Supplemental Information

Aryl-aldehyde formation in fungal polyketides: Discovery and characterization of a distinct biosynthesis mechanism

Meng Wang, Mirko Beissner, Huimin Zhao*

Inventory of Supplemental Information

Supplemental Tables and Figures

Table S1, related to Plasmid Construction. Primers used in this study

Primer name	Sequence(5'-3')	Remarks
ATEG_03629-F	aaagaaagcatagcaatctaatctaagttttaattacaaaatgactccaaccctgctgct	Exon1 fragment
ATEG_03629-exon1-R:	gtgtcgccgatgccacatcggaccgcacagataaatatgcaccattgtatgtctccaaga	
ATEG_03629-exon2-F:	tgaacgaaaagtcaaagatatcttggagacatacaatggtgcatatttatctgtgcggtc	Exon2 fragment1*
ATEG_03629-exon2-M-R	cgctaagggagaagctctgc	
ATEG_03629-exon2-M-F	gcaagtcgactactggccgt	Exon2 fragment2*
ATEG_03629-exon2-R	cgagagagcatgatgtggcctccgccatggatcatgagcgcaatggggcgcttgacgcct	
ATEG_03629-exon3-F:	cctacgaagcccgacgcaccaggcgtcaagcgccccattgcgctcatgatccatggcgga	Exon3 fragment
ATEG_03629-R	aattagagcgtgatcatgaattaataaaagtgttcgcaaactaaaggtgcctcaacaaaa	
ATEG_03630-F	ttttttagttttaaaacaccagaacttagtttcgacggatatgtcgcccatcgccatcga	Exon1 fragment
ATEG_03630-exon1-R	caactgtcatggcgagacatggtgcgaccaacccgaatccgtgaaataacggtagagcag	
ATEG_03630-exon2-F	cgcgctcggaactgtactctctgctctaccgttatttcacggattcgggttggtcgcacc	Exon2 fragment
ATEG_03630-exon2-R	gcacagtagccgcattgtcaacagcgatccaggagagagtctgtttaatagctggcagtt	
ATEG_03630-exon3-F	gtttgcaagcatggttggtcaactgccagctattaaacagactctctcctggatcgctgt	Exon3 fragment
ATEG_03630-R	ttcaaaaaaataatatcttcattcaatcatgattctttttcattctagtggtgccctca	
GPM1p-F	attgggtaccgggccccccctcgaggtcgacggtatcgattagtcgtgcaatgtatgact	GPM1p promoter fragment
GPM1p-R	aaattggcgaagtgcttgcgcttgatgtgtcttgcaccattattgtaatatgtgtgtttg	
npgA-F	attcttcttaataatccaaacaaacacacatattacaataatggtgcaagacacatcaag	*npgA* fragment
npgA-R	ggagggaaaaagaaatcatcaaatcattcattcttcagacttaggataggcaattacaca	
GPM1t-F	ccagccctgtgcgactggggtgtgtaattgcctatcctaagtctgaagaatgaatgattt	GPM1t terminator fragment
GPM1t-R	tagaactagtggatcccccgggctgcaggaattcgatatctattcgaactgcccattcag	
long homology-F	cgaggtgccgtaaagcacta	ATEG_03630 fragment
PYKt-R	catttatgtacccatgtata	
PYKt-F	aatcatgattgaatgaagat	ATEG_03629 fragment 1*
03629-mid-R	ttgcaacttttcacttgtca	
03630-mid-F	gctagtctactcgttccgga	ATEG_03629 fragment 2*
long homology-R	ggaagcggaagagcgcccaa	
03629-S1651A-R	tttctctgatggcgcttagaagctctgctgccatgag*GGC*atcaatgcccaagtcctcca**	S1651A
03629-S1651A-F	cagcagagcttctaagcgcc	
03629-S1772A-R	gttggatatctgcagccaattccatgccaaccaaCGCgtccagaccgatatcacggagtg	S1772A
03629-S1772A-F	ttggcatggaattggctgca	
03630-S611A-R	aagaaagtagaagtcgatgcaggcggattgcctgcagAGCattcataccgagttcaaaca	S611A
03630-S611A-F	caatccgcctgcatcgactt	

ADH2p-ateg-03630-his-F	atcaactattaactatatcgtaataccatcatgggcagcagccatcatca	
ADH2p-ateg-03630-his-R	ataatgaaaactataaatcgtgaaggcatgtcattctagtggtgccctca	
ateg-03630-Y863F-F	cattgaagactttggaTTCgggaaggccaagtatgtatgc	Y863F
ateg-03630-Y863F-R	gcatacatacttggccttcccGAAtccaaagtcttcaatg	
ateg-03630-K867A-F	tttggatacgggaaggccGCGtatgtatgcgaagagatta	K867A
ateg-03630-K867A-R	taatctcttcgcatacataCGCggccttcccgtatccaaa	
ateg-03630-T690A-F	atgccacagttctcttgGCTggtagcacaggcaatctcgg	T690A
ateg-03630-T690A-R	ccgagattgcctgtgctaccAGCcaagagaactgtggcat	
qPCR-A.terreus-actin-F	gatctggcatcacacgttct	
qPCR-A.terreus-actin-R	catcttctcacggttggactt	
03628-qPCR-F	gcaagcaacccagtacagcgaaat	
03628-qPCR-R	ttacctgtccgtggcatctccatt	
03629-qPCR-F	acgggagctcgtgaatgacatctt	
03629-qPCR-R	caaaccgcaatccaaacagggtca	
03630-qPCR-F	aaagcctgttcccaggtccgataa	
03630-qPCR-R	actggagattcaatgcggtcgtct	
03631-qPCR-F	tcaacttcgctacgctcaatccga	
03631-qPCR-R	tgcgctttgaagctcaatacccac	
03632-qPCR-F	aatgtcaaagggccgatgttcctc	
03632-qPCR-R	agcactcgggtcatttgctcgata	
03635-qPCR-F	aatgggctcgggctaactcatctt	
03635-qPCR-R	aaactaattccctggcgggctagt	

* The long PKS gene was broken into two fragments for ready amplification by PCR.
** Upper-case letters are mutated codons.

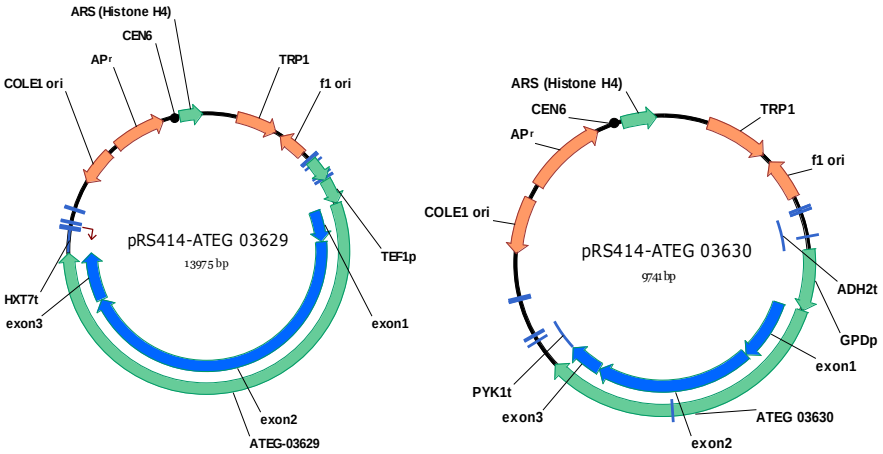

Figure S1A, related to Plasmid Construction. Maps of the plasmids carrying the NR-PKS or NRPS-like gene.

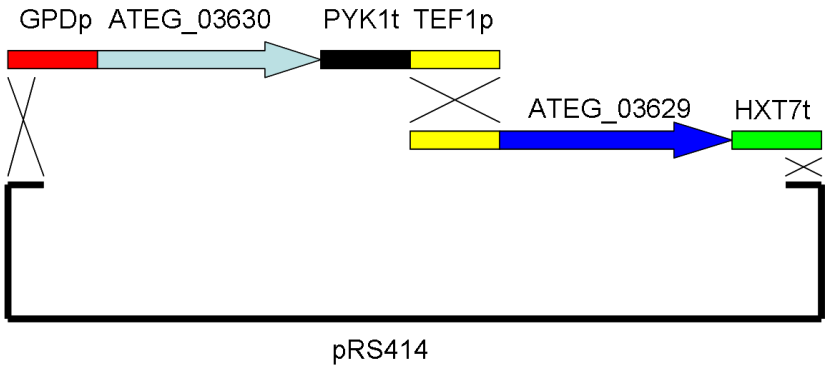

Figure S1B, related to Plasmid Construction. Strategy for assembling the two-gene ATEG_03629_03630 cassette.

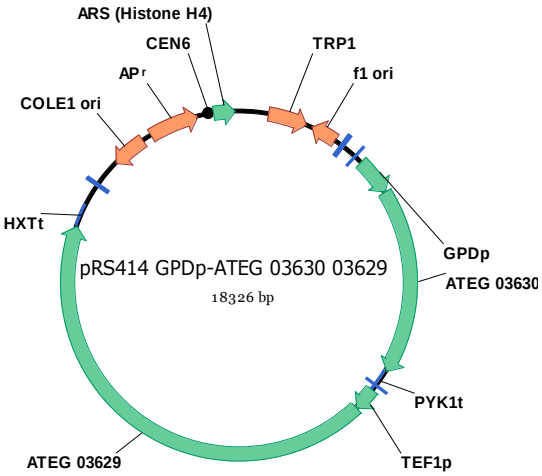

Figure S1C, related to Plasmid Construction. Map of the plasmid carrying the two-gene cassette consisting of the NR-PKS and NRPS-like genes.

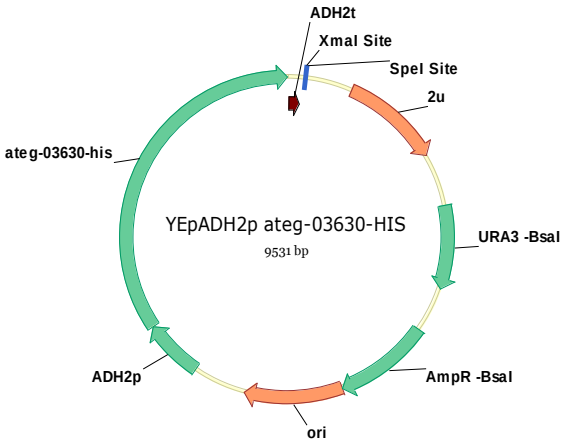

Figure S1D, related to Plasmid Construction. Map of the plasmid for over-expressing NRPS-like protein ATEG_03630.

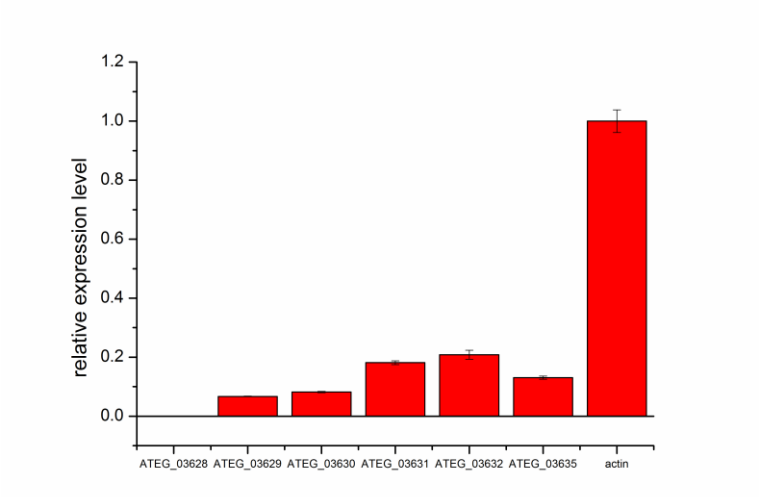

Figure S1E, related to Quantitative PCR Analysis. Quantitative PCR analysis of the expression level of the target gene cluster in the native host *A. terreus*. Data are represented as mean ± SD.

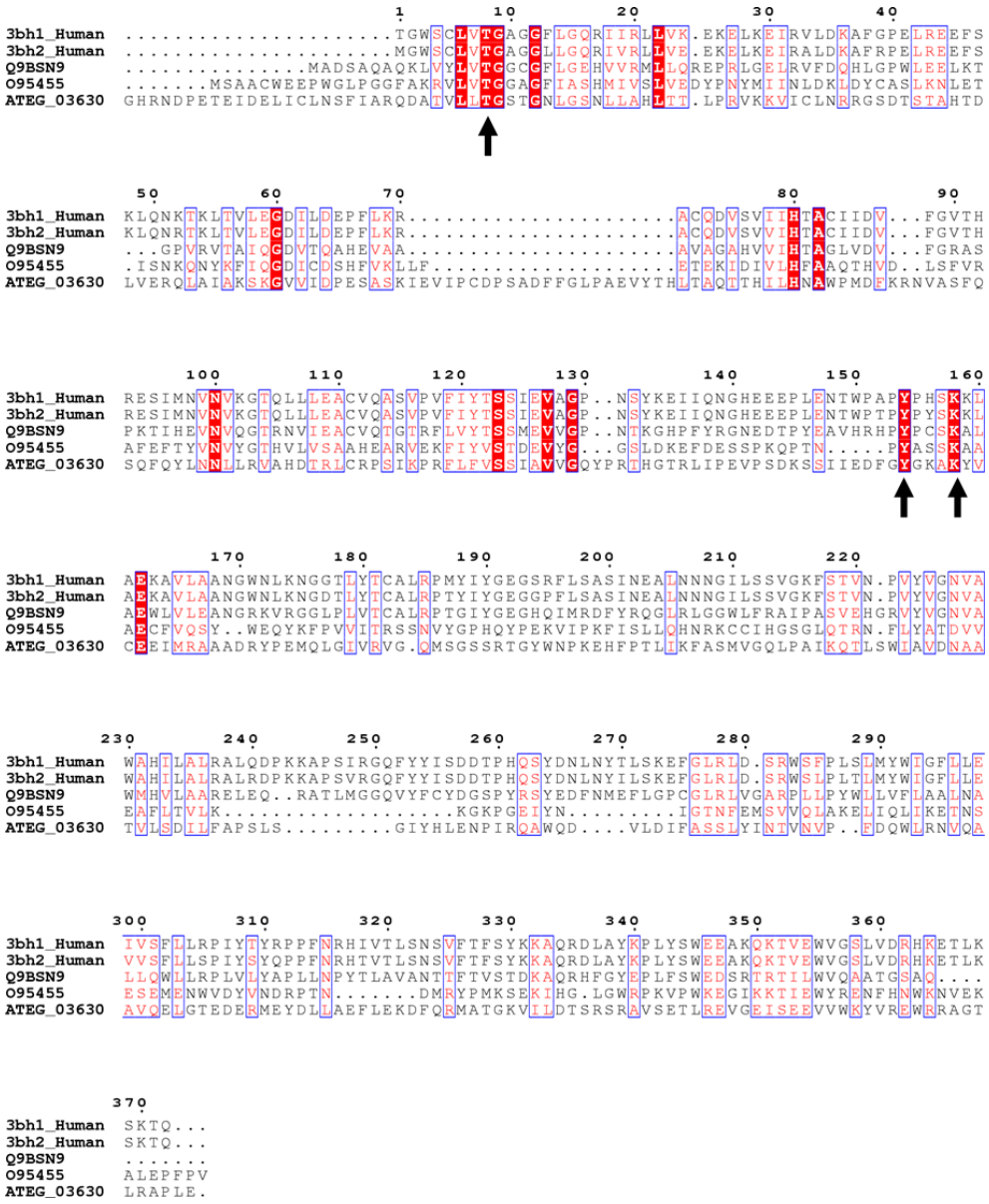

Figure S1F, related to Figure 2. Multiple sequence alignment of R domain of ATEG_03630 with different short-chain dehydrogenases/reductases family proteins. Black arrow:mutated residues in this study. Abbreviation are as follows: 3bh1_Human: 3 beta-hydroxysteroid dehydrogenase type 1, 3bh2_Human: 3 beta-hydroxysteroid dehydrogenase type 2, O95455: dTDP-D-glucose 4,6-dehydratase, Q9BSN9: 3 beta-hydroxy-delta 5-c27-steroid oxidoreductase.

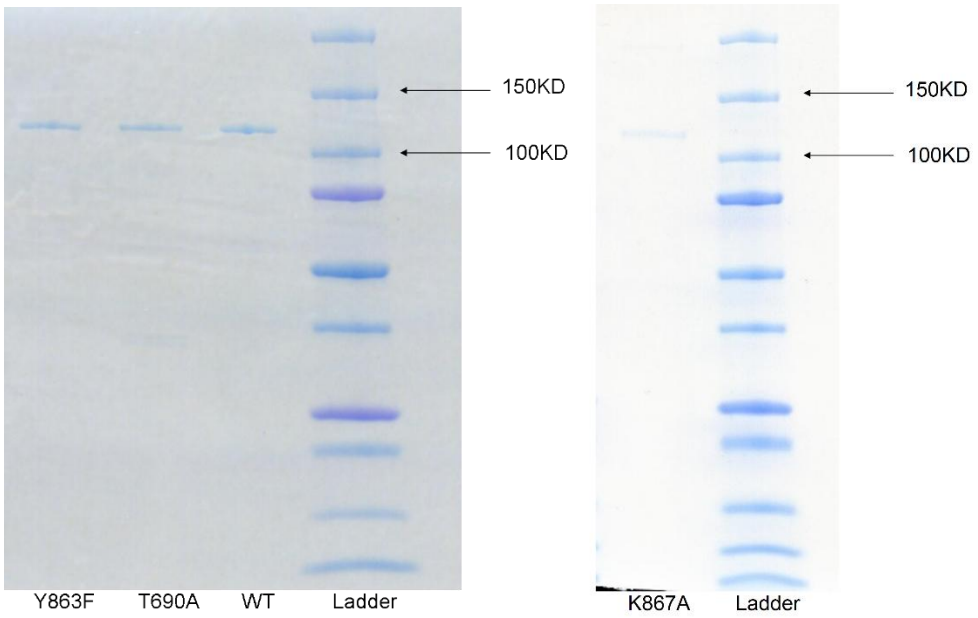

Figure S1G, related to Figure 2. SDS-PAGE analysis of purified ATEG_03630 WT and mutant proteins.

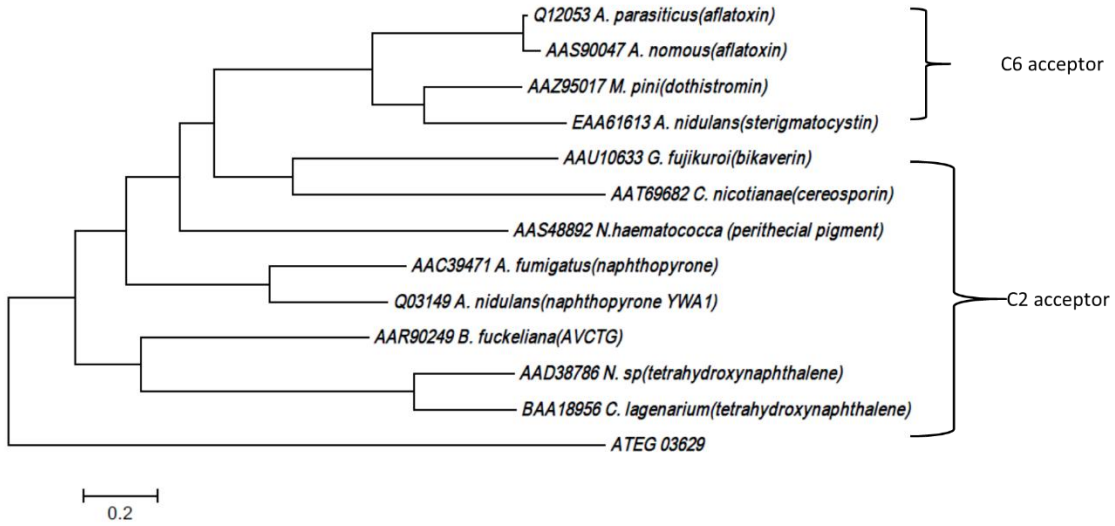

Figure S1H, related to Figure 2. Phylogenetic analysis of the SAT domains of homologous NR-PKSs. Starter unit of ATEG_03629 is more likely to be acetyl-CoA than other long chain CoA.

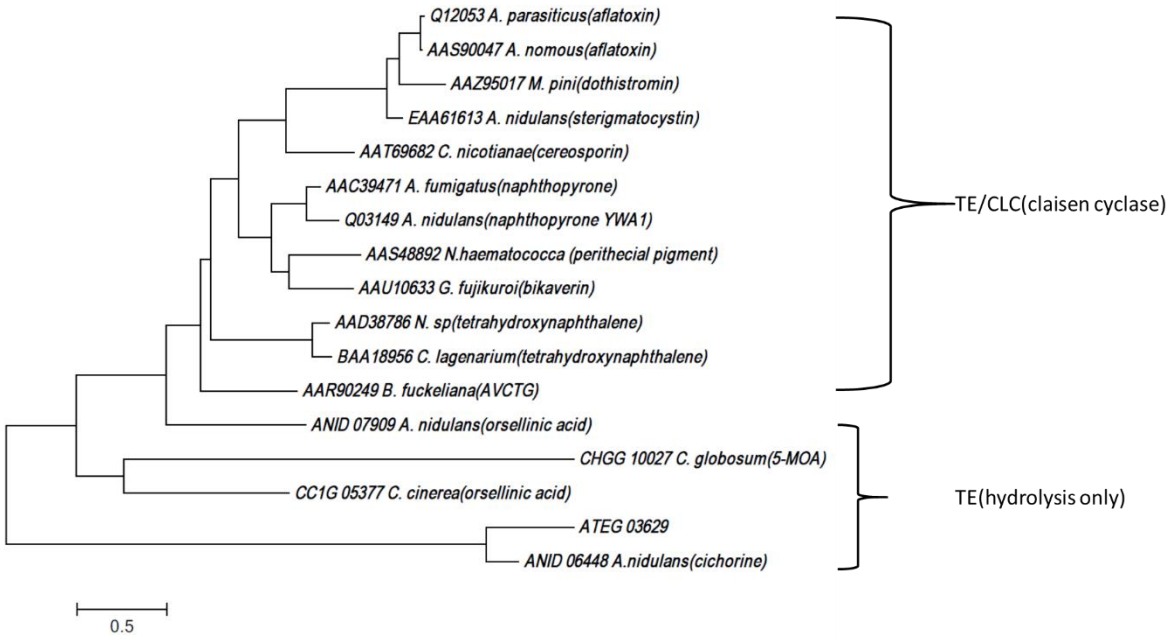

Figure S1I, related to Figure 2. Phylogenetic analysis of the TE domains of homologous NR-PKSs. The TE domain of ATEG_03629 is more likely to perform only hydrolysis reaction than claisen condensation.

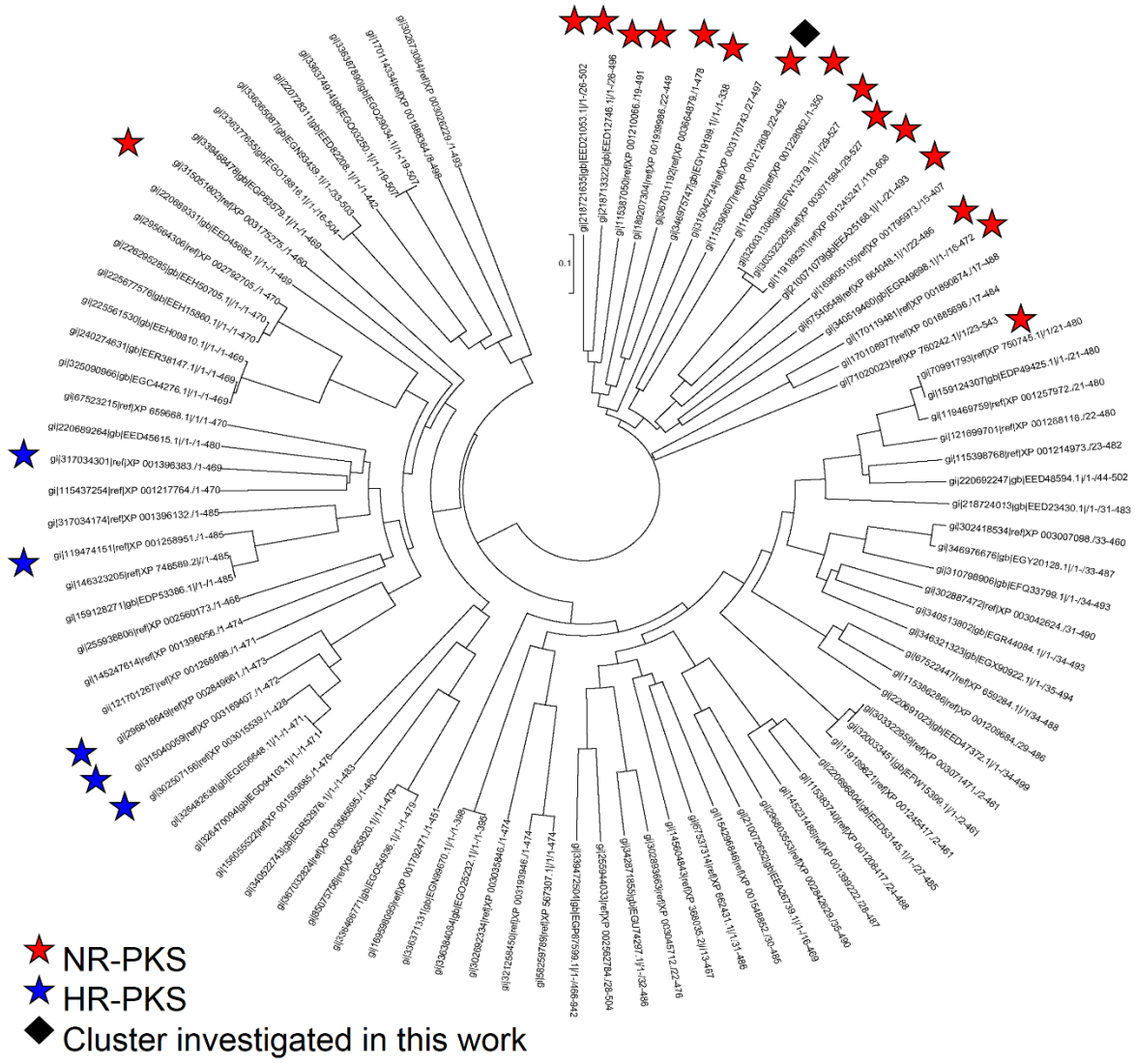

Figure S1J, related to Figure 2. Phylogenetic analysis of the A domain of homologous NRPS-like genes. Red star: NRPS-like gene with a NR-PKS in its vicinity. Blue star: NRPS-like gene with a HR-PKS in its vicinity.

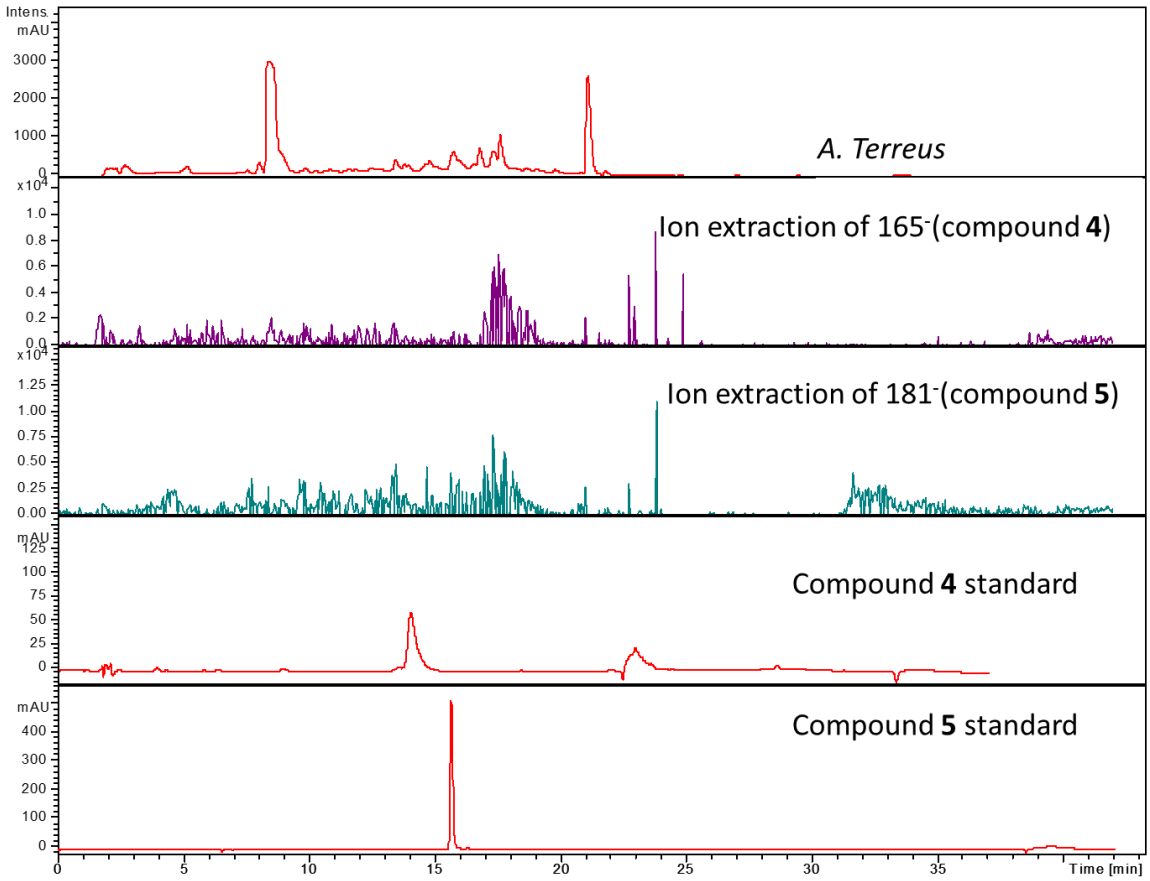

Figure S1K, related to Figure 2. HPLC-MS analysis of culture extraction of native host *A. terreus*.

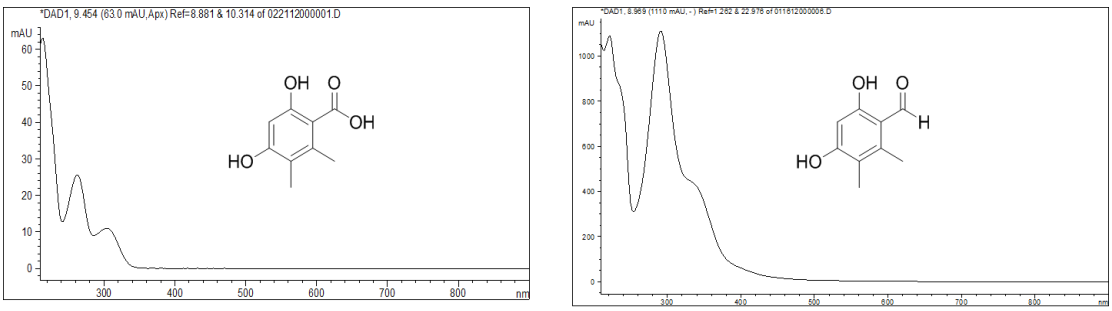

Figure S1L, related to Figure 2. UV-Vis spectra of compounds **4** and **5**.

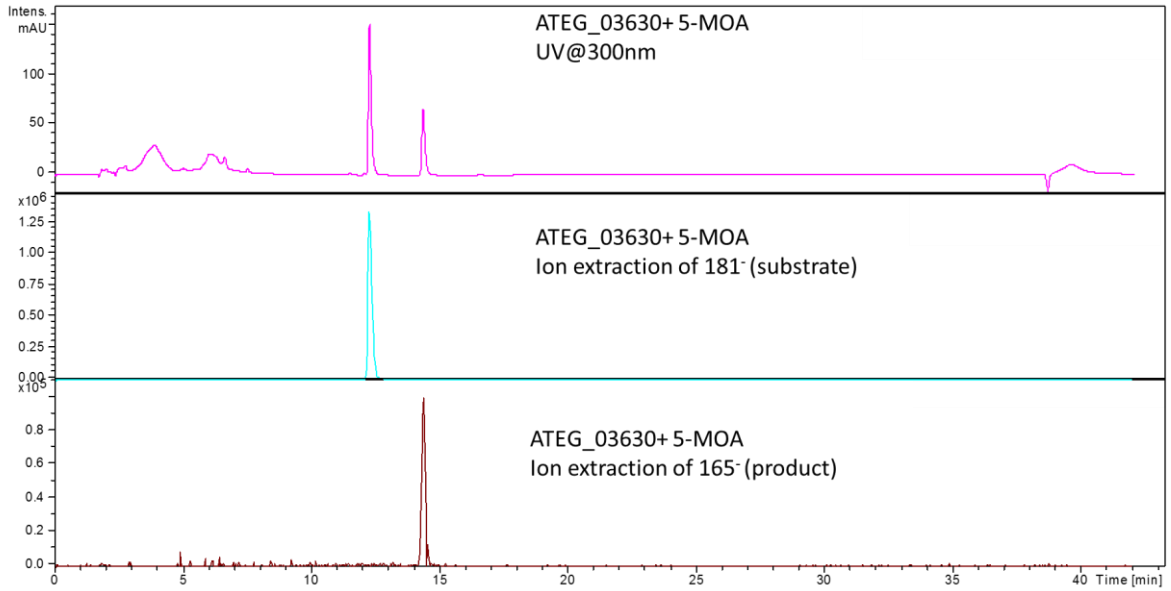

Figure S1M, related to *In vitro* Characterization of ATEG_03630 and its Mutants. HPLC-MS analysis of *in vitro* assay of NRPS-like protein ATEG_03630 catalyzed reduction of 5-MOA after 15min incubation. Compound **5** can be detected.

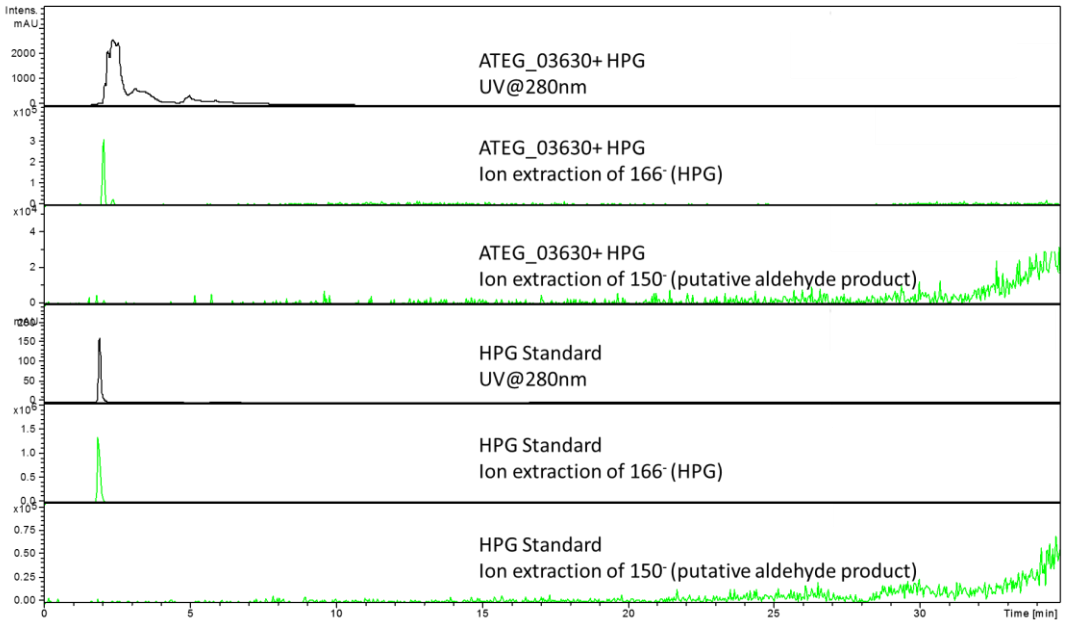

Figure S1N, related to *In vitro* Characterization of ATEG_03630 and its Mutants. HPLC-MS analysis of *in vitro* assay of NRPS-like protein ATEG_03630 catalyzed reduction of HPG. No aldehyde product can be detected.

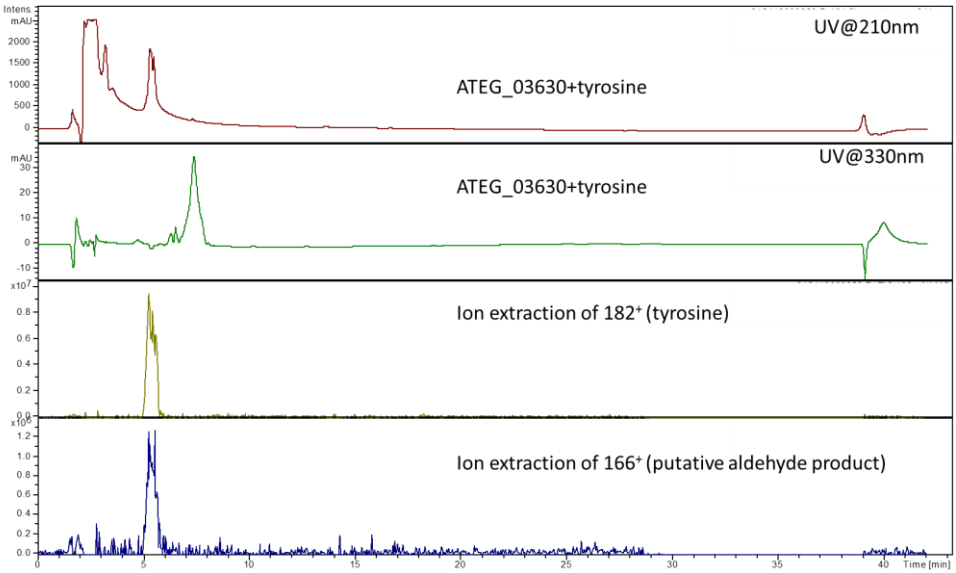

Figure S1O, related to *In vitro* Characterization of ATEG_03630 and its Mutants. HPLC-MS analysis of *in vitro* assay of NRPS-like protein ATEG_03630 catalyzed reduction of tyrosine. No aldehyde product can be detected.

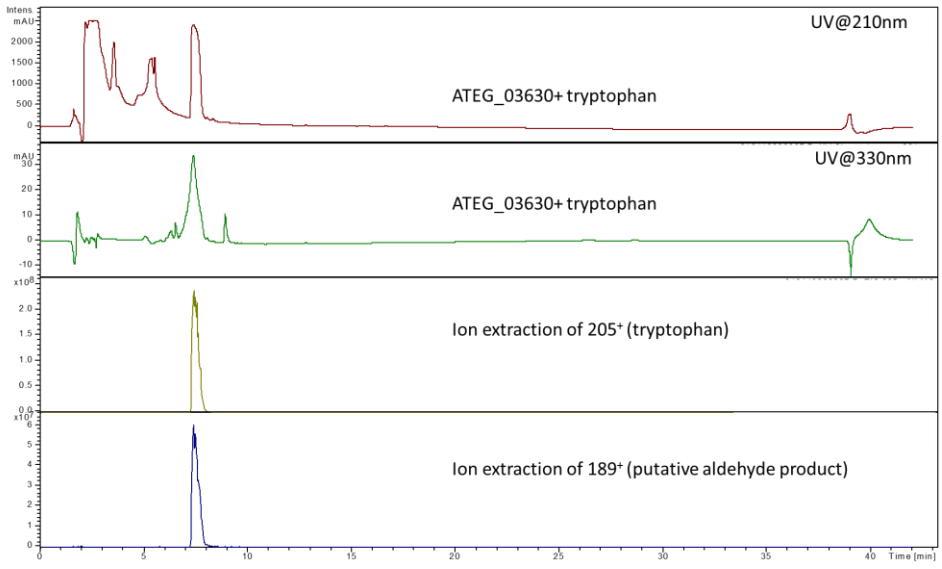

Figure S1P, related to *In vitro* Characterization of ATEG_03630 and its Mutants. HPLC-MS analysis of *in vitro* assay of NRPS-like protein ATEG_03630 catalyzed reduction of tryptophan. No aldehyde product can be detected.

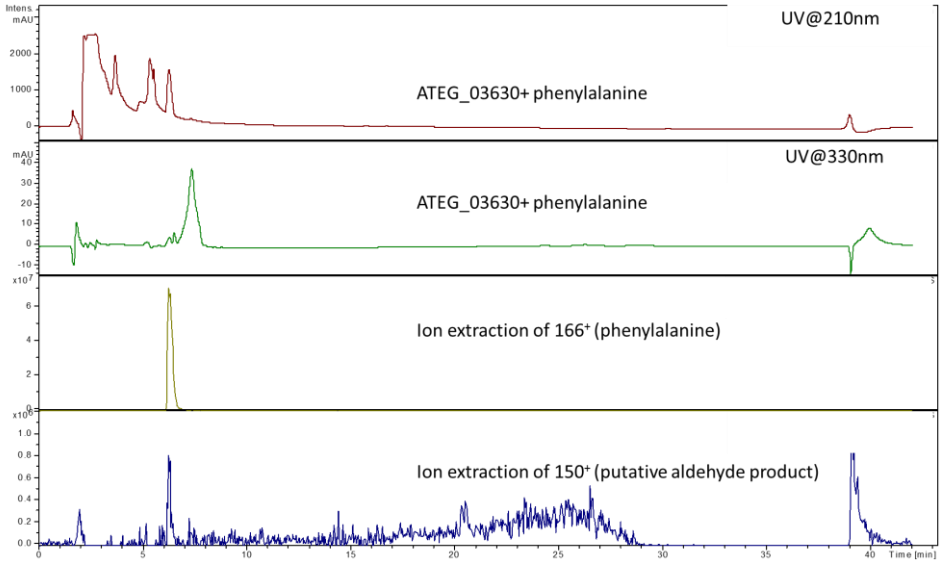

Figure S1Q, related to *In vitro* Characterization of ATEG_03630 and its Mutants. HPLC-MS analysis of *in vitro* assay of NRPS-like protein ATEG_03630 catalyzed reduction of phenylalanine. No aldehyde product can be detected.

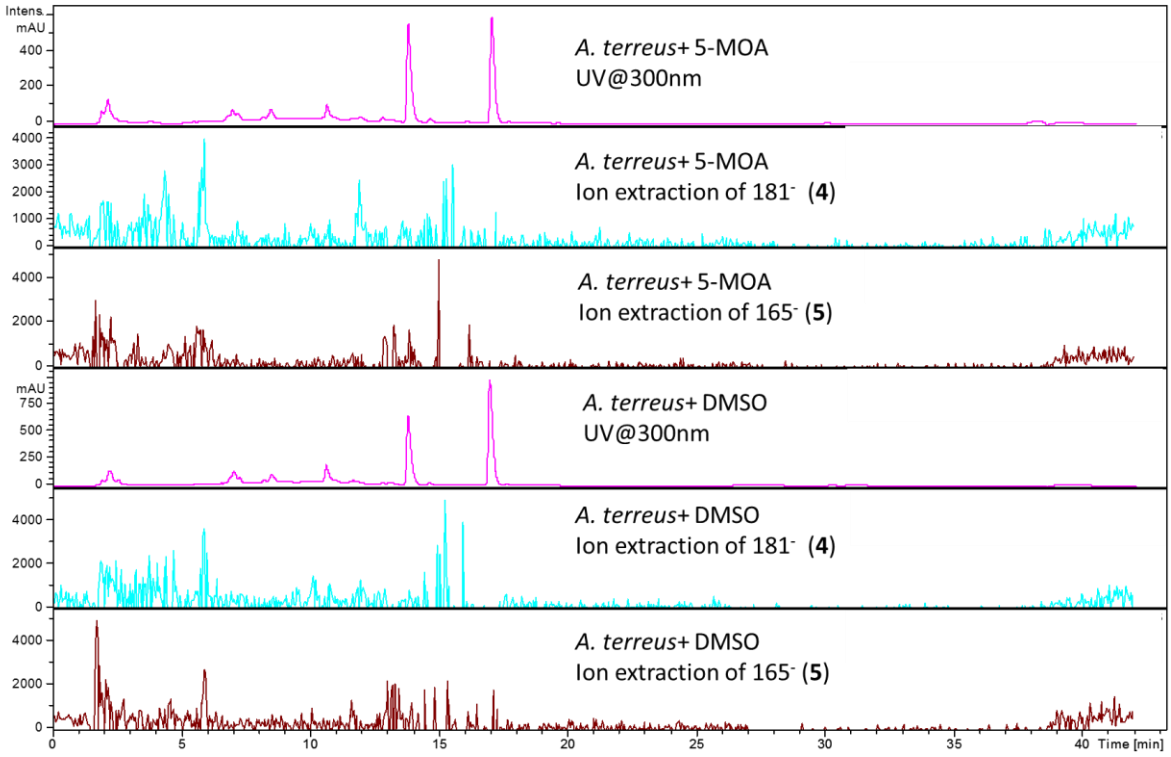

Figure S1R, related to Figure 2. HPLC-MS analysis of culture extraction of native host *A. terreus* fed with 5-MOA or DMSO (control).

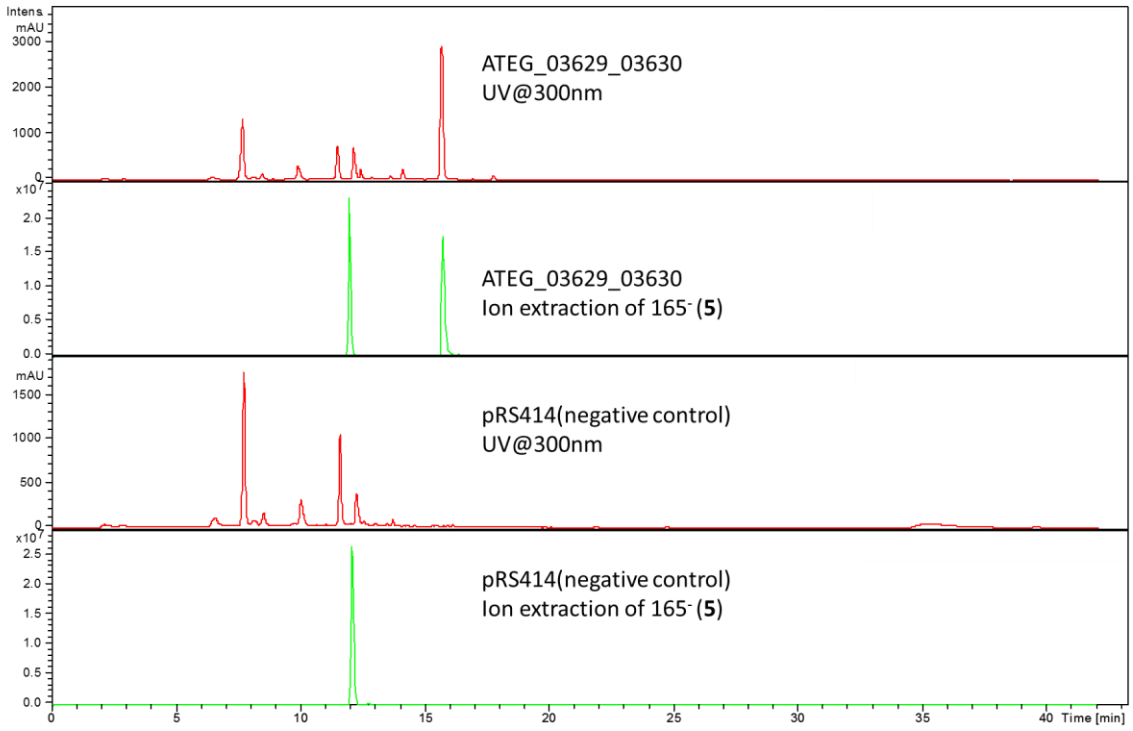

Figure S1S, related to Figure 2. HPLC-MS analysis of extraction product from large scale culture (800 ml).

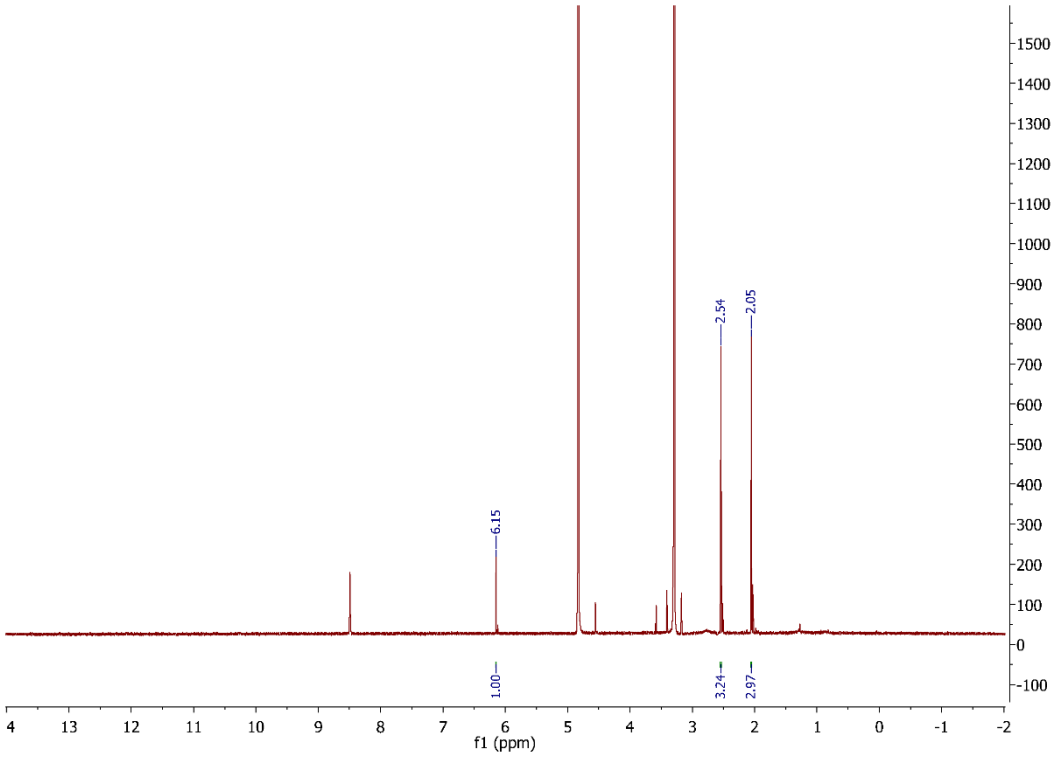

Figure S2A, related to Figure 2. Proton NMR (600 MHz) spectrum of compound **4** in CD_3OD.

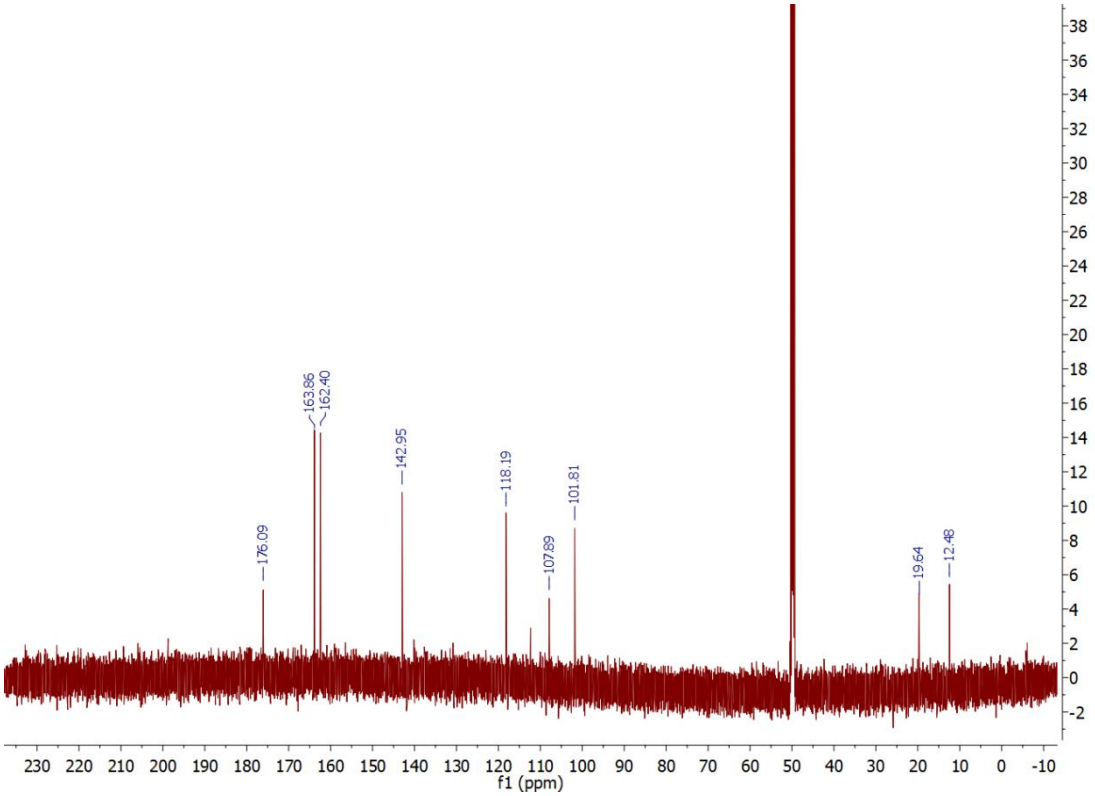

Figure S2B, related to Figure 2. Carbon NMR (600 MHz) spectrum of compound **4** in CD_3OD.

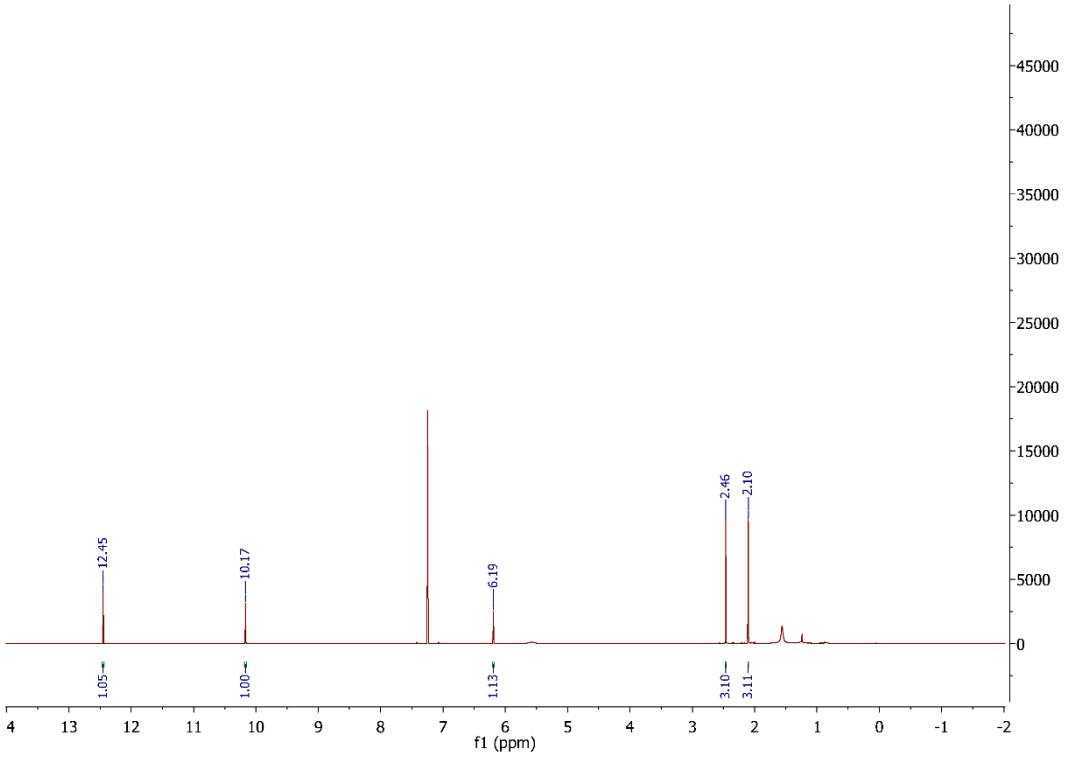

Figure S2C, related to Figure 2. Proton NMR (600 MHz) spectrum of compound **5** in $CDCl_3$.

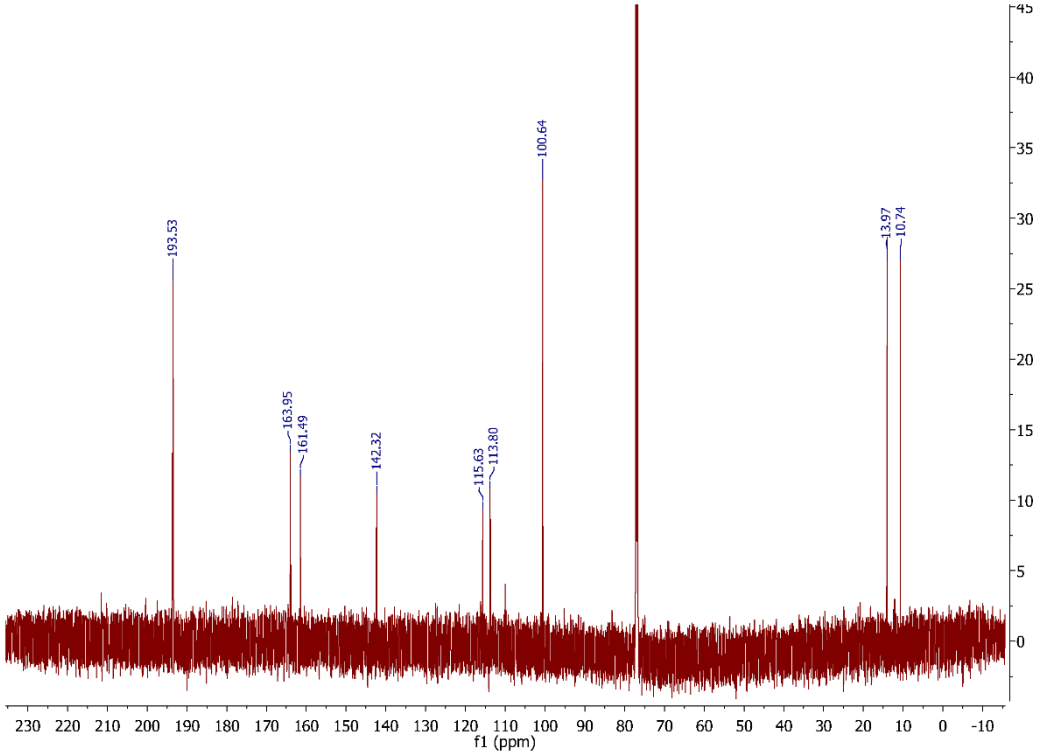

Figure S2D, related to Figure 2. Carbon NMR (600 MHz) spectrum of compound **5** in $CDCl_3$.

Supplemental Experimental Procedures

I. Molecular Biology

1. Plasmid Construction

Cloning of *ATEG_03629*. PCR was used to amplify each of the three exons that in combination encode the ATEG_03629 gene from the genomic DNA of *A. terreus* using the corresponding primers listed in Table S1. Helper plasmid pRS414-TEF1p-HXT7t was linearized by *Xho*I digestion. The full ATEG_03629 gene under the control of the TEF1p promoter was assembled via the DNA assembler method[1] to obtain the plasmid pRS414-ATEG_03629.

Cloning of *ATEG_03630*. PCR was used to amplify each of the three exons that in combination encode the ATEG_03630 gene from the genomic DNA of *A. terreus* using the corresponding primers listed in Table S1. Helper plasmid pRS414-GPDp-PYK1t was linearized by *Xho*I digestion. The full ATEG_03630 gene under the control of the GPDp promoter was assembled via the DNA assembler method to obtain plasmid pRS414-ATEG_03630.

Cloning of *npgA*. PCR was used to amplify to the *npgA* gene from the genomic DNA of *A. nidullans* using the corresponding primers listed in Table S1. DNA fragments encoding the promoter GPM1p and the terminator GPM1t were amplified from plasmid pRS414-GPM1p-GPM1t. Plasmid pRS416 was linearized by *Hind*III digestion. The full *npgA* gene under the control the GPM1p promoter was assembled via the DNA assembler method to obtain plasmid pRS416-GPM1p-npgA-GPM1t.

Construction of plasmid pRS414-ATEG_03629_03630. Backbone fragment was obtained by restriction digestion of plasmid pRS414 with *BamH*I and *EcoR*I. DNA fragment containing the *ATEG_03629* gene was PCR-amplified from the plasmid pRS414-ATEG_03629 using the corresponding primers listed in Table S1. DNA fragment containing the *ATEG_03630* gene was PCR-amplified from the plasmid pRS414-ATEG_03630 using the primers listed in Table S1. The plasmid pRS414-ATEG_03629_03630 was constructed via DNA assembler as shown in Figure S1A.

Construction of plasmid pRS414-ATEG_03629m-S1651A-S1772A. DNA fragments encoding double mutant ATEG_03629m-S1651A-S1772A were PCR-amplified from plasmid pRS414-ATEG_03629 using the corresponding primers listed in Table S1. Helper plasmid pRS414-TEF1p-HXT7t was linearized by *Xho*I digestion. The plasmid pRS414-ATEG_03629m-S1651A-S1772A was constructed via DNA assembler.

Construction of plasmid pRS414-ATEG_03630m-S611A. DNA fragments encoding the mutant *ATEG_03630m-S611A* gene were PCR-amplified from plasmid pRS414-ATEG_03630 using the corresponding primers listed in Table S1. Helper plasmid pRS414-GPDp-PYK1t was linearized by *Xho*I digestion. The plasmid pRS414-ATEG_03630m-S611A was constructed via DNA assembler.

Construction of plasmid pRS414-ATEG_03629m-S1651A. DNA fragments encoding the mutant ATEG_03629m-S1651A were PCR-amplified from plasmid pRS414-ATEG_03629 using the corresponding primers listed in Table S1. Helper plasmid pRS414-TEF1p-HXT7t was linearized by *Xho*I digestion. The plasmid pRS414- ATEG_03629m-S1651A was constructed via DNA assembler.

Construction of plasmid pRS414-ATEG_03629m-S1772A. DNA fragments encoding the mutant ATEG_03629m-S1772A were PCR-amplified from plasmid pRS414-ATEG_03629 using the corresponding primers listed in Table S1. Helper plasmid pRS414-TEF1p-HXT7t was linearized by *Xho*I digestion. The plasmid pRS414- ATEG_03629m-S1772A was constructed via DNA assembler.

Construction of plasmid pRS414-ATEG_03629m-S1651A-S1772A_03630. The vector backbone fragment was obtained by restriction digestion of plasmid pRS414 with *BamH*I and *EcoR*I. DNA fragment containing the ATEG_03629m-S1651A-S1772A gene was PCR-amplified from the plasmid pRS414-ATEG_03629m-S1651A-S1772A using the corresponding primers listed in Table S1. DNA fragment containing the mutant ATEG_03630 gene was PCR-amplified from the plasmid pRS414 -ATEG_03630 using the primers listed in Table S1. The plasmid pRS414-ATEG_03629m-S1651A-S1772A_03630 was constructed via DNA assembler.

Construction of plasmid pRS414-ATEG_03629_03630m-S611A. Backbone fragment was obtained by restriction digestion of plasmid pRS414 with *BamH*I and *EcoR*I. DNA fragment containing the ATEG_03629 gene was PCR-amplified from the plasmid pRS414-ATEG_03629 using the corresponding primers listed in

Table S1. DNA fragment containing the mutant ATEG_03630m-S611A gene was PCR-amplified from the plasmid pRS414-ATEG_03630m-S611A using the primers listed in Table S1. The plasmid pRS414-ATEG_03629_03630m-S611A was constructed via DNA assembler.

Construction of plasmid YEp-ADH2p-ATEG_03630-HIS. Backbone fragment was obtained by restriction digestion of plasmid YEp-ADH2p with *BsaI*. DNA fragment containing the ATEG_03630 gene with 6×His-tag was PCR-amplified from the plasmid pRS414-ATEG_03630 using the corresponding primers listed in Table S1. The plasmid YEp-ADH2p-ATEG_03630-HIS was constructed via DNA assembler.

Construction of plasmid YEp-ADH2p-ATEG_03630-HIS. Backbone fragment was obtained by restriction digestion of plasmid YEp-ADH2p with *BsaI*. DNA fragment containing the ATEG_03630 mutant with 6×His-tag was PCR-amplified from the plasmid pRS414-ATEG_03630 using the corresponding primers listed in Table S1. The plasmid YEp-ADH2p-ATEG_03630-HIS was constructed via DNA assembler.

Construction of plasmid YEp-ADH2p-ATEG_03630-HIS-T690A. Backbone fragment was obtained by restriction digestion of plasmid YEp-ADH2p with *BsaI*. DNA fragment containing the ATEG_03630 mutant with 6×His-tag was PCR-amplified from the plasmid YEp-ADH2p-ATEG_03630-HIS using the corresponding primers listed in Table S1. The plasmid YEp-ADH2p-ATEG_03630-HIS-T690A was constructed via DNA assembler.

Construction of plasmid YEp-ADH2p-ATEG_03630-HIS-Y863F. Backbone fragment was obtained by restriction digestion of plasmid YEp-ADH2p with *BsaI*. DNA fragment containing the ATEG_03630 mutant with 6×His-tag was PCR-amplified from the plasmid YEp-ADH2p-ATEG_03630-HIS using the corresponding primers listed in Table S1. The plasmid YEp-ADH2p-ATEG_03630-HIS-Y863F was constructed via DNA assembler.

Construction of plasmid YEp-ADH2p-ATEG_03630-HIS-K867A. Backbone fragment was obtained by restriction digestion of plasmid YEp-ADH2p with *BsaI*. DNA fragment containing the ATEG_03630 mutant with 6×His-tag was PCR-amplified from the plasmid YEp-ADH2p-ATEG_03630-HIS using the corresponding primers listed in Table S1. The plasmid YEp-ADH2p-ATEG_03630-HIS-K867A was constructed via DNA assembler.

2. Restriction Digestion Analysis and Mutant Confirmation

Colonies were randomly picked and grown in drop-out liquid media for two days, after which the plasmids from yeast were isolated using Zymoprep II Yeast Plasmid Miniprep kit (Zymo Research, CA). Yeast plasmids were transformed to *E. coli* strain BW25141 and selected on Luria Broth (LB) agar plates supplemented with 50 μg/mL ampicillin. Colonies were inoculated into 5 mL of LB media supplemented with 50 μg/mL ampicillin, and plasmids were isolated from the liquid culture using the plasmid miniprep kit from Qiagen (Valencia, CA). Plasmids isolated from *E. coli* were then subjected to restriction digestion. Usually, one or two enzymes cutting the target molecule at multiple sites were chosen. The reaction mixtures were loaded to 0.7% agarose gels to check for the correct restriction digestion pattern by DNA electrophoresis. The site-directed mutations in the ATEG_03629m-S1651A-S1772A double mutant, ATEG_03630m-S611A, ATEG_03629m-S1651A, ATEG_03629m-S1772A, YEp-ADH2p-ATEG_03630-HIS-T690A, YEp-ADH2p-ATEG_03630-HIS -Y863F and YEp-ADH2p-ATEG_03630-HIS-K867A were confirmed by DNA sequencing.

II. Compound Purification and Characterization

1. Purification and characterization of compound 4

The yeast transformant producing compound **4** was grown in 1.6 L SC-TRP-URA dropout media for 5 days. Cell pellets were collected by centrifugation. The supernatants were extracted with 2 volume of ethyl acetate and cell pellets were extracted with acetone. The organic fractions of both extractions were combined and evaporated to dryness. It was re-dissolved in 1 mL of methanol. Compound **4** was then purified by semi-preparative HPLC purification using a Gemini C6-phenyl column (Phenomenex, Torrance, CA). HPLC parameters for separation were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 20 min, maintain at 100% B for 15 min, return and maintain at 0% B for 15 min; flow rate 3 mL/min. The collected fractions were extracted with 2 volume of ethyl-acetate and the organic phase was dried under vacuum and yielded the pure compound as white power. ^{1}H NMR (600 MHz, CD_3OD) δ 2.05 (s, 3H), 2.54 (s, 3H), 6.15 (s, 1H); ^{13}C NMR (600 MHz, CD_3OD) δ 12.48, 19.63, 101.81, 107.89, 118.18, 142.95, 162.40, 163.86, 176.08. HRMS (ESI) Calcd. for $[C_9H_9O_4]^-$ 181.0501, found 181.0463.

2. Purification and characterization of compound 5

Yeast transformant producing compound **5** was grown in 1.6 L SC-TRP-URA dropout media for 5 days. Cell pellets were collected by centrifugation. The supernatants were incubated with 10g of XAD-16 resin overnight. XAD-16 resins were collected via centrifugation and eluted with 200 mL of methanol. Cell pellets were extracted with acetone. The organic fractions of both extractions were combined and re-dissolved in 1 mL of methanol. Compound **5** was then purified by semi-preparative HPLC purification using a Luna C18 column (Phenomenex, Torrance, CA). HPLC parameters for separation were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 20 min, maintain at 100% B for 15 min, return and maintain at 0% B for 15 min; flow rate 3 mL/min. The collected fractions were extracted with 2 volume of ethyl-acetate and the organic phase was dried under vacuum and yielded the pure compound as white power. ^{1}H NMR (600 MHz, $CDCl_3$) δ 2.10 (s, 3H), 2.46 (s, 3H), 6.19 (s, 1H), 10.17 (s, 1H), 12.45 (s, 1H); ^{13}C NMR (600 MHz, $CDCl_3$) δ 10.74, 13.97, 100.64, 113.8, 115.63, 142.32, 161.49, 163.95, 193.53. HRMS (ESI) Calcd. for $[C_9H_{11}O_3]^+$ 167.0708, found 167.0711.

3. Quantification of compound 4 production from PKS gene ATEG_03629 and its mutants

Plasmid carrying ATEG_03629 or its single mutant ATEG_03629m-S1651A, ATEG_03629m-S1772A were co-transformed with plasmid pRS416-GPM1p-npgA-GPM1t to HZ848 strain. Each strain were grown in 3ml of SC-TRP-URA dropout media for 5 days. Cell pellets were collected by centrifugation. The supernatants were extracted by 2 volume of ethyl acetate and the cell pellets were extracted by acetone. All extracts were combined, dried and re-dissolved in 200ul of methanol. HPLC-ESI-MS was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA) with a reverse-phase kinetex C18 column (Phenomenex, Torrance, CA). HPLC parameters for detection of compounds **4** were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 60 min, maintain at 100% B for 10 min, return and maintain at 10% B for 7 min; flow rate 0.3 ml/min; detection by **UV** spectroscopy at 300 nm or 330nm. Isolated compound **4** were used as standard in the same HPLC condition for quantification.

4. Feeding study of the native host *A. terreus*

A. terreus was inoculated to 4 ml of YPAD medium and grow at 30°C and 250 rpm. 0.2 mg of compound **4** or DMSO (control) was added to the culture broth. After four days, the supernatants were extracted by 2 volume of ethyl acetate and the mycelia were extracted by methanol. All extracts were combined, dried and re-dissolved in 100 μl of methanol for HPLC-MS analysis.