

## **Supporting Information**

### **Evolution of chemical diversity by coordinated gene swaps in type II polyketide gene clusters**

Maureen E. Hillenmeyer<sup>1</sup>, Gergana A. Vandova<sup>1,2</sup>, Erin E. Berlew<sup>3</sup>, Louise K. Charkoudian<sup>3</sup>

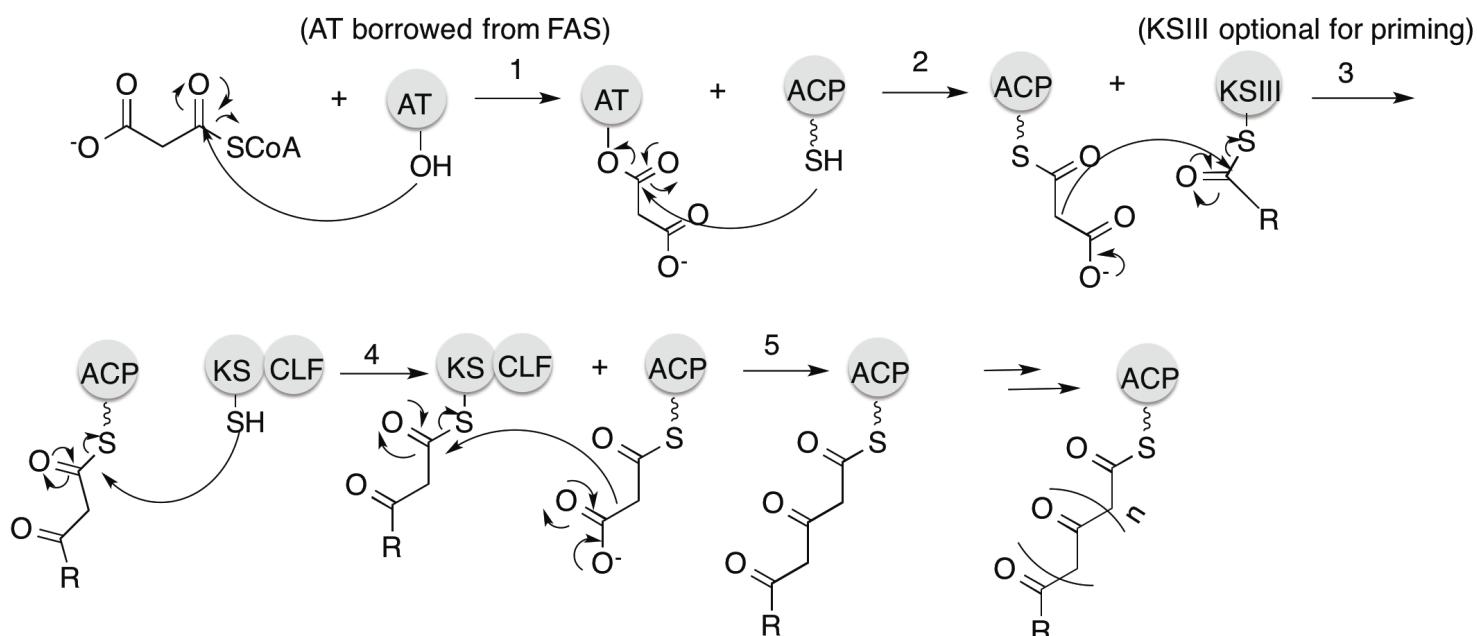
<sup>1</sup> Stanford Genome Technology Center, Stanford University, Palo Alto CA; <sup>2</sup> Department of Biochemistry, Stanford University, Palo Alto CA; <sup>3</sup> Department of Chemistry, Haverford College, Haverford PA.

Page 2 ..... Supplementary Figures

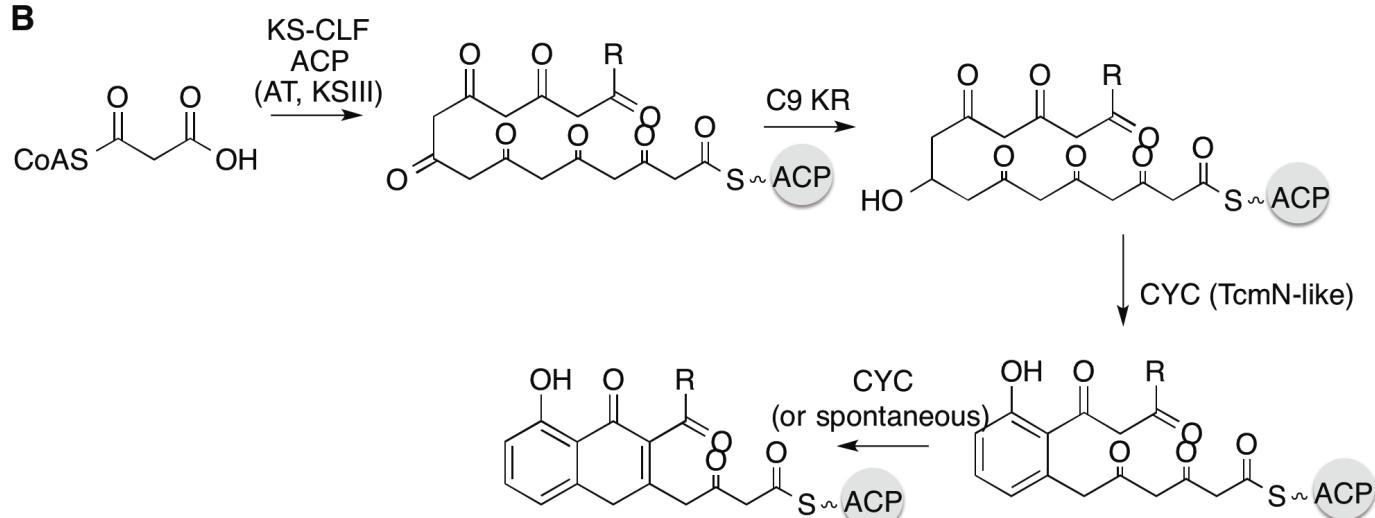
Page 20 ... .... Materials and Methods

Page 26... .... Supplementary Table S5: Reference Set of Curated Type II Polyketide Synthase Gene Clusters

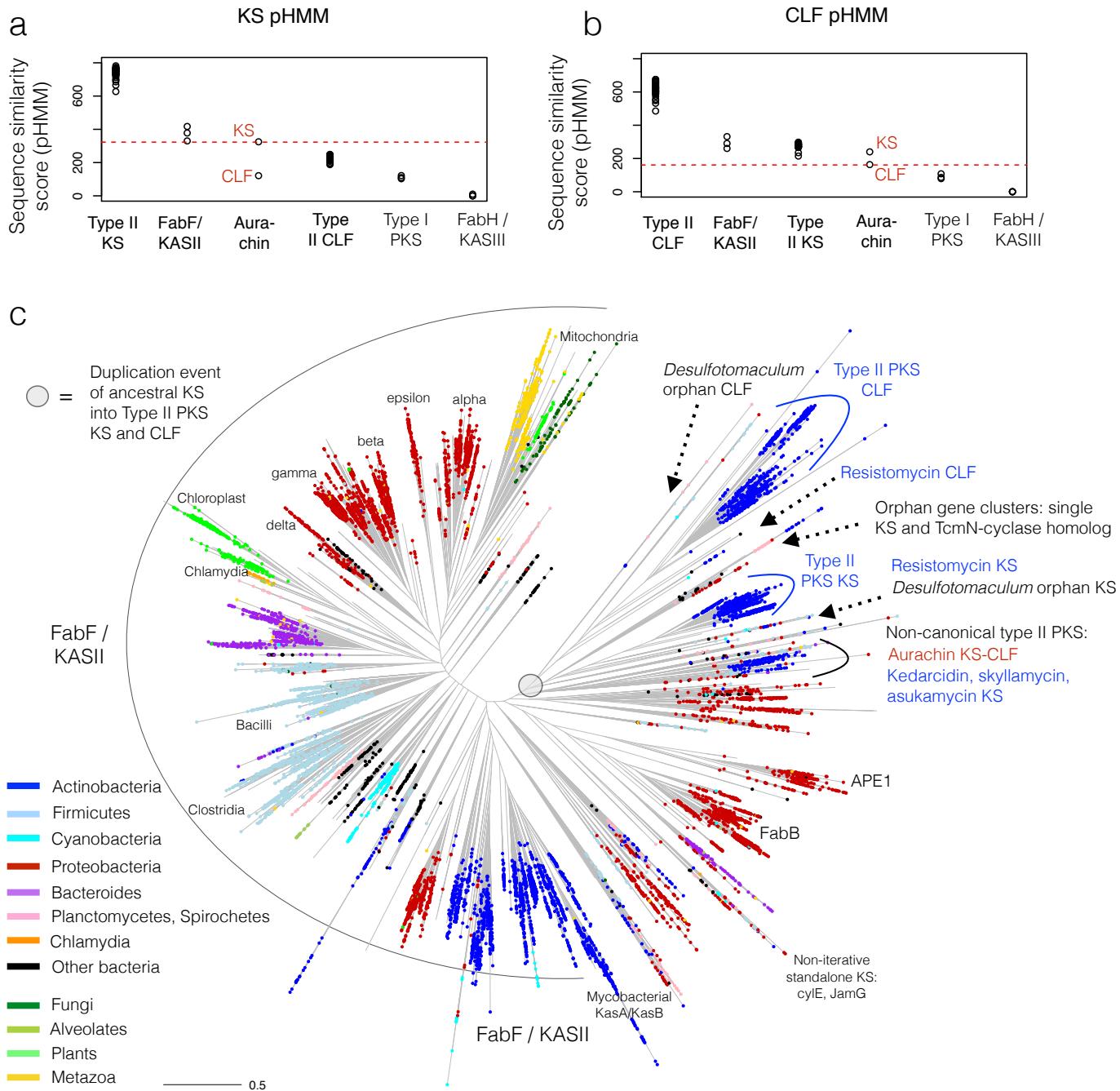
A



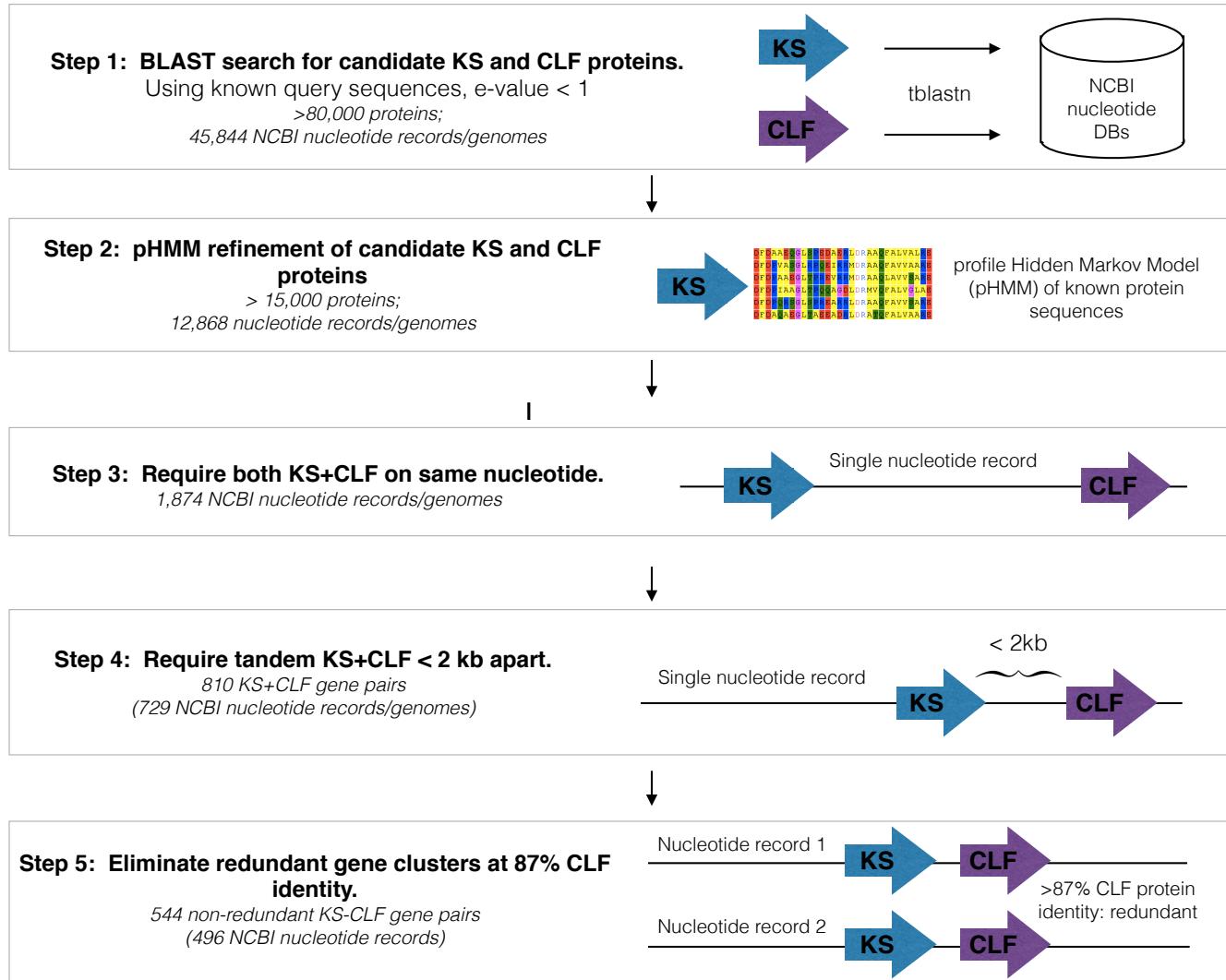
B



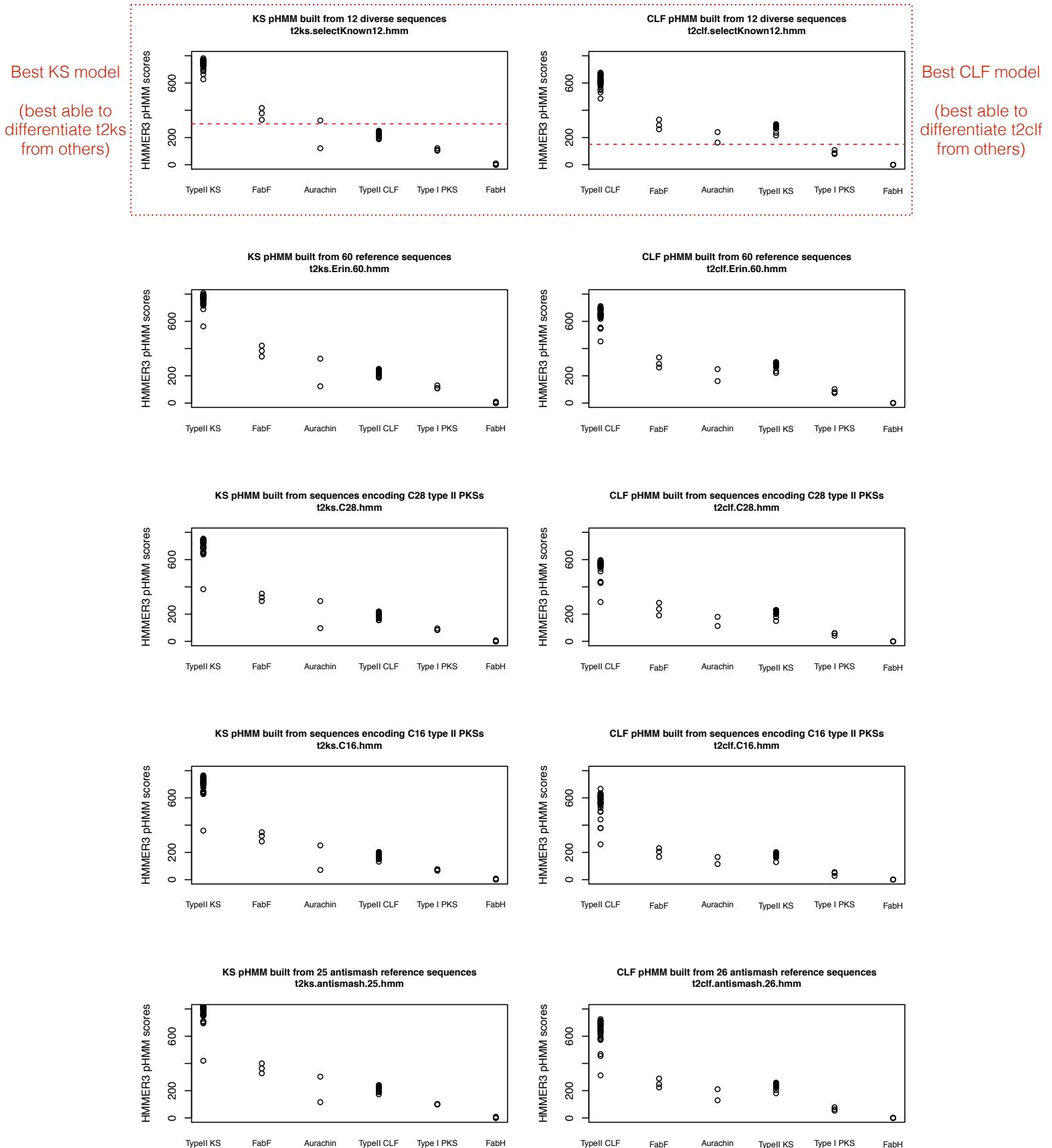
**Figure S1.** A. Basic mechanisms involved in type II polyketide biosynthesis. The AT (likely borrowed from FAS) selects the malonyl CoA building block (Step 1) and transfers it to the ACP (Step 2). In the presence of the initiation module, the starter unit is presented by the KSIII and undergoes a decarboxylative Claisen condensation with malonyl ACP (Step 3). In the absence of the initiation module, malonyl-ACP reacts with the acetyl-primed KS of the KS-CLF. The polyketide chain is then transferred to the KS-CLF via a transacylation reaction (Step 4). Decarboxylative Claisen condensation reactions with malonyl-ACP (Step 5) followed by transacylation of the polyketide intermediate to the KS-CLF repeats until the programmed chain length (determined in part by the cavity volume of the KS-CLF) is reached. B. Role of commonly found accessory enzymes in polyketide biosynthesis. The nascent polyketide chain can be reduced stereo- and regio-specifically by the action of ketoreductase (KR) enzymes and/or cyclized by the action of region-specific cyclases.



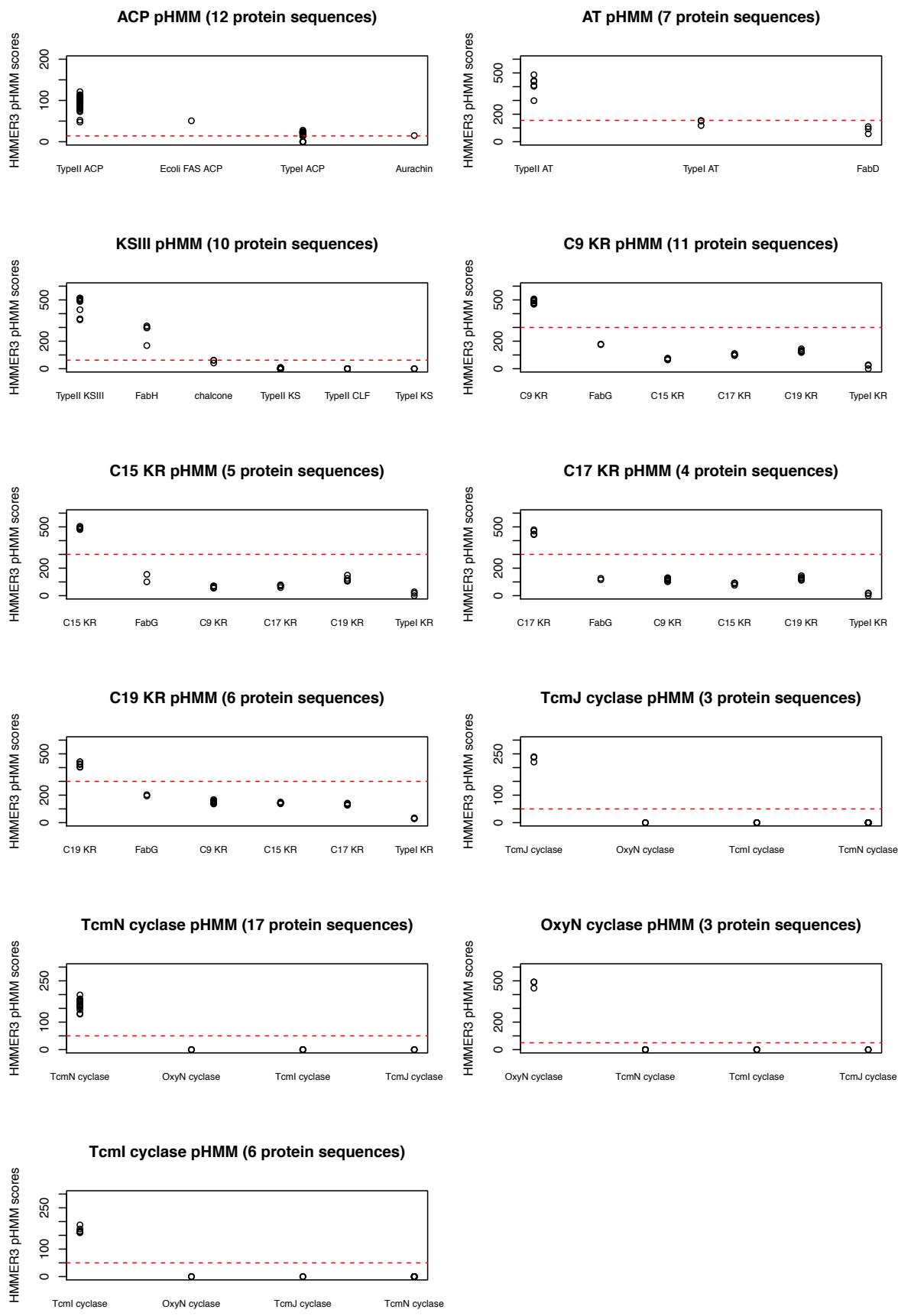
**Figure S2.** Homologs of type II PKS KS and CLF proteins. (a) and (b) Similarity of type II PKS KS and CLF proteins to homologs, measured by profile hidden Markov model (pHMM) scores. (c) Phylogeny of proteins scoring above KS and CLF pHMM score thresholds.



**Figure S3.** Search pipeline to identify type II PKS gene clusters.

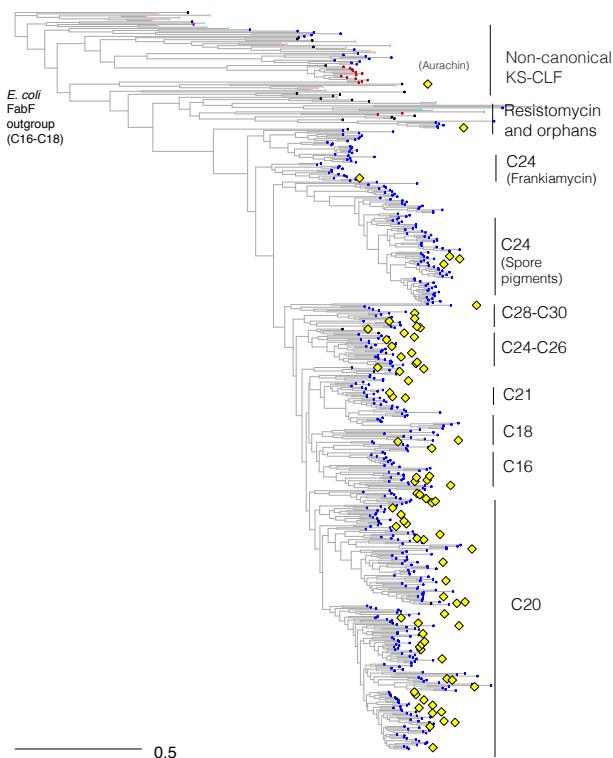


**Figure S4A.** Scores of various KS and CLF pHMMS when applied to protein sequences from different classes of ketosynthases. Each plot represents a different pHMMS scored against various test sets, using the program HMMER3. The top model was selected for its ability to best differentiate the true Type II KS (or CLF) from the nearest homologs (FabF). The Aurachin KS and CLF scores were selected as thresholds for catalog generation in this study.

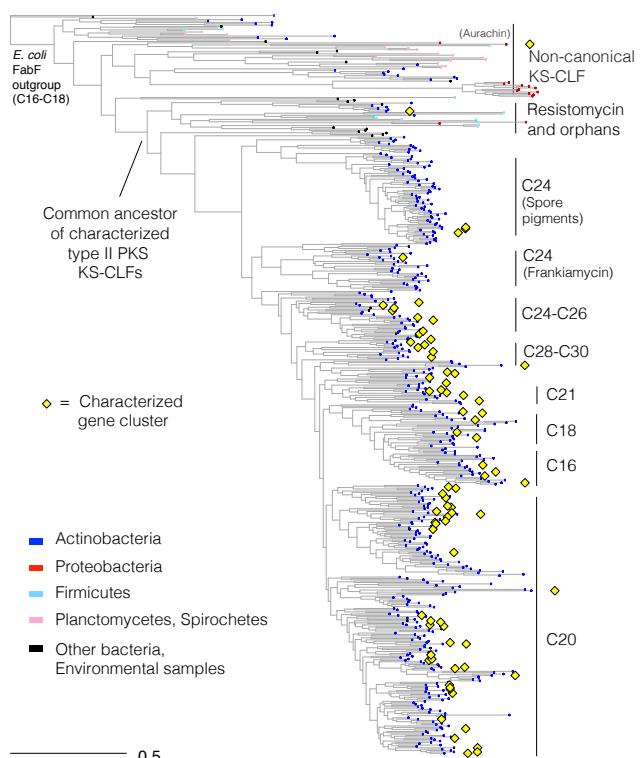


**Figure S4B.** Sequence similarity scores for accessory enzyme pHMM sequence models tested on various test sets.

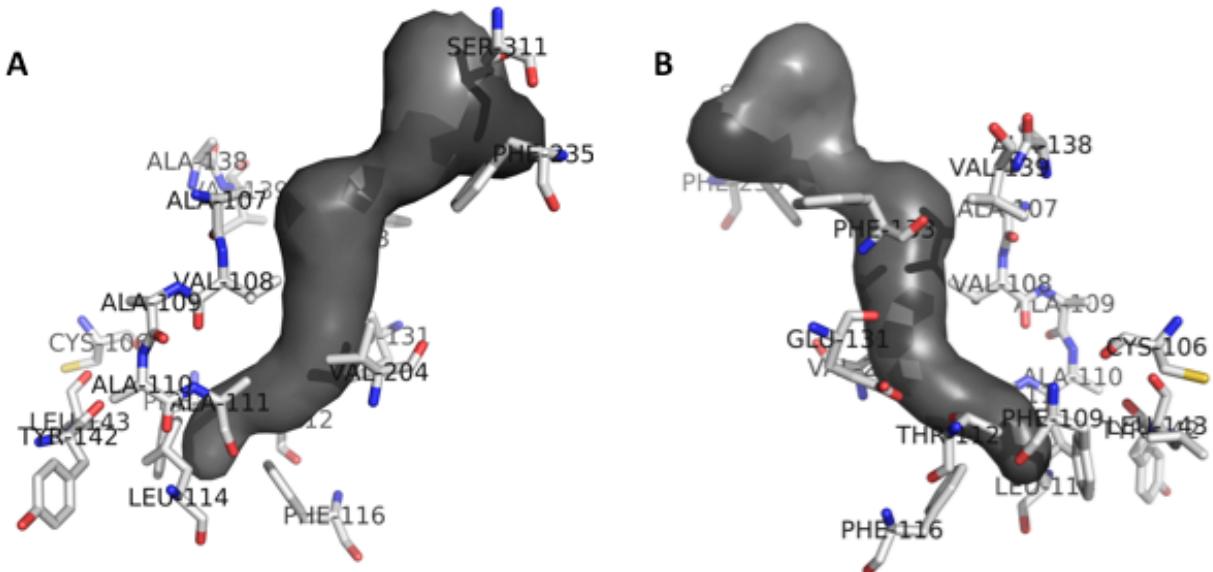
## KS Phylogeny



## CLF Phylogeny

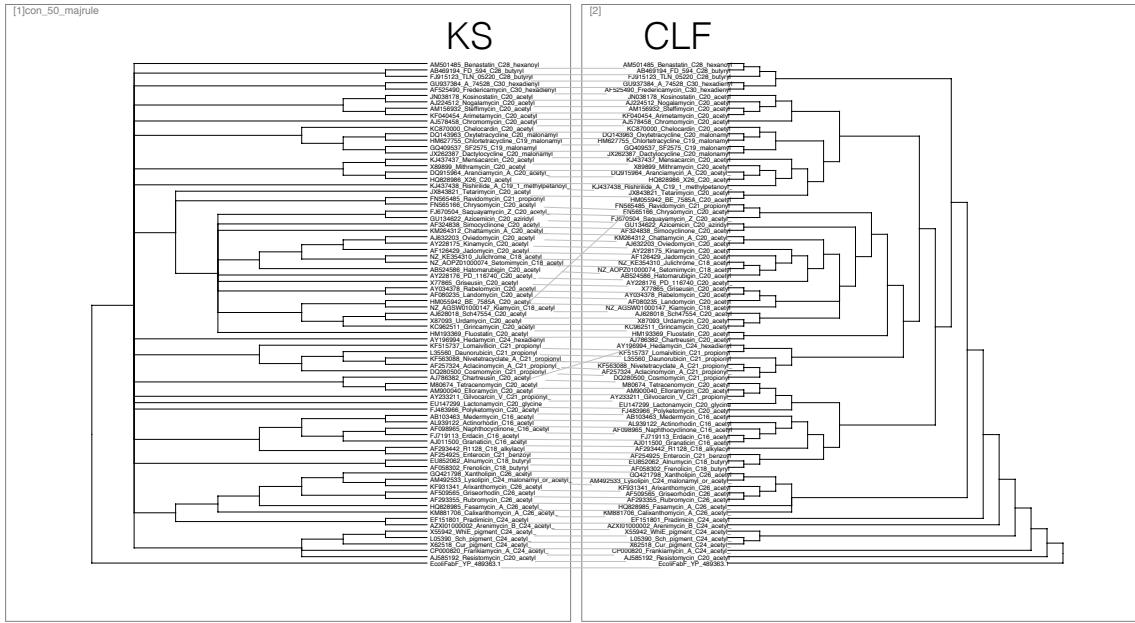


**Figure S5.** Approximate maximum likelihood phylogenetic trees of KS and CLF proteins from 544 putative type II PKS gene clusters. Multiple sequence alignment was performed using MAFFT (1), and trees were generated using FastTree2 (2).

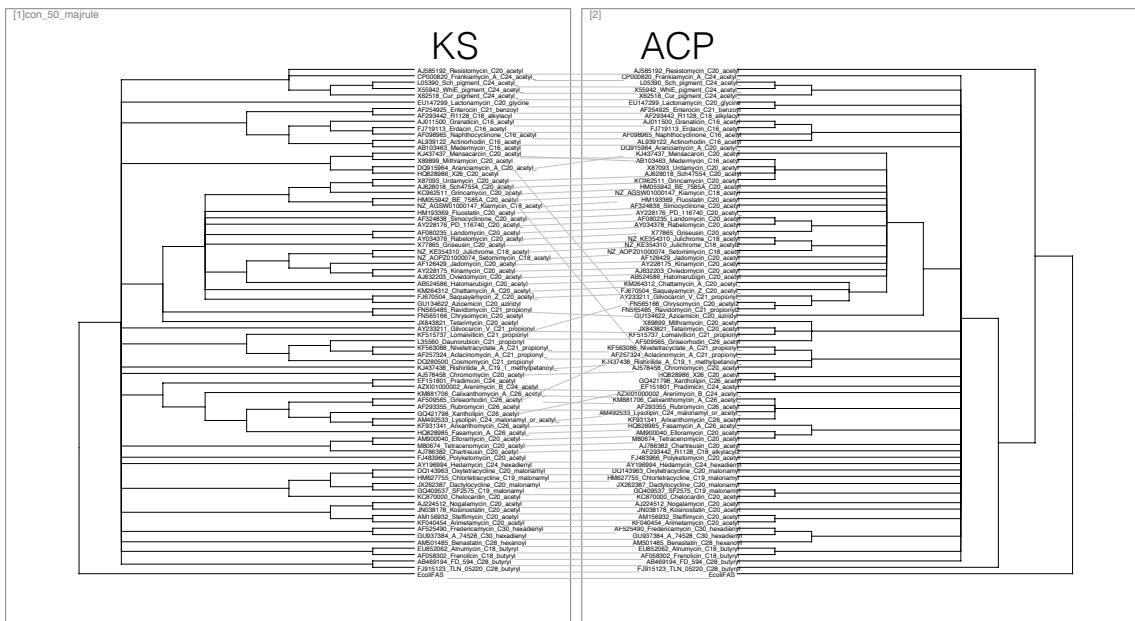


**Figure S6.** The side chains of cavity-lining residues(CLRS) deemed to be of interest (see Methods) are shown along the actinorhodin KS-CLF structure (PDB ID 1TQY). The volume of the cavity in which the nascent polyketide sides (grey) was determined using CASTp as 706.2 Å. Mutations were made in PyMol to represent each consensus sequence of the 5 categories of chain length and the cavity volume for each mutant was CASTp.

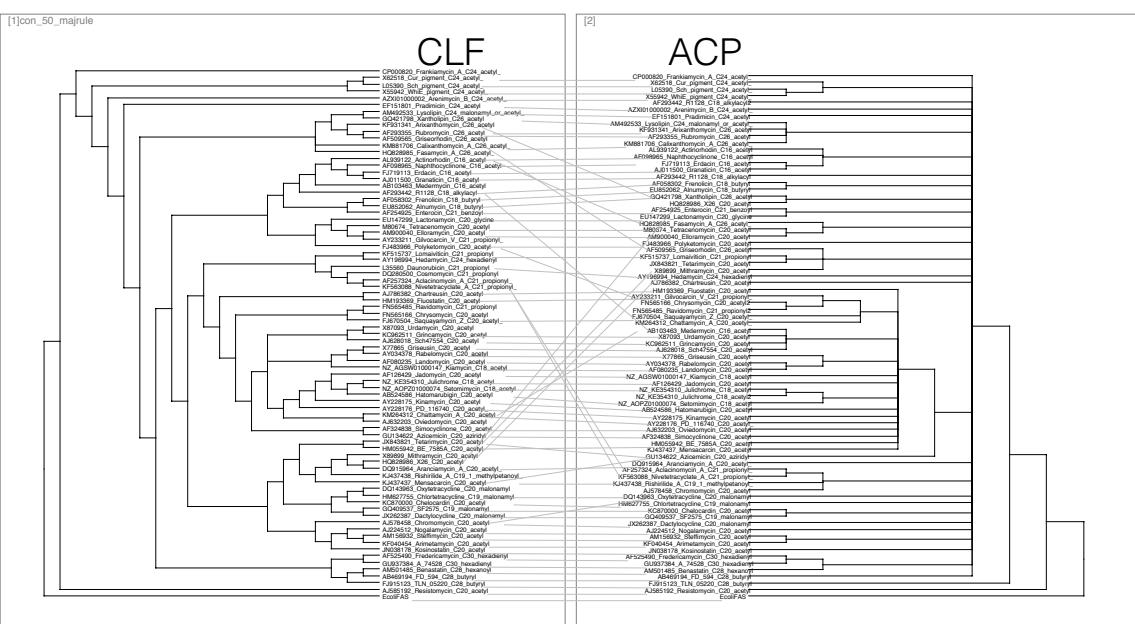
a



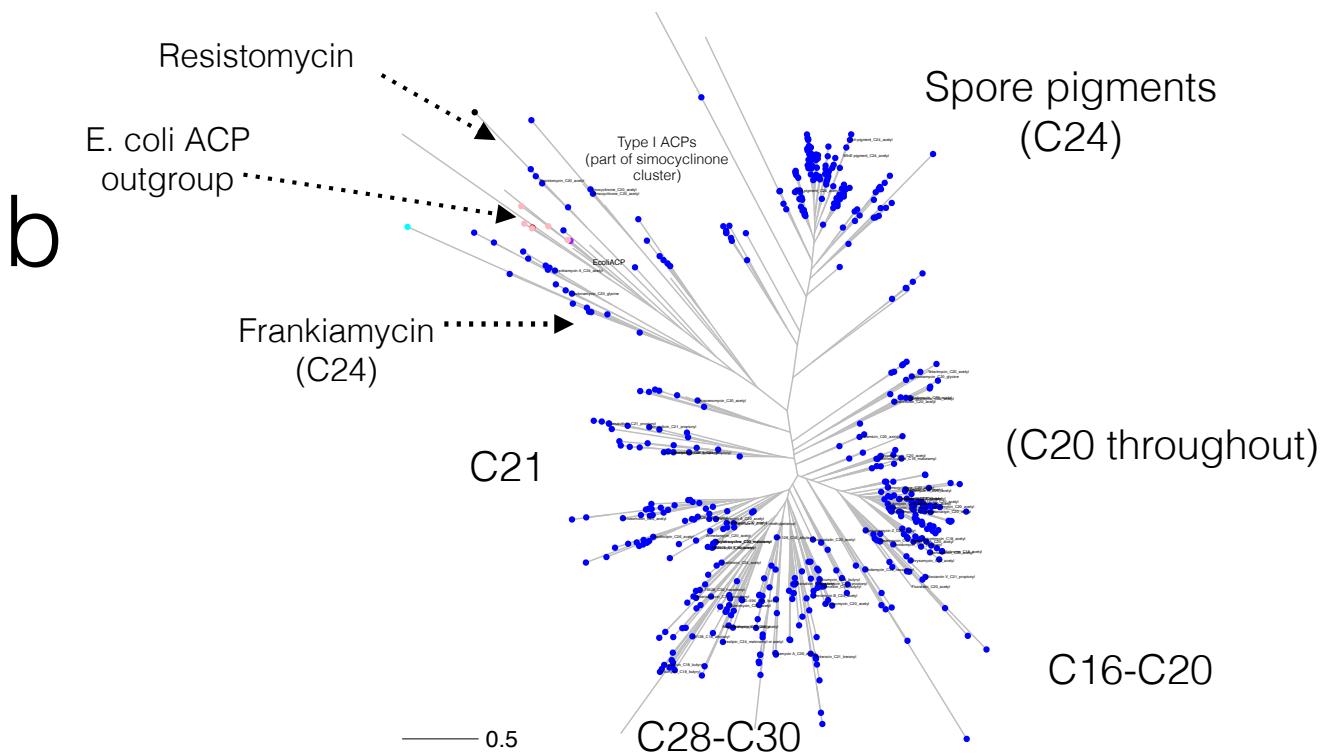
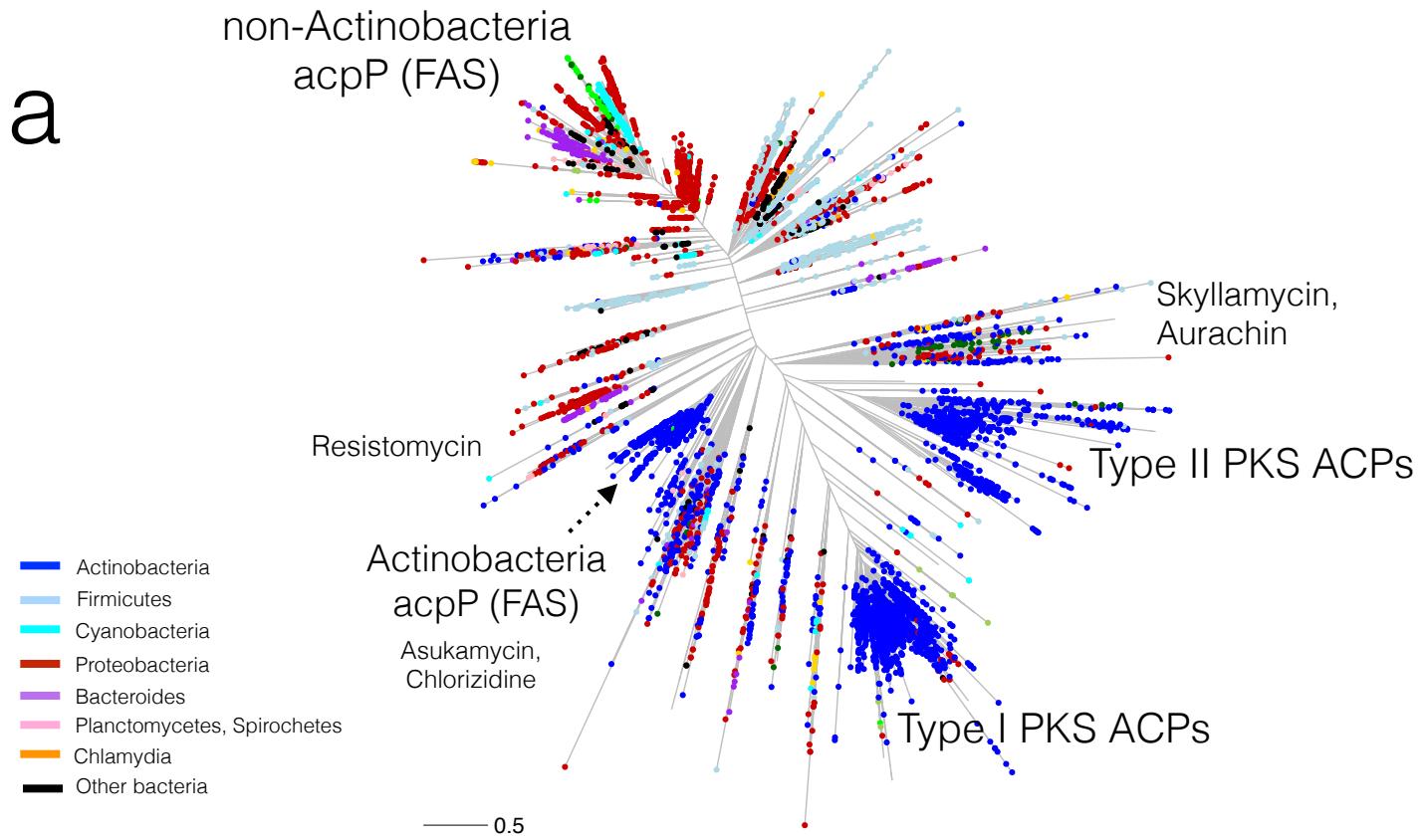
b



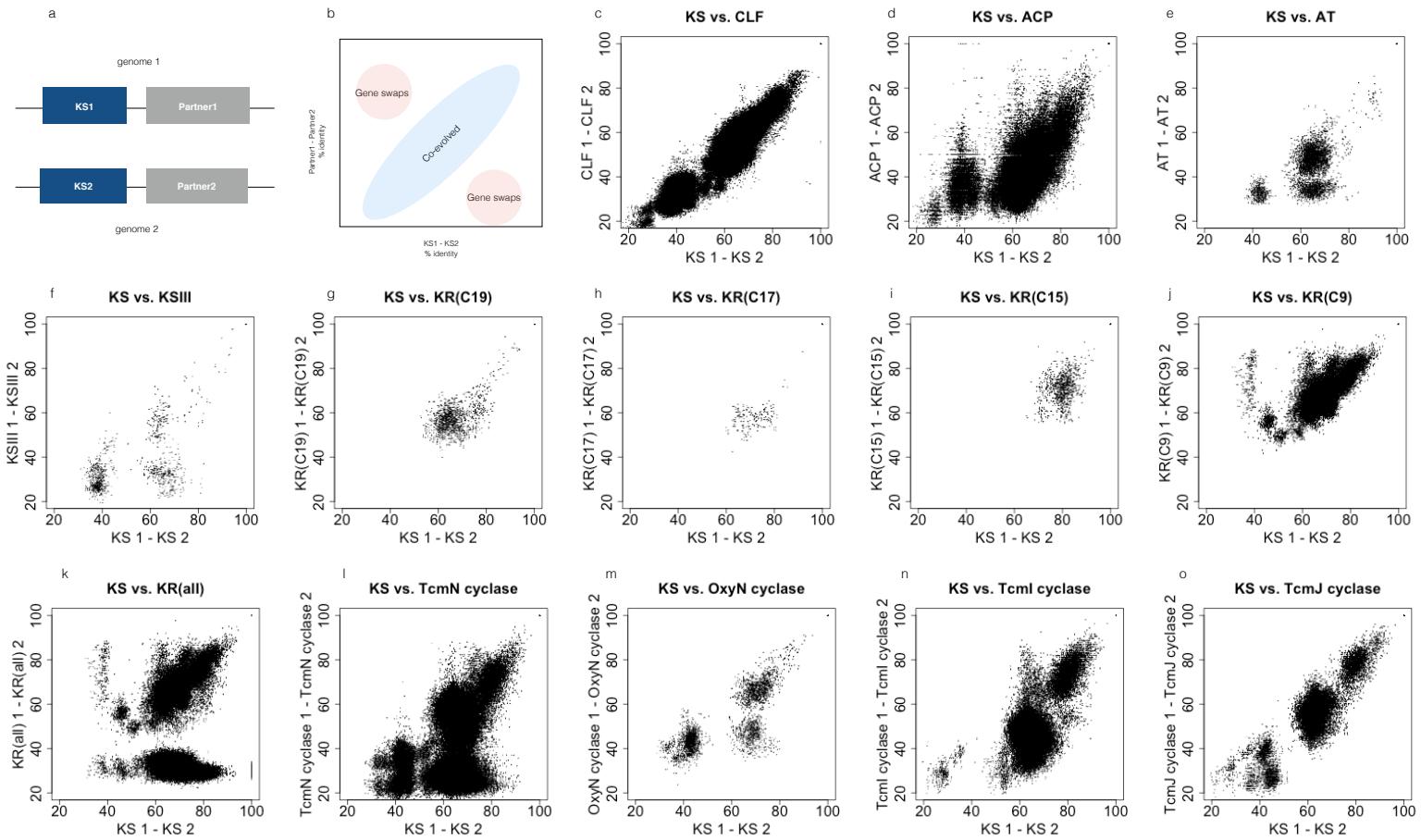
c



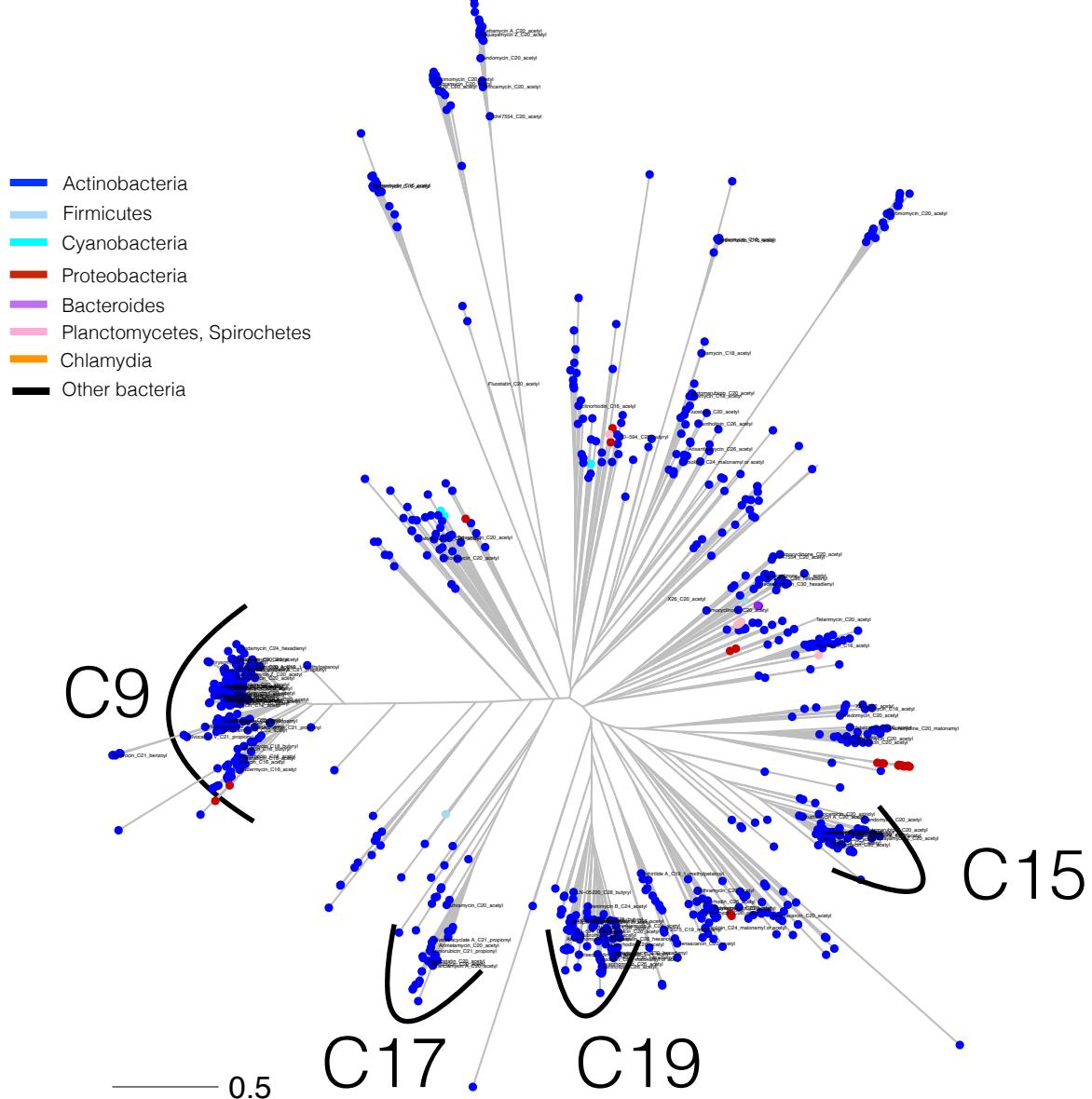
**Figure S7.** Comparison of KS, CLF, and ACP phylogenies by tanglegram analysis of proteins from characterized type II PKS gene clusters. Lines between genes from the same gene cluster are drawn between trees. The KS and ACP phylogenies are not well-resolved due to low sequence diversity, leading to a large polytomy (unresolved clades comprising more than two descending branches). Trees were generated using MUSCLE multiple sequence alignment and MrBayes Bayesian tree inference. Tanglegram was generated using Dendroscope.



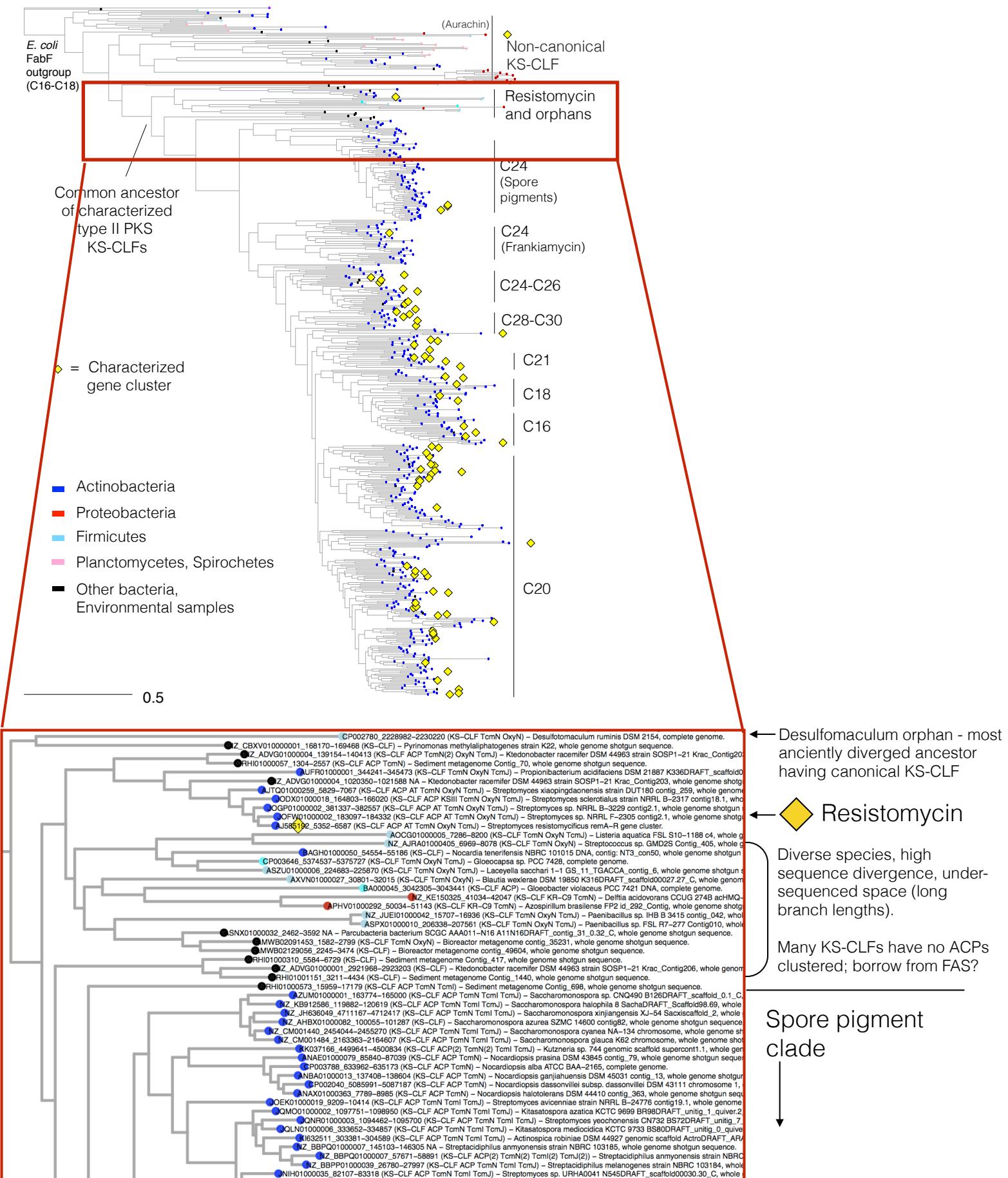
**Figure S8.** (a) Phylogeny of all ACP protein homologs with (HMMER score > 14, see Methods). The type II PKS ACP proteins are highlighted. The other clades harbor proteins whose genes are not found clustered with KS-CLF genes.  
(b) Phylogeny of subset of ACP homologs that are found clustered with type II PKS tandem KS-CLF genes.



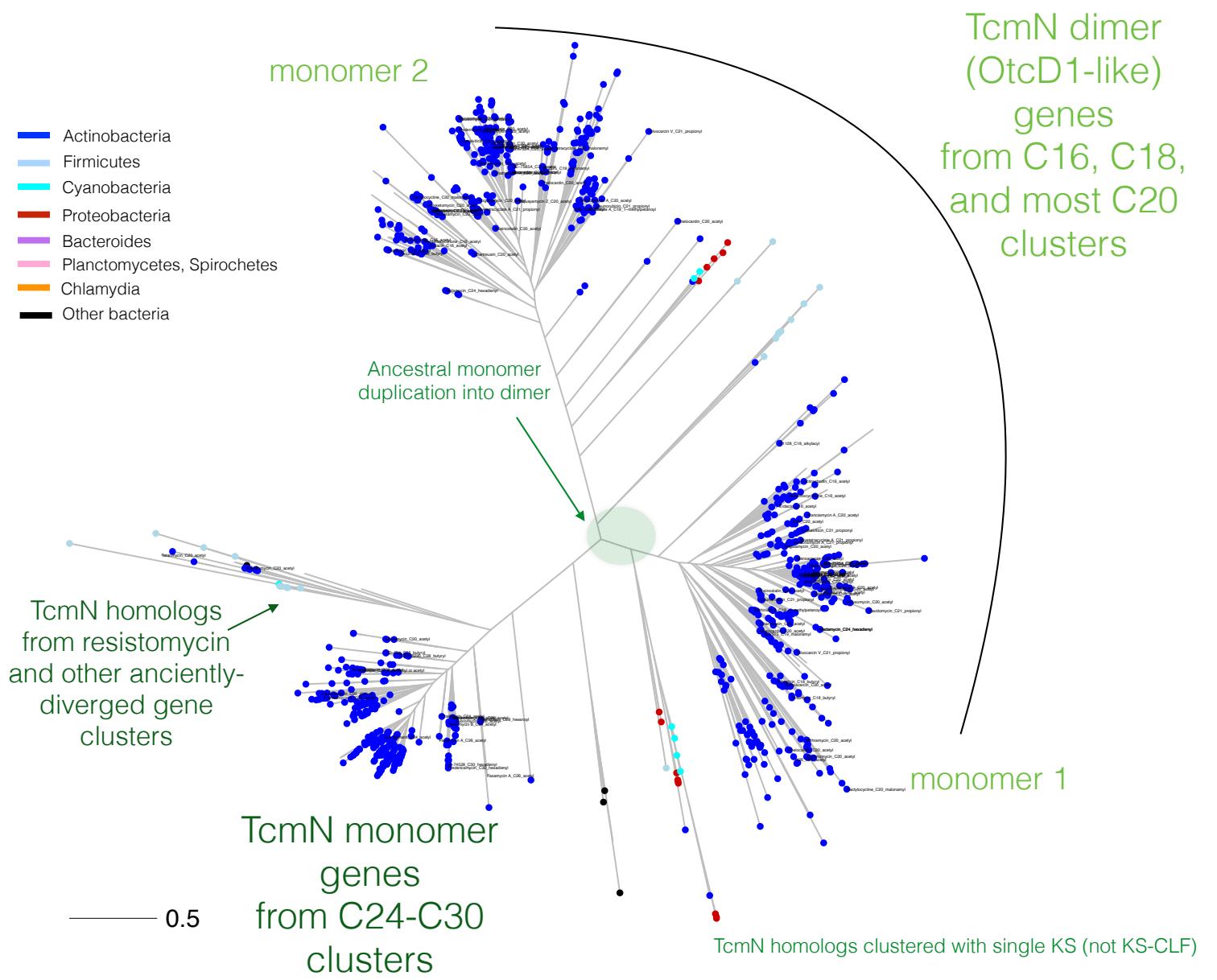
**Figure S9.** Co-evolution of type II PKS KS with partner genes. Schematic illustrating two nucleotide records and the clustered (within 30kb) KS+Partner on each record. (b) KS1-KS2 pairwise amino acid identities are plotted vs. pairwise identities of a clustered partner. (c-o) Correlation of evolutionary histories of the KS with each partner gene from Table 2.



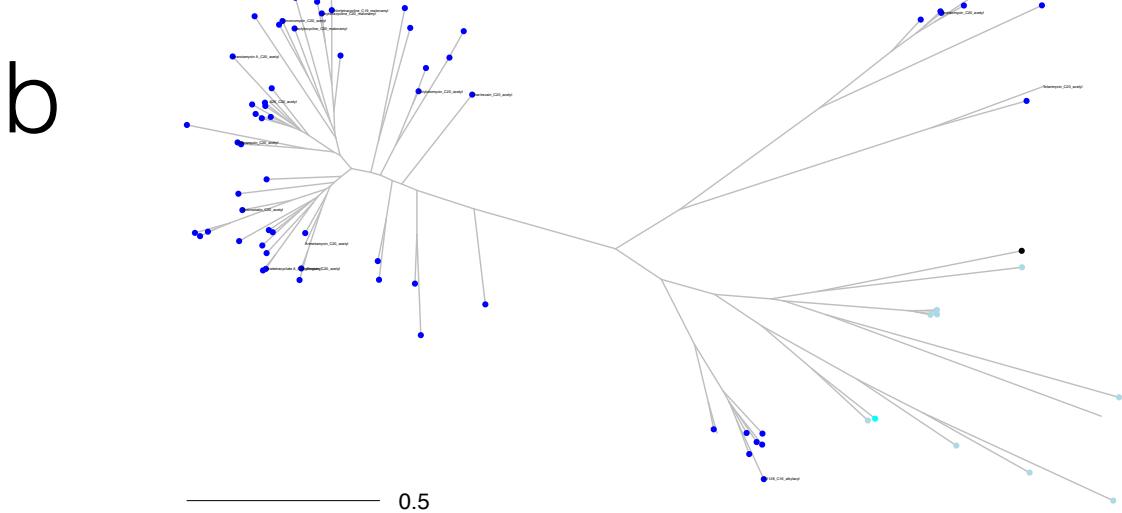
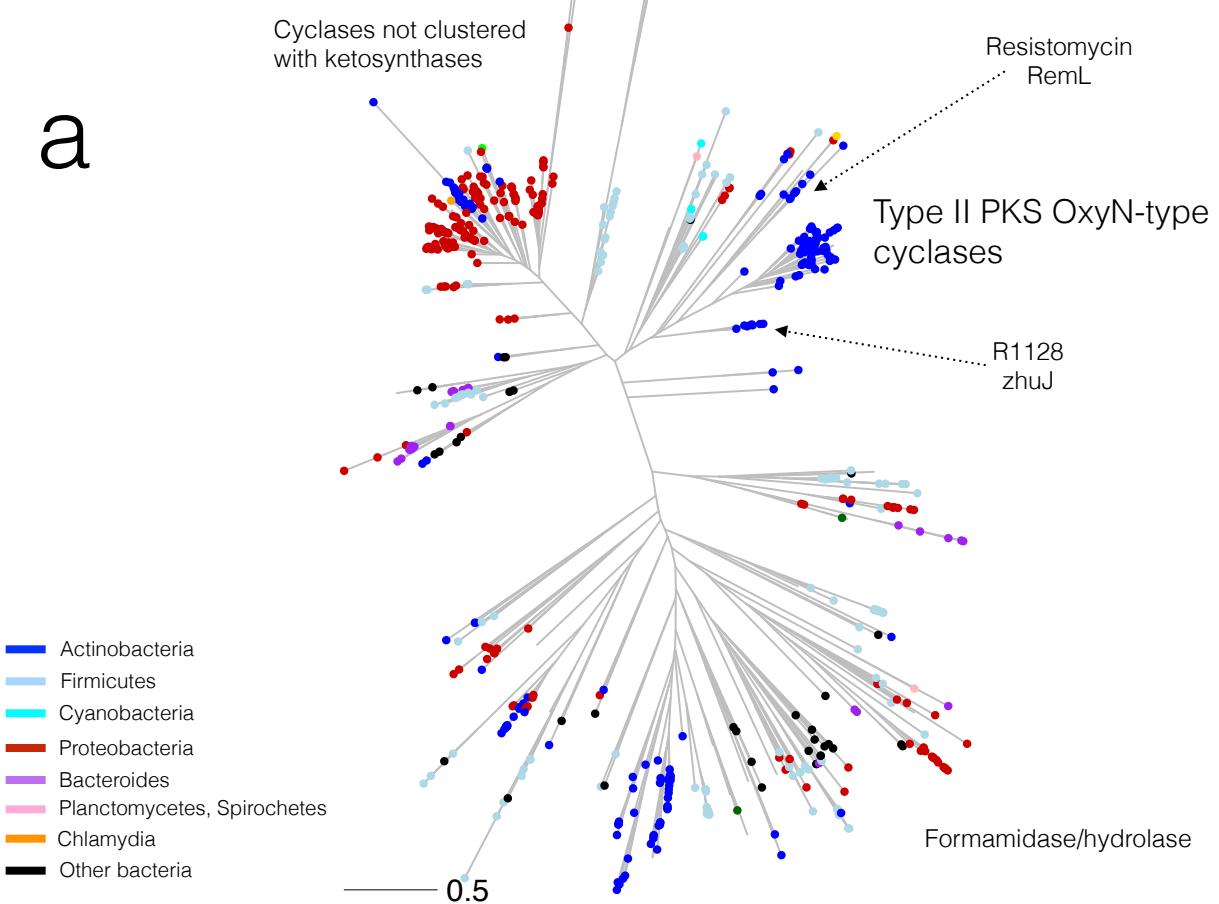
**Figure S10.** Phylogeny of KR homologs found clustered with type II KS-CLF genes. KR sequences were identified by a broad HMM KR search for diverse KR types. The C9, C17, C19, and C15 KR types form well-supported clades on the larger phylogeny. Contrary to phylogenies of other accessory enzymes ACP and cyclases, this phylogeny is limited to KRs clustered with a KS-CLF, in order to focus on the 4 classes highlighted. The phylogeny of all KRs (including those not clustered with a tandem KS-CLF) is much larger.



**Figure S11.** Phylogeny of CLF sequences (same as Figure 1 in main text), with the region around resistomycin enlarged at bottom. Sequences from diverse phyla are similar to the resistomycin CLF. The region is relatively poorly sequenced, as compared with sequences from spore pigment and antibiotic clusters.



**Figure S12.** Phylogeny of TcmN protein homologs (HMMER score > 50, see Methods).



**Figure S13.** (a) Phylogeny of all OxyN protein homologs with (HMMER score > 50, see Methods). The type II PKS cyclase proteins are highlighted. The other clades harbor proteins whose genes are not found clustered with KS-CLF genes. (b) Phylogeny of subset of OxyN homologs that are found clustered with type II PKS tandem KS-CLF genes.

**a**

	KS-CLF	ACP	KR	AT	KSIII	TcmN	Tcml	TcmJ	OxyN
KS-CLF	544	421	353	81	39	418	250	174	69
ACP	421	-	302	73	28	376	234	149	57
KR	353	302	-	69	36	299	170	82	47
AT	81	73	69	-	13	69	28	18	18
KSIII	39	28	36	13	-	25	17	11	3
TcmN	418	376	299	69	25	-	238	169	65
Tcml	250	234	170	28	17	238	-	152	0
TcmJ	174	149	82	18	11	169	152	-	12
OxyN	69	57	47	18	3	65	0	12	-

Observed number of gene clusters (out of 544 nonredundant) in which the listed pairs of gene co-occur within a cluster.

**b**

	KS-CLF	ACP	KR	AT	KSIII	TcmN	Tcml	TcmJ	OxyN
KS-CLF	100.0%	77.4%	64.9%	14.9%	7.2%	76.8%	46.0%	32.0%	12.7%
ACP	77.4%	-	55.5%	13.4%	5.1%	69.1%	43.0%	27.4%	10.5%
KR	64.9%	55.5%	-	12.7%	6.6%	55.0%	31.3%	15.1%	8.6%
AT	14.9%	13.4%	12.7%	-	2.4%	12.7%	5.1%	3.3%	3.3%
KSIII	7.2%	5.1%	6.6%	2.4%	-	4.6%	3.1%	2.0%	0.6%
TcmN	76.8%	69.1%	55.0%	12.7%	4.6%	-	43.8%	31.1%	11.9%
Tcml	46.0%	43.0%	31.3%	5.1%	3.1%	43.8%	-	27.9%	0.0%
TcmJ	32.0%	27.4%	15.1%	3.3%	2.0%	31.1%	27.9%	-	2.2%
OxyN	12.7%	10.5%	8.6%	3.3%	0.6%	11.9%	0.0%	2.2%	-

Observed fraction of gene clusters (out of 544 nonredundant) in which the gene pairs co-occur.

**c**

	KS-CLF	ACP	KR	AT	KSIII	TcmN	Tcml	TcmJ	OxyN
KS-CLF	100.0%	77.4%	64.9%	14.9%	7.2%	76.8%	46.0%	32.0%	12.7%
ACP	77.4%	-	50.2%	11.5%	5.5%	59.5%	35.6%	24.8%	9.8%
KR	64.9%	50.2%	-	9.7%	4.7%	49.9%	29.8%	20.8%	8.2%
AT	14.9%	11.5%	9.7%	-	1.1%	11.4%	6.8%	4.8%	1.9%
KSIII	7.2%	5.5%	4.7%	1.1%	-	5.5%	3.3%	2.3%	0.9%
TcmN	76.8%	59.5%	49.9%	11.4%	5.5%	-	35.3%	24.6%	9.7%
Tcml	46.0%	35.6%	29.8%	6.8%	3.3%	35.3%	-	14.7%	5.8%
TcmJ	32.0%	24.8%	20.8%	4.8%	2.3%	24.6%	14.7%	-	4.1%
OxyN	12.7%	9.8%	8.2%	1.9%	0.9%	9.7%	5.8%	4.1%	-

Expected fraction of gene clusters (out of 544 nonredundant) in which the gene pairs co-occur. This is calculated as the product of frequency of the first gene multiplied by the frequency of the second gene. E.g., KRs are observed in 65% of gene clusters, while ACPs are observed in 77% of gene clusters, so the expected frequency of clusters having both KR and ACP is  $.65 \times .77 = .50$

**d**

The ratio of the *observed vs. expected* co-occurrences of gene pairs in type II PKS gene clusters. Green represents pairs of genes that co-occur more frequently than expected by chance, and Red represents gene pairs that co-occur less frequently than expected by chance.

**Figure S14.** Co-occurrence of pairs of genes in 544 putative type II PKS gene clusters.

**Table S1.** List of all KS-CLF residues that are not conserved across all chain length categories.\*

Residue #	Chain length category				Residue #	Chain length category				Residue #	Chain length category				Residue #	Chain length category				
	C16-18	C20	C24	C26		C16-18	C20	C24	C26		C16-18	C20	C24	C26	C28-30	C16-18	C20	C24	C26	C28-30
8 D 100.000	D 100.000	D 100.000	D 100.000	DM 50.000	D 100.000	88 D 60.000	D 69.565	D 57.143	E 62.500	N 50.000	182 T 70.000	Q 56.522	A 42.857	L 87.500	V 66.667	284 A 100.000	A 100.000	A 100.000	AT 37.500	A 83.333
9 K 50.000	T 44.444	IK 50.000	KT 50.000	K 66.667	89 W 50.000	W 63.043	L 57.143	L 37.500	M 50.000	183 IL 40.000	I 58.696	V 42.857	IL 37.500	L 66.667	285 A 70.000	I 54.348	I 57.143	I 50.000	AI 50.000	
10 R 50.000	PRS 20.000	R 66.667	RST 33.333	GST 33.333	95 G 40.000	G 69.565	GKR 28.571	DK 25.000	A 33.333	185 R 60.000	K 45.652	R 57.143	T 37.500	KR 33.333	286 R 50.000	E 69.565	R 42.857	ER 37.500	R 66.667	
11 A 50.000	V 63.636	PTV 33.333	V 66.667	A 50.000	96 V 70.000	V 58.696	L 42.857	A 50.000	V 66.667	187 T 50.000	T 67.391	T 71.429	S 50.000	T 50.000	287 L 100.000	L 60.870	AG 28.571	R 50.000	L 66.667	
13 I 55.556	V 73.333	V 57.143	I 57.143	I 66.667	98 P 70.000	P 78.261	P 57.143	P 100.000	LP 50.000	188 PR 30.000	P 26.087	DPR 28.571	R 87.500	R 66.667	290 A 70.000	A 56.522	AD 42.857	A 75.000	A 50.000	
16 IV 33.333	L 51.111	L 42.857	L 71.429	IL 33.333	99 AE 30.000	A 34.783	A 85.714	A 87.500	AET 33.333	189 L 90.000	L 76.087	T 42.857	L 87.500	L 83.333	294 L 60.000	L 30.435	C 42.857	L 37.500	TV 33.333	
19 AV 44.444	A 40.000	M 57.143	V 71.429	V 66.667	101 L 100.000	L 80.000	D 42.857	E 87.500	L 50.000	190 V 60.000	M 41.304	V 100.000	V 100.000	V 66.667	306 A 70.000	A 63.043	A 57.143	A 75.000	G 83.333	
22 N 60.000	N 80.435	HNT 28.571	F 62.500	N 33.333	102 P 30.000	P 75.556	S 42.857	P 75.000	P 50.000	191 V 60.000	V 63.043	V 71.429	LV 50.000	V 66.667	307 A 100.000	A 82.609	A 57.143	A 37.500	AY 33.333	
24 I 90.000	L 71.739	IL 28.571	IV 37.500	L 66.667	103 D 70.000	E 39.130	P 57.143	E 75.000	DE 50.000	195 V 70.000	V 67.391	T 57.143	T 87.500	T 83.333	314 R 50.000	R 84.783	R 71.429	LR 37.500	R 83.333	
25 G 90.000	G 91.304	S 85.714	G 100.000	DG 50.000	104 Y 90.000	FY 43.478	Y 100.000	Y 66.667	196 D 70.000	D 97.826	E 57.143	D 100.000	D 83.333	315 A 60.000	V 34.783	A 71.429	A 75.000	Q 50.000		
28 E 60.000	A 34.783	AT 28.571	AER 25.000	A 33.333	105 D 50.000	D 40.435	E 42.857	E 62.500	DE 50.000	197 S 90.000	A 47.826	A 57.143	A 87.500	A 50.000	320 L 60.000	L 58.696	I 75.000	IL 50.000	IL 66.667	
29 V 80.000	Y 63.043	Y 57.143	H 62.500	WY 50.000	106 M 50.000	M 56.522	D 42.857	M 75.000	M 50.000	198 A 70.000	S 84.783	P 57.143	S 75.000	S 50.000	327 R 30.000	R 54.348	H 57.143	R 37.500	R 50.000	
31 K 40.000	A 52.174	AK 42.857	AES 25.000	S 50.000	*107 G 90.000	G 97.826	G 57.143	A 87.500	A 66.667	200 D 90.000	C 89.130	A 42.857	C 75.000	S 83.333	328 G 70.000	G 78.261	A 57.143	G 62.500	G 66.667	
33 T 50.000	T 76.087	TV 42.857	V 75.000	T 66.667	109 V 90.000	V 65.271	V 100.000	V 66.667		202 W 90.000	W 80.435	V 85.714	YE 87.500	A 50.000	333 A 70.000	A 65.217	J 71.429	A 62.500	V 83.333	
35 T 30.000	R 26.087	DR 42.857	A 50.000	D 66.667	111 S 70.000	A 93.478	A 85.714	A 87.500	A 100.000	203 G 90.000	G 89.130	G 42.857	G 100.000	G 83.333	337 L 50.000	M 78.261	G 42.857	L 75.000	M 66.667	
37 R 50.000	ER 21.739	T 28.571	R 50.000	K 83.333	112 N 70.000	S 69.565	M 42.857	S 75.000	N 100.000	204 W 60.000	W 86.957	ILM 28.571	L 87.500	L 33.333	338 T 80.000	T 93.478	IT 42.857	T 87.500	T 100.000	
39 G 100.000	G 71.739	GSV 28.571	G 75.000	G 100.000	113 A 90.000	S 63.043	G 57.143	S 87.500	A 50.000	205 V 60.000	V 56.522	V 71.429	D 62.500	V 100.000	343 S 50.000	S 52.174	C 42.857	A 62.500	A 66.667	
40 I 70.000	I 91.304	L 57.143	IL 50.000		114 T 30.000	A 43.478	S 71.429	S 75.000	S 66.667	206 S 80.000	A 71.739	C 85.714	A 87.500	ACS 33.333	346 S 70.000	A 76.087	A 42.857	A 87.500	G 66.667	
42 P 70.000	P 54.348	P 71.429	R 62.500	R 66.667	117 F 100.000	F 97.826	G 42.857	T 87.500	V 66.667	207 HQ 50.000	Q 67.391	A 75.000	Q 100.000	Q 100.000	347 P 90.000	P 52.174	A 42.857	A 62.500	A 66.667	
43 L 60.000	I 143.478	I 57.143	I 87.500	I 66.667	120 T 90.000	G 93.478	T 87.500	G 84.783	G 83.333	209 A 60.000	A 54.343	G 57.143	A 50.000	A 50.000	350 V 80.000	L 56.522	V 57.143	V 87.500	V 50.000	
44 T 60.000	T 60.870	T 71.429	T 75.000	ST 33.333	121 H 100.000	Q 82.609	P 85.714	S 87.500	Q 83.333	210 S 70.000	S 50.000	S 57.143	T 50.000	T 50.000	351 V 70.000	A 80.435	N 42.857	A 75.000	A 66.667	
46 F 70.000	F 100.000	EF 42.857	P 100.000	F 100.000	122 R 80.000	R 63.043	R 100.000	H 62.500	R 83.333	213 IV 40.000	L 67.391	L 71.429	L 87.500	L 83.333	352 A 70.000	P 52.174	T 71.429	T 87.500	AGT 33.333	
47 D 100.000	D 89.130	DG 42.857	D 100.000	D 100.000	124 F 60.000	L 97.826	L 57.143	M 62.500	L 83.333	220 A 40.000	A 26.667	A 57.143	A 71.429	D 50.000	357 L 70.000	L 41.304	D 50.000	L 75.000	L 66.667	
50 R 60.000	G 43.478	HR 42.857	G 37.500	RS 33.333	126 K 80.000	A 39.130	K 42.857	R 37.500	K 100.000	224 L 90.000	L 91.304	LR 42.857	L 87.500	L 83.333	358 R 100.000	R 78.261	ER 42.857	L 75.000	R 100.000	
53 S 50.000	A 58.696	L 42.857	V 75.000	A 66.667	128 W 100.000	W 100.000	W 100.000	H 62.500	W 66.667	229 R 60.000	R 34.783	A 57.143	A 37.500	R 66.667	359 V 80.000	D 60.870	DH 42.857	D 50.000	D 83.333	
54 R 70.000	R 50.000	R 57.143	R 75.000	Q 66.667	129 S 50.000	S 76.087	G 42.857	S 37.500	RS 33.333	233 Y 50.000	Y 58.696	F 57.143	V 87.500	Y 83.333	360 G 60.000	G 73.913	G 71.429	G 87.500	Q 50.000	
55 L 110.000	L 91.304	V 42.857	I 87.500	I 66.667	130 Q 40.000	K 56.522	P 42.857	K 37.500	E 50.000	241 I 50.000	F 95.652	A 85.714	I 87.500	I 66.667	366 A 50.000	V 63.043	P 42.857	A 50.000	V 50.000	
58 Q 80.000	E 91.304	E 57.143	E 75.000	E 66.667	132 P 100.000	S 36.957	P 42.857	P 87.500	P 100.000	246 D 80.000	D 54.348	D 42.857	AD 25.000	D 33.333	367 G 40.000	HN 41.304	N 42.857	G 87.500	H 50.000	
59 I 90.000	V 56.522	V 71.429	V 87.500	AV 33.333	134 HRSY 20.000	V 58.696	FH 28.571	W 62.500	A 33.333	248 A 30.000	E 41.304	A 57.143	E 37.500	E 50.000	368 TV 30.000	V 71.739	V 57.143	V 37.500	V 100.000	
60 D 40.000	P 36.957	P 57.143	P 62.500	DT 33.333	136 S 100.000	S 100.000	G 71.429	G 75.000	S 66.667	249 AS 40.000	A 36.957	A 57.143	A 50.000	DT 33.333	369 T 60.000	T 23.810	L 28.571	T 50.000	T 50.000	
61 D 60.000	G 70.732	G 57.143	G 75.000	D 50.000	137 V 100.000	A 89.130	LR 42.857	A 87.500	A 66.667	251 R 50.000	R 60.870	R 28.571	E 37.500	R 66.667	370 E 30.000	E 26.087	D 57.143	R 50.000	EGST 25.000	
63 E 60.000	P 93.478	V 57.143	P 50.000	SV 33.333	138 A 50.000	G 80.435	G 71.429	G 87.500	G 66.667	252 A 50.000	A 43.478	ER 28.571	A 62.500	AR 33.333	371 V 80.000	P 58.696	V 57.143	P 62.500	P 50.000	
64 A 90.000	A 56.522	A 57.143	A 100.000	AP 50.000	141 F 100.000	F 95.652	A 85.714	I 87.500	I 66.667	255 A 70.000	A 78.261	A 50.000	AV 42.857	A 50.000	372 P 30.000	A 43.478	DP 50.000	A 87.500	A 50.000	
66 D 40.000	D 42.222	E 57.143	E 62.500	D 66.667	147 V 100.000	V 97.826	A 85.714	A 87.500	A 83.333	256 EHP 22.222	R 50.000	A 66.667	R 42.857	EG 33.333	373 D 30.000	P 32.609	P 42.857	P 62.500	P 66.667	
67 H 100.000	H 80.435	LR 42.857	ST 37.500	H 83.333	152 I 70.000	I 86.957	I 57.143	V 50.000	L 66.667	258 IV 30.000	V 43.478	V 66.667	AIV 33.333	I 50.000	375 Y 90.000	Y 57.778	H 57.143	C 50.000	H 50.000	
68 I 60.000	L 80.435	V 57.143	V 75.000	AV 33.333	153 M 80.000	M 65.217	FM 42.857	S 87.500	M 66.667	261 E 77.778	E 65.217	T 42.857	E 37.500	E 83.333	376 E 70.000	D 26.087	DG 28.571	E 37.500	G 33.333	
69 P 100.000	P 97.826	EP 42.857	P 100.000	P 83.333	158 M 80.000	N 65.217	FM 42.857	P 87.500	M 66.667	262 L 55.556	I 67.391	IV 42.857	I 37.500	I 50.000	377 IL 50.000	L 65.217	L 71.429	I 62.500	L 66.667	
70 G 60.000	S 50.000	DG 28.571	G 50.000	S 33.333	159 R 100.000	R 73.913	K 57.143	R 100.000	R 100.000	263 A 88.889	A 82.609	A 85.714	D 62.500	A 100.000	381 LR 40.000	T 60.870	RT 28.571	L 62.500	LRT 33.333	
73 L 110.000	L 65.217	L 71.429	IL 25.000	LV 50.000	162 S 70.000	S 65.217	C 71.429	C 87.500	C 83.333	265 Y 70.000	Y 84.783	HY 42.857	V 87.500	Y 83.333	382 G 70.000	P 60.870	P 58.696	V 57.143	P 50.000	
74 P 100.000	P 93.478	V 57.143	P 50.000	SV 33.333	163 A 50.000	G 80.435	G 71.429	G 87.500	G 66.667	268 T 90.000	T 95.652									

**Table S2.** Predicted volume of the cavity (Å<sup>3</sup>) of KS-CLFs encoding polyketides of various chain lengths.

C16-C18	C20	C24	C26	C28-C30
706	887	1053	1049	1134

**Table S3.** Chemotype and cyclization pattern of C20 polyketides with OxyN versus TcmI cyclases.\*

C20 polyketides					
OxyN			TcmI		
Secondary Metabolite	Chemotype	Shape	Secondary Metabolite	Chemotype	Shape
Chromomycin	aureolic acid	linear	Fluostatin	angucycline	bent
Kosinostatin	anthracycline	linear	Saquayamycin Z	angucycline	bent
Arimetamycin	anthracycline	linear	Chrysomycin	gilvocarcin	bent
Steffimycin	anthracycline	linear	Ravidomycin	gilvocarcin	bent
Dactylocycline	tetracycline	linear	Simocyclinone	angucycline	bent
Chlorotetracycline	tetracycline	linear	Azicemicin	angucycline	bent
Oxytetracycline	tetracycline	linear	Sch47554	angucycline	bent
Tetaramycin	tetracycline	linear	Urdamycin	angucycline	bent
Mithramycin	aureolic acid	linear	Grincamycin	angucycline	bent
Aranciamycin	anthracycline	linear	Kinamycin	angucycline	bent
X26	aureolic acid	linear	Landomycin	angucycline	bent
Chartreusin	other	bent	Rabelomycin	angucycline	bent
Resistomycin	other	bent	Griseusin	isochromanequinone	linear
Nivetetracyclate	other	linear	Chattamycin	angucycline	bent
Polyketomycin	tetracycline	linear	Oviedomycin	angucycline	bent
			Kiamycin	angucycline	bent
			PD 116740	angucycline	bent
			Hatomarubigin	angucycline	bent
			BE 7585A	angucycline	bent
			Lactonamycin	other	bent
			Tetracenomycin	tetracenomycin	linear
			Elloramycin	tetracenomycin	linear
			Gilvocarcin	gilvocarcin	bent

\*Type II polyketides with clusters outside the canonical clade are highlighted in grey.

**Table S4.** Regioselectivity of TcmN-catalyzed first-ring cyclization in the absence of C9 KR

Polyketide	First-ring Cyclization
Spore pigments	No structure
Frankiamycin A	C9-C14
Pradimicin	C9-C14
Fasamycin A	Unknown
Arenimycin B	C9-C14
Calixanthomycin A	C9-C14
Xantholipin	C9-C14
Lysolipin	C9-C14
Arixanthomycin	C9-C14
Griseorhodin	C9-C14
Rubromycin	C9-C14
A 74528	C9-C14
Fredericamycin	C9-C14
TLN 05220	C9-C14
Benastatin	C9-C14
FD 594	C9-C14
Polyketomycin	C7-C12
Lactonamycin	C9-C14
Tetracenomycin	C9-C14
Elloramycin	C9-C14
Resistomycin	C9-C14

## Materials and Methods

### Resources

Data are available at <http://sequence.stanford.edu/TypeIIPKS/>.

### Defining Type II PKS gene clusters

To generate a comprehensive yet non-redundant catalog of orphan Type II PKS gene clusters, we defined an algorithm of five steps (Figure S1).

**Step 1: BLAST search for KS and CLF homologs.** All NCBI nucleotide and genome databases were searched using KS and CLF protein queries (*tblastn*) (4). The *tblastn* algorithm translates proteins from nucleotide records, without using predicted proteins. This allowed inclusion of even unannotated databases such as the NCBI whole-genome shotgun (WGS) database. Query KS and CLF protein sequences were selected from the fredericamycin, resistomycin, and WhiE spore pigment gene clusters owing to their sequence diversity among the set of known clusters. The *tblastn* algorithm was used to search the major nucleotide databases in NCBI (nt, wgs, env\_nt, patnt, htgs, tsa\_nt, other\_genomic, refseq\_genomic, sts, gss).

Because KS and CLF proteins share sequence similarity with distant ketosynthases involved in diverse biosynthetic processes, an initial relaxed (e-value < 1) BLAST search identified a large number of homologs (>100,000) (Figure ). In addition to the expected type II PKS genes, the set also included KASII genes (FabF homologs) from the fatty acid biosynthesis pathway, and type I PKS genes (e.g., the erythromycin KSs). The type II PKS genes were more closely related to the KASII/FabF homologs than to type I PKSs, consistent with previous phylogenetic analysis (5).

**Step 2: Refined pHMM definition of KS and CLF homologs.** To filter the large set of BLAST-identified PKS and FAS KS homologs, we used profile hidden Markov models (pHMMs), statistical models of position-specific conservation in multiple sequence alignments. pHMMs are richer sequence models than the single-sequence queries that are used in BLAST (4). pHMM searches can be computationally-intensive, but the searches were tractable on the BLAST-filtered (and *tblastn*-translated) set of proteins. pHMM searches depend on the sequences that serve as inputs to the model (4).

To select a KS sequence model and input sequences, we created several versions of KS and CLF pHMMs, and compared them in terms of ability to differentiate known type II PKS protein sequences from distant homologs, including fatty acid synthase KS, type I PKS KSs, and related homologs (Figure S2). The diversity of input sequences varied among the models, as well as the total number of sequences. The five models were compared by searching against "positive" and "negative" test sets of genes. The positive set included 31 known type II KS genes, and the negative set genes included (a) FabF/KASII homologs (b) FabH/KASIII homologs, and (c) type I PKS genes. Models were judged by their ability to differentiate the positive set from the negative set. The same procedure was carried out in order to identify an optimal CLF pHMM sequence model. A model built using 12 diverse type II KS (or CLF) proteins demonstrated the

greatest ability to differentiate known type II PKS KSs (or CLFs) from more distantly-related ketosynthases (Figure S2).

To define the sequence similarity threshold for the KS and CLF pHMM similarity search, we examined a distant relative of Type II PKS gene clusters, a gene cluster encoding the isoprenoid quinoline alkaloid aurachin (6-8). This gene cluster harbors an ACP, a distant KS homolog, and a distant CLF homolog (in that order). We defined the KS and CLF thresholds as just allowing the KS and CLF scores in this gene cluster, but no protein sequences more distant.

Using aurachin KS and CLF scores as a threshold, we identified 12,868 nucleotides having a match to KS or CLF. The Type I PKS proteins that had appeared in the BLAST step were eliminated here (scored below the aurachin threshold) (Figure S2). However, this threshold did not eliminate the FabF/KASII genes; indeed, the scores in Figure S2 suggest that type II KS and CLF proteins are more closely related to FabF homologs than to the aurachin KS and CLF. Phylogenetic analysis of proteins scoring as well or better than aurachin KS and CLF homologs similarly revealed a close relationship of the FabF proteins to the type II KS and CLF proteins, with aurachin and other ketosyntheses appearing relatively distant (Figure S3). Besides FabF homologs, this set also contained some related KS genes involved in secondary metabolism, such as the recently-discovered APE genes (9), as well as some standalone non-iterative KSs such as cylIE from the cylindrocyclophane gene cluster (10). These genes usually only had one KS, with no apparent CLF nearby, and were therefore eliminated in subsequent steps.

*Step 3: Require both KS and CLF.* Requiring that a nucleotide record have two distinct ketosynthase ORFs, one of which matched KS and one of which matched CLF, reduced the set to 1,874 nucleotide records.

*Step 4: KS + CLF spatially clustered on chromosome.* In all characterized type II PKS gene clusters, the KS and CLF are physically clustered on the genome, directly in tandem with no intervening genes. We defined the maximum allowed distance between KS and CLF as 2 kb. This filter reduced the set to 810 KS+CLF proteins, in 729 nucleotide records.

*Step 5: Eliminate redundant gene clusters.* To select a threshold for defining redundant gene clusters in the type II PKS catalog, we manually examined pairwise sequence similarities of two pairs of characterized gene clusters that encode similar but non-identical compounds. (1) Julichrome and setomimycin are closely related compounds, with same chain length (11); their CLF proteins are 86% identical, and KS 90% identical. (2) Fredericamycin and A-74528 are of the same chain lengths but are structurally distinct from one another (12); they are encoded by CLF proteins that are 84% identical, and KS proteins that are 93% identical. We selected the CLF protein for the redundancy calculations, as it displays the greater sequence variability. At a redundancy threshold of 87% CLF amino acid and 93% KS amino acid identity, we identified 544 non-redundant CLF protein sequences, each of which has a KS gene in tandem. Given that all characterized CLFs at  $\leq$  86% identity produce unique compounds and that CLF pairs of higher sequence identity produced identical compounds, we set our threshold for de-replication at 87%. These 544 putative type II gene clusters exist on 495 NCBI nucleotide records/genomes, as some genomes harbor two type II PKS gene clusters.

## **Reference set of 78 characterized Type II PKS gene clusters**

A reference set of 78 characterized type II PKS gene clusters (Table S1) was compiled manually by browsing the set of 544 putative type II PKS clusters, and referring to primary literature and reviews (7), ClusterMine360 (13), and the Dictionary of Natural Products (<http://dnp.chemnetbase.com>). The compound name, product structure, accession number, and producing organism were obtained from literature searches and the GenBank record of each gene cluster. The carbon chain length and priming unit were determined by manual structural analysis of each polyketide. The year of structure publication for each cluster was found by literature search and augmented by literature references contained within the Dictionary of Natural Products entry for each polyketide. The year of cluster sequencing was provided by information requests to the National Center for Biotechnology Information.

## **Phylogenetic analysis**

For the large trees, multiple sequence alignment was performed using MAFFT (1), and large maximum likelihood phylogenetic trees were generated using FastTree2 (2). Trees were plotted using the APE package (14). For the small trees, multiple sequence alignment was performed using MUSCLE (15), and Bayesian inference of phylogeny was performed using MrBayes (16), using the mixed model of amino acid substitution. The WAG (17) model was identified as the best by MrBayes.

## **Identifying cavity-lining residues (CLRs) of interest**

The KS and CLF sequences represented in the reference set were binned into 5 categories based on phylogenetic distribution of CLFs encoding products of varying chain length: C16-C18, C20, C24, C26, and C28-C30. For each amino acid site position, the most commonly-occurring amino acid was identified using Consurf (18). If the amino acid position was conserved (defined as >50% conservation) within at least one bin but not conserved across all bins, it was defined as a potential polyketide-length determining residue (See Table S1 for a comprehensive list of all residues that were not conserved across all bins).

Next we sought to identify cavity-lining amino acid residues that could directly affect the volume of the KS/CLF active site cavity if mutated (and therefore might direct the length of the polyketide). ACT KS-CLF CLRs were obtained by uploading the actinorhodin KS-CLF crystal structure (PDB ID 1TQY) to MOLE 2.0 (19) and using the cavity search feature with the default probe radius of 3.0 and interior threshold of 1.25, and the following selections (based on the solves KS/CLF structure): C169 A, H309 A, H346 A, F117 B, T120 B, F124 B, E139 B, S140 B, and W202 B. The KS-CLF cavity was consistent with that identified by previously (20) and the 43 inner residues were defined as CLRs. Residues important for the polyketide size were defined as CLRs with variable intra-bin consensus sequences (as defined above) and were used to calculate the effect of CLR sequence variation on the cavity volume (See residues highlighted in grey in Table S1).

## **Volume analysis of the CLRs of interest**

The effect of CLR sequence variation on the KS-CLF cavity was analyzed by creating KS-CLF PDB files representing mutations observed in the CLF for each of the 5

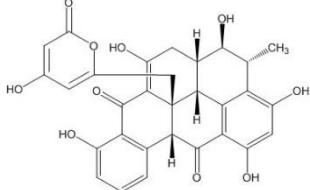
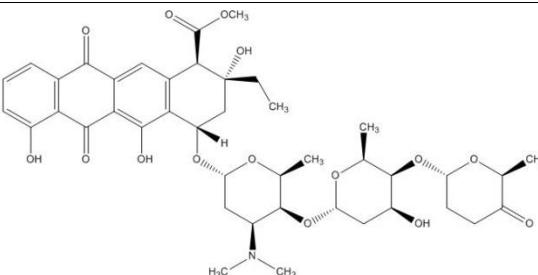
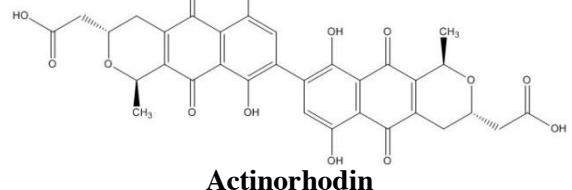
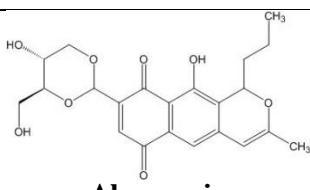
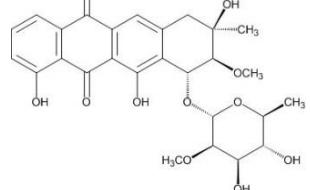
categories of chain length. The following mutations were made using the mutagenesis wizard in PyMol: C16-C16 (C114T); C20 (C114A, T120G, F124L, E139Q); C24 (C114S, F117G, T120G, F124L, E139Q, F141I); C26 (C114S, F117T, T120G, F124M, E139Q, F141I, W202Y); C28-C30 (C114S, F117V, T120G, F124L, E139M, F141I, W202A). The volume of the cavity was then predicted for each mutant using CASTp (21) and compared to the volume of the original actinorhodin KS/CLF cavity (706.2 Å) to determine the effect of KS-CLF sequence variation on cavity volume (See Figure S6).

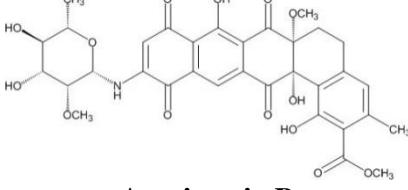
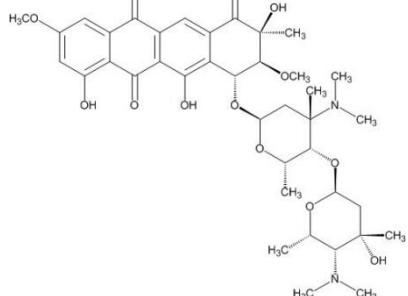
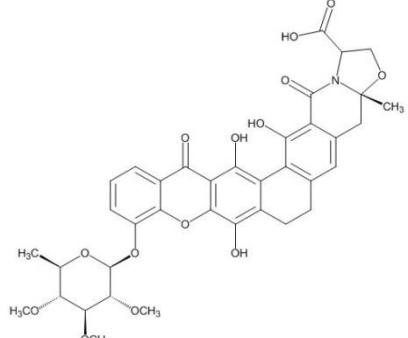
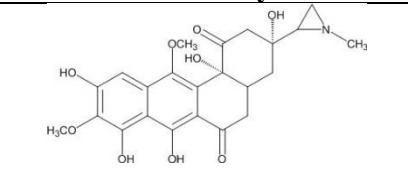
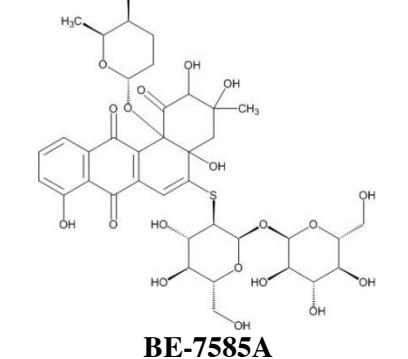
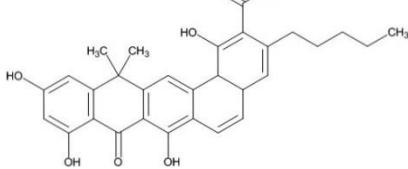
## References

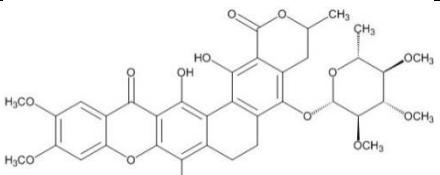
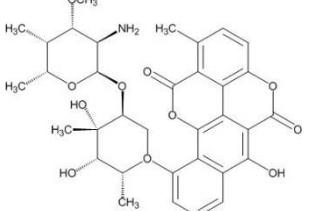
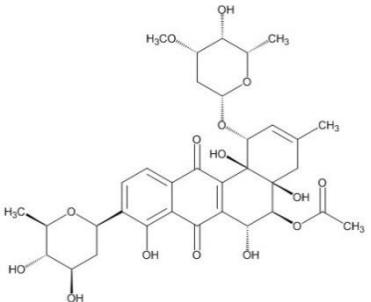
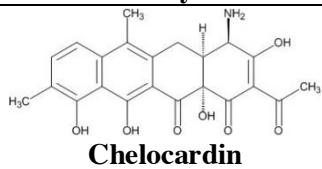
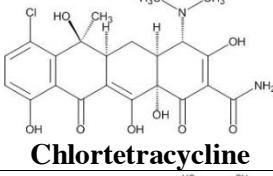
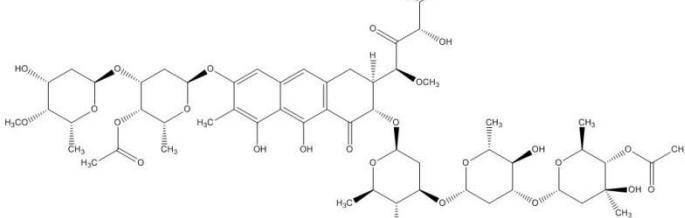
1. Kato K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780.
2. Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5(3):e9490.
3. Scornavacca C, Zickmann F, Huson DH (2011) Tanglegrams for rooted phylogenetic trees and networks. *Bioinformatics* 27(13):i248–56.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3):403–410.
5. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* 22(10):2027–2039.
6. Höfle G, Kunze B (2008) Biosynthesis of aurachins A-L in *Stigmatella aurantiaca*: a feeding study. *J Nat Prod* 71(11):1843–1849.
7. Ridley CP, Lee HY, Khosla C (2008) Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci USA* 105(12):4595–4600.
8. Pistorius D, Li Y, Sandmann A, Müller R (2011) Completing the puzzle of aurachin biosynthesis in *Stigmatella aurantiaca* Sg a15. *Mol BioSyst* 7(12):3308–3315.
9. Cimermancic P, et al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158(2):412–421.
10. Nakamura H, Hamer HA, Sirasani G, Balskus EP (2012) Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J Am Chem Soc* 134(45):18518–18521.
11. Präg A, et al. (2014) Regio- and stereoselective intermolecular oxidative phenol coupling in *Streptomyces*. *J Am Chem Soc* 136(17):6195–6198.
12. Zaleta-Rivera K, Charkoudian LK, Ridley CP, Khosla C (2010) Cloning, sequencing, heterologous expression, and mechanistic analysis of A-74528 biosynthesis. *J Am Chem Soc* 132(26):9122–9128.
13. Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res* 41(Database issue):D402–7.
14. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.
15. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
16. Ronquist F, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539–542.

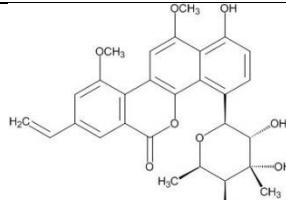
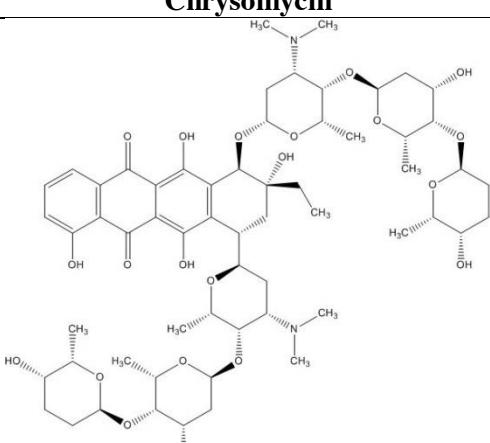
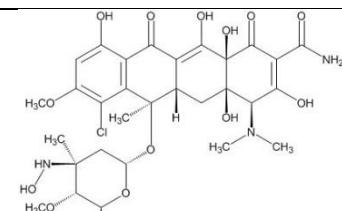
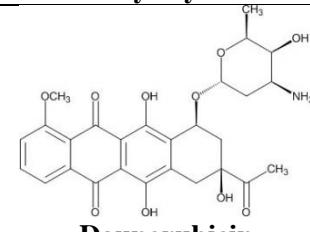
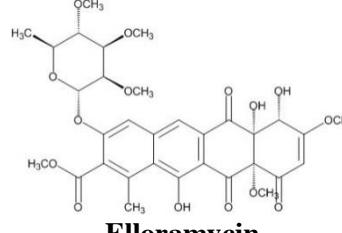
17. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18(5):691–699.
18. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–33.
19. Sehnal D, et al. (2013) MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *Journal of Cheminformatics* 5(1):39.
20. Keatinge-Clay AT, Maltby DA, Medzihradszky KF, Khosla C, Stroud RM (2004) An antibiotic factory caught in action. *Nat Struct Mol Biol* 11(9):888–893.
21. Dundas J, et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34(Web Server issue):W116–8.

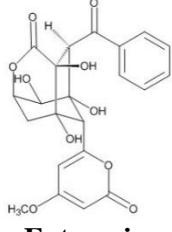
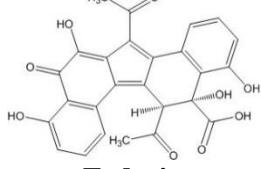
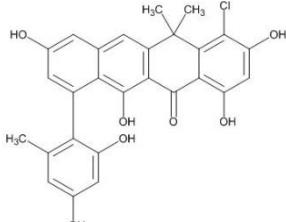
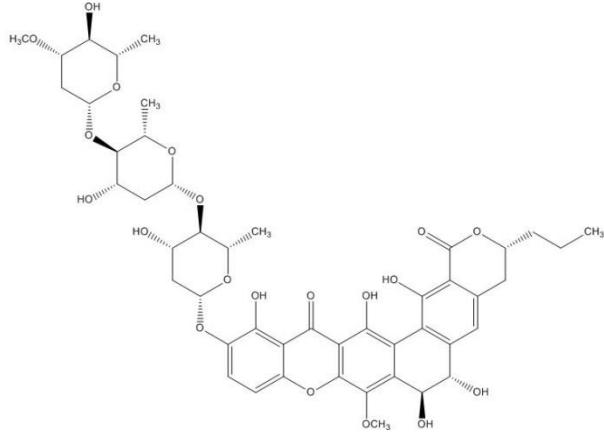
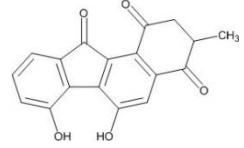
**Table S5.** The type II polyketide validated reference set.

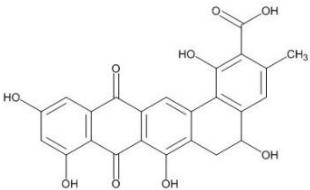
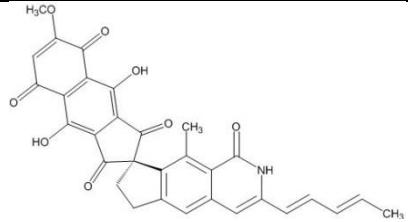
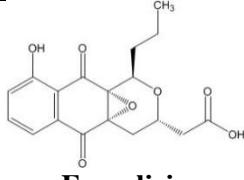
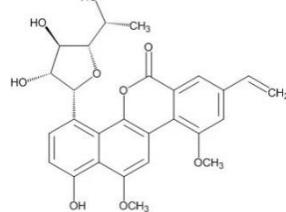
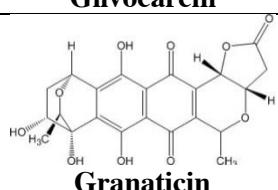
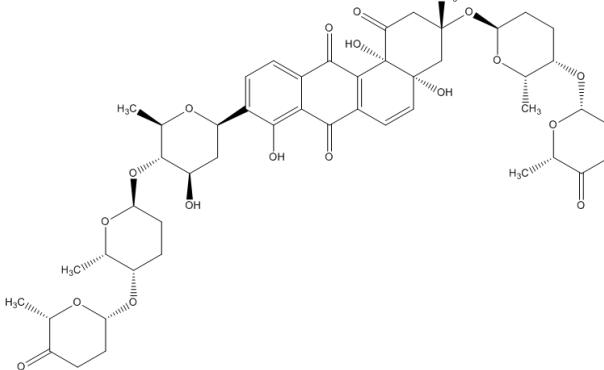
Structure Compound Name	Chain length, Priming (Year structure solved; structure PMID)	Source (Accession number) [Year sequence made public; sequence PMID]
 <p><b>A-74528</b></p>	C30, Hexadienyl <b>(2005; 16061376)</b>	<i>Streptomyces</i> sp. SANK 61196 (GU937384) <b>[2010; 20550125]</b>
 <p><b>Aclacinomycin A</b></p>	C21, Propionyl <b>(1975; 1184473)</b>	<i>Streptomyces galilaeus</i> (AF257324) <b>[2000; 12137949]</b>
 <p><b>Actinorhodin</b></p>	C16, Acetyl <b>(1966; Brockmann, et al.)</b>	<i>Streptomyces coelicolor</i> (AL939122, X63449) <b>[1992; 12000953]</b>
 <p><b>Alnumycin</b></p>	C18, Butyryl <b>(1998; 9589078)</b>	<i>Streptomyces</i> sp. CM020 (EU852062) <b>[2008; 18940666]</b>
 <p><b>Aranciamycin A</b></p>	C20, Acetyl <b>(1970; 5433438)</b>	<i>Streptomyces echinatus</i> (DQ915964) <b>[2006; GenBank record only; no publication]</b>

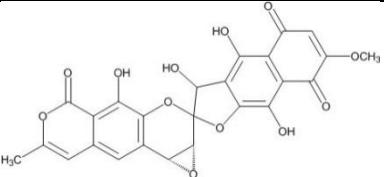
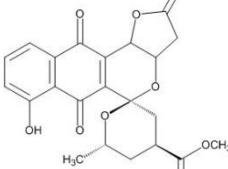
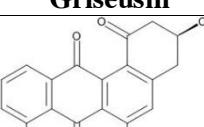
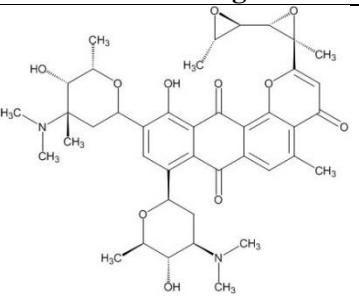
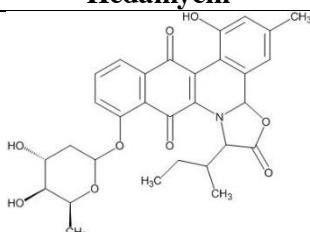
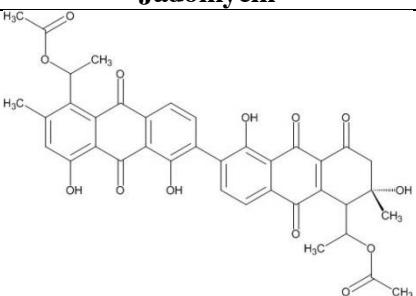
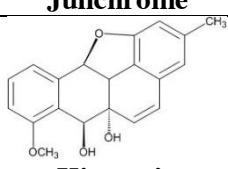
 <p><b>Arenimycin B</b></p>	C24, Acetyl <b>(2009; 19927167)</b>	<i>Salinispora arenicola</i> CNB527 (AZXI01000002) [2014; GenBank record only; no publication]
 <p><b>Arimetamycin</b></p>	C20, Acetyl <b>(2013; 24038656)</b>	Uncultured bacterium clone AZ129 (KF040454) [2013; 24038656]
 <p><b>Arixanthomycin</b></p>	C26, Acetyl <b>(2014; 24730509)</b>	Uncultured bacterium (KF931341) [2014; 24730509]
 <p><b>Azicemicin</b></p>	C20, Aziridyl <b>(1995; 7490223)</b>	<i>Kibdelosporangium</i> sp. MJ126-NF4 (GU134622) [2010; 19928906]
 <p><b>BE-7585A</b></p>	C20, Acetyl <b>(1990; Okabe, et al.)</b>	<i>Amycolatopsis orientalis</i> subsp. <i>vinearia</i> strain BA-07585 (HM055942) [2010; 20443562]
	C28, Hexanoyl <b>(1993; 8514625)</b>	<i>Streptomyces</i> sp. A2991200 (AM501485) [2008; 17439117]

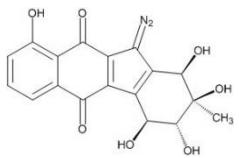
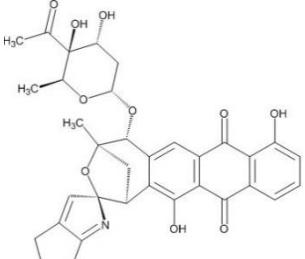
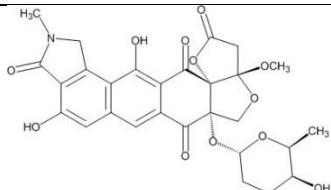
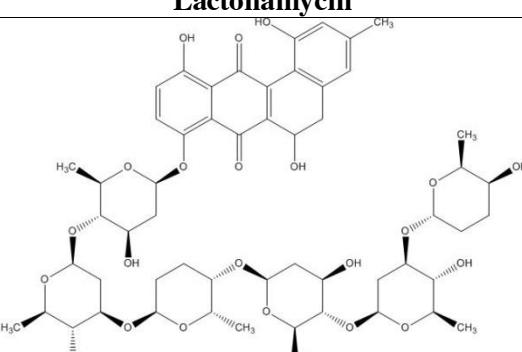
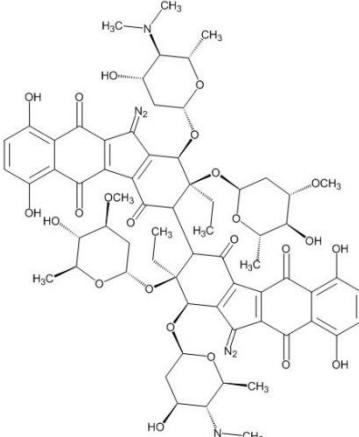
<b>Benastatin</b>		
 <b>Calixanthomycin A</b>	C26, Acetyl <b>(2014; 25521786)</b>	Uncultured bacterium clone BAC AB1692/916/170 (KM881706) <b>[2014; 25521786]</b>
 <b>Chartreusin</b>	C20, Acetyl <b>(1960; Simonitsch, et al.)</b>	<i>Streptomyces chartreusis</i> HKI-249 (AJ786382) <b>[2005; 15911378]</b>
 <b>Chattamycin A</b>	C20, Acetyl <b>(2015; 25511454)</b>	<i>Streptomyces chattanoogensis</i> strain L10 (KM264312) <b>[2014; 25511454]</b>
 <b>Chelocardin</b>	C20, Acetyl <b>(1970; 5459200)</b>	<i>Amycolatopsis sulphurea</i> strain NRRL 2822 (KC870000) <b>[2013; 24043447]</b>
 <b>Chlortetracycline</b>	C19, Malonamyl <b>(1952; Stephens, et al.)</b>	<i>Streptomyces aureofaciens</i> strain F3 (HM627755) <b>[2013; 23800859]</b>
 <b>Chromomycin</b>	C20, Acetyl <b>(1966; 5905316)</b>	<i>Streptomyces griseus</i> subsp. <i>griseus</i> (AJ578458) <b>[2004; GenBank record only; no publication]</b>

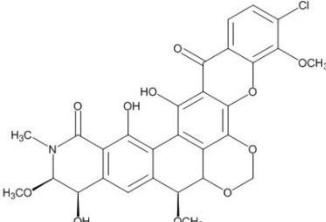
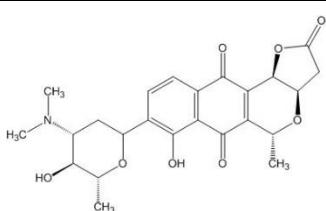
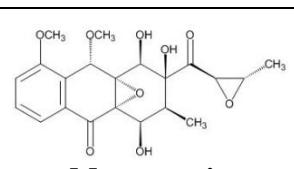
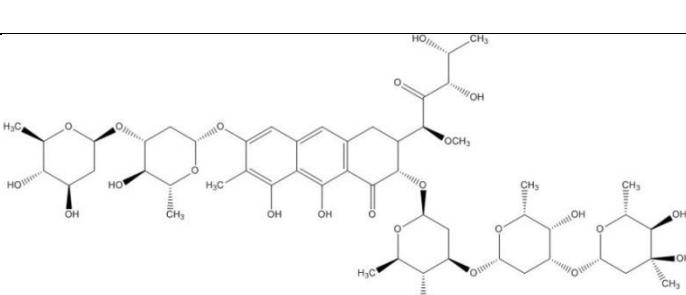
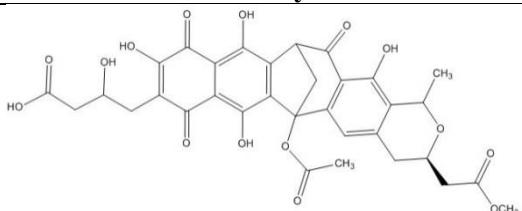
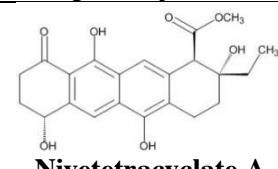
 <p><b>Chrysomycin</b></p>	C20, Acetyl <b>(1982; 7142022)</b>	<i>Streptomyces albaduncus</i> (FN565166) [2009; GenBank record only; no publication]
 <p><b>Cosmomycin</b></p>	C21, Propionyl <b>(1986; 3700244)</b>	<i>Streptomyces olindensis</i> DAUPE 5622 (DQ280500) [2005; 16810496]
<p><i>(Definitive structure unknown)</i></p> <p><b>Cur pigment</b></p>	C24, Acetyl	<i>Streptomyces curacoi</i> (X62518) [1991; 1644304]
 <p><b>Dactyloclavine</b></p>	C20, Malonamyl <b>(1992; 1490880)</b>	<i>Dactylosporangium</i> sp. SC14051 (JX262387) [2012; 23024027]
 <p><b>Daunorubicin</b></p>	C21, Propionyl <b>(1968; Arcamone, et al.)</b>	<i>Streptomyces peucetius</i> (U77891) [1996; 8955419]
 <p><b>Elloramycin</b></p>	C20, Acetyl <b>(1985; 3840789)</b>	<i>Streptomyces olivaceus</i> strain Tu 2353 (AM900040) [2007; 18310024]

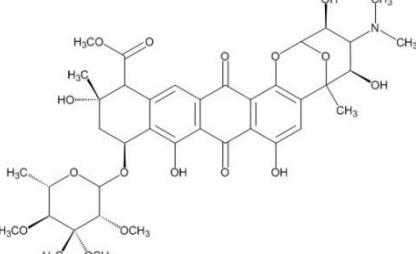
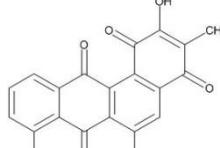
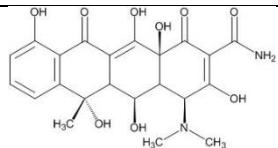
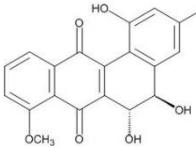
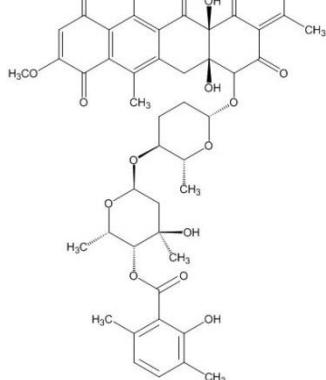
 <p><b>Enterocin</b></p>	C21, Benzoyl <b>(1976; 770404)</b>	<i>Streptomyces maritimus</i> (AF254925) [2000; 11137817]
 <p><b>Erdacin</b></p>	C16, Acetyl <b>(2009; 19621341)</b>	Uncultured bacterium V167 (FJ719113) [2010; 19621341]
 <p><b>Fasamycin A</b></p>	C26, Acetyl <b>(2012; 22224500)</b>	Uncultured bacterium clone AZ154 (HQ828985) [2011; 21768346]
 <p><b>FD-594</b></p>	C28, Butyryl <b>(1998; 9589064)</b>	<i>Streptomyces</i> sp. TA-0256 (AB469194) [2011; GenBank record only; no publication]
 <p><b>Fluostatin</b></p>	C20, Acetyl <b>(1998; 9711223)</b>	Uncultured bacterium BAC AB649/1850 (HM193369) [2011; GenBank record only; no publication]

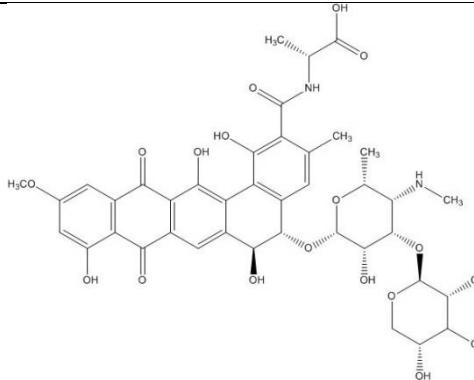
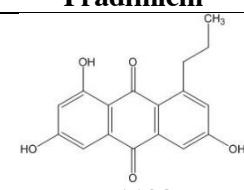
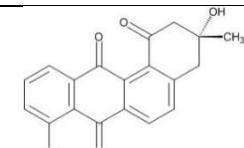
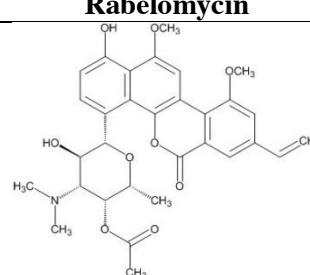
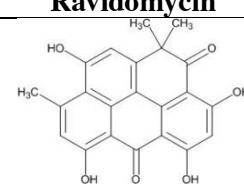
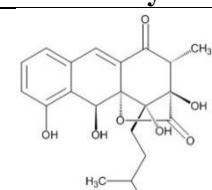
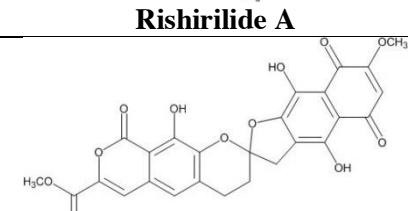
 <p><b>Frankiamycin</b></p>	C24, Acetyl <b>(2015; 25837682)</b>	<i>Frankia</i> sp. EAN1pec (CP000820) [2007; GenBank record only; no publication]
 <p><b>Fredericamycin</b></p>	C30, Hexadienyl <b>(1987; 3112081)</b>	<i>Streptomyces griseus</i> (AF525490) [2006; 16305230]
 <p><b>Frenolicin</b></p>	C18, Butyryl <b>(1966; 5636537)</b>	<i>Streptomyces roseofulvus</i> (AF058302) [1998; 8181754]
 <p><b>Gilvocarcin</b></p>	C21, Propionyl <b>(1981; 7275808)</b>	<i>Streptomyces griseoflavus</i> strain Goe 3592 (AY233211) [2003; 12822997]
 <p><b>Granaticin</b></p>	C16, Acetyl <b>(1968; 5680736)</b>	<i>Streptomyces violaceoruber</i> Tu22 (AJ011500) [1965; 2583128]
 <p><b>Grincamycin</b></p>	C20, Acetyl <b>(1987; 3429346)</b>	<i>Streptomyces lusitanus</i> strain SCSIO LR32 (KC962511) [2013; 23782455]

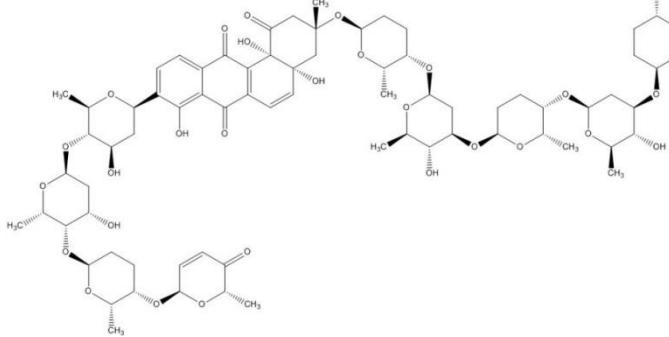
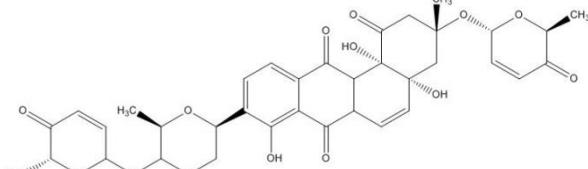
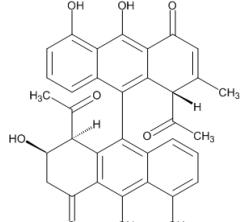
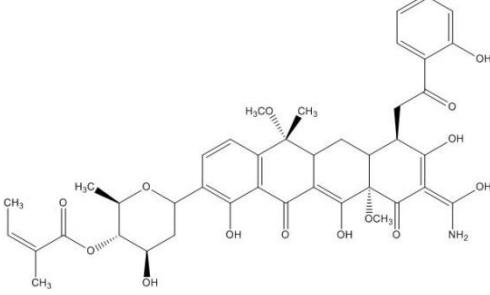
 <p><b>Griseorhodin</b></p>	C26, Acetyl <b>(1965; 5881054)</b>	<i>Streptomyces</i> sp. JP95 (AF509565) <b>[2002; 12323376]</b>
 <p><b>Griseusin</b></p>	C20, Acetyl <b>(1976; 819406)</b>	<i>Streptomyces griseus</i> (X77865) <b>[1994; 8169211]</b>
 <p><b>Hatomarubigin</b></p>	C20, Acetyl <b>(1991; 1761414)</b>	<i>Streptomyces</i> sp. 2238-SVT4 (AB524586) <b>[2010; 20453135]</b>
 <p><b>Hedamycin</b></p>	C24, Hexadienyl <b>(1977; 863729)</b>	<i>Streptomyces griseoruber</i> (AY196994) <b>[2004; 15271354]</b>
 <p><b>Jadomycin</b></p>	C20, Acetyl <b>(1993; 8514643)</b>	<i>Streptomyces venezuelae</i> (AF126429) <b>[1999; 7881555]</b>
 <p><b>Julichrome</b></p>	C18, Acetyl <b>(1969; 5802377)</b>	<i>Streptomyces afghaniensis</i> (NZ_KE354310) <b>[2013; GenBank record only; no publication]</b>
 <p><b>Kiamycin</b></p>	C18, Acetyl <b>(2012; 22611353)</b>	<i>Streptomyces</i> sp. W007 (AGSW01000147) <b>[2011; 22374958]</b>

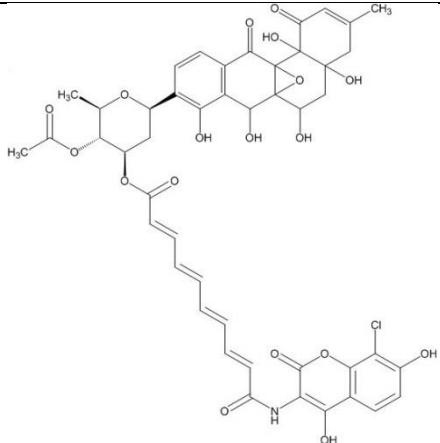
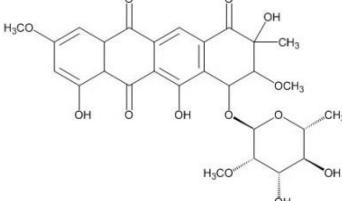
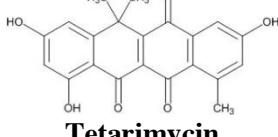
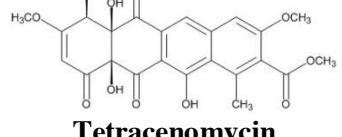
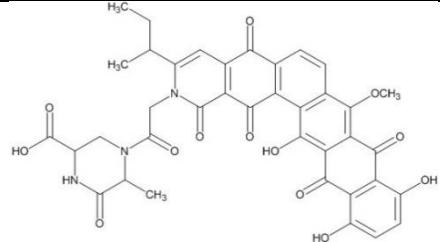
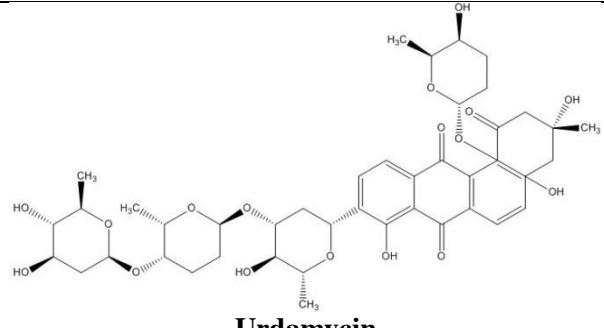
 <p><b>Kinamycin</b></p>	C20, Acetyl <b>(1970; 5458310)</b>	<i>Streptomyces murayamaensis</i> (AY228175) [2003; GenBank record only; no publication]
 <p><b>Kosinostatin</b></p>	C20, Acetyl <b>(2002; 12002994)</b>	<i>Micromonospora</i> sp. TP-A0468 (JN038178) [2014; 23790490]
 <p><b>Lactonamycin</b></p>	C20, Glycine <b>(1999; 10348043)</b>	<i>Streptomyces sanglieri</i> strain AK 623 (EU147299) [2008; 18070976]
 <p><b>Landomycin</b></p>	C20, Acetyl <b>(1990; 2358402)</b>	<i>Streptomyces cyanogenus</i> (AF080235) [1999; 9933932]
 <p><b>Lomaiviticin</b></p>	C21, Propionyl <b>(2001; 11457405)</b>	<i>Salinispora pacifica</i> strain DPJ-0019 (KF515737) [2014; GenBank record only; no publication]

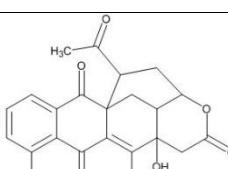
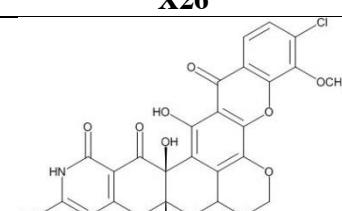
 <p><b>Lysolipin</b></p>	C24, Malonamyl or Acetyl <b>(1977; 838594)</b>	<i>Streptomyces tendae</i> , strain Tu 4042 (AM492533) [2007; 20399259]
 <p><b>Medermycin</b></p>	C16, Acetyl <b>(1985; Tanaka, et al.)</b>	<i>Streptomyces</i> sp. AM-7161 (AB103463) [2003; 12855716]
 <p><b>Mensacarcin</b></p>	C20, Acetyl <b>(2004; 15372645)</b>	<i>Streptomyces bottropensis</i> strain Goe C4/4 (KJ437437) [2014; GenBank record only; no publication]
 <p><b>Mithramycin</b></p>	C20, Acetyl <b>(1968; 5654066)</b>	<i>Streptomyces argillaceus</i> (X89899) [1995; 8807845]
 <p><b>Naphthocyclinone</b></p>	C16, Acetyl <b>(1974; Zeeck, et al.)</b>	<i>Streptomyces arenae</i> strain DSM40737 (AF098965) [1999; 10206788]
 <p><b>Nivetetrcyclate A</b></p>	C21, Propionyl <b>(2013; 24160456)</b>	<i>Streptomyces</i> sp. Ls2151 (KF563088) [2013; 24160456]

 <p><b>Nogalamycin</b></p>	C20, Acetyl <b>(1977; 830691)</b>	<i>Streptomyces nogalater</i> (AJ224512) [1998; 8668120]
 <p><b>Oviedomycin</b></p>	C20, Acetyl <b>(2002; 12027768)</b>	<i>Streptomyces antibioticus</i> strain ATCC11891 (AJ632203) [2004; GenBank record only; no publication]
 <p><b>Oxytetracycline</b></p>	C20, Malonamyl <b>(1953; Hochstein, et al.)</b>	<i>Streptomyces rimosus</i> (DQ143963) [2005; 16597959]
 <p><b>PD 116740</b></p>	C20, Acetyl <b>(1985; 3841121)</b>	<i>Streptomyces</i> WP 4669 PD 116740 (AY228176) [2003; GenBank record only; no publication]
 <p><b>Polyketomycin</b></p>	C20, Acetyl <b>(1998; 9531984)</b>	<i>Streptomyces diastatochromogenes</i> strain Tu6028 (FJ483966) [2009; 19266534]

 <p><b>Pradimicin</b></p>	C24, Acetyl <b>(1993;</b> Tsunakawa, <i>et al.</i> )	<i>Actinomadura hibisca</i> strain P157-2 (EF151801) <b>[2007;</b> GenBank record only; no publication]
 <p><b>R1128</b></p>	C18, Alkylacyl <b>(1993;</b> 8360100)	<i>Streptomyces</i> sp. R1128 (AF293442) <b>[2000;</b> 10931852]
 <p><b>Rabelomycin</b></p>	C20, Acetyl <b>(1970;</b> 5459625)	<i>Streptomyces</i> sp. PGA64 (AY034378) <b>[2001;</b> 12654660]
 <p><b>Ravidomycin</b></p>	C21, Propionyl <b>(1981;</b> Findlay <i>et al.</i> )	<i>Streptomyces ravidus</i> strain C23201NS3 (FN565485) <b>[2009;</b> 20140934]
 <p><b>Resistomycin</b></p>	C20, Acetyl <b>(1967;</b> 6073224)	<i>Streptomyces resistomycificus</i> (AJ585192) <b>[2006;</b> 14982421]
 <p><b>Rishirilide A</b></p>	C19, 1- Methylpentanoyl <b>(1984;</b> 6209258)	<i>Streptomyces bottropensis</i> strain Goe C4/4 (KJ437438) <b>[2014;</b> GenBank record only; no publication]
	C26, Acetyl <b>(1966;</b> 5955233)	<i>Streptomyces collinus</i> DSM2012 (AF293355) <b>[2000;</b> GenBank record only; no

Rubromycin		publication]
	C20, Acetyl <b>(2005; 15835722)</b>	<i>Micromonospora</i> sp. Tu 6368 (FJ670504) <b>[2009; 17497146]</b>
<b>Saquayamycin Z</b>		
	C20, Acetyl <b>(1993; 8514642)</b>	<i>Streptomyces</i> sp. SCC 2136 (AJ628018) <b>[2006; 17085966]</b>
<b>Sch47554</b>  (Definitive structure unknown) <b>Sch pigment</b>	C24, Acetyl	<i>Streptomyces halstedii</i> (L05390) <b>[1993; 8344517]</b>
	C18, Acetyl <b>(1980; Kakinuma, et al.)</b>	<i>Streptomyces aurantiacus</i> JA 4570 (NZ_AOPZ01000074 .1) <b>[2013; GenBank record only; no publication]</b>
	C19, Malonamyl <b>(1992; 1577661)</b>	<i>Streptomyces</i> sp. SF2575 (GQ409537) <b>[2010; 19908837]</b>

 <p><b>Simocyclinone</b></p>	C20, Acetyl <b>(2000; 11079799)</b>	<i>Streptomyces antibioticus</i> (AF324838) <b>[2001; 11959542]</b>
 <p><b>Steffimycin</b></p>	C20, Acetyl <b>(1967; 4964307)</b>	<i>Streptomyces steffisburgensis</i> strain NRRL 3193 (AM156932) <b>[2006; 16751529]</b>
 <p><b>Tetramycin</b></p>	C20, Acetyl <b>(2012; 23157252)</b>	Uncultured bacterium clone AZ60 (JX843821) <b>[2012; 23157252]</b>
 <p><b>Tetracenomycin</b></p>	C20, Acetyl <b>(1979; 485765)</b>	<i>Streptomyces glaucescens</i> (M80674) <b>[1993; 1548230]</b>
 <p><b>TLN-05220</b></p>	C28, Butyryl <b>(2009; 19680283)</b>	<i>Micromonospora echinospora</i> subsp. challisensis strain NRRL 12255 (FJ915123) <b>[2010; 19680283]</b>
 <p><b>Urdamycin</b></p>	C20, Acetyl <b>(1988; 3346183)</b>	<i>Streptomyces fradiae</i> strain T#2717 (AF164961) <b>[1999; 10658661]</b>

<i>(Definitive structure unknown)</i> <b>WhiE pigment</b>	C24, Acetyl	<i>Streptomyces coelicolor</i> (X55942) [1990; 2077356]
 <b>X26</b>	C20, Acetyl (2011; 21768346)	Uncultured bacterium clone X26 (HQ828986) [2011, 21768346]
 <b>Xantholipin</b>	C26, Acetyl (2003; Terui, et al.)	<i>Streptomyces flavogriseus</i> strain SIIA-A02191 (GQ421798) [2012; 22444597]