spectrum; *C* is a zoom in the spectrum with annotations of the sugar spin systems. (*D* and *E*) ^{1}H-^{13}C heteronuclear multiple-bond correlation spectroscopy (HMBC) spectrum and annotations. The spectrum was observed with the polarization transfer delay optimized for $^{2,3}J_{1H/13C} = 7$ Hz. *D* is a full spectrum, and *F* shows HMBC annotations. (*F*) ^{1}H-^{13}C heteronuclear single-quantum coherence (HSQC) spectrum with annotations. The spectrum was observed with the polarization transfer delay optimized for $^{1}J_{1H/13C} = 145$ Hz. *E* is a full spectrum, and *G* shows HMBC annotations. (*G*) ^{1}H-^{1}H NOESY spectrum and annotations.

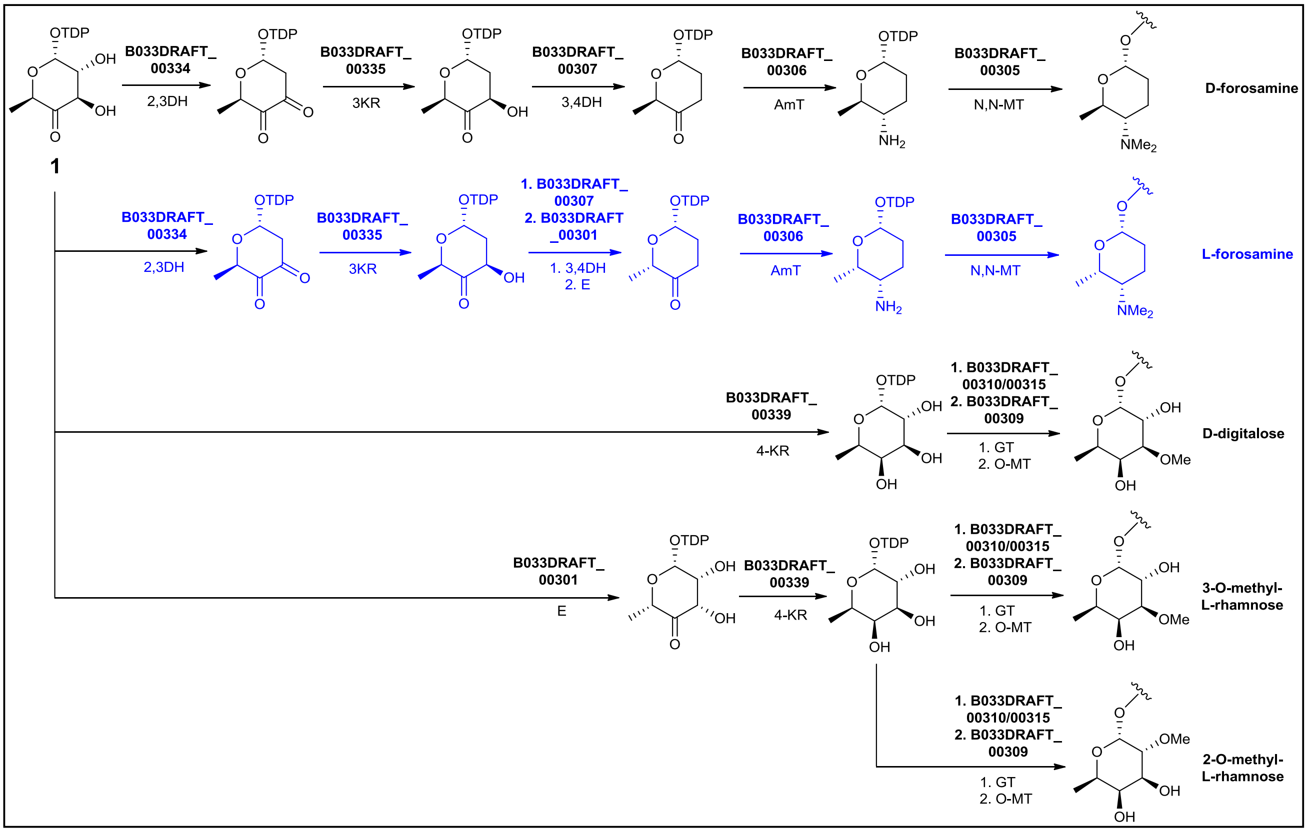


Gene	Size [aa]	Predicted function	Closest homolog (Sim/Id, %/%)	Reference	Pradimicin pathway homolog
B033DRAFT_00300	330	NT	dTDP-glucose synthase [Micromonospora griseorubida] (82/73)	BAC57039	
B033DRAFT_00301	218	E	dTDP-4-dehydrorhamnose 3,5-epimerase [Micromonospora lupini str. Lupac 08] (82/68)	ZP_21030398	
B033DRAFT_00302	169	regulator	MarR family transcriptional regulator [Stackebrandtia nassauensis DSM 44728] (59/46)	YP_003512399	
B033DRAFT_00303	316	transporter	daunorubicin ABC transporter ATPase [Actinosynnema mirum DSM 43827] (74/64)	YP_003098925	
B033DRAFT_00304	280	transporter	ABC transporter [Actinosynnema mirum DSM 43827] (71/53)	YP_003098924	
B033DRAFT_00305	247	N,N-MT	probable N-dimethyltransferase [Saccharopolyspora spinosa] (75/60)	AAG23280	
B033DRAFT_00306	386	AmT	probable aminotransferase [Saccharopolyspora spinosa] (79/65)	AAG23279	
B033DRAFT_00307	448	3,4-DH	putative 3-dehydratase [Streptomyces galilaeus] (86/77)	AAF73452	
B033DRAFT_00308	401	GT-like (P450)	AknT [Streptomyces galilaeus] (44/36)	AAF73456	
B033DRAFT_00309	376	O-MT	putative sugar O-methyltransferase [Streptomyces sp. SANK 60405] (56/43)	BAJ05903	
B033DRAFT_00310	411	GT	glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (63/45)	YP_001102993	
B033DRAFT_00311	490	deacetylase	acetylornithine deacetylase [Microvirga sp. WSM3557] (56/36)	ZP_10180744	
B033DRAFT_00312	114	NK	nucleoside diphosphate kinase [Listeria ivanovii FSL F6-596] (54/30)	ZP_07874356	
B033DRAFT_00313	411	cytochrome P450	cytochrome P450 [Streptomyces davawensis JCM 4913] (57/40)	CCK25399	
B033DRAFT_00314	243	hypothetical protein	LmbE family protein [Acidothermus cellulolyticus 11B] (68/58)	YP_872134	
B033DRAFT_00315	383	GT	ChaGT1 protein [Streptomyces chartreusis] (56/41)	CAH10164	
B033DRAFT_00316	325	reductase	aldo/keto reductase [Streptomyces flavogriseus ATCC 33331] (64/51)	YP_004926245	
B033DRAFT_00317	346	methyltransferase	PdmT [Actinomadura hibisca] (60/47)	ABK58689	
B033DRAFT_00318	540	monooxygenase	putative FAD-dependent monooxygenase GrhO8 [Streptomyces sp. JP95] (69/54)	AAM33675	
B033DRAFT_00319	276	regulator	DNA-binding transcriptional activator of the SARP family [Frankia sp. QA3] (55/42)	ZP_10309867	
B033DRAFT_00320	117	cyclase	polyketide cyclase [Streptomyces sp. SANK 61196] (74/59)	ADG86313	PdmK
B033DRAFT_00321	131	cyclase	Cupin 2 barrel domain-containing protein [Micromonospora aurantiaca ATCC 27029] (69/53)	YP_003836600	PdmL
B033DRAFT_00322	161	cyclase	Aromatase WhiE VI [Streptomyces gancidicus BKS 13-15] (78/67)	EMF26754	PdmD
B033DRAFT_00323	120	oxygenase	antibiotic biosynthesis monooxygenase [Amycolatopsis mediterranei U32] (68/47)	YP_003767590	PdmI
B033DRAFT_00324	120	biotin carboxylase	acetyl-CoA carboxylase, carboxyl transferase, beta subunit [Frankia sp. EUN1f] (80/67) (PrmP1)	ZP_06417970	
B033DRAFT_00325	286	regulator	SARP family regulator [Streptomyces lavendulae] (68/50)	BAG74714	
B033DRAFT_00326	280	methyltransferase	ChaI protein [Streptomyces chartreusis] (56/42)	CAH10176	
B033DRAFT_00327	139	hypothetical protein	hypothetical protein RER_29830 [Rhodococcus erythropolis PR4] (55/40)	YP_002766430	
B033DRAFT_00328	151	oxygenase	RubQ [Streptomyces collinus] (64/49)	AAM97373	
B033DRAFT_00329	412	cytochrome P450	cytochrome P450 [Streptomyces avermitilis MA-4680] (57/43)	NP_823237	
B033DRAFT_00331	422	minPKS, KS	PdmA [Actinomadura hibisca] (83/74)	ABM21747	PdmA
B033DRAFT_00332	405	minPKS, CLF	PdmB [Actinomadura hibisca] (74/65)	ABM21748	PdmB
B033DRAFT_00333	241	reductase	putative 3-oxoacyl-ACP reductase [Streptomyces sp. TA-0256] (75/61)	ABM21753	PdmG
B033DRAFT_00334	458	2,3DH	NDP-hexose 2,3-dehydratase [Frankia sp. CcI3] (75/62)	YP_483211	
B033DRAFT_00335	332	3KR	NDP-hexose-3-ketoreductase [Streptomyces lydicus] (68/57)	CBA11563	
B033DRAFT_00336	84	minPKS, ACP	acyl carrier protein [Streptomyces arenae] (71/54)	AAD20269	PdmC
B033DRAFT_00337	108	oxygenase	PdmH [Actinomadura hibisca] (72/58)	ABM21754	PdmH
B033DRAFT_00338	339	4,6DH	dTDP-glucose 4,6-dehydratase [Streptomyces sp. SirexAA-E] (83/73)	YP_004802450	
B033DRAFT_00339	283	4KR	putative dTDP-4-keto-6-deoxyhexose reductase [Actinoplanes missouriensis 431] (80/74)	YP_005462169	
B033DRAFT_00340	206	regulator	Transcriptional regulator ArsR family [Patulibacter sp. I11] (61/42)	ZP_09470860	
B033DRAFT_00341	69	carboxylase	carboxylase [Streptomyces ghanaensis ATCC 14672] (65/54)	ZP_06581973	
B033DRAFT_00342	243	reductase	putative NADPH-dependent FMN reductase [Actinoplanes missouriensis 431] (80/72)	YP_005464808	

Fig. S5. (Continued)

b

Y-Mass shift [Da]	B-mass shift [Da]	C-mass shift [Da]	Sugar	Common genes	Specific BS genes									
					DH			KR		IM	AmT	MT		
				NT	4,6-DH	2,3-DH	3,4-DH	3-KR	4-KR	(3,5-)E	4-AmT	C3-MT	N,N-MT	O-MT
141.115363	142.122639	160.133204	D-forosamine	▨	▨	▨	▨	▨			▨		▨	
160.07356	161.080836	179.091401	D-digitalose	▨	▨				▨					▨
160.07356	161.080836	179.091401	3-O-methyl-rhamnose	▨	▨				▨	▨				▨
160.07356	161.080836	179.091401	2-O-methyl-L-rhamnose	▨	▨				▨	▨				▨
160.07356	161.080836	179.091401	6-deoxy-3-C-methyl-L-mannose	▨	▨				▨	▨		▨		



c

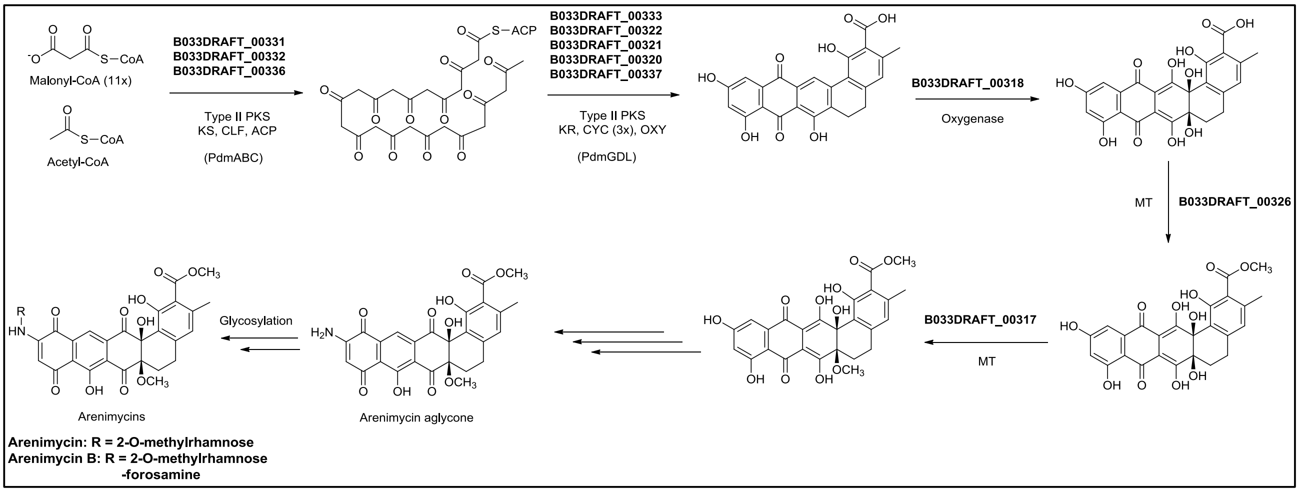


Fig. S5. Glycogenomic connection of arenimycin B with its biosynthetic gene cluster from *Salinispora arenicola* CNB-527. (*A*) Gene cluster analysis of candidate arenimycin B pathway, with highlighted glycosylation genes (red) and aglycone biosynthetic genes (gray). (*B*) Matching pathways of specific glycosylation genes with candidate MS/MS forosamine sugar and O-methyldeoxysugars from arenimycin B starting at deoxysugar biosynthetic intermediate TDP-4-keto-6-deoxy-α-D-glucose (1). Characterized L-forosamine would include a putative epimerase-catalyzed step in its biosynthesis (blue). (*C*) Biosynthetic proposal of the arenimycins based on their biosynthetic genes. Abbreviations: ACP, acyl carrier protein; CLF, chain length factor; CYC, cyclase; KR, ketoreductase; MT, methyltransferase; OXY, oxygenase; PKS, polyketide synthase.

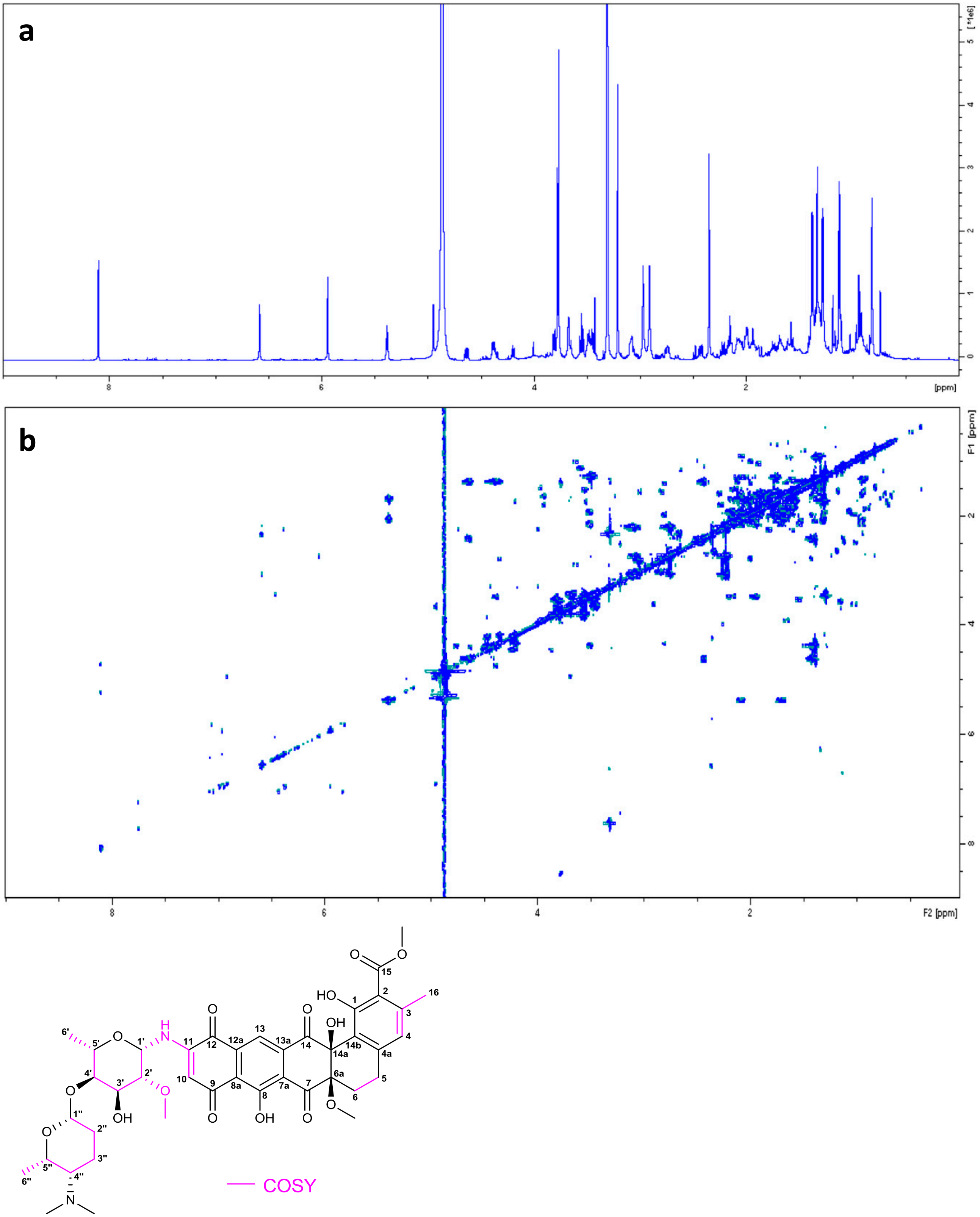


Fig. S6. (Continued)

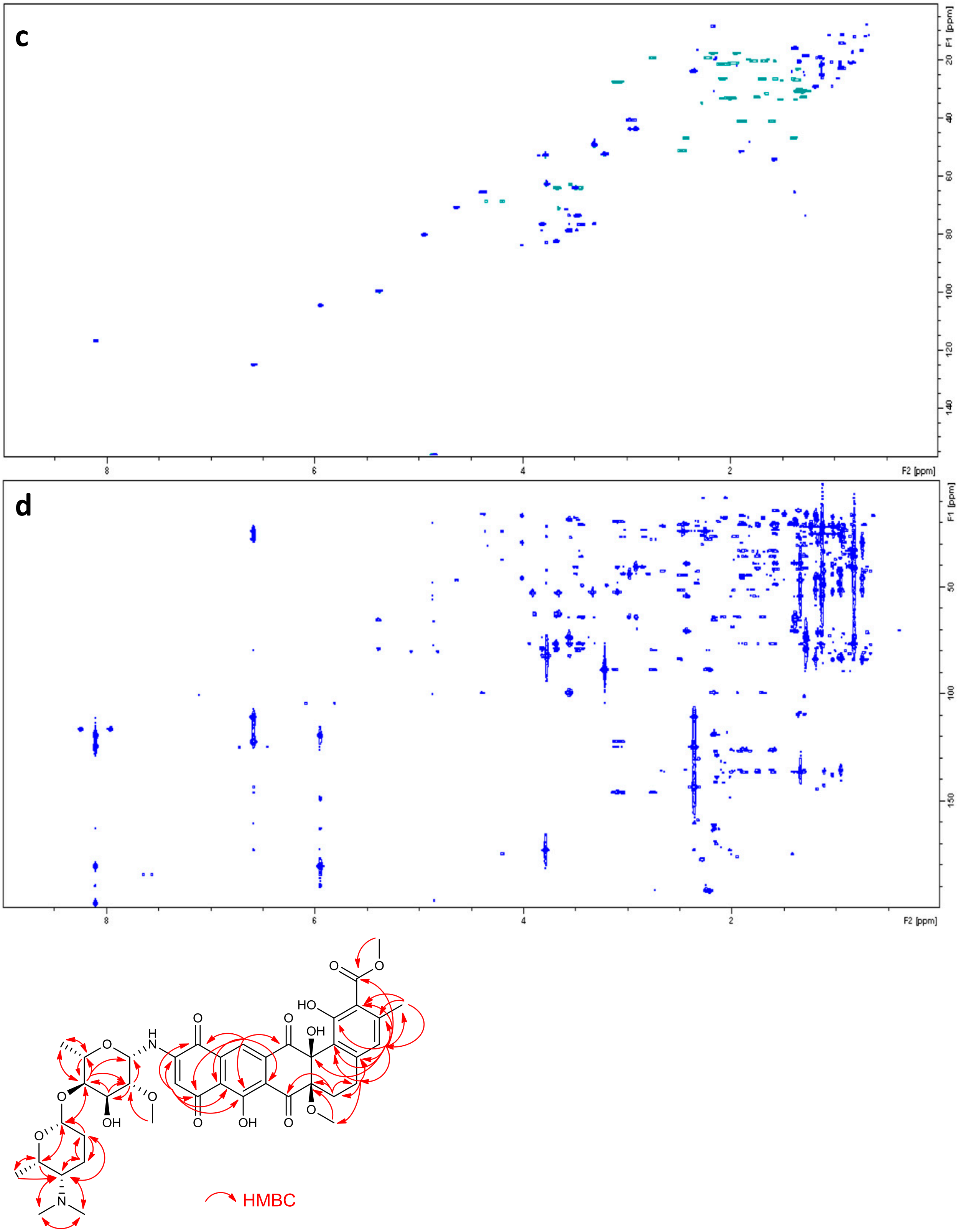


Fig. S6. (Continued)

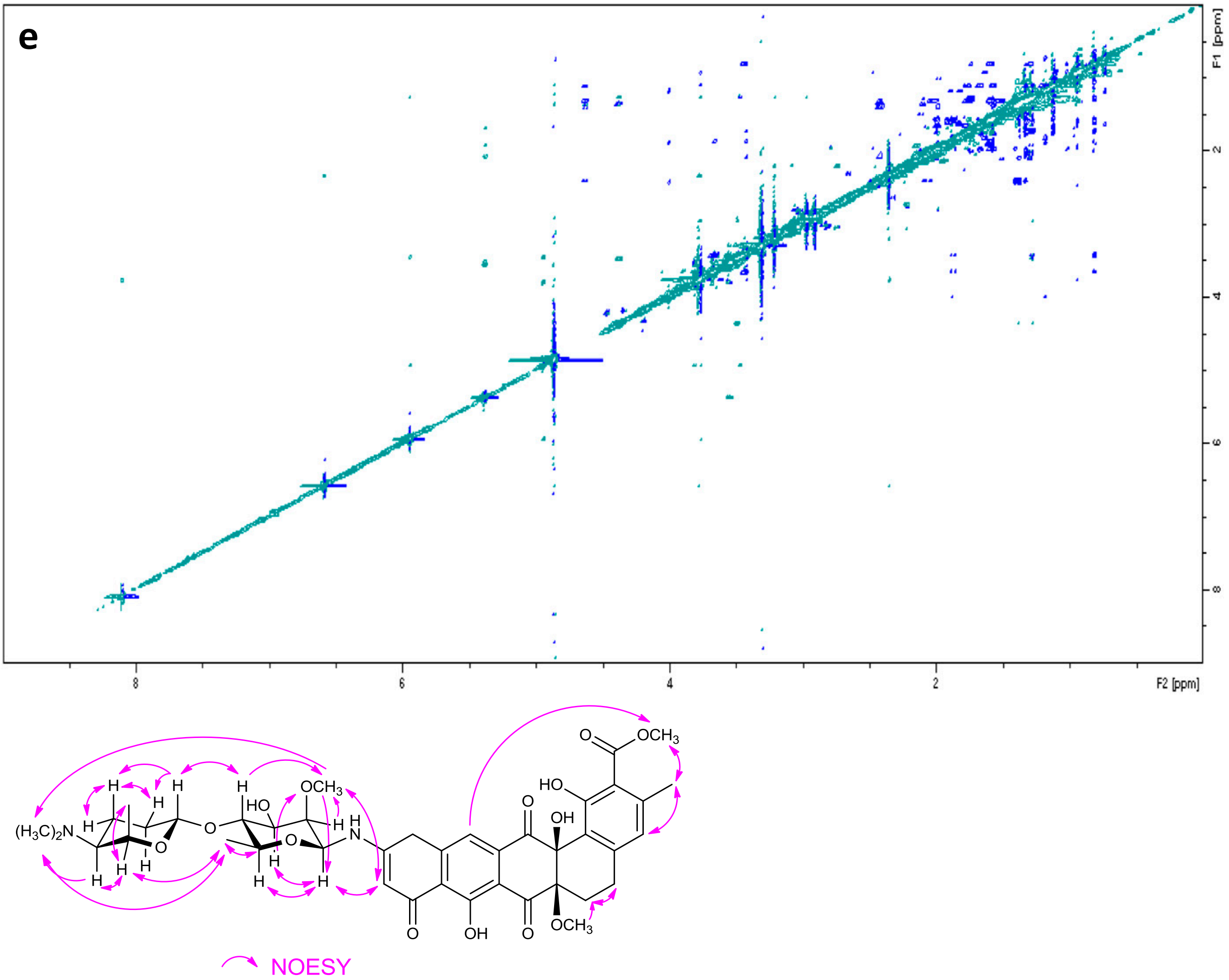


Fig. S6. NMR analysis of arenimycin B. All spectra were observed in MeOD-d4, 600 MHz, 298 K. (*A*) ^{1}H NMR spectrum. The detailed annotations are listed in Table S2. (*B*) ^{1}H-^{1}H double-quantum–filtered correlation spectroscopy (DQF-COSY) spectrum and annotations. (*C*) ^{1}H-^{13}C heteronuclear single-quantum coherence (HSQC) spectrum. (*D*) ^{1}H-^{13}C HMBC spectrum and annotations. (*E*) ^{1}H-^{1}H NOESY spectrum and annotations.

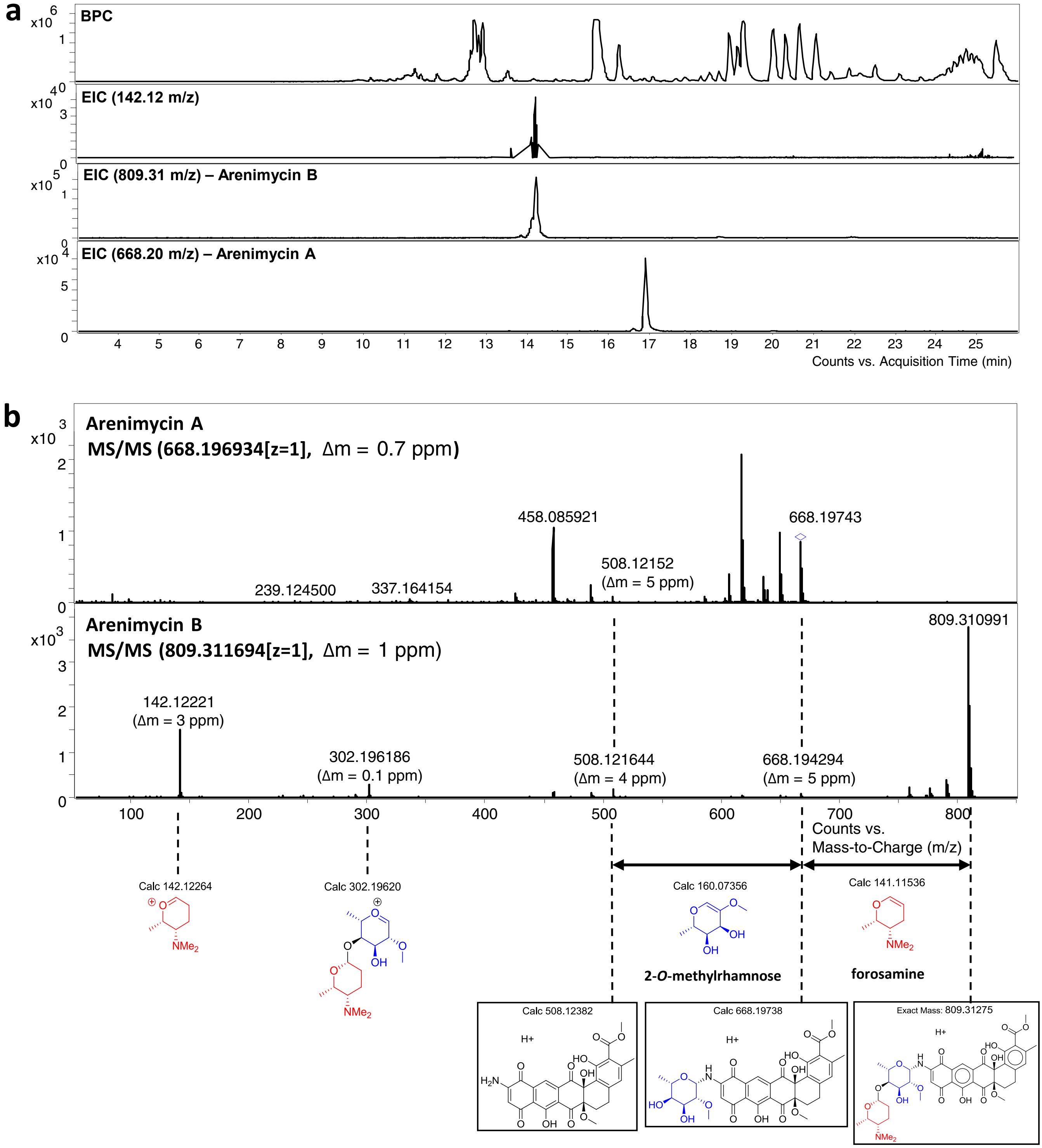


Fig. S7. Liquid chromatography (LC)-MS and MS/MS characterization of arenimycins. (*A*) LC-MS profiles of a crude ethylacetate extract of *S. arenicola* CNB-527. Abbreviations: BPC, base peak chromatogram; EIC, extracted ion chromatogram. (*B*) MS/MS spectra of arenimycin A (*Upper*) and arenimycin B (*Lower*) with structural peak assignments.

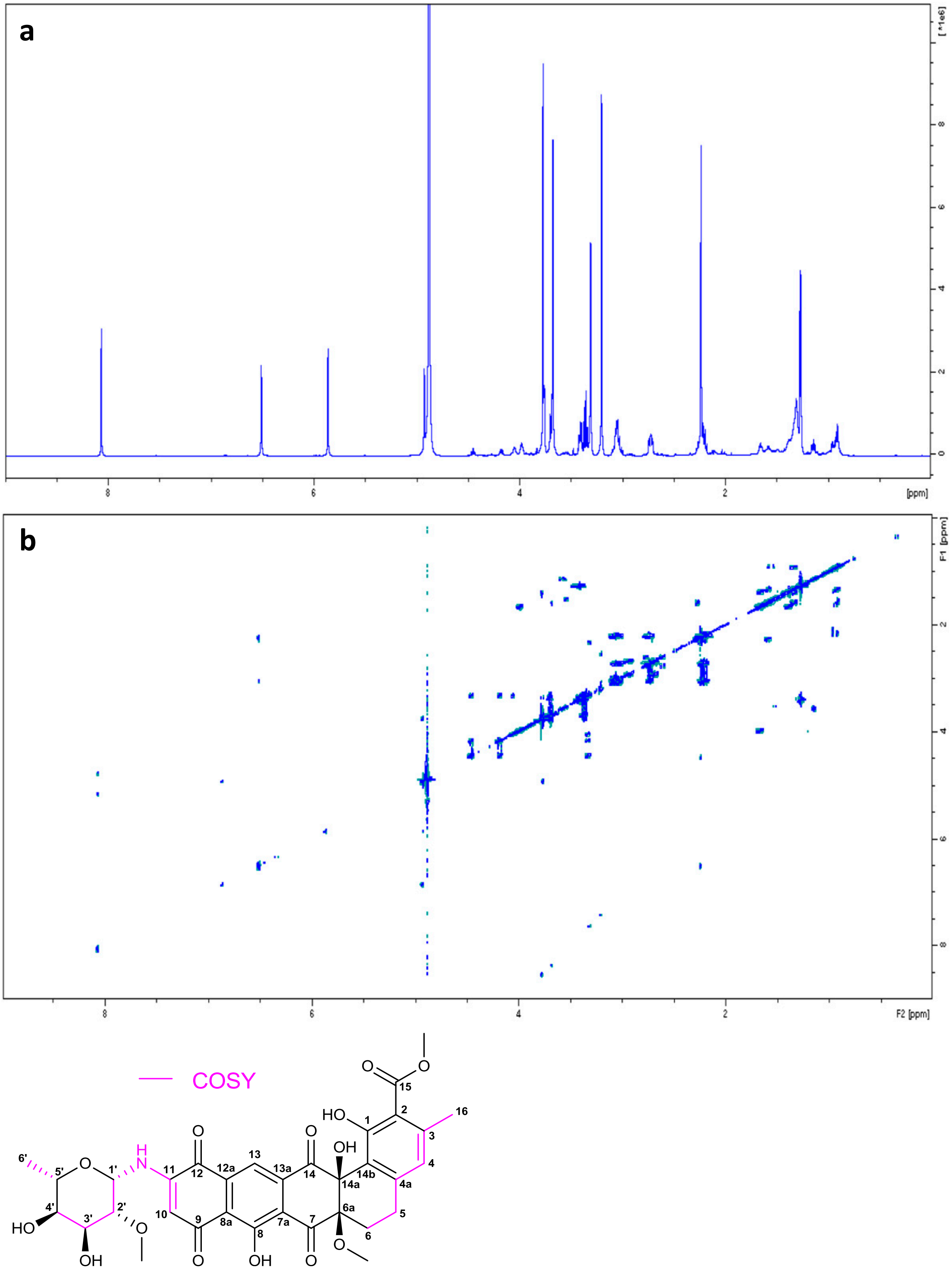


Fig. S8. (Continued)

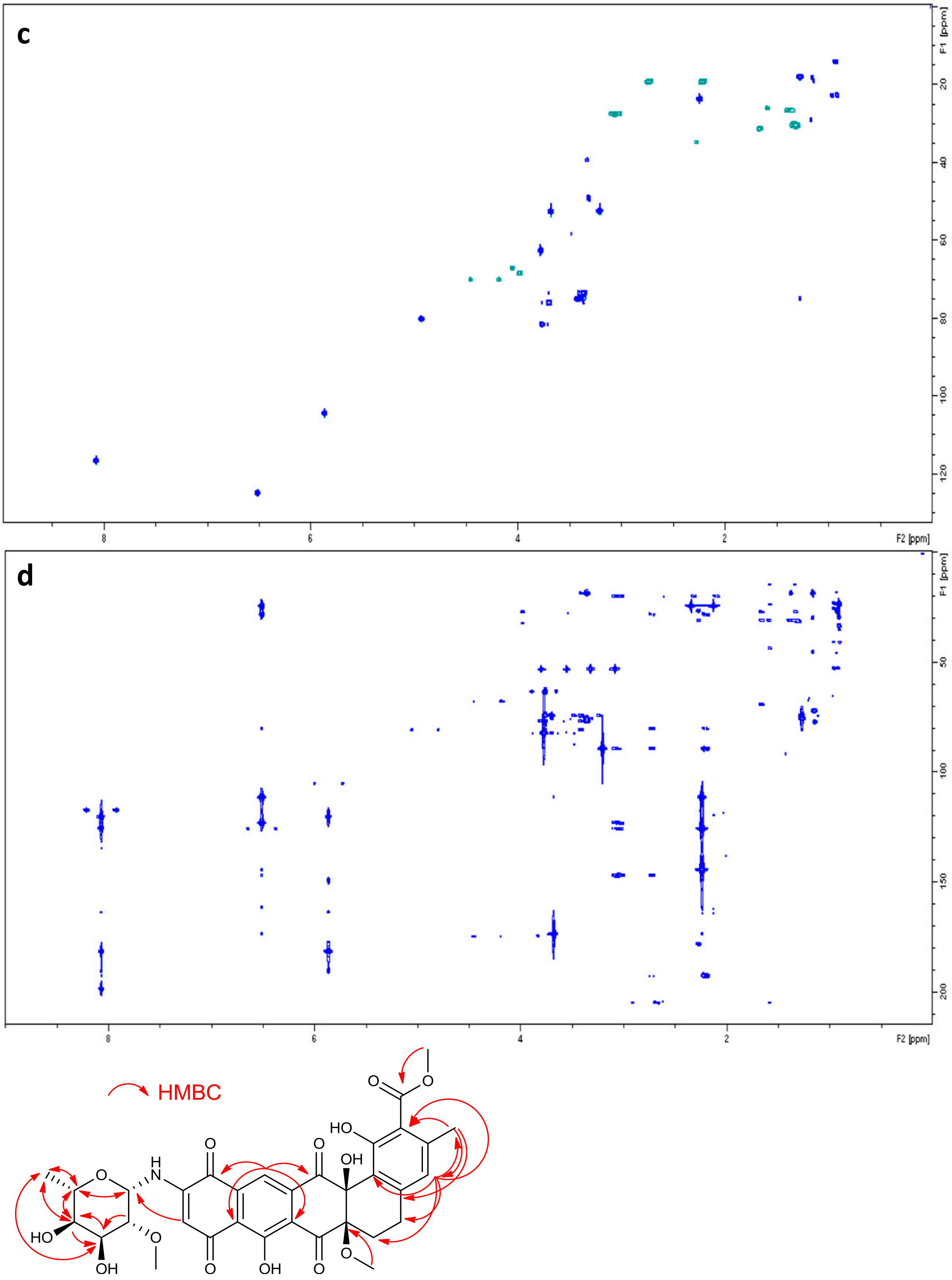


Fig. S8. (Continued)

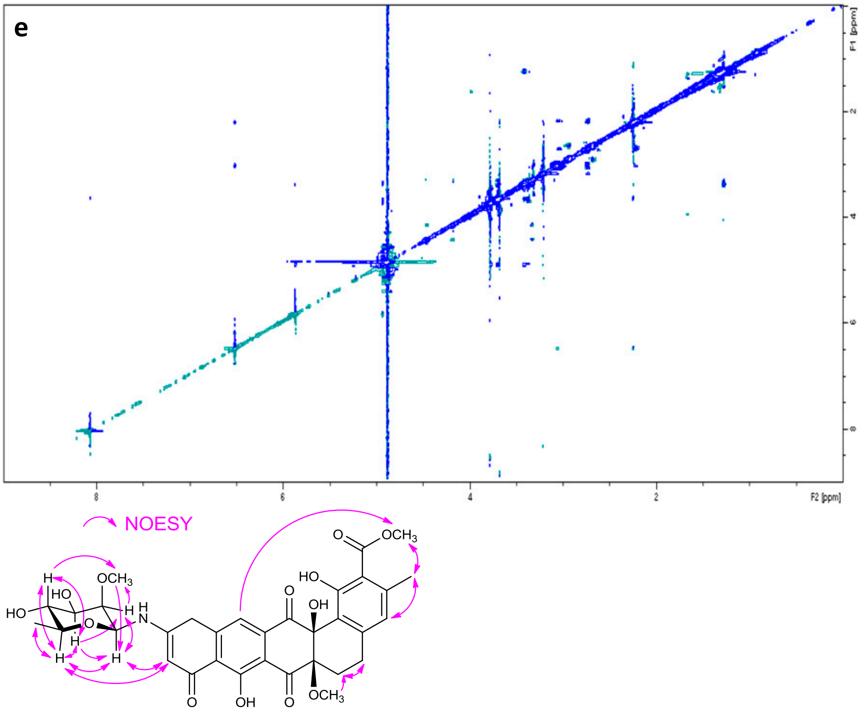


Fig. S8. NMR analysis of arenimycin A. All spectra were observed in MeOD-d4, 600 MHz, 298 K. (*A*) ^{1}H NMR spectrum. The detailed annotations are listed in Table S5. (*B*) ^{1}H-^{1}H DQF-COSY spectrum and annotations. (*C*) ^{1}H-^{13}C HSQC spectrum. (*D*) ^{1}H-^{13}C HMBC spectrum and annotations. (*E*) ^{1}H-^{1}H NOESY spectrum and annotations.

Table S1. NMR analysis of cinerubin B (1-hydroxyaclacinomycin A) in MeOD-d4

Site		δC, ppm	δH, ppm	Signal	*J*(H-H), Hz	Site		δC, ppm	δH, ppm	Signal	*J*(H-H), Hz
1		158.9				Rhodosamine					
2		130.3	7.39	dd	9.5, 13.4	1′		101.3	5.59	d	4
3			7.39	dd	9.5, 13.4	2′	(e)	28.6	2.04	dt	4, 12.8
4		158.9					(f)		2.22	m	4
4-OH			N/A			3′		63.8	12.8	d	12.8
4a		113.9				4′		74.5		br	
5		192.2				5′		68.4	4.24	q	6.6
5a						6′	(Me)	17.8	1.38	d	6.6
6		163.1				3′-NMe_2	(Me)	2.89	43	s	
6-OH			N/A					3.05	40.7	s	
6a		143.9				2′-Deoxyfucose					
7		72.2	5.15	d	4.4	1″		101.3	5.38	d	4.4
8	(a)	35.6	2.5	m		2″	(g)	26.5	2.65	dt	4.4, 18.9
	(b)		2.23	m			(h)		1.95	dd	4.8, 12.3
9		71.7				3″		67.5	4.29	dt	4.0, 12.3
9-OH			N/A			4″		67.5	4.09	br	
10		58.1	4.11	s		5″		67.5	4.13	q	6.6
10a		132.7				6″	(Me)	16.4	1.26	d	6.6
11		120.5	7.72	s		3″-OH			N/A		
11a		116.3				Cinerulose B					
12		187				1‴		92.4	5.2	d	3
12a		113.9				2‴		64.2	4.45	dd	3
13	(c)	33.3	1.77	dt	21.6, 7.0	3‴	(i)	53	2.48	m	
	(d)		1.55	dt	21.8, 7.0		(j)		2.77	dd	2.6, 14.5
15						4‴		210			
21	(Me)	6.6	1.09	t	7.3	5‴		79.2	4.74	q	6.6
22	(MeO)	51.7	3.69	s		6‴	(Me)	16.4	1.31	d	6.6

Abbreviations: br, broad; d, doublet; m, multiplet; Me, methyl; MeO, methoxy; N/A, not annotated, the hydroxyl-protons were solvent exchanged and, thus, not observed; q, quartet; s, singlet; t, triplet.

Table S2. NMR analysis of arenimycin B at 600 MHz in MeOD-d4

Position	δH	δH—integral, signal (*J* value)	δC	HMBC (H→C)	COSY	NOESY
1	—	—	160.9	—	—	—
2	—	—	111.0	—	—	—
3	—	—	143.7	—	—	—
4	6.58	1H, s	125.0	C-1, -2, -3, -4a, -5, -14a, -14b, -15, -16	H-16	H-16
4a	—	—	146.3	—	—	—
5	3.09/3.13	1H, m	27.6	C-4a, -6, -6(OCH$_3$), -14b	H-6	H-6
6	2.23/2.75	1H, m/1H, m	19.5	C-5, -6a, -7	H-5	H-5
6a	—	—	88.9	—	—	—
6a (OCH$_3$)	3.21	3H, s	52.6	C-6a	—	H-6
7	—	—	192.2	—	—	—
7a	—	—	124.5	—	—	—
8	—	—	163.1	—	—	—
8a	—	—	119.6	—	—	—
9	—	—	190.0	—	—	—
10	5.94	1H, s	104.7	C-8, -8a, -9, -11, -12	H-1′	H-1′, -5′, -6′
11	—	—	149.0	—	—	—
12	—	—	180.7	—	—	—
12a	—	—	N/A	—	—	—
13	8.10	1H, s	116.8	C-7a, -8, -8a, -9, -12, -14	—	H-15(OCH$_3$)
13a	—	—	N/A	—	—	—
14	—	—	197.9	—	—	—
14a	—	—	79.6	—	—	—
14b	—	—	122.6	—	—	—
15	—	—	173.1	—	—	—
15 (OCH$_3$)	3.78	3H, s	52.9	C-15	—	H-4, -16
16	2.35	3H, s	23.9	C-2, -3, -4	H-4	H-4, -15(OCH$_3$)
1′	4.95	1H, br	80.4	—	H-10, -2′	H-10, -3′, -5′
2′	3.67	1H, m	82.6	H-1′, -3′	H-2′(OCH$_3$), -3′, -4′	H-2′(OCH$_3$)
2′ (OCH$_3$)	3.77	3H, s	63.0	C-2′	—	H-1′, -6′, -1″, -4″(N(CH$_3$)$_2$)
3′	3.79	1H, dd (3.3, 9.6)	76.6	C-2′, -4′	H-2′, -4′	H-1′, -2′(OCH$_3$)
4′	3.55	1H, m	78.9	C-2′, -3′, -5′, -6′, -1″	H-3′, -5′	H-2′(OCH$_3$), -1″
5′	3.46	1H, m	73.7	C-4′, -6′	H-4′, -6′	H-1′, -6′
6′	1.28	3H, d (5.9)	18.6	C-4′, -5′	H-5′	H-5′, -5′′, -4″(NCH$_3$)$_2$)
1″	5.38	1H, t (5.0)	99.6	C-4′, -5″	H-2′	H-4′, -2″, -3″ -5″
2″	1.67/2.05	m	26.6	C-1″, -3″, -4″	H-1″, -3″	H-3″
3″	1.94/2.16	m	17.8	C-2″, -4″	H-2″, -4″	H-2″
4″	3.49	m	64.2	C-2″, -3″, -4″(N(CH$_3$)$_2$)	H-3″, -5″	H-4″(N(CH$_3$)$_2$), -5″
4″ (N(CH$_3$)$_2$)	2.91/2.97	3H, s/3H, s	43.9/40.8	C-4″(N(CH$_3$)$_2$), -4″	—	H-6′
5″	4.38	m	65.7	C-1″, -4″, -6″	H-4″, -6″	H-6′, -4″, -6″
6″	1.38	3H, d (7.0)	16.1	C-4″, -5″	H-5″	H-4″(N(CH$_3$)$_2$), -5″

Abbreviations: br, broad; d, doublet; *J*, coupling constant in Hertz [Hz]; m, multiplet; s, singlet; t, triplet. See Fig. S6 for positions.

Table S3. Comparison of NMR data of arenimycins

	δH, ppm (integral; signal; *J* value, Hz)			δC, ppm		
Position	Arenimycin A [Lit. (1), 500 MHz, $CDCl_3$]	Arenimycin A (600 MHz, MeOD-d4)	Arenimycin B (600 MHz, MeOD-d4)	Arenimycin A [Lit. (1), 500 MHz, $CDCl_3$]	Arenimycin A (600 MHz, MeOD-d4)	Arenimycin B (600 MHz, MeOD-d4)
1	—	—	—	160.3	160.9	160.9
2	—	—	—	109.6	110.9	111.0
3	—	—	—	142.8	143.6	143.7
4	6.49 (1H, s)	6.51 (1H, s)	6.58 (1H, s)	124.1	125.0	125.0
4a	—	—	—	145.0	146.4	146.3
5	3.08/3.15 (2H, ddd, 19.2, 8.6, 2.0/ 19.2, 9.8, 7.2)	3.06/3.06 (2H, m)	3.09/3.13 (1H, m)	26.8	27.7	27.6
6	2.24/2.75 (2H, ddd, 12.3, 9.8, 8.6/ 12.3, 7.2, 2.0)	2.21/2.73 (1H/1H, m)	2.23/2.75 (1H/1H, m)	18.9	19.3	19.5
6a	—	—	—	87.4	88.9	88.9
6a (OCH_3)	3.22 (3H, s)	3.20 (3H, s)	3.21 (3H, s)	52.2	52.7	52.6
7	—	—	—	190.1	192.0	192.2
7a	—	—	—	124.0	124.6	124.5
8	—	—	—	162.8	163.1	163.1
8a	—	—	—	118.8	119.6	119.6
9	—	—	—	189.0	189.9	190.0
10	5.86 (1H, s)	5.86 (1H, s)	5.94 (1H, s)	104.6	104.7	104.7
11	—	—	—	147.4	148.9	149.0
12	—	—	—	179.1	180.7	180.7
12a	—	—	—	136.5	135.5	N/A
13	8.20 (1H, s)	8.07 (1H, s)	8.10 (1H, s)	116.2	116.9	116.8
13a	—	—	—	140.9	N/A	N/A
14	—	—	—	198.1	197.8	197.9
14a	—	—	—	78.0	79.6	79.6
14b	—	—	—	121.2	122.5	122.6
15	—	—	—	172.4	173.1	173.1
15 (OCH_3)	3.81 (3H, s)	3.68 (3H, s)	3.78 (3H, s)	51.9	52.8	52.9
16	2.35 (3H, s)	2.24 (3H, s)	2.35 (3H, s)	23.9	23.9	23.9
1′	4.69 (1H, d, 9.0)	4.93 (1H, br)	4.95 (1H, br)	79.4	80.4	80.4
2′	3.67 (1H, m)	3.76 (1H, m)	3.67 (1H, m)	80.2	81.9	82.6
2′ (OCH_3)	3.75 (3H, s)	3.77 (3H, s)	3.77 (3H, s)	62.8	62.9	63.0
3′	3.46 (1H, m)	3.36 (1H, m)	3.79 (1H, dd, 3.3, 9.6)	72.1	73.8	76.6
4′	3.65 (1H, m)	3.70 (1H, m (9.2))	3.55 (1H, m)	75.4	76.3	78.9
5′	3.34 (1H, dq, 9.2, 5.9)	3.41 (1H, dq, 9.2, 6.1)	3.46(1H, m)	73.7	75.2	73.7
6′	1.34 (3H, d, 5.9)	1.28 (3H, d, 6.1)	1.28 (3H, d, 5.9)	18.4	18.1	18.6
1″	—	—	5.38 (1H, t, 5.0)	—	—	99.6
2″	—	—	1.67/2.05 (m)	—	—	26.6
3″	—	—	1.94/2.16 (m)	—	—	17.8
4″	—	—	3.49 (m)	—	—	64.2
4″ ($N(CH_3)_2$)	—	—	2.91/2.97 (3H/3H, s)	—	—	43.9/40.8
5″	—	—	4.38 (m)	—	—	65.7
6″	—	—	1.38 (3H, d, 7.0)	—	—	16.1

Abbreviations: br, broad; d, doublet; *J*, coupling constant in Hertz; m, multiplet; N/A, not available; s, singlet; t, triplet. See Fig. S8 for positions.

1. Asolkar RN, Kirkland TN, Jensen PR, Fenical W (2010) Arenimycin, an antibiotic effective against rifampin- and methicillin-resistant *Staphylococcus aureus* from the marine actinomycete *Salinispora arenicola*. *J Antibiot (Tokyo)* 63(1):37–39.

Table S4. Antibacterial and anticancer bioactivity tests of arenimycins

Antibacterial bioactivity test		MIC, μg/mL		MAC, μg/mL	
Strain	Details	Are A	Are B	Are A	Are B
Staphylococcus aureus TCH1516	USA300 CA-MRSA (ATCC)	67	25	50	25
Staphylococcus aureus Sanger 252	USA200 HA-MRSA (NARSA)	42	17	42	17
Staphylococcus aureus MSSA	ATCC 29213	83	33	100	33
Escherichia coli MCR106 *imp*		100	100	100	100
Pseudomonas aeruginosa		>100	>100	>100	>100
Anticancer bioactivity test		IC50, μg/mL			
HCT-116 adenocarcinoma cells		Are A	Are B		
		0.251	0.411		

Abbreviations: Are A, arenimycin A; Are B, arenimycin B; IC50, half-maximal inhibitory concentration; MBC, minimum bactericidal concentration; MIC, minimum inhibitory concentration.

Table S5. NMR analysis of arenimycin A at 600 MHz in MeOD-d4

Position	δH, ppm	δH—integral, signal, *J* value	δC, ppm	HMBC (H→C)	COSY	NOESY
1	—	—	160.9	—	—	—
2	—	—	110.9	—	—	—
3	—	—	143.6	—	—	—
4	6.51	1H, s	125.0	C-1, -2, -3, -4a, -5, -14a, -14b, -15, -16	H-5, -16	H-5, -15($\mathrm{OCH_3}$)
4a	—	—	146.4	—	—	—
5	3.06/3.06	2H, m	27.7	—	H-6	H-4, -6
6	2.21/2.73	1H/1H, m	19.3	C-2, -4a, -5, -6a, -7, -14a	H-5	H-5
6a	—	—	88.9	—	—	—
6a ($\mathrm{OCH_3}$)	3.20	3H, s	52.7	C-6a	—	—
7	—	—	192.0	—	—	—
7a	—	—	124.6	—	—	—
8	—	—	163.1	—	—	—
8a	—	—	119.6	—	—	—
9	—	—	189.9	—	—	—
10	5.86	1H, s	104.7	C-8, -8a, -9, -11, -12a, -12	H-1′	H-1′, -5′
11	—	—	148.9	—	—	—
12	—	—	180.7	—	—	—
12a	—	—	135.5	—	—	—
13	8.07	1H, s	116.9	C-7a, -7, -8a, -8, -9, -12, -14	—	H-15($\mathrm{OCH_3}$)
13a	—	—	N/A	—	—	—
14	—	—	197.8	—	—	—
14a	—	—	79.6	—	—	—
14b	—	—	122.5	—	—	—
15	—	—	173.1	—	—	—
15 ($\mathrm{OCH_3}$)	3.68	3H, s	52.8	C-15	—	—
16	2.24	3H, s	23.9	C-1, -2, -3, -4, -15	H-4	H-4, -15($\mathrm{OCH_3}$)
1′	4.93	1H, br	80.4	—	H-2′, -10	H-10, -2′, -5′
2′	3.76	1H, m	81.9	C-2′($\mathrm{OCH_3}$), -3′, -4′	H-3′	H-1′
2′ ($\mathrm{OCH_3}$)	3.77	3H, s	62.9	C-2′	—	—
3′	3.36	1H, m	73.8	C-4′, -5′, 6′	H-4′	H-1′, -2′, -5′
4′	3.70	1H, m (9.2)	76.3	C-2′, -3′	H-3′	H-3′, -5′
5′	3.41 (3.50)	1H, dq (9.2, 6.1)	75.2	C-1′, -3′, -4′, -6′	H-6′	H-1′, -3′, -6′
6′	1.28 (1.27)	3H, d (6.1)	18.1	C-3′, -4′, -5′	H-5′	H-5′

Abbreviations: br, broad; d, doublet; *J*, coupling constant in Hertz; m, multiplet; N/A, not available; s, singlet; t, triplet. See Fig. S8 for positions.

Dataset S1. Prediction of gene clusters of glycosylated natural products in finished actinobacterial genomes (October 2012, Joint Genome Institute database) by antiSMASH analysis of GenBank genome files and subsequent analysis of glycosylation genes in predicted gene clusters

Dataset S1

Predicted GNP pathways were differentiated by the presence or absence of specific glycosylation genes. Predicted GNP pathways are highlighted in gray and corresponding strain genomes and families in yellow.

Dataset S2. MS-glycogenetic code

Dataset S2

The gray square indicates enzyme is present in corresponding sugar pathway. The white square indicates enzyme is not present in corresponding sugar pathway. References are listed on the second sheet.

Dataset S3. Connection of known GNP chemotypes and genotypes by the MS-glycogenetic code

Dataset S3

Reference GNP chemotypes were analyzed in sugar-specific MS^n neutral losses or B-/C-ion fragments. MS/MS candidate sugars were identified based on observed sugar masses (Dataset S2). Corresponding reference GNP genotypes from GenBank were then analyzed in predicted common and specific glycosylation genes (Dataset S2). Predicted glycosylation genes were analyzed in matching the biosynthesis of the MS^n candidate sugars using Dataset S2. A GNP chemotype-to-genotype connection by the MS-glycogenetic code was a match of MS^n and genetic candidate sugars (green). GNPs with an unsuccessful chemotype–genotype connection were highlighted in red. For abbreviations, see Dataset S2.

Dataset S4. MS/MS-fragmentation of known GNP chemotypes and gene predictions of corresponding GNP genotypes

Dataset S4

MS/MS-sugar fragmentation and glycosylation gene prediction from chemotypes and genotypes of characterized glycosylated natural products (GNPs) from databases (Dataset S3) or self-acquired MS/MS data.

Dataset S1. Prediction of gene clusters of glycosylated natural products in finished actinobacterial genomes (Oct 2012, JGI database) by antiSMASH analysis of GenBank genome files and subsequent analysis of glycosylation genes in predicted gene clusters. Predicted GNP pathways were differentiated by presence or absence of specific glycosylation genes. Predicted GNP pathways are highlighted in grey and corresponding strain genomes and families in yellow.

Strain	Genbank	Genus	Family	Putative GNP pathway - no specific genes (# - AntiSMASH gene cluster)	Putative GNP pathway - with specific genes (# - AntiSMASH gene cluster)
Acidothermus cellulolyticus 11B	NC_008578.1	*Acidothermus*	*Acidothermaceae*	other - 1GT	none
Arcanobacterium haemolyticum CCM, DSM 20595	CP002045.1	*Arcanobacterium*	*Actinomycetaceae*	none	none
Mobiluncus curtisii ATCC 43063	CP001992.1	*Mobiluncus*	*Actinomycetaceae*	none	none
Actinosynnema mirum 101, DSM 43827	CP001630.1	*Actinosynnema*	*Actinosynnemataceae*	10 - nrps, 1GT 12 - nrps-t1pks - 1GT 14 - nrps, 1GT 15 - t1pks, 1GT	5 - nucleoside, 1GT, 1 spec gene 22 - oligosaccharide-t1pks, 2GT, 1 spec gene
Beutenbergia cavernae HKI 0122, DSM 12333	CP001618.1	*Beutenbergia*	*Beutenbergiaceae*	none	none
Catenulispora acidiphila ID139908, DSM 44928	CP001700.1	*Catenulispora*	*Catenulisporaceae*	6 - nrps, 1GT 7 - nrps-lant-t1pks, 1GT	8 - terpene, 2GT, 2 spec genes
Cellulomonas flavigena 134, DSM 20109	CP001964.1	*Cellulomonas*	*Cellulomonadaceae*	t2pks - 1GT	none
Cellulomonas fimi NRS 133, ATCC 484	CP002666.1	*Cellulomonas*	*Cellulomonadaceae*	none	none
Cellvibrio gilvus ATCC 13127	CP002665.1	*Cellulomonas*	*Cellulomonadaceae*	terpene - 1GT	none
Corynebacterium aurimucosum CN-1, ATCC 700975	CP001601.1	*Corynebacterium*	*Corynebacteriaceae*	t1pks - 1GT	none
Corynebacterium diphtheriae gravis NCTC 13129	BX248353.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium efficiens YS-314	NC_004369.1	*Corynebacterium*	*Corynebacteriaceae*	terpene - 1GT	none
Corynebacterium glutamicum Kalinowski ATCC 13032	NC_006958.1	*Corynebacterium*	*Corynebacteriaceae*	terpene - 1GT t1pks - 1GT	none
Corynebacterium glutamicum Nakagawa ATCC 13032	NC_003450.3	*Corynebacterium*	*Corynebacteriaceae*	terpene - 1GT t1pks - 1GT	none
Corynebacterium glutamicum R	NC_009342.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium jeikeium K411	NC_007164.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium kroppenstedtii DSM 44385	CP001620.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium pseudotuberculosis 1002	CP001809.2	*Corynebacterium*	*Corynebacteriaceae*	4 -t1pks, 1GT	none
Corynebacterium pseudotuberculosis C231	CP001829.1	*Corynebacterium*	*Corynebacteriaceae*	4 -t1pks, 1GT	none
Corynebacterium pseudotuberculosis FRC41	CP002097.1	*Corynebacterium*	*Corynebacteriaceae*	4 -t1pks, 1GT	none
Corynebacterium pseudotuberculosis I19	CP002251.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium resistens DSM 45100	CP002857.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium ulcerans 809	CP002790.1	*Corynebacterium*	*Corynebacteriaceae*	5 -t1pks, 1GT	none
Corynebacterium ulcerans BR-AD22	CP002791.1	*Corynebacterium*	*Corynebacteriaceae*	5 -t1pks, 1GT	none
Corynebacterium urealyticum DSM 7109	NC_010545.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium diphtheriae 31A	CP003206.1	*Corynebacterium*	*Corynebacteriaceae*	4 - nrps, 1GT 6 - t1pks, 1GT	none
Corynebacterium diphtheriae BH8	CP003209.1	*Corynebacterium*	*Corynebacteriaceae*	5 -t1pks, 1GT	none
Corynebacterium diphtheriae C7 (beta)	NC_016801.1	*Corynebacterium*	*Corynebacteriaceae*	5 - nrps, 1GT 6 - t1pks, 1GT	none
Corynebacterium diphtheriae CDCE 8392	CP003211.1	*Corynebacterium*	*Corynebacteriaceae*	5 - nrps, 1GT 7 - t1pks, 1GT	none
Corynebacterium diphtheriae HC01	CP003212.1	*Corynebacterium*	*Corynebacteriaceae*	4 -t1pks, 1GT	none
Corynebacterium diphtheriae HC02	CP003213.1	*Corynebacterium*	*Corynebacteriaceae*	6 - t1pks, 1GT	none
Corynebacterium diphtheriae HC03	CP003214.1	*Corynebacterium*	*Corynebacteriaceae*	6 - t1pks, 1GT	none
Corynebacterium diphtheriae HC04	CP003215.1	*Corynebacterium*	*Corynebacteriaceae*	7 - t1pks, 1GT	none
Corynebacterium diphtheriae INCA 402	CP003208.1	*Corynebacterium*	*Corynebacteriaceae*	2 - nrps, 1GT 6 - t1pks, 1GT	none
Corynebacterium diphtheriae PW8	CP003216.1	*Corynebacterium*	*Corynebacteriaceae*	5 - t1pks, 1GT	none
Corynebacterium diphtheriae VA01	CP003217.1	*Corynebacterium*	*Corynebacteriaceae*	6 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis 1/06-A	CP003082.1	*Corynebacterium*	*Corynebacteriaceae*	3 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis 267	CP003407.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Corynebacterium pseudotuberculosis 3/99-5	CP003152.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis 316	CP003077.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis 42/02-A	CP003062.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis CIP52.97	CP003061.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis P54B96	CP003385.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium pseudotuberculosis PAT10	CP002924.1	*Corynebacterium*	*Corynebacteriaceae*	4 - t1pks, 1GT	none
Corynebacterium ulcerans 0102	AP012284.1	*Corynebacterium*	*Corynebacteriaceae*	5 - t1pks, 1GT	none
Corynebacterium variabile DSM 44702	CP002917.1	*Corynebacterium*	*Corynebacteriaceae*	none	none
Brachybacterium faecium 6-10, DSM 4810	CP001643.1	*Brachybacterium*	*Dermabacteraceae*	none	none
Kytococcus sedentarius 541, DSM 20547	CP001686.1	*Kytococcus*	*Dermacoccaceae*	2 - t1pks, 1GT	none
Frankia alni ACN14a	CT573213.2	*Frankia*	*Frankiaceae*	GC1 - t1pks - 1GT GC4 - terpene - 1GT GC5 - t1pks - 1GT GC6 - t1pks - 1GT GC7 - nrps - 1GT GC11 - t1pks -1GT GC13 - t1pks - 1GT GC18 - other - 1GT	none
Frankia sp. CcI3	CP000249.1	*Frankia*	*Frankiaceae*	GC2 - terpene - 1GT GC3 - t1pks - 1GT GC4 - t1pks - 1GT GC8 - other - 5GTs GC19 - lant - 1GT GC24 - terpene - 1GT	none
Frankia sp. EAN1pec	CP000820.1	*Frankia*	*Frankiaceae*	none	none
Frankia sp. EuI1c	CP002299.1	*Frankia*	*Frankiaceae*	2 - t1pks, 1GT 4 - t1pks, 1GT	none
Frankia symbiont of *Datisca glomerata*	NC_015656.1	*Frankia*	*Frankiaceae*	2 - t2pks, 2GTs 5 - t1pks, 1GT 11 - t2pks, 1GT	none
Blastococcus saxobsidens DD2	FO117623.1	*Blastococcus*	*Geodermatophilaceae*	none	none
Geodermatophilus obscurus G-20, DSM 43160	CP001867.1	*Geodermatophilus*	*Geodermatophilaceae*	GC2 - other, 6GTs	none
Stackebrandtia nassauensis LLR-40K-21, DSM 44728	CP001778.1	*Stackebrandtia*	*Glycomycetaceae*	GC2 - nrps-t1pks - 1GT GC15 - other - 1GT	none
Gordonia bronchialis 3410, DSM 43247	CP001802.1	*Gordonia*	*Gordoniaceae*	1 - t1pks, 1GT 6 -terpene, 1GT	none
Gordonia polyisoprenivorans VH2, DSM 44266	CP003119.1 CP003120.1	*Gordonia*	*Gordoniaceae*	6 - butyrolactone, 1GT 7 - terpene, 5GTs 9 - nrps, 1GT 11 - nrps, 3GTs 13 - terpene, 1GT	4 - nrps, 5GT, 1NT, spec genes
Intrasporangium calvum 7KIP, DSM 43043	CP002343.1	*Intrasporangium*	*Intrasporangiaceae*	none	none
Jonesia denitrificans 55134, DSM 20603	CP001706.1	*Jonesia*	*Jonesiaceae*	none	none
Kineococcus radiotolerans SRS30216	NC_009664.2	*Kineococcus*	*Kineosporiaceae*	3 - other, 1GT	none
Clavibacter michiganensis michiganensis NCPPB 382	NC_009480.1 NC_009478.1 NC_009479.1	*Clavibacter*	*Microbacteriaceae*	none	none
Leifsonia xyli xyli CTCB07	NC_006087.1	*Leifsonia*	*Microbacteriaceae*	none	none
Microbacterium testaceum StLB037	NC_015125.1	*Microbacterium*	*Microbacteriaceae*	1- t3pks, 1GT	none
Tropheryma whipplei TW08/27	BX072543.1	*Tropheryma*	*Microbacteriaceae*	none	none
Tropheryma whipplei Twist	NC_004572.3	*Tropheryma*	*Microbacteriaceae*	none	none
Arthrobacter arilaitensis re117, CIP108037	NC_014550.1	*Arthrobacter*	*Micrococcaceae*	1 - siderophore, 2GT	none
Arthrobacter aurescens TC1	NC_008711.1 NC_008712.1 NC_008713.1	*Arthrobacter*	*Micrococcaceae*	4 - t3pks, 1GT	none
Arthrobacter chlorophenolicus A6	NC_011886.1 NC_011881.1 NC_011879.1	*Arthrobacter*	*Micrococcaceae*	none	none
Arthrobacter phenanthrenivorans Sphe3	NC_015145.1 NC_015146.1 NC_015147.1	*Arthrobacter*	*Micrococcaceae*	2 - t3pks, 1GT	none
Arthrobacter sp. FB24	NC_008541.1 NC_008537.1 NC_008538.1 NC_008539.1	*Arthrobacter*	*Micrococcaceae*	none	none
Kocuria rhizophila DC2201	NC_010617.1	*Kocuria*	*Micrococcaceae*	none	none
Micrococcus luteus Fleming NCTC 2665	CP001628.1	*Micrococcus*	*Micrococcaceae*	terpene - 1GT	none
Renibacterium salmoninarum ATCC 33209	CP000910.1	*Renibacterium*	*Micrococcaceae*	4 - t1pks, 1GT	none
Rothia dentocariosa ATCC 17931	CP002280.1	*Rothia*	*Micrococcaceae*	none	none
Rothia mucilaginosa DY-18	NC_013715.1	*Rothia*	*Micrococcaceae*	none	none
Actinoplanes missouriensis NBRC 102363	AP012319.1	*Actinoplanes*	*Micromonosporaceae*	GC2 - nrps - 1GT GC7 - t1pks-nrps-terpene - 1GT GC8 - other - 1GT	GC6 - t2pks - 1GT, 1NT, 1 4,6DH, 5 spec genes
Actinoplanes sp. SE50/110	CP003170.1	*Actinoplanes*	*Micromonosporaceae*	GC12 - t1pks - 1GT	GC4 - nrps(LPS) - 2GT, spec genes GC5 - amglyccycl - 2GTs, 1NT, 1 4,6DH, 1 spec genes

Dataset S1. Prediction of gene clusters of glycosylated natural products in finished actinobacterial genomes (Oct 2012, JGI database) by antiSMASH analysis of GenBank genome files and subsequent analysis of glycosylation genes in predicted gene clusters. Predicted GNP pathways were differentiated by presence or absence of specific glycosylation genes. Predicted GNP pathways are highlighted in grey and corresponding strain genomes and families in yellow.

Strain	Genbank	Genus	Family	Putative GNP pathway - no specific genes (# - AntiSMASH gene cluster)	Putative GNP pathway - with specific genes (# - AntiSMASH gene cluster)
Micromonospora aurantiaca ATCC 27029	CP002162.1	*Micromonospora*	*Micromonosporaceae*	GC4 - NRPS-PKS - 2GT GC6 - hybrid - 1GT GC10 - oligosaccharide - 3GT	GC12 - oligosaccharide-type II PKS - 7GT, 1NT, 1 4,6DH, 5 spec genes
Micromonospora sp. L5	CP002399.1	*Micromonospora*	*Micromonosporaceae*	GC10 - nrps-oligosaccharide-terpene - 3 GT GC14 - lant-nrps-t1pks - 1GT GC17 - NRPS-PKS - 2GT	GC8 - oligosaccharide-type II PKS - 5GT, 1 4,6DH, 1NT, 6 spec genes
Salinispora arenicola CNS-205	CP000850.1	*Salinispora*	*Micromonosporaceae*	GC12 - amglyccycl	GC4 - type I pks-nrps - 2GT, 1NT, spec genes GC7 - oligosaccharide-t1pks - 4GT, 5 spec genes GC10 - indole - 1GT, 1NT, 1 4,6DH, 4 spec genes
Salinispora tropica CNB-440	NC_009380.1	*Salinispora*	*Micromonosporaceae*	none	GC4 - type II pks -2GT, 1NT, 1 4,6-DH, 4 spec genes
Verrucosispora maris AB-18-032	CP002638.1	*Verrucosispora*	*Micromonosporaceae*	none	none
Amycolicicoccus subflavus DQS3-9A1	NC_015564.1 NC_015560.1 NC_015561.1	*Amycolicicoccus*	*Mycobacteriaceae*	6 terpene, 1GT 8 - t1pks, 1GT	2 - nrps, 1GT, 1NT, 1spec gene
Mycobacterium africanum GM041182	NC_015758.1	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks-nrps, 2GTs	6 - t1pks, 5GTs, 1spec gene
Mycobacterium avium 104	CP000479.1	*Mycobacterium*	*Mycobacteriaceae*	11 - nrps-oligosaccharide, 4GTs	13 - terpene, 1GT, 1spec gene
Mycobacterium leprae Br4923	NC_011896.1	*Mycobacterium*	*Mycobacteriaceae*	2 - oligosaccharide-t1pks, 3GTs	5 - t1pks, 1GT
Mycobacterium avium paratuberculosis K-10	NC_002944.2	*Mycobacterium*	*Mycobacteriaceae*	none	none
Mycobacterium bovis AF2122/97	NC_002945.3	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks-nrps, 3GTs	6 - t1pks, 5GTs, 1spec gene
Mycobacterium bovis BCG Pasteur 1173P2	NC_008769.1	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks-nrps, 2GTs	6 - t1pks, 5GTs, 1spec gene
Mycobacterium bovis BCG Tokyo 172	NC_012207.1	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks-nrps, 2GTs	6 - t1pks, 5GTs, 1spec gene
Mycobacterium canettii CIPT 140010059	HE572590.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 5GTs 12 - t1pks-nrps-oligosaccharide, 3GTs 14 - t1pks, 1GT	none
Mycobacterium gilvum PYR-GCK	NC_009338.1 NC_009339.1 CP000658.1 CP000659.1	*Mycobacterium*	*Mycobacteriaceae*	2 - nrps, 1GT 7 - terpene, 1GT 11 - t13pks, 1GT	4 - t1pks, spec gene
Mycobacterium leprae TN	NC_002677.1	*Mycobacterium*	*Mycobacteriaceae*	2 - oligosaccharide-t1pks, 2GTs 5 - t1pks, 1GT	none
Mycobacterium marinum M, ATCC BAA-535	NC_010612.1 NC_010604.1	*Mycobacterium*	*Mycobacteriaceae*	10 - t1pks-nrps, 1GT 14 - t1pks-nrps, 4GTs	none
Mycobacterium smegmatis MC2 155	NC_008596.1	*Mycobacterium*	*Mycobacteriaceae*	2 - nrps-t1pks, 3GTs, 1 4,6DH 10 - t1pks-nrps, 2GTs 11 - nrps, 1GTs 14 - nrps, 3GTs	none
Mycobacterium gilvum Spyr1	NC_014814.1 NC_014811.1 NC_014812.1	*Mycobacterium*	*Mycobacteriaceae*	1 - nrps, 1GT 4 - terpene, 1GT 7 - t3pks, 1GT 13 - t1pks, 1GT	none
Mycobacterium sp. JDM601	CP002329.1	*Mycobacterium*	*Mycobacteriaceae*	none	7 - other, 1GT, 1spec gene
Mycobacterium sp. JLS	CP000580.1	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks, 1GT	none
Mycobacterium sp. KMS	CP000518.1 CP000519.1 CP000520.1	*Mycobacterium*	*Mycobacteriaceae*	7 - nrps-t1pks, 3GTs	none
Mycobacterium sp. MCS	CP000384.1 CP000385.1	*Mycobacterium*	*Mycobacteriaceae*	3 - nrps-t1pks, 3GTs	none
Mycobacterium tuberculosis CCDC5079	CP001641.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 5GTs 12 - t1pks-nrps-oligosaccharide, 2GTs	none
Mycobacterium tuberculosis CCDC5180	CP001642.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 5GTs	none
Mycobacterium tuberculosis CDC1551	NC_002755.2	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks-nrps-oligosaccharide, 2GTs	
Mycobacterium tuberculosis F11 (ExPEC)	NC_009525.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 5GTs 12 - t1pks-nrps-oligosaccharide, 3GTs	none
Mycobacterium tuberculosis H37Ra	CP000611.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 5GT 12 - t1pks-nrps-oligosaccharide, 2GTs	none
Mycobacterium tuberculosis KZN 1435 (MDR)	CP001658.1	*Mycobacterium*	*Mycobacteriaceae*	4 - oligosaccharide-t1pks-nrps, 3GTs 11 - t1pks, 5GTs	none
Mycobacterium ulcerans Agy99	NC_008611.1 NC_005916.1	*Mycobacterium*	*Mycobacteriaceae*	none	3 - nrps, 3GTs, 2spec genes
Mycobacterium vanbaalenii PYR-1	NC_008726.1	*Mycobacterium*	*Mycobacteriaceae*	7 - other, 4GTs 13 - t1pks, 1GT	none
Mycobacterium bovis BCG Mexico	CP002095.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 4GTs 12 - t1pks-nrps, 3GTs	none
Mycobacterium chubuense NBB4	CP003053.1 CP003054.1 CP003055.1	*Mycobacterium*	*Mycobacteriaceae*	12 - t1pks, 2GT	11 - other, 1GT, 1 spec gene
Mycobacterium intracellulare MOTT-02	CP003323.1	*Mycobacterium*	*Mycobacteriaceae*	3 - nrps, 1GT	15 - terpene, 1GT, 2 spec genes
Mycobacterium intracellulare MOTT-64	CP003324.1	*Mycobacterium*	*Mycobacteriaceae*	none	15 - terpene, 1GT, 2 spec genes
Mycobacterium massiliense GO 06	CP003699.1	*Mycobacterium*	*Mycobacteriaceae*	1 - t1pks, 5GT 13 - nrps, 2GTs	12 - nrps, 3GT, 1NT, 1 spec gene
Mycobacterium rhodesiae NBB3	CP003169.1	*Mycobacterium*	*Mycobacteriaceae*	4 - lant, 1GT 6 - t1pks, 1GT 13 - t2pks, 1GT	2 - t1pks-nrps, 5GTs, 1 4,6DH, spec genes
Mycobacterium tuberculosis CTRI-2	CP002992.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 3GTs 13 - t1pks-nrps, 3GTs	none
Mycobacterium tuberculosis KZN 4207 (DS)	CP001662.1	*Mycobacterium*	*Mycobacteriaceae*	4 - oligosaccharide-t1pks-nrps, 4GTs 11 - t1pks, 5GTs	none
Mycobacterium tuberculosis RGTB327	CP003233.1	*Mycobacterium*	*Mycobacteriaceae*	none	none
Mycobacterium tuberculosis RGTB423	CP003234.1	*Mycobacterium*	*Mycobacteriaceae*	8 - t1pks-nrps, 1GT	none
Mycobacterium tuberculosis UT205	NC_016934.1	*Mycobacterium*	*Mycobacteriaceae*	6 - t1pks, 4GTs	none
Nakamurella multipartita Y-104, DSM 44233	CP001737.1	*Nakamurella*	*Nakamurellaceae*	terpene - 1GT	none
Nocardia farcinica IFM 10152	AP006618.1 AP006619.1 AP006620.1	*Nocardia*	*Nocardiaceae*	GC1 - t1pks - 5GTs GC17 - nrps - 1GT	none
Nocardia cyriacigeorgica GUH-2	NC_016887.1	*Nocardia*	*Nocardiaceae*	1 - t1pks, 3GTs 6 - nrps, 3GTs 15 - nrps, 1GT, 1 4,6DH 19 - nrps, 1GT	none
Rhodococcus equi 103S	FN563149.1	*Rhodococcus*	*Nocardiaceae*	GC2 - terpene - 1GT GC7 - terpene - 1GT GC10 - other - 1GT GC11 - t1pks - 4GTs	GC14 - nrps - 2GT, 1NT, 1 4,6DH, 1 spec gene
Rhodococcus erythropolis PR4	AP008957.1 AP008931.1 AP008932.1 AP008933.1	*Rhodococcus*	*Nocardiaceae*	GC1 - t1pks - 6GTs GC2 - t1pks - 1GT GC5 - nrps - 2GTs GC8 - nrps - 1GT GC10 - nrps - 1GT GC12 - terpene - 1GT GC15 - amglyccycl - 1GT	none
Rhodococcus opacus B4	AP011115.1 AP011116.1 AP011117.1	*Rhodococcus*	*Nocardiaceae*	GC1 - nrps - 1GT GC3 - terpene - 1GT GC12 - t1pks - 12GTs GC19 - nrps - 4GTs GC20 - nrps - 3GTs	GC13 - t1pks - 1GT, 1 spec gene
Rhodococcus jostii RHA1	CP000431.1 CP000432.1 CP000433.1 CP000434.1	*Rhodococcus*	*Nocardiaceae*	GC11 - t1pks - 5GTs	GC17 - nrps - 2GTs, 2NTs, 1 4,6DH, spec genes GC18 - nrps - 1GT, 2NTs, 1 4,6DH, spec genes GC20 - nrps - 1GT, 1NT, spec gene
Kribbella flavida IFO 14399, DSM 17836	CP001736.1	*Kribbella*	*Nocardioidaceae*	none	none
Nocardioides sp. JS614	NC_008699.1 NC_008697.1	*Nocardioides*	*Nocardioidaceae*	none	none
Nocardiopsis dassonvillei dassonvillei DSM 43111	NC_014210.1 NC_014211.1	*Nocardiopsis*	*Nocardiopsaceae*	none	13 - lant-oligosaccharide, 8GTs, 1NT, 1 4,6DH, spec genes
Thermobifida fusca YX	CP000088.1	*Thermobifida*	*Nocardiopsaceae*	3 - bcin, 1GT	none
Isoptericola variabilis 225	CP002810.1	*Isoptericola*	*Promicromonosporaceae*		
Xylanimonas cellulosilytica XIL07, DSM 15894	CP001821.1 CP001822.1	*Xylanimonas*	*Promicromonosporaceae*	none	none
Microlunatus phosphovorus NM-1	AP012204.1	*Microlunatus*	*Propionibacteriaceae*	none	none
Propionibacterium acnes 266	CP002409.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes 6609	CP002815.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes KPA171202	NC_006085.1	*Propionibacterium*	*Propionibacteriaceae*	none	none

Dataset S1. Prediction of gene clusters of glycosylated natural products in finished actinobacterial genomes (Oct 2012, JGI database) by antiSMASH analysis of GenBank genome files and subsequent analysis of glycosylation genes in predicted gene clusters. Predicted GNP pathways were differentiated by presence or absence of specific glycosylation genes. Predicted GNP pathways are highlighted in grey and corresponding strain genomes and families in yellow.

Strain	Genbank	Genus	Family	Putative GNP pathway - no specific genes (# - AntiSMASH gene cluster)	Putative GNP pathway - with specific genes (# - AntiSMASH gene cluster)
Propionibacterium acnes SK137	CP001977.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium freudenreichii shermanii CIRM-BIA1	NC_014215.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes ATCC 11828	CP003084.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes TypeIA2 P.acn17	CP003196.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes TypeIA2 P.acn31	CP003197.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Propionibacterium acnes TypeIA2 P.acn33	CP003195.1	*Propionibacterium*	*Propionibacteriaceae*	none	none
Amycolatopsis mediterranei U32	CP002000.1	*Amycolatopsis*	*Pseudonocardiaceae*	GC13 - nrps-t1pks-terpene - 1GT GC14 - t1pks - 1GT GC15 - t1pks - 1GT	GC1 - type I PKS - 1GT, 2 spec genes GC22 - nrps-oligosaccharide-t2pks - 9GT, 1NT, spec genes
Amycolatopsis mediterranei S699	CP003729.1	*Amycolatopsis*	*Pseudonocardiaceae*	14 - t1pks, 1GT	22 - nrps-oligosaccharide-t1pks - 9GTs, 1NT, 3 spec genes
Pseudonocardia dioxanivorans CB1190	CP002593.1 CP002594.1 CP002595.1 CP002596.1 CP002597.1 CP002598.1	*Pseudonocardia*	*Pseudonocardiaceae*	none none none none none none	none none none none none none
Saccharomonospora viridis P101, DSM 43017	CP001683.1	*Saccharomonospora*	*Pseudonocardiaceae*	none	none
Saccharopolyspora erythraea NRRL 2338 white	AM420293.1	*Saccharopolyspora*	*Pseudonocardiaceae*	GC7 - t1pks - 1GT GC14 - terpene - 1GT GC25 - t1pks - 1GT GC27 - terpene - 1GT	GC3 - t1pks - 2GT, 11 spec genes
Sanguibacter keddieii ST-74, DSM 10542	CP001819.1	*Sanguibacter*	*Sanguibacteraceae*	none	none
Segniliparus rotundus CDC 1076, DSM 44985	CP001958.1	*Segniliparus*	*Segniliparaceae*	8 -t1pks, 1GT	none
Kitasatospora setae KM-6054, NBRC 14216	NC_016109.1	*Kitasatospora*	*Streptomycetaceae*	8 - terpene, 1GT, 1NT 10 - t3pks, 1GT, 1NT, 1 4,6DH	7 - siderophore, 1GT, 1 spec gene
Streptomyces avermitilis MA-4680	NC_003155.4 NC_004719.1	*Streptomyces*	*Streptomycetaceae*	GC5 - NRPS-PKS - 1GT GC27 - melanin - 2GT, polysaccharide? GC31 - type I PKS - 1GT GC33 - type I PKS - 1GT	GC6 -type I PKS - 1GT, 1NT, 1 4,6DH, 5 spec genes GC7 - terpene - 1GT, 1NT, 2 spec genes GC11 - hopane - 1GT, 1NT, 2 spec genes
Streptomyces bingchenggensis BCW-1	NC_016582.1	*Streptomyces*	*Streptomycetaceae*	GC7- type I PKS - 1GT, 1 4,6DH GC19 - terpene - 1GT GC26 - NRPS-PKS - 1GT GC27 - NRPS - 1GT	GC34 - type II PKS - 5GT, 1NT, 1 4,6DH, 5 spec genes GC39 - type I PKS - 3GT, 1NT, 4,6DH - 3 spec genes
Streptomyces coelicolor A3(2)	NC_003888.3 NC_003904.1	*Streptomyces*	*Streptomycetaceae*	GC8 - melanin - 1GT GC21 - hopane - 1GT, 1NT	none
Streptomyces griseus griseus NBRC 13350	NC_010572.1	*Streptomyces*	*Streptomycetaceae*	GC6 - t1pks - 1GT GC33 - nrps - 1GT	GC7 - oligosaccharide - 1GT, 1NT, 1 4,6DH, 4 spec genes GC12 - terpene - 2GT, 1NT, 1 spec gene GC27 - amglyccycl - 2GTs, 1 NT, 1 4,6DH, spec genes
Streptomyces scabiei 87.22	NC_013929.1	*Streptomyces*	*Streptomycetaceae*	none	GC7- terpene - 1GT, 1NT, 1 spec genes GC26 - type I PKS - 1GT, 1 4,6DH, 5 spec genes
Streptomyces cattleya NRRL 8057	CP003219.1 CP003229.1	*Streptomyces*	*Streptomycetaceae*	GC13 - terpene - 1GT GC19 - hopene - 1GT, 1NT GC20 - t3pks - 1GT none	GC24 - butyrolactone-t1pks-nrps - 1GT, 1NT, 2 spec genes none
Streptomyces sp. Tu6071	CM001165.1	*Streptomyces*	*Streptomycetaceae*	GC5 - hopene - 1GT, 1NT	GC8 - terpene - 1GT, 1NT, 1 4,6DH, spec genes
Streptomyces flavogriseus ATCC 33331	CP002475.1 CP002476.1 CP002477.1	*Streptomyces*		GC16 - terpene - 1GT	GC6- terpene - 1GT, 1NT, 1 spec genes
Streptomyces venezuelae ATCC 10712	FR845719.1	*Streptomyces*	*Streptomycetaceae*	none	GC18 - t2pks - 1GT, 1NT, 1 4,6DH, 3 spec genes
Streptomyces hygroscopicus jinggangensis 5008	NC_017765.1 NC_017766.1 NC_016972.1	*Streptomyces*	*Streptomycetaceae*	none	2 - aminoglycoside, 1GT, 1NT, spec genes 32 - terpene - 1GT, 1NT, 1spec gene
Streptosporangium roseum NI 9100, DSM 43021	NC_013595.1 NC_013596.1	*Streptosporangium*	*Streptosporangiaceae*	none none	none none
Thermomonospora curvata DSM 43183	NC_013510.1	*Thermomonospora*	*Thermomonosporaceae*	3 - nrps, 1GT 9 - t1pks, 1GT	4 - t2pks-t1pks, 2GT, 1NT, 1 4,6DH, 6 spec genes
Tsukamurella paurometabola 33, DSM 20162	NC_014158.1 NC_014159.1	*Tsukamurella*	*Tsukamurellaceae*	5 - t3pks, 1GT	8 - t1pks, 1NT, 1GT, 2 spec genes
Thermobispora bispora R51, DSM 43833	CP001874.1	*Thermobispora*	*unclassified*	none	none
Bifidobacterium adolescentis ATCC 15703	NC_008618.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis AD011	CP001213.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis BB-12	CP001853.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis Bl-04, ATCC SD5219	CP001515.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis DSM 10140	NC_017834.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis V9	CP001892.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis subsp. lactis CNCM I-2494	CP002915.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium bifidum PRL2010	CP001840.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium bifidum S17	CP002220.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium breve ACS-071-V-Sch8b	CP002743.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium breve UCC2003	CP000303.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium dentium Bd1	CP001750.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum DJO10A	NC_010816.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum NCC2705	NC_004307.2	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum infantis 157F-NC	NC_015052.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum infantis ATCC 15697	NC_017219.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum longum BBMN68	CP002286.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum longum JDM301	CP002010.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium longum subsp. longum KACC 91563	NC_017221.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis animalis ATCC 25527	CP002567.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Bifidobacterium animalis lactis BLC1	CP003039.1	*Bifidobacterium*	*Bifidobacteriaceae*	none	none
Gardnerella vaginalis ATCC 14019	NC_014644.1	*Gardnerella*	*Bifidobacteriaceae*	none	none
Gardnerella vaginalis 409-05	NC_013721.1	*Gardnerella*	*Bifidobacteriaceae*	none	none
Gardnerella vaginalis HMP9231	NC_017456.1	*Gardnerella*	*Bifidobacteriaceae*	none	none
Atopobium parvulum IPP 1246, DSM 20469	CP001721.1	*Atopobium*	*Coriobacteriaceae*	none	none
Coriobacterium glomerans PW2, DSM 20642	NC_015389.1	*Coriobacterium*	*Coriobacteriaceae*	none	none
Cryptobacterium curtum 12-3, DSM 15641	NC_013170.1	*Cryptobacterium*	*Coriobacteriaceae*	none	none
Eggerthella lenta VPI 0255, DSM 2243	CP001726.1	*Eggerthella*	*Coriobacteriaceae*	none	none
Eggerthella sp. YY7918	NC_015738.1	*Eggerthella*	*Coriobacteriaceae*	none	none
Gordonibacter pamelaeae 7-10-1-bT, DSM 19378	FP929047.1	*Gordonibacter*	*Coriobacteriaceae*	none	none
Olsenella uli VPI, DSM 7084	CP002106.1	*Olsenella*	*Coriobacteriaceae*	none	none
Slackia heliotrinireducens RHS 1, DSM 20476	CP001684.1	*Slackia*	*Coriobacteriaceae*	none	none
Rubrobacter xylanophilus DSM 9941	NC_008148.1	*Rubrobacter*	*Rubrobacteraceae*	none	none
Conexibacter woesei ID131577, DSM 14684	NC_013739.1	*Conexibacter*	*Conexibacteraceae*	none	none

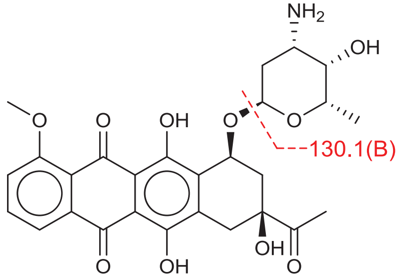
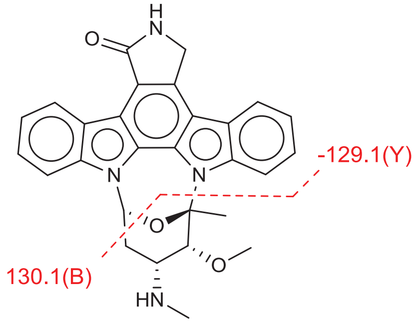
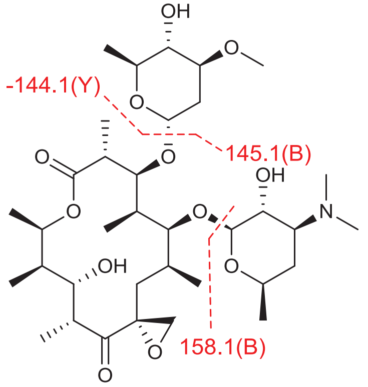
Dataset S3. Connection of known GNP chemo- and genotypes by the MS-glycogenetic code. Reference GNP chemotypes were analyzed in sugar-specific MS^n neutral losses or B-/C-ion fragments. MS/MS candidate sugars were identified based on observed sugar masses (see **Dataset S4**). Corresponding reference GNP genotypes from GenBank were then analyzed in predicted common and specific glycosylation genes (see **Dataset S4**). Predicted glycosylation genes were analyzed in matching the biosynthesis of the MS^n candidate sugars utilizing **Dataset S2**. A GNP chemotype-to-genotype connection by the MS-glycogenetic code was a match of MS^n and genetic candidate sugars (green). GNPs with an unsuccessful chemotype-genotype connection were highlighted in red. Abbreviations: see **Dataset S2**.

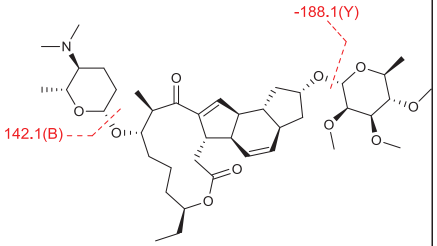
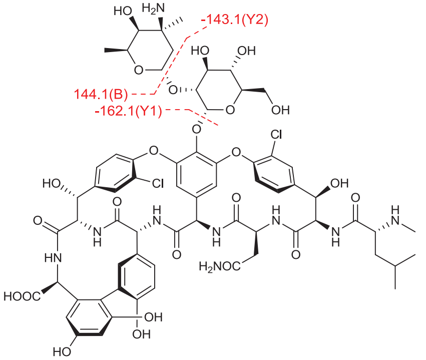
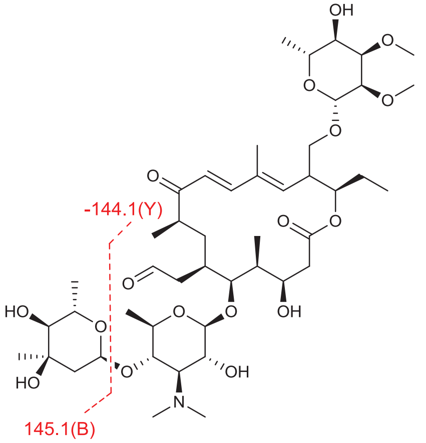
#	Reference GNP chemotype	Instrument	Reference	Observed MS/MS sugar footprint [Da] (see Dataset S4)		MS/MS candidate sugars	Genetic candidate sugars	Matching reference pathway based on Dataset S2	BLAST analysis of gene clusters (see Dataset S4)		Gene cluster (GenBank #)
				Y-ion neutral loss	B-ion				Specific glycosylation genes	Common glycosylation genes	
1	phenalinolactone	ESI-LTQ-FT-MS	this study	128.072 (Y)	129.094 (B)	O-methyl-L-amicetose	O-methyl-L-amicetose	2,3DH, 3,4DH, 3KR, 4KR, E, O-MT	2,3DH, 3,4DH, 3KR, 4KR, E, O-MT	NT, 4,6DH, GT	DQ230532
						4-O-methyl-L-rhodinose	4-O-methyl-L-rhodinose				
2	daunomycin	ESI-Q-TOF-MS	Metlin, ID 550		130.0852 (B)	L-daunosamine	L-daunosamine	2,3DH, AmT, E, 4KR	2,3DH, AmT, E, 4KR	NT, 4,6DH, GT	STMDNRLM STMDNRQ SPU77891 STMDNRI
						L-acosamine					
						L-ristosamine	L-ristosamine	2,3DH, AmT, E, 4KR			
						L-ristosamine					
3	staurosporine	ESI-Q-TOF-MS	Metlin, ID 3307	129.080 (Y)	130.086 (B)	L-daunosamine	L-daunosamine	2,3DH, AmT, E, 4KR	2,3DH, AmT, E, 4KR, O-MT, MT	NT, 4,6DH, GT (2x)	AB088119
						L-acosamine					
						L-ristosamine	L-ristosamine	2,3DH, AmT, E, 4KR			
						L-ristosamine					
4	oleandomycin	ESI-Q-TOF-MS	Metlin, ID 44103	144.084 (Y)	145.086 (B)	D-chalcose			2,3DH, 3,4DH, 3KR, E, 4KR, E, 3,4DH/AmT, O-MT, N,N-MT, oxDA	NT, 4,6DH, GT (4x)	AF055579 AJ002638
						D-mycarose					
						L-oleandrose	L-oleandrose	2,3DH, 3KR, E, 4KR, O-MT			
						olivomose	olivomose	2,3DH, 3KR, 4KR, O-MT			
					158.116 (B)	4-N-ethyl-4-amino-3-O-methoxy-2,4,5-trideoxypentose					
						D-3-N-methyl-4-O-methyl-L-ristosamine					
						D-desosamine	D-desosamine	3,4DH, oxDA, AmT, N,N-MT			
						N,N-dimethyl-L-pyrrolosamine		2,3DH, 3KR, E, AmT, N,N-MT			
						L-megosamine	L-megosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-nogalamine	nogalamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-rhodosamine	L-rhodosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						D-angolosamine	D-angolosamine	2,3DH, AmT, 4KR, N,N-MT			
						kedarosamine		2,3DH, 3KR, AmT, N,N-MT			
5	spinosyn A	ESI-TOF-MS	Ferrer, I. *et al. Trends Anal. Chem.* **24**, 671-682 (2005)	188.105 (Y)		2,3,4-tri-O-methylrhamnose			2,3DH, 3,4DH, AmT, 3KR, O-MT (4x), N,N-MT	GT	AY007564
					142.123 (B)	D-forosamine	D-forosamine	2,3DH, 3KR, 3,4DH, AmT, N,N-MT			
6	vancomycin	ESI-Q-TOF-MS	Metlin, ID 582	143.082 (Y1)	144.100 (B)	L-actinosamine			2,3DH, 4KR, E, C-MT	GT (2x)	HE589771
						3-epi-L-vancosamine	3-epi-L-vancosamine	2,3DH, 3KR, E, 4KR, C-MT			
						L-vancosamine	L-vancosamine				
						L-vicenisamine					
						calicheamicin deoxyaminopentose					
						L-actinosamine					
				162.061 (Y2)		hexose					
7	tylosin	ESI-Q-TOF-MS	Metlin, ID 40964	144.076 (Y)	145.085 (B)	D-chalcose			2,3DH, 3KR, 4KR (2x), E (2x), AmT, C-MT, O-MT	NT, 4,6DH, GT (2x)	AF055922 AF147704 SFU08223
						D-mycarose	D-mycarose	2,3DH, 3KR, C-MT, E, 4KR			
						L-oleandrose	L-oleandrose	2,3DH, 3KR, O-MT, E, 4KR			
						olivomose	olivomose	2,3DH, 3KR, O-MT, 4KR			
8	avermectin B1a	ESI-Q-TOF-MS	Metlin, ID 40988	144.075 (Y1)		D-chalcose			2,3DH, 3KR, 4KR, E, O-MT	NT, 4,6DH, GT	AB032523
				144.077 (Y2)		D-mycarose					
						L-oleandrose	L-oleandrose	2,3DH, 3KR, O-MT, E, 4KR			
						olivomose	olivomose	2,3DH, 3KR, O-MT, 4KR			
9	nystatin	ESI-Q-TOF-MS	Metlin, ID 40970	no sugar fragmentation		no sugar observed					
10	chartreusin	ESI-MS	Xu, Z. *et al. Chem. Biol.* **12**, 579-588 (2005).	160(Y)		4-deoxy-4-methylthio-D-digitoxose	4-deoxy-4-methylthio-D-digitoxose		4KR, E/KR, MT	NT, 4,6DH, GT(2x)	AJ786382 AJ786383
						D-digitalose	D-digitalose	4KR, O-MT			
						3-O-methyl-rhamnose	3-O-methyl-rhamnose	4KR, E, O-MT			
						2-O-methyl-L-rhamnose	2-O-methyl-L-rhamnose	4KR, E, O-MT			
						6-deoxy-3-C-methyl-L-mannose	6-deoxy-3-C-methyl-L-mannose				
11	aclacinomycin A	ESI-Q-TOF-MS	Metlin, ID 551	112.0495(Y2)	113.0598(B) 131.0697 (C)	L-cinerulose A	L-cinerulose A	2,3DH, 3,4DH, 3KR, 4KR, E	DH, 3,4DH, 3KR, 4KR, E, AmT, N,N-MT	NT, 4,6DH, GT (3x)	AF264025 AF257324
				130.0675 (Y1)		D-digitoxose	D-digitoxose	2,3DH, 3KR, 4KR			
						L-digitoxose	L-digitoxose	2,3DH, 3KR, 4KR, E			
						2-deoxy-L-fucose	2-deoxy-L-fucose	2,3DH, 3KR, 4KR, E			
						D-olivose	D-olivose	2,3DH, 3KR, 4KR, E			
						D-oliose	D-oliose	2,3DH, 3KR, 4KR			
						esparamicin A1 sugar 1					
12	novobiocin	ESI-Q-TOF-MS	Metlin, ID 43281	217.0938(Y)	218.1035(B)	3-O-carbamoyl-4-O-methyl-L-noviose	3-O-carbamoyl-4-O-methyl-L-noviose	E, 4KR, C-MT, O-MT, CarbT	E, 4KR, C-MT, O-MT, CarbT, MT	NT, 4,6DH, GT	AF170880
13	neocarzinostatin	FAB-MS	Edo, K., Mizugaki, M. *Tetrahedron Lett.* **26**, 331-334 (1985)		160 (B)	2'-N-methyl-D-fucosamine	2'-N-methyl-D-fucosamine	2,3DH, 4KR, AmT, N-MT	2,3DH, 4KR, AmT, N-MT, O-MT	NT, 4,6DH, GT	AY117439

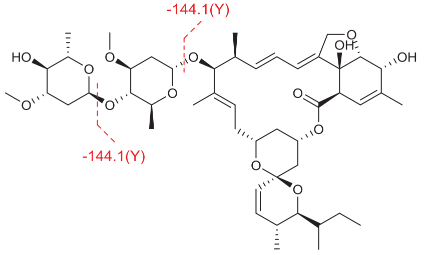
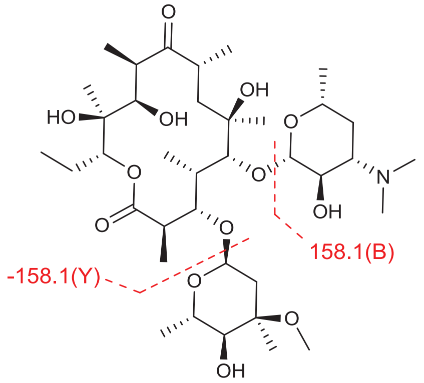
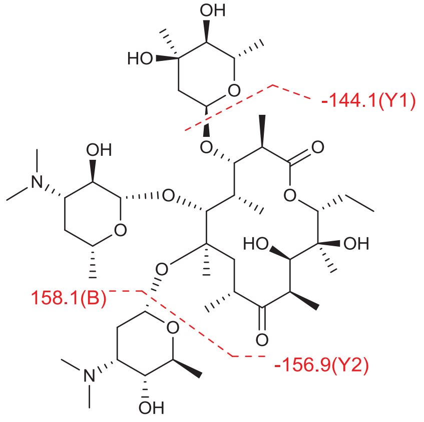
Dataset S3 (continued). Connection of known GNP chemo- and genotypes by the MS-glycogenetic code. Reference GNP chemotypes were analyzed in sugar-specific MS^n neutral losses or B-/C-ion fragments. MS/MS candidate sugars were identified based on observed sugar masses (see **Dataset S4**). Corresponding reference GNP genotypes from GenBank were then analyzed in predicted common and specific glycosylation genes (see **Dtaset S4**). Predicted glycosylation genes were analyzed in matching the biosynthesis of the MS^n candidate sugars utilizing **Dataset S2**. A GNP chemotype-to-genotype connection by the MS-glycogenetic code was a match of MS^n and genetic candidate sugars (green). GNPs with an unsuccessful chemotype-genotype connection were highlighted in red. Abbreviations: see **Dataset S2**.

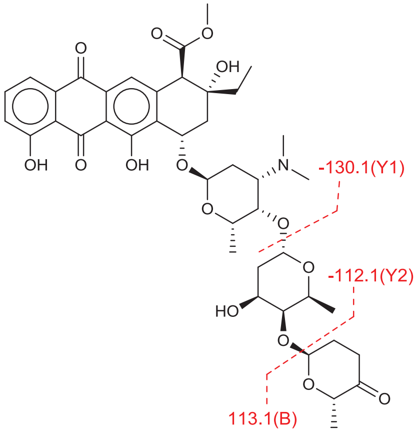
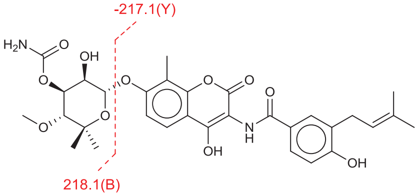
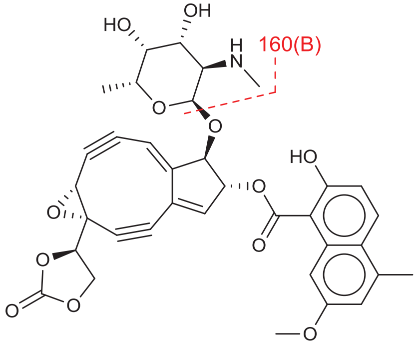
#	Reference GNP chemotype	Instrument	Reference	Observed MS/MS sugar footprint [Da] (see Dataset S4)		MS/MS candidate sugars	Genetic candidate sugars	Matching reference pathway based on Dataset S2	BLAST analysis of gene clusters (see Dataset S4)		Gene cluster (GenBank #)
				Y-ion neutral loss	B-ion				Specific glycosylation genes	Common glycosylation genes	
14	erythromycin A	ESI-Q-TOF-MS	Metlin, ID 2573	158.0928 (Y)		L-cladinose	L-cladinose	2,3DH, 3KR, E, C-MT, 4KR, MT	2,3DH, 3,4DH, 3KR, 4KR, E, 3,4IM, 3,4DH/AmT, C-MT,	GT (2x)	AM420293 SEU77459
					158.1168 (B)	4-N-ethyl-4-amino-3-O-methoxy-2,4,5-trideoxypentose					
						D-3-N-methyl-4-O-methyl-L-ristosamine					
						D-desosamine	D-desosamine	3,4DH, oxDA, AmT, N,N-MT			
						N,N-dimethyl-L-pyrrolosamine		2,3DH, AmT, E, 3KR, N,N-MT			
						L-megosamine	L-megosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-nogalamine	nogalamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-rhodosamine	L-rhodosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						D-angolosamine	D-angolosamine	2,3DH, AmT, 4KR, N,N-MT			
						kedarosamine		N/A			
15	megalomicin	N/A	Useglio, M. *et al. Appl. Env. Microbiol.* **76**, 3869-3877 (2010)	144.08 (Y1)		D-chalcose			2,3DH (2x), 3KR, 4KR (4x), 3,4IM, E, AmT, N,N-MT, AcT	GT (4x)	AF263245
						D-mycarose					
						L-oleandrose	L-oleandrose	2,3DH, 3KR, O-MT, E, 4KR			
						olivomose	olivomose	2,3DH, 3KR, O-MT, 4KR			
				156.91 (Y2)	158.12 (B), 174.11 (C)	4-N-ethyl-4-amino-3-O-methoxy-2,4,5-trideoxypentose					
						D-3-N-methyl-4-O-methyl-L-ristosamine					
						D-desosamine					
						N,N-dimethyl-L-pyrrolosamine	N,N-dimethyl-L-pyrrolosamine	2,3DH, AmT, E, 3KR, N,N-MT			
						L-megosamine	L-megosamine	2,3DH, 4KR, E, AmT, N,N-MT			
						L-nogalamine	nogalamine	2,3DH, 4KR, E, AmT, N,N-MT			
						L-rhodosamine	L-rhodosamine	2,3DH, 4KR, E, AmT, N,N-MT			
						D-angolosamine	D-angolosamine	2,3DH, 4KR, AmT, N,N-MT			
						kedarosamine					
16	amphotericin B	ESI-Q-TOF-MS	this study (Dataset S4)	163.084(Z)		D-mycosamine	D-mycosamine	3,4IM (CytP450), AmT	3,4IM (CytP450), AmT	4,6DH, GT	AF357202
17	lankamycin	CI-MS	Edo, K. *et al. Tetrahedron Lett.* **59**, 1886-1894 (1976)	200 (Y1)	201 (B)	4-O-acetyl-L-arcanose	4-O-acetyl-L-arcanose	2,3DH, 3KR, 4KR, E, C-MT, O-MT, AcT	2,3DH, 3KR (2x), 4KR, 3,4IM, E, 3,4DH/AmT, C-MT, oxDA, AcT, O-MT (2x),	NT, 4,6DH, GT (3x)	AB088224
				144 (Y2)	145 (B)	D-chalcose					
						D-mycarose					
						L-oleandrose	L-oleandrose	2,3DH, 3KR, O-MT, E, 4KR			
						olivomose	olivomose	2,3DH, 3KR, O-MT, 4KR			
18	chalcomycin	ESI-Q-TOF-MS	this study (Dataset S4)	174.062 (Y1)		2,3-O-dimethyl-L-rhamnose	2,3-O-dimethyl-L-rhamnose	4KR, E, O-MT	3KR, 4KR, E, 3,4DH/AmT, O-MT (3x), oxDA	NT, 4,6DH, GT (2x)	AY509120
						2,4-O-dimethyl-L-rhamnose	2,4-O-dimethyl-L-rhamnose				
						3,4-O-dimethyl-L-rhamnose	3,4-O-dimethyl-L-rhamnose				
				144.072 (Y2)	145.071 (B)	D-chalcose	D-chalcose	AmT, oxDA, 3KR, O-MT			
						D-mycarose					
						L-oleandrose					
						olivomose					
19	Sch40832	ESI-MS	Puar, M.S. *et al*. *J. Antibiot* **51**, 221-224 (1998).	130 (Y)		D-digitoxose			none	GT	GG657738.1
						L-digitoxose					
						2-deoxy-L-fucose					
						D-olivose					
						D-oliose					
						L-boivinose					
20	lomaiviticin C	ESI-Q-TOF-MS	Kersten, R. D. *et al. Chembiochem* **14**, 955-962 (2013).	144.077 (Y)	145.086 (B)	D -chalcose			2,3DH, 3KR, 4KR, E, AmT, O-MT, MT	NT, 4,6DH, GT (2x)	CP000667
						L -oleandrose	L -oleandrose	2,3DH, 3KR, 4KR, E, O-MT			
						D -mycarose					
						olivomose	olivomose	2,3DH, 3KR, 4KR, O-MT			
				175.121 (Z)	158.117 (B)	4-N-ethyl-4-amino-3-O-methoxy-2,4,5-trideoxypentose					
						D-3-N-methyl-4-O-methyl-L-ristosamine	D-3-N-methyl-4-O-methyl-L-ristosamine	2,3DH, 4KR, E, AmT, N-MT, O-MT			
						D-desosamine					
						N,N-dimethyl-L-pyrrolosamine	N,N-dimethyl-L-pyrrolosamine	2,3DH, 3KR, E, AmT, N,N-MT			
						L-megosamine	L-megosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-nogalamine	L-nogalamine	2,3DH, AmT, E, 4KR, N,N-MT			
						L-rhodosamine	L-rhodosamine	2,3DH, AmT, E, 4KR, N,N-MT			
						D-angolosamine	D-angolosamine	2,3DH, AmT, 4KR, N,N-MT			
						kedarosamine	kedarosamine	2,3DH, 3KR, AmT, N,N-MT			

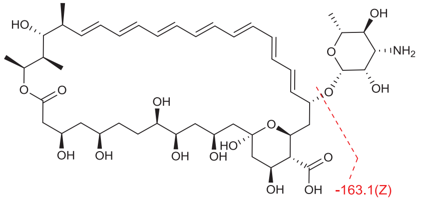
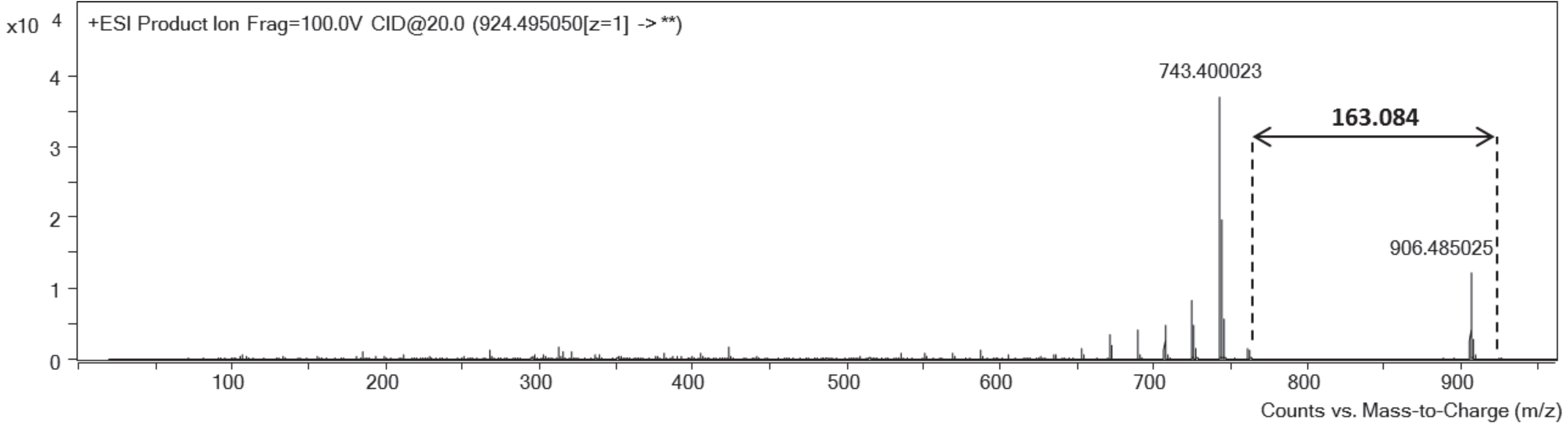
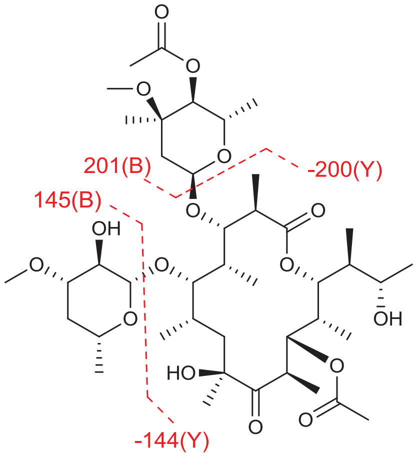
Dataset S4. MS/MS-fragmentation of known GNP chemotypes and gene predictions of corresponding GNP genotypes. MS/MS-sugar fragmentation and glycosylation gene prediction from chemotypes and genotypes of characterized glycosylated natural products (GNPs) from databases (Dataset S3) or self-acquired MS/MS data.

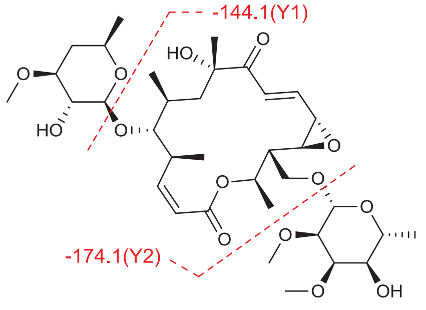
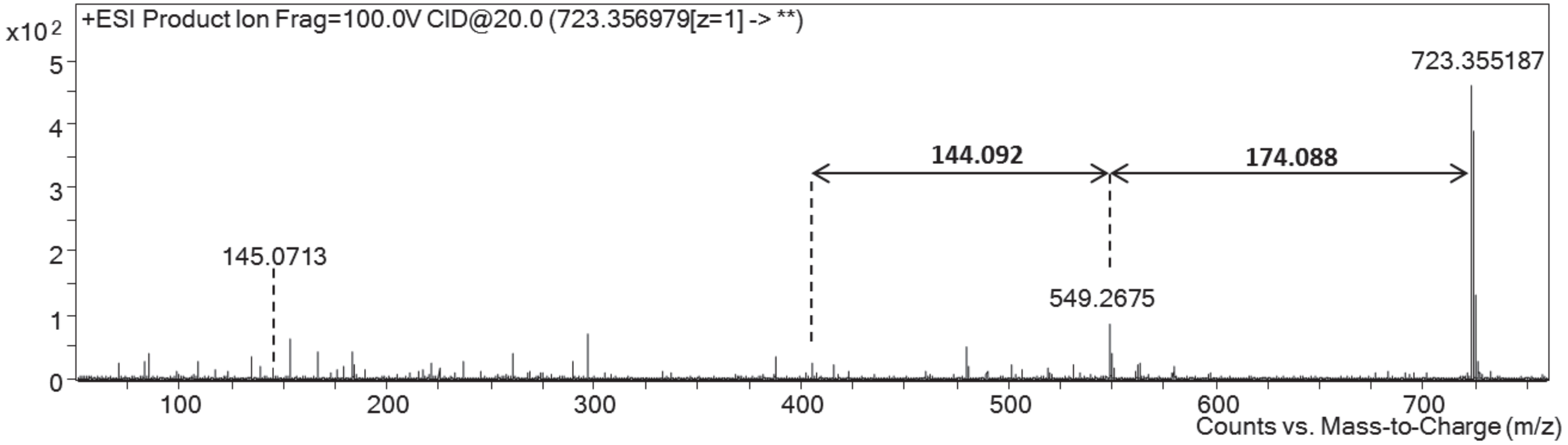
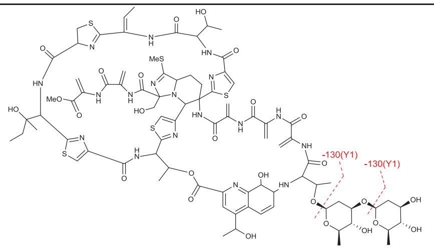
<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>daunomycin</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>dnmT</td><td>2,3DH</td><td>BAD08363.1| dTDP-4-keto-6-deoxyglucose 2,3-dehydratase [Streptomyces halstedii] (73/57)</td></tr>
<tr><td>dnrH</td><td>GT</td><td>AAB08020.1| glycosyltransferase [Streptomyces sp. C5] (93/93)</td></tr>
<tr><td>dnrS</td><td>GT</td><td>ABC00729.1| CosG (glycosyltransferase) [Streptomyces olindensis] (75/61)</td></tr>
<tr><td>dnmU</td><td>E</td><td>AF257324_2 AknL (epimerase) [Streptomyces galilaeus] (83/73)</td></tr>
<tr><td>dnmV</td><td>4KR</td><td>CAA12010.1| SnogG [Streptomyces nogalater] (63/53)</td></tr>
<tr><td>dnrM</td><td>4,6DH</td><td>ZP_06913915.1| dTDP-glucose 4,6-dehydratase [Streptomyces pristinaespiralis ATCC 25486] (62/57)</td></tr>
<tr><td>dnmL</td><td>NT</td><td>ZP_06913914.1| dTDP-glucose synthase [Streptomyces pristinaespiralis ATCC 25486] (89/79)</td></tr>
<tr><td>dnrJ</td><td>AmT</td><td>CCD33157.1| putative C-3 aminotransferase [Amycolatopsis orientalis] (85/74)</td></tr>
</table>
</td></tr>
<tr><td>staurosporine</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>staB</td><td>4,6DH</td><td>ZP_05008524.1| dTDP-glucose 4,6-dehydratase [Streptomyces clavuligerus ATCC 27064] (89/81)</td></tr>
<tr><td>staA</td><td>NT</td><td>ZP_05008523.1| glucose-1-phosphate thymidyltransferase [Streptomyces clavuligerus ATCC 27064] (87/78)</td></tr>
<tr><td>staG</td><td>GT</td><td>CAD58668.1| putative glycosyltransferase [Streptomyces longisporoflavus]</td></tr>
<tr><td>staN</td><td>GT</td><td>CAD58669.1| putative P450 protein [Streptomyces longisporoflavus] (99/98)</td></tr>
<tr><td>staMA</td><td>O-MT</td><td>ZP_06776292.1| Staurosporine biosynthesis (O-)methyltransferase StaMA [Streptomyces clavuligerus ATCC 27064] (81/71)</td></tr>
<tr><td>staI</td><td>AmT</td><td>CCD33157.1| putative C-3 aminotransferase [Amycolatopsis orientalis] (92/85)</td></tr>
<tr><td>staJ</td><td>2,3DH</td><td>CAC48374.1| putative NDP-hexose 2,3-dehydratase [Amycolatopsis balhimycina DSM 5908] (85/75)</td></tr>
<tr><td>staK</td><td>4KR</td><td>ZP_06776290.1| Staurosporine biosynthesis 4-ketoreductase StaK [Streptomyces clavuligerus ATCC 27064] (84/76)</td></tr>
<tr><td>staE</td><td>E</td><td>ZP_06776288.1| Staurosporine biosynthesis 3,5-epimerase StaE [Streptomyces clavuligerus ATCC 27064] (92/82)</td></tr>
<tr><td>staMB</td><td>MT</td><td>YP_001537181.1| type 11 methyltransferase [Salinispora arenicola CNS-205] (92/81)</td></tr>
</table>
</td></tr>
<tr><td>oleando-mycin</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>oleW</td><td>3KR</td><td>CCH33151.1| NDP-hexose 3-ketoreductase [Saccharothrix espanaensis DSM 44229] (69/55)</td></tr>
<tr><td>oleV</td><td>2,3DH</td><td>ZP_10457036.1| NDP-hexose 2,3-dehydratase [Streptomyces acidiscabies 84-104] (74/60)</td></tr>
<tr><td>oleL</td><td>E</td><td>ADI50280.1| sugar 3,5-epimerase [Streptomyces sp. MK730-62F2] (66/56)</td></tr>
<tr><td>oleS</td><td>NT</td><td>ZP_07308385.1| glucose-1-phosphate thymidylyltransferase [Streptomyces viridochromogenes DSM 40736] (85/75)</td></tr>
<tr><td>oleU</td><td>4KR</td><td>ZP_07308387.1| dTDP-4-dehydrorhamnose reductase [Streptomyces viridochromogenes DSM 40736]</td></tr>
<tr><td>oleE</td><td>4,6DH</td><td>ZP_07308386.1| dTDP-glucose 4,6-dehydratase [Streptomyces viridochromogenes DSM 40736] (84/76)</td></tr>
<tr><td>oleNI</td><td>3,4DH/AmT</td><td>YP_001102983.1| eryCIV NDP-6-deoxyhexose 3,4-dehydratase [Saccharopolyspora erythraea NRRL 2338] (79/66)</td></tr>
<tr><td>oleT</td><td>ox.DA</td><td>YP_001102982.1| EryCV NDP-4,6-dideoxyhexose 3,4-enoyl reductase [Saccharopolyspora erythraea NRRL 2338] (76/65)</td></tr>
<tr><td>oleNII</td><td>AmT</td><td>YP_001103001.1| erythromycin biosynthesis transaminase EryCI [Saccharopolyspora erythraea NRRL 2338]</td></tr>
<tr><td>oleI</td><td>GT</td><td>ZP_07289948.1| oleandomycin glycosyltransferase [Streptomyces sp. C] (59/44)</td></tr>
<tr><td>oleD</td><td>GT</td><td>YP_006248311.1| oleandomycin glycosyltransferase [Streptomyces hygroscopicus subsp. jinggangensis 5008] (92/87)</td></tr>
<tr><td>oleY</td><td>O-MT</td><td>BAC57026.1| methyltransferase [Micromonospora griseorubida] (58/41)</td></tr>
<tr><td>oleM1</td><td>N,N-MT</td><td>YP_001102985.1| TDP-desosamine-N-dimethyltransferase [Saccharopolyspora erythraea NRRL 2338] (77/68)</td></tr>
<tr><td>oleG2</td><td>GT</td><td>YP_001102993.1| glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (71/54)</td></tr>
<tr><td>oleG1</td><td>GT</td><td>YP_001102993.1| glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (70/54)</td></tr>
</table>
</td></tr>
</table>

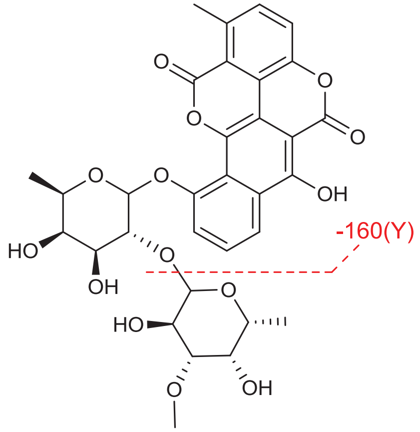
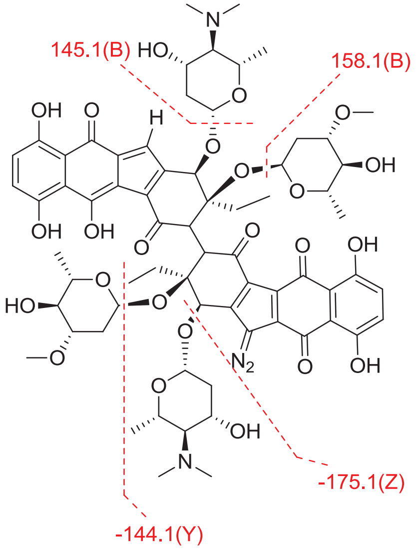
<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>spinosyn A</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>spnS</td><td>N,N-MT</td><td>CBH32796.1| putative N,N-dimethyltransferase [Streptomyces ravidus] (67/50)</td></tr>
<tr><td>spnR</td><td>AmT</td><td>ADM72812.1| putative NDP-hexose aminotransferase [Streptomyces aureofaciens] (78/64)</td></tr>
<tr><td>spnQ</td><td>3,4DH</td><td>AF264025_3 putative 3-dehydratase [Streptomyces galilaeus] (85/72)</td></tr>
<tr><td>spnP</td><td>GT</td><td>ZP_06826086.1| glycosyltransferase family 28 domain-containing protein [Streptomyces sp. SPB74] (61/45)</td></tr>
<tr><td>spnO</td><td>2,3DH</td><td>AF324838_29 putative dNDP-4-keto-6-deoxy-glucose-2,3-dehydratase SimB3 [Streptomyces antibioticus] (67/53)</td></tr>
<tr><td>spnN</td><td>3KR</td><td>AF264025_4 putative 3-ketoreductase [Streptomyces galilaeus] (68/53)</td></tr>
<tr><td>spnK</td><td>O-MT</td><td>BAJ05901.1| putative sugar O-methyltransferase [Streptomyces sp. SANK 60405] (69/53)</td></tr>
<tr><td>spnI</td><td>GT</td><td>YP_004965109.1| Gene info linked to YP_004965109.1glycosyl transferase [Streptomyces bingchenggensis BCW-1] (54/41)</td></tr>
<tr><td>spnI</td><td>O-MT</td><td>ZP_05000469.1| O-methyltransferase [Streptomyces sp. Mg1] (59/42)</td></tr>
<tr><td>spnH</td><td>O-MT</td><td>AEP40941.1| sugar O-methyltransferase [Nocardiopsis sp. FU40] (77/62)</td></tr>
<tr><td>spnF</td><td>O-MT</td><td>ACY01395.1| O-methyl transferase [Streptomyces platensis subsp. rosaceus] (55/39)</td></tr>
</table>
</td></tr>
<tr><td>vancomycin</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>gtfD</td><td>GT</td><td>CAA76553.1| glycosyltransferase [Amycolatopsis balhimycina DSM 5908] (81/71)</td></tr>
<tr><td>gtfE</td><td>GT</td><td>CAA76552.1| glycosyltransferase [Amycolatopsis balhimycina DSM 5908] (88/83)</td></tr>
<tr><td>vcaC</td><td>C-MT</td><td>CAC48364.1| putative C-3 methyl transferase [Amycolatopsis balhimycina DSM 5908] (96/93)</td></tr>
<tr><td>vcaA</td><td>2,3DH</td><td>CAC48374.1| putative NDP-hexose 2,3-dehydratase [Amycolatopsis balhimycina DSM 5908] (92/87)</td></tr>
<tr><td>vcaE</td><td>4KR</td><td>AEI58885.1| 4-ketoreductase [Amycolatopsis orientalis HCCB10007] (85/76)</td></tr>
<tr><td>vcaD</td><td>E</td><td>CAC48377.1| putative 3,5 epimerase [Amycolatopsis balhimycina DSM 5908] (94/88)</td></tr>
<tr><td>vcaB</td><td>AmT</td><td>CAC48376.1| putative C-3 amino transferase [Amycolatopsis balhimycina DSM 5908] (96/90)</td></tr>
</table>
</td></tr>
<tr><td>tylosin</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>tylN</td><td>GT</td><td>ABV49604.1| glycosyltransferase [Streptomyces eurythermus] (83/75)</td></tr>
<tr><td>tylE</td><td>O-MT</td><td>ABV49603.1| O-methyltransferase [Streptomyces eurythermus] (83/75)</td></tr>
<tr><td>tylD</td><td>4KR</td><td>ABV49602.1| NDP-4-ketoreductase [Streptomyces eurythermus] (74/62)</td></tr>
<tr><td>tylJ</td><td>E</td><td>ABV49598.1| NDP-hexose 3,(5)-epimerase [Streptomyces eurythermus] (84/72)</td></tr>
<tr><td>tylCII</td><td>3KR</td><td>YP_001102994.1| TDP-4-keto-6-deoxyhexose 2,3-reductase [Saccharopolyspora erythraea NRRL 2338] (84/74)</td></tr>
<tr><td>tylCIV</td><td>4KR</td><td>ABW91155.1| NDP-hexose 4-ketoreductase [Streptomyces eurythermus] (70/60)</td></tr>
<tr><td>tylCIII</td><td>C-MT</td><td>ABW91157.1| NDP-hexose 3-C-methyltransferase [Streptomyces eurythermus] (88/78)</td></tr>
<tr><td>tylCV</td><td>GT</td><td>ABW91158.1| glycosyltransferase [Streptomyces eurythermus] (78/68)</td></tr>
<tr><td>tylCVII</td><td>E</td><td>ABW91159.1| NDP-hexose 3,5-epimerase [Streptomyces eurythermus] (80/74)</td></tr>
<tr><td>tylB</td><td>AmT</td><td>AAF59939.1| aminotransferase-like protein [Streptomyces antibioticus] (73/63)</td></tr>
<tr><td>tylA1</td><td>NT</td><td>ABV49608.1| NDP-hexose synthase [Streptomyces eurythermus] (83/76)</td></tr>
<tr><td>tylA2</td><td>4,6DH</td><td>ABV49607.1| NDP-hexose 4,6-dehydratase [Streptomyces eurythermus] (83/74)</td></tr>
<tr><td>tylCVI</td><td>2,3DH</td><td>ABV49606.1| NDP-hexose 2,3-dehydratase/thioesterase [Streptomyces eurythermus] (77/66)</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>avermectin B1a</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>aveBI</td><td>GT</td><td>YP_004967697.1|glycosyl transferase family protein [Streptomyces bingchenggensis BCW-1] (59/47)</td></tr>
<tr><td>aveBII</td><td>4,6DH</td><td>YP_003132507.1| dTDP-glucose 4,6-dehydratase [Saccharomonospora viridis DSM 43017] (84/73)</td></tr>
<tr><td>aveBIII</td><td>NT</td><td>CAJ88197.1| putative glucose-1-phosphate thymidyltransferase [Streptomyces ambofaciens ATCC 23877] (85/75)</td></tr>
<tr><td>aveBIV</td><td>4KR</td><td>BAC55215.1| 4-ketoreductase [Streptomyces sp. TP-A0274] (64/54)</td></tr>
<tr><td>aveBV</td><td>E</td><td>NP_851468.1| putative NDP-4-keto-6-deoxyhexose 3,5-epimerase [Streptomyces rochei] (73/65)</td></tr>
<tr><td>aveBVI</td><td>2,3DH</td><td>AAD13549.1| NDP-hexose 2,3-dehydratase homolog [Streptomyces cyanogenus] (79/71)</td></tr>
<tr><td>aveBVII</td><td>O-MT</td><td>NP_851467.1| putative NDP-hexose 3-O-methyltransferase [Streptomyces rochei] (92/85)</td></tr>
<tr><td>aveBVIII</td><td>3KR</td><td>ZP_06594449.1| dTDP-4-keto-6-deoxy-L-hexose 2,3-reductase [Streptomyces albus J1074] (88/81)</td></tr>
</table>
</td></tr>
<tr><td>nystatin</td><td>No sugar fragmentation</td><td>-</td></tr>
<tr><td>erythromycin A</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>eryBVII</td><td>E</td><td>ZP_10551370.1| dTDP-4-deoxyglucose 3,5-epimerase [Streptomyces auratus AGR0001] (82/76)</td></tr>
<tr><td>eryCV</td><td>ox.DA</td><td>NP_851461.1| putative NDP-4,6-dideoxyhexose 3,4-enoyl reductase [Streptomyces rochei] (77/66)</td></tr>
<tr><td>eryCIV</td><td>3,4DH/AmT</td><td>NP_851460.1| putative NDP-6-deoxyhexose 3,4-dehydratase [Streptomyces rochei] (81/71)</td></tr>
<tr><td>eryBVI</td><td>2,3DH</td><td>ZP_10457036.1| NDP-hexose 2,3-dehydratase [Streptomyces acidiscabies 84-104] (68/53)</td></tr>
<tr><td>eryCVI</td><td>N,N-MT</td><td>CAA05643.1| methyltransferase [Streptomyces antibioticus] (76/68)</td></tr>
<tr><td>eryBV</td><td>GT</td><td>AAG13915.1|AF263245_11 TDP-mycarose glycosyltransferase [Micromonospora megalomicea subsp. nigra] (86/75)</td></tr>
<tr><td>eryBIV</td><td>4KR</td><td>AAG13916.1|AF263245_12 TDP-4-keto-6-deoxyhexose 4-ketoreductase [Micromonospora megalomicea subsp. nigra] (81/70)</td></tr>
<tr><td>eryCII</td><td>3,4IM</td><td>AAG13920.1|AF263245_16 TDP-4-keto-6-deoxyglucose 3,4-isomerase [Micromonospora megalomicea subsp. nigra] (79/73)</td></tr>
<tr><td>eryCIII</td><td>GT</td><td>AAG13921.1|AF263245_17 TDP-desosamine glycosyltransferase [Micromonospora megalomicea subsp. nigra] (90/84)</td></tr>
<tr><td>eryBII</td><td>3KR</td><td>AAG13914.1|AF263245_10 TDP-4-keto-6-deoxyhexose 2,3-reductase [Micromonospora megalomicea subsp. nigra] (89/81)</td></tr>
<tr><td>eryG</td><td>MT</td><td>AAU93801.1| methylase [Aeromicrobium erythreum] (83/73)</td></tr>
<tr><td>eryBIII</td><td>C-MT</td><td>AAD41823.1|AF147704_3 NDP-hexose 3-C-methyltransferase TylCIII [Streptomyces fradiae] (85/72)</td></tr>
<tr><td>eryCI</td><td>AmT</td><td>AAC68680.1| transaminase [Streptomyces venezuelae] (76/67)</td></tr>
</table>
</td></tr>
<tr><td>megalomicin</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>megT</td><td>2,3DH</td><td>YP_001102984.1| NDP-4-keto-6-deoxy-glucose 2,3-dehydratase [Saccharopolyspora erythraea NRRL 2338] (63/56)</td></tr>
<tr><td>megDVI</td><td>3,4IM</td><td>YP_001102992.1| TDP-4-keto-6-deoxy-glucose 3,4-isomerase [Saccharopolyspora erythraea NRRL 2338] (61/51)</td></tr>
<tr><td>megDI</td><td>GT</td><td>YP_001102993.1| glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (82/68)</td></tr>
<tr><td>megY</td><td>AcT</td><td>ZP_09183446.1| acyltransferase 3 [Streptomyces sp. S4] (62/43)</td></tr>
<tr><td>megDIII</td><td>N,N-MT</td><td>YP_005464796.1| putative methyltransferase [Actinoplanes missouriensis 431] (75/66)</td></tr>
<tr><td>megDIV</td><td>E</td><td>YP_006881544.1| dTDP-4-dehydrorhamnose 3,5-epimerase [Streptomyces venezuelae ATCC 10712] (72/58)</td></tr>
<tr><td>megDVI</td><td>4KR</td><td>AAL14256.1| NDP-4-keto-6-deoxyhexose 4-ketoreductase [Streptomyces venezuelae ATCC 10712] (67/55)</td></tr>
<tr><td>megDVII</td><td>2,3DH</td><td>YP_001102994.1| TDP-4-keto-6-deoxyhexose 2,3-reductase [Saccharopolyspora erythraea NRRL 2338] (89/81)</td></tr>
<tr><td>megBV</td><td>GT</td><td>YP_001102986.1| 6-DEB TDP-mycarosyl glycosyltransferase [Saccharopolyspora erythraea NRRL 2338] (86/75)</td></tr>
<tr><td>megBIV</td><td>4KR</td><td>YP_001102987.1| dTDP-4-keto-6-deoxy-L-hexose 4-reductase [Saccharopolyspora erythraea NRRL 2338] (81/70)</td></tr>
<tr><td>megDII</td><td>AmT</td><td>ACB46490.1| sugar 3-aminotransferase [Actinomadura kijaniata] (85/74)\</td></tr>
<tr><td>megCII</td><td>GT</td><td>AAB84066.1| EryCII [Saccharopolyspora erythraea NRRL 2338] (79/72)</td></tr>
<tr><td>megCIII</td><td>GT</td><td>YP_001102993.1|glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (90/84)</td></tr>
<tr><td>megBII</td><td>3-KR</td><td>YP_001102994.1| TDP-4-keto-6-deoxyhexose 2,3-reductase [Saccharopolyspora erythraea NRRL 2338] (73/63)</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>aclacino-mycin A</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>aknP</td><td>3,4DH</td><td>AAL24451.1| RdmI [Streptomyces purpurascens] (87/76)</td></tr>
<tr><td>aknQ</td><td>3KR</td><td>AAA83425.2| RdmF [Streptomyces purpurascens] (79/64)</td></tr>
<tr><td>aknR</td><td>4,6DH</td><td>ZP_06826098.1| dTDP-glucose 4,6-dehydratase [Streptomyces sp. SPB74] (90/85)</td></tr>
<tr><td>aknS</td><td>GT</td><td>ZP_06826097.1| glycosyltransferase family 28 domain-containing protein [Streptomyces sp. SPB74] (89/83)</td></tr>
<tr><td>aknT</td><td>GT</td><td>ABC00727.1| CosT [Streptomyces olindensis] (55/45)</td></tr>
<tr><td>aknX2</td><td>N,N-MT</td><td>ZP_06826092.1| methyltransferase type 11 [Streptomyces sp. SPB74] (92/86)</td></tr>
<tr><td>aknY</td><td>NT</td><td>ZP_06826091.1| glucose-1-phosphate thymidylyltransferase [Streptomyces sp. SPB74] (96/91)</td></tr>
<tr><td>aknZ</td><td>AmT</td><td>CCD33157.1| putative C-3 aminotransferase [Amycolatopsis orientalis] (88/80)</td></tr>
<tr><td>aknM</td><td>4KR</td><td>CAA12010.1| SnogG [Streptomyces nogalater] (63/53)</td></tr>
<tr><td>aknL</td><td>E</td><td>ZP_06826087.1| dTDP-4-dehydrorhamnose 3,5-epimerase [Streptomyces sp. SPB74] (94/88)</td></tr>
<tr><td>aknK</td><td>GT</td><td>ZP_06826086.1| glycosyltransferase family 28 domain-containing protein [Streptomyces sp. SPB74] (95/90)</td></tr>
<tr><td>aknN</td><td>DH</td><td>AEM44313.1| goadsporin biosynthetic protein [uncultured bacterium] (64/56)</td></tr>
</table>
</td></tr>
<tr><td>novobiocin</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>novM</td><td>GT</td><td>AAN65229.1|AF329398_19 glycosyltransferase [Streptomyces roseochromogenes subsp. oscitans] (89/84)</td></tr>
<tr><td>novN</td><td>CarbT</td><td>AAO06921.1| GdmN [Streptomyces hygroscopicus] (74/62)</td></tr>
<tr><td>novO</td><td>MT</td><td>AAG29793.1|AF235050_16 putative methyltransferase [Streptomyces rishiriensis]</td></tr>
<tr><td>novP</td><td>O-MT</td><td>AAG29794.1|AF235050_17 O-methyltransferase [Streptomyces rishiriensis] (97/95)</td></tr>
<tr><td>novS</td><td>4KR</td><td>AAN65241.1|AF329398_31 dTDP-4-keto-6-deoxyhexose reductase [Streptomyces roseochromogenes subsp. oscitans] (91/84)</td></tr>
<tr><td>novT</td><td>4,6DH</td><td>AAG29802.1| dTDP-glucose 4,6-dehydratase [Streptomyces rishiriensis] (93/89)</td></tr>
<tr><td>novU</td><td>C-MT</td><td>AAN65243.1|AF329398_33 C-methyltransferase [Streptomyces roseochromogenes subsp. oscitans] (93/89)</td></tr>
<tr><td>novV</td><td>NT</td><td>AAG29804.1| dTDP-glucose synthase [Streptomyces rishiriensis] (96/91)</td></tr>
<tr><td>novW</td><td>E</td><td>AAG29805.1| dTDP-4-keto-6-deoxyglucose 3,5-epimerase [Streptomyces rishiriensis] (93/88)</td></tr>
</table>
</td></tr>
<tr><td>neocarzino-statin</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>AY117439.1:10097..10894</td><td>4-KR</td><td>CAJ88311.1| putative NDP-hexose 4-ketoreductase [Streptomyces ambofaciens ATCC 23877] (80/75)</td></tr>
<tr><td>AY117439.1:14674..15672</td><td>O-MT</td><td>YP_001539692.1| O-methyltransferase family protein [Salinispora arenicola CNS-205] (77/66)</td></tr>
<tr><td>AY117439.1:26475..27179</td><td>NT</td><td>CCK25159.1| mannose-1-phosphate guanyltransferase [Streptomyces davawensis JCM 4913] (82/70)</td></tr>
<tr><td>AY117439.1:27203..28198</td><td>4,6DH</td><td>ZP_11210703.1| NAD-dependent epimerase/dehydratase [Streptomyces somaliensis DSM 40738] (78/67)</td></tr>
<tr><td>AY117439.1:28341..29060</td><td>N-MT</td><td>ZP_10580916.1| methylase involved in ubiquinone/menaquinone biosynthesis [Bradyrhizobium sp. YR681] (62/41)</td></tr>
<tr><td>AY117439.1:29285..30493</td><td>GT</td><td>CCK29032.1| glycosyl transferase [Streptomyces davawensis JCM 4913] (64/52)</td></tr>
<tr><td>AY117439.1:67831..68811</td><td>2,3DH</td><td>ZP_05000564.1| NDP-hexose-2,3-dehydratase [Streptomyces sp. Mg1] (73/59)</td></tr>
<tr><td>AY117439.1:11170..12750</td><td>AmT</td><td>ZP_05001621.1| histidine ammonia-lyase [Streptomyces sp. Mg1] (69/54)</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>amphotericin B</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>amphDIII</td><td>4,6DH</td><td>AAF71765.1|AF263912_4 NysDIII [Streptomyces noursei ATCC 11455] (94/89)</td></tr>
<tr><td>amphDII</td><td>AmT</td><td>AAF71772.1|AF263912_11 NysDII [Streptomyces noursei ATCC 11455] (96/90)</td></tr>
<tr><td>amphDI</td><td>GT</td><td>AAF71773.1|AF263912_12 NysDI [Streptomyces noursei ATCC 11455] (93/85)</td></tr>
<tr><td>amphN</td><td>3,4IM (CytP450)</td><td>AAF71771.1|AF263912_10 NysN [Streptomyces noursei ATCC 11455] (91/84)</td></tr>
</table></td></tr>
<tr><td></td><td colspan="2">
</td></tr>
<tr><td>lankamycin</td><td>
</td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>lkmD</td><td>4,6DH</td><td>YP_003103995.1| dTDP-glucose 4,6-dehydratase [Actinosynnema mirum DSM 43827] (79/67)</td></tr>
<tr><td>lkmBIII</td><td>C-MT</td><td>AAD41823.1|AF147704_3 NDP-hexose 3-C-methyltransferase TylCIII [Streptomyces fradiae] (82/71)</td></tr>
<tr><td>lkmG</td><td>O-MT</td><td>ADU56358.1| putative D-glucose O-methyltransferase [Streptomyces tacrolimicus] (58/43)</td></tr>
<tr><td>lkmBII</td><td>3KR</td><td>YP_001102994.1| TDP-4-keto-6-deoxyhexose 2,3-reductase [Saccharopolyspora erythraea NRRL 2338] (84/75)</td></tr>
<tr><td>lkmI</td><td>GT</td><td>YP_001102993.1| glycosyl transferase [Saccharopolyspora erythraea NRRL 2338] (78/64)</td></tr>
<tr><td>lkmCII</td><td>3,4IM</td><td>AAG13907.1|AF263245_3 TDP-4-keto-6-deoxyhexose 3,4-isomerase [Micromonospora megalomicea subsp. nigra] (56/46)</td></tr>
<tr><td>lkmJ</td><td>AcT</td><td>P_06711012.1| acyltransferase MdmB [Streptomyces sp. e14] (59/45)</td></tr>
<tr><td>lkmCIV</td><td>3,4DH/AmT</td><td>YP_001102983.1| eryCIV NDP-6-deoxyhexose 3,4-dehydratase [Saccharopolyspora erythraea NRRL 2338] (81/71)</td></tr>
<tr><td>lkmCV</td><td>ox.DA</td><td>AAB84076.1| EryCV [Saccharopolyspora erythraea NRRL 2338] (76/65)</td></tr>
<tr><td>lkmL</td><td>GT</td><td>CAA05642.1| glycosyltransferase [Streptomyces antibioticus] (68/56)</td></tr>
<tr><td>lkmBIV</td><td>4KR</td><td>ABW91155.1| NDP-hexose 4-ketoreductase [Streptomyces eurythermus] (63/54)</td></tr>
<tr><td>lkmCVI</td><td>3KR</td><td>ADI71457.1| putative sugar 3-ketoreductase [Amycolatopsis orientalis subsp. vinearia] (64/54)</td></tr>
<tr><td>lkmM</td><td>GT</td><td>YP_004818444.1| MGT family glycosyltransferase [Streptomyces violaceusniger Tu 4113] (82/73)</td></tr>
<tr><td>lkmO</td><td>O-MT</td><td>ZP_06594450.1| dTDP-6-deoxy-L-hexose 3-O-methyltransferase [Streptomyces albus J1074] (95/88)</td></tr>
<tr><td>lkmBVII</td><td>E</td><td>NP_822124.1| dTDP-4-keto-6-deoxyhexose 3,5-epimerase [Streptomyces avermitilis MA-4680] (73/65)</td></tr>
<tr><td>lkmU</td><td>NT</td><td>ADO32770.1| putative dTDP-1-glucose synthase [Streptomyces vietnamensis] (74/62)</td></tr>
<tr><td>lkmBVI</td><td>2,3DH</td><td>AAD55451.1| 2,3-dehydratase [Streptomyces antibioticus] (61/49)</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td rowspan="2">chalcomycin</td><td> </td><td><table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>chmCIV</td><td>3,4DH/AmT</td><td>ABB52533.1| 3,4-dehydratase-like protein [Streptomyces sp. KCTC 0041BP] (98/93)</td></tr>
<tr><td>chmCI</td><td>O-MT</td><td>ZP_05000456.1| sugar-O-methyltransferase [Streptomyces sp. Mg1] (93/87)</td></tr>
<tr><td>chmCV</td><td>ox.DA</td><td>ZP_05000457.1| NDP-4,6-dideoxyhexose-3,4-enoyl reductase [Streptomyces sp. Mg1] (92/89)</td></tr>
<tr><td>chmAII</td><td>4,6DH</td><td>ZP_05000462.1| dTDP-glucose-4,6-dehydratase [Streptomyces sp. Mg1] (97/97)</td></tr>
<tr><td>chmAI</td><td>NT</td><td>ABB52525.1| alpha-D-glucose-1-phosphate thymidylyltransferase [Streptomyces sp. KCTC 0041BP] (99/97)</td></tr>
<tr><td>chmJ</td><td>E</td><td>ZP_05000464.1| NDP-hexose-3-epimerase [Streptomyces sp. Mg1] (98/96)</td></tr>
<tr><td>chmMII</td><td>O-MT</td><td>ABB52523.1| 3-O-methyltransferase [Streptomyces sp. KCTC 0041BP] (98/96)</td></tr>
<tr><td>chmD</td><td>4-KR</td><td>ABB52541.1| hexose-4-ketoreductase [Streptomyces sp. KCTC 0041BP] (96/93)</td></tr>
<tr><td>chmMI</td><td>O-MT</td><td>ABB52542.1| O-methyltransferase [Streptomyces sp. KCTC 0041BP] (97/94)</td></tr>
<tr><td>chmN</td><td>GT</td><td>ZP_05000470.1| 6-deoxy-D-allosyltransferase [Streptomyces sp. Mg1] (97/95)</td></tr>
<tr><td>chmCII</td><td>3,4IM</td><td>ZP_05001883.1| NDP-hexose-3,4-isomerase [Streptomyces sp. Mg1]</td></tr>
<tr><td>chmCIII</td><td>GT</td><td>ABB52547.1| chalcosyltransferase [Streptomyces sp. KCTC 0041BP] (98/94)</td></tr>
<tr><td>chmU</td><td>3KR</td><td>ABB52548.1| 3-oxoacyl-(acyl-carrier-protein)-reductase [Streptomyces sp. KCTC 0041BP] (96/90)</td></tr>
</table></td></tr>
<tr><td colspan="2"> </td></tr>
<tr><td>Sch40832</td><td> </td><td><table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>MCAG_03952</td><td>GT</td><td>YP_003102514.1 Sterol 3-beta-glucosyltransferase [Actinosynnema mirum DSM 43827] (60/49)</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>GNP</th><th>MS/MS glycosylation footprints of characterized GNP</th><th>Glycosylation genes in gene cluster of characterized GNP</th></tr>
<tr><td>chartreusin</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>ChaGT2</td><td>GT</td><td>ZP_09566415.1| glycosyl transferase family 2 [Singulisphaera acidiphila DSM 18658] (72/54)</td></tr>
<tr><td>ChaM</td><td>MT</td><td>YP_002518358.1| methyltransferase [Caulobacter crescentus NA1000] (57/41)</td></tr>
<tr><td>ChaGT1</td><td>GT</td><td>CAE17548.1| glycosyltransferase [Streptomyces griseus subsp. griseus] (58/44)</td></tr>
<tr><td>ChaS1</td><td>NT</td><td>ADO32770.1| putative dTDP-1-glucose synthase [Streptomyces vietnamensis] (72/54)</td></tr>
<tr><td>ChaS3</td><td>4KR</td><td>YP_003301686.1| GtrA family protein [Thermomonospora curvata DSM 43183] (68/52)</td></tr>
<tr><td>ChaS2</td><td>4,6DH</td><td>YP_005464806.1| putative dTDP-glucose-4,6-dehydratase [Actinoplanes missouriensis 431] (82/69)</td></tr>
<tr><td>ChaS4</td><td>E/KR</td><td>ZP_04607399.1| nucleoside-diphosphate-sugar epimerase [Micromonospora sp. ATCC 39149] (56/44)</td></tr>
</table>
</td></tr>
<tr><td>lomaiviticin C</td><td> </td><td>
<table>
<tr><th>Gene</th><th>Predicted function</th><th>Closest functional homolog by BLAST (similiarity/identity) [%/%]</th></tr>
<tr><td>Strop2181</td><td>2,3DH</td><td>3-dehydratase [Streptomyces coelicoflavus ZG0656] (81/72)</td></tr>
<tr><td>Strop2213</td><td>GT</td><td>Glycosyltransferase [uncultured bacterium] (57/41)</td></tr>
<tr><td>Strop2215</td><td>O-MT</td><td>NanM [Streptomyces nanchangensis] (77/60)</td></tr>
<tr><td>Strop2216</td><td>4KR</td><td>4-ketoreductase [Streptomyces sp. TP-A0274] (68/56)</td></tr>
<tr><td>Strop2217</td><td>E</td><td>3,5-epimerase [Streptomyces sp. TP-A0274] (77/67)</td></tr>
<tr><td>Strop2218</td><td>3KR</td><td>Putative 3-ketoreductase [Streptomyces galilaeus] (68/55)</td></tr>
<tr><td>Strop2219</td><td>GT</td><td>Hypothetical protein [Salinispora arenicola CNS-205] (88/80)</td></tr>
<tr><td>Strop2220</td><td>GT</td><td>Glycosyltransferase [Streptomyces cyanogenus] (55/42)</td></tr>
<tr><td>Strop2221</td><td>3,4DH/AmT</td><td>3,4-dehydratase-like protein [Streptomyces sp. KCTC 0041BP] (74/61)</td></tr>
<tr><td>Strop2222</td><td>4,6DH</td><td>Putative dTDP-glucose-4,6-dehydratase [Actinoplanes missouriensis 431] (79/69)</td></tr>
<tr><td>Strop2230</td><td>NT</td><td>G1P thymidylyltransferase [Thermomonospora curvata DSM 43183] (87/72)</td></tr>
</table>
</td></tr>
</table>