Supplementary Information for

Temporal and fluoride control of secondary metabolism regulates cellular organofluorine biosynthesis

Mark C. Walker, Miao Wen, Amy M. Weeks, and Michelle C. Y. Chang*

Materials and Methods

Supplementary Results and Discussion

Materials and Methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth, yeast extract, malt extract, glycerol, and triethylamine were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 25 phenol: 24 chloroform: 1 isoamyl alcohol, magnesium chloride hexahydrate, kanamycin, acetonitrile, ethylene diamine tetracetic acid disodium dihydrate (EDTA), and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Streptomycin sulfate, ammonium persulfate, sodium fluoroacetate, CoA trilithium salt (CoA), acetyl-CoA, methylthioadenosine (MTA), dihydroxyacetone phosphate (DHAP), oxaloacetic acid (OAA), phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), adenosine monophosphate sodium salt (AMP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), nicotinamide adenine dinucleotide phosphate ($NADP^+$), sodium fluoride, myokinase, pyruvate kinase, lactate dehydrogenase, isocitrate dehydrogenase, lysozyme, DOWEX 50WX8-100 ion exchange resin, Q Sepharose Fast Flow ion exchange resin, magnesium sulfate heptahydrate, zinc chloride, poly(ethyleneimine) solution (PEI), sodium citrate, 5-fluorouracil, fluorocitric acid barium salt, sodium sulfide, ammonium iron (II) sulfate hexahydrate, β-mercaptoehtanol, sodium phosphate dibasic hepthydrate, D-mannitol, apramycin sulfate salt, hygromycin, and N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED) were purchased from Sigma-Aldrich (St. Louis, MO). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Acrylamide/Bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), Platinum Taq High-Fidelity polymerase (Pt Taq HF), PCR2.1-TOPO TA cloning kit, and Trizol™ reagent were purchased from Invitrogen (Carlsbad, CA). DNase and Spectra multicolor low range protein ladder were purchased from Fermentas (Glen Burnie, Maryland). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4 °C for immediate usage or -20 °C for longer term usage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators and 5,000 MWCO regenerated cellulose ultrafiltration membranes were purchased from

Millipore (Billerica, MA). Bacto™ Agar was purchased from BD (Sparks, Maryland) Soy flour was purchased from Berkeley Bowl (Berkeley, California). Deuterium oxide was purchased from Cambridge Isotope Laboratories (Andover, MA). Streptomyces ReDirect kit was purchased from the Biotechnology and Biological Research Council (Wiltshire, UK). Illumina Genomic DNA Sample Prep Kit, Paired-End Adapter Oligo Mix, and PE-A1/PE-A2 primers were purchased from Illumina (San Diego, CA).

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA). *E. coli* DH10B-T1^R was used for plasmid construction and BL21(de3) was used for heterologous protein production. *E. coli* GM272 [*1*] harboring the non-transmissible, *oriT*-mobilizing plasmid pUZ8002 [*2*] was used for conjugative plasmid transfer into *S. cattleya.*

Preparation of paired-end genomic libraries for Illumina sequencing. *S. cattleya* genomic DNA was isolated using a modified salting-out protocol [*3*] from a 48 h culture grown in YEME (30 mL) [*3*] supplemented with 0.5% glycine at 30°C. The cell pellet was collected by centifugation at 15,316 × *g* for 15 min and resuspended in SET buffer (5 mL; 75 mM NaCl, 25 mM EDTA, 20 mM Tris-HCl pH 7.5). At this time, lysozyme (2 mg) was added and the solution was incubated at 37 °C for 1 h. The solution was incubated for an additional 2 h at 55°C following addition of proteinase K (2.8 mg) and 10% SDS (0.6 mL). Next, sodium chloride (5.0 M, 2 mL) and chloroform (5 mL) were added after cooling on ice and rewarming to 37°C. After mixing for 30 min at room temperature, the aqueous and organic layer were separated by centrifugation for 20 min at 5,000 × *g*. The genomic DNA was precipitated from the aqueous layer by addition of 0.6 volumes isopropanol (4.5 mL) and collected by spooling, washed with ethanol (70% v/v), air dried, and resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 (2.0 mL) by heating overnight at 55°C. The purity of the genomic DNA was assayed by amplification of the 16s rRNA gene sequence with Platinum Taq HF DNA polymerase using the EUB R933/EUB R1387 [*4*] and U1 F/U1 R [*5*] primer sets and sequencing 10 individual colonies from each PCR reaction after insertion into pCR2.1-TOPO. All 20 16s rRNA sequences were found to match with those from *S. cattleya* A and B from the NCBI database. The genomic library for Illumina sequencing was prepared using the Illumina Genomic DNA Sample Prep Kit (Illumina) with some modifications to the manufacturer specifications. Genomic DNA (25 μg) was nebulized with Ar at 35 psi for 2 min in nebulizing buffer (700 μL). The fragmented DNA was end-polished before ligation to the Paired-End Adapter Oligo Mix and separation on an agarose gel (2%). Gel bands corresponding to fragment sizes of 150-200 bp, 2 × 250-300 bp, and 450-500 bp according to the 100 bp ladder (Fermentas) were excised and used as a template (1.5

μL) for amplification with Phusion DNA polymerase with the P1 and P2 primers (13-15 cycles, 200 μL reaction) after gel purification. The libraries were then concentrated using the Qiagen PCR Purification Kit and eluted in Buffer EB (30 μL). The libraries were analyzed on an Agilent 2100 BioAnalyzer (Santa Clara, CA) using a DNA 1000 Series II chip (Agilent) (actual average sizes, 200, 250, 292 and 465 bp; Figure S1).

Library sequencing and read trimming. Two paired-end libraries (284 bp, insert size) were sequenced on an Illumina Genome Analyzer (Hayward, CA) at the Vincent J. Coates Genome Sequencing Center at UC Berkeley using the Cluster Generation Kit v1 (Illumina) and Sequencing Kit v1 (Illumina) to generate 45 base forward and reverse reads (1.5 GB). An additional two paired-end libraries (163 and 318 bp, insert sizes) were sequenced at the UC Davis Genome Center using the Cluster Generation Kit v2 (Illumina) and Sequencing Kit v3 (Illumina) to generate 61 base forward and 45 base reverse reads (2.7 GB). Libraries were titrated by PCR against a known standard before sequencing. Data processing was performed on a Linux PC with dual 64 bit 2.66 GHz quad-core processors and 32 Gb of RAM. Prior to assembly, the reads were trimmed based on the average Illumina quality score for bases in a sliding window one tenth the length of the read. The reads were truncated to the base where the average quality score in that window dropped below 20. Reads were trimmed to 20 bases if it would have been truncated to <20 bases based on this criterion. These reads were effectively eliminated from the assembly as hash lengths <20 were not used, while maintaining read pair information for the assembly software to use.

Short-read assembly using Velvet. The short reads were assembled using VELVET v. 1.0.14, a program developed for the *de novo* assembly of short-read sequencing information [*6*]. An initial assembly was performed using the VelvetOptimizer v. 2.1.7 program using default settings. Next, BWA [*7*] was used to map the short reads to this assembly to accurately determine the average length and standard deviation of the inserts for use in future assemblies. VelvetOptimizer was then run again with a defined insert length and standard deviation so that the program would perform optimization of expected coverage and coverage cutoff values on each hash length from 23 to 43. The assembly with the largest N_{50} and maximum scaffold length was chosen for further assembly. Contigs that were both <100 bases and not included in a scaffold were assumed to be misassembled and therefore discarded at this time. Next, Image [*8*] was used to close gaps in the scaffolds generated by Velvet after dissolving them into their constituent contigs. Image was run iteratively on these contigs, one read library at a time, starting with the library with the smallest insert length. Image was run using hash lengths of 43, 33 and 23. Image was run at each hash length until no further gaps could be closed. Next overlaps

between remaining contigs were joined by CAP3 followed by manual joining of overlapping contigs in Consed to produce the final draft genome.

Short-read assembly using SOAP. The *S. cattleya* genome was also assembled using SOAP v. 1.05 for comparison. Using the trimmed reads and insert lengths determined above, assemblies were performed with hash lengths ranging from 21 to 43. The assembly with the largest N_{50} (hash length 37) was chosen for further assembly. Gapcloser from the SOAP package was used to close gaps present in the scaffolds from this assembly using first the trimmed reads used for the assembly followed by a round with the untrimmed reads. Finally, BWA was used to map the untrimmed reads to the assembled draft genome and the remaining gaps were closed by manually examining the sequence of reads that spanned these gaps.

Genome annotation. Putative open reading frames as well as structural RNAs were detected and annotated by submission to the IMG Expert Review pipeline at the Joint Genome Institute [*9*].

***S. cattleya* biomass preparation.** *S. cattleya* spores were heat-shocked at 55ºC for 5 min before inoculating into GYM media (yeast extract, 4 g/L; malt extract, 10 g/L; glucose, 4 g/L) (pH 5.0) for overnight growth at 30ºC at 200 rpm. The overnight cultures were subcultured into fresh GYM media (pH 5.0, 1 L) to OD_{600} = 0.05 in a 2.8 L Fernbach baffled flask containing glass beads (5 mm) to obtain dispersed growth. Sodium fluoride (2 mM final concentration) was added to the appropriate cultures 24 h after inoculation, which is well before fluoride uptake and organofluorine production begins. Organofluorine physiological markers (fluoride uptake and fluoroacetate/fluorothreonine production) were found to be unchanged compared to growths where sodium fluoride was added at inoculation. Biomass was collected at 0.5, 2, 6, 24, and 48 h after the addition of fluoride by filtering cultures (50-80 mL) through ashless filter paper (GE Healthcare; Buckinghamshire, UK). The cell pellet and culture supernatants were flash frozen and stored at –80ºC. Fluoride concentration was measured using a fluoride ion selective electrode (Mettler Toledo; Schwerzenbach, Switzerland) using a standard curve in GYM media (pH 5.0). Organofluorine production was monitored by lyophilizing the supernatant (50 mL) and resuspending in a mixture of 80% Tris-HCl pH 7.5 (100 mM) with 5-fluorouracil (1 mM)/20% D_2O (1 mL) for ^{19}F NMR using 5-fluorouracil as an internal standard. ^{19}F NMR spectra were collected at 25ºC on a Bruker AVQ-400 spectrometer at the College of Chemistry NMR Facility at the University of California, Berkeley.

Total RNA isolation and rRNA depletion. Biomass (from ~5 mL culture) was ground to fine powder in liquid nitrogen. The powder (50 mg) was then transferred into a microfuge tube (1.7 mL) pre-chilled in liquid nitrogen. Trizol™ reagent (1 mL) was then added to the powder and mixed by vortexing for 30 s. The sample was incubated at room temperature for 10 min before centrifuging at 20,817 × *g* at 4ºC for 10 min. The supernatant was transferred to a fresh tube containing chloroform (200 μL), mixed by vortexing (15 s), and incubated at room temperature for 10 min. The sample was centrifuged at 20,817 × *g* at 4ºC for 10 min to separate the organic and aqueous layers. The aqueous layer (200 μL) was removed and used for RNA isolation using Qiagen RNeasy Mini Kit according to the manufacturer's protocol with the following modifications: RLT buffer (700 μL) was mixed with β-mercaptoethanol (7 μL) immediately before being added to the 200 μL supernatant. Ethanol (500 μL) was then added to the mixture and mixed by pipetting. The resulting solution was loaded on the column and removed by spinning. The column was washed with RW1 buffer (500 μL) and RPE buffer (700 μL) consecutively. RNA was then eluted into 50 μL RNAse free water.

rRNA was removed using *MICROBE*xpress Kit (Ambion; Austin, TX) according to a modified manufacturer protocol. First, total RNA (10 μg) was concentrated by ethanol precipitation and resuspended in nuclease-free water (2.5 μL). Two rounds of rRNA depletion were carried out to enrich mRNA from total RNA. For the first round of depletion, binding buffer (200 μL) and Capture Oligonucleotide Mix (4 μL) were mixed with total RNA and denatured at 70ºC for 10 min before being incubated at 37ºC for 15 min. The equilibrated Oligo MagBeads (50 μL) were added and incubated with the RNA sample at 37ºC for 15 min. The supernatant was removed while using a magnet to pull down the Oligo MagBeads and transferred to a fresh tube for the second round of rRNA depletion using half the amount of Capture Oligonucleotide Mix and Oligo MagBeads compared to the first round. After two rounds to rRNA removal, mRNA was recovered by ethanol precipitation and resuspended in nuclease-free water (5-10 μL). RNA was analyzed on a BioRad Experion (Model 700-7000) using RNA HighSens Analysis Kit in order to assess the quality and concentration of individual samples.

Preparation of cDNA libraries for RNA sequencing. mRNA (9 μL) was fragmented by mixing with fragmentation buffer (1 μL, Ambion) and heating at 70ºC for 10 min. Fragmented RNA was recovered by ethanol precipitation and reverse transcribed using SuperScript III (200 U, Invitrogen) in a reaction (20 μL) containing random hexamer primers (200 ng), dNTPs (0.5 mM), DTT (10 mM), DMSO (5%) and SuperScript III (200 U). The first strand cDNA was used directly for second strand cDNA synthesis by adding RNAse H (2.5 U, Invitrogen), *E. coli* DNA polymerase I (42.5 U, Invitrogen), and dNTPs (30 nmol) in Buffer 1 (NEB, 100 μL total reaction

volume). The reaction mixture was incubated at 16ºC for 2.5 h and purified using the QIAquick PCR Purification Kit. The cDNA ends were polished at 20ºC for 30 min using Klenow (5 U, NEB), T4 phosphonucleotide kinase (50 U, NEB) and dNTPs (40 nmol) in T4 DNA ligation buffer (NEB, 100 μL total reaction volume). The blunt-ended cDNAs were isolated using the QIAquick PCR Purification Kit and incubated at 72ºC for 10 min (50 μL total reaction volume) with Taq polymerase (1 μL) and dATP (20 nmol). After another QIAquick PCR purification step, adapters (20 nmol) were ligated with T4 DNA ligase (2,000 U) at room temperature for 30 min (30 μL total reaction volume). The libraries were enriched by 16-18 cycles of amplification with Platinum Taq High-Fidelity after QIAquick PCR purification.

RNA sequencing analysis. Each lane of RNA-seq reads were mapped to the completed genome of *S. cattleya* [*10*] using BWA [*7*] allowing for no more than 2 mismatched bases. Reads per feature were then extracted from the alignment file using HTseq-count [*11*] and the annotation of the completed genome. Values for reads mapped to rRNA sequences were removed prior to further data processing. Next, the counts per feature for each condition and time point were quantile normalized using the preprocessCore R package [*12*]. For differential expression analysis, Multiexperiment Viewer (MeV) was used to perform statistical analysis using edgeR [*13*].

DNA microarray design, hybridization, and analysis. Custom DNA microarrays (Agilent Technologies) were designed with 2-4 unique probes for each predicted opening reading frame based on the Velvet draft genome using eArray (https://earray.chem.agilent.com/earray) with a theoretical melting temperature of 85ºC. Custom microarrays containing *in situ* synthesized 60-mer oligos were then produced by Agilent Technologies (Santa Clara, CA) with a total of 15,744 probes randomly printed on the array (Agilent controls, 536; replicated *S. cattleya* probes, 240). cDNA was generated by a two step labeling method of mRNA wherein the ARES DNA labeling Kit (Invitrogen) was used to synthesize amine modified cDNA by incorporating aminoallyl-dUTP according to the manufacturer protocol from 20 μg of RNA. The remaining RNA was then hydrolyzed using NaOH (1 M), which was subsequently removed using a QIAquick column. Amine-modified cDNA (3~5 μg) was then labeled with Alexa Fluor 555 or Alexa Fluor 647 according to manufacturer protocol. Equal amounts (3~5 μg) of dye labeled cDNA from the reference and sample were mixed and used for hybridization in a G2545A hybridization oven (Agilent) at 10 rpm for 17 h at 65ºC. The slides were subsequently washed twice each with Solution I for 5 min and Solution II for 5 min followed by drying and scanning by a G2565BA scanner (Agilent). The features were extracted using Agilent Feature extraction software (version 10.7.3.1). Dye bias was normalized by applying the Lowess algorithm and the $\log_2$ fold changes

were calculated using the processed signals of each probe. The reproducibility was assessed by plotting the $\log_2$ fold change ratios of technical replicates against each other. The following probes were removed before statistical analysis: (1) probes with surrogated signals; (2) probes with processed signal intensity lower than 18 in any two replicates. One sample t-test was applied to $\log_2$ fold changes to determine differentially expressed genes. The probes with $\log_2$ ratios >1 or <-1 and with *p* values <0.01 were considered showing significant changes between conditions.

Construction of plasmids for protein expression. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase, Platinum Taq High Fidelity or Taq polymerase using the oligonucleotides listed in Table S1. Briefly, PCR reactions containing 5 – 10% DMSO were cycled with melting temperatures of 95°C or 98°C as appropriate for 30 s, followed by annealing at temperatures ranging from 62°C - 72°C (10°C below the calculated T_m of the primers) for 30 s and then extension at 72°C for a length of time appropriate for the gene for 25 rounds. pET16b-His$_{10}$-ACS.EC, pET16b-His$_{10}$-GltA.EC, pET16b-His$_{10}$-FucA.EC, pET16b-His$_{10}$-ACS.Scat, pET16b-His$_{10}$-AckA.Scat, pET16b-His$_{10}$-PTA.Scat, pET16x-His$_{10}$-Cit1.Scat, pET16b-His$_{10}$-MRI1.Scat, pET16b-His$_{10}$-MRI2.Scat, pET16b-His$_{10}$-CitA.Sco, and pET16b-His$_{10}$-MtnK.BS were constructed by amplification out of the appropriate genomic DNA and pET16b-His$_{10}$-AckA.EC and pET16b-His$_{10}$-PTA.EC were amplified from the plasmid pRSFDuet-ackA.pta using the respective primers given in Table S1 and inserted directly into the NdeI and BamHI restriction sites of pET16b, except for Cit1.Scat, which was inserted into the NdeI and SpeI sites of pET16x. 1b-His$_6$-Cit2.Scat was constructed by amplification out of the appropriate genomic DNA and inserted into the SspI site of Addgene plasmid 29653 (1b) by Ligation Independent Cloning (LIC) by the QB3 MacroLab (Berkeley, CA). Following plasmid construction, all cloned inserts were sequenced at Sequetech (Mountainview, CA) or Quintara Biosciences (Berkeley, CA).

Expression and purification of His-tagged proteins. *E. coli* BL21(de3) was transformed with the protein expression plasmid as well as pRARE2 for the expression of *Streptomyces* and *Bacillus* proteins. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 μg/mL each) in a 2.8 L- Fernbach baffled shake flask to OD_{600} = 0.05. The cultures were grown at 37 °C at 200 rpm to OD_{600} = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.2 mM) and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min at 4ºC and stored at -80ºC.

Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in Buffer A (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, 10% glycerol, 20 mM β-mercaptoethanol, pH 7.5) supplemented with PMSF (0.5 mM). The cell paste was homogenized before lysis by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 14,000 psi. The lysate was then centrifuged at 15,300 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.8% (w/v) streptomycin sulfate and stirring at 4°C for 30 min and the precipitated DNA was then removed by centrifugation at 15,300 × *g* for 20 min at 4°C. The remaining soluble lysate was passed over a Ni-NTA agarose column (Qiagen, 5-10 mL) for isolation of the target protein.

His_{10}-ACS.EC and His_{10}-ACS.Scat were purified by gravity flow as follows: After loading the cleared cell lysate, the column was washed with Buffer A until the $A_{280\ nm}$ dropped below 0.05 (Nanodrop 1000; Thermo Fisher) followed by Buffer B (50 mM sodium phosphate, 300 mM sodium chloride, 25 mM imidazole, 10% glycerol, 20 mM β-mercaptoethanol, pH 7.5) until the $A_{280\ nm}$ again dropped below 0.05. Protein was then eluted from the column by with Buffer C (50 mM sodium phosphate, 300 mM sodium chloride, 300 mM imidazole, 10% glycerol, 20 mM β-mercaptoethanol, pH 7.5). Finally, the protein was concentrated in an Amicon spin concentrator with a molecular weight cutoff of 30 kD.

His_{10}-GltA.EC, His_{10}-AckA.EC, His_{10}-PTA.EC, His_{10}-AckA.Scat, His_{10}-PTA.Scat, His_{10}-Cit1.Scat, His_6-Cit2.Scat, His_{10}-CitA.Sco and His_{10}-MtnK.BS were purified on an ÄKTApurifier FPLC (GE Healthcare Life Sciences; Piscataway, NJ) as follows: After loading the cleared cell lysate, the column was washed with 15 column volumes of Buffer A, followed by a linear gradient to 30% Buffer C over 30 column volumes (2 mL/min). Finally, the protein was eluted with 10 column volumes of Buffer C. Fractions containing the purified protein were pooled and concentrated in an Amicon spin concentrator with the appropriate molecular weight cutoff.

Proteins were then exchanged into Buffer D (50 mM HEPES, 100 mM sodium chloride, 2.5 mM DTT, 10% glycerol, pH 7.5) by gravity flow passage over a Sephadex G-25 column (Sigma-Aldrich, bead size 50-150 μm, 75 mL) and concentrated in an Amicon spin concentrator with the appropriate molecular weight cutoff. Final protein concentration was determined using a calculated $\varepsilon_{280\ nm}$ (ExPASY ProtParam) and were measured as follows: His_{10}-ACS.EC: 14.3 mg/mL (ε_{280} = 138,770 M^{-1} cm^{-1}), His_{10}-ACS.Scat: 1.9 mg/mL (ε_{280} = 150,800 M^{-1} cm^{-1}), GltA.EC: 2.3 mg/mL (ε_{280} = 40,340 M^{-1} cm^{-1}), His_{10}-AckA.Scat: 1.9 mg/mL ($\varepsilon_{280\ nm}$ = 13,410 M^{-1} cm^{-1}), His_{10}-PTA.Scat: 3.2 mg/mL ($\varepsilon_{280\ nm}$ = 20,860 M^{-1} cm^{-1}), His_{10}-Cit1.Scat: 0.8 mg/mL ($\varepsilon_{280\ nm}$ = 44,810 M^{-1} cm^{-1}), His_6-Cit2.Scat: 1.7 mg/mL ($\varepsilon_{280\ nm}$ = 50,420 M^{-1} cm^{-1}), His_{10}-CitA.Sco:

2.3 mg/mL ($\varepsilon_{280\ nm}$ = 49,280 $M^{-1}\ cm^{-1}$) and His_{10}-MtnK.BS: 2 mg/mL ($\varepsilon_{280\ nm}$ = 39,880 $M^{-1}\ cm^{-1}$). All proteins were then stored at -80°C.

His_{10}-MRI2.Scat and His_{10}-FucA.EC were purified by FPLC as follows: After loading the cleared cell lysate, the column was washed with 15 column volumes of Buffer F (50 mM sodium phosphate, 10 mM imidazole, 20% glycerol, 20 mM β-mercaptoethanol, pH 7.5) followed by a linear gradient to 100% Buffer G (50 mM sodium phosphate, 300 mM imidazole, 20% glycerol, 20 mM β-mercaptoethanol, pH 7.5) over 30 column volumes (2 mL/min). Fractions containing the purified protein were pooled and stirred at 4°C for 30 min with zinc chloride (1 mM final concentration). Protein was then loaded onto a SepharoseQ column and eluted with a linear gradient from 100% Buffer H (50 mM HEPES, 20% glycerol, pH 7.5) to 100% Buffer I (50 mM HEPES, 500 mM sodium chloride, 20% glycerol, pH 7.5) over 30 column volumes. Protein was then concentrated in an Amicon spin concentrator with a molecular weight cutoff of 15 kD to 0.8 mg/mL for His_{10}-MRI2.Scat ($\varepsilon_{280\ nm}$ = 50, 420 $M^{-1}\ cm^{-1}$) and 1.5 mg/mL for His_{10}-FucA.EC ($\varepsilon_{280\ nm}$ = 21,430 $M^{-1}\ cm^{-1}$).

Preparation of methylthioribose-1-phosphate. Methylthioribose-1-phosphate was synthesized according a modified literature procedure [*14*]. Methylthioadenosine (22.2 mg, 75 μmol) was dissolved in 50 mM sulfuric acid (25 mL) and heated to 100°C for 3 h. The reaction was then allowed to cool to room temperature followed by passage over a column of Dowex 50WX8-100 (2 mL). The column was washed with water until no further methylthioribose eluted as determined by a reducing sugar assay [*15*]. The eluent was neutralized with barium hydroxide, filtered to remove precipitate, and lyophilized overnight. The remaining solid was dissolved in a reaction mixture (15 mL) with 5 mM ATP, 5 mM magnesium chloride, 1 mg of His_{10}-BS.MtnK and 100 mM HEPES, pH 7.5. The reaction was run overnight at 37°C and then purified using a SepharoseQ column (10 mL) using a linear gradient of 0 – 200 mM sodium chloride over 30 column volumes. Fractions containing the methylthioribose-1-phosphate as determined by LC-MS were lyophilized overnight before resuspending in water (1 mL) and desalted by HPLC using an Eclipse XDB C-18 column (5 μm, 9.4 × 250 mm, Agilent) with a gradient from water to 100% acetonitrile over 30 min (3 mL/min). Fractions containing the MTR1P as determined LC-MS were pooled, lyophilized, and resuspended in 100 mM HEPES, pH 7.5 for storage at -80°C.

Biochemical characterization of the isomerase/dehydratase fusion (MRI2). The reaction product of MRI2 was identified by an LC-MS assay in negative ion mode. The reaction (500 μL) containing 1 mM MTR1P, 20 mM HEPES, 5 mM magnesium sulfate, pH 7.5 was initiated by the addition of MRI2 (0.1 mg/mL) at 30°C. Aliquots (50 μL) were removed at the appropriate

times and the reaction was quenched by heating the aliquot to 100°C for 30 s. The product was analyzed by LC-MS with a Zorbax-Eclipse C-18 column (3.5 μm, 3.0 × 150 mm; Agilent) using a linear gradient to 50% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 20 mM triethylamine with the pH adjusted to 4.5 with formic acid as the mobile phase.

Preparation of fluoroacetyl-CoA. Fluoroacetyl-CoA was synthesized according to a modified literature procedure [*16*]. Sodium fluoroacetate (100 mg, 1mmol) was dried under vacuum in an oven dried round bottom flask equipped with a stir bar and distillation apparatus. Phosphorus pentachloride (208 mg, 1 mmol) was added to the flask and the two powders were mixed by shaking the flask, which was subsequently heated at 100°C with aggressive stirring. Fluoroacetylchloride was collected by distillation. Aliquots of fluoroacetylchloride (2.1 μL, 0.03 mmol) were added to a stirring solution of coenzyme A trilithium salt (25 mg, 0.03 mmol) dissolved in a 10% solution of sodium bicarbonate (1 mL) until the pH of the solution dropped below 7 (litmus paper). The fluoroacetyl-CoA was purified by HPLC using an Eclipse XDB C-18 column (5 μm, 9.4 × 250 mm, Agilent) with a gradient from 0-50% acetonitrile in water over 30 min at a flow rate of 3 mL/min. Fractions containing <5% free CoA were identified by LC-MS, snap frozen in liquid nitrogen and lyophilized.

Enzyme assays. Acetyl-CoA synthetase activity was measured using a modified literature method [*17*] to monitor the production of AMP by coupling ADP production by myokinase to phosphoenolpyruvate dephosphorylation by pyruvate kinase to NADH dependent reduction of pyruvate to lactate by lactate dehydrogenase and monitoring NADH consumption spectrophotometerically. Assays were performed at 30°C in a total volume of 1000 μL containing 50 mM HEPES, pH 7.5, CoA (0.5 mM), ATP (2 mM), magnesium chloride (5 mM), phosphoenolpyruvate (1 mM), NADH (0.16 mM), myokinase (10 U), pyruvate kinase (10 U), lactate dehydrogenase (10 U), and ACS (ACS.Scat, 25 nM; ACS.EC, 29 nM). The reaction was initiated with addition of acetate (5-500 μM) or fluoroacetate (0.5 – 1000 mM) and monitored at 340 nm in a Beckman Coulter DU-800 spectrophotometer. The error was derived from the nonlinear curve fitting.

Acetate kinase activity was measured by a modified literature method [*17*] to monitor the production of ADP, coupled to pyruvate formation by pyruvate kinase and lactate formation by lactate dehydrogenase and monitoring NADH consumption. Assays were performed at 30^{o}C in a total volume of 500 μL containing 100 mM HEPES, pH 7.5, ATP (10 mM), magnesium chloride (20 mM), phosphoenolpyruvate (2 mM), NADH (0.25 mM), pyruvate kinase (approx. 23 U), lactate dehydrogenase (approx. 33 U), and AckA (AckA.Scat: 3.4 nM, acetate; 1.1 μM,

fluoroacetate. AckA.EC: 0.3 nM, acetate; 40 nM, fluoroacetate). The reaction was initiated with addition of acetate (1 – 150 mM) or fluoroacetate (1 - 200 mM) and monitored at 340 nm in a Beckman Coulter DU-800 spectrophotometer.

Phosphotransacetylase activity in the acetyl-phosphate forming direction was monitored by a modified literature method to monitor the release of free CoA by DTNB [*18*]. Assays were performed at 30°C in a total volume of 500 μL containing 100 mM HEPES, pH 7.5, acetyl-CoA (25 – 1000 μM), DTNB (0.5 mM), sodium phosphate (30 mM), magnesium chloride (PTA.EC, 0 mM; PTA.Scat, 40 mM) and PTA (PTA.EC, 65 nM; PTA.Scat, 130 nM). The reaction was initiated by the addition of phosphate and monitored at 412 nm in a Beckman Coulter DU-800 spectrophotometer

Phosphotransacetylase activity in the acetyl-CoA forming direction was determined using a discontinuous HPLC assay. Reactions consisting of 100 mM HEPES, pH 7.5, acetate or fluoroacetate (100 μM), CoA (500 μM), magnesium chloride (PTA.EC: 40 mM; PTA.Scat, 40 mM) and AckA (1 μM) and PTA (PTA.Scat, 250 nM; PTA.EC, 25 nM) at 30°C were initiated by the addition of ATP (10 mM). Aliquots (49 μL) were removed at 30 s intervals over 3 min and quenched by the addition of 70% perchloric acid (1 μL). Aliquots were then analyzed by HPLC on a Zorbax Eclipse Plus Rapid Resolution HD C-18 column (1.8 μm, 2.1 × 50 mm; Agilent) using a linear gradient from 2% to 20% methanol over 3 min at a flow rate of 0.8 mL/min with 50 mM sodium phosphate with 0.1% trifluoroacetic acid, pH 4.5 as the mobile phase.

Citrate synthase activity was measured by a modified literature method to monitor the release of free CoA by DTNB [*18*]. Assays were performed at 30°C in a total volume of 500 μL containing 50 mM HEPES, pH 7.5, oxaloacetate (0.5 mM), DTNB (0.5 mM), potassium chloride (200 mM), acetyl-CoA (CitA.Sco, 5 – 300 μM; GltA.EC, 10 – 500 μM; CS1.Scat, 5 – 200 μM; CS2.Scat, 1 – 50 μM) or fluoroacetyl-CoA (CitA.Sco, 10 – 75 μM; GltA.EC, 5 – 200 μM; CS1.Scat, 5 – 200 μM; CS2.Scat, 1 – 15 μM). The reaction was initiated by addition of enzyme (CitA.Sco: 20 nM, acetyl-CoA; 200 nM, fluoroacetyl-CoA. GltA.EC: 10 nM, acetyl-CoA; 200 nM, fluoroacetyl-CoA. Cit1.Scat: 20 nM, acetyl-CoA, 100 nM, fluroacetyl-CoA. Cit2.Scat: 2.5 nM, acetyl-CoA; 25 nM fluoroacetyl-CoA) and monitored at 412 nm in a Beckman Coulter DU-800 spectrophotometer.

Kinetic parameters (k_{cat}, K_M, $K_{0.5}$, and n) were determined by fitting the data using Microcal Origin to the equations:

$$v_o = v_{max}\,[S]/\,K_M + [S]$$

$$v_o = v_{max}\,[S]^n/\,(K_{0.5})^n + [S]^n$$

where v is the initial rate and [S] is the substrate concentration. Standard error generated by non-linear curve fitting (n = 3).

Measurement of aconitase assay in *S. cattleya* cell lysates. *S. cattleya* was cultured as described and harvested 48 h after fluoride addition by centrifugation at 9,800 × *g* for 7 min at 4ºC. Cell pellets were then resuspended in 50 mM sodium phosphate, pH 7.5 (5 mL/g cell paste) with lysozyme (1 mg/mL) and incubated at room temperature for 30 min. Following homogenization, cells were lysed by passage through a French Pressure cell at 25,000 Psi and supplemented with ammonium ferrous sulfate (0.25 mM final concentration) before removal of insoluble material by centrifugation at 15,300 × *g* for 20 min at 4°C. The supernatant was then placed in a round-bottom flask on ice and degassed by 15 cycles of gentle vacuum followed by flushing with Ar over 30 min. The degassed cell lysate was then sparged gently with Ar for an additional 20 min before recharging the aconitase by addition of ammonium ferrous sulfate (0.1 mM) and sodium sulfide (0.1 mM) followed by incubation at room temperature for 20 min. Aconitase activity was measured using a modified literature method [*19*] to monitor the production of isocitrate coupled to the production of α-ketoglutarate and NADPH by isocitrate dehydrogenase. Assays (2 mL) were performed at 30°C in a airtight cuvette (Starna Cells; Atascadero, CA) containing 50 mM sodium phosphate, pH 7.5, $NADP^+$ (0.18 mM), magnesium sulfate (1.5 mM), isocitrate dehydrogenase (0.2 U/mL), fluorocitrate (0 – 100 μM) and recharged cell lysate (100 μL). The reaction was initiated with addition of citrate (2 mM) and monitored at 340 nm in a Beckman Coulter DU-800 spectrophotometer.

Construction of plasmids for gene disruption in *S. cattleya*. Plasmids for gene disruption contained a cassette consisting of an apramycin (Am) or hygromycin (Hm) resistance marker and an origin of transfer (*oriT*) flanked by *S. cattleya* genomic DNA sequence derived from the upstream and downstream regions of the open reading frame to be disrupted.

pCR2.1-TOPO-Δ*mri2*::Am^R: *S. cattleya* NRRL 8057 cosmid library was constructed using the SuperCos 1 Cosmid Vector Kit (Stratagene). SuperCos 1 cosmid vector was digested with *Xba*I, dephosphorylated using Antarctic phosphatase (New England BioLabs) and then digested with *Bam*HI. *S. cattleya* genomic DNA was partially digested with *Sau*3A and dephosphorylated (Figure S10). Linearized SuperCos I vector and partially digested genomic DNA was then

ligated and packaged using Gigapack III XL packaging extract (Stratagene). The cosmid library was screened with primers Iso.Sc F1 and Iso.Sc R1 using the pooled PCR approach (Figure S10). The colonies (10-12) were combined for the first round of screening in a 96-well PCR plate. The colonies in the positive colony pools were screened individually for the second round to identify positive clones. The cosmid (C6-1) was isolated using Qiagene QIAprep Spin Miniprep Kit for sequencing and confirmed to harbor a *S. cattleya* genomic fragment containing *mri2*. The Am^R/OriT cassette was amplified from pIJ773 using primers MRI2 LR F and MRI2 LR R and used to disrupt the *mri2* gene in C6-1 using the REDIRECT method according to protocol in *E. coli* DH10B pIJ790 host cells [*20*]. Apramycin resistant colonies were picked and disruption of *mri2* was confirmed by sequencing. As the upstream and downstream regions were determined to be too large for efficient double crossovers in *S. cattleya*, the disrupted *mri2* with 2 kb upstream and downstream flanking regions were amplified from C6-1-Δ*mri2*::Am^R using primers MRI2 UF and MRI2 DR and inserted pCR2.1-TOPO using TA cloning. The resulting plasmid was confirmed by sequencing and used to generate the *S. cattleya* Δ*mri2*::Am^R strain.

pIJ10701-Δ*mri1*::Hm^R-(+) and pIJ10701-Δ*mri1*::Hm^R-(-): The *mri1* gene was disrupted with the Hm^R cassette in two different orientations (pIJ10701-Δ*mri1*::Hm^R-(+), Hm^R cassette on the coding strand; pIJ10701-Δ*mri1*::Hm^R-(-), Hm^R cassette on the non-coding strand) in order to control for possible polar effects reported in the literature [*21-23*].

pIJ10701-Δ*mri1*::Hm^R-(+) was constructed by amplifying flanking genomic sequences (2 kb) immediately upstream (MRI1 UF and MRI1 UR primers) and downstream (MRI1 DF and MRI1 DR primers) of the *mri1* open reading frame and inserting into the NotI-SpeI and HindIII-ClaI sites, respectively, of pIJ10701. The resulting plasmid was confirmed by sequencing and used to generate the *S. cattleya* -Δ*mri1*::Hm^R-(+) strain. pIJ10701-Δ*mri1*::Hm^R-(-) was constructed by amplification of flanking genomic sequences (2 kb) immediately upstream (MRI1 G UF and MRI1 G UR primers) and downstream (MRI1 G DF and MRI1 G DR primers) of the *mri1* open reading frame and inserting into the XhoI and NotI site, respectively, of pIJ10701 using the Gibson protocol [*24*]. The resulting plasmid was confirmed by sequencing and used to generate the *S. cattleya* -Δ*mri1*::Hm^R-(-) strain.

pRS316-Δ*flK*::Am^R. Flanking DNA sequences (2 kb) immediately upstream (flK UF1 and flK UR1 primers) and downstream (flK DF1 and flK DR1 primers) of the *flK* open reading frame were amplified. The Am^R/*oriT* cassette was amplified from pIJ773 using primers flK KF1/flK KR1. Primers for all amplification steps contained extensions at the 5′ and 3′ end such that all PCR products to be assembled contained 50 nt of sequence identity with the preceding and

following pieces. The three PCR products and KpnI-linearized pRS316 were transformed into *Saccharomyces cerevisiae* L4581 for assembly by *in vivo* homologous recombination. Transformants containing the circularized plamid and the desired insert were selected on solid synthetic complete dextrose (SCD)–uracil plates. DNA was recovered by phenol/chloroform extraction, resuspended in TE buffer (50 μL; 10 mM Tris-HCl, pH 7.5, 1 mM EDTA), and transformed into *E. coli* DH10B-T1^R. Transformants were selected on LB agar supplemented with carbenicillin (50 μg/mL) and apramycin (50 μg/mL) and plasmids were verified by sequencing and used to generate the *S. cattleya* Δ*flK*::Am^R strain.

Conjugative transfer of plasmids from *E. coli* to *S. cattleya*. Plasmids for gene disruption were transferred into *S. cattleya* by conjugation with *E. coli* GM272 harboring the *oriT*-mobilizing plasmid pUZ8002 using a modified literature protocol [*3*]. *E. coli* GM272 pUZ8002 were transformed with plasmids for gene disruption and transformants were selected on LB agar supplemented with kanamycin (50 μg/mL) and apramycin (50 μg/mL) or carbenicillin (50 μg/mL). Single colonies were inoculated into LB supplemented with the appropriate antibiotics and were grown overnight at 37°C shaking at 200 rpm. The overnight *E. coli* culture was inoculated into LB supplemented with the appropriate antibiotics to OD_{600} = 0.1. The culture (10 mL) was grown to an OD_{600} of 0.4-0.6 and cells were pelleted by centrifugation, washed twice with sterile water (10 mL), and resuspended in GYM media (0.5 mL). *S. cattleya* spores (10^8 in 50 μL) were prepared for conjugation by heat shock at 55°C for 5 min followed by addition of GYM (500 μL) and incubation for 4 h at 30°C with shaking at 200 rpm. *E. coli* (500 μL) were added to spores and the mixture was incubated at 30°C for 1 h with shaking at 200 rpm. The mixture was plated on mannitol-soy flour (MS) media. The plate was incubated for at 30°C for 36 h, and then overlaid with nalidixic acid in 2.5 M $MgCl_2$ (640 μL, 1.5 mg/mL). After a second incubation for 16 h, the plate was then overlaid with apramycin (20 μL, 50 mg/mL) or hygromycin (20 μL, 50 mg/mL). Ex-conjugants appeared within 5-7 d and gene disruption was verified by PCR amplification of the targeted locus followed by sequencing. Consistent with phenotypes observed in other actinomycetes, Δ*mri1*-(+):*Hm*^{R} demonstrates a severe growth defect and is also not competent for organofluorine production. Thus *S. cattleya* Δ*mri1* Δ*mri2* was constructed as the Δ*mri1*-(-):*Hm*^{R} strain.

Characterizing organofluorine production in *S. cattleya* wildtype, Δ*mri1*, Δ*mri2*, and Δ*mri1*Δ*mri2* strains. Single colonies of wildtype, Δ*mri1*, Δ*mri2*, and Δ*mri1*Δ*mri2*, were picked from SMMS [*3*] plates and inoculated into GYM pH 7 (2 mL) with glass beads (5 mm) to obtain dispersed growth. After 4 d of growth, cultures were sub-cultured into GYM pH 7 (25 mL) and grown overnight. These overnight cultures were then used to inoculate GYM pH 5 (1 L) to an

OD_{600} of 0.05. After 24 h, sodium fluoride (2 mM final concentration) was added to the cultures. At appropriate time points, aliquots of the cultures (50 mL) were removed, OD_{600} was determined, and cells were pelleted by centrifugation at 9,800 × *g* for 7 min at 4ºC. Fluoride concentration in the supernatant was determined using a fluoride electrode against a standard curve in GYM pH 5. The supernatant (45 mL) was lyophilized and resuspended in 80% Tris-HCl pH 7.5 (100 mM) with 5-fluorouracil as an internal standard (1 mM)/20% D_2O (1 mL). Insoluble material was removed by centrifugation before collecting ^{19}F NMR data on a Bruker AVQ-400 (400 scans; o1p, -170 ppm; d1, 11 s).

Characterizing fluoroacetate sensitivity in *S. cattleya* wildtype and Δ*flK* strains. Spores of *S. cattleya* wild-type and Δ*flK*::Am^R (40 μL) were heat-shocked at 55° C for 5 min and inoculated into GYM (pH 7.0). Cultures were grown at 30°C, 200 rpm, to OD_{600} = 1.5-2 and then were diluted to OD_{600} = 0.1 in GYM (pH 5.0, 50 mL) with and without sodium fluoride (2 mM) or sodium fluoroacetate (2 mM). Glass beads (4 per flask, 5 mm), were added to promote dispersed growth. Growth was monitored by measuring OD_{600} of a sample (1 mL) withdrawn from the culture every 2 h until cultures reached saturation.

Supplementary Results and Discussion

Figure S1. Electropherograms of the paired-end genomic libraries prepared from *Streptomyces cattleya* NRRL 8057 (ATCC 35852) utilized for Illumina sequencing.

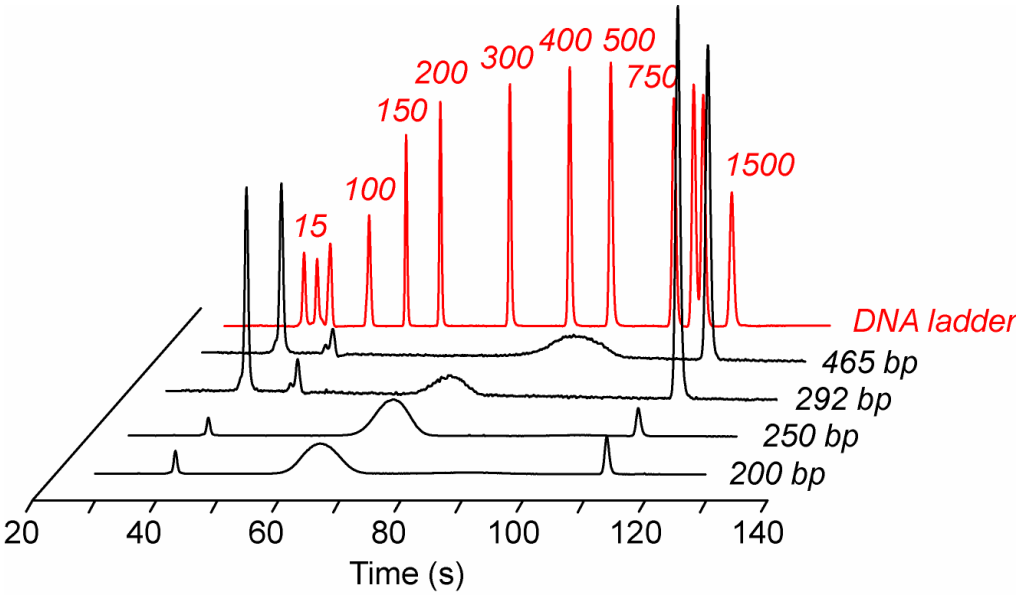


Table S1. Oligonucleotides used for plasmid construction, genomic library construction, and cosmid screening/sequencing

Name	Sequence
EUB F933	gcacaagcggtggagcatgtgg
EUB R1387	gcccgggaacgtattcaccg
U1 F	ccagcagccgcggtaatacg
U1 R	atcggctaccttgttacgacttc
Iso2.SC F1	accgcgctcggcgcgattctg
Iso2.SC R1	acgacgacgaagggtatgcggtg
SuperCosISeq F1	cacctgacgtctaagaaacc
CosmidSeq F1	gcgcgcccggcgacgaagagg
CosmidSeq F2	ccgccacagcgcggtcagcc
CosmidSeq F5	cgccgcccgtcgcggcacc
CosmidSeq F6	cccgggacggcgggcctgg
CosmidSeq F7	ggagcacgccgaactgctgg
CosmidSeq F8	gccgtatctggctgtacgag
CosmidSeq F9	tcctcgtcggcgcggccaagc
CosmidSeq F13-1	gcagcgtcgtcgtccccgac
CosmidSeq F14	cgccgagcgggtagcggcgc
CosmidSeq F15	atgaccacggccacccggtc
CosmidSeq F20	gccggacgtggccgatcacg
CosmidSeq F21	aggtgccgccgtgctcgaag
CosmidSeq F22	tgccggatcccgggcagctc
CosmidSeq F24	gggctgccgtcacggctgtggtc
CosmidSeq R1	cggtaggactcggggagtcgc
CosmidSeq R2	gcgcgcgccgcgtgctgcgc
CosmidSeq R3	ggcaccctggagtacctgg
CosmidSeq R4	cgatgccggcggcgccttcg
CosmidSeq R5	gcctggccgaggcgatggc
CosmidSeq R6	ctgtgcgtggagggttgcgg
CosmidSeq R8	accgctggtgcgggacagcc
CosmidSeq R9-1	cgaagccgttgctcgtgggc
CosmidSeq R12	ttcgaggggcccaggcgggc
CosmidSeq R12-1	ggcctcgccgagccccgagg
CosmidSeq R13	gcaacgacgccttcgggctg
CosmidSeq R16	ctgggagctgatctgggtcg
CosmidSeq R19	gtacccgttccgcgtcgcac
CosmidSeq R21	tggcagggcggaccgctggc
CosmidSeq R23	atgggcgccctggccaagag
PKSSeq R1	cctcgacgcggcgctgcacgc
PKSSeq R2	cgttcatcgaacccgtcacggc
PKSSeq R3	gctggtcgcccgccatctgg
PKSSeq4	gtcccgacgccctcttcccg
PKSSeq6	tcgtcggcggatccggtgac
PKSSeq7	cggctggtggccgaggggc
PKSSeq8	caactacgccgccgccaacg
PKSSeq9	cacggcgacgccgaggggatcg
PKSSeq10	cgccgctcggcgggcacgtcgg
PKSSeq11	cgaggccgttgtggtggcttcg
PKSSeq12	cgcccacgcccctggccatacg
PKSSeq13	tggccgaacggcttgcggaagg
PKSSeq14	tgggatcctgcgcctccatgg

PKSSeq15	cgcaccgtggagctgctctacc
PKSSeq16	ttcttcggtatctcgccgcgtgag
PKSSeq17	acactcaccgccggaccctcaag
PKSSeq18	acccggcgggctcggccgtcgg
PKSSeq19	tcccaggccaccgcctccgtcaac
ACS.EC F	ggaattccatatgagccaaattcacaaacacaccattcctgccaacatcgc
ACS.EC R	attggatccttacgatggcatcgcgatagcctgcttctcttcaagcagc
GltA.EC F	ggaattccatatggctgatacaaaagcaaaactcaccctcaacggggatacagc
GltA.EC R	ttggatccttaacgcttgatatcgcttttaaagtcgcgtttttcatatcctgtatacagc
AckA.EC F	aaggagatatacatatgtcgagtaagttagtactggttctgaactgcggtagttcttcactg
AckA.EC R	attggatcctctagatcaggcagtcaggcg
FucA.EC F	aaggactcgagatggaacgaaataaacttgctcgtcagattattgacacttgcctgg
FucA.EC R	aataggatccttactcttcaattcgtaacccataggttttgaatttctccagcactacggc
ACS.Scat F	ggacatatgagctatgccctgggacaaatggcaacacccgaggg
ACS.Scat R	attggatccgtaacgcctcgctcacccgcccggggtcagtcctcgctg
AckA.Scat F	attcatatgaccgcgccccgtcccgccccgaaggagacc
AckA.Scat R	attggatccttatccctcgcggacggaatgggacgacgtacgaacgagcg
PTA.Scat F	ggacggagcaggcatatggacgcgcagcgtgtacataaccg
PTA.Scat R	attggatccgacggggcgcggtcatcgggcgtcgccctc
AckA.EC F	ggagatatacatatgtcgagtaagttag
AckA.EC R	attggatcctctagatcaggcagtcaggcg
PTA.EC F	attcatatgtcccgtattattatgctgatc
PTA.EC R	attctcgaggagggtaccgacgtcttac
Cit1.Scat F	attcatatgagcgacaactctgtagtactgcggtacggggacggcgagtacagctacccg gtggtcgagagcaccgttgg
Cit1.Scat R	attaactagttcagcgggcctctacggggacgtagtcgc
MRI1.Scat F	aaggagatatacatatgggtgatcagtccgtacagcctttggccaagggcacgg
MRI1.Scat R	attggatcctcacggctgggcgcggacggg
MRI2.Scat F	gagatatacatatgccccggcgtgccaccaccccgtcgctcacctgggacg
MRI2.Scat R	attggatcctcatccggtcaccccggggccggcggcgtccagggcgagctg
CitA.Sco F	gatatacatatgagcgacaactctgtagtactgcggtacggcgacggc
CitA.Sco R	ataggatcctcagcgctcctcgacggggacgaagtcg
MtnK.BS F	gacatatgatgatgatgggagtcacaaaaacacc
MtnK.BS R	gtgaattctcatgaatagatcattcctttagaag
MRI2 LR F	cgctcgccgcgaccttcgccgacccacggaggccccatgattccggggatccgtcgacc
MRI2 LR R1	tgtaggctggagctgcttcggtgtctcctcttcgtcgccgggcgcgcggtcccgg
MRI2 UF	attggatcccgaggccctggtggccgtcatggccagctccgtcaactacaac
MRI2 DR	attaagcttagccgctcgatgttgcgtgcgaacaccgcgaagacctcctc
MRI1 UF	ttaatgcggccgcaatcagacgattccccggaccccgccctcttt
MRI1 UR	attaactagtccgtgcccttggccaaaggctgtacggactgatcaccgat
MRI1 DF	attaaagcttcccgtccgcgcccagccgtga
MRI1 DR	tatatatcgattgaccttcatccgttcctggaggaggccgtc
MRI1 G UF	tatatactttagattgattatcaagcttgggcccttcgggaggctcttcccagcggagag
MRI1 G UR	cgactcaggtaccgggccccccctcgaggggaccccgtgggagagtaggacgccgccg
MRI1 G DF	ctggagctccaccgcggtggcggccgcccgcggccatccgcaccggggtgac
MRI1 G DR	ctgcagcccgggggatccactagttctagacggcctcgacggtcgtcatccgccgcc
flK UF1	aggaattcgatatcaagcttatcgataccgtcgacctcgaggggggggcccccgtccggatcgccgtaccgggcggtgg
flK UR1	aacttcgaagcagctccagcctacagggccactcctcactggtgccggggcgaag
flK DF1	ctgcaggtcgacggatccccggaatgcggatccggccccggcccgcgcc
flK DR1	tacgccaagctcggaattaaccctcactaaagggaacaaaagctgggtacccgtctccgggtggagttcgcggggcgtt
flK KF1	ccccggcaccagtgaggagtggccctgtaggctggagctgcttc
flK KR1	ggggcgcgggccggggccggatccgcattccggggatccgtcgacct

Supplementary Discussion 1. *De novo* assembly of the *S. cattleya* genome from short reads. During the course of this effort, four paired-end libraries were generated from *S. cattleya* genomic DNA and sequenced using the Illumina platform, providing 8.4 Gb of sequence information that represented an estimated 1,000-fold coverage of the predicted genome size [*3*]. Although the read length and quality have continued to increase since these libraries were sequenced, the resulting 45- and 61-base paired-end reads were sufficient to generate a final assembly on par with a draft genome sequence assembled from 84-base paired-end reads for a related actinomycetes [*25*].

Velvet assembly. The reads from all four libraries were combined and assembled using VelvetOptimiser, which allows for automated optimization of parameters such as hash length (*Table S2*, **1**) [*6*]. Although trimming the reads based on quality scores resulted in an increase in the total number of contigs and scaffolds in the assembly, this step helps to remove low quality bases and thus reduces the errors in the assembly. With the processed reads, VelvetOptimizer produced an 8.04 Mb genome assembly comprised of 343 total scaffolds and contigs with a slightly longer maximum scaffold length compared to the assembly generated from untrimmed reads (*Table S2*, **2**). After mapping reads using BWA [*7*] to obtain an accurate measure of the insert lengths and its standard deviation, the assembly was reduced to 205 contigs and scaffolds (*Table S2*, **3**). At this point, contigs that were both less than 100 bases in length and not included in a scaffold were removed from the assembly as they were assumed to originate from the termination of a missassembly at the short read level, leaving 8.05 Mb bases in 73 scaffolds and contigs. (*Table S2*, **4**)

With an initial draft in hand generated from VelvetOptimizer, we next used IMAGE [*8*] in order to extend the ends and close the gaps between contigs that had been placed together on a scaffold by Velvet. First, the scaffolds produced by Velvet were broken down into their constituent contigs, leaving the assembly with 7.96 Mb assembled in 2,281 contigs (*Table S2*, **5)**. After processing this assembly with IMAGE, a significant number of gaps were closed, increasing the assembly size to 8.09 Mb and reducing the number of contigs over ten-fold to 217 (*Table S2*, **6**). At this time, in order to join any new overlaps between contigs produced by IMAGE's extension of contig ends, an assembly with CAP3 [*26*] was performed. The resulting assembly was 8.08 Mb in a total of 195 contigs (*Table S2*, **7**). As a final assembly the contigs were examined using Consed [*27*] to explore any remaining overlaps between contigs. Surprisingly, a substantial number of overlaps between contigs were identified in Consed. Contigs that overlapped on the appropriate strand were joined if the ends of those contigs matched exactly and the overlaps were longer than 100 bases. This assembly resulted in 8.07

Mb in 154 contigs with a maximum contig length 327 kb and N_{25}, N_{50} and N_{75} of 188 kb, 133 kb and 78 kb respectively (*Table S2*, **8**).

SOAPdenovo. For comparison, we also tested a newer and more automated short-read assembler, SOAPdenovo63mer v 1.05 [*28*], to produce an initial assembly of 8.01 Mb over 252 scaffolds and contigs (Table S2, **9**). Gaps (71 gaps) were then closed in the resultant scaffolds using the Gapcloser utility in SOAPdenovo. The remaining gaps in scaffolds (62 gaps) were then manually closed by aligning reads to the *de novo* SOAPdenovo assembly with BWA, to produce an assembly of 192 contigs after removal of contigs <200 bases (Table S2, **10**).

Table S2. Stepwise assembly statistics for the *S. cattleya* draft genome. Descriptions of the state of the draft genome after each step in the assembly process are given. Initial assembly with unprocessed reads (**1**), assembly with trimmed reads (**2**) and assembly with defined insert length (**3**). Followed by removal of presumably misassembled short contigs (**4**), scaffolds broken up into their constitutive contigs (**5**), after automated gap closing (**6**), after automated joining of overlapping contigs (**7**) and after manual joining of overlapping contigs (**8**). Initial assembly by SOAPdenovo63mer (**9**) and final SOAP assembly after gap closing and elimination of short contigs (**10**).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Total bases (Mb)	8.02	8.04	8.06	8.05	8.00	8.09	8.08	8.07	8.01	8.00
No. contigs or scaffolds	198	343	205	73	2,281	217	195	154	252	192
Max scaffold or contig (kb)	503	522	1,787	1,787	55	263	263	370	301	301
N_{75} (kb)	103	78	181	181	3	56	52	78	72	73
N_{50} (kb)	180	152	1,065	1,065	7	91	93	133	143	158
N_{25} (kb)	281	327	1,346	1,346	11	157	151	188	210	211

Figure S2. Comparison of steps in the assembly of the *S. cattleya* draft genome with the completed genome. (A) Scaffolds and contigs produced by Velvet. (B) Contigs produced by Velvet after scaffolding information was removed. (C) After gap closing with Image. (D) After joining overlapping contigs with Cap3. The contigs produced by Velvet (*B*) showed fairly complete coverage of the genome, but the scaffolds those contigs were assembled into (*A*) were fairly inaccurate. Draft genome contigs were mapped to the completed genome using the NUCMER from the MUMMER package.

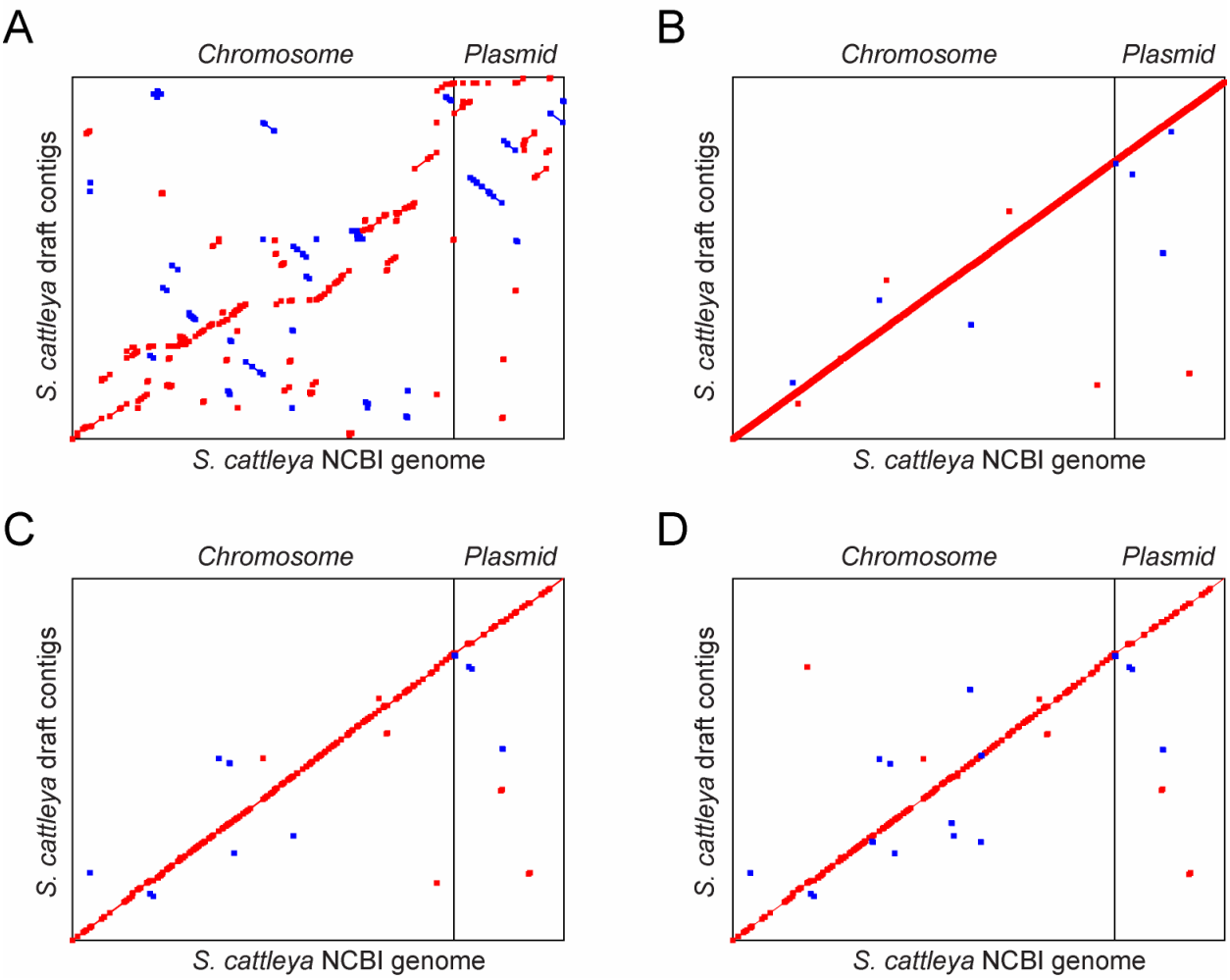
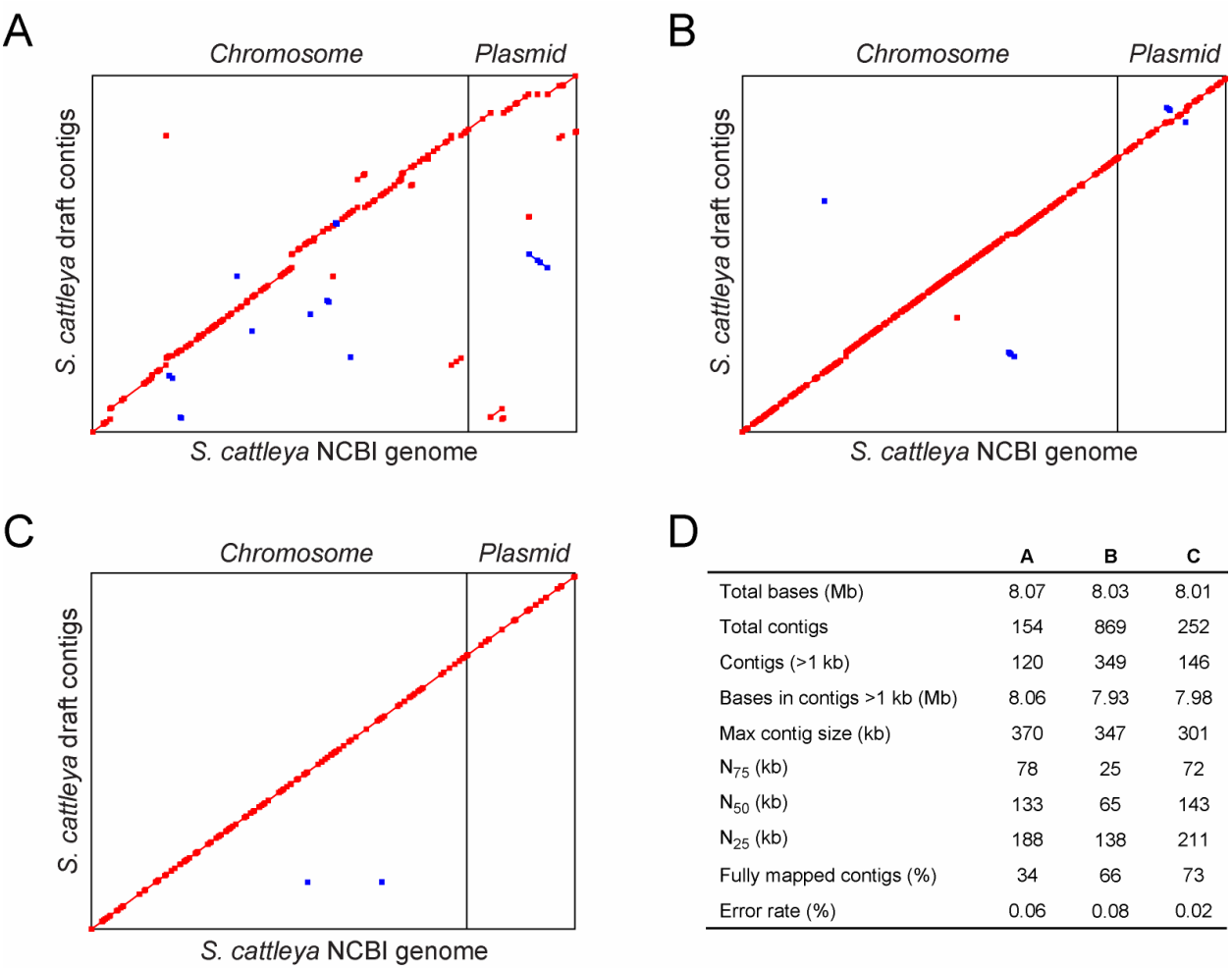


Figure S3. Comparison of draft genomes assembled with different versions of short read assemblers. (A) Velvet v 1.0.14 using VelvetOptimiser v 2.1.7. (B) Velvet v 1.2.03 using VelvetOptimiser v 2.2.0. (C) SOAPdenovo63mer v 1.05. (D) Statistics for assemblies.



D

	A	B	C
Total bases (Mb)	8.07	8.03	8.01
Total contigs	154	869	252
Contigs (>1 kb)	120	349	146
Bases in contigs >1 kb (Mb)	8.06	7.93	7.98
Max contig size (kb)	370	347	301
N_{75} (kb)	78	25	72
N_{50} (kb)	133	65	143
N_{25} (kb)	188	138	211
Fully mapped contigs (%)	34	66	73
Error rate (%)	0.06	0.08	0.02

Table S3. Identity of gaps in coverage of the *Streptomyces cattleya* draft genomes.

	Velvet assembly			SOAPdenovo assembly		
	Gaps	**Length of gaps (b)**	**% Total**	**Gaps**	**Length of gaps (b)**	**% Total**
rRNA	5	23,252	46.8	11	28,134	27.8
Transposase	25	11,295	22.7	28	28,167	27.8
PKS	6	4,700	9.5	20	7,798	7.7
Chromosome and plasmid ends	4	2,788	5.6	4	3,352	3.3
Hypothetical proteins	11	2,292	4.6	23	12,070	11.9
Transporter	1	1,509	3.0	2	3,746	3.7
Intergenic regions	14	1,508	3.0	27	10,420	10.3
Crotonyl-CoA reductase	1	1,071	2.2	2	1,859	1.8
Cytochrome *c* oxidase	1	782	1.6	5	811	0.8
Oxidoreductase	1	440	0.9	0	0	0
Nitrate reductase	1	84	0.2	0	0	0
NRPS	0	0	0	4	1,325	1.3
Transcriptional regulator	0	0	0	3	1,293	1.3
Sensor protein	0	0	0	2	149	0.1
tRNA	0	0	0	2	84	0.1
4-hydroxy-3-methylbut-2-ene-1-yl diphosphate synthase	0	0	0	1	1,010	1.0
Epsilon-poly-L-lysine synthase	0	0	0	1	1,011	1.0
Thymidine phosphorylase	0	0	0	1	54	0.1
Total	**70**	**49,721**	**100**	**136**	**101,283**	**100**

Table S4. Paralogs of proposed pathway genes in comparison to other sequenced streptomycetes. Numbers of predicted orthologs of proposed fluoroacetate biosynthetic pathway in selected *Streptomyces* spp. with fully sequenced genomes were identified based on their Pfam families, TIGRfam families or COG families.

	S. cattleya	*S. bingchenggensis*	*S. coelicolor*	*S. scabiei*	*S. griseus*	*S. avermitilis*
Fluorinase (FlA)	1	0	0	0	0	0
MTAP	3	1	1	1	1	1
MTR1P isomerase	2	1	1	2	1	2
F1P aldolase	2	2	1	1	1	1
ALDH	24	37	24	39	26	25

Figure S4. Preparation of cDNA libraries for RNA sequencing and DNA microarray experiments. Electropherogram showing RNA extracted from *S. cattleya* and removal of rRNA.

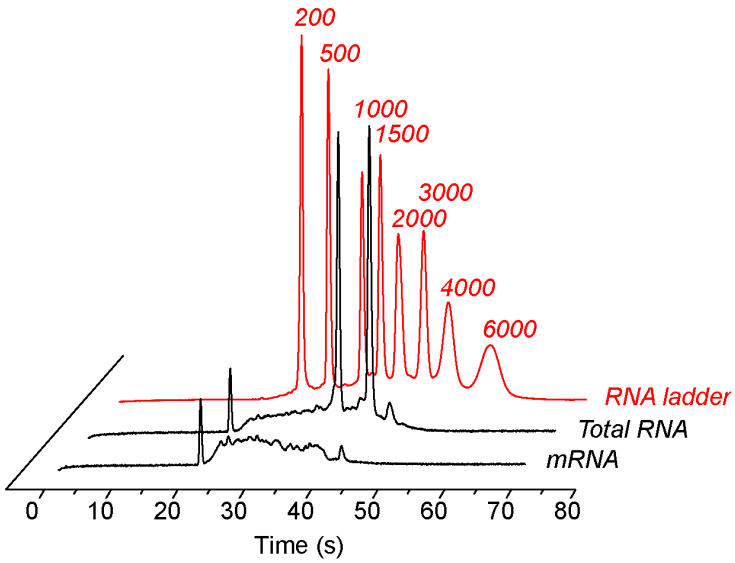


Figure S5. Comparison of RNA-seq and microarray data. Fold changes for genes that were identified as differentially expressed in both the microarray and RNA-seq data are compared.

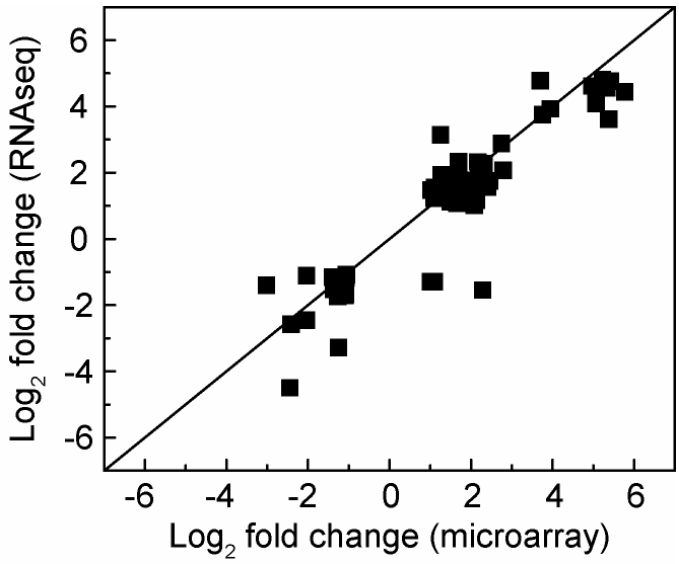


Figure S6. Genomic contexts of methylthioadenosine phosphorylases. *S. cattleya* contains two methylthioadenosine (MTA) phosphorylases in addition to FlB (SCAT_4166). SCAT_3528 is conserved across all of the fully sequenced and annotated *Streptomyces* spp. genomes while SCAT_2201 is located in a putative methionine salvage pathway context. SCAT_2201 is absent even in *S. avermitilis* and *S. scabiei*, which contain putative methionine salvage pathways as well as *S. coelicolor* and *S. griseus*, which have no predicted salvage pathway.

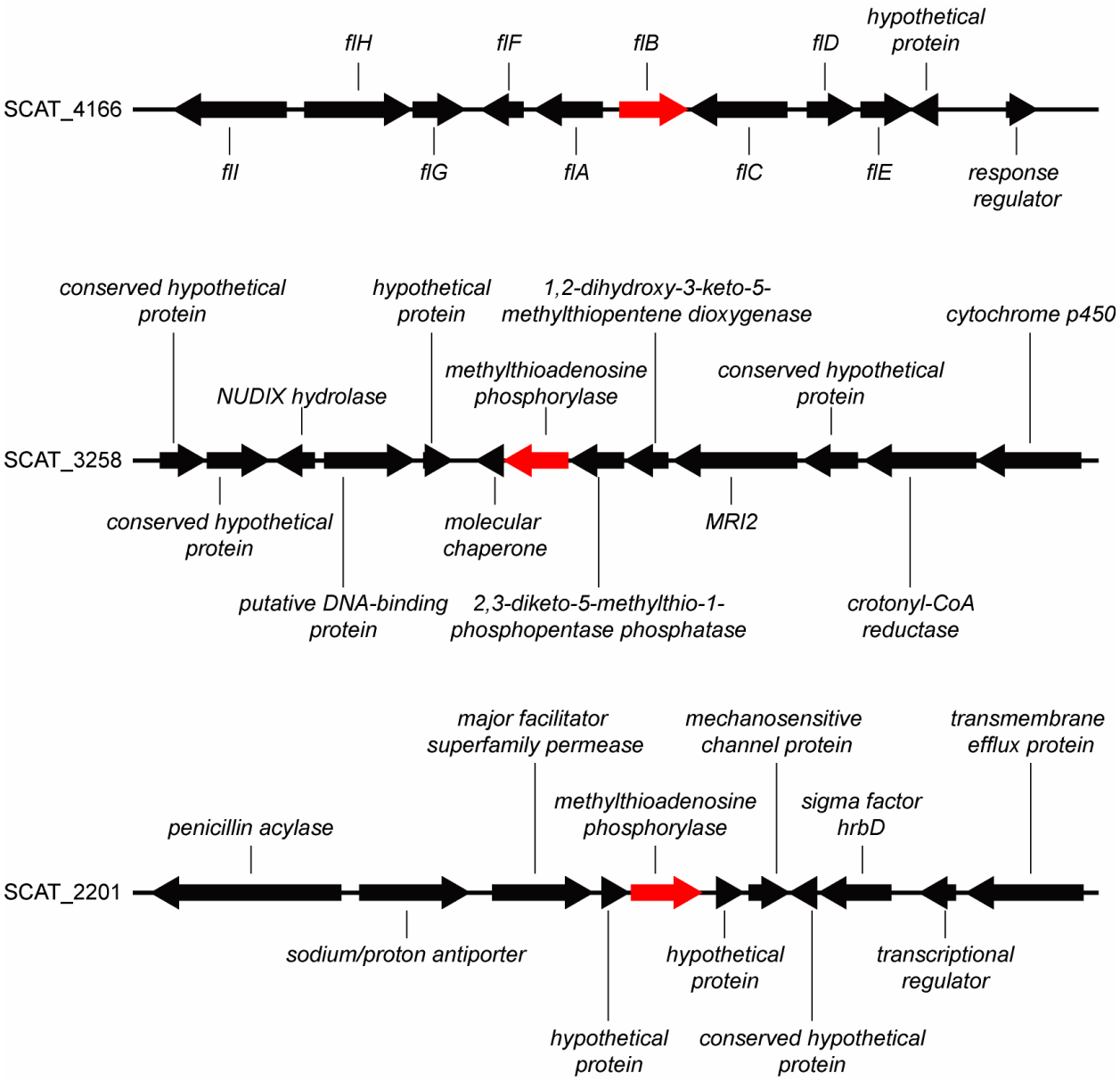


Figure S7. Genomic contexts of methylthioribose-1-phosphate isomerases.

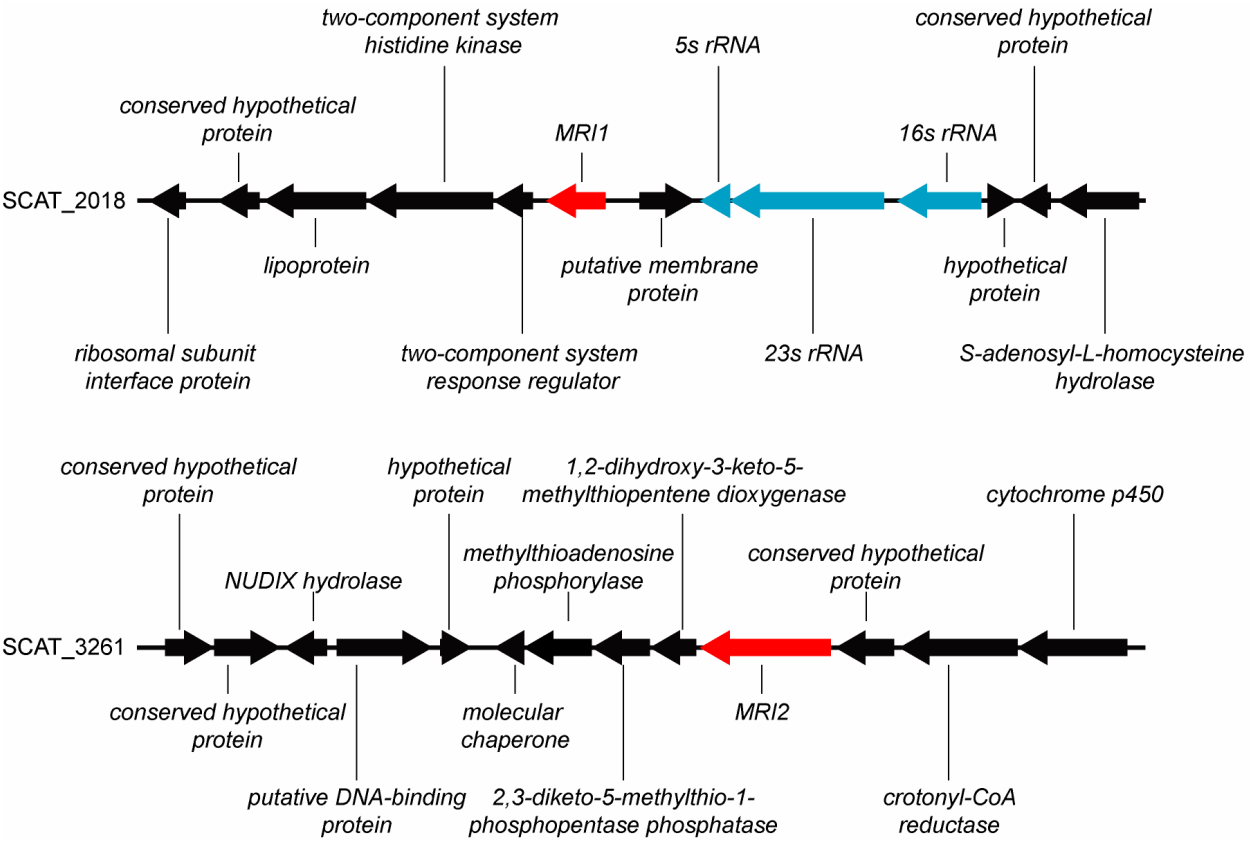


Figure S8. Genomic contexts of *mri2* homologs in *Nocardia farcinica*.

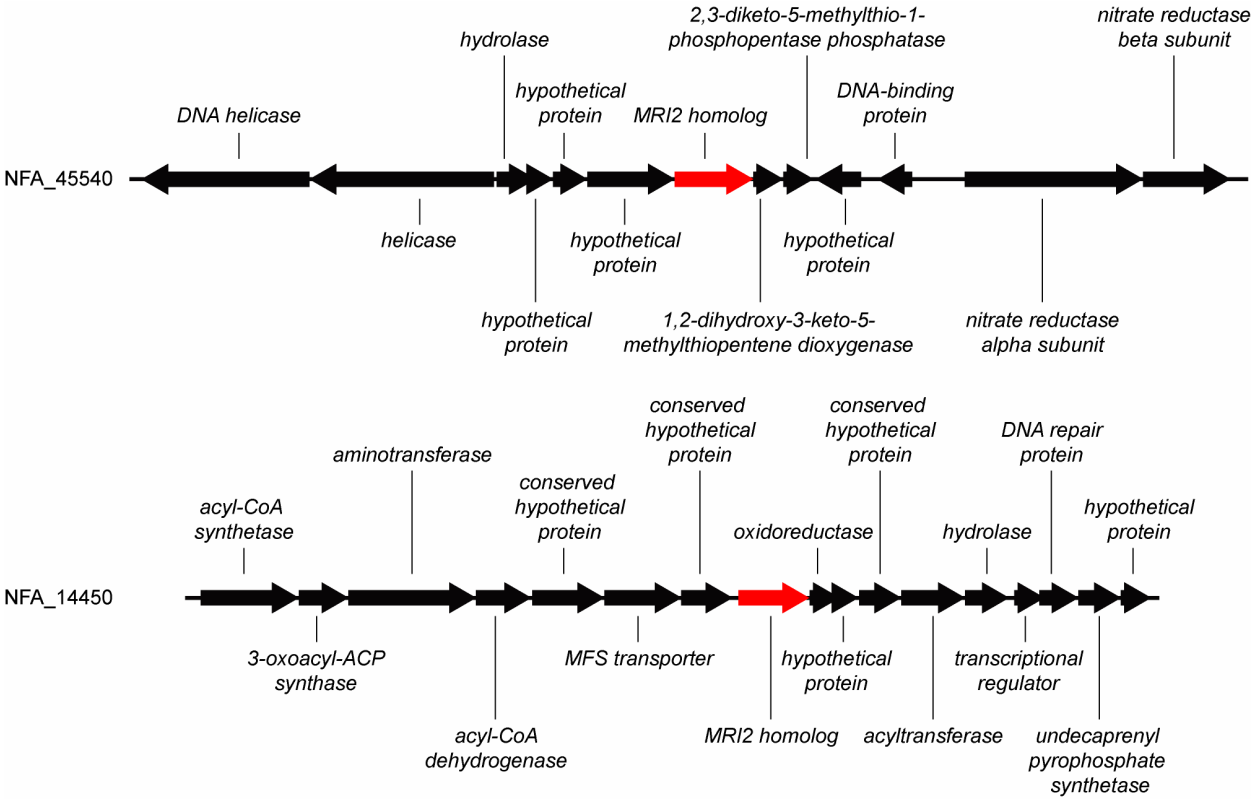


Figure S9. The *mri2* gene is adjacent to a predicted polyketide synthase/non-ribosomal peptide synthetase biosynthetic cluster. The sequence of the cluster was independently verified by Sanger sequencing of cosmids containing the *mri2* gene prepared from an *S. cattleya* genomic library.

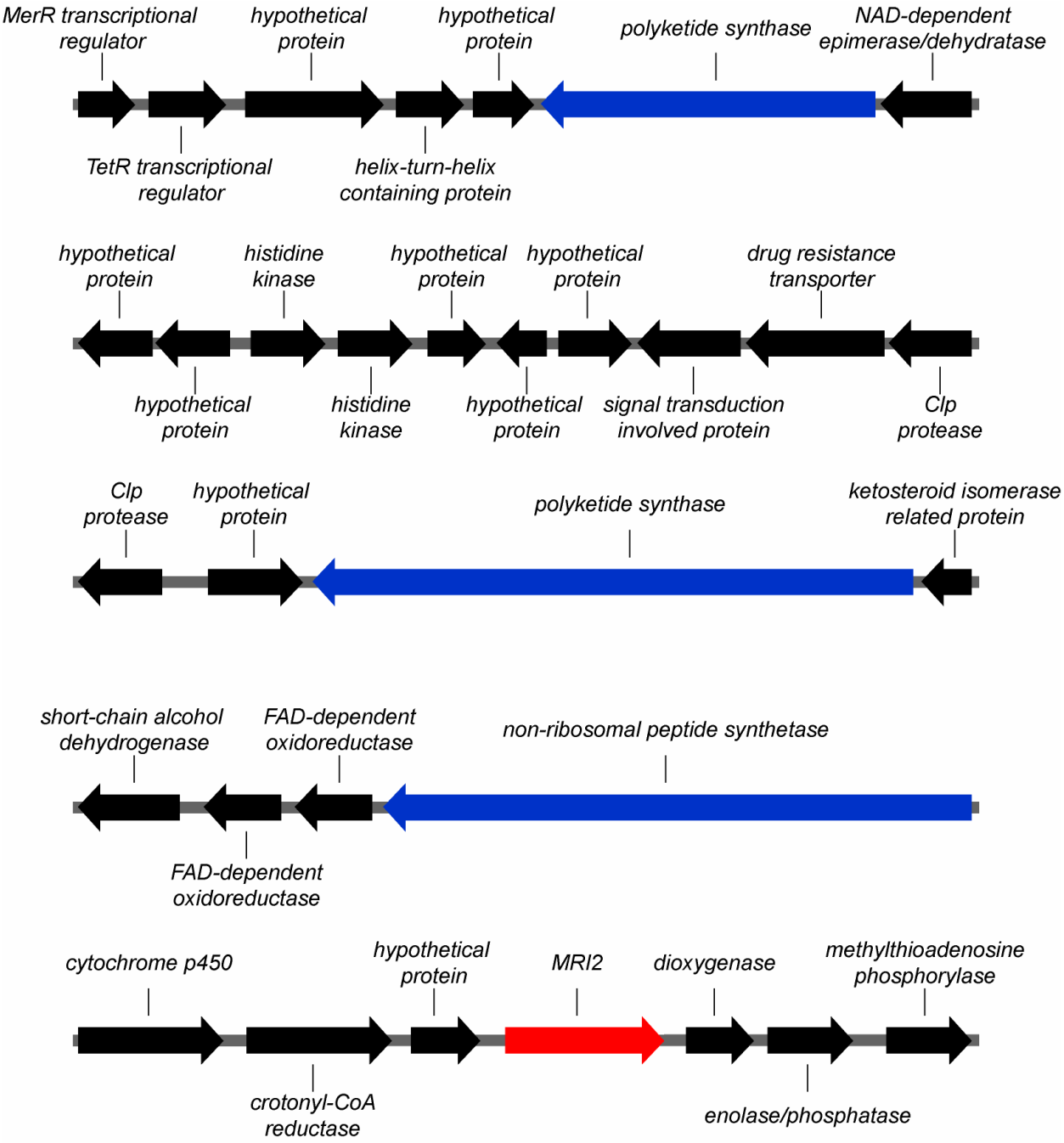


Figure S10. Generation and screening of a cosmid-based genomic library. (A) Optimization of Sau3A partial digest of genomic DNA from *S. cattleya* NRRL 8057 (ATCC 35852). (B) Screening of pooled cosmid libraries for the presence of the *mri2* fusion gene to identify cosmids for confirmation of its genomic context by Sanger sequencing and gene disruption using the REDIRECT method.

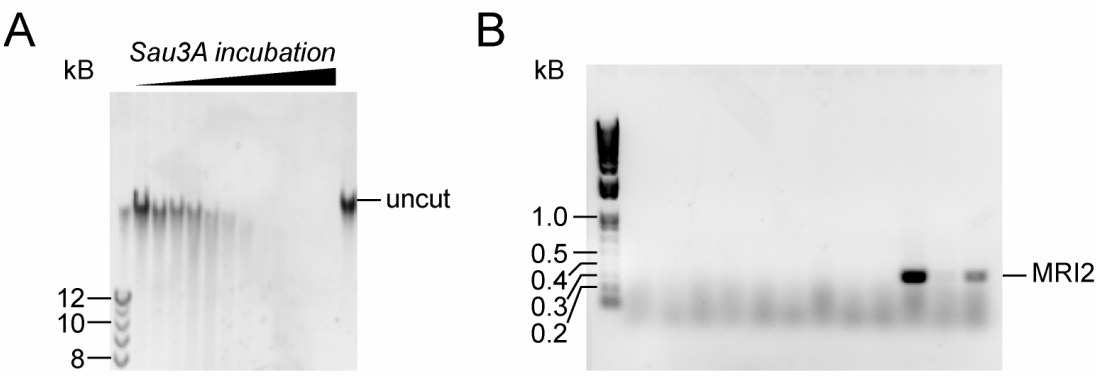


Figure S11. Characterization of the MTR1P isomerase/dehydratase fusion enzyme (MRI2) from *S. cattleya*. RP HPLC-MS trace of the MRI1 and MRI2 reactions. Traces shown is the sum of the extracted masses of the substrate (MTR1P, *m/z* 259) and the product (2,3-diketo-5-methylthiopentyl-1-phosphate, *m/z* 241).

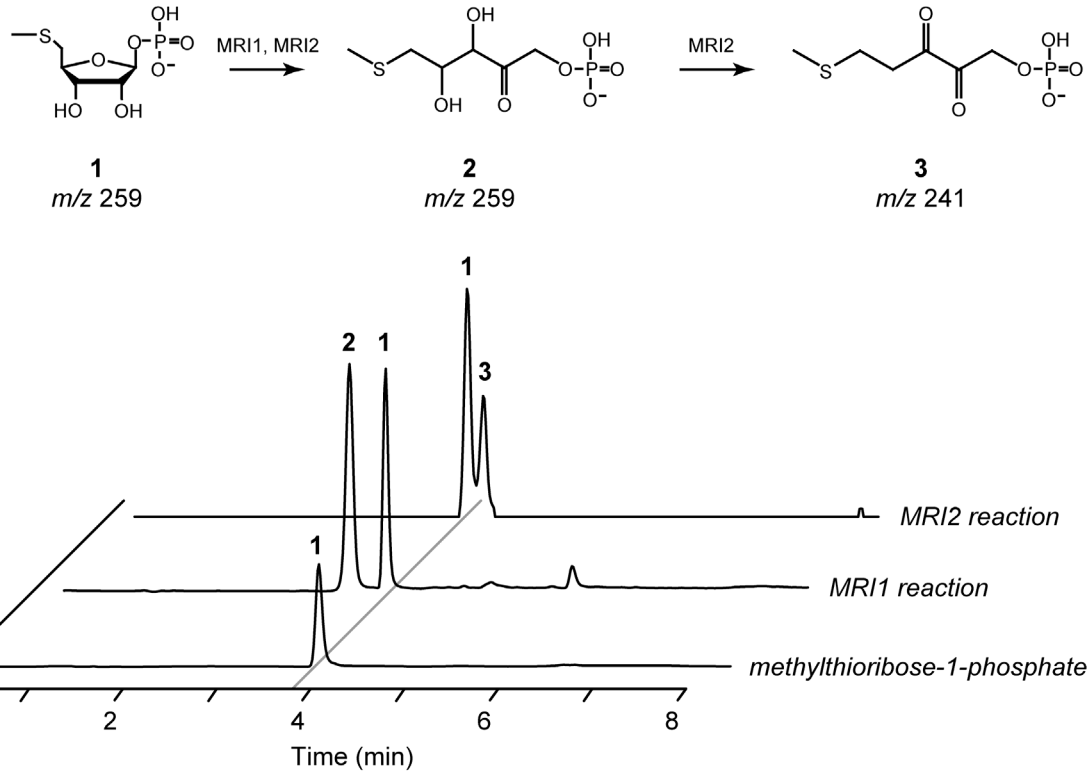


Figure S12. Characterization of *S. cattleya* wildtype, Δ*mri1*, Δ*mri2*, and Δ*mri1*Δ*mri2* strains. (A) Growth of the strains with the addition of fluoride (2 mM) on day 0. (B) Concentration of fluoride remaining in the media over time.

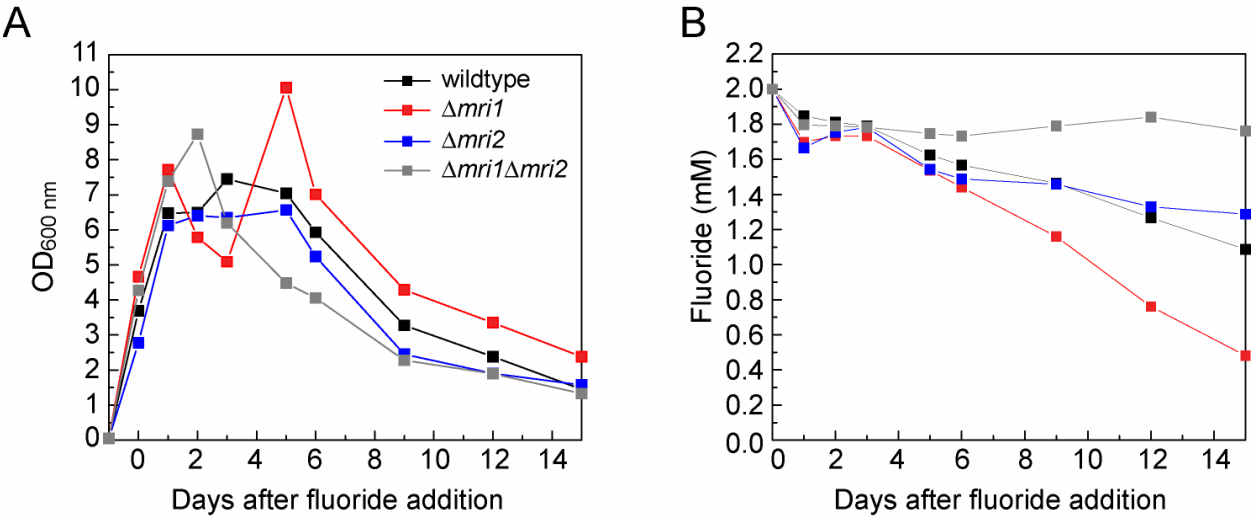


Figure S13. Growth curves for (A) *S. cattleya* wildtype and (B) Δ*flK:Am*^{R} strains in GYM (pH 5) in the absence (■) and presence of 2 mM sodium fluoride (■) or 2 mM sodium fluoroacetate (■). (C) Production levels of fluoroacetate and fluorothreonine by Δ*flK:Am*^{R} is similar to wt after 2 d.

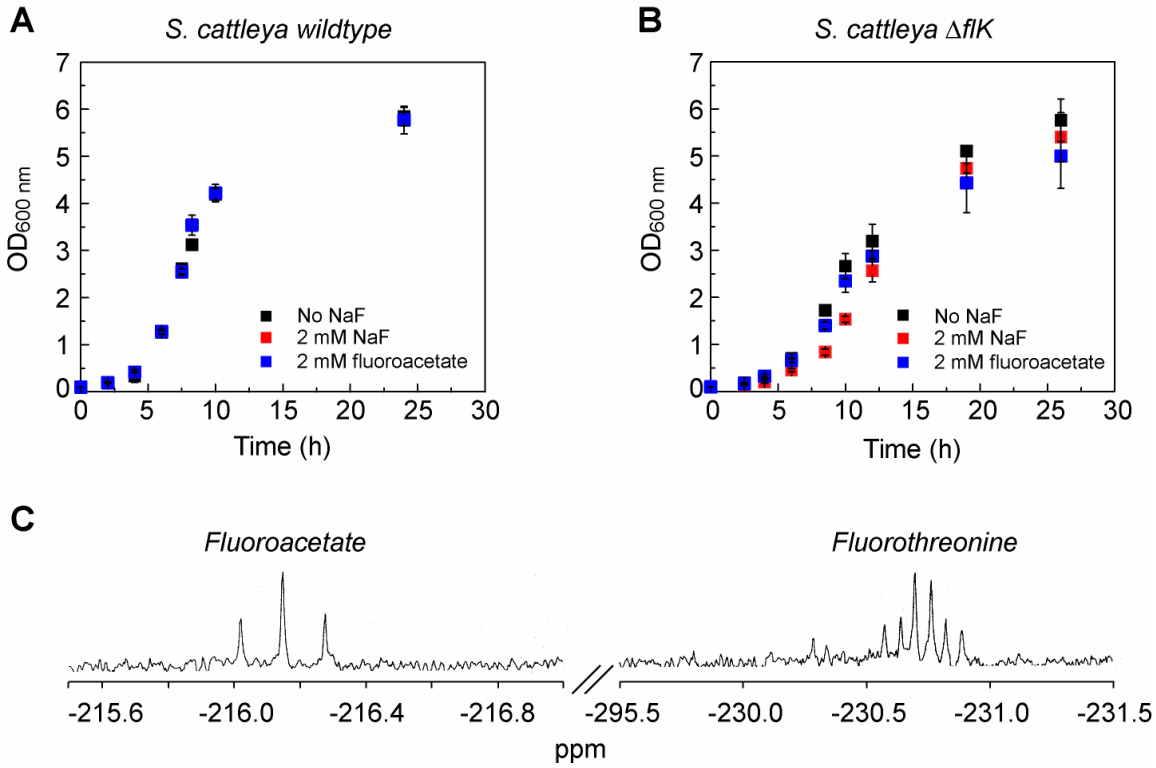


Figure S14. Heterologous expression and purification of acetate assimilation enzymes from *S. cattleya* and *E. coli*. Acetyl-CoA synthetase from (A) *S. cattleya* and (B) *E. coli*. Acetate kinase (AckA) from (C) *S. cattleya* and (D) *E. coli*. Phosphotransacetylase (PTA) from (E) *S. cattleya* and (F) *E. coli*. All gels are pre-induction (*lane 1*), post-induction (*lane 2*) and purified enzyme (*lane 3*).

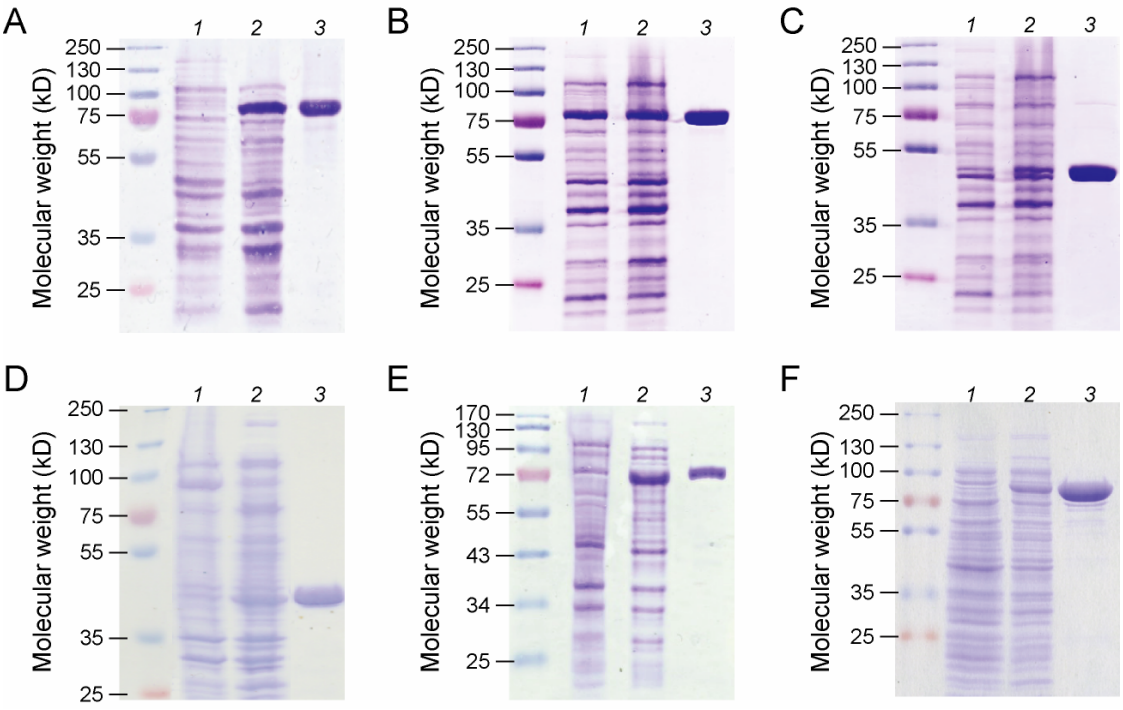


Figure S15. Phylogenetic tree for citrate synthases. *S. cattleya* contains four citrate synthases, which falls in the expected range based on other sequenced and annotated *Streptomyces* spp. genomes (*S. coelicolor, S. avermitilis,* and *S. griseus,* 4 CS; *S. scabiei* and *S. bingchenggensis*, 2 CS). Top 10 blast hits from the NCBI non-redundant protein database for each of the 4 *S. cattleya* citrate synthases as well as GltA from *E. coli* and CitZ from *B. subtilis* were aligned with ClustalW [*29*]. Tree was produced using the BioNJ algorithm with 10,000 replicate bootstraps in Seaview [*30*]. By homology to the major vegetative citrate synthases from *S. coelicolor* (92%) and *S. hygroscopis* (93%) as well as distance relationships with canonical hexameric and dimeric CSs, Cit1 should be the major CS in *S. cattleya*.

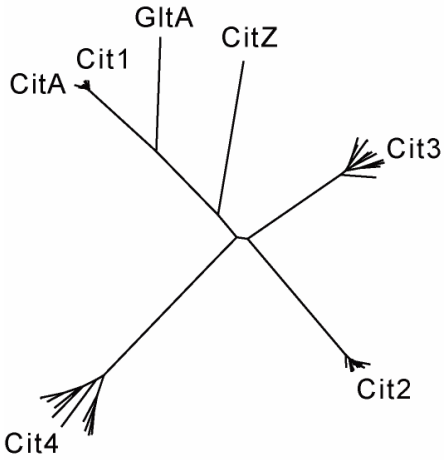


Figure S16. Heterologous expression in *E. coli* and purification of citrate synthases (A) GltA from *E. coli*, (B) CitA from *S. coelicolor*, (C) Cit1 and (D) Cit2 from *S. cattleya*. All gels are pre-induction (*lane 1*), post-induction (*lane 2*) and purified enzyme (*lane 3*).

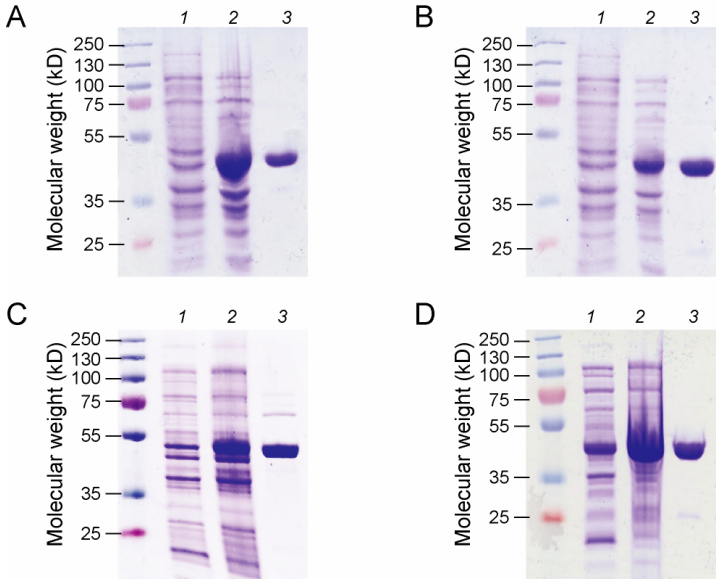


Figure S17. Dose-response curves of assayed citrate synthases fit to the Michaelis-Menten equation (black curve) and the Hill equation (red curve). *E. coli* GltA with respect to (A) acetyl-CoA and (B) fluoroacetyl-CoA. *S. coelicolor* CitA with respect to (C) acetyl-CoA and (D) fluoroacetyl-CoA. *S. cattleya* Cit1 with respect to (E) acetyl-CoA and (F) fluoroacetyl-CoA. *S. cattleya* Cit2 with respect to (G) acetyl-CoA and (H) fluoroacetyl-CoA.

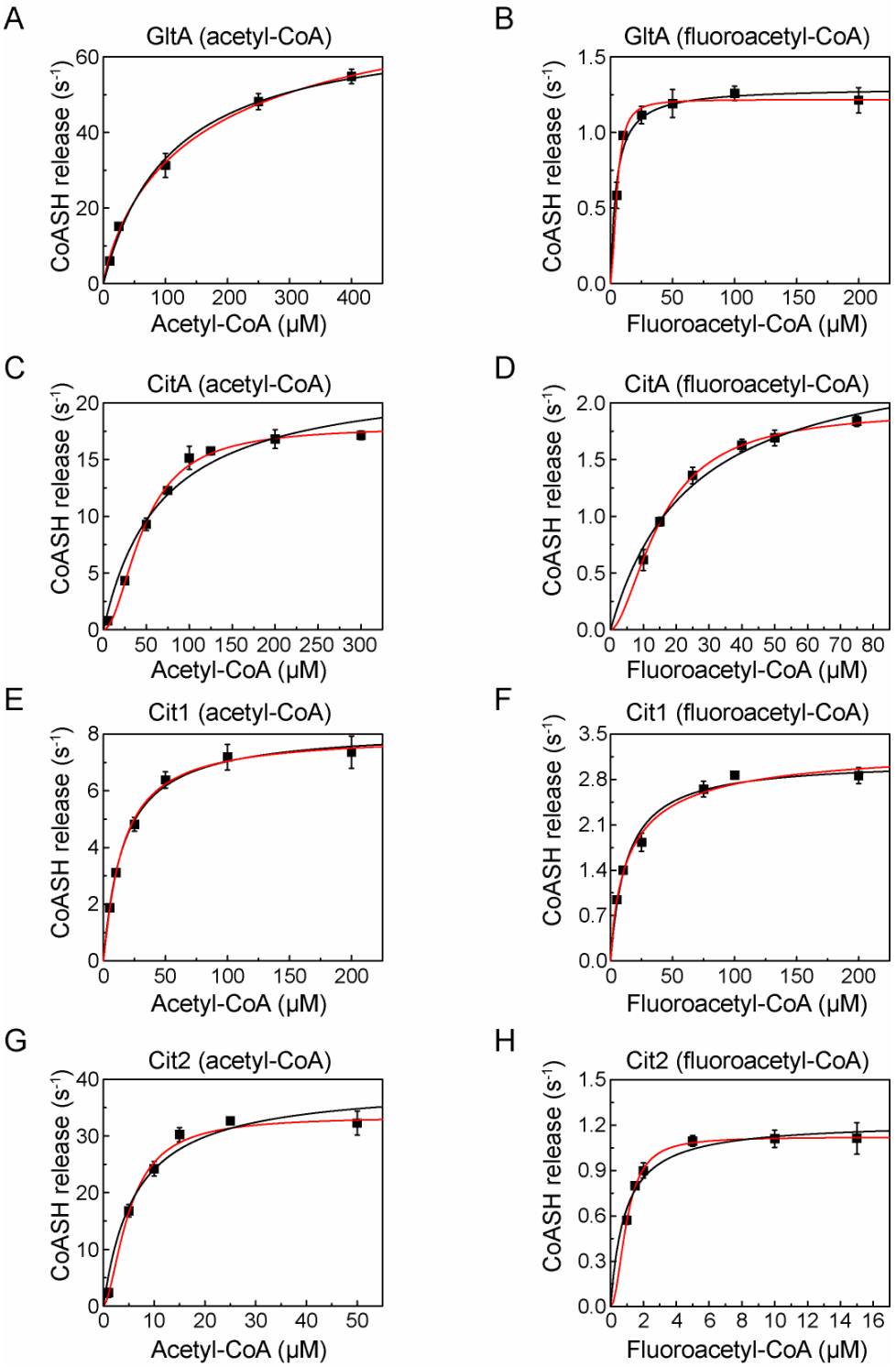


Figure S18. Multiple sequence alignment of Cit1 with related citrate synthases and GltA from *E. coli*. Residues important for dimer-dimer interactions in GltA are indicated with stars. Residues in this region that are conserved in Cit1's relatives, but not Cit1 are highlighted in red. Sequences were aligned using Muscle [*31*].

```
E. coli (GltA)            …GNFLNMMFSTPCEPYEVNPILERAMDRILILHADHEQNASTSTVRTAGSSGANPFACIAA…
                                  ****
S. cattleya (Cit1)        …ENFLRMTFAVPAEEYELDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANLFASISA…
S. hygroscopicus          …ENFLRMTFSVPAQEYDLDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. bingchenggensis        …ENFLRMTFSVPAQEYELDPIVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. violaceusniger         …ENFLRMTFSVPAQEYDLDPTVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. coelicolor (CitA)      …ENFLRMTFSVPAQEYELDPTVVAALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. pristinaespiralis      …ENFLRMTFSVPAQEYELDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. clavuligerus           …ENFLRMTFSVPAQEYDLDPVVVAALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. griseoaurantiacus      …ENFLRMTFSVPAQEYDLDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. avermitilis            …ENFLRMTFSVPAQDYDLDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
S. scabiei                …ENFLRMTFSVPAQEYDLDPVVVSALDKLLILHADHEQNCSTSTVRLVGSSQANMFASISA…
```

Figure S19. Decoupling of acetyl-CoA and fluoroacetyl-CoA hydrolysis from citrate and fluorocitrate synthesis. As the literature CS assay monitors CoA release, which can also arise from hydrolysis of the acyl-CoA, the extent of conversion of acetyl-CoA and fluoroacetyl-CoA to citrate (red) and fluorocitrate (gray), respectively, was quantified after the reaction was allowed to run to completion. All citrate synthases demonstrated substantial decoupling of fluoroacetyl-CoA hydrolysis from fluorocitrate production, but the level of decoupling was not substantially different for the *S. cattleya* homologs.

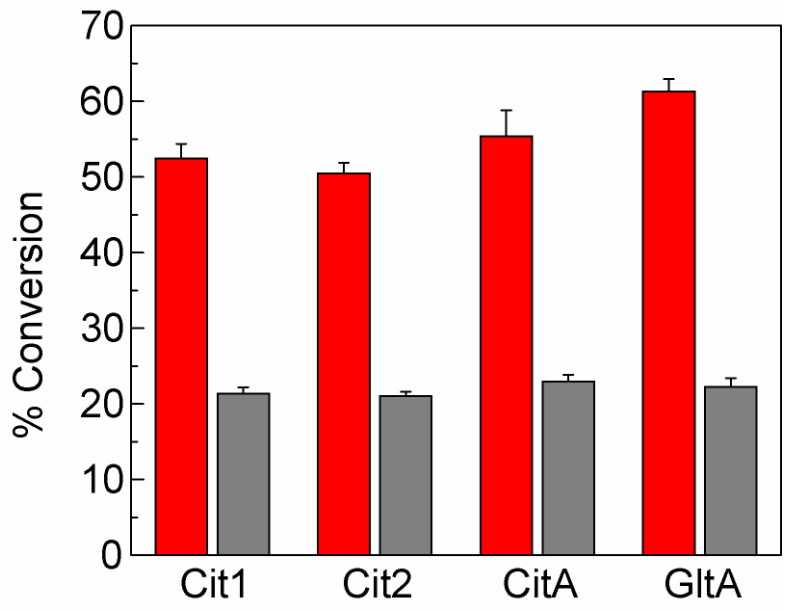


Figure S20. Fluorocitrate inhibition studies of aconitase from *S. cattleya*. *S. cattleya* cultures grown in the presence (A) and absence (B) of 2 mM fluoride. Aconitase activity was monitored in cell lysates by a coupled assay with isocitrate dehydrogenase in the presence of 0 μM fluorocitrate (■), 10 μM fluorocitrate (■), and 100 μM fluorocitrate (■).

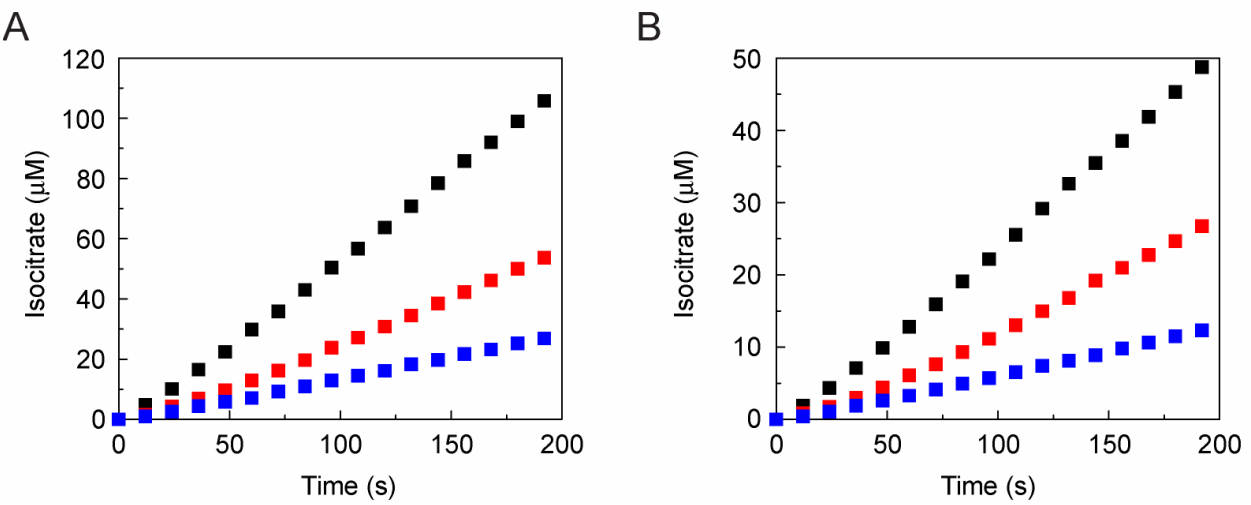
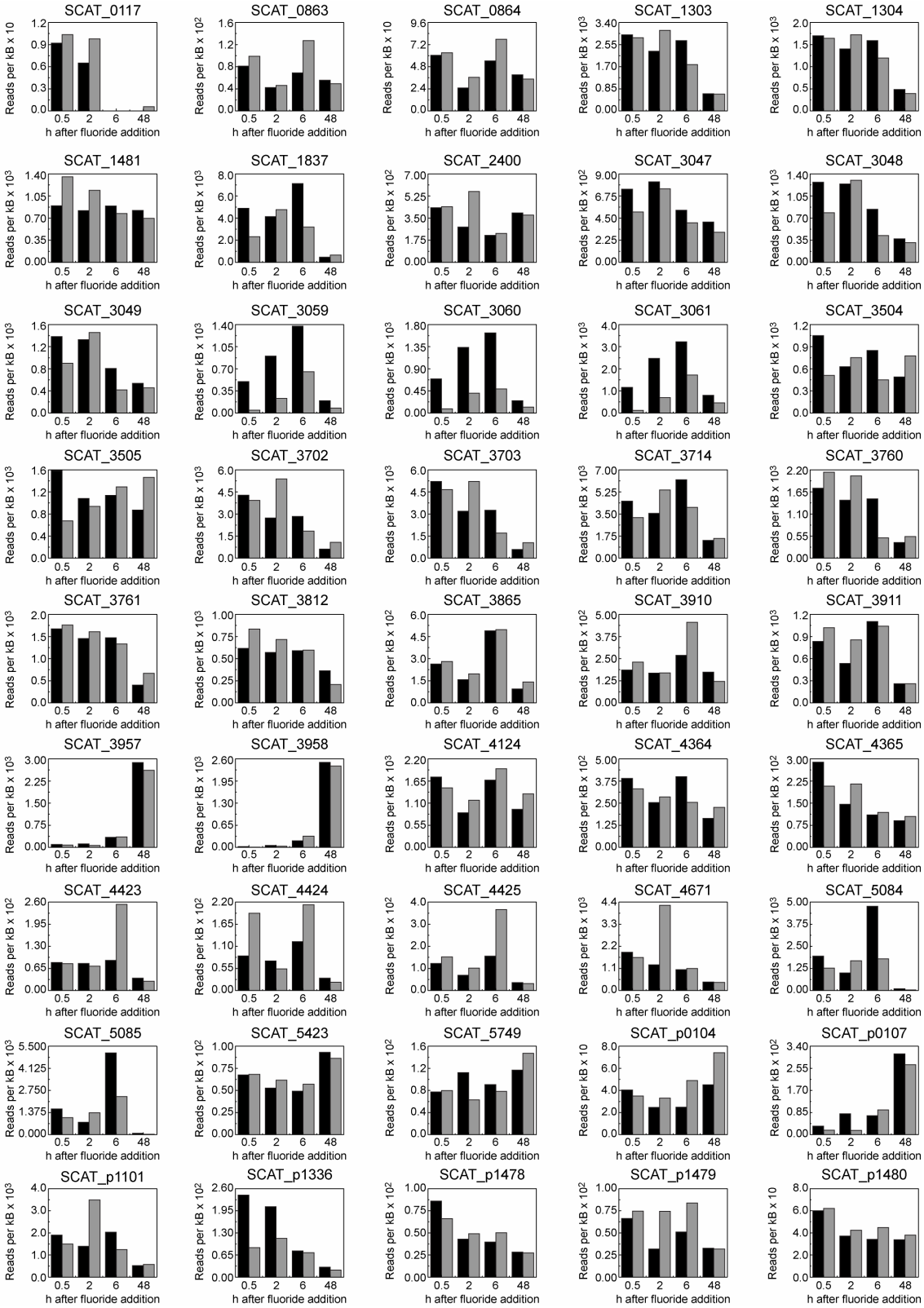


Figure S21. Time courses of transcription levels for predicted TCA cycle and related genes in the presence (black) and absence (grey) of fluoride.



SCAT_0117	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component
SCAT_0863	Cit3
SCAT_0864	Cit4
SCAT_1303	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component
SCAT_1304	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component
SCAT_1481	Pyruvate dehydrogenase complex, dehydrogenase (E1) component
SCAT_1837	Cit1
SCAT_2400	Cit2
SCAT_3047	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component
SCAT_3048	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_3049	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_3059	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component
SCAT_3060	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_3061	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_3504	Pyruvate:ferredoxin oxidoreductase
SCAT_3505	Pyruvate:ferredoxin oxidoreductase
SCAT_3702	Succinyl-CoA synthetase, beta subunit
SCAT_3703	Succinyl-CoA synthetase, alpha subunit
SCAT_3714	Malate dehydrogenase
SCAT_3760	Succinate dehydrogenase/fumarate reductase
SCAT_3761	Succinate dehydrogenase/fumarate reductase
SCAT_3812	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component
SCAT_3865	Phosphoenolpyruvate carboxykinase
SCAT_3910	Fumarase
SCAT_3911	Fumarase
SCAT_3957	Succinate dehydrogenase
SCAT_3958	Succinate dehydrogenase
SCAT_4124	α-Ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_4364	Citrate lyase
SCAT_4365	Citrate lyase
SCAT_4423	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component
SCAT_4424	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_4425	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component,
SCAT_4671	Aconitase
SCAT_5084	Succinate dehydrogenase
SCAT_5085	Succinate dehydrogenase
SCAT_5423	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component
SCAT_5749	Citrate lyase
SCAT_p0104	Citrate lyase
SCAT_p0107	Citrate lyase
SCAT_p1101	Isocitrate dehydrogenase
SCAT_p1336	Citrate lyase
SCAT_p1478	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_p1479	Pyruvate/α-ketoglutarate dehydrogenase complex, dehydrogenase (E1) component
SCAT_p1480	Pyruvate/α-ketoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component

Figure S22. Transcriptional response of the *crcB* locus in the presence (red) and absence (grey) of fluoride (2 mM) at 48 h. Recently, structural RNAs with the *crcB* motif have been reported to be fluoride-responsive riboswitches [*32*]. Searching the *S. cattleya* genome with the co-variance model for the *crcB* motif using Infernal [*33*] did not yield any hits; however, a *crcB* gene in *S, cattleya* is upregulated 2.1-fold in the presence of fluoride suggesting that *S. cattleya* may export fluoride under these growth conditions. Additionally, a substantial number of reads map to the genome in the intergenic region upstream from the *crcB* gene possibly indicating the presence of a structured 5´-UTR that could be a fluoride riboswitch. RNA-seq reads were aligned using BWA and the resulting BAM files were visualized with Artemis.

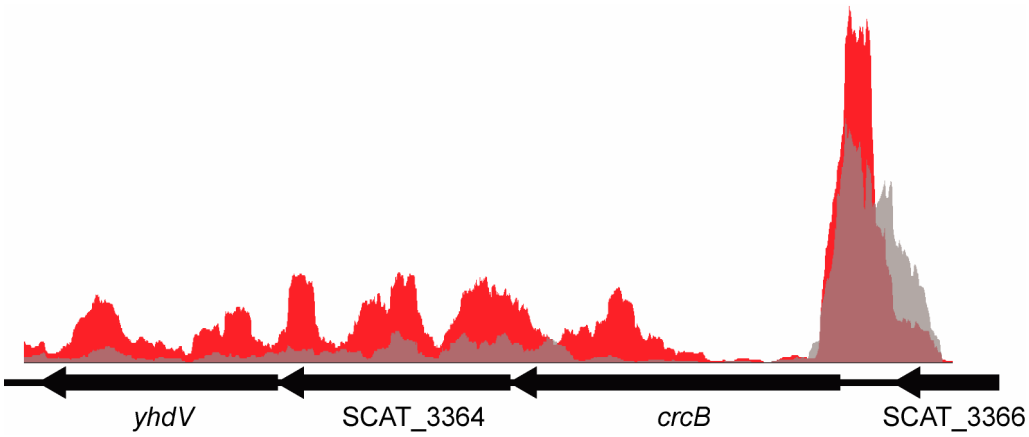


Figure S23. Alignment of "selectivity filter" residues from select *eriC* orthologs. *S. cattleya* also contains an ortholog of *eriC,* another gene often found under the control of a fluoride-responsive riboswitch. The *S. cattleya* EriC appears to contain "selectivity filter" residues that are more similar to those of chloride specific transporters than those whose substrate is fluoride [*32*] and is not differentially expressed in the presence of fluoride under these conditions. Alignments were generated with Muscle. Residues from the *S. cattleya eriC* that are conserved among the chloride specific and not the fluoride specific *eriC*s are highlighted in red.

Pseudomonas syringae (F)	GNNLI	GREGT	GEVTP	Y
Clostridium difficile (F)	GMNLI	GREGV	GEVTP	Y
Arabidopsis lyrata (nitrate)	GPGIP	GKEGP	GLFLP	Y
Escherichia coli (Cl)	GSGIP	GREGP	GIFAP	Y
Homo sapiens (Cl)	GSGIP	GKEGP	GLFIP	Y
Streptomyces cattleya	GHGMP	GAEGP	GVLAP	L

Table S5. Alignment and predicted substrate of acetyl-transferase (AT) domains in predicted polyketide synthases. Forty-four AT domains were identified using Hmmer and aligned with ClustalW. Substrate specificity was predicted based on previously reported patterns for AT domain specificity [*34*], suggesting that 27 likely use a malonyl-CoA extender unit and 8 likely use a methylmalonyl-CoA extender unit, leaving 9 utilizing an ambiguous extender unit

	Active site residues	Predicted substrate
gi|337763529|emb|CCB72237.1|/87-98	-QGMSFGACHSGA	Unclear
gi|337762631|emb|CCB71339.1|/2628-2632	--GHGVG------	Unclear
gi|337767566|emb|CCB76277.1|/2103-2109	-HGDGAGH-----	Unclear
gi|337764166|emb|CCB72875.1|/547-558	QQGHSVGRFH-HV	Malonyl-CoA
gi|337764166|emb|CCB72875.1|/1535-1546	QQGHSVGRFH-HV	Malonyl-CoA
gi|337764165|emb|CCB72874.1|/3358-3369	QQGHSVGRFH-HV	Malonyl-CoA
gi|337769649|emb|CCB78362.1|/573-585	QQGHSFGRFHDQV	Malonyl-CoA
gi|337764165|emb|CCB72874.1|/5074-5086	QQGHSVGRFHSHV	Malonyl-CoA
gi|337764212|emb|CCB72921.1|/636-648	NQGHSVGRFHSNV	Malonyl-CoA
gi|337763789|emb|CCB72499.1|/1577-1589	QQGHSQGRSHTNV	Methylmalonyl-CoA
gi|337764088|emb|CCB72797.1|/635-647	YQGHSMGRGHTNV	Unclear
gi|337767569|emb|CCB76280.1|/357-369	QQGHSLGRFHNQV	Malonyl-CoA
gi|337762633|emb|CCB71341.1|/542-554	QQGHSVGRFHNHV	Malonyl-CoA
gi|337767571|emb|CCB76282.1|/305-317	QQGHSMGRSHTNV	Methylmalonyl-CoA
gi|337763797|emb|CCB72507.1|/2230-2242	QQGHSVGRFHNHV	Malonyl-CoA
gi|337763789|emb|CCB72499.1|/3279-3291	QQGHSIGRFHDHV	Malonyl-CoA
gi|337764170|emb|CCB72879.1|/575-587	QQGHSVGRFHLQV	Possibly malonyl-CoA
gi|337767263|emb|CCB75974.1|/1565-1577	QQGHSVGRFHNHV	Malonyl-CoA
gi|337762738|emb|CCB71446.1|/570-582	QQGHSCGRFHCQV	Possibly malonyl-CoA
gi|337764169|emb|CCB72878.1|/2625-2637	QQGHSQGRSHTNV	Methylmalonyl-CoA
gi|337764165|emb|CCB72874.1|/573-585	QQGHSVGRFHTQV	Malonyl-CoA
gi|337762631|emb|CCB71339.1|/575-587	QQGHSQGRSHANV	Possibly methylmalonyl-CoA
gi|337762615|emb|CCB71323.1|/3215-3227	QQGHSVGRFHDSV	Possibly malonyl-CoA
gi|337763983|emb|CCB72693.1|/476-488	NQGHSVGRFHTNV	Malonyl-CoA
gi|337763789|emb|CCB72499.1|/554-566	QQGHSIGRFHDHV	Malonyl-CoA
gi|337767566|emb|CCB76277.1|/621-627	-RAADPVE-----	Unclear
gi|337762630|emb|CCB71338.1|/546-558	VQGYSVGRFHNHV	Malonyl-CoA
gi|337767263|emb|CCB75974.1|/4977-4989	QQGHSVGRFHNHV	Malonyl-CoA
gi|337765502|emb|CCB74211.1|/11-23	QQGHSVGRFHNQV	Malonyl-CoA
gi|337762615|emb|CCB71323.1|/1241-1253	QQGHSQGRSHTNV	Methylmalonyl-CoA
gi|337764166|emb|CCB72875.1|/3280-3292	QQGHSQGRSHTNV	Methylmalonyl-CoA
gi|337762632|emb|CCB71340.1|/2137-2149	QQGHSVGRFHNHV	Malonyl-CoA
gi|337763798|emb|CCB72508.1|/574-586	QQGHSIGRFHNHV	Malonyl-CoA
gi|337767263|emb|CCB75974.1|/3268-3280	QQGHSQGRSHANV	Possibly methylmalonyl-CoA
gi|337762632|emb|CCB71340.1|/549-561	QQGHSVGRFHNHV	Malonyl-CoA
gi|337763797|emb|CCB72507.1|/471-483	QQGHSVGRFHNHV	Malonyl-CoA
gi|337762615|emb|CCB71323.1|/4701-4713	QQGHSVGRFHDHV	Malonyl-CoA
gi|337769734|emb|CCB78447.1|/537-549	QQGHSQGRSHTNV	Methylmalonyl-CoA
gi|337762630|emb|CCB71338.1|/2038-2050	VQGYSLGRFHDHV	Malonyl-CoA
gi|337767569|emb|CCB76280.1|/12-24	QQGASLGQFHGVF	Possibly malonyl-CoA
gi|337767263|emb|CCB75974.1|/546-558	GLGHSMGRGHNHV	Unclear
gi|337764169|emb|CCB72878.1|/569-581	QHGQSCGRGHPQV	Unclear
gi|337762615|emb|CCB71323.1|/268-280	QQGHCLGWGHTNV	Unclear
gi|337764165|emb|CCB72874.1|/1581-1593	QQAHSVGRTHTNV	Unclear

Table S6. Predicted substrate of adenylation (A) domains in predicted non-ribosomal synthetases (NRPS) and hybrid PKS/NRPSs. Twenty-eight A domains were identified using Hmmer. Substrate specificity for all but three domains was predicted by NRPSpredictor2 [*35*].

	Specificity
gi|337766678|emb|CCB75389.1|/48-450	2,3-Dihydroxybenzoic acid
gi|337766678|emb|CCB75389.1|/1106-1508	Asparigine
gi|337766678|emb|CCB75389.1|/2176-2572	Tyrosine
gi|337769658|emb|CCB78371.1|/286-703	3-Amino-adipic acid
gi|337769658|emb|CCB78371.1|/1379-1789	Cysteine
gi|337769658|emb|CCB78371.1|/2451-2859	Valine
gi|337766677|emb|CCB75388.1|/493-891	Phenylalanine
gi|337766677|emb|CCB75388.1|/1548-1946	Phenylalanine
gi|337762498|emb|CCB71206.1|/28-421	Ornithine
gi|337762498|emb|CCB71206.1|/1120-1517	Serine
gi|337767258|emb|CCB75969.1|/457-862	Serine
gi|337767258|emb|CCB75969.1|/1519-1926	Proline
gi|337762953|emb|CCB71661.1|/42-477	Unclear
gi|337762953|emb|CCB71661.1|/1156-1562	β-Hydroxy-tyrosine
gi|337763497|emb|CCB72205.1|/41-441	Isoleucine
gi|337763497|emb|CCB72205.1|/1086-1489	Proline
gi|337762947|emb|CCB71655.1|/539-954	3,5-Dihydroxyphenylglycine
gi|337762497|emb|CCB71205.1|/497-578	Unclear
gi|337762497|emb|CCB71205.1|/588-888	Unclear
gi|337763496|emb|CCB72204.1|/527-919	Isoleucine
gi|337762967|emb|CCB71675.1|/26-419	β-Lysine
gi|337766688|emb|CCB75399.1|/282-688	Glutamine
gi|337762359|emb|CCB71065.1|/94-525	Threonine
gi|337769735|emb|CCB78448.1|/64-466	Cysteine
gi|337762361|emb|CCB71067.1|/489-885	Leucine
gi|337766676|emb|CCB75387.1|/1049-1449	Arginine
gi|337767563|emb|CCB76274.1|/532-929	Glycine
gi|337762934|emb|CCB71642.1|/501-900	D-Lysergic acid

Literature Cited

1. Palmer, B.R. and Marinus, M.G. The dam and dcm strains of *Escherichia coli* - A review. *Gene* **143**, 1-12 (1994).
2. Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. Evidence that the extracytoplasmic function sigma factor ς^E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**, 204-211 (1999).
3. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and D.A, H. Practical *Streptomyces* genetics. (The John Innes Foundation, Norwich, England; 2000).
4. Ji, N., Peng, B., Wang, G., Wang, S., and Peng, X. Universal primer PCR with DGGE for rapid detection of bacterial pathogens. *J. Microbiol. Methods* **57**, 409-413 (2004).
5. Lu, J.-J., Perng, C.-L., Lee, S.-Y., and Wan, C.-C. Use of PCR with universal primers and restriction endonuclease digestions for detection and identification of common bacterial pathogens in cerebrospinal fluid. *J. Clin. Microbiol.* **38**, 2076-2080 (2000).
6. Zerbino, D.R. and Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829 (2008).
7. Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
8. Tsai, I., Otto, T., and Berriman, M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* **11**, R41 (2010).
9. Markowitz, V.M. et al. IMG ER: A system for microbial genome annotation expert review and curation. *Bioinformatics* **25**, 2271-2278 (2009).
10. Barbe, V. et al. Complete genome sequence of *Streptomyces cattleya* NRRL 8057, a producer of antibiotics and fluorometabolites. *J. Bacteriol.* **193**, 5055-5056 (2011).
11. Anders, S. and Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, 1-12 (2010).
12. Bolstad, B.M., Irizarry, R.A., Åstrand, M., and Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).
13. Saeed, A. et al. TM4: A free, open-source system for microarray data management and analysis. *BioTechniques* **34**, 374-378 (2003).
14. Ashida, H. et al. A functional link between RuBisCO-like protein of *Bacillus* and photosynthetic RuBisCO. *Science* **302**, 286-290 (2003).
15. Waffenschmidt, S. and Jaenicke, L. Assay of reducing sugars in the nanomole range with 2,2'-bicinchoninate. *Anal. Biochem.* **165**, 337-340 (1987).
16. Huang, F. et al. The gene cluster for fluorometabolite biosynthesis in *Streptomyces cattleya*: A thioesterase confers resistance to fluoroacetyl-coenzyme A. *Chem. Biol.* **13**, 475-484 (2006).
17. Williamson, J.R. and Corkey, B.E. Assays of intermediates of the citric acid cycle and related compounds by fluorometric enzyme methods. *Method. Enzymol.*, 434-513 (1969).
18. Srere, P.A., Brazil, H., and Gonen, L. The citrate condensing enzyme of pigeon breast muscle and moth flight muscle. *Acta Chem. Scand.* **19**, 129-134 (1963).

19. Henson, C.P. and Cleland, W.W. Purification and kinetic studies of beef liver cytoplasmic aconitase. *J. Biol. Chem.* **242**, 3833-3838 (1967).
20. Gust, B. et al. λ Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv. Appl. Microbiol.*, 107-128 (2004).
21. Moker, N. et al. Deletion of the genes encoding the MtrA-MtrB two-component system of *Corynebacterium glutamicum* has a strong influence on cell morphology, antibiotics susceptibility and expression of genes involved in osmoprotection. *Mol. Microbiol.* **54**, 420-438 (2004).
22. Zahrt, T.C. and Deretic, V. An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *J. Bacteriol.* **182**, 3832-3838 (2000).
23. Hutchings, M.I. Unusual two-component signal transduction pathways in the actinobacteria. *Adv. Appl. Microbiol.* **61**, 1-26 (2007).
24. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).
25. Brown, M.E., Walker, M.C., Nakashige, T.G., Iavarone, A.T., and Chang, M.C.Y. Discovery and characterization of heme enzymes from unsequenced bacteria: Application to microbial lignin degradation. *J. Am. Chem. Soc.* **133**, 18006-18009 (2011).
26. Huang, X. and Madan, A. CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868-877 (1999).
27. Gordon, D., Abajian, C., and Green, P. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**, 195-202 (1998).
28. Li, R., Li, Y., Kristiansen, K., and Wang, J. SOAP: Short Oligonucleotide Alignment Program. *Bioinformatics* **24**, 713-714 (2008).
29. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
30. Gouy, M., Guindon, S., and Gascuel, O. SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221-224 (2010).
31. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792-1797 (2004).
32. Baker, J.L. et al. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**, 233-235 (2012).
33. Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).
34. Yadav, G., Gokhale, R.S., and Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335-363 (2003).
35. Rottig, M. et al. NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362-W367 (2011).