Supporting Information Appendix

Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes

Wei Ding, Wan-Qiu Liu, Youli Jia, Yongzhen Li, Wilfred A van der Donk*, and Qi Zhang*

Materials and General Methods

Culture media were from Sinopharm Chemical Reagent Co. Ltd or from Difco laboratories. Chemicals were purchased from Fisher Scientific or from Aldrich unless noted otherwise. Enzymes were from Takara Biotechnology or from Vazyme Biotech Co. Ltd. *Phomopsis leptostromiformis* ATCC 26115 and *Aspergillus oryzae* ATCC 22788 were purchased from ATCC, *Neurospora crassa* CGMCC 3.1604, *Beauveria bassiana* CGMCC 3.3577, and *Metarhizium anisopliae* CGMCC 3.4607 were purchased from China General Microbiological Culture Collection Center (CGMCC). DNA sequencing was performed by the Biotechnology Center at the University of Illinois at Urbana-Champaign. High resolution mass spectra (HRMS) were acquired either using a Synapt ESI quadrupole ToF Mass Spectrometry System (Waters) equipped with an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters), or using a Q-Exactive™ Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher) equipped with a Dionex Ultimate 3000 HPLC system (Thermo Fisher).

Genome sequencing and analysis

Genomic DNA was prepared from stationary liquid cultures using the UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA). Three libraries were used for genome sequencing. The 800 nt shotgun library was constructed with the TruSeq DNA Sample Prep kit (Illumina, CA), whereas the 3 kb and 8 kb mate-pair libraries were constructed with the Nextera Mate Pair library Sample Prep kit

(Illumina, CA). The libraries were pooled, quantitated by qPCR and sequenced on two lanes for 100 cycles from each end of the fragments on a HiSeq2000 using a TruSeq SBS sequencing kit version 3 and demultiplexed with Casava1.8.2. The genome assemblies were performed using the assembly programs SOAPdeNOVO (1) and Velvet version 1.0.15 (2) at the Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign. The assembled sequences were converted to BLAST database format for local BLAST search using standalone BLAST software (v. 2.2.18). Gene prediction was performed using Geneid (3) and Augustus (4) using *Neurospora crassa* as a reference genome (5), and was manually checked by comparing with homologous gene/proteins in the GenBank database. The data was deposited into the Genbank database under accession number KU645826-KU645844.

Construction of the *phomQ*-knockout mutant

The *hgh*-containing plasmid pQZ101 was constructed according to a previous report (*hgh* is the hygromycin B phosphotransferase gene) (6). Briefly, a 4123 bp fragment containing the whole *trpC* gene (7) (including the promoter to terminator) was amplified from genomic DNA of *Aspergillus nidulans* by PCR using a primer pair *trpC-F* and *trpC-R*, and the resulting fragment was cloned into pMD19-T to generate a subclone construct pTrpC-MD19T. After sequencing to firm the sequence fidelity, the *trpC* cassette was amplified by PCR using a primer pair *re-trpC-F* and *re-trpC-R* and pTrpC-MD19T as the template. The resulting fragment was cloned into the EcoRI/KpnI site of pUC18 by using In-fusion HD Cloning Kit (Clontech) to give a plasmid pTrC-UC18. pQZ101 was constructed from pTrC-UC18 by using the λ-RED mediated PCR-targeting (8). A primer pair *hgh-F* and *hgh-R* was used to amplify the *hgh* gene from pcDNAT3.1/Hygro(+) (Invitrogen). The resulting DNA fragment was then transferred into *E. coli* BW25113/pIJ790/pTrC-UC18 competent cell via electroporation (2.5 kV/cm, 25μF, 200 Ω, 2 mm) and the cells were plated on LB-agar containing hygromycin B (100 μg/mL) to screen for positive clones, which were

confirmed by DNA sequencing.

To inactivate *phomQ*, a primer pair *phomQ-F* and *phomQ-R* was used to amplify the 710 bp internal fragment of *phomQ* from the genomic DNA of *P. leptostromiformis*. The resulting PCR product was treated with the restriction enzymes BamH and HindIII, and was then inserted into the same restriction site of pQZ101 to yield the gene inactivation construct pPhomQ-dis1-QZ101. Preparation of *P. leptostromiformis* protoplast and transformation were performed similarly to a procedure described previously (9). Transformants were grown on Potato Dextrose Agar (PDA) supplemented with 100 μg/mL hygromycin B, and integration of the disruption cassette into the genome was confirmed by PCR using a primer pair *id-php-F* and *id-phomQ-R*.

RNA preparation and cDNA synthesis

For RNA preparation, all tubes and pipette tips used in the experiments were treated with 0.1% Diethylpyrocarbonate (DEPC) before autoclaving. *P. leptostromiformis* ATCC 26115 was grown in PD liquid medium at room temperature without shaking for 15 days, and the mycelium was collected in sterilized Miracloth, washed two times with saline, pressed dry and massed. *Neurospora crassa* CGMCC 3.1604 and *Beauveria bassiana* CGMCC 3.3577 were grown on PDA plates with cellophane at room temperature, and the mycelia from the cultures grown for 4, 8, and 16 days were collected and combined. Mycelia were frozen by liquid nitrogen, ground extensively, and suspended in ～2 mL of RNA extraction reagent (Vazyme Biotech Co., Ltd) at room temperature for 5 min. The sample was centrifuged at 13,800 x g for 5 min at 4 ^{o}C and the supernatant was mixed with 300 μL of chloroform. The resulting solution was vortexed vigorously for 15 s and incubated for 5 min at RT. After centrifugation at 5000 x g for 5 min at 4 ^{o}C, the aqueous phase was mixed with an equal volume of isopropyl alcohol and the resulting mixture was placed at -20 ^{o}C for 10 min. RNA was collected by centrifugation (13,800 x g, 10 min), washed with 0.5

mL of 75% ethanol, and dissolved in double distilled water (ddH_2O).

Synthesis of the 1st strand of cDNA was performed by using HiScript II 1st Strand cDNA Synthesis Kit (Vazyme Biotech Co., Ltd) according to the manufacturer's instructions. Briefly, 1 μL of Oligo $dT_{23}VN$ (50 uM), 10 pg-5 μg of total RNA and RNase free ddH_2O were added together to make a 8 μL mixture, and the mixture was incubated at 65 °C for 5 min before placing on ice for 2 min. Then, 10 μL of 2×RT Mix and 2 μL of HiScript II Enzyme Mix were added to the mixture, and the 1st cDNA was reverse transcribed by a two-step procedure at 50 °C for 45 min and then 85 °C for 5 min. The primers used in the transcriptional analysis are given in Table S3.

Construction of the plasmid for expressing S-adenosylhomocysteine (SAH) hydrolase

The SAH hydrolase gene was amplified from the genomic DNA of *Streptomyces coelicolor* A3(2) by using a primer pair *SahH-F* and *SahH-R*. The PCR amplified products were digested with NdeI and HindIII, purified using a Qiagen PCR purification kit, and inserted into the same restriction site of pET28a (Novagen). Chemically competent *E. coli* DH5α cells were transformed with the ligation mixture and plated on LB-agar containing kanamycin (50 μg mL-1) to screen for positive clones, which were confirmed by DNA sequencing.

Construction of the NeurM- and BeauM-expression plasmids

The CDS (coding domain sequence) fragments of NeurM and BearM were amplified from the cDNA by using a primer pair *NeurM-F* and *NeurM-R*, and a primer pair *BeauM-F* and *BeauM-R*, respectively. The resulting PCR products were inserted in the NdeI/XhoI restriction site of pET28a (Novagen) by homologous recombination using the One Step Cloning Kit (Vazyme Biotech Co., Ltd). Typically, 20 μL of a mixture containing 4 μL of 5×CE II Buffer, 2 μL of Exnase II, 50-200 ng linear plasmid, 20-200 ng PCR fragment and ddH_2O was incubated at 37 °C for 30

min, and the resulting solution was used to transform chemically competent *E. coli* DH5α cells. Colony PCR and sequencing were carried out to confirm the sequence fidelity of the recombinant plasmids.

Construction of the PhomM- and MetaM-expression plasmids

Codon-optimized (for *E. coli*) *phomM* (Sequence I) and *metaM* (Sequence II) were ordered from Genewiz Inc. (Suzhou, China). *phomM* and *metaM* were amplified by PCR using a primer pair *PhomM-F* and *PhomM-R*, and a primer pair *MetaM-F* and *MetaM-R*, respectively. The PCR-amplified products were digested with NdeI and HindIII, purified using a Qiagen PCR purification kit, and inserted into the same restriction site of pET28a (Novagen). Chemically competent *E. coli* DH5α cells were transformed with the ligation mixture and plated on LB-agar containing kanamycin (50 μg mL^{-1}) to screen for positive clones, which were confirmed by DNA sequencing.

Enzyme overexpression and purification

Chemically competent *E. coli* BL21 (DE3) cells (for *phomM* and *metaM*) or *E. coli* Rosetta (DE3) cells (for *sahH*, *neurM* and *beauM*) were transformed with each protein-expression construct. A single colony transformant was used to inoculate a 30 mL culture of LB supplemented with 100 μg/mL kanamycin. The culture was grown at 37 °C for 12 h and was used to inoculate 3 L of LB medium containing 100 μg/mL kanamycin. Cells were grown at 37 °C and 220 rpm to an OD_{600} ～ 0.8, and were then chilled on ice for 10 min. IPTG was added to the culture to a final concentration of 0.1 mM before additional 15-20 h of incubation at 18 °C and 220 rpm. The cells were harvested by centrifugation at 4000 x g for 15 min at 4 °C. The pellet was resuspended in 30 mL of start buffer (20 mM Tris, 500 mM NaCl, 1 mM TCEP, 10% glycerol, pH 8.0) and were lysed using a high pressure homogenizer (FB-110X, Shanghai Litu Mechanical Equipment Engineering Co., Ltd, China or JN-02HC, JNBIO, China). Cell debris was removed via centrifugation at 23,800 x g for 30 min

at 4 °C. The supernatant was passed through a column containing 5 mL of high-affinity Ni-NTA resin (Qiagen Co. Ltd) equilibrated with start buffer, and the column was then washed using 50 mL of wash buffer (20 mM Tris, 500 mM NaCl, 50 mM imidazole, 10% glycerol, pH 8.0). The protein fractions were eluted by using 10 mL of elution buffer (20 mM Tris, 500 mM NaCl, 500 mM imidazole, 10% glycerol, pH 8.0). The desired fractions were combined and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit and analyzed by SDS-PAGE (12% Tris-glycine gel). Protein concentration was determined using a Bradford Assay Kit (Promega) using bovine serum albumin (BSA) as a standard. Protein solution was then subjected to desalting on a DG-10 column (Bio-Rad) pre-equilibrated with the elution buffer (50 mM Tris, 25 mM NaCl, 10 mM DTT and 10% (v/v) glycerol, pH 8.0). The protein fraction was collected and concentrated, and was used directly for in vitro assay or stored at −80 °C until further use.

Genome mining for putative dikaritin biosynthetic gene clusters

The nucleotide and amino acid sequences of putative dikaritin biosynthetic gene clusters were obtained from the National Center for Biotechnology Information (NCBI) sequence database. To identify putative dikaritin biosynthetic gene clusters, BlastP searches were performed using the PhomA, PhomQ, and PhomM protein sequences as different queries. Hits were selected with E-value < 1E-10, 1E-20 and 1E-15, respectively. The genes around the identified hits were inspected, and the gene clusters were recognized as dikaritin biosynthetic gene clusters only when all the proteins listed below were found to be encoded in the gene cluster. These proteins are 1) a PhomA-like protein containing an N-terminal leader peptide and several repeated core/recognition motifs; 2) a PhomQ analogous protein; 3) a methyltransferase (may or may not be homologous to PhomM); 4) multiple DUF3328 proteins; and 5) transporter proteins or peptidases. The sequences of the putative dikaritin precursor peptides and the accession numbers of the gene clusters are summarized in Supplementary Table 2.

Phylogenetics

Bayesian MCMC inference analyses were performed using the program MrBayes (version 3.2) (10). Final analyses consisted of two sets of eight chains each (one cold and seven heated), run for about 2 million generations with trees saved and parameters sampled every 100 generations. Analyses were run to reach a convergence with standard deviation (SD) of split frequencies < 0.004. Posterior probabilities were averaged over the final 75% of trees (25% burn in).

Fig. S1 A hypothetical biosynthetic pathway to phomopsin A. The order of the posttranslational modifications is not known and just one sequence is shown here. The tyrosinase PhomQ likely forms the Tyr-Ile crosslink in the precursor peptide PhomA, but the identity of the enzyme(s) responsible for the dehydrogenation reactions is at present still unclear. The S41 family peptidase PhomP1 may excise the core peptides from the precursor peptide. Since both the core peptide and the recognition sequence immediately upstream of the core peptide end in Asp, it is possible that PhomP1 excises the core peptide by cleaving twice. The methyltransferase PhomM necessarily needs to act after the N-terminus of the core peptide is liberated by proteolysis. We have not detected any obvious candidates for the enzymes that need to carry out the chlorination and hydroxylation reactions. These transformations may either be carried out by the ORFs of unknown function in the clusters (including predicted hydroxylase and oxidoreductase proteins) or by proteins encoded outside of the clusters.

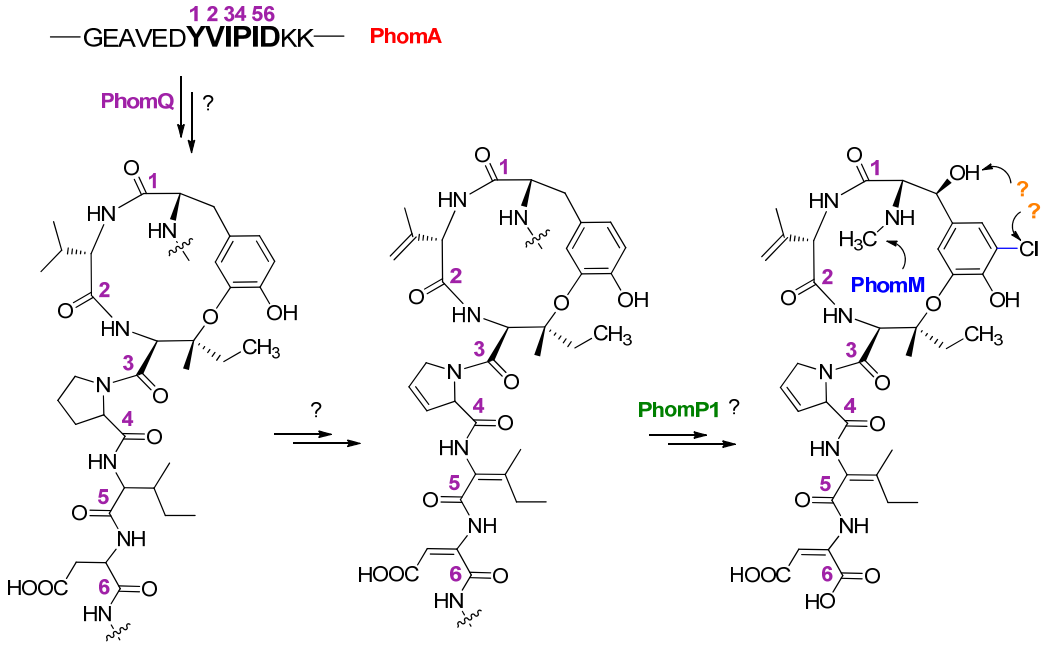
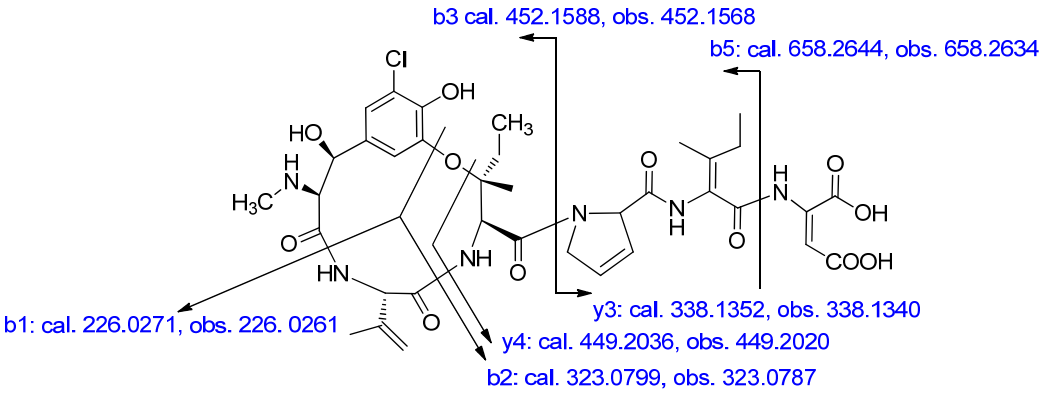


Fig. S2. HR-MSMS analysis of phomopsin A, showing its structure and the ion fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B).

A



B

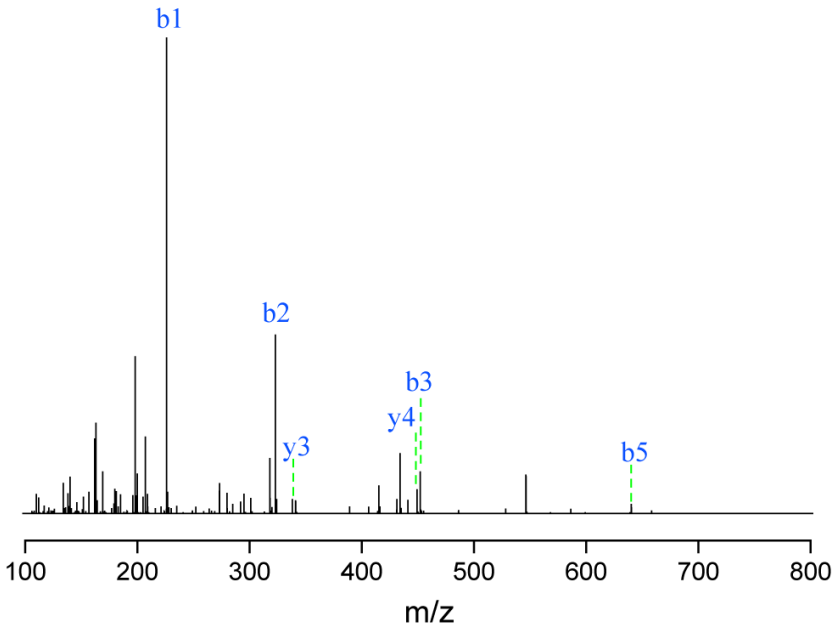


Fig. S3. UV-Vis spectrum of phomopsin A.

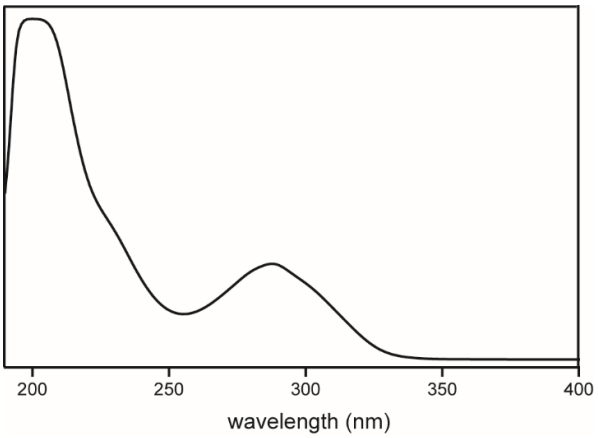


Fig. S4. ^{1}H NMR spectrum (500 MHz, d_6-DMSO) of phomopsin A.

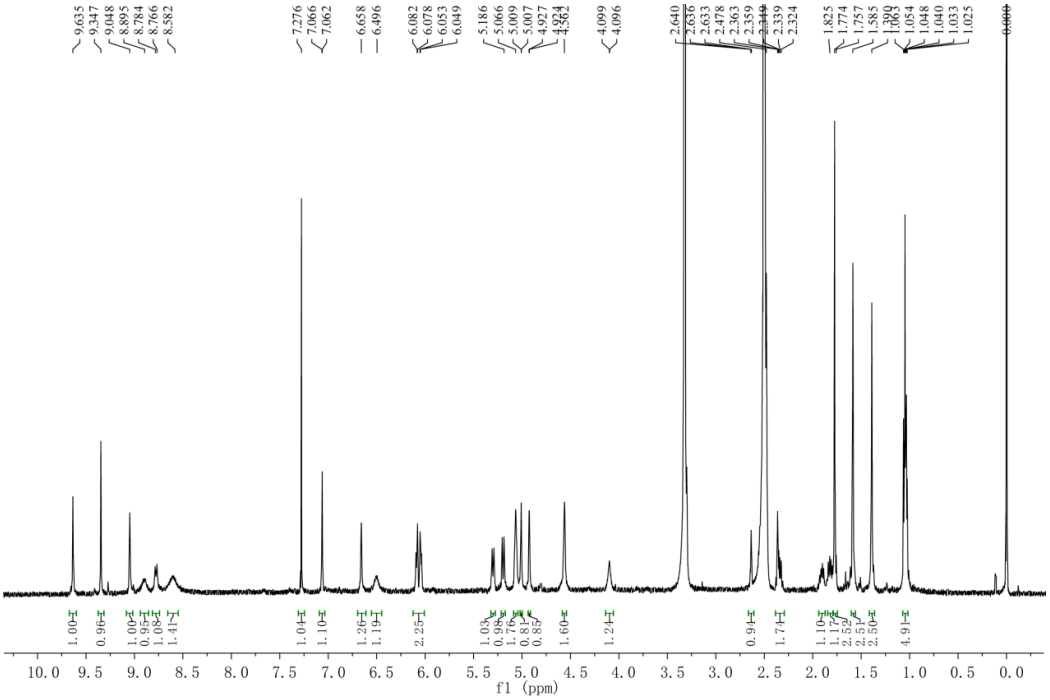


Fig. S5. ^{13}C NMR spectrum (126 MHz, d_6-DMSO) of phomopsin A.

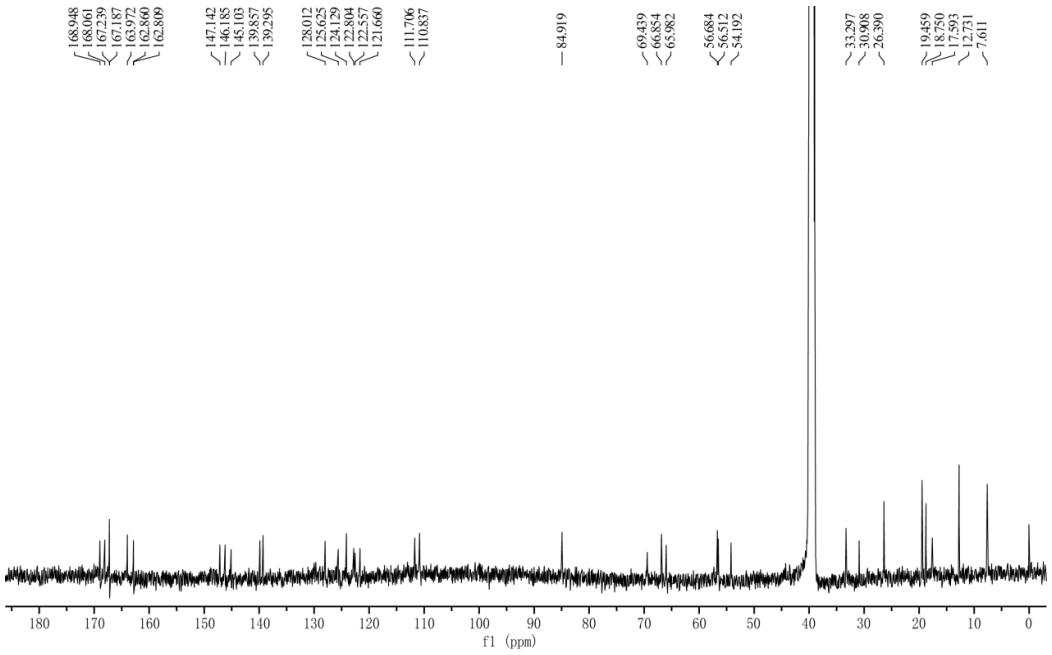


Fig. S6 gCOSY spectrum (d_6-DMSO) of phomopsin A.

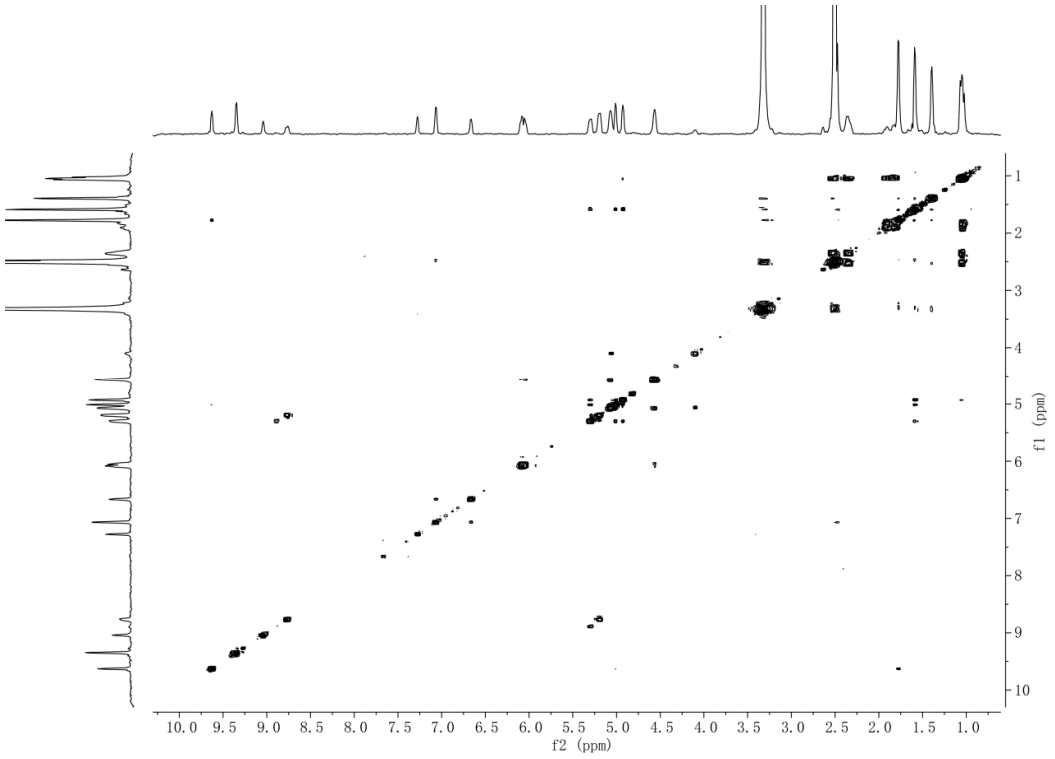


Fig. S7 HSQC spectrum (d_6-DMSO) of phomopsin A.

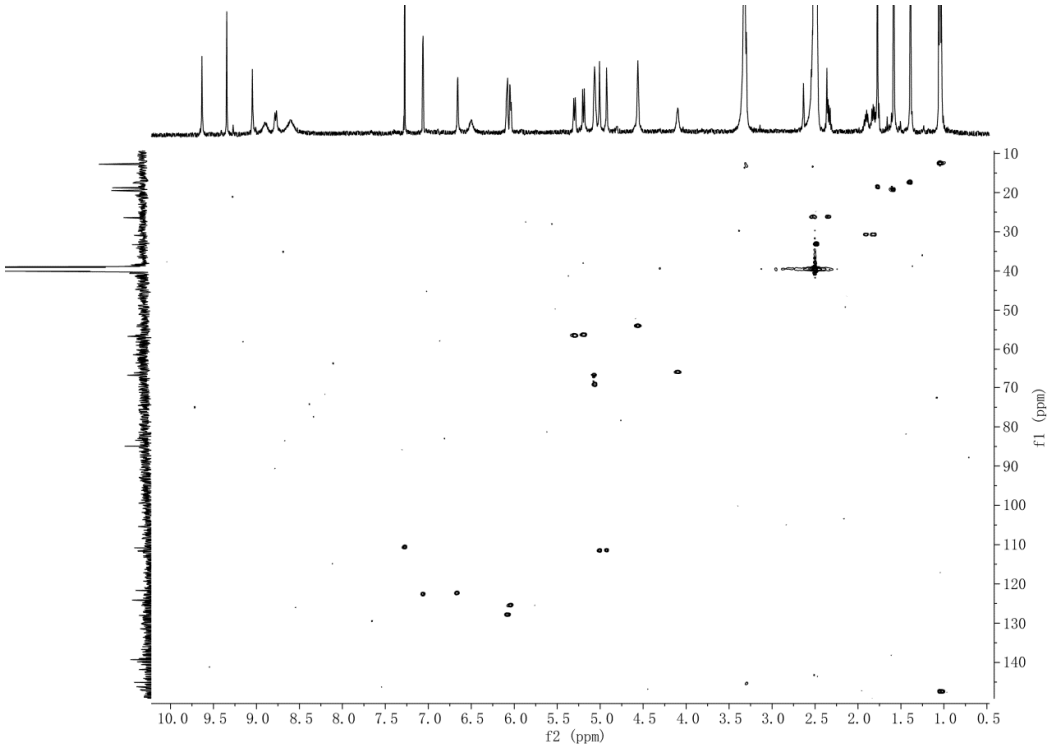
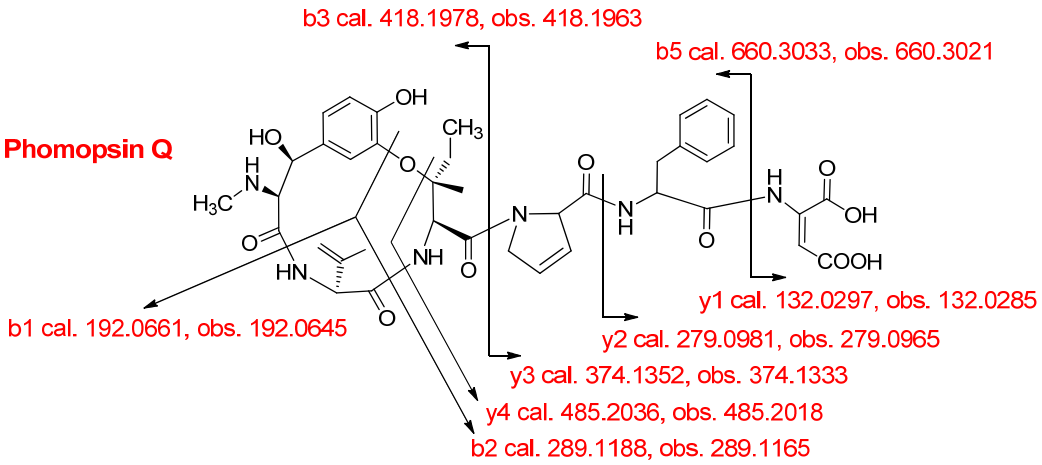


Fig. S8. HR-MSMS analysis of phomopsin Q, showing its structure and the ion fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B). The stereochemistry shown is based on analogy with phomopsin A as well as the ribosomal origin of the precursor peptide.

A



B

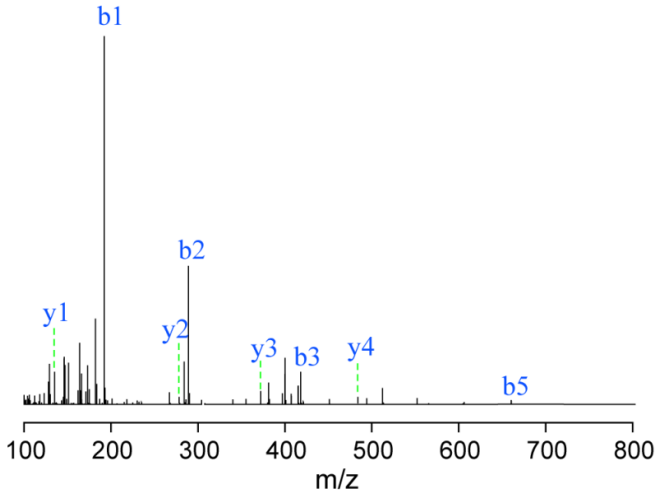
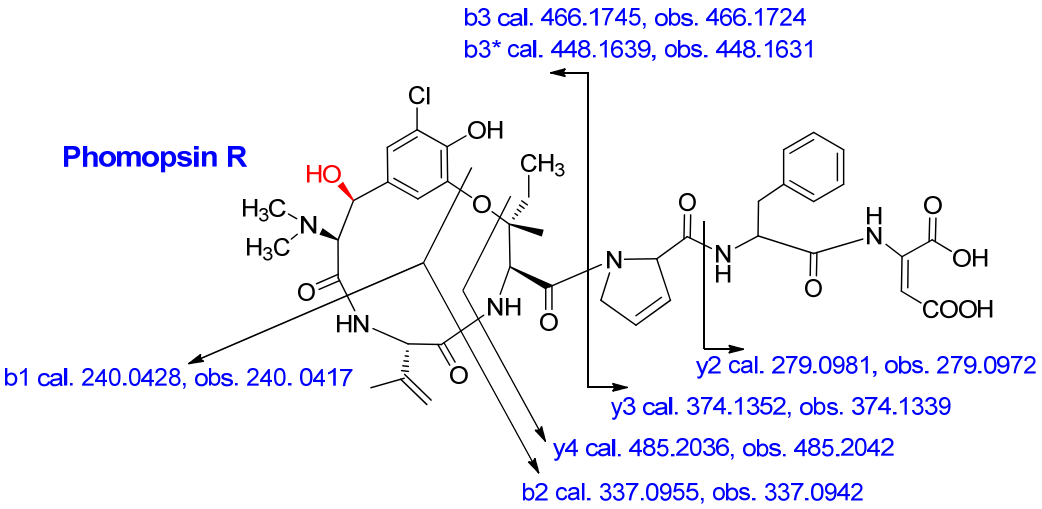


Fig. S9. HR-MSMS analysis of phomopsin R, showing its structure and the fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B). b* represents $[b - 18]^+$ ion, which likely derived from dehydration of the hydroxyl group highlighted in red in panel A. The stereochemistry shown is based on analogy with phomopsin A as well as the ribosomal origin of the precursor peptide.

A



B

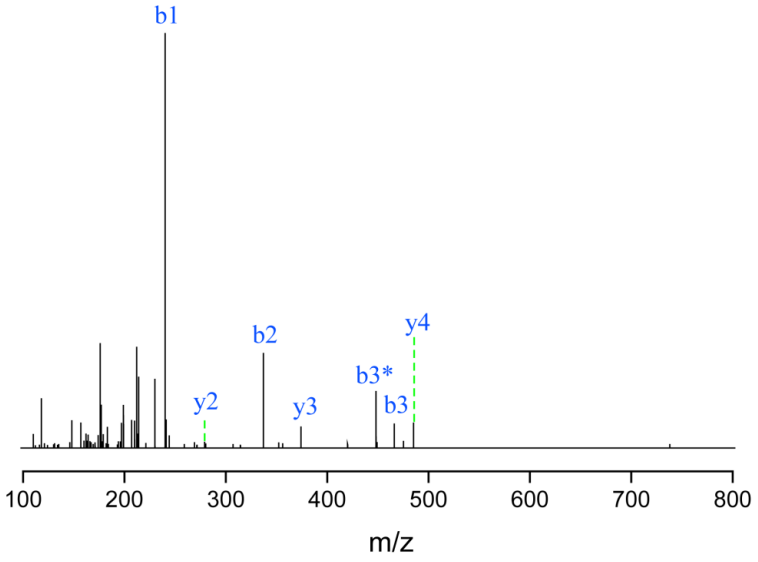
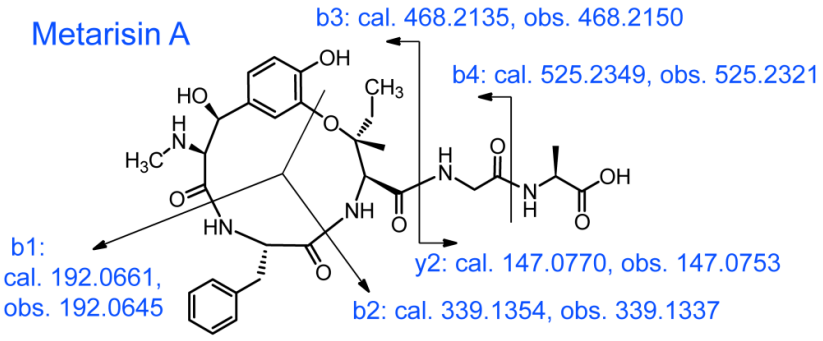


Fig. S10. HR-MSMS analysis of metarisin A, showing its structure and the ion fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B). The stereochemistry shown is based on analogy with phomopsin A as well as the ribosomal origin of the precursor peptide.

A



B

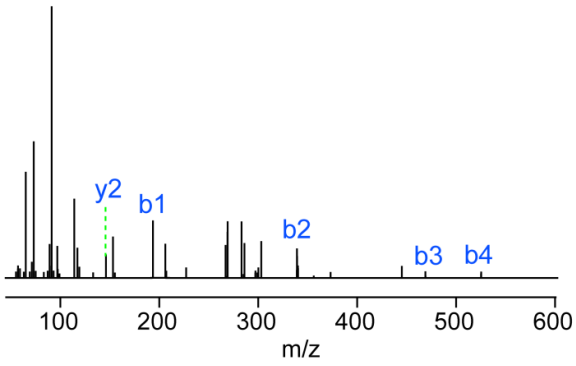
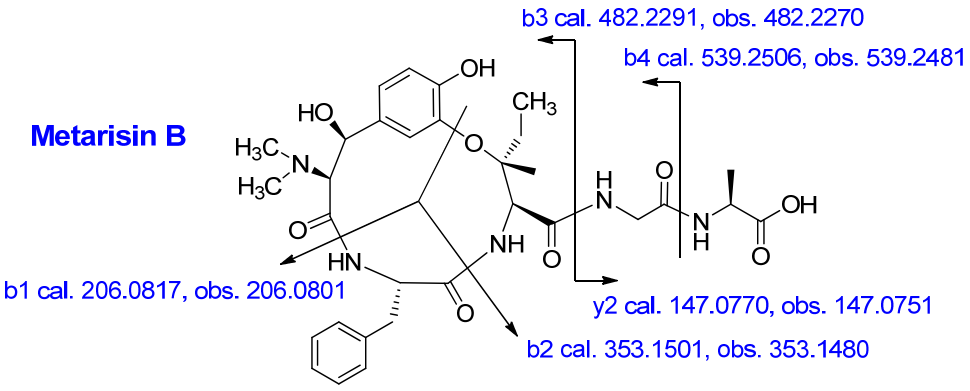


Fig. S11. HR-MSMS analysis of metarisin B, showing its structure and the ion fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B). The stereochemistry shown is based on analogy with phomopsin A as well as the ribosomal origin of the precursor peptide.

A



B

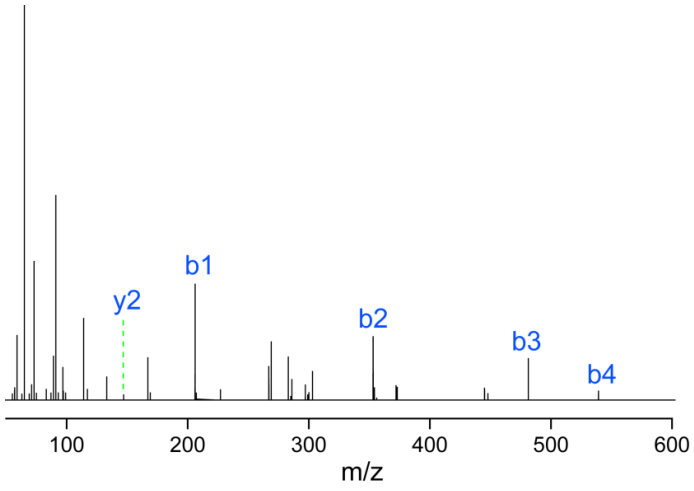
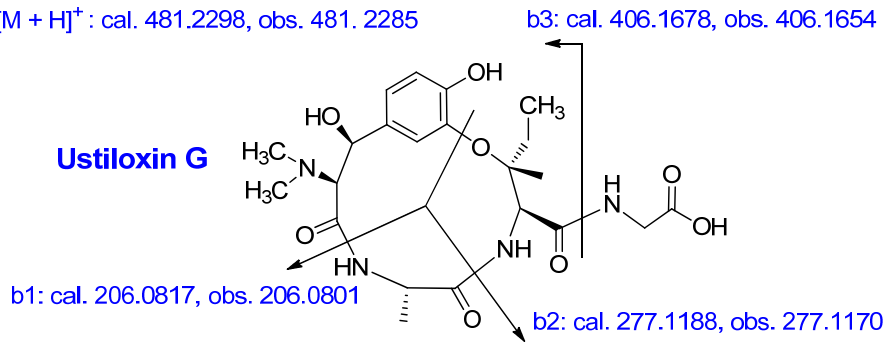


Fig. S12. HR-MSMS analysis of ustiloxin G, showing its structure and the ion fragments generated upon collision induced dissociation (A), and the MS/MS spectrum (B). The stereochemistry shown is based on analogy with phomopsin A as well as the ribosomal origin of the precursor peptide.

A



B

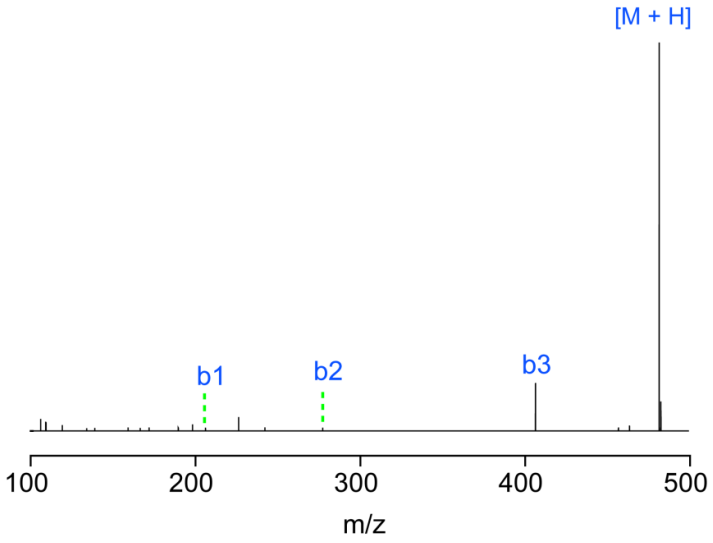


Fig. S13. The biosynthetic gene cluster of phomopsin Z from *Beauveria bassiana*, showing the Genbank accession number and the predicted function of each gene. The genes are shown in different colors in a similar way to the *phom* gene cluster shown in Figure 2B in the main text. EJP63461.1 is the precursor peptide to phomopsin Z shown in Figure 4C of the main text.

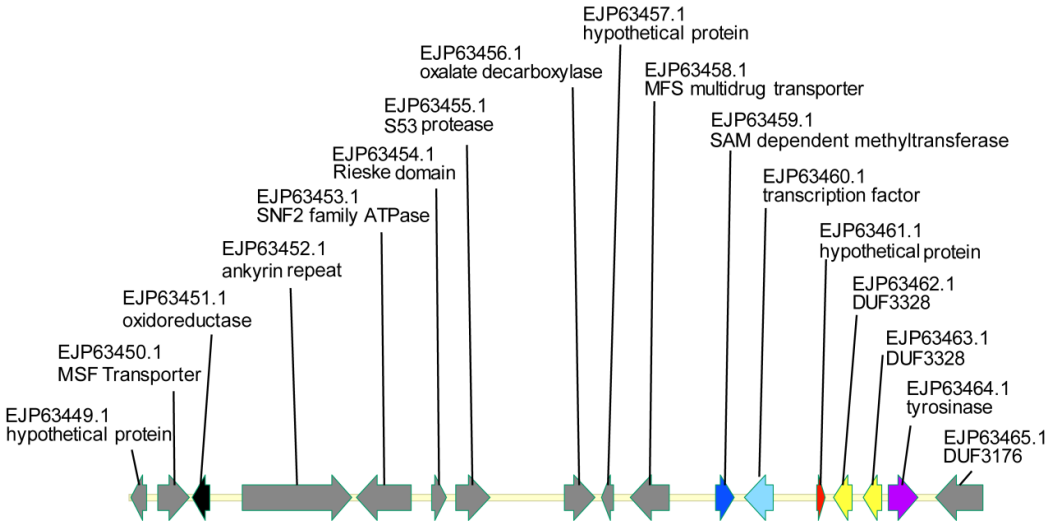


Fig. S14. Unrooted Bayesian Markov chain Monte Carlo (MCMC) tree of the methyltransferases involved in dikaritin biosynthesis. Enzymes are shown according to their NCBI accession numbers, and the Bayesian posterior probability (bPP) is shown along the branches.

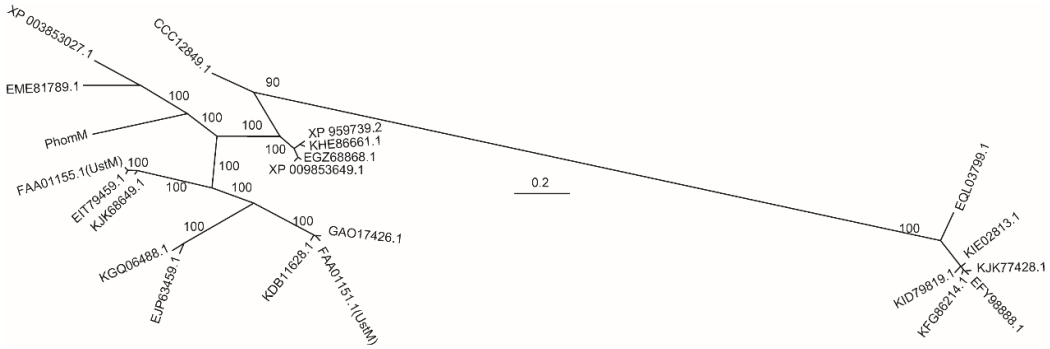


Fig. S15 HR-LCMS analysis of methyltransferase assay mixtures, showing the extracted ion chromatograms (EICs) of m/z 803.3 (corresponding to phomopsin E) and the corresponding mass spectra of each assay.

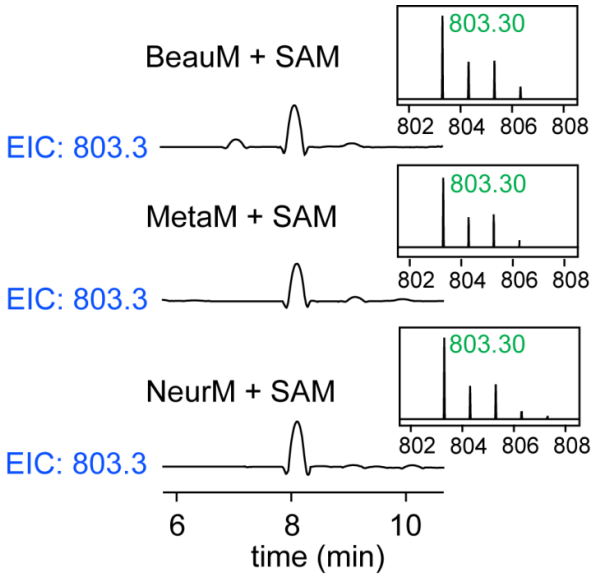


Table S1. Analysis of the phomopsin biosynthetic gene cluster using Geneid (3) with *Neurospora crassa* as a reference genome (5) followed by conserved domain analysis (11, 12) at the NCBI website. The size of the proteins is predicted and could be incorrect.

Predicted gene	Number of amino acids	Putative function
phomC	189	Cupin family protein
phomD	469	Zinc Finger protein
phomT	595	MFS multidrug transporter
phomE	396	Hydroxylase
phomF	293	Oxidoreductase
phomYa	121	DUF3328 protein
phomP1	666	Peptidase S41 family protein
phomYb	251	DUF3328 protein
phomA	144	Precursor peptide
phomR	639	Transcription Factor
phomO	1561	ABC transporter
phomM	387	Methyltransferase
phomB	342	Hypothetic protein
phomP2	842	Oligopeptide transporter
phomYc	254	DUF3328 protein
phomQ	378	Tyrosinase
phomYd	234	DUF3328 protein
phomYe	285	DUF3328 protein
phomG	1213	Peptidase

Table S2. Putative dikaritin biosynthetic precursor peptides and their associated gene clusters.

Accession Number*	Length/ repeats	Sequence Repeated core sequences are highlighted in yellow Deviations from the ED-YXIXX(D/N)-KR motif are highlighted in blue
Gene cluster: CM001199.1 from *Zymoseptoria tritici* IPO323		
EGP88004.1	196/9	MHTSSATWALFALSCTGALSAPIFNGPPLPALAPRHEAVTAVEDYLDKT KRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKT KRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKT KRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKTKRAEAVEDYVIPVDKT KRA
Gene cluster: ANFO01000810.1 from *Beauveria bassiana* D1-5		
KGQ06486.1	131/6	MKLSFISVLAAGVAIAAPTPPDNAIEDYVISVNKRSDAIEDYAIPVNKRG DAIEDYAIPVNKRGDAIEDYAIPVNKRSDAIEDYAIPVNKRGDAVEDYA IPVNKRGDAIEDYAIPVNKRGDAIEDYVISVN
Gene cluster: GL891218.1 from *Neurospora tetrasperma* FGSC 2509		
EGZ72581.1	162/6	MKYSSSILIAAFCVSVLAAPAAKRSSVEDYVIEVDKRSSVEDKRSSVED YVISVDNKRSSVEDYAIGVDKRGAVEDKRSAVEDYVIEVDKRHNSVED YAIGVDKRGSVEDKRSAVEDYVIEVDKRHNSVEDYAIGVDKRGSVED
Gene cluster: KN389650.1 from *Neurospora crassa*		
KHE86654.1	197/8	MKYSSSILIAAFCVSVLAAPAAKRSSVEDYVIEVDKRSSVEDKRGAVED KRHNSVEDYVIAVDRRGSVEDKRSSVEDYVISVDNKRSSVEDYAIGVD KRGAVEDYVIEVDKRHNSVEDYAIGVDKRGAVEDKRGAVEDKRHNSV EDYAIGVDKRGAVEDYVIEVDKA
Gene cluster: CM002242.1 from *Neurospora crassa* OR74A		
EAA30507.1	238/10	MKYSSSILIAAFCVSVLAAPAAKRSSVEDYVIEVDKRSSVEDKRGAVED KRHNSVEDYVIAVDRRGSVEDKRSSVEDYVISVDNKRSSVEDYAIGVD KRGAVEDYVIEVDKRHNSVEDYAIGVDKRGAVEDKRGAVEDKRHNSV EDYAIGVDKRGAVEDYVIEVDKRHGGVEDYAIGVDKRGSVEDKRHNS VEDYVIEVDKA
Gene cluster:GL891306.1 from *Neurospora tetrasperma* FGSC 2508		
EGO55879.1	223/9	MKYSSSILIAAFCVSVLAAPAAKRSSVEDYVIEVDKRSSVEDKRSSVED YVISVDNKRSSVEDYAIGVDKRGAVEDKRSAVEDYVIEVDKRHNSVED YAIGVDKRGSVEDKRSAVEDYVIEVDKRHNSVEDYAIGVDKRGSVED KRSAVEDYVIEVDKRHNSVEDYAIGVDKRGSVEDKRSAVEDYVIEVDK A
Gene cluster: JZEE01000032.1 from *Aspergillus parasiticus* SU-1		
KJK68632.1	253/17	MKLMLTLLVSGLCALAAPTAKRDGIEDYAIGIDKRNSVEDYAIGIDKRN SVEDYAIGIDKRNSVEDYAIGIDKRGGSVEDYAIGIDKRGGSVEDYAIGI DKRGGSVEDYAIGIDKKRNSVEDYAIGIDKRNSVEDYAIGIDKRNSVED YAIGIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRNSV EDYAIGIDKKRNSVEDYAIGIDKRGSVEDYAIGIDKRGGSVE DYAIGIDKRHGGH
Gene cluster: AP007166.1 from *Aspergillus oryzae* RIB40		
BAE63088.1	224/15	MKLILTLLVSGLCALAAPAAKRDGVEDYAIGIDKRNSVEDYAIGIDKRN SVEDYAIGIDKRNSVEDYAIGIDKRNTVEDYAIGIDKRNSVEDYAIGIDK RNTVEDYAIGIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRGGSVEDYAI GIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRGSVEDYAIGIDKKRGTVE DYAIGIDKRGGSVEDYAIGIDKRHGGH
Gene cluster:JHTR01000063.1 from *Ustilaginoidea virens*		
KDB11630.1	134/7	MKFSLISILATGFPAAVVAAPTFQENVMAKRASVEDYAIGVDKRDAVE DYVIGVDKRDAVEDYVIGVDKRDAVEDYVIGVDKRDAVEDYAIGVDK RDAVEDYVIGVDKRDAVEDYVIGVDKRDAVEDYAIGVNK
Gene cluster: JH725175.1 from *Beauveria bassiana* ARSEF 2860		
EJP63461.1	113/7	MKLSFISVLAAGVAIAAPTPPDNAIEDYVISVNKRGDAIEDYAIPVNKRG DAIEDYAIPVNKRDDAIEDYVISVDKRDDYAIPVNKRSDEIEDYAIPVNK RGDAVEDYVISVN
Gene cluster: AKHY01000126.1 from *Aspergillus oryzae* 3.042		

EIT79501.1	127/7	MKLILTLLVSGLCALAAPAAKVSKSYLIQRDGVEDYAIGIDKRNSVEDY AIGIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRNSVEDYAIGIDKRGGSV EDYAIGIDKRGGSVEDYAIGIDKRHGGH
Gene cluster: : CABT02000031.1 from *Sordaria macrospora* k-hell		
CCC12843.1	101/5	MKYSSSVLIAAFCVSVLAAPVAKSNAVEDYVIPIDKRSSVEDYVIPIDKR GSVEDYVIPIDKRSSVEDYVIPIDKRNGAVEDYAIPIDKAQLRRGLCDPD R
Gene cluster: : JYNM01000637.1from *Penicillium solitum*		
KJJ13508.1	469/10	MEGVTNAVYNNAGRGSRGNGRGVVEVTEMAEVVTTAVTTTTSSNPTS CLHTTSQCVTKSRETISIESPPVPISQEATATTAPIPNRYHDPQEKRLAVE YGKRDRDIRRAAARAKAKETGGNKLTEGGGQGFAGPGERAMAVTGD VTTIVTTDSSTDSTTHWIDETSNDVTADEIMTMDRVDVDELFLHDGEPL HAMDVELPSTYNQAMRDQWYDAMRKQLDDLAAKGTWSLIRKPPKAK VLPGKWRLTLKTNPSDYVYSFKARWVVCGNFQRIGLLLKFNNLFFFRL NLQKRAKMKLTMISILAPCLSTLVLAAPAIQERAMAKRGVEDYAIGVD KRNVEDYAIPVDKRGSVEDYAIGVDKRNVEDYAIGVDKRGSVEDYAIG VDKRGSVEDYAIGVDKRNVEDYAIPVDKRGSVEDYAIGVDKRNVEDY AIPADKRGSVEDYAIGVDKRGVEDYAIGVDKRHGNHGAV
Gene cluster:DF952562.1from *Ustilaginoidea virens*		
GAO17429.1	121/7	MKFSLISILATGFPAAVVAAPTFQENVMAKRASVEDYAIGVDKRDAVE DYVIGVDKRDAVEDYVIGVDKRDAVEDYVIGVDKRDAVEDYAIGVDK RDAVEDYVIGVDKRDAVEDYAIGVNK
Gene cluster: ADNJ02000002.1 from *Metarhizium robertsii* ARSEF 23		
EFY98885.1	132/8	MKFSTAIVVTAVAVGAAAAPTPGKTVEDYFIGADKRDKTVEDYFIGAD KRDKTVEDYFIGADKRDKTVEDYFIGADKRDKTVEDYFIGADKRDKTV EDYFIGADKRDKTVEDYFIGADKRDKTVEDYFIGAD
Gene cluster:KE384738.1 from *Metarhizium anisopliae* BRIP 53293		
KJK77431.1	132/8	MKFSTAIVVTALAVGAAAAPTPGKTVEDYFIGADKRDKTVEDYFIGAD KRDKTVEDYFIGADKRDKTVEDYFIGADKRDKTVEDYFIGADKRDKTV EDYFIGAEKRDKTVEDYFIGADKRDKTVEDYFIGAD
Gene cluster:JNNZ01000019.1from *Metarhizium anisopliae*		
KFG86213.1	118/7	MKFSTAIVVTALAVGAAAAPTPGKTVEDYFIGADKRDKTVEDYFIGAD KRDKTVEDYFIGADKQDKTVEDYFIGADKRDKTVEDYFIGADKQDKT VEDYFIGADKRDKTVEDYFIGAD
Gene cluster:CACQ02004680.1 from *Colletotrichum higginsianum*		
CCF41414.1	98/4	MKAFDIAFVLAFAAGALAMPTQGGHLATTGEAQGGIENSNGNPVEDY VIPIDKRDEAPVEDYVIPIDKRDEAPVEDYVIPIDKRDEAPVEDYVIPIDK
Gene cluster:GG697444.1 from *Colletotrichum graminicola* M1.001		
EFQ36603.1	116/5	MQIMKVSEISFVLAVAAGALALPAQGENSLMARKSGGEMAKGDGDPV VDYVIPIDKRDKDPVVDYVIPIDKRDDDPVVDYVIPIDKRNEGPVVDYV IPIDKRDEDPVVDYVIPIDK
Gene cluster: KB020750.1 from *Colletotrichum gloeosporioides* Nara gc5		
ELA31405.1	121/5	MKLSSVTAVLAFVISSLALPNMSINNGLVERDGDEAVASGESSPATDYV IPIDKRDESPATDYVIPIDKRDENPATDYVIPIDKRDDSPATDYVIPIDKR EDSADDGGNSPATDYVIPIDK
Gene cluster:AZNH01000104.1 from *Metarhizium guizhouense* ARSEF 977		
KID82083.1	118/7	MKFSTAIVVTALAVGAAAAPTPGKTVEDYFIGADKRDKTVEDYFIGAD KRDKTVEDYFIGADKRDKTVEDYFIGAEKRDKTVEDYFIGADKRDKTV EDYFIGADKRDKTVEDYFIGAD
Gene cluster:KE652192.1 from *Ophiocordyceps sinensis* CO18		
EQL03795.1	104/6	MRFSVTFIIGALAVCSNAAPTPGMSVEDYTIGPDKRGMSVEDYTIGPEK RSKTVEDYTIGPEKRGMSVEDYTIGPEKRSKTVEDYTIGPDKRGMSVED YTIGPV
Gene cluster:AZNE01000003.1 from *Metarhizium majus* ARSEF 297		
KIE02816.1	104/6	MKFSTAIVMTALAVGAAAAATPGKTVEDYFIGADKRDKTVEDYFIGA DKRDKTVEDYFIGADKRDKTVEDYFIGADKRDKTVEDYFIGADKRDKT VEDYFIGAD
Gene cluster:AZNG01000001.1from *Metarhizium brunneum* ARSEF 3297		
KID79822.1	104/6	MKFSTAIVVTALAVGAAAAPTPGKTVEDYFIGADKRDKTVEDYFIGAD

		KRDKTVEDYFIGADKRDKTVEDYFIGAEKRDKTVEDYFIGADKRDKTVEDYFIGAD
Gene cluster:AMYD01004223.1from *Colletotrichum gloeosporioides* Cg-14		
EQB43629.1	122/5	MKLSCVTAVLAFVAGSLALPNMPINNVLVERDGGEGVASGESSPATDYVIPIDKRDDSPDTDYVIPIDKRDEKPATDYVIPIDKRDDSPATDYVIPIDKREEPGADPGGDSPATDYVIPIDK
Gene cluster: : KB446559.1from *Pseudocercospora fijiensis* CIRAD86		
EME81786.1	87/3	MHKNYAAWLLIALSSGAFGAAIGLAPIIARVDKRAEAVEDYIIPIDKRHGAEAIEKRHGAEAVEDYIIPIDKRHGAEAVDDYIIPID
Gene cluster: : KB725822.1 from *Colletotrichum orbiculare* MAFF 240422		
ENH84133.1	121/4	MKFSGITTTFLALATGTLALPTLSINNNLVGRDGGEGVAVGENNPATDYVIPIDKRQEPTDPTGGSPATDYVIPIDKRDDPAQPAGESPATDYVIPIDKRQEPTDPTGGNPATDYVIPIDK

* The accession number for the gene cluster denotes the genome scaffold sequence that contains the full or partial gene cluster.

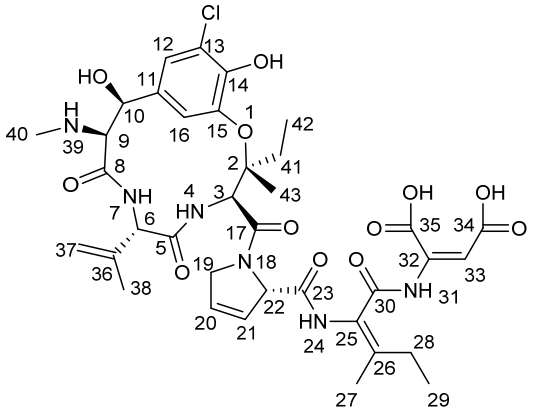
Table S3. PCR primers used in this study.

Primer	Sequence*	Function
Rt-phomA-F	GGACTATGTGATCCCTATTGAC	Transcriptional analysis of *phomA*
Rt-phomA-R	GGGATCACATAGTCCTCAACAGC	Transcriptional analysis of *phomA*
Rt-phomM-F	CACCGTCACCCCCGGGGTCC	Transcriptional analysis of *phomM*
Rt-phomM-R	GAAGAGGTCGGCCACGGACTGC	Transcriptional analysis of *phomM*
Rt-phomQ-F	GTGGGGCGGCTGTACCAGGTC	Transcriptional analysis of *phomQ*
Rt-phomQ-R	CCCCCCATCGGCATCACATCC	Transcriptional analysis of *phomQ*
trpC-F	ATGCCAGTTGTTCCAGTGATCTTCGT	*trpC* subclone
trpC-R	AAGAAGGATTACCTCTAAACAAGTG	*trpC* subclone
re-trpC-F	CTATGACCATGATTACGAATTCATGCCAGTTGTTCCAGTGATCTTCGT	*trpC* cloning
re-trpC-R	CTCTAGAGGATCCCCGGGTACCAAAGAAGGATTACCTCTAAACAAGTG	*trpC* cloning
hgh-F	GATTCAATCTGACTTACCTATTCTACCCAAGCCTCGATCATGAAAAAGCCTGAACTCACCGCGACG	pQZ101 construction
hgh-R	AACGCGTTTTATTCTTGTTGACATGGAGCTATTAAATCACTATTCCTTTGCCCTCGGACGAGTGCTG	pQZ101 construction
phomQ-F	AAAGGATCCTCCGAGACTTCGTGCCCTTCATGG	*phomM* disruption
phomQ-R	AAAAAGCTTCCTGCCATGGCACCGCATGTATC	*phomM* disruption
id-hph-F	CCGACGCCCCAGCACTCGTCCGAGGGC	*ΔphomQ* verification
id-phomQ-R	CCGGAAGCTACTCCCACTACGCGGCATC	*ΔphomQ* verification
NeurM-F	CCGCGCGGCAGCCATATGGATGATGATGATAAAATGGCCATCCCAGGAATGGACTC	*neurM* expression
NeurM-R	GGTGGTGGTGGTGCTCGACTACTCCAACTCAAACACC	*neurM* expression
BeauM-F	CCGCGCGGCAGCCATATGGATGATGATGATAAAATGGGATTAGAAAACGTCAAACCG	*BeauM* expression
BeauM-R	GGTGGTGGTGGTGCTCGATCAATACTTGTAGCACAGAAAGCC	*BeauM* expression
PhomM-F	AAACATATGGCTTCTGTTACCATCACCTCTG	*PhomM* expression

Primer	Sequence*	Function
PhomM-R	AAAAAGCTTAACGTTCCAGTTCAGCCCACCAAACG	*PhomM* expression
MetaM-F	AAACATATGCAGTCTCCGGTTAAAATCCACACC	*MetaM* expression
MetaM-R	AAAAAGCTTAGTCTTTGGTACGCTGAGCAAC	*MetaM* expression
SahH-F	AAAACATATGACGACTGTCGACAACCGACAGGACTTC	*SahH* expression
SahH-R	AAAAAAGCTTCAGTAGCGGTAGTGGTCCGCCTTGTAGGG	*SahH* expression

*Restriction sites for KpnI (GGTACC), EcoRI (GAATTC), BamHI (GGATCC), HindIII (AAGCTT), NdeI (CATATG), are underlined.

Table S4. Assignment of ^{1}H and ^{13}C NMR spectra of phomopsin A.



Position	$^{13}C(\delta)$	$^{1}H(\delta)$
1		
2	84.92	
3	56.51	5.194 (d, 1H, *J* = 9.5 Hz, CH)
4		8.582 (brs, 1H, NH)
5	168.95	
6	56.68	5.299 (d, 1H, *J* = 9.5 Hz, CH)
7		8.895 (brs, 1H, NH)
8	162.81	
9	65.98	4.097 (d, 1H, *J* =1.5 Hz, CH)
10	69.44	5.066 (s, 1H, CH)
11	125.63	
12	122.80	7.064 (d, 1H, *J* = 2.0 Hz, CH)
13	121.66	
14	147.14	9.048 (s, 1H, OH)
15	145.10	
16	122.56	6.659 (s, 1H, CH)
17	167.19	
18		
19	54.19	4.562 (s, 2H, CH_2)
20	125.63	6.035-6.042 (m, 1H, CH)
21	128.01	6.078-6.095 (m, 1H, CH)
22	66.85	5.066 (s, 1H, CH)
23	168.06	
24		9.635 (s, 1H, NH)
25	146.19	
26	124.13	
27	18.75	1.774 (s, 3H, CH_3)

28	26.39	2.324-2.363 (m, 1H, CH_{2a}), 2.515-2.545 (m, 1H, CH_{2b})
29	7.61	1.048 (t, 3H, J = 7.5 Hz, CH_3)
30	162.86	
31		9.347 (s, 1H, NH)
32	139.29	
33	110.84	7.276 (s, 1H, CH)
34	167.24	
35	163.97	
36	139.86	
37	111.71	4.925 (d,1H, J = 1.5 Hz, CH_{2a}), 5.008 (d,1H, J = 1.5 Hz, CH_{2b})
38	19.46	1.585 (s, 3H, CH_3)
39		6.496 (brs, 1H, NH)
40	33.30	2.478 (s, 3H, CH_3)
41	30.91	1.818 (q, 1H, J = 7.5 Hz, CH_{2a}), 1.908(q, 1H, J = 7.5 Hz, CH_{2b})
42	12.73	1.040 (t, 3H, J = 7.5 Hz, CH_3)
43	17.59	1.390 (s, 3H, CH_3)

Sequence I. The codon-optimized sequence of *phomM* for expression in *E. coli.*

ATGGCTTCTGTTACCATCACCTCTGTTCACGAACCGGCTAACCCGCTGGCTGTTGCTCACAC
CGGTGAACGTGAAGTTGAAGGTGACCAGCTGATCAAAACCGACACCAACAACGACAACG
ACATCATCAACACCCTGCTGAACAACTTCCTGTACCCGCAGGAACTGAAACACAACGTTTT
CGTTCCGCGTTTCCAGCAGCGTATCACCATCGTTAAAGCGTGGGACATCCTGTCTCTGTCTA
CCTCTCCGCCGAACGGTACCAACGGTGCTAACGGTACCAACTCTGCTCCGCACCGTCCGAA
AGACCTGCACATCCTGGACATCGGTTGCGGTCAGGGTGAATCTGCTGCTACCATGGCTGCT
CTGCTGCAGCCGCACATGCACGGTGCTCGTCTGCACATCACCGGTATCGACACCGCTCGTC
CGGACTACGGTACCCCGTACACCGTTGCTGAAACCCACGCTCACCTGACCGCTTCTGCTCT
GGGTCGTCACATCTCTTTCCGTCGTGAAGACGCTGCTGCTTTCTTCTCTCCGTCTCGTCTGT
CTTCTCCGTCTCCGCCGGGTTCTTGGCCGTCTGCTGCTAACGTTGACGCTGTTACCCTGTGC
CACTCTCTGTGGTACTTCCCGACCCCGCAGTCTGTTGCTGACCTGTTCACCACCCTGGCTGG
TGCTCGTGTTCCGCGTGTTTACCTGGCTGAATACTCTTTCCGTGGTTCTCTGCCGGGTGGTC
AGCAGGACGCTCACATCCTGGCTGCTCGTGCTCAAGCTCTGCTGCACGCTTCTGTTCTGGA
AAAACTGTCTGCTGACTCTTCTCAACAGAACCACCAGGGTCGTGAACCGGGTCCGCGTGC
TCCGAACGTTCGTGCTGCTCTGGACGTTGGTTCTATCGTTGAAGCTGCTGCTGCTGCTGGTT
GGGCTGTTCGTCGTCAGGGTACCTTCGTTCCGGCTGCTGACATGATCGAAGGTCACCTGGA
AGCTCGTCTGGTTGTTAAAGACGCTTTCGCTGAAGCTGTTCGTGCTGAAGGTCTGTCTCCG
GAACGTGAACACGAAGTTCTGGGTCTGGTTCCGGGTGTTAAAGAAGCGTTCGCTCGTCTGG
CTGAAGCTGGTGTTGCTAAAGGTCGTGCTATGGACGTTTGGTGGGCTGAACTGGAACGTTA

Sequence II. The codon-optimized sequence of *metaM* for expression in *E. coli.*

ATGCAGTCTCCGGTTAAAATCCACACCGTTGAATCTTCTTACTCTGACGCTTACCAGGAAGC
TGTTATCGCTAACTACGACGACCCGCCGGCTGTTTGGGAAAAAGTTCTGGGTGAAACCCTG
TCTTTCAACGGTGGTCTGTTCGACGAAGCTGAACTGACCGCTGGTCCGAAACCGGGTTCTG
TTGGTGCTTCTGAATTTCGTGGTATCAACCGTCAGCTGGAACTGGCTGGTCTGCTGTCTCCG
GACCGTCAGCTGCTGCGTCGTATCCTGGACCTGGGTTGCGGTTGGGGTGTTCTGACCCAGC
ACCTGGCTAAAGTTTTCCCGGAATGCCAGTGCATCGACGCTATCAACATCTCTCAACAGCA
GCTGGACTACTGCGCTGAAAAACTGCCGCCGGAACTGCGTAAACGTGTTAACCTGTACCTG
TGCAACGCTCAAGACGTTGACCGTCTGCCGGACCCGACCGAACCGTACGACTTCGTTTTCG
TTCGTGGTGTTTACTTCCACCTGCTGCCGTCTGTTTTCGAAGCTAGTGTTGCTCGTCTGGCT
CAACGTATCCGTCCGGGTGGTATCCTGCTGCTGTCTGACCCGCTGTACCGTGACGCTGACGT
TGACGCTCCGGTTTCTACCGCTTCTGAACCGTCTGACGGTCTGGGTACCGGTGACCACAAA
TCTCCGCAGTACTACACCTCTGTTCTGAAAAAACACGGTTTCCAGATCCAGGACCTGCGTG
TTCTGCCGTCTAACGCTGAATTTATCCACTGGTTCCGTGTTCTGCGTCTGAACATCGAAGCT
AACTTCCCGACCTTCCCGAACGGTGTTTCTGTTCCGGTTAAAGACCTGCACGAATTTGCTG
AATCTTTCGCTCAAAAACTGGCTGAAGACAAAGTTTCTATGTACTCTATCGTTGCTCAACGT
ACCAAAGACTAA

References

1. Li R*, et al.* (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20(2):265-272.
2. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18(5):821-829.
3. Blanco E, Parra G, & Guigo R (2007) Using geneid to identify genes. *Curr Protoc Bioinformatics* Chapter 4:Unit 4 3.
4. Stanke M & Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33(Web Server issue):W465-467.
5. Galagan JE*, et al.* (2003) The genome sequence of the filamentous fungus Neurospora crassa. *Nature* 422(6934):859-868.
6. Cullen D, Leong SA, Wilson LJ, & Henner DJ (1987) Transformation of Aspergillus nidulans with the hygromycin-resistance gene, hph. *Gene* 57(1):21-26.
7. Mullaney EJ, Hamer JE, Roberti KA, Yelton MM, & Timberlake WE (1985) Primary structure of the trpC gene from Aspergillus nidulans. *Mol Gen Genet* 199(1):37-45.
8. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97(12):6640-6645.
9. Anco DJ, Kim S, Mitchell TK, Madden LV, & Ellis MA (2009) Transformation of Phomopsis viticola with the green fluorescent protein. *Mycologia* 101(6):853-858.
10. Ronquist F*, et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539-542.
11. Pruitt KD, Tatusova T, & Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35(Database issue):D61-65.
12. Marchler-Bauer A*, et al.* (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222-226.