Supplementary Material for:

Natural product discovery through improved functional metagenomics in Streptomyces.

Hala A. Iqbal, Lila Low-Beinart, Joseph U. Obiajulu, and Sean F. Brady[1]

Author Affiliation
Laboratory of Genetically Encoded Small Molecules, The Rockefeller University 1230 York Avenue, New York, NY 10065

Corresponding Author: Sean F. Brady

Contact
Laboratory of Genetically Encoded Small Molecules
The Rockefeller University
1230 York Avenue
New York, NY 10065

Phone: 212-327-8280
Fax: 212-327-8281
Email: sbrady@rockefeller.edu

Table of Contents

A. Experimental Section
- I. General Methods....................S1
- II. Identification of Streptomyces Strain for use as Metagenomic Library Host S1
- III. Two-step Library Construction and Screening in *S. albus*....................S2
- IV. Analysis of Phenotypic Screening Hits....................S5

B. Tables
- S1: Conjugation Frequency of Streptomyces strains....................S7
- S2: Gene table of Metatricycloene-producing Clone M13.................... S8
- S3: Carotenoid Identification Data S9
- S4: Metatricyclone NMR Assignment Table....................S10

C. Figures
- S1: Optimization of mating efficiency of library hosted in *E. coli* into *S. albus*...............S11
- S2: Phenotypic hits obtained in Functional Screening of *S. albus* Libraries S12
- S3: Carotenoid pathway hits....................S13
- S4: Metatricyclone 1H NMR Spectrum....................S14
- S5: Metatricyclone 13C NMR Spectrum....................S15
- S6: HPLC analysis of Thioesterase knock out clones....................S16
- S7: Alignment of Metatricycloene KS/KS* with representative KSα....................S17

D. References

Experimental Section

I. General Methods:

Making Streptomyces Spore Stocks: Single Streptomyces colonies were spread using sterile wooden sticks onto the solid media recommended for the strain by NRRL. Plates were incubated at 30 ^{o}C for 10 – 14 days. Each mature plate was flooded with 4 mL dH_2O and scraped using a flame-sterilized spreader. To remove mycelia and agar, the resulting suspension was filtered through an autoclaved 10 ml syringe plugged with cotton wool. This filtrate was centrifuged (3,200 g, 5 minutes) and the spore pellet was resuspended in 10% glycerol for storage at -20 ^{o}C. OD_{450} spectroscopic measurements of the spore stock diluted in dH_2O were used to estimate the number of spores (OD_{450} 0.3 = 10^8 spores mL^{-1}).[1]

Streptomyces Cosmid DNA Extraction: Streptomyces harboring cosmid clones were inoculated into 50 ml TSB liquid media and grown until confluence. 4 ml of culture was pelleted, resuspended in 1.2 ml of P1 buffer (Qiagen) supplemented with 1 mg ml^{-1} Lysozyme and incubated at 37 ^{o}C for 2 hours. DNA was miniprepped using the Qiagen Spin Miniprep kit (Qiagen) according to the manufacturer's protocol with minor modifications: i) 1.2 ml and 1.68 ml of buffers P2 and N3 were used, respectively, and ii) the lysis reaction was allowed to proceed for 5 minutes. One column was used per cosmid prep. Immobilized DNA was washed 2x with 0.5 ml PB buffer and 6x with 0.75 ml PE buffer. DNA was eluted with 40 µl of 50 ^{o}C EB.

Standard Mating into Streptomyces: An overnight culture of *E. coli* ET12567 harboring individual or library clones was diluted 1:100 into fresh LB media and grown to OD_{600} 0.5. *E. coli* cells were pelleted (4,000 g, 20 minutes), washed twice with cold LB and resuspended in LB at 1/100 volume of original LB culture volume. 10^8 Streptomyces spores were added to 0.5 ml Tryptic Soy Broth (TSB) (Oxoid) and heat shocked for 10 min at 50 °C. Spores were spun down (21,130 g, 1 min) and the pellet was mixed with 0.5 ml of washed and concentrated *E. coli* culture. This mixture was centrifuged (21,130 g, 1 min), resuspended in 100 µl of the supernatant and plated on 150 mm plates containing 30 ml ISP4 (Soluble Starch 10 g L^{-1}, K_2HPO_4 1 g L^{-1}, $MgSO_4.7H_2O$ 1 g L^{-1}, NaCl 1 g L^{-1}, $(NH_4)_2SO_4$ 1 g L^{-1}, $CaCO_3$ 2 g L^{-1}, Agar 20 g L^{-1}, Trace salt solution 1 ml L^{-1} added after autoclaving: $FeSO_4.7H_2O$ 0.1 g L^{-1}, $MnCl_2.4H_2O$ 0.1 g L^{-1}, $ZnSO_4.7H_2O$ 0.1 g L^{-1}) + 10 mM $MgCl_2$. Plates were incubated at 30 ^{o}C for 14 hours and then overlaid with 2 ml dH_2O containing 3 mg apramycin (for Streptomyces exconjugant selection) and 1.5 mg nalidixic acid (to eliminate *E. coli*). Plates were allowed to dry at room temperature and then placed at 30 ^{o}C.

Preparation of Electrocompetent *E. coli*: Electrocompetent *E. coli* cells were prepared as previously described with minor modifications[2]. Briefly, an overnight culture of cells grown in YENB media (7.5% Yeast, 8% Nutrient Broth, 37 °C, 250 rpm) was diluted 1:100 into fresh YENB and allowed to grow to OD_{600} 0.7. Cells were pelleted (4,000 g, 20 min) and gently washed twice with 50 ml 4 °C water. Cells were washed with 20 ml of 10% glycerol (4 °C) and resuspended in 4 °C using 1/1000 of the original LB culture volume. 75 µl aliquots of electrocompetent cells were flash frozen using liquid nitrogen and stored at -80 °C.

II. Identification of Streptomyces Strain for use as Metagenomic Library Host:

Assembling a Streptomyces Strain Collection: 16S rRNA sequences of Streptomyces strains were obtained from the Ribosomal Database Project (Release 11, version 3, https://rdp.cme.msu.edu/index.jsp) for phylogenetic and molecular evolutionary analysis. Type strain sequences from individual isolates (size > 1,200 bp, good quality) were downloaded and aligned using the Maximum Composite Likelihood algorithm (MUSCLE) in MEGA (version 5.2.2)[3]. A phylogenetic tree was drawn using the Neighbor-Joining method. 38 strains of Streptomyces were chosen to maximize phylogenetic diversity. Strains were obtained from either ATCC or the Agricultural Research Service Culture (ARCS/NRRL) Collection.

Choosing Streptomyces Strain for use as Metagenomic Library Host: Streptomyces strains were assayed for aptitude as metagenomic library hosts by testing for transformation efficiency and heterologous expression capabilities. To interrogate strains for these characteristics, a test biosynthetic library was constructed from 97 individual eDNA-derived Type II minimal PKS containing cosmids found in previous studies using homology-based screening of metagenomic libraries[4-6]. Each cosmid was electroporated into *E. coli* ET12567. To create

our test biosynthetic library, equal volumes of overnight cultures of individual clones were pooled and this pool was store as a glycerol stock (15%).

The test library and an empty pWEB436 vector control culture were conjugated separately using standard mating conditions into each Streptomyces strain. Exconjugant colonies were allowed to grow for 14 days, following which colonies were counted to determine the conjugation frequency as a proportion of the 10^8 Streptomyces spores used in the mating reaction. Plates from strains displaying conjugation frequencies of $>10^{-6}$ were scraped to make spore stocks to plate on rich media (Table S1). OD_{450} values were used to calculate spores ml^{-1}, and spore stocks were diluted and plated to obtain 1,000 colonies on R5A media plates (Sucrose 100 g L^{-1}, K_2SO_4 0.25 g L^{-1}, $MgCl_2.6H_2O$ 10.12 g L^{-1}, Glucose 10 g L^{-1}, Casamino acids 0.1 g L^{-1}, Yeast Extract 5 g L^{-1}, MOPS 21 g L^{-1}, NaOH 2 g L^{-1}, Agar 20 g L^{-1}, R2YE Trace elements solution 2ml L^{-1}: $ZnCl_2$ 40 mg L^{-1}, $FeCl_3.6H_2O$ 200 mg L^{-1}, $CuCl_2.2H_2O$ 10 mg L^{-1}, $MnCl_2.4H_2O$ 10 mg L^{-1}, $Na_2B_4O_7.10H_2O$ 10 mg L^{-1}, $(NH_4)_6Mo_7O_{24}.4H_2O$ 10 mg L^{-1}).[1] After 14 days of growth, colonies were visually inspected for changes in color compared to the natural variation seen on vector control plates.

Pink-violet and blue colored hits found in all Streptomyces strains were miniprepped and restriction-mapped to confirm that the phenotypes arose from 1 clone each. *S. albus* brown phenotypes were de-replicated by growing 30 brown colonies and a negative control colony from the pWEB436 vector mating plate in 50 ml R5A liquid media for 14 days. 30 ml of cultures were first extracted with 10 ml ethyl acetate, then extracted with 10 ml acidified ethyl acetate (0.5 ml 88% formic acid). Neutral and acidified ethyl acetate extracts were dried down, resuspended in 50 ul methanol and run on Waters LCMS with a C18 analytical column (Xbridge 10 x 150 mm) as follows: 1.5 ml min^{-1} flow rate, solvents dH_2O + 0.1% formic acid and methanol + 0.1% formic acid. 5 minutes at 90:10 dH_2O:methanol, 15 minutes linear gradient from 90:10 dH_2O:methanol to 100% methanol, 5 minutes at 100% methanol, 1 minute from 100% methanol to 90:10 dH_2O:methanol, 4 minutes at 90:10 dH_2O:methanol. Mass Spectrometry data was collected on a Waters Micromass ZQ instrument. Resulting chemotypes were de-replicated by comparing HPLC traces, UV data and masses of clone-specific peaks. To further investigate the heterologous expression capabilities of *S. albus*, 100 random clones from the test biosynthetic library hosted in *S. albus* were grown in liquid R5A, extracted and analyzed as described above.

III. Two-step Library Construction and Screening in *S. albus*

Construction of vector pWEB436: A new *E. coli*-*Streptomyces* shuttle vector (pWEB436) was constructed for this project. This vector was constructed by ligating the blunt end repaired (End-It, Epicentre Biotechnologies) MfeI fragment of pWEB (*E. coli* origin of replication, cos site and ampicillin resistance genes) to the DraI fragment from the integrative *Streptomyces* vector pOJ436 (ΦC31 integration system, apramycin resistance gene and oriT). Cut vector fragments were ligated (FastLink-DNA ligation kit, Epicentre Biotechnologies), transformed into EC100 cells and selected on LB plates containing Ampicillin (100 ug/ml) and Apramycin (75 ug/ml).

eDNA Extraction from Soil: To make an eDNA library, eDNA was first extracted from Texas desert soil (TX, USA) as previously described.[7] Briefly, 250 g of soil was sieved through a 1/8 inch mesh screen to remove debris. 15 ml lysis buffer (100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCL, 1% w/v cetyl trimethyl ammonium bromide, 2% w/v SDS, pH 8.0) was added and inverted to mix. Bottles were incubated at 70 °C for 2 hours, and the lysate was centrifuged at 4,000 g for 10 minutes at 4 °C. The supernatant was centrifuged for 20 minutes to remove any residual debris and eDNA was isolated by isopropanol precipitation as follows: 0.7 volumes of isopropanol was added to the crude lysate. Bottles were gently inverted to mix and incubated at room temperature for 30 minutes. DNA was pelleted by centrifuging at 4,000 g for 30 minutes (4 °C). The pellet was rinsed with 100 ml of 70% ethanol and centrifuged at 4,000 g for 30 minutes (4 °C). After decanting, the pellet was allowed to air dry for 3 hours at room temperature. DNA was resuspended in 10 ml of TE (10 mM Tris, 1 mM EDTA, pH 8) warmed to 50 °C. Gel purification was used to obtain High Molecular Weight (HMW) DNA from the crude soil eDNA prep by running the eDNA on 0.7% ethidium bromide-free agarose gel (100 V for 1 hour then 20 V for 5 hours). The band corresponding to HMW eDNA was cut from the gel and transferred to dialysis tubing (MWCO 10,000) for electroelution (100 V for 3 hours). Electroeluted HMW eDNA was concentrated using an Amicon centrifugal concentrator (MWCO 30,000), rinsed with TE and then resuspended in 1 ml of TE.

Two-step eDNA library construction in *E. coli*:

Step 1: The pWEB436 vector was prepped for cloning by digestion with ScaI (New England Biolabs) and overnight dephosphorylation with Calf Intestinal Phosphatase (New England Biolabs). Blunt end repaired eDNA (End-It, Epicentre Biotechnologies) was ligated to pWEB436 (FastLink DNA ligation kit, Epicentre Biotechnologies). Ligation reactions were packaged into MaxPlax™ Lambda Packaging Extracts (Epicentre). Packaging Extracts were titered by transfecting into *E. coli* EC100 cells grown with 10 mM $MgSO_4$ to OD_{600} 1.0, and plating dilutions on LB plates containing 75 ug/ml apramycin. The number of colonies obtained on titer plates was used to calculate the potential number of library clones per ligation reaction. Library production was scaled to obtain 1.5 million clones. This library was stored in 15% glycerol as three megapools containing 500,000 clones each. Each megapool was constructed in 100 unique aliquots containing 5,000 colonies each.

Step 2: The second step of the library construction protocol involves the removal of empty vector and small-insert cosmids from the library prior to transformation into *E. coli* S17.1 (ATCC 47055) for mating into Streptomyces. Overnight cultures of library megapools in *E. coli* EC100 were midiprepped (Machery-Nagel) and 50 ug of DNA from each megapool was run separately on a 0.7% ethidium bromide-free agarose gel for 2 hours at 80V, followed by 4-6 hours at 40 V. The sides of the gel were cut and stained with ethidium bromide and the gel was reassembled to guide excision of the band corresponding to full-length cosmid DNA. The gel slice containing cosmids was transferred into dialysis membrane tubing (12-14 kD, 25 mm width, 16 mm diameter, Spectrumlabs) containing 8 ml 0.5x TBE Buffer. Electroelution was allowed to proceed for 3 hours at 100V followed by 1 minute at 30V with the electrodes switched to dislodge DNA stuck to the membrane. DNA was concentrated in a 30,000 MWCO concentrator (Pall Corporation) by centrifuging at 4,000 g, rinsed twice with 10 ml TE buffer and then concentrated to a final volume of 0.5 ml. The resulting size-selected library DNA was transformed into *E. coli* S17.1 cells by electroporation. Dilution titers of the transformation reaction were plated on LB with Apramycin (75 ug/ml) and used to calculate the number of reactions needed to obtain 1.5x fold coverage of the library. Plates from the transformation reactions were scraped into 1 ml LB and pooled to make 15% glycerol stocks of library megapools in S17.1.

Optimization of Mating into *S. albus*: Using the standard *E. coli*-Streptomyces mating protocol described in the general experimental methods section, the conjugation frequency into *S. albus* was too low (10^{-6}) to shuttle a large (>$1x10^6$ membered) cosmid library using a manageable number of mating reactions. We therefore set out to optimize *S. albus* mating conditions as follows (Figure S1):

1. Matings were carried out on different media: ISP4, R5A, MS (20 g L^{-1} Mannitol, 20 g L^{-1} Soya flour, Tap water), and AS-1 (Yeast Extract 1 g L^{-1}, L-Alanine 0.2 g L^{-1}, L-Arginine 0.2 g L^{-1}, L-Asparagine 0.5 g L^{-1}, Soluble Starch 5 g L^{-1}, NaCl 2.5 g L^{-1}, Na_2SO_4 10 g L^{-1}, Agar 20 g L^{-1}). ISP4 consistently produced the highest number of exconjugants.
2. Mating reactions were plated on ISP4 media containing different concentrations of $MgCl_2$. Increasing the concentration of $MgCl_2$ to 30 mM resulted in highest conjugation rates.
3. Mating were carried out using different ratios and amounts of Streptomyces and *E. coli*. The optimal amount and ratio of *S. albus* and *E. coli* for mating were found to be 2.5 x 10^9 *S. albus* spores to 1.6 x 10^8 *E. coli* cells per mating reaction.
4. To test the efficiency of conjugation using different *E. coli* strains, size-selected library DNA was transformed not only into *E. coli* S17.1 but also *E. coli* ET12567 [selection: chloramphenicol (12.5 ug/ml), kanamycin (30 ug/ml), apramycin (75 ug/ml)]. For the triparental mating using the *E. coli* helper strain DH5α (pRK2013), the TX library in EC100 and DH5α (pRK2013) cells were mixed in equivalent amounts and used in the improved Streptomyces mating protocol. S17.1 showed the highest conjugation efficiency.

The cumulative effect of these changes was a 300-fold increase in conjugation frequency (~7,000 exconjugants/mating reaction).

Construction and Phenotypic Screening of the TX *S. albus* library: The 1.5 million membered high-quality TX library was mated from *E. coli* S17.1 into *S. albus* using optimized mating conditions. Library conjugation plates were incubated at 30 °C and screened visually for phenotypic variation starting 7 days after the mating reaction. 48 colored hits were picked from the conjugation plates and struck on to fresh ISP4 plates. To confirm that the observed color phenotype was conferred by the eDNA cosmid, colonies that retained the colored phenotype when re-struck were grown in 50 ml TSB. DNA was extracted from each culture using the

Streptomyces cosmid DNA extraction protocol described above. DNA was retransformed into *E. coli* S17.1 cells, re-mated into *S. albus* and the cosmid was considered a phenotypic hit if the resulting exconjugants produced color.

IV. Analysis of Phenotypic Screening Hits

Sequencing and Bioinformatic Analysis of color-conferring cosmids: Each unique color-conferring cosmid was sequenced on the IonTorrent Personal Genome Machine platform (Life Technologies). Reads were assembled using Newbler GS De Novo Assembly Software (version 2.6, Roche). For clones that assembled with gaps, primers were designed from the ends of contigs to sequence all missing segments by primer walking. Complete cosmid sequences were annotated as follows: open reading frames were identified using Meta GeneMark software (http://exon.gatech.edu/GeneMark/meta_gmhmmp.cgi) and gene functions were predicted by homology searches to the BLAST (blast.ncbi.nlm.nih.gov) and pfam (pfam.xfam.org) databases. Sequences of unique, reproducible hits J10, L10, M13, N1, P1, P8, P11, R15 and R16 were deposited in GenBank (Accession numbers KX523612 – KX523620). Cosmid sequences were also analyzed using antiSMASH[8] to search for relationships to known biosynthetic clusters. Clones P1 and P8 were found to contain identical gene clusters and a small region of overlapping genomic sequences.

Carotenoid Isolation and Identification: Spore stocks of *S. albus* harboring carotenoid clusters were used to start liquid cultures in R5A media alongside a pWEB436 vector control culture. Cultures were grown at 30 °C with 200 rpm shaking for 10 days. To extract carotenoids, cells were pelleted (4,000 g, 20 min) then washed with water and methanol. The pellet was then resuspended in 0.5 volumes of acetone. After 3 hr, crude extracts were filtered using filter paper (Whatman #3), dried and resuspended in 0.002 volume acetone. All carotenoid samples were stored in the dark at -20 °C.

To identify carotenoids, extracts were run on a Waters LCMS with a C18 column (Xbridge 4.6 x 150 mm): 1.5 ml min^{-1} flow rate for 30 min using a 92% Acetonitrile, 6% Methanol, 2% Propanol isocratic solvent system. β-carotene (Sigma) and Lycopene (Sigma) were used as standards. Low-resolution mass spectrometry data was obtained on Waters Micromass ZQ instrument. For high-resolution mass spectrometry data, clone-specific peaks were purified on an Agilent Technologies 1200 Series HPLC (Waters XBridge C18 column 10 x 150 mm) using the solvent conditions described above. HRMS data was collected using Waters LCT Premier XE mass spectrometer (Memorial Sloan-Kettering Institute).

Isolation and Purification of Metatricycloene:

Small-scale analytical analysis: To analyze the small molecule produced by clone M13, a 50 ml culture was grown (30 °C, 200 rpm shaking, 21 days) in liquid media R5A alongside pWEB436 negative control. 30 ml of cultures were extracted with 15 ml neutral ethyl acetate, followed by 15 ml acidified ethyl acetate (1 ml 88% Formic Acid). Extracts were dried and resuspended in methanol using 0.002 of the original culture volume. Extracts were analyzed using the following HPLC conditions: C18 column (Xbridge 10 x 150 ml) 1.5 ml min^{-1}, solvents dH_2O and methanol with 0.1% formic acid, 5 min at 90:10 dH_2O:methanol, 15 min linear gradient from 90:10 dH_2O:methanol to 100% methanol, 5 min at 100% methanol. Low resolution Mass Spectrometry data was collected on a Waters Micromass ZQ instrument.

Large-scale isolation: To obtain large amounts of compound for structure determination, 6L of culture was grown in multiple 125 ml baffled flasks containing 50 ml of R5A liquid media. After 21 days (30 °C, 200 rpm) cultures were pooled and extracted using neutral ethyl acetate as described above. The crude extract was partitioned using the following conditions: CombiFlash (Teledyne), 50 g RediSep C18 column, 40 ml min^{-1} flow rate, dH_2O:methanol (0.1% formic acid), 10 min 90:10, 30 min linear gradient from 90:10 to 100% methanol, 10 min 100% methanol. Fractions containing metatricycloene were pooled and metatricycloene (retention time = 16 mins) was purified from the combined fraction using the following HPLC conditions: Agilent Technologies 1200 Series HPLC, Waters XBridge C18 column 10 x 150 mm, 3.5 ml/min 70:30 dH_2O:acetonitrile (0.1% formic acid) isocratic. HRMS data was collected using Waters LCT Premier XE mass spectrometer. NMR data was collected on a Bruker 600 MHz instrument.

Phylogenetic Analysis of M13 Di-domain Ketosynthase: The set of 500 closest BLAST hits to the predicted M13 KS:KS* protein was downloaded and curated to remove single domain KS sequences <650 amino acids

as well as redundant sequences from strains of the same bacterial species. Sequences were aligned using the Maximum Composite Likelihood algorithm (MUSCLE) in MEGA (version 5.2.2)[3] and a phylogenetic tree was drawn using the Neighbor-Joining method. Representative gene clusters from each branch of the resulting phylogenetic tree were examined for polyene natural product biosynthetic potential. We defined this as the presence of a di-domain ketosynthase, acyl carrier protein, transacylase, ketoreductase and dehydratase genes.

Transposon Mutagenesis of M13 clone: A transposon library was generated using the EZ-Tn5 <kan-2> kit (Epicentre). Using PCR and primers flanking each gene of interest (Cytochrome P450, FAD-oxidase, β-carotene monooxygenase), individual clones of this library were screened for oxidase gene knockout mutants. Briefly, *E. coli*-based clones were picked into 50 ul LB (Apramycin 50 ug ml^{-1}, Kanamycin 30 ug ml^{-1}), grown overnight and screened using the following touchdown PCR conditions: denaturation (95 °C, 2 min), 10 touchdown cycles (95 °C, 30s; 69 °C, 30s, -1 °C per cycle; 72 °C, 40s), 29 standard cycles (95 °C, 30s; 59 °C, 30s; 72 °C, 2.5 min) and a final extension (72 °C, 7 min). As the transposon is 1.2 kb in size, mutants with transposons inserted in each oxidase gene were identified by the appearance of amplicons 1.2 kb larger than predicted from the M13 cosmid sequence. Mutants were transformed into *E. coli* S17 and conjugated into *S. albus* using the standard mating protocol. Exconjugants were grown in R5A, extracted using ethyl acetate, and extracts were analyzed by LCMS as described above.

Table S1. Conjugation frequency of Streptomyces strains.

Streptomyces Strain	**Conjugation Frequency**
S. coelicolor	10^{-4}
S. bikiniensis	10^{-4}
S. vinaceusdrappus	10^{-4}
S. griseoviridus	10^{-4}
S. lividans	10^{-4}
S. tendae	10^{-4}
S. malachiticus	10^{-5}
S. coerulescens	10^{-5}
S. coeruleorubidus	10^{-5}
S. albus	10^{-6}
S. violatus	10^{-6}
S. viridochromogenes	10^{-6}
S. antibioticus	10^{-7}
S. peucetius	10^{-7}
S. pseudovenezulae	10^{-7}
S. venezulae	10^{-7}
S. tauricus	10^{-7}
S. spectabilis	10^{-7}
S. griseus	10^{-7}
S. capoamus	10^{-7}
S. nodosus	10^{-7}
S. bottropenis	10^{-7}
S. roseosporus	10^{-7}
S. albidoflavus	10^{-7}
S. ghanaensis	10^{-7}
S. avermitilis	10^{-7}
S. alboniger	10^{-7}
S. achromogenes	10^{-7}
S. globisporous	10^{-7}
S. melanosporalaciens	10^{-7}
S. almquisti	10^{-7}
S. fumanus	10^{-7}
S. spheroides	10^{-8}
S. longispuroflavus	10^{-8}
S. fradiaei	10^{-8}
S. hygroscopicus	0
S. platensis	0
S. lincolnensis	0

Table S2. Gene table for M13 Clone

ORF	Size (aa)	Annotation	Organism	% Identity
a	400	Trehalose-phosphate synthase	*Dactylosporangium aurantiacum*	80
b	299	Thiamine biosynthesis protein	*Kibdelosporangium sp.*	59
1	448	Cytochrome P450	*Kibdelosporangium sp.*	70
2	461	MtcA FAD-linked oxidase	*Kribbella catacumbae*	66
3	457	Tryptophan synthase	*Kibdelosporangium sp.*	76
4	170	MtcB Thioesterase Type II	*Thermomonospora curvata*	37
5	142	MtcC 4'-phosphopantetheinyl transferase	*Thermomonospora curvata*	41
6	309	MtcD ACP S-malonyltransferase	*Streptomyces sp.*	61
7	251	MtcE 3-ketoacyl-ACP reductase	*Kibdelosporangium sp.*	70
8	95	MtcF Acyl carrier protein	*Streptomyces sp.*	75
9	169	MtcG β-hydroxyacyl-ACP dehydratase	*Streptomyces sp.*	77
10	311	MtcH Aminomethyltransferase[1]	*Thermomonospora curvata*	50
11	809	MtcI Ketosynthase KS/KS*	*Kibdelosporangium sp.*	71
12	136	Hypothetical protein	*Acidobacteriaceae bacterium*	35
13	247	Short chain dehydrogenase	*Streptomyces sp.*	60
14	356	Amidohydrolase	*Mesorhizobium sp*	48
15	127	MtcJ 4'-phosphopantetheinyl transferase	*Sphingomonas sp.*	40
16	605	MtcK Asparagine synthase	*Kibdelosporangium sp.*	73
17	370	Hypothetical protein	*Kibdelosporangium sp.*	56
c	469	β-carotene monooxygenase	*Saccharopolyspora erythraea*	57
d	223	Acetyltranferase	*Kibdelosporangium sp.*	64
e	252	Hypothetical protein	*Kibdelosporangium sp.*	44
f	402	Cytochrome P450	*Streptomyces actuosus*	47
g	363	Aminotransferase	*Kibdelosporangium sp.*	54
h	78	Hypothetical protein	*Streptomyces bicolor*	58
i	353	Hypothetical protein	*Microbispora rosea*	48
j	183	PadR transcriptional regulator	*Actinoplanes friuliensis*	69
k	175	Membrane protein	*Hamadaea tsunoensis*	44
l	319	Membrane protein	*Catelliglobosispora koreensis*	48
m	210	Membrane protein	*Catelliglobosispora koreensis*	68
n	402	Hypothetical protein	*Streptomyces viridochromogenes*	37

[1] This gene is present in other characterized di-domain KS/KS* pathways. Methylated derivative not seen in LCMS analysis of the Metatricycloene-producing clone.

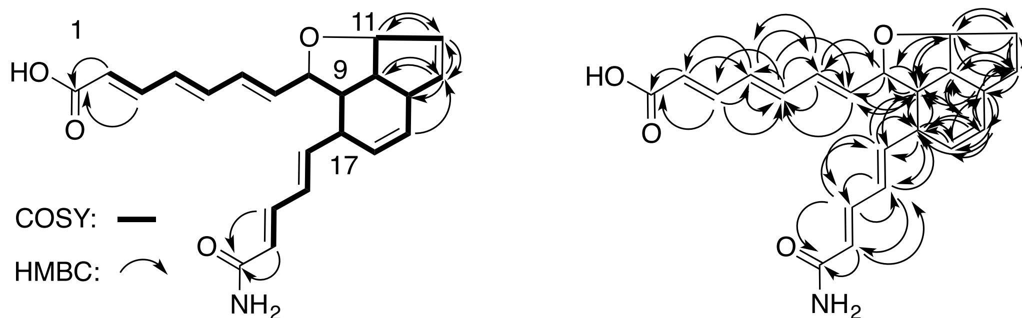
Table S3. Analytical data of carotenoids produced by colored clones

Compound	Name	Clone	LR/HR Mass	Formula	UV Maxima Reported	UV Maxima Seen
2	Dihydroxylycopene	R16	572.4599	$C_{41}H_{60}O_2$	445, 474, 506	447, 473, 504
3	Rhodopin	R16	554.4481	$C_{40}H_{58}O$	445, 472, 504	448, 476, 507
4	Isorenieratene	J10 R15	528.3747 528.3766	$C_{40}H_{48}$	(430), 453, 480	(432), 453, 482
5	Lycopene	J10 L10 N1 P1 P8 P11 R15 R16	536.84 536.83 536.84 536.87 536.89 536.91 536.83 536.83	$C_{40}H_{56}$	448, 474, 505	448, 476, 505
6	β-carotene	J10 N1 P1 P8 P11 R15	536.84 536.84 536.87 536.89 536.91 536.83	$C_{40}H_{56}$	425, 447, 475	429, 449, 480
7		P11	564.4043	$C_{40}H_{52}O_2$		462, 474, 491
8		P11	552.4327	$C_{40}H_{56}O$		(447), 458, 479
9		P11	550.4230	$C_{40}H_{54}O$		458, 474, 491

Table S4. Metatricycloene NMR Table in CD_3OD

C/H	δ_H	Multiplicity (*J*, Hz)	δ_C	HMBC
1			169.0	
2	5.91	d (15.1)	121.0	C1, C4
3	7.23	dd (15.2, 11.2)	144.8	C1, C4, C5
4	6.29	dd (10.9, 3.70)	130.4	C2, C6
5	6.54	dd (14.9, 11.0)	139.8	C3, C4, C6, C7
6	6.32	dd (10.8, 3.80)	132.6	C4, C5, C7, C8
7	5.73	m	137.5	C5, C9
8	4.31	dd (10.8, 8.0)	84.5	C6, C7, C9
9	1.65	q (11.0)	51.8	C7, C8, C11, C17, C18
10	2.73	ddd (12.1, 12.1, 6.5)	45.3	C9, C13, C15
11	5.01	d (6.8)	85.7	C12, C13
12	5.81	m	134.5	C11, C13
13	5.81	m	133.4	C10, C12, C14
14	3.19	m	45.37	C9, C13, C15, C16
15	5.71	m	133.2	C13, C16, C17
16	5.95	m	121.0	C14, C17
17	3.07	t (10.5)	42.5	C9, C15, C16, C18, C19
18	5.73	t (12.9)	143.8	C15, C17, C20
19	6.19	dd (15.2, 10.9)	128.7	C17, C21
20	7.01	d (15.2, 11.2)	141.2	C18, C19, C22
21	5.91	d (15.1)	122.4	C19, C22
22			167.6	
N22[1]	7.41	s		C22

[1] Amine exchangeable proton: δ_H, δ_C and HMBC correlations in DMSO-*d6*



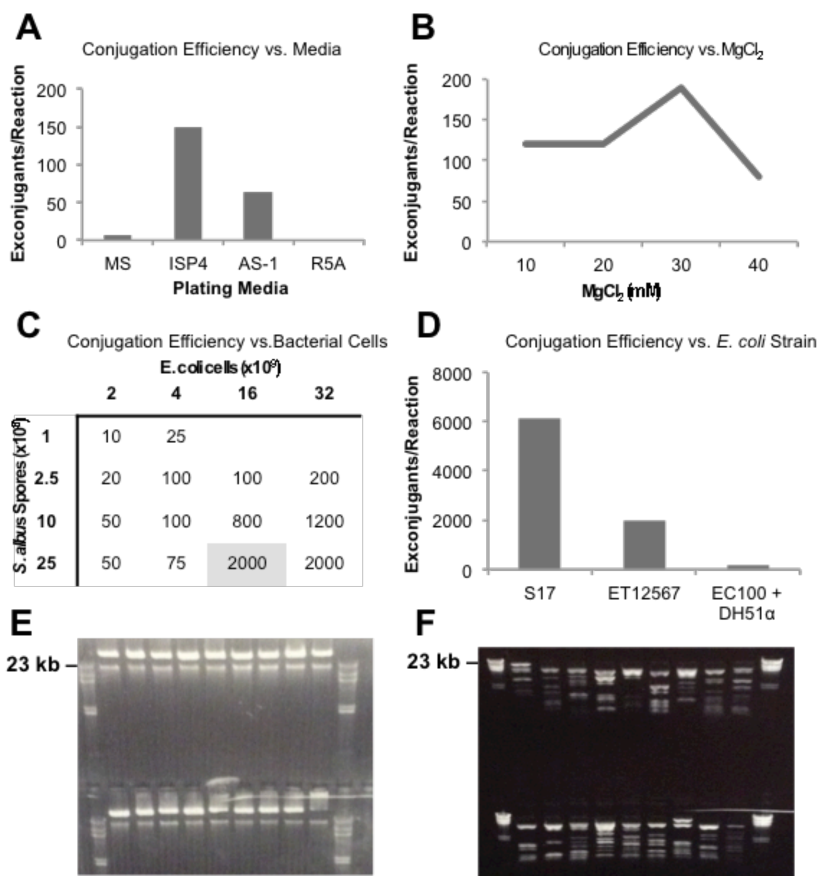


Figure S1. Optimization of *E. coli* and *S. albus* mating. (A) The dependence of conjugation efficiency on the media used for mating. (B) The dependence of conjugation efficiency on the $MgCl_2$ concentration used in ISP4 plates. (C) The dependence of conjugation efficiency on *E.coli*:*S. albus* cell ratio and amount. (D) The dependence of conjugation efficiency on the *E. coli* strain used. (E) Cosmid DNA found in 18 randomly picked exconjugants from the metagenomic library host in *S. albus*. All clones are expected size for full-length cosmid clones (40-50kb), confirming the quality of our library. (F) Restriction maps (NotI, XhoI) of the same 18 randomly picked exconjugants. All show unique restriction patterns.

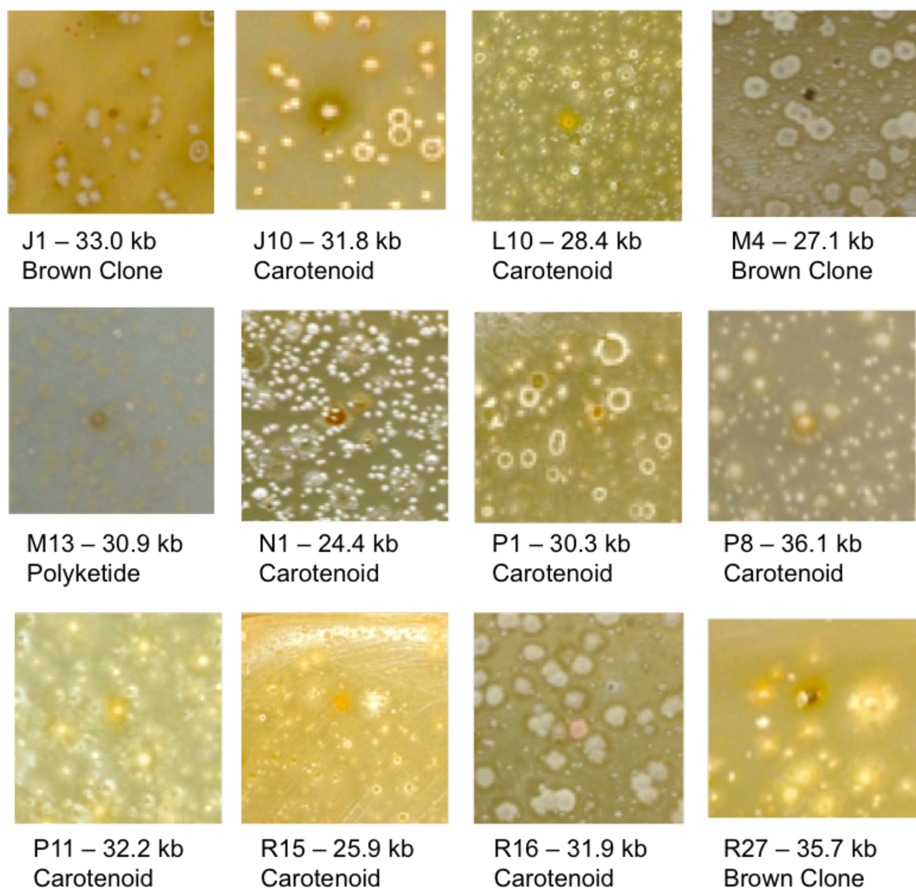


Figure S2. Phenotypes of reproducible hits obtained in color-based functional screen of *S. albus* based metagenomic library

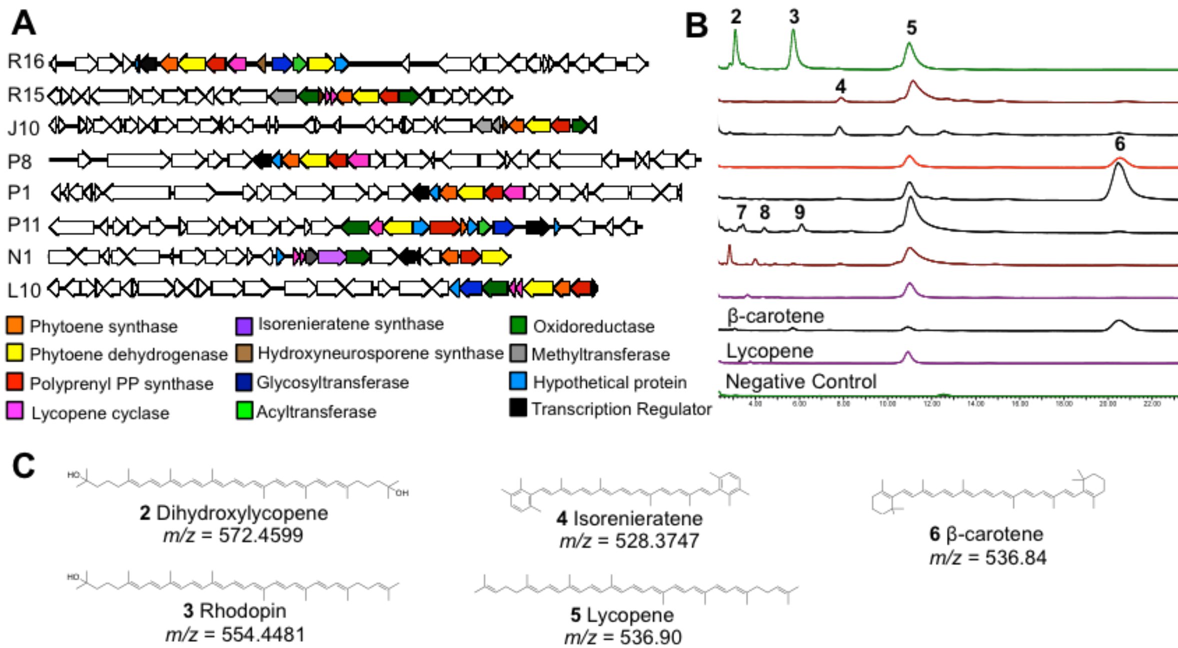


Figure S3: Carotenoids obtained from color-based functional screening of *S. albus* metagenomic library. (A) Gene clusters with core carotenoid and tailoring enzymes highlighted. (B) LCMS traces of culture broth extracts from carotenoid-producing clones (C) Predicted carotenoids structures for peaks seen in B.

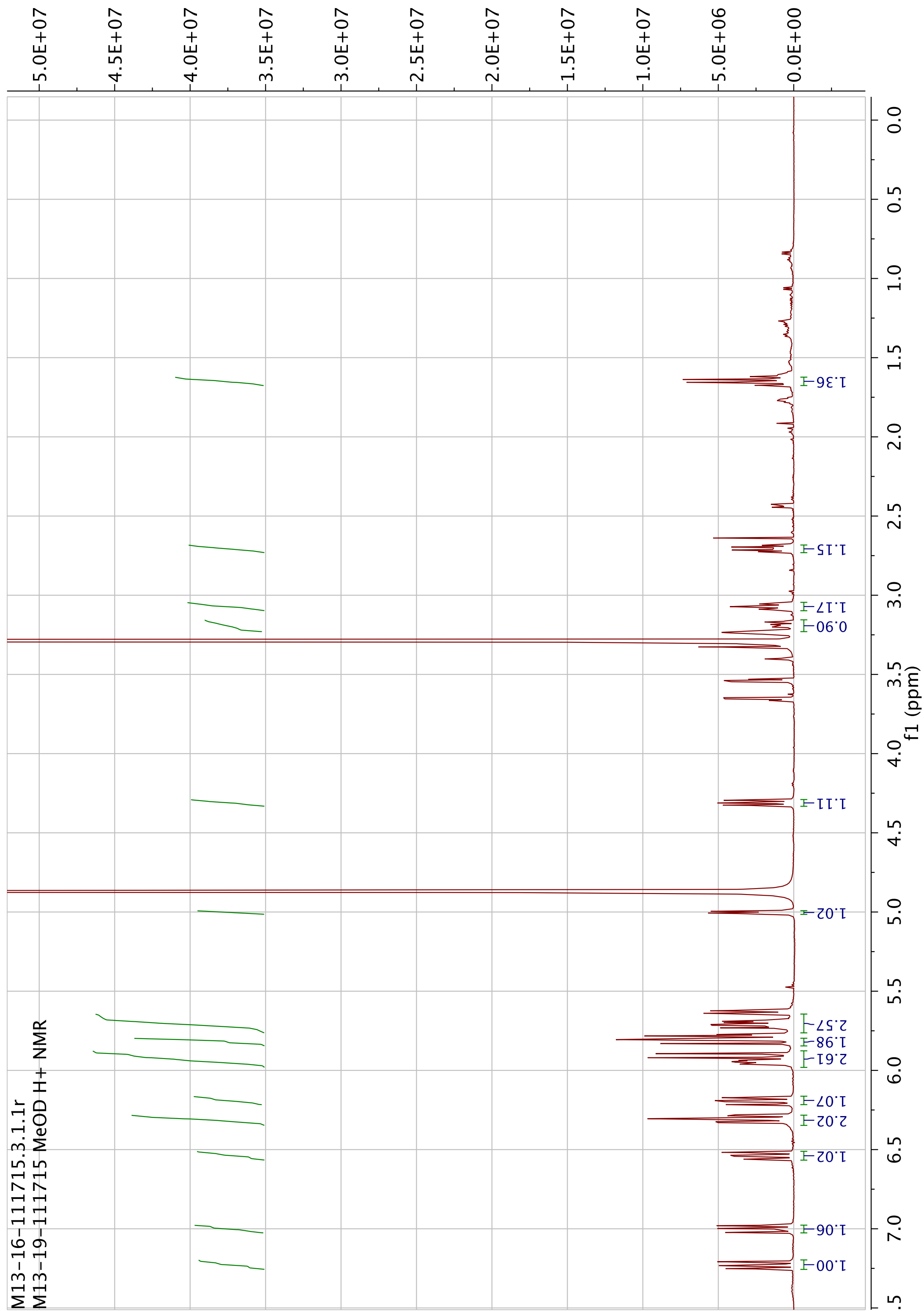


Figure S4. Metatricycloene ^{1}H NMR in CD_3OD

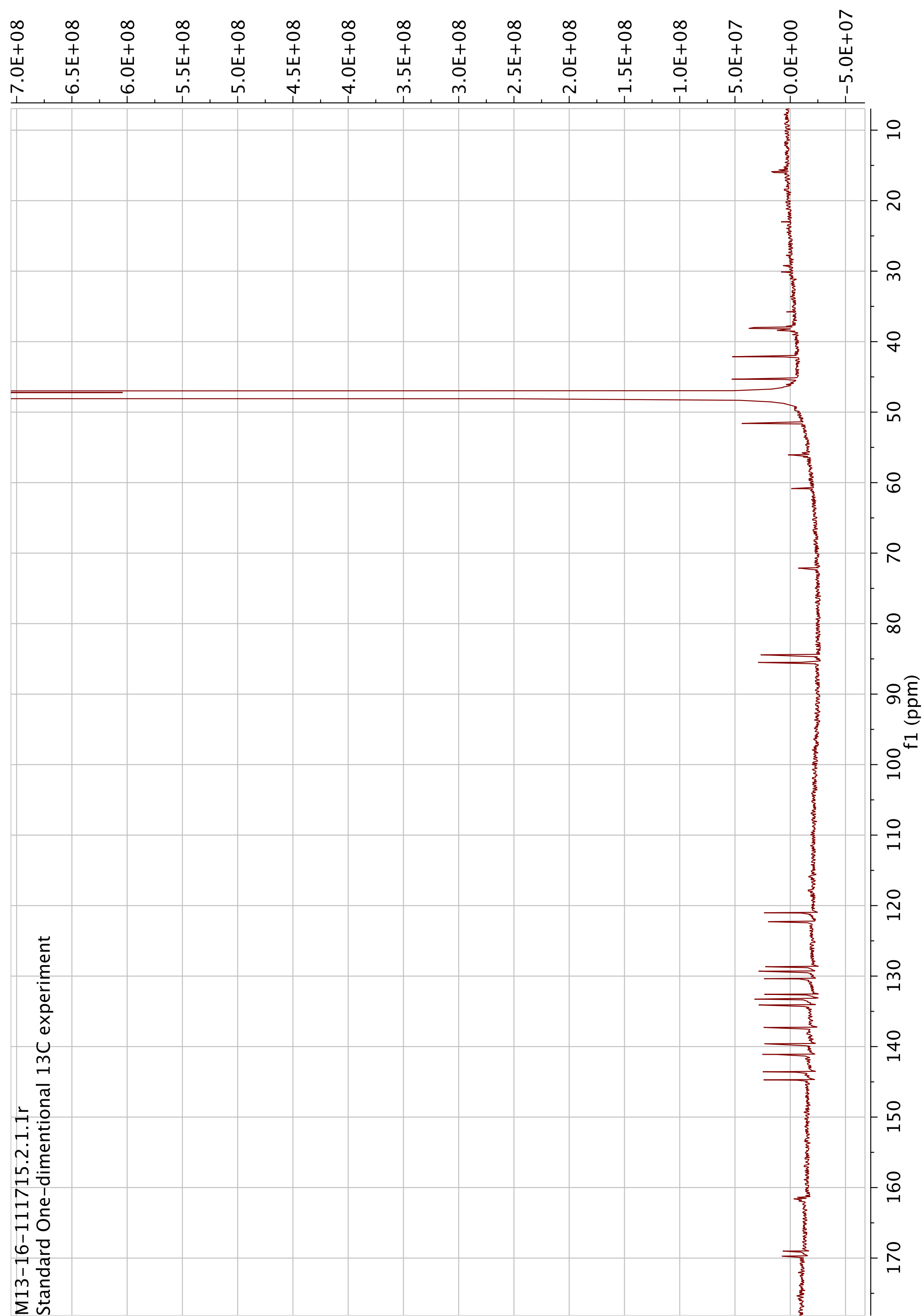


Figure S5. Metatricycloene ^{13}C NMR in CD_3OD

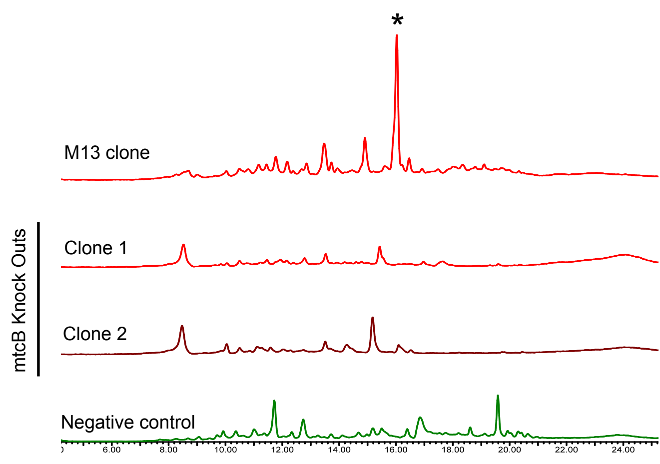


Figure S6. HPLC UV (270 nm) traces of culture broth extracts from M13 and thioesterase gene (mtcB) knockout mutants.

```
Arimetamycin KSa     117 EYVVVSDGGRVWQVDERYAVPHLFNYFVPSSLGVEVAWTVGAEGPVAVVSTGCTSGLDSV 176
Jadomycin KSa        117 EYRVVSDGGRLDLVDHAYAVPHLYDYMVPSSFSAEVAWAVGAEGPNTVVSTGCTSGLDSV 176
Tetracenomycin KSa   121 EYARVSEGGSRWLVDHTLAVEQLFDYFVPTSICREVAWEAGAEGPVTVVSTGCTSGLDAV 180
M13 KS               116 FHWELLAGGIASTGRHHLLVYPLY------TAADAVSAAYGAKGPKVVISNACAAGANSI 169
M13 KS*              115 LTETIGTQG-----PDRVNPRLFPNSVMNAAAGHACLSLQLRGPLSTLATGCASGVAGL 168
                         EY  VS.GG .   VDH  AVP.LF Y VP.S...EVAW .GAEGP..VVSTGCTSGLDSV

Arimetamycin KSa     296 AAVLPEDIDYVNAHGSGTVQNDRHETAAFKRSLKEHAYAVPISSIKSMIGHSLGAIGSLE 355
Jadomycin KSa        297 ARMNPTEIDYINAHGSGTKQNDRHETAAFKKSLGDHAYRTPVSSIKSMVGHSLGAIGSIE 356
Tetracenomycin KSa   300 ARRTGDDLHYINAHGSGTRQNDRHETAAFKRSLGQRAYDVPVSSIKSMIGHSLGAIGSLE 359
M13 KS               281 AGLTPADVDYVNGHGTGTPANDSAETKAADALFGGSP--PPMSSTKSQIGHTLGAAGAIE 338
M13 KS*              281 GEVTADQVGTVYGDARGTATLDLAEARAVGHVWAPGA--VRLANIGGQVGHLHSTTSLLS 338
                         A  .TP.D.DYVNAHGSGT QNDRHETAAFK.SLG .AY VP.SSIKSMIGHSLGAIGSLE
                                      *
```

Figure S7. Alignment of representative Type II Polyketide KSα and Metatricycloene KS/KS* proteins. Catalytic residues are highlighted in grey. The histidine missing in KS* is starred.

References

(1) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; John Innes Foundation, 2000.
(2) Sharma, R. C.; Schimke, R. T. *BioTechniques* **1996**, *20*, 42.
(3) Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. *Molecular biology and evolution* **2011**, *28*, 2731.
(4) Feng, Z.; Kallifidas, D.; Brady, S. F. *Proc Natl Acad Sci U S A* **2011**, *108*, 12629.
(5) Feng, Z.; Kim, J. H.; Brady, S. F. *Journal of the American Chemical Society* **2010**, *132*, 11902.
(6) King, R. W.; Bauer, J. D.; Brady, S. F. *Angewandte Chemie (International ed. in English)* **2009**, *48*, 6257.
(7) Brady, S. F. *Nat Protoc* **2007**, *2*, 1297.
(8) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res* **2011**, *39*, W339.