Supplementary data for

Biological characterization of the hygrobafilomycin antibiotic JBIR-100 and bioinformatic insights into the hygrolide family of natural products.

Evelyn M. Molloy[a,†], *Jonathan I. Tietz*[b,†], *Patricia M. Blair*[b], *Douglas A. Mitchell*[a,b,c,*]

[a]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[b]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[c]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

*Corresponding author. Tel.: +1-217-333-1345; Fax: +1-217-244-8024; Email: douglasm@illinois.edu
[†] These two authors contributed equally to this manuscript.

Abbreviations

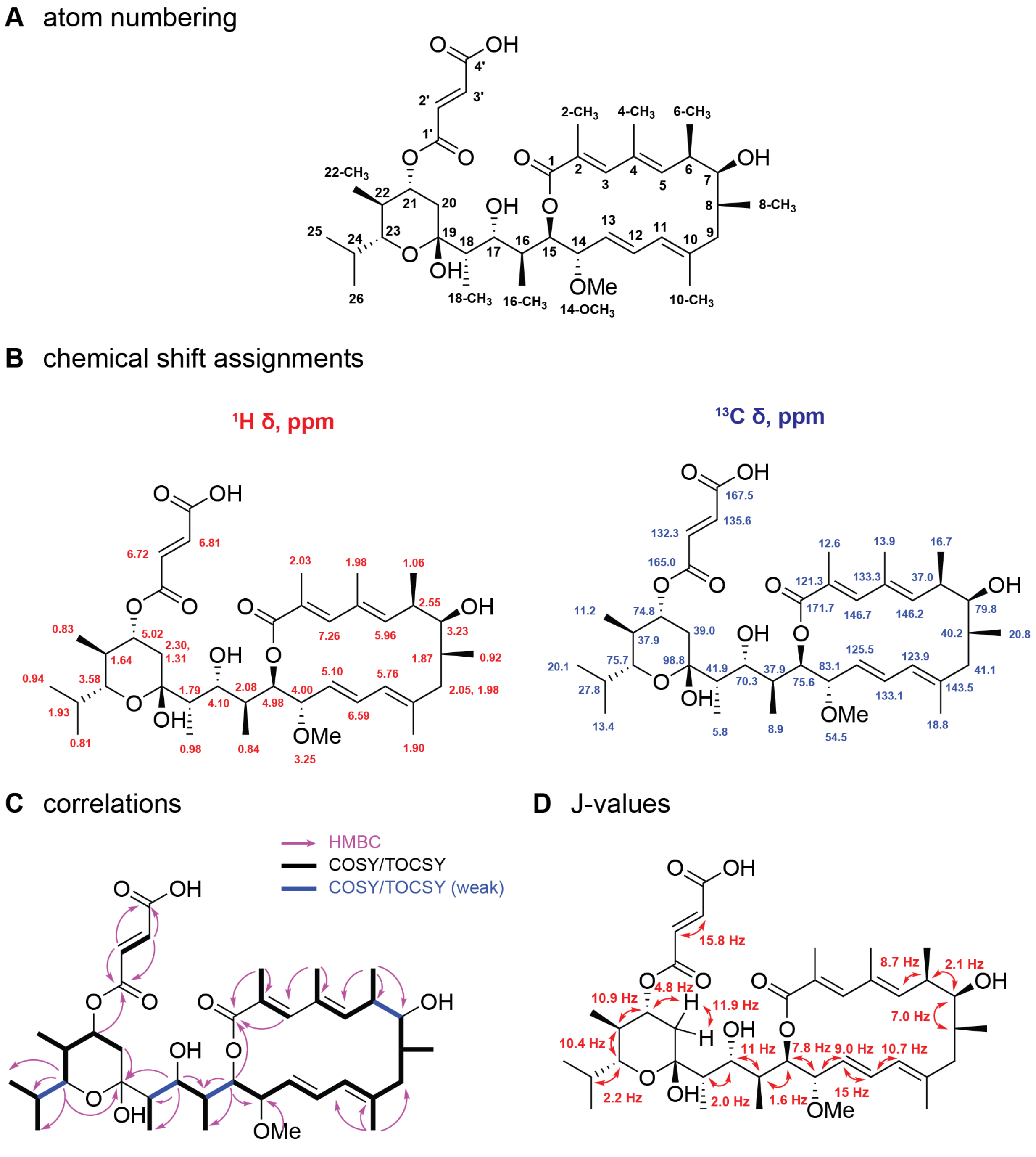
Biosynthetic gene cluster, BGC; ketoreductase, KR.

Table of Contents

Table S1. ^{1}H and ^{13}C NMR chemical shift assignments of JBIR-100.

Position	Group	13C δ (ppm)	1H δ (ppm)	m (J, Hz)
1	C=O	171.7	-	-
2	C	121.3	-	-
2-CH3	CH3	12.6	2.03	s
3	CH	146.7	7.26	br s
4	C	133.3	-	-
4-CH3	CH3	13.9	1.98	s
5	CH	146.2	5.96	d (8.7)
6	CH	37.0	2.55	dq (7.1, 8.7)
6-CH3	CH3	16.7	1.06	d (7.0)
7	CH	79.8	3.23	m
8	CH	40.2	1.87	m
8-CH3	CH3	20.8	0.92	d (6.9)
9	CH(a)	41.1	2.05	m
	CH(b)		1.98	m
10	C	143.5	-	-
10-CH3	CH3	18.8	1.90	s
11	CH	123.9	5.76	d (10.7)
12	CH	133.1	6.59	dd (15.0, 10.8)
13	CH	125.5	5.10	dd (14.9, 9.0)
14	CH	83.1	4.00	dd (9.1, 7.7)
14-OCH3	CH3	54.5	3.25	s
15	CH	75.6	4.98	dd (7.8, 1.6)
16	CH	37.9	2.08	m
16-CH3	CH3	8.9	0.84	d (6.5)
17	CH	70.3	4.10	dd (10.8, 1.9)
18	CH	41.9	1.79	qd (7.1, 2.0)
18-CH3	CH3	5.8	0.98	d (7.1)
19	C	98.8	-	-
20	CH(a)	39.0	2.30	dd (11.9, 4.8)
	CH(b)		1.31	t (11.6)
21	CH	74.8	5.02	td (10.9, 4.8)
22	CH	37.9	1.64	ddq (10.9, 10.5, 6.5)
22-CH3	CH3	11.2	0.83	d (6.2)
23	CH	75.7	3.58	dd (10.4, 2.2)
24	CH	27.8	1.93	td (6.8, 2.2)
25	CH3	20.1	0.94	d (6.9)
26	CH3	13.2	0.81	d (6.8)
1'	C=O	165	-	-
2'	CH	132.3	6.72	d (15.8)
3'	CH	135.6	6.81	d (15.8)
4'	C=O	167.5	-	-

Figure S1. Atom numbering (A), chemical shift assignments (B), correlations (C), and J-values of JBIR-100 (D).



A atom numbering

B chemical shift assignments

C correlations

D J-values

Figure S2. ^{1}H NMR spectrum of JBIR-100 (750 MHz, CD_3OD).

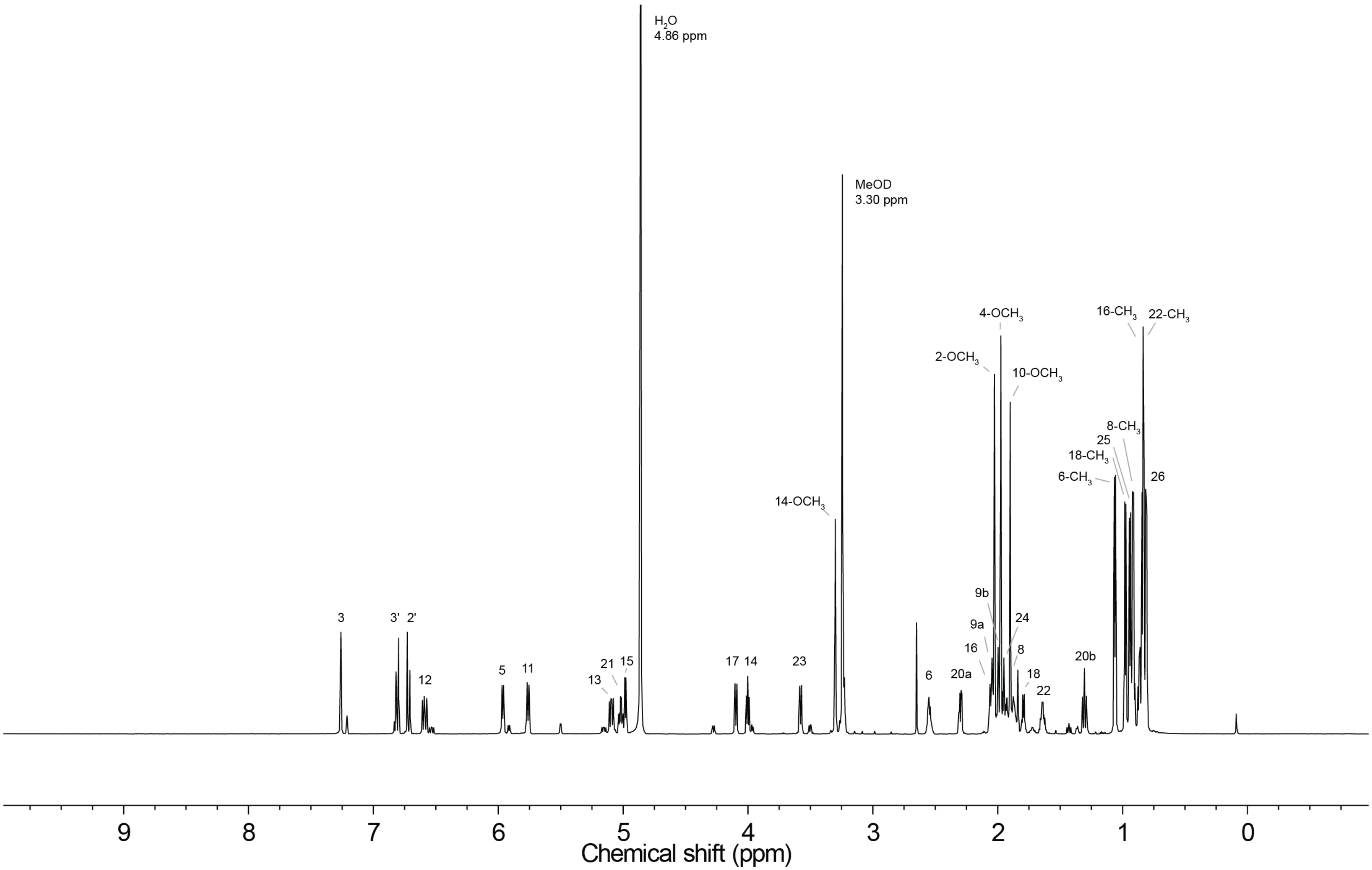


Figure S3. ^{1}H-^{1}H DQF-COSY spectrum of JBIR-100 (750 MHz, CD$_3$OD).

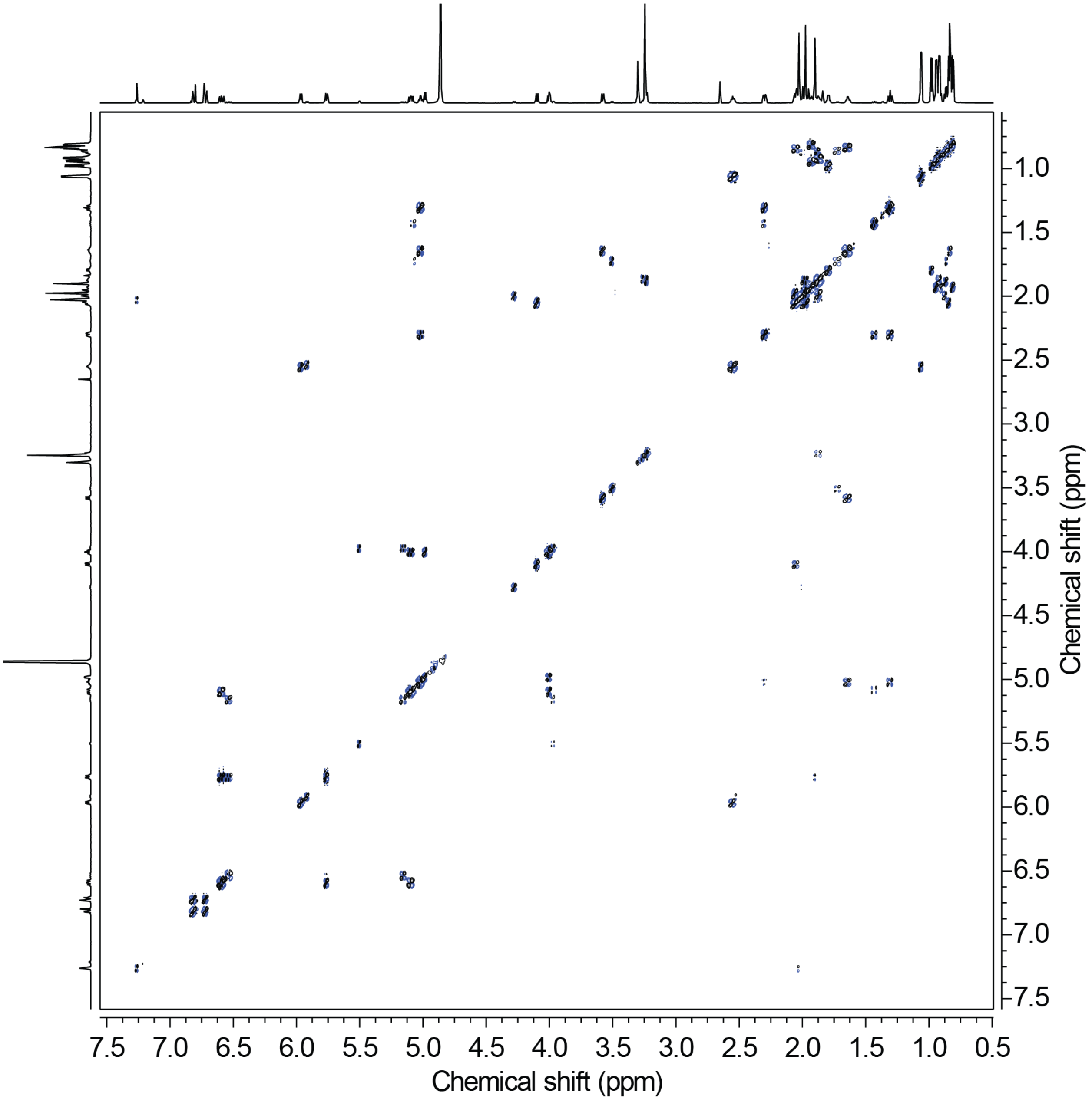


Figure S4. ^{1}H-^{1}H TOCSY spectrum of JBIR-100 (750 MHz, CD$_3$OD).

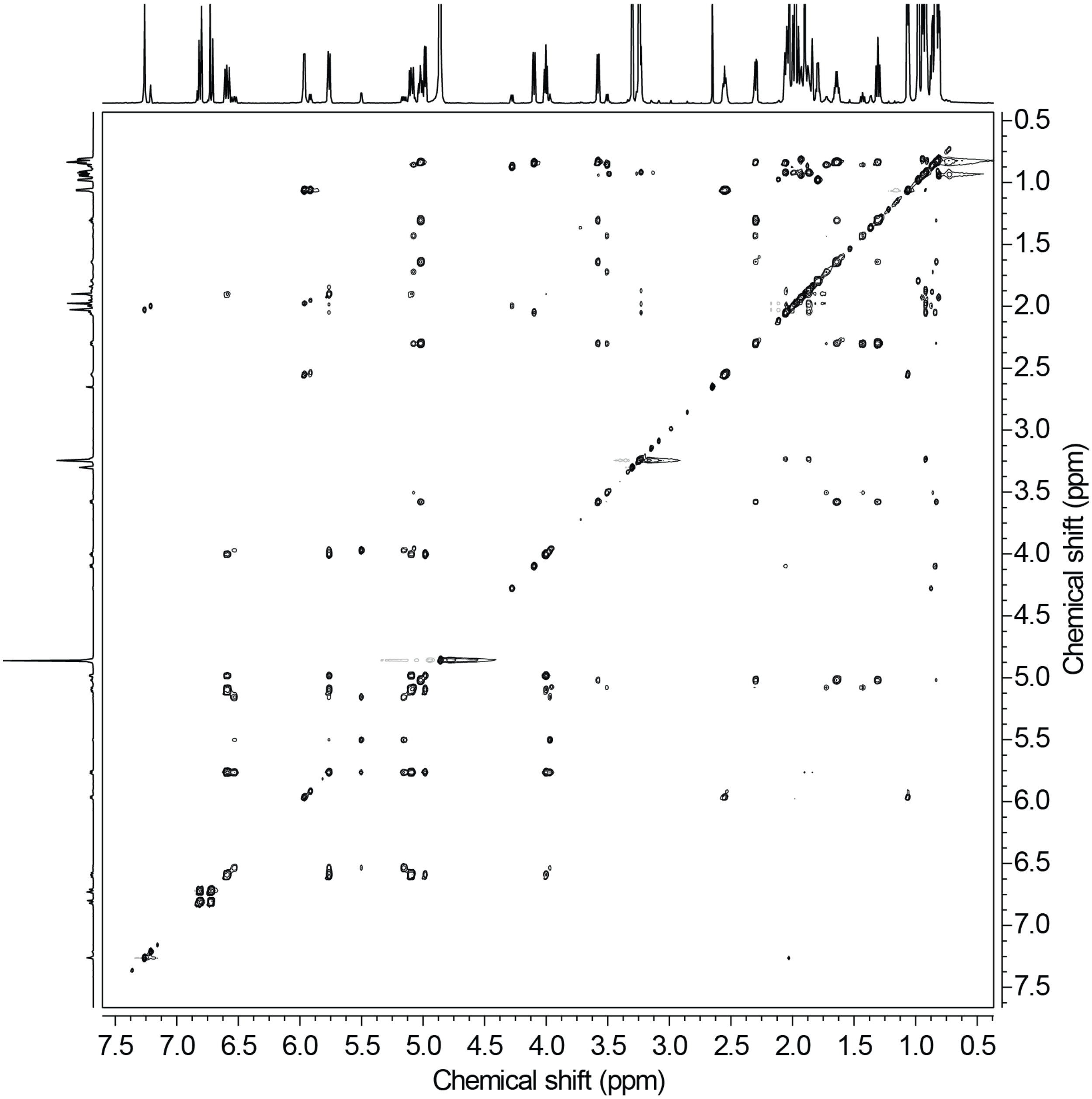


Figure S5. ^{1}H-^{13}C HSQC spectrum of JBIR-100 (750 MHz, CD_3OD).

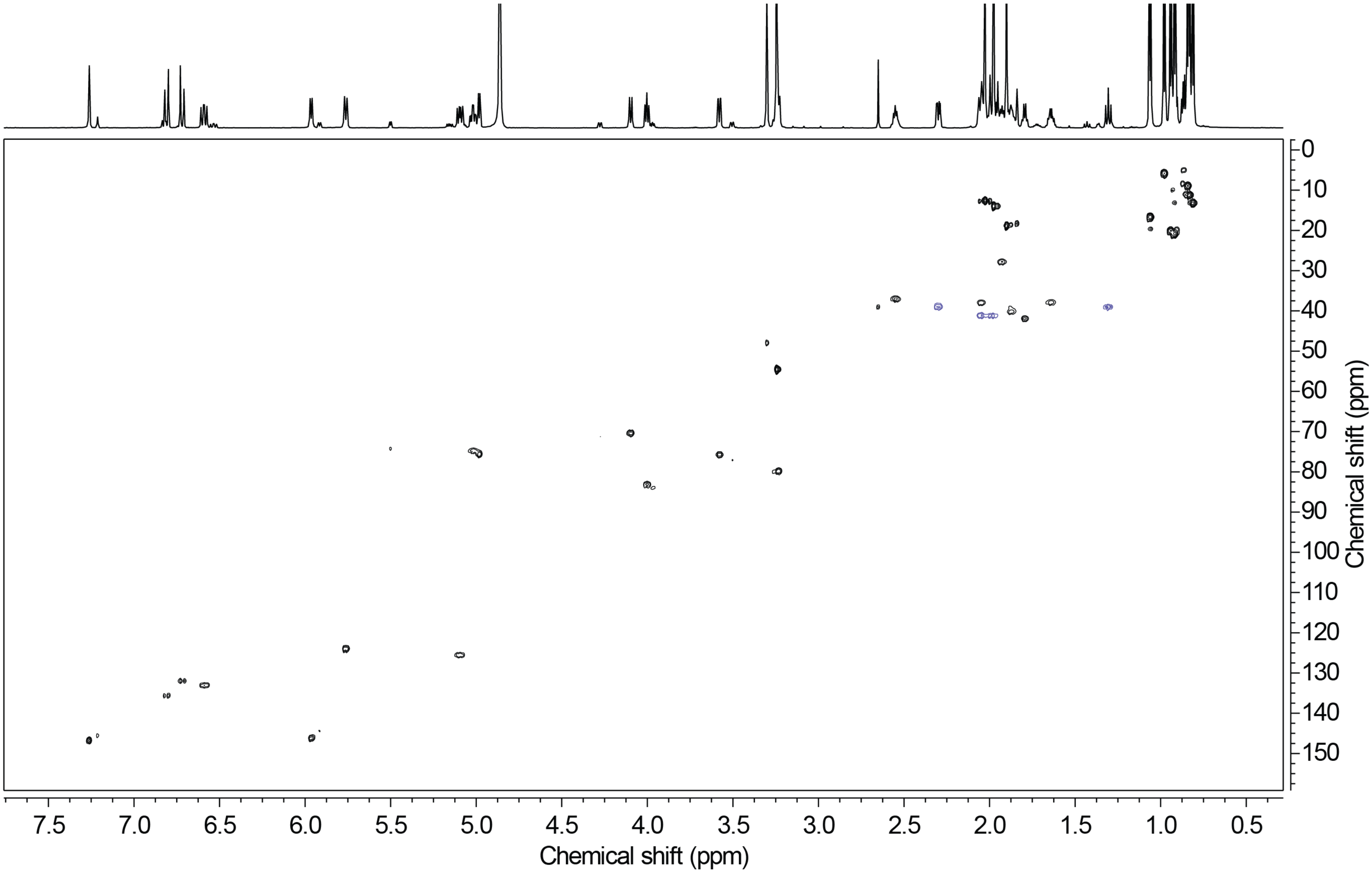


Figure S6. ^{1}H-^{13}C HMBC spectrum of JBIR-100 (750 MHz, CD_3OD)

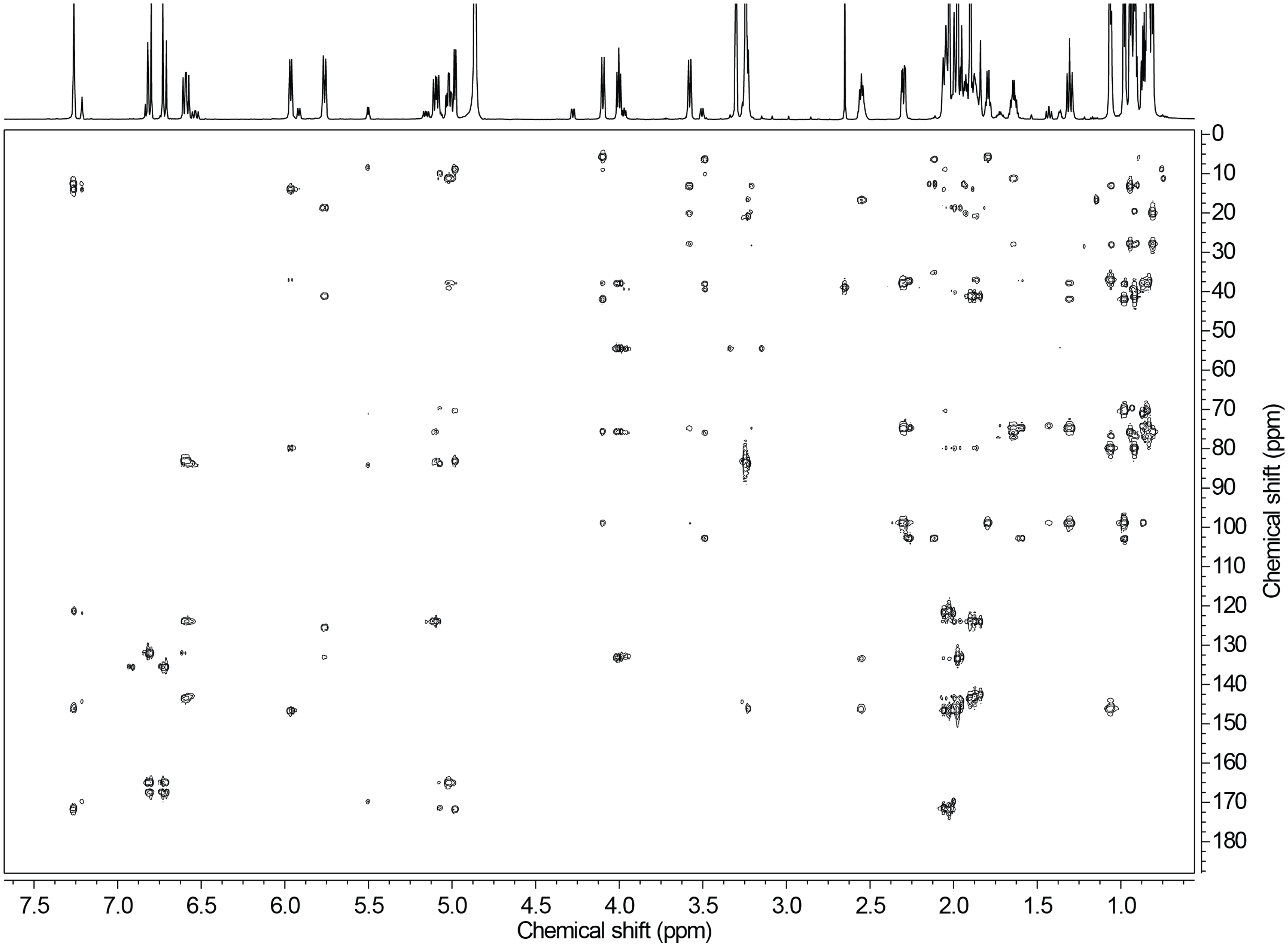


Figure S7. Growth defects of JBIR-100-resistant *B. subtilis* mutants. Growth rates (monitored at optical density at 595 nm) depicting the growth defects of JBIR-100-resistant mutants *B. subtilis* P10A, P10B and P10C compared to the parental strain, *B. subtilis* ATCC 47096. Data depicted is the average of two independent replicates.

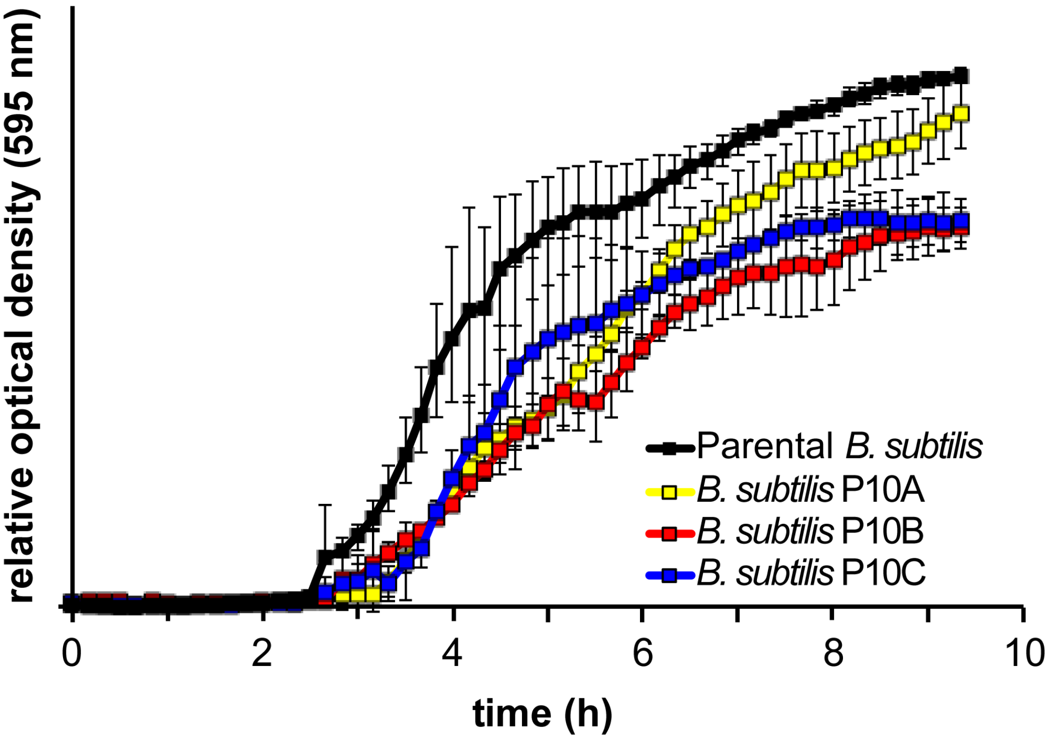


Figure S8. Representative plecomacrolide BGCs and their products. (A) Alignment of representative plecomacrolide BGCs. BGCs from *S. griseus, S. hygroscopicus, S. neyagawaensis,* and *S. lohii* were previously confirmed by genetic manipulation. BGCs from *S. halstedii, K. cheerisanensis, K. setae*, and *S. varsoviensis* were identified bioinformatically by homology to reported BGCs and biosynthetic logic. In *S. halstedii,* the BGCs are split between multiple partial-cluster contigs, as indicated. Interestingly, the methoxymalonate biosynthesis cassette is absent in the BGC from *S. hygroscopicus* despite the presence of this moiety in the molecule. *S. varsoviensis* lacks *bafXYZ* homologues thought to be responsible for 5-aminolevinulate installation; this structural moiety is not present in identified metabolites from this organism. (B) Representative natural products from the above BGCs are shown.

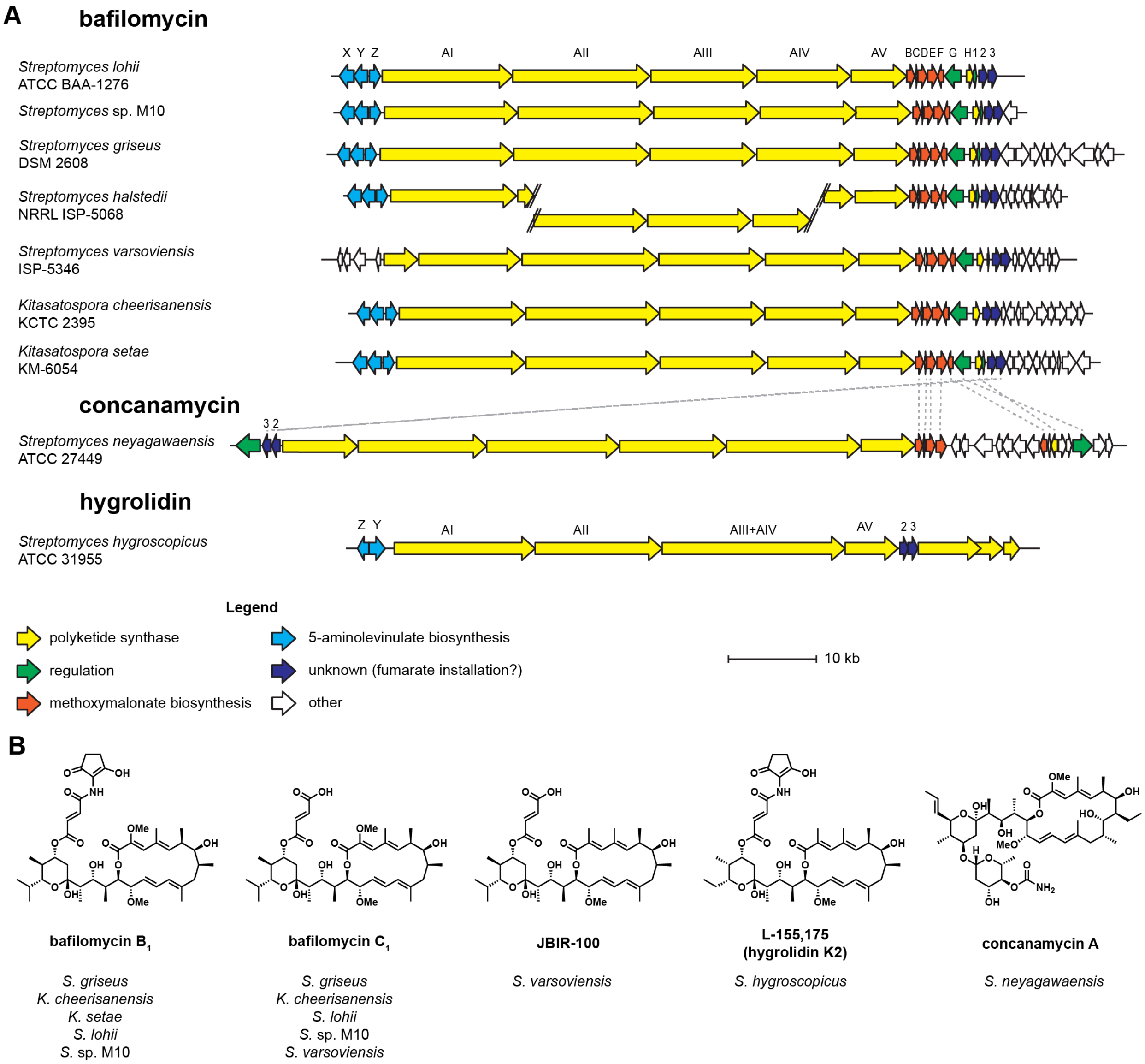
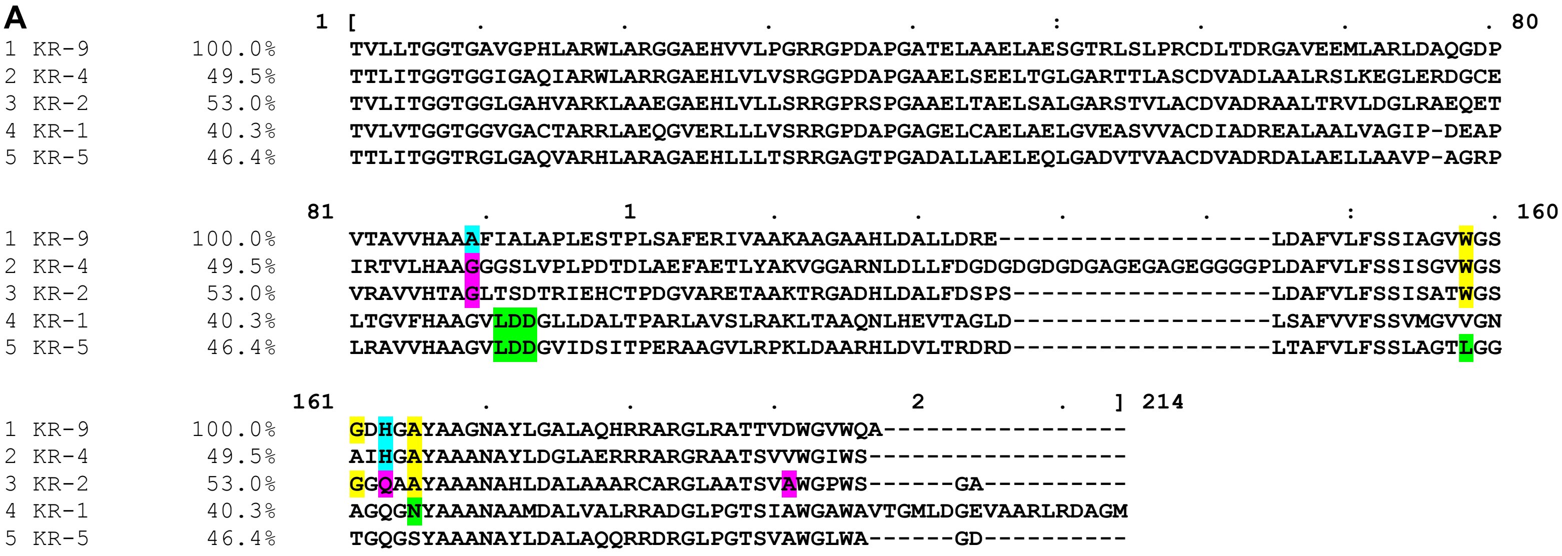


Figure S9. Stereospecificity prediction of KR domains. (A) Alignment of KR domains from the JBIR-100 BGC of *S. varsoviensis* NRRL ISP-5346. Sequence alignment was generated using MUSCLE (v. 3.8.31) on default settings and visualized using MView (v. 1.54). Key diagnostic residues are highlighted. (B) ScoreDiff values using prediction tool by Kitsche and Kalesse (https://akitsche.shinyapps.io/profileHMM_App) for classification of KR domains.[1] Positive ScoreDiff values indicate predicted D stereochemistry, whereas negative ScoreDiff values predict L stereochemistry. ScoreDiff values with larger magnitudes indicate increased certainty. Methyl stereochemistry of KR2, KR4, and KR5 could not be confidently predicted. (C) Structure of JBIR-100 labeled with KRs responsible for setting stereocenters.



```
                        1 [        .         .         .         .         :         .         .         . 80
1 KR-9       100.0%       TVLLTGGTGAVGPHLARWLARGGAEHVVLPGRRGPDAPGATELAAELAESGTRLSLPRCDLTDRGAVEEMLARLDAQGDP
2 KR-4        49.5%       TTLITGGTGGIGAQIARWLARRGAEHLVLVSRGGPDAPGAAELSEELTGLGARTTLASCDVADLAALRSLKEGLERDGCE
3 KR-2        53.0%       TVLITGGTGGLGAHVARKLAAEGAEHLVLLSRRGPRSPGAAELTAELSALGARSTVLACDVADRAALTRVLDGLRAEQET
4 KR-1        40.3%       TVLVTGGTGGVGACTARRLAEQGVERLLLVSRRGPDAPGAGELCAELAELGVEASVVACDIADREALAALVAGIP-DEAP
5 KR-5        46.4%       TTLITGGTRGLGAQVARHLARAGAEHLLLTSRRGAGTPGADALLAELEQLGADVTVAACDVADRDALAELLAAVP-AGRP

                       81          .         1         .         .         .         .         :         . 160
1 KR-9       100.0%       VTAVVHAAAFIALAPLESTPLSAFERIVAAKAAGAAHLDALLDRE-------------------LDAFVLFSSIAGVWGS
2 KR-4        49.5%       IRTVLHAAGGGSLVPLPDTDLAEFAETLYAKVGGARNLDLLFDGDGDGDGDGAGEGAGEGGGGPLDAFVLFSSISGVWGS
3 KR-2        53.0%       VRAVVHTAGLTSDTRIEHCTPDGVARETAAKTRGADHLDALFDSPS------------------LDAFVLFSSISATWGS
4 KR-1        40.3%       LTGVFHAAGVLDDGLLDALTPARLAVSLRAKLTAAQNLHEVTAGLD------------------LSAFVVFSSVMGVVGN
5 KR-5        46.4%       LRAVVHAAGVLDDGVIDSITPERAAGVLRPKLDAARHLDVLTRDRD------------------LTAFVLFSSLAGTLGG

                      161          .         .         .         2         .    ] 214
1 KR-9       100.0%       GDHGAYAAGNAYLGALAQHRRARGLRATTVDWGVWQA-----------------
2 KR-4        49.5%       AIHGAYAAANAYLDGLAERRRARGRAATSVVWGIWS------------------
3 KR-2        53.0%       GGQAAYAAANAHLDALAAARCARGLAATSVAWGPWS------GA----------
4 KR-1        40.3%       AGQGNYAAANAAMDALVALRRADGLPGTSIAWGAWAVTGMLDGEVAARLRDAGM
5 KR-5        46.4%       TGQGSYAAANAYLDALAQQRRDRGLPGTSVAWGLWA------GD----------
```

Key:

Type A (L-alcohol) **X** L-methyl **X** D-methyl **X** Type B (D-alcohol) **X**

B

Ketoreductase domain	Prediction	ScoreDiff
Domain 1 ketoreductase (KR1)	D-alcohol	29.6
Domain 2 ketoreductase (KR2)	L-alcohol	-25.6
Domain 4 ketoreductase (KR4)	L-alcohol	-77.09
Domain 5 ketoreductase (KR5)	D-alcohol	35.3
Domain 9 ketoreductase (KR9)	L-alcohol	-80.46
	L-methyl	-74.95

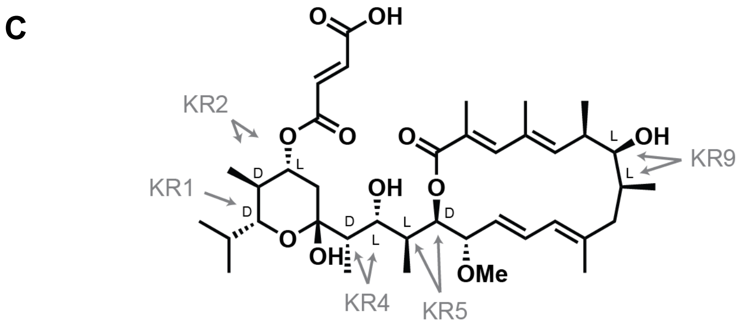
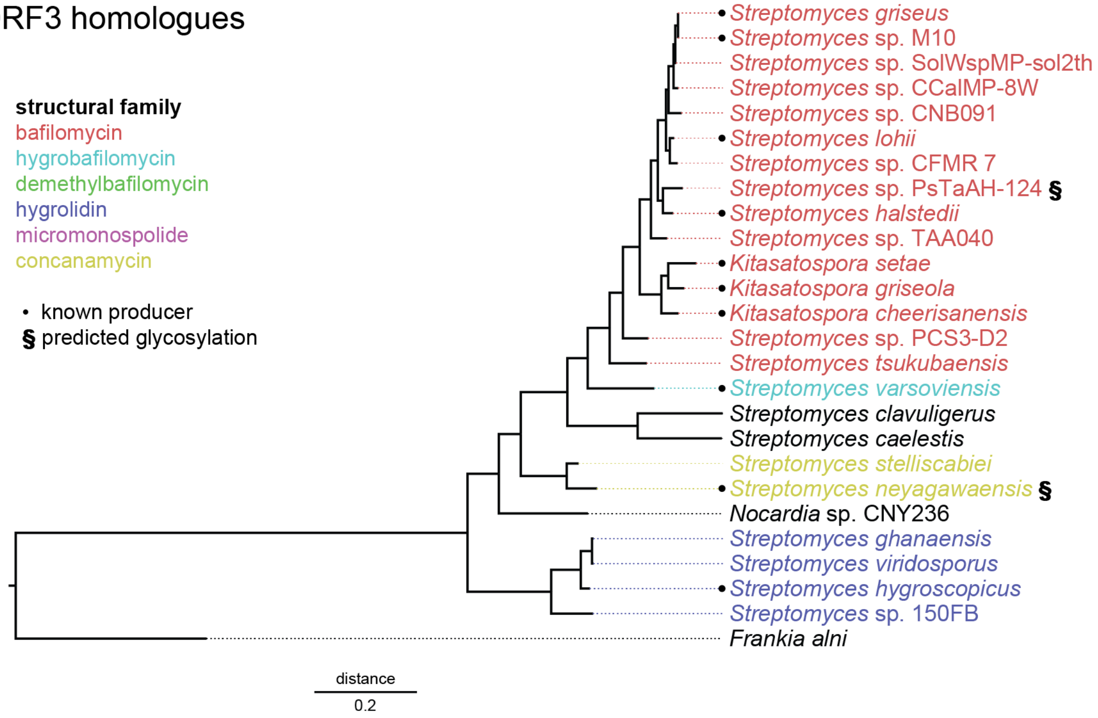


Table S2. Accession numbers for hygrolide BGCs. Nearest homologues of compounds from the *baf* BGC in *S. lohii* are shown. In some cases, locus tags or nucleotide coordinates are given in lieu of accessions for nucleotide-only records. The *bafAIV* homologue in *S. halstedii* is split between two contigs and so its percent identity was not assessed. In *S. hygroscopicus*, the protein corresponding to CCF2302 is a fusion of BafAIII and BafAIV.

	S. lohii		*S. varsoviensis*			*S.* sp. M10			*S. griseus*			*K. setae*			*K. cheerisanensis*			*S. halstedii*			*S. hygroscopicus*			*S. neyagawaensis*		
Gene	**Accession**	**Length (aa)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**Identity (%)**	**Accession**	**Length (aa)**	**I Identity (%)**
bafX	ADC79613	528	-	-	-	AIY53018	530	94	AGK25217	461	94	BAJ33096	538	88	KDN87768	518	89	WP_030630771	531	93	-	-	-	-	-	-
bafY	ADC79614	514	-	-	-	AIY53019	514	95	AGK25192	514	96	BAJ33097	517	87	KDN87767	514	89	WP_030630774	514	93	CCF23199	640	56	-	-	-
bafZ	ADC79615	414	-	-	-	AIY53020	414	98	AGK25193	414	98	BAJ33098	410	88	KDN87766	410	89	WP_030630777	414	93	CCF23198	426	71	-	-	-
bafAI	ADC79616	4884	IF55_RS32365 (locus)	5174	68	AIY53021	5007	93	AGK25194	5015	93	BAJ33099	4833	82	KDN87765	4759	84	WP_030817189	4807	89	CCF23200	5282	55	AAZ94387	4825	61
bafAII	ADC79617	5145	WP_048832936	5105	69	AIY53022	5064	95	AGK25213	5088	94	BAJ33100	5004	86	KDN87764	4988	86	WP_030818068 (partial)	4268	89	CCF23201	4763	62	AAZ94388	4970	68
bafAIII	ADC79618	3968	IF55_RS32375 (locus)	3935	67	AIY53023	3980	94	AGK25214	3980	94	BAJ33101	3964	85	KDN87763	3915	87	IG73_RS35125 (locus)	3880	90	CCF23202	6871	58	AAZ94389	3982	65
bafAIV	ADC79619	3511	WP_037966908	3473	68	AIY53024	3614	90	AGK25195	3626	89	BAJ33102	3436	85	KDN87762	3392	85	partial	ND	ND			55	AAZ94390	5006	62
bafAV	ADC79620	2103	IF55_RS32385 (locus)	2098	62	KM245330.1 [58129..64485]	ND	ND	AGK25196	2105	93	BAJ33103	2110	83	KDN87761	2097	85	WP_030631833	2123	90	CCF23203	1983	52	AAZ94391	2056	61
bafB	ADC79621	296	WP_048832937	294	78	AIY53025	296	97	AGK25191	296	97	BAJ33104	301	88	KDN87760	289	91	WP_028443380	289	91	-	-	-	AAZ94392	290	70
bafC	ADC79622	93	WP_048832938	100	73	AIY53026	93	97	AGK25197	93	97	BAJ33105	93	86	KDN87759	89	88	WP_028443381	93	94	-	-	-	AAZ94393	92	72
bafD	ADC79623	363	IF55_RS32400 (locus)	385	73	AIY53027	363	98	AGK25198	363	98	BAJ33106	371	90	KDN87758	371	90	WP_030631836	371	94	-	-	-	AAZ94394	371	73
bafE	ADC79624	365	WP_048832952	374	80	AIY53028	366	98	AGK25199	366	97	BAJ33107	383	88	KDN87757	366	93	WP_030631838	366	93	-	-	-	AAZ94395	380	72
bafF	ADC79625	220	WP_030890472	220	77	AIY53029	220	96	AGK25200	220	96	BAJ33108	220	91	KDN87756	220	92	WP_028443384	220	94	-	-	-	AAZ94403	222	69
bafG	ADC79626	609	WP_030890475	637	67	AIY53030	609	95	AGK25201	609	95	BAJ33109	622	84	KDN87755	639	84	WP_030631840	627	91	-	-	-	AAZ94408	701	53
bafH	ADC79627	253	WP_030890478	249	67	AIY53031	253	96	AGK25202	253	96	BAJ33110	253	82	KDN87754	213	83	WP_028443386	253	92	-	-	-	AAZ94405	270	48
orf1	ADC79628	117	WP_037966912	72	72	AIY53032	117	92	AGK25203	117	92	BAJ33111	105	74	-	-	-	WP_028443387	122	88	-	-	-	-	-	-
orf2	ADC79629	320	WP_030890486	319	79	AIY53033	348	97	AGK25204	348	97	BAJ33112	320	88	KDN87753	320	90	WP_030631842	320	92	CCF23204	324	57	AAZ94385	320	68
orf3	ADC79630	332	WP_048832939	366	82	AIY53034	359	97	AGK25205	359	97	BAJ33113	363	89	KDN87752	359	90	WP_028443389	359	94	CCF23205	362	67	AAZ94384	368	74

Figure S10. Phylogenetic tree of homologues from hygrolide/concanamycin BGCs. Shown are maximum-likelihood trees for (A) ORF3 homologues and (B) BafD homologues. The structural class of compound predicted or known to correspond to an organism is indicated. For organisms listed in black, a compound class could not be predicted and novelty is suspected. The actinomycete *Frankia alni* was chosen as an outgroup and is not predicted to contain a hygrolide BGC.

A ORF3 homologues



B BafD homologues

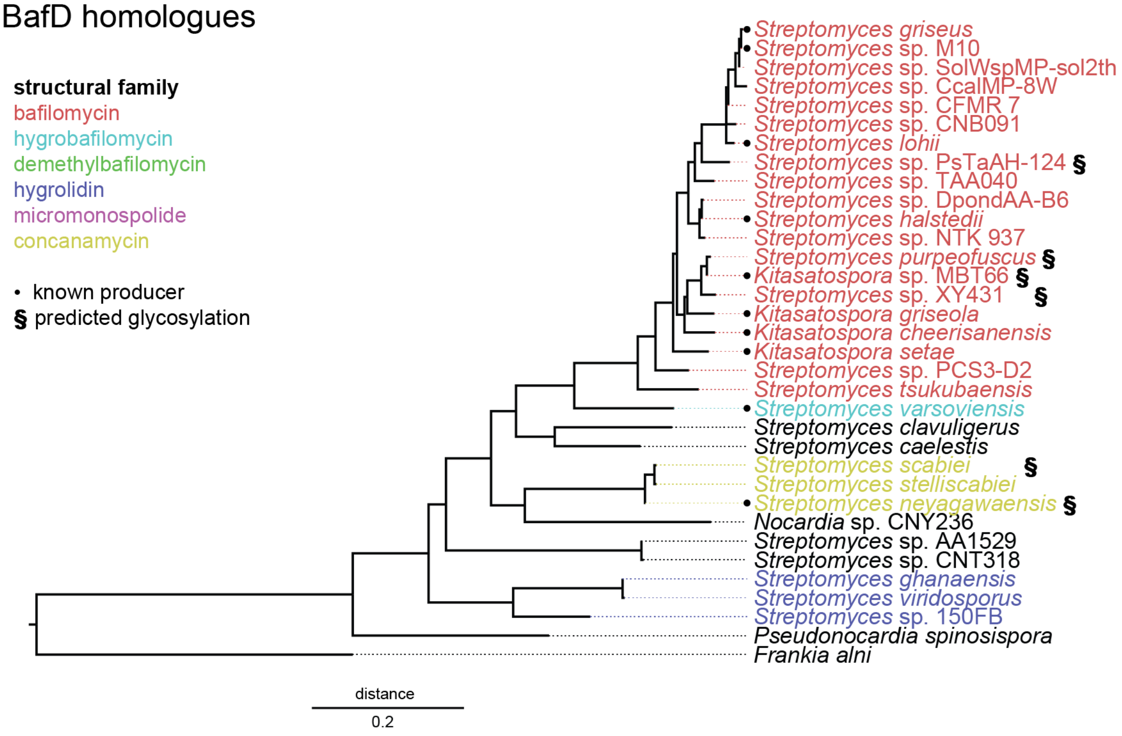


Figure S11. Phylogenetic tree of hygrolide producers. (A) A maximum-likelihood phylogenetic tree was generated using 16S rRNA gene sequences of known producing organisms for various hygrolides. For the genus *Micromonospora*, no 16S sequence corresponding to a confirmed producer was available, so *M. inyonensis* and *M. peucetia* were used as placeholders. For organisms listed in black, a compound class could not be predicted and novelty is suspected. The actinomycete *Frankia alni* was chosen as an outgroup and is not predicted to contain a hygrolide BGC. (B) Core scaffolds of the families shown in (A) are given. Within each family, variation is seen at various positions of the core. Features distinguishing the families from each other are shown in red.

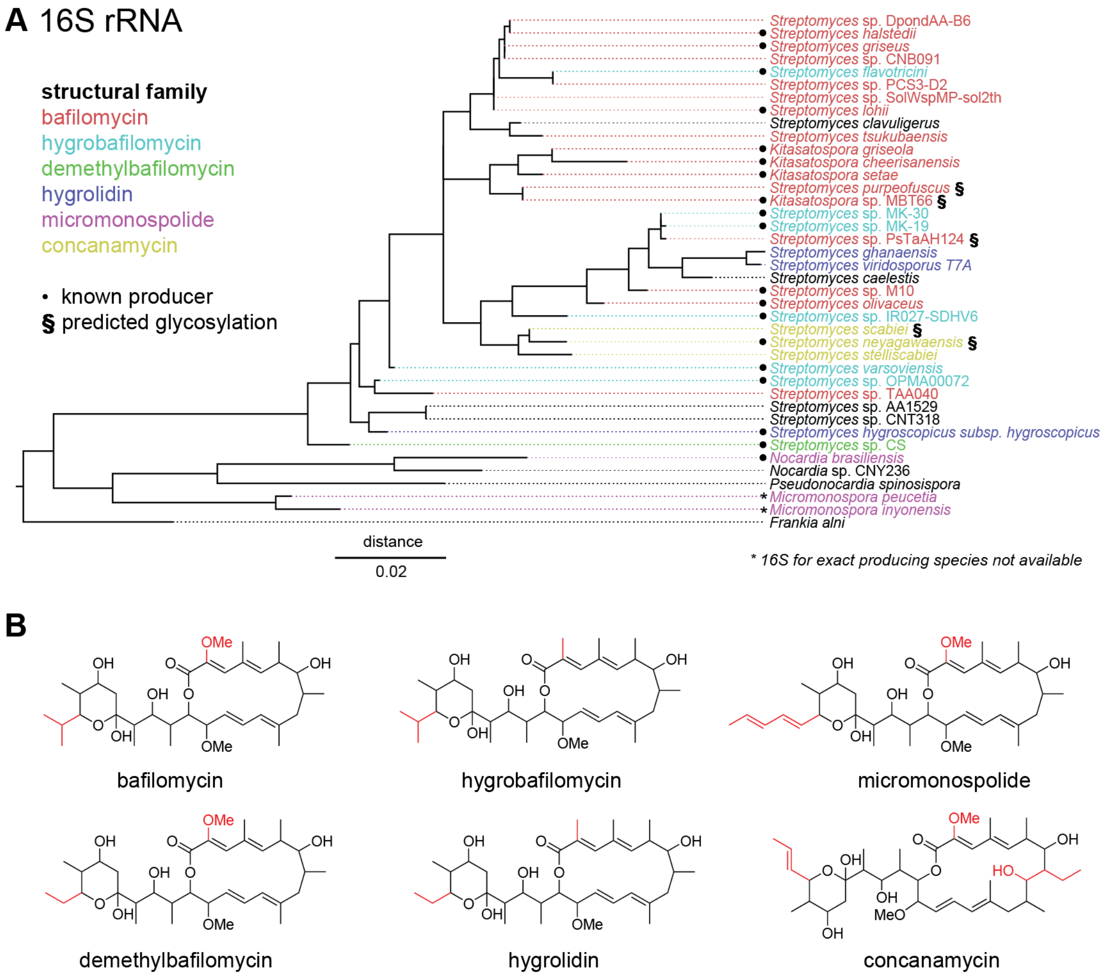


Table S3. Phylogenetic classification of known and predicted hygrolide BGCs. BafD and ORF3 homologues are given from each organism as GI numbers. The percent amino acid identity of each gene to a set of representative genes from organisms harboring bafilomycin, hygrobafilomycin, hygrolidin, or concanamycin BGCs was found using BLAST and is indicated by the depth of color shade. BGCs were classified by similarity. BGCs where a hygrolide compound has been isolated are indicated, along with the structural family of compounds. For several BGCs (bottom of the table), a definitive assignment could not be made due to a lack of strong relatedness to any of the model sequences. n.d., not determinable. Twenty-two BGCs have no known associated natural product(s). For BafD, where no homologue exists in the *S. hygroscopicus* BGC, the homologue from *S. ghanaensis* was used instead for comparison.

Homologue (GI number)			**Identity (% aa) to ORF3**					**Identity (% aa) to BafD**						
BafD	**ORF3**	**Organism**	***S. lohii***	***K. setae***	***S. varsovoviensis***	***S. hygroscopicus***	***S. neyagawaensis***	***S. lohii***	***K. setae***	***S. varsoviensis***	***S. ghanaensis***	***S. neyagawaensis***	**Scaffold Prediction**	**Known Natural Product Structural Family**
288965957	288965964	*Streptomyces lohii*	100	89	82	67	74	100	90	75	69	73	bafilomycin	bafilomycin
482661949	482661956	*Streptomyces griseus*	97	88	80	67	74	98	90	76	69	73	bafilomycin	bafilomycin
493388797	493388790	*Streptomyces tsukubaensis*	88	86	83	68	75	85	86	74	68	72	bafilomycin	–
517298471	517298478	*Streptomyces* sp. CcalMP-8W	97	89	80	66	74	97	90	73	68	72	bafilomycin	–
517394815	517394807	*Streptomyces* sp. PsTaAH-124	93	88	81	66	74	92	91	74	68	71	bafilomycin	–
517791715	517791722	*Streptomyces* sp. CNB091	96	88	79	66	74	98	90	73	67	72	bafilomycin	–
654969407	654969414	*Streptomyces* sp. SolWspMP-sol2th	97	89	80	67	74	98	90	73	68	72	bafilomycin	–
654979441	654979377	*Streptomyces* sp. TAA040	94	88	80	66	74	93	91	74	68	72	bafilomycin	–
654993951	n.a.	*Streptomyces* sp. DpondAA-B6	-	-	-	-	-	94	92	74	70	71	bafilomycin	–
663401407	n.a.	*Streptomyces purpeofuscus*	-	-	-	-	-	91	92	74	71	74	bafilomycin	–
664093809	654993958	*Streptomyces halstedii*	94	90	82	67	75	94	93	74	70	72	bafilomycin	bafilomycin
726969199	726969206	*Streptomyces* sp. M10	97	88	80	67	74	98	90	76	69	73	bafilomycin	bafilomycin
737891878	737891867	*Kitasatospora cheerisanensis*	90	93	80	68	73	90	92	74	71	74	bafilomycin	bafilomycin
740031905	654993958	*Streptomyces* sp. NTK 937	94	90	82	67	75	93	93	74	70	72	bafilomycin	–
740068110	740068128	*Streptomyces* sp. PCS3-D2	89	86	84	67	76	88	90	75	71	73	bafilomycin	–
759764416	n.a.	*Kitasatospora* sp. MBT66	-	-	-	-	-	91	92	74	71	74	bafilomycin	bafilomycin
763023462	763023454	*Kitasatospora griseola*	90	92	79	67	72	91	93	74	70	73	bafilomycin	bafilomycin
799307788	n.a.	*Streptomyces* sp. XY431	-	-	-	-	-	91	91	74	71	74	bafilomycin	–
916321000	503906410	*Kitasatospora setae*	89	100	79	67	73	90	100	75	69	73	bafilomycin	bafilomycin
924906653	924906647	*Streptomyces* sp. CFMR 7	98	89	80	66	74	99	91	73	68	72	bafilomycin	–
502770800	n.a.	*Streptomyces scabiei*	-	-	-	-	-	73	72	69	67	98	concanamycin	–
822059630	822063703	*Streptomyces stelliscabiei*	76	76	74	69	93	73	72	69	67	98	concanamycin	–
943900294	943905490	*Streptomyces neyagawaensis*	74	75	73	69	100	73	72	69	68	100	concanamycin	concanamycin
n.a.	889609429	*Streptomyces varsoviensis*	82	81	100	65	73	75	72	100	64	67	hygrobafilomycin	hygrobafilomycin
491120173	491120212	*Streptomyces ghanaensis*	67	67	66	96	68	69	69	65	100	68	hygrolidin	hygrolidin
515242709	515242729	*Streptomyces viridosporus T7A*	67	67	66	96	68	69	69	65	100	68	hygrolidin	–
748767887	748767878	*Streptomyces* sp. 150FB	68	67	66	86	69	68	68	68	79	70	hygrolidin	–
n.a.	466860673	*Streptomyces hygroscopicus*	67	67	65	100	69	-	-	-	-	-	hygrolidin	hygrolidin
655029786	655029979	*Nocardia sp. CNY236*	71	72	68	69	75	71	70	65	66	70	n.d.	–
490056513	490056911	*Streptomyces clavuligerus*	70	72	74	64	68	70	71	70	64	71	n.d.	–
739855312	n.a.	*Streptomyces* sp. AA1529	-	-	-	-	-	67	67	64	66	69	n.d.	–
739934956	n.a.	*Streptomyces* sp. CNT318	-	-	-	-	-	67	67	64	66	69	n.d.	–
916734751	n.a.	*Pseudonocardia spinosispora*	-	-	-	-	-	64	64	62	63	64	n.d.	–
925291620	917236896	*Streptomyces caelestis*	71	71	72	65	68	74	75	73	67	73	n.d.	–

References

(1) Kitsche, A.; Kalesse, M. "Configurational assignment of secondary hydroxyl groups and methyl branches in polyketide natural products through bioinformatic analysis of the ketoreductase domain" *ChemBioChem* **2013**, *14*, 851.