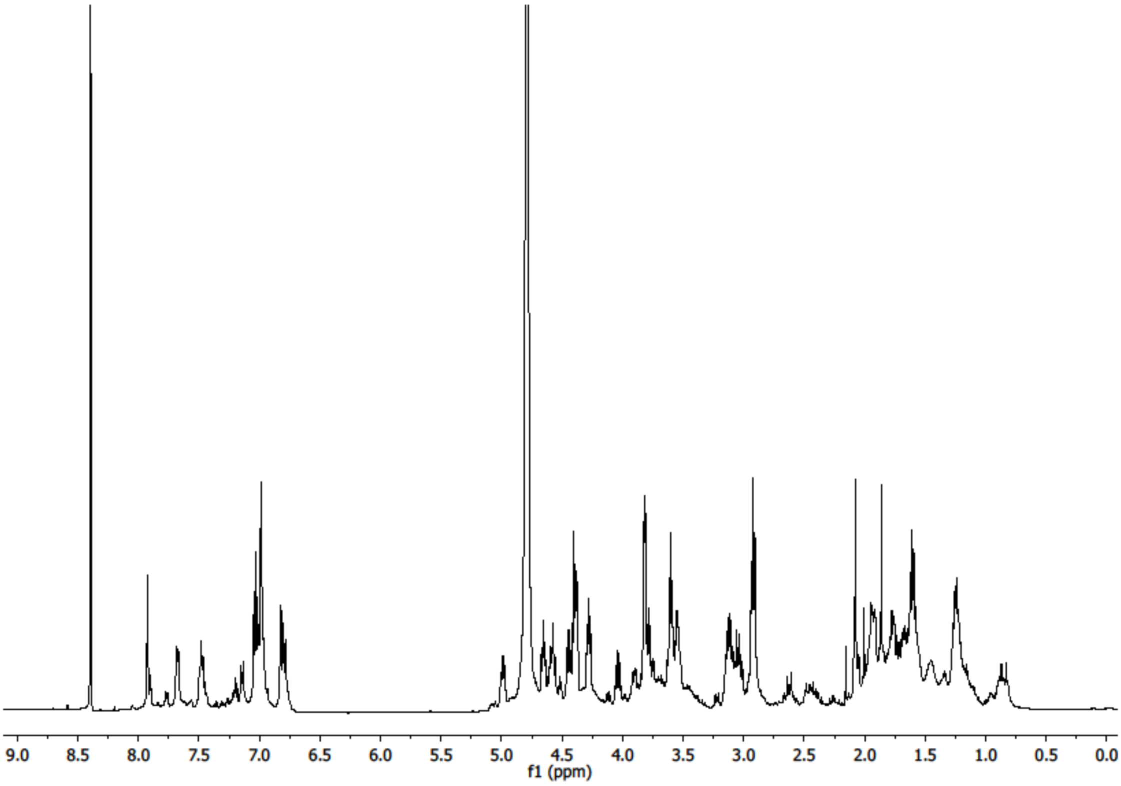
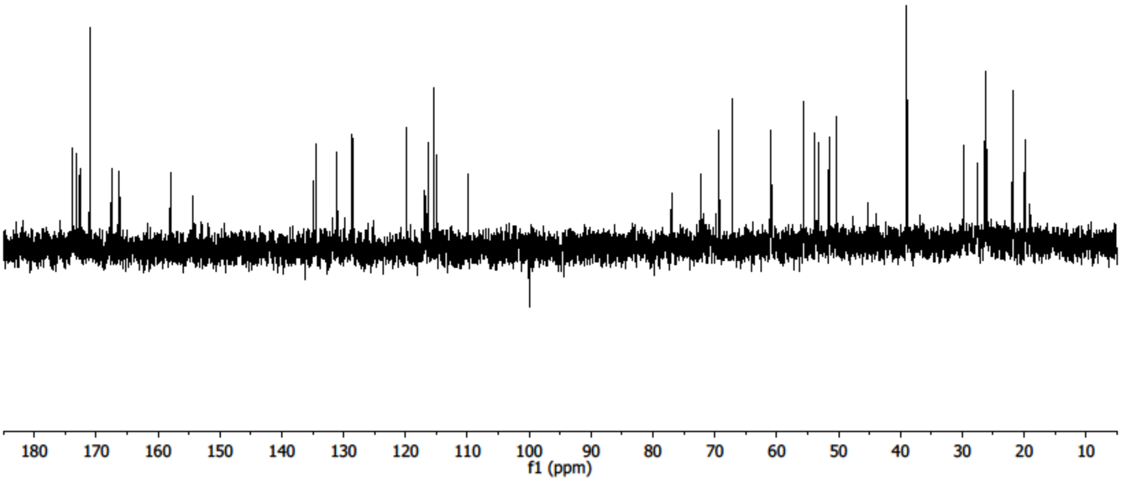
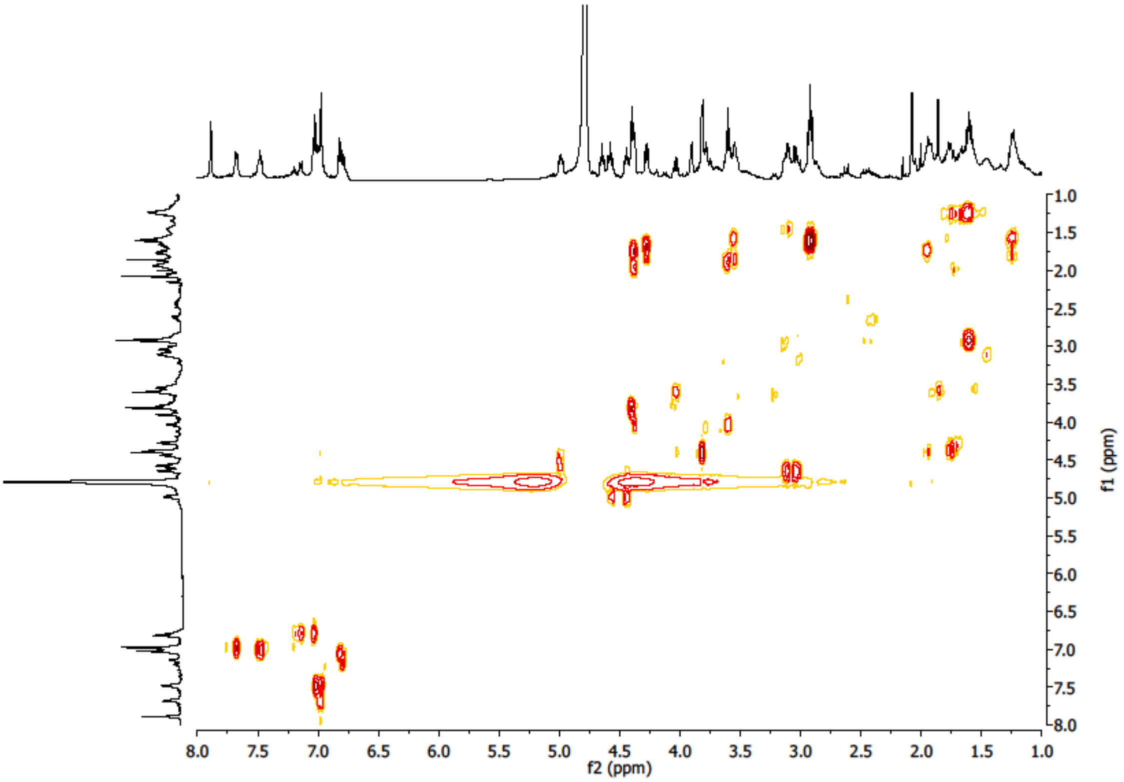
Electronic Supplementary Information



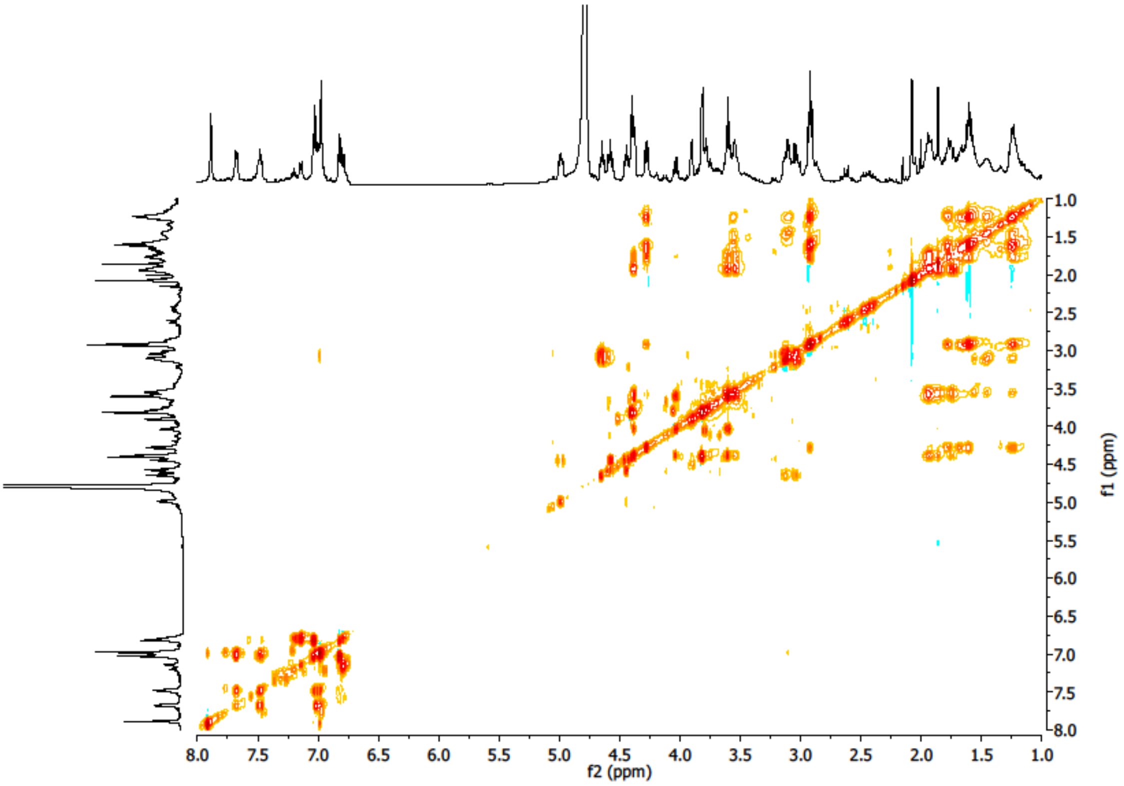
Supplementary Figure 1. ^{1}H NMR spectrum of gobichelin A in D_2O.



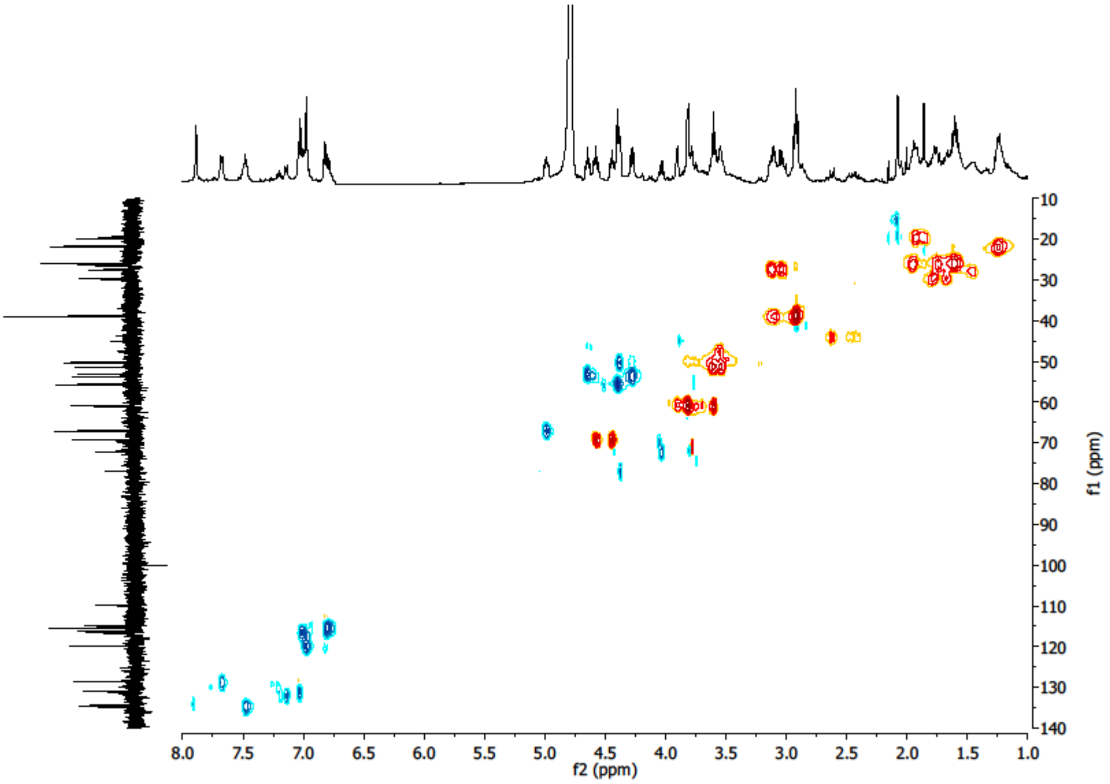
Supplementary Figure 2. ^{13}C NMR spectrum of gobichelin A in D_2O



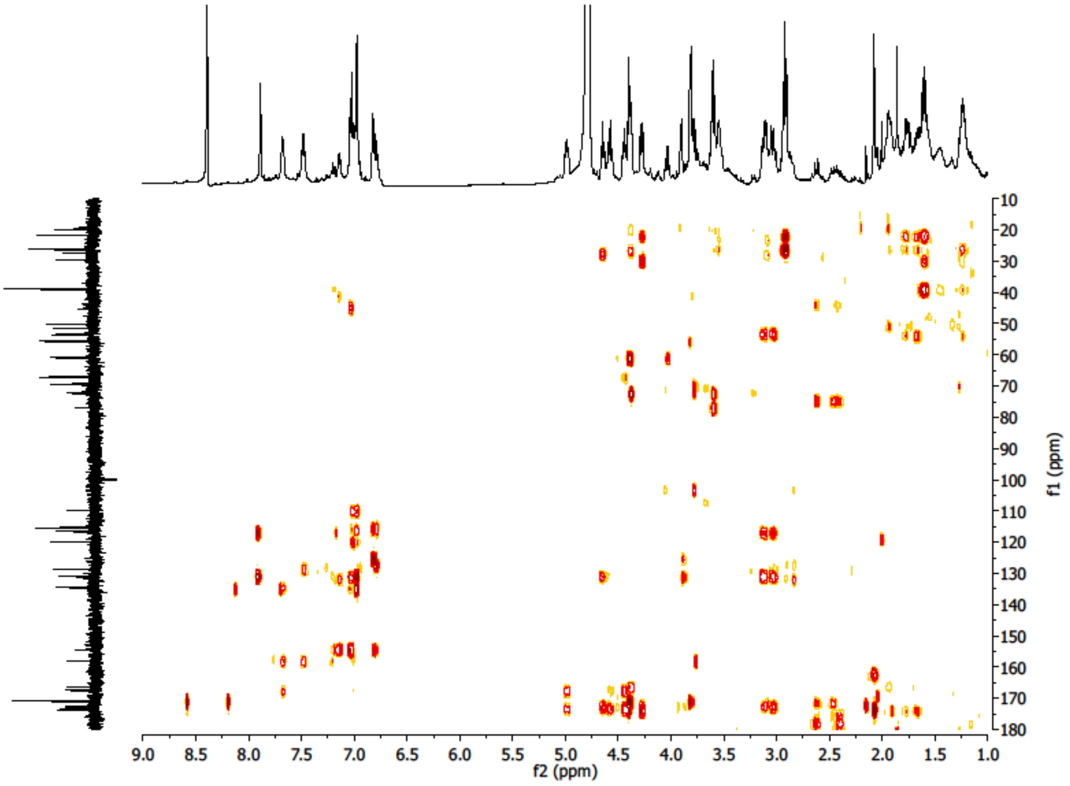
Supplementary Figure 3. COSY spectrum of gobichelin A in D_2O.



Supplementary Figure 4. TOCSY spectrum of gobichelin A in D_2O.



Supplementary Figure 5. HSQC spectrum of gobichelin A in D_2O



Supplementary Figure 6. HMBC NMR spectrum of gobichelin A in D_2O

Supplementary Table 1. Annotations of the gobichelin gene cluster.

Gene	Predicted Function	Length (aa)	Sequence simiarity	Identity/similarity
A	TetR transcriptional regulator	153	YP_004965285 *Stretpomyces bingchenggensis* BCW-1	52/64
B	NH_3 dependent NAD+ synthase	277	ZP_07288305 *streptomyces sp.*C	92/96
C	hypothetical protein	218	ZP_09180682 *Streptomyces sp*. S4	80/89
D	acetyltransferase	156	ZP_06775522 *Streptomyces clavuligerus* ATCC 27064	51/65
E	isochorismatase	219	YP_004801102 *Streptomyces sp. SirexAA-E*	72/81
F	putative ORF	99	No significant similarity to proteins in nr	
G	amidinotransferase	253	ZP_04713053 *Streptomyces roseosporus* NRRL 11379	80/88
H	salicylate synthase	533	ZP-06417730 *Frankia sp.* EUN1f	44/53
I	type II thioesterase	261	ref|YP_481508.1 *Frankia sp.* CcI3	45/59
J	NRPS (T-Cy-A-T)	1181	ref|ZP_07278495.1 *Streptomyces sp*. AA4	58/69
K	2,3-dihydroxybenzoate-AMP ligase	578	ref|YP_001104901.1 *Saccharopolyspora erythraea* NRRL 2338	62/72
L	MbtH	81	emb|CAE53354.1 *Actinoplanes teichomyceticus*	73/80
M	iron siderophore uptake ABC system	354	ref|YP_075115.1 *Symbiobacterium thermophilum* IAM 1486	45/65
N	ABC transporter	279	ref|YP_003339144.1 *Streptosporangium roseum* DSM 43021	70/81
O	ABC transporter	362	ref|ZP_08235100.1 *Streptomyces griseus* XylebKG-1	56/69
P	iron siderophore transporter	278	ref|ZP_08806318.1 *Streptomyces zinciresistens* K42	64/77
Q	enterobactin transporter	331	ref|ZP_10053435.1	46/56

			Amycolatopsis sp. ATCC 39116	
R	NRPS (C-A-T-C-A-T-E)	2583	ref|ZP_06575792.1 pstC *Streptomyces ghanaensis* ATCC 14672	48/59
S	NRPS (C-A-T-C-A-T-TE)	2277	ref|YP_005057340.1 *Granulicella mallensis* MP5ACTX8	50/64
T	L-ornithine 5-monooxygenase	456	ref|ZP_06712113.1 *Streptomyces sp*. e14	67/77

Supplementary Table 2. Peptide identifications matching known NRPS proteins from *Streptomyces sp.* NRRL F-4415.

Peptide Sequence	**Protein Description**	**Accession #**
LGAGDDIPIGTPVAGR	daptomycin biosynthetic protein subunit [*Streptomyces filamentosus*]	AAB96629
AALGDVVMRHEALR	putative NRPS [*Streptomyces bingchenggensis* BCW-1]	ADI09919
GVVVPHSAIVNR	amino acid adenylation domain protein [*Acidovorax avenae subsp*. avenae ATCC 19860]	ADX48474
WTADGVIEYLGR	non-ribosomal peptide synthetase [*Streptomyces noursei*]	BAH68405
ETDLAAYAHQDLPFER	pstC [*Actinoplanes friuliensis*	CAM56770
TVGWFTTVHPVR	putative non-ribosomal peptide synthetase [*Streptomyces collinus*]	CAN89661
LRPVPPGVTGELYIAGAGLAR	PI synthetase 3 and 4 [*Streptomyces pristinaespiralis*]	CBW45647
AGAAYLPVDPGYPADR	amino acid adenylation domain-containing protein [*Frankia sp.* EAN1pec]	YP_001510190
IEPGEIEARLR	amino acid adenylation domain-containing protein [*Pseudomonas putida* GB-1]	YP_001670309
LVAYVVPAAGSTIDGATLR	non ribosomal peptide synthase, antibiotic synthesis	YP_002007886
AAVADLTDR	non-ribosomal peptide synthetase [*Rhodococcus erythropolis* PR4]	YP_002766916
DDVVFGTVTSGR	nonribosomal peptide synthetase [*Teredinibacter turnerae* T7901]	YP_003073479
SADLAAFEHQDLPFER	non-ribosomal peptide synthetase-like protein [*Streptosporangium roseum* DSM 43021]	YP_003339147
VRSADLAAFEHQDLPFER	non-ribosomal peptide synthetase-like protein [*Streptosporangium roseum* DSM 43021]	YP_003339147
FVADPYGPPGTR	non-ribosomal peptide synthetase [*Streptomyces scabiei* 87.22]	YP_003486112

FVADPFGAPGSR	non-ribosomal peptide synthetase [*Amycolatopsis mediterranei* U32]	YP_003765298
LVEILNPAR	non-ribosomal *peptide synthetase [Rhodococcus equi* 103S]	YP_004008264
RPFDLTAESPLR	amino acid adenylation domain protein [*Frankia sp.* EuI1c]	YP_004018396
LTYAELDER	putative non-ribosomal peptide synthetase [*Nocardia farcinica* IFM 10152]	YP_119006
LPLTSNGK	amino acid adenylation [*Pseudomonas syringae pv. syringae* B728a]	YP_234879
VVAVALPR	amino acid adenylation [*Frankia sp*. CcI3]	YP_481555
FSDFAVWQR	non-ribosomal peptide synthetase [*Rhodococcus jostii* RHA1]	YP_700226
IVALAIPR	amino acid adenylation domain-containing protein [*Acidovorax avenae subsp. citrulli* AAC00-1]	YP_972967
IEPGEIEAVLAR	amino acid adenylation domain protein [*bacterium Ellin*514]	ZP_03628586
TTALDAYAHQDIPFER	CDA peptide synthetase I [*Streptomyces roseosporus* NRRL 15998]	ZP_04691538
FDEANELGR	PstC protein [*Streptomyces clavuligerus* ATCC 27064]	ZP_05003829
TGAAHLPVDPGYPADR	amino acid adenylation domain protein [*Streptomyces sp.* ACT-1]	ZP_06276625
LRNGDIEYLGR	Non-ribosomal peptide synthase [*Cylindrospermopsis raciborskii* CS-505]	ZP_06307642
GYLGRPGLTSER	peptide synthetase 3 [*Streptomyces ghanaensis* ATCC 14672]	ZP_06575050
VRETDLAAYAHQDLPFER	linear gramicidin synthetase LgrD [*Streptomyces griseoflavus* Tu4000]	ZP_0731401

Supplementary Table 3. PCR primers used in this work. Primers starting with the same number are reverse translated from the same peptide, and “-F” means forward direction and “-R” means reverse direction. Primers with a “D” are reverse translated using degenerate codons while primers without a “D” are reverse translated using the most likely codons from *Streptomyces coelicolor* A3(2). A3F, A3R, A7F and A7R are degenerate primers for the conserved regions of NRPS A domains.

NAME	SEQUENCE	NAME	SEQUENCE
1F	GACATCCCGATCGGCACC	**1DF**	GAYATCCCSATCGGNAC
1R	GGTGCCGATCGGGATGTC	**1DR**	SGTNCCGATSGGGATRTC
2R	GCAGGGCCTCGTGGCG	**2DR**	CKSAGSGCYTCRTGVCK
3F	CGCACTCCGCCATCGTCA	**3DF**	CSCAYWSSGCSATCGTSAA
4R	ACGCCGTCGGCGGTCCA	**4DR**	ACNCCRTCSGCSGTCCA
5F	TACGCCCACCAGGACCTGCC	**5DF**	TAYGCSCAYCARGAYCTSCC

5R	GGCAGGTCCTGGTGGGCGTA	**5DR**	GGSAGRTCYTGRTGSGCRTA
6R	GTGGACGGTGGTGAACCA	**6DR**	RTGSACSGTSGTRAACCA
7R	GATGTACAGCTCGCCGGTGAC	**7DR**	GCGATRTASAGYTCNCCSGTSAC
8F	CCGGTCGACCCGGGCTAC	**8DF**	TAYCTSCCSGTSGAYCCSG
9R	CTCGATCTCGCCCGGCTC	**9DR**	GCYTCGATYTCNCCSGGYTC
10R	GGCCGGGACGACGTAGGC	**10DR**	SGCSGGSACSACRTASGC
11R	TCGGTCAGGTCGGCGACG	**11DR**	TCSGTSAGRTCSGCSAC
12F	GTCGTCTTCGGCACCGT	**12DF**	GTSGTSTTYGGNACSGT
13F	CTTCGAGCACCAGGACCTGC	**13DF**	GCSTTYGARCAYCARGAYCTSC
16F	GTCGAGATCCTGAACCCGGC	**16DF**	GTSGARATCCTSAAYCCSGC
17F	CGCCCGTTCGACCTGAC	**17DF**	GBCCSTTYGAYCTSACSGC
18F	ACCTACGCCGAGCTGGACGAG	**18DF**	ACSTAYGCSGARCTSGAYGA
19R	TGCCGTTGGAGGTCAGCG		
20F	GTCGCCGTCGCCCTGCC	**20DF**	GTSGCSGTSGCSCTSCC
20R	GGCAGGGCGACGGCGAC	**20DR**	GGSAGSGCSACSGCSAC
21F	TCCGACTTCGCCGTCTGG	**21DF**	SGAYTTYGCSGTSTGGCA
21R	CCAGACGGCGAAGTCGGA	**21DR**	TGCCASACSGCRAARTC
22F	GTCGCCCTGGCCATCCC	**22DF**	GTSGCSCTSGCSATCCC
22R	GGGATGGCCAGGGCGAC	**22DR**	GGGATSGCSAGSGCSAC
23R	TCGATCTCGCCCGGCTC	**23DR**	TCGATYTCNCCSGGYTC
24F	CACCAGGACATCCCGTTCGAG	**24DF**	CAYCARGAYATCCCSTTYGA
24R	CTCGAACGGGATGTCCTGGTG	**24DR**	TCRAASGGGATRTCYTGRTG
25F	TTCGACGAGGCCAACGAG	**25DF**	TTYGAYGARGCSAAYGARC
25R	CTCGTTGGCCTCGTCGAA	**25DR**	GYTCRTTSGCYTCRTCRAA
26F	CCGGTCGACCCGGGCTAC	**26DF**	CCSGTSGAYCCSGGNTA
27R	CGGCCCAGGTACTCGAT		
A3F	GCSTACSYSATSTACACSTCSGG	**A3R**	CCSGASGTGTASATSRSGTASGC
A7F	STACCGSACSGGBGACSTS	**A7R**	SASGTCVCCSGTSCGGTAS

Supplementary Table 4. PCR reactions performed to retrieve the DNA sequences of *gob* gene cluster. For each PCR reaction number, two parallel reactions were performed, using primers designed from the degenerate codons and the most likely codons.

rxn#	Forward Primer	Reverse Primer	rxn#	Forward Primer	Reverse Primer
1	A3F	1R	24	8F	A3R
2	A3F	2R	25	12F	A3R
3	A3F	5R	26	13F	A3R
4	A3F	6R	27	16F	A3R
5	A3F	7R	28	17F	A3R
6	A3F	19R	29	18F	A3R
7	A3F	20R	30	21F	A3R
8	A3F	21R	31	24F	A3R
9	A3F	23R	32	26F	A3R

10	A3F	27R	33	1F	A7R
11	A7F	1R	34	3F	A7R
12	A7F	2R	35	5F	A7R
13	A7F	5R	36	8F	A7R
14	A7F	6R	37	12F	A7R
15	A7F	10R	38	13F	A7R
16	A7F	11R	39	16F	A7R
17	A7F	19R	40	17F	A7R
18	A7F	20R	41	18F	A7R
19	A7F	21R	42	20F	A7R
20	A7F	23R	43	21F	A7R
21	A7F	27R	44	24F	A7R
22	1F	A3R	45	26F	A7R
23	5F	A3R			

Supplementary Table 5. Sequencing of successful PCR reactions. The most similar homolog is based on BLASTX search result.

PCR rxn #	Most Similar Homolog	Length (bp)	Identity/Similarity (%)
Contig of 38, 35, 44, 45	gb|AEK43604.1|non-ribosomal peptide synthetase [Amycolatopsis mediterranei S699]	1470	51/63
Contig of 29 and 41	ref|ZP_08234019.1| amino acid adenylation domain protein [Streptomyces cf. griseus XylebKG-1]	996	63/73
37	ref|YP_004018398.1| amino acid adenylation domain protein [Frankia sp. EuI1c]	707	74/81