Heterologous pathway assembly reveals molecular steps of fungal terreic acid biosynthesis

Chuixing Kong¹, Hezhou Huang¹, Ying Xue¹, Yiqi Liu¹, Qiangqiang Peng¹, Qi Liu¹, Qin Xu¹, Qiaoyun Zhu¹, Ying Yin¹, Xiangshan Zhou¹, Yuanxing Zhang¹,² & Menghao Cai¹

Terreic acid is a potential anticancer drug as it inhibits Bruton’s tyrosine kinase; however, its biosynthetic molecular steps remain unclear. In this work, the individual reactions of terreic acid biosynthesis were determined by stepwise pathway assembly in a heterologous host, *Pichia pastoris*, on the basis of previous knockout studies in a native host, *Aspergillus terreus*. Polyketide synthase AtX was found to catalyze the formation of partially reduced polyketide 6-methylsalicylic acid, followed by 3-methylcatechol synthesis by salicylate 1-monoxygenase AtA-mediated decarboxylative hydroxylation of 6-methylsalicylic acid. Our results show that cytochrome P450 monooxygenase AtE hydroxylates 3-methylcatechol, thus producing the next product, 3-methyl-1,2,4-benzenetriol. A smaller putative cytochrome P450 monooxygenase, AtG, assists with this step. Then, AtD causes epoxidation and hydroxyl oxidation of 3-methyl-1,2,4-benzenetriol and produces a compound terremutin, via which the previously unknown function of AtD was identified as cyclooxygenation. The final step involves an oxidation reaction of a hydroxyl group by a glucose-methanol-choline oxidoreductase, AtC, which leads to the final product: terreic acid. Functions of AtD and AtG were determined for the first time. All the genes were reanalyzed and all intermediates and final products were isolated and identified. Our model fully defines the molecular steps and corrects previous results from the literature.

Fungal secondary metabolites are well known for their wide-ranging biological activities. Terreic acid (TA, compound 1, Fig. 1) is a polyketide that was originally isolated from *Aspergillus terreus* and has an inhibitory effect against bacteria. The compound also selectively inhibits the catalytic activity of Bruton’s tyrosine kinase (Btk), and this kinase significantly affects mast cell activation and B-cell development. Recently, a selective inhibitor of Btk, ibrutinib, was approved by the US FDA for the treatment of mantle cell lymphoma and chronic lymphocytic leukemia. Thus, TA or its derivative Btk inhibitors have a good potential as anticancer pharmaceuticals and arouse interest at present; characterization of TA’s biosynthetic mechanism will facilitate industrial biosynthesis and experiments with TA or screening of its bioactive derivatives.

Although a radiolabeled-precursor approach already demonstrated in the 1960s that TA derives from 6-methylsalicylic acid (6-MSA) via decarboxylation and a series of oxidation steps, the whole sequential biosynthetic pathway has not been deciphered until recently. In 2014, Boruta and Bizukojc reported an at gene cluster for TA biosynthesis by bioinformatic analysis of *A. terreus* genome. Guo and coworkers then proposed a biosynthetic pathway for TA by means of a gene knockout approach to this at cluster. The 6-MSA synthase (6-MSAS) encoded by atX, which was identified by Fujii and coworkers for the first time, first utilizes one acetyl-CoA as a starter unit and three malonyl-CoA molecules as extension units and catalyzes a series of programmed reactions including Claisen condensation, dehydration, reduction, and cyclization to generate 6-MSA (compound 2, Fig. 1). The atA-encoded 6-MSA decarboxylase then catalyzes decarboxylation and hydroxylation reactions to form a predicted compound: 3-methylcatechol (compound 5, Fig. 1), followed by a hydroxylation reaction catalyzed by the atE-encoded cytochrome P450 monooxygenase to produce a predicted compound, 3-methyl-1,2,4-benzenetriol (compound 6, Fig. 1). This reaction could be catalyzed by a catechol 1,2-dioxygenase encoded by a gene outside the at cluster resulting in formation of a nonaromatic compound: (2E,4Z)-2-methyl-2,4-hexadienedioic acid

¹State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai, 200237, China. ²Shanghai Collaborative Innovation Center for Biomanufacturing, 130 Meilong Road, Shanghai, 200237, China. Correspondence and requests for materials should be addressed to M.C. (email: cmh022199@ecust.edu.cn)
Another cytochrome P450 monooxygenase, the one encoded by \( atG \), was predicted to drive the next step (epoxidation) generating terremutin (compound 3, Fig. 1). Then, a glucose-methanol-choline (GMC) oxidoreductase encoded by \( atC \) is thought to catalyze a reaction of oxidation of terremutin, thereby yielding the final product: TA 7.

Because of the complicated metabolic background in native \( A. \) terreus, compounds 5 and 6 could not be isolated from the strain and identified 7. Although deficiency in \( atE \) leads to accumulation of a shunt product 7, the relation between AtE and its substrate 3-methylcatechol still needs to be verified. Moreover, deficiency in \( atD \) and \( atG \) blocks TA synthesis but no intermediates or shunt products have been identified, leaving a gap in functional characterization of both enzymes 7. Of note, AtD shows homology to PatJ from the patulin cluster 9, but the functions of both putative enzymes are still unknown 7; however, no characterized homologue can be identified for AtG even though it contains a conserved cytochrome P450 monooxygenase domain 7.

To exactly determine the functions of \( atA \), \( atE \), \( atG \), \( atD \), and \( atC \), a stepwise pathway assembly in a heterologous host may work. In a previous study, we successfully constructed a \( Pichia \) \( pastoris \) (\( Komagataella \) \( phaffii \)) strain carrying \( A. \) \( terreus \) \( atX \) encoding 6-MSA synthase (6-MSAS) 10. When \( A. \) \( nidulans \) \( npgA \) is introduced next, which encodes a phosphopantetheiny1 transferase (PPTase) for activation of acyl carrier protein (ACP) domain in polyketide synthase (PKS), the target polyketide product 6-MSA is efficiently synthesized (2.2 g/L) 10. The high activity of 6-MSAS in \( P. \) \( pastoris \) indicates that this host may be a suitable chassis organism for proteins from \( A. \) \( terreus \), and it was therefore chosen for heterologous expression of the TA pathway in the present study.

Guo et al. reported that eight genes in the \( at \) cluster participate in TA biosynthesis, and \( atB \) located there was proposed to be a putative gene for a transporter, whereas \( atF \) was suggested to encode a putative zinc family transcription factor 7. Because heterologous pathway assembly usually involves promoters and transcription factors from a chassis microorganism, \( atF \) is not necessary for the TA pathway expression in \( P. \) \( pastoris \). The transporter protein is also dispensable for the TA synthesis process regardless of the producing ability. Besides, functions of \( atX \) were fully characterized by both a knockout 7 and \( P. \) \( pastoris \) expression 10 elsewhere. Thus, we mainly focused on \( atA \), \( atC \), \( atD \), \( atE \), and \( atG \) in this study. The functional genes of the \( at \) cluster were expressed here separately and combinatorially to clarify their roles in TA biosynthetic steps. The heterologous biosynthesis of TA and of the intermediates was realized via combinatorial expression of various functional genes, and the functions of the biosynthetic genes were finally confirmed and redefined, thereby correcting previous results from the literature and describing all the reactions of the TA biosynthesis pathway.

**Results**

**Cloning and intron identification of TA biosynthetic genes.** The mRNAs of \( atA \), \( atE \), \( atG \), \( atD \), and \( atC \) were obtained, reversely transcribed to cDNA and sequenced. The protein-coding sequence and introns of each gene were then identified (Supplementary Fig. S1), thus correcting the previous results of genome shotgun sequencing (GenBank accession No. CH476602.1). The newly identified sequences were deposited in the...
### Table 1. Gene characteristics and functions in the at gene cluster. *Cupin and PatJ* are not characterized.

| Gene   | CDS length /Intron(s) position | GenBank accession No. | Cofactors (putative) | Protein blast homologs (% identity, GenBank or UniProtKB accession No.) | Function assigned |
|--------|-------------------------------|-----------------------|----------------------|---------------------------------------------------------------------|-------------------|
| atA    | 1405bp/847–910                | KY950680              | FAD/ NADH            | NahG (34%, P23262.4) 
Orf4 (25%, 1HVM7) 
SalA (32%, AAG33865.1) | salicylate 1-monoxygenase |
| atE    | 1580bp/140–203; 268–328; 706–769; 1191–1243 | KY950681              | NADH/ NADPH          | Pat (63%, A1CF6.2) 
Pah (54%, A1CF7.1) | cytochrome P450 monoxygenase |
| atD    | 1160bp/81–145; 597–650        | KY950682              | NADH/ NADPH          | Cupin (70%, CDM36831.1) 
PatJ (62%, A1CF7.1) | epoxidase |
| atG    | 583bp/142–191; 353–411        | KY950683              | NADH/ NADPH          | unknown                 | cytochrome P450 monoxygenase |
| atC    | 2277bp/364–415; 469–525; 556–613; 1125–1180; 1292–1341; 1490–1562; 1989–2048; 2184–2233 | KY950684              | FAD                  | VBS (41%, AAC49318.1) 
XPS (41%, XP_002379930.1) | GMC oxidoreductase |

Figure 3. Gene cluster (ATEG_062XX.1).
Identification of functions of AtD, AtG, and AtC. Guo et al. knocked out atG and atD in native A. terreus, and each procedure caused a loss of the TA-biosynthetic ability. Nevertheless, because no intermediates or shunt products were identified, functions of atG and atD could not be confirmed. Therefore, strains GS-NXAEG and GS-NXAED were constructed, and the heterologous expression results confirmed that AtD but not AtG works in this reaction, in contrast to the previous prediction of AtG function (Fig. 3A–C). The product was identified as terremutin (3, 49.0 mg/L) by LC-MS and 1H NMR analysis (Supplementary Fig. S5) and was consistent with the literature data. Moreover, to test whether AtG assists AtD at this step, we then introduced atG into strain GS-NXAED to generate GS-NXAED-G transformants, which contain all other intact biosynthetic genes of strain GS-NXAED. Several GS-NXAED-G strains were selected randomly, and none of them produced a higher titer of terremutin (3) than GS-NXAED did (Supplementary Table S1). Therefore, AtG did not assist AtD in this reaction. Strains GS-NXAEDC and GS-NXAEDC were constructed by introducing atD and atC simultaneously into GS-NXAED and GS-NXAE, respectively. As compared to GS-NXAED, a specific product was produced by both strain GS-NXAEDC (0.9 mg/L) and strain GS-NXAEGDC (5.8 mg/L) at retention time 25.8 min (Fig. 4A,B,E). The new product had the same retention time and ultraviolet (UV) absorption spectrum as did the TA standard (1), and feeding the TA standard into the extracted sample enhanced the compound absorption peak as expected. Besides, LC-MS results indicated m/z of 154 for this compound, in line with TA’s m/z (Supplementary Fig. S6).
These results revealed that the newly produced compound was probably the final product: TA. Nonetheless, because production of this newly generated compound was very low in strains GS-NXAEDC and GS-NXAEGDC in shake flask culture, it was not easy to obtain enough of this compound for 1H NMR analysis. Consequently, the strain was fermented in a 5 L bioreactor and the target compound was purified. 1H NMR data (Supplementary Fig. S6) were in agreement with other results 7, confirming that AtC catalyzes transformation of terremutin (3) into TA (1). As shown in Fig. 4B and E, higher TA production was observed in strain GS-NXAECDGC compared with GS-NXAEDC. Nevertheless, given that strains GS-NXAEDC and GS-NXAEGDC were constructed separately, they may contain different biosynthetic gene copies, which affected the concentration of intermediates and final products. Thus, we introduced $atG$ into the GS-NXAE strain to obtain GS-NXAE-G transformants and compared the TA production levels. Three GS-NXAE-G strains were selected for culture randomly and none of them produced a higher titer of TA (1) than GS-NXAEDC did (Supplementary Table S2). This finding indicated that AtG neither catalyzed nor assisted AtC in catalyzing the final reaction.

**AtG boosted AtE catalysis.** In the native strain of *A. terreus*, accumulation of intermediates or shunt products was not detected after a knockout of $atG$. Additionally, AtG has a putative function of cytochrome P450 monooxygenase, which might work with the other cytochrome P450 monooxygenase: AtE. Accordingly, we introduced $atG$ into the GS-NXAE strain to generate GS-NXAE-G transformants. Three strains were selected randomly for analysis of production of compounds 6 and 7. Both compounds were highly accumulated in GS-NXAE-G strains compared with GS-NXAE, especially 6 after 24 h methanol induction and 7 after 48 h methanol induction (Fig. 5). Levels of 6 were higher even in GS-NXAE-G and GS-NXAECDGC transformants than in their parent strains GS-NXAEDC and GS-NXAECDG (Supplementary Table S3). Thus, we may conclude that AtG assists AtE but not AtD or AtC in the TA-biosynthetic pathway. Accordingly, the molecular steps for TA biosynthesis were clarified (Fig. 6), and AtX, AtA, AtE/AtG, AtD, and AtC were found to function stepwise in this process. Nevertheless, how AtG works with AtE to improve the reaction step still kept unknown. To test if AtG interacts with AtE and thereby forms a protein complex, we then conducted a yeast two-hybrid (Y2H) assay 12 on both proteins. However, the interaction between AtG and AtE was not observed (Supplementary Fig. S7), indicating that they probably not form protein complex and work in other way that needs further deep work to clarify.

**Discussion**

To determine the biosynthetic molecular steps for TA (1) in *A. terreus*, stepwise pathway assembly of TA was performed here in a heterologous host: methylotrophic yeast *P. pastoris*. Coding sequences and introns of biosynthetic genes $atA$, $atE$, $atD$, $atG$, and $atC$ were identified by reverse transcription, and our findings corrected the results previously submitted to databases. As expected, *P. pastoris* expresses AtA, AtE, AtD, AtG, and AtC correctly according to western blotting analysis (Supplementary Fig. S8). Conserved domain analysis indicated that the identified AtA contains a FAD-dependent salicylate 1-monooxygenase domain catalyzing the conversion of salicylate to catechol. Coexpression of $atX$ and PPTase-encoding gene $npgA$ produced 6-MSA (2), and introducing $atA$ next led to the biosynthesis of 3-methylcatechol (5) by a decarboxylative hydroxylation reaction. Quick BLASTp results on AtA matched several functional uncharacterized salicylate 1-hydroxylases/1-monooxygenases. As reported elsewhere, some salicylate 1-monooxygenase NahG (1-hydroxylating, decarboxylating, EC 1.14.13.1) was identified in *Pseudomonas putida* 13. A hydrolase OpS4 (UniProtKB accession No. J4VWM7) catalyzing orsellinic acid to 6-methyl-1,2,4-benzenetriol by decarboxylative hydroxylation was also identified in *Beauveria bassiana* 14. Nevertheless, AtA shares only 34% identity with NahG and 25% identity with OpS4. BLAST results indicates that AtA also shares 32% identity with an *A. nidulans* salicylate 1-monooxygenase,
SalA, which was characterized biologically but not chemically (Table 1). Moreover, the oxidation behavior of AtA is similar to that of three other reported FAD-dependent monooxygenases, TropB in tropolone biosynthesis, SorbC in sorbicillinoid biosynthesis, and AzaH in azaphilone biosynthesis, where they perform oxidative dearomatization of their specific substrates. AtA may hydroxylate 6-MSA (2) to form an unstable intermediate, which would easily undergo decarboxylation to generate 3-methylcatechol (5). For biosynthesis of TA in this case, sorbicillinoids such as sorbicillactones and azaphilones such as rubropunctatin require only ring oxidation. Nonetheless, tropolone biosynthesis requires oxidation of both the ring itself and the ring methyl group of a polyketide aldehyde, in contrast to the biosynthesis of citrinin, where only the ring methyl is oxidized.

Recently, a 6-MSA decarboxylase, PatG, was identified in patulin biosynthesis and was found to catalyze the first biosynthetic step, namely, decarboxylation but not hydroxylation of 6-MSA to form m-cresol. Thus, these results finally confirmed AtA as a 6-MSA 1-monooxygenase but not the previously predicted 6-MSA decarboxylase.

Quick BLASTp of AtE revealed good identity to cytochrome P450 monooxygenases, among which, PatI and PatH (Table 1) in the biosynthesis of patulin have been chemically identified. Introduction of atE next led to production of 3-methyl-1,2,4-benzenetriol (6) via hydroxylation of 3-methylcatechol, also in agreement with the putative function of cytochrome P450 monooxygenase AtE. Furthermore, 6 easily converted to a specific compound (7) during an active culture phase with decreasing pH, and 7 quickly converted back to 6 after purification, allowing us to infer that it is 3,4-dihydroxy-2-methylcyclohexa-2,5-diene-1-one (7), a tautomer of 6 whose chemical structure could not be identified precisely. This uncertainty did not affect the biosynthetic pathway analysis in this case.

The most uncertain molecular step for this biosynthetic pathway is the conversion of 3-methyl-1,2,4-benzenetriol (6) to terremutin (3) as reported in another work. By experimental gene identification in our study, AtG was confirmed as a protein only 157 aa long. Although AtG contains a conserved cytochrome P450 monooxygenase domain, it shows low identity to other proteins in BLAST results. Particularly, AtG shows much smaller molecular weight than the proteins from BLAST results. Therefore, we predicted that AtG may not work for terremutin production, and introduction of atG into one of our strains finally proved this conclusion. By contrast, introduction of atD into one of our strains successfully generated terremutin. A conserved-domain analysis in AtD suggested that this protein contains a cupin_2 domain, and Quick BLASTp search yielded ~20 hypothetical homologues with undefined function (from filamentous fungi) with high identity to AtD, including a putative cupin protein (identity of 70%) and hypothetical dioxygenase PatJ (identity of 62%) (Table 1).

Considering its identified function in this case, it may be designated as an epoxidase with epoxidation activity.

Figure 4. The HPLC chromatograms of organic extracts from culture broth. (A) Strain GS-NXAED; (B) strain GS-NXAEGDC; (C) strain GS-NXAEGDC supplemented with a terreic acid (TA) standard; (D) TA standard; (E) strain GS-NXAEDC. A UV spectrum of the specific peak and TA standard are shown. For HPLC, samples extracted from culture broth after methanol induction for 48 h were analyzed for UV absorbance at 330 nm.
functions. On the other hand, AtD showed no significant similarity with some reported epoxidation-mediated epoxidases or cytochrome P450 enzymes, e.g., those participating in the biosynthesis of squalene, lasalocid, mycinamicin, and FD-891. Its enzymatic mechanism and specific biosynthetic roles in other fungi will be an interesting topic for future research.

Introduction of atC into one of our strains next produced the final product, TA, proving that AtC functions at this step, in line with the gene knockout results. Conserved-domain analysis indicated that AtC is a GMC oxidoreductase that matches many homologues. Nevertheless, only the versicolorin B synthase (VBS) from Aspergillus parasiticus and that from Aspergillus flavus (Table 1) have been characterized.

Fungal cytochrome P450 monooxygenases usually have versatile biocatalytic activities. Elsewhere, a knockout of atA in native A. terreus caused potent accumulation of 6-MSA (2) whereas a knockout of atG did not, meaning that the reaction of 6-MSA (2) producing 3-methylcatechol (5) is independent of AtG. To further clarify the specific function of atG, it was then introduced here into the terremutin-producing or TA-producing strains. These experiments yielded no new compound or improvement of terremutin or TA biosynthesis, suggesting that atE but not atG plays the essential role in this reaction. Our finding that strain GS-NXAE but not GS-NXAG (genes are transcribed and proteins are expressed correctly, Supplementary Figs S8 and S9).

Figure 5. Introduction of atG into strain GS-NXAE improved the biosynthesis of compounds 6 and 7. Three resulting strains (GS-NXAE-G #18, #33, and #36) were selected randomly and tested. Gene copies of atG in each strain were not determined, and the production levels of the two compounds differed among the three GS-NXAE-G strains. Compound 6 was quantified properly, but compound 7 was quantified only as the relative HPLC peak area (the highest titer of GS-NXAE-G #18 after 48 h induction was set to 100%). One-way analysis of variance (ANOVA) was employed to determine significant production differences of compounds 6 and 7 between GS-NXAE and each GS-NXAE-G strain. The P-value was used to check the significance, and it was significant at P < 0.05. OriginPro 8.0 (OriginLab Corporation, USA) was used for ANOVA. **P < 0.01 at 24 h; ***P < 0.01 at 48 h. Detailed P-value for each run was shown in Supplementary Table S8.
provides 3-methyl-1,2,4-benzenetriol (6) confirmed this notion. Of note, when atG was introduced into the 3-methyl-1,2,4-benzenetriol (6)-producing strain (GS-NXAE), it highly improved the biosynthesis of this compound. BLAST searches revealed that AtE and AtG share very low identity. As opposed to AtG, many homologues of AtE (with relatively high identity) were found by Quick BLASTp. Thus, AtG could be a putative cytochrome P450 monooxygenase assisting AtE at the hydroxylation step. Fungal cytochrome P450 usually contains four kinds of conserved motifs30,31, and sequence analysis revealed that AtG possesses PER and EXXR motifs whereas AtE contains only PER motifs. To date, a vast number of cytochromes P450 classified into ~400 families have been identified in >2500 fungal species30–32. Nonetheless, cooperation of the two types of cytochrome P450 monooxygenase at a single biocatalytic step has seldom been reported. Our protein-protein interaction analysis by Y2H assay preliminarily showed that AtG and AtE did not combine with each other, while how these enzymes (with widely divergent molecular weights) work together at this catalytic step is still an interesting topic for a future study.

After these efforts, we finally clarified the molecular steps in the TA biosynthetic pathway. These data show a fundamental pathway for biosynthesis of TA derivatives, which can be screened for anticancer pharmaceuticals. Moreover, because 6-MSA is an abundant primary intermediate in fungal secondary metabolism30,25,33, the results we reported in this work may be useful for analysis of the biosynthetic mechanism for other 6-MSA-derived bioactive compounds. In addition, the successful heterologous expression proved that P. pastoris is a good chassis organism maintaining correct bioactivity of fungal proteins, and these properties certainly facilitate heterologous biosynthesis of fungal secondary metabolites. Moreover, with a short culture phase, clean metabolic background, and easy genetic manipulations34–37, this host may be a good choice for either biosynthetic analysis or improvement of production of fungal secondary metabolites.

**Methods**

**Strains, plasmids, media, and culture conditions.** Genes for TA biosynthesis were cloned from the at cluster of *A. terreus* NIH2624. *Escherichia coli* TOP10 served as a storage host for plasmids. *P. pastoris* GS115 was used as the basic host for heterologous expression of TA-biosynthetic genes. Vectors pAG32 (kindly provided by Prof. Saurabh Joshi in University of California, San Diego)38, pPIC3.5 K (Invitrogen), and pPICZ B (Invitrogen) were employed for gene expression. Primers used for identification of introns and construction of expression strains are listed in Supplementary Tables S4 and S5. Plasmids and expression strains in this study are listed in Supplementary Tables S6 and S7. *A. terreus* was cultivated at 28 °C in the PDB medium (Hangzhou Microbial Reagent Co., Ltd., China). *E. coli* was cultured at 37 °C in the Luria-Bertani (LB) medium consisting of 0.5% yeast
extract, 1% tryptone, and 0.5% NaCl. *P. pastoris* was cultivated at 30°C in the YPD medium consisting of 1% yeast extract, 2% tryptone, and 2% glucose for seed preparation, and then cultivated in the minimal medium (MM) composed of 1.34% YNB (Sigma) and methanol for protein expression and compound biosynthesis. Methanol was added to 0.5% (v/v) every 24 h as a carbon source and inducer.

**Molecular biological techniques.** For PCR experiments, standard protocols were applied with a PCR amplification kit (TaKaRa, Cat. # R011). Fungal RNA was extracted by means of the RNAiso Total RNA Kit (TIANGEN Cat. # DP419). Plasmid DNA was isolated from *E. coli* using the TIANprep Rapid Mini Plasmid Kit (TIANGEN Cat. # DP105–03). DNA fragments separated in an agarose gel were extracted with the Universal DNA Purification Kit (TIANGEN Cat. # DP214–03). Multiple fragments were assembled via the ClonExpress™ II One Step Cloning Kit (Vazyme Biotech Co., Ltd., China). Strains *P. pastoris* GS115 and *E. coli* TOP10 and yeast vectors pPICZ B and pPIC3.5 K were purchased from Invitrogen. Transformation of yeast cells and screening of transformants were executed according to *Pichia* protocols[9]. Yeast two-hybrid (Y2H) assay were described in detail in supplementary data file (Supplementary Fig. S7).

**Identification of introns of genes within the at cluster.** The mRNA sequences of genes within the at cluster are already predicted in GenBank (GenBank accession No. CH476602.1), but many of them are different from the prediction results of the SoftBerry software. To confirm the exact positions of introns and express correct enzymes in *P. pastoris* for TA biosynthesis, cDNA for each gene was obtained and analyzed by reverse transcription of RNA. An *A. terreus* strain was cultivated at 28°C and 120 rpm in the PDB medium for 7 days, and total RNA was then extracted. A series of primers (Supplementary Table S4) for each gene were used to amplify cDNA of each gene, and the intron positions were then confirmed after DNA sequencing.

**Construction of the GS-NX strain.** In our previous study, we successfully implemented 6-MSA biosynthesis in an engineered *P. pastoris* carrying *Aspergillus nidulans* PPTase–encoding gene *npgA* and *A. terreus* 6-MSAS–encoding gene *atX*[9]. Given that several genes need to be expressed in *P. pastoris* and selective markers were limited, *npgA* and *atX* were then inserted into one plasmid in this case. The *npgA* and *atX* expression cassettes with the *AOX1* promoter (*P* _AOX1) and *AOX1* terminator were amplified from plasmids pPIC3.5K-npgA and pPICZ B-*atX*[9], respectively. Two pairs of primers TT-*AOX1-F/TT-HIS4-R and Amp-*AOX1-F/AOX1-**X**R were employed in PCR, and DNA fragment 1 (2353 bp) and fragment 2 (6785 bp) were obtained. Moreover, the selective marker *HIS4* was amplified from plasmid pPIC3.5 K with primers TT-*HIS4-F* and *ori-HIS4-R* (fragment 3). Replicon *ori* and a selective marker—ampicillin resistance gene *AmpR* with the *AmpR* promoter—(fragment 4) were amplified together from plasmid pPIC3.5 K with primers HIS4-ori-*F* and *AOX1-Amp-*R. After that, fragments 1, 2, 3, and 4 were assembled, leading to expression plasmid pPIC3-npgA-*atX*. It was transformed into *E. coli* TOP10. After PCR verification with primers 5AOX1 and 3AOX1 and DNA sequencing, the correct plasmid was obtained by means of *BglII* and transfected into wild-type *P. pastoris* GS115 by electroporation. The histidine auxotroph was used for screening of transformants for those positive for GS115-NpgA-AtX (GS-NX). The strains were then verified by genotyping PCRs (Supplementary Fig. S10).

**Construction of strains GS-NX, GS-NXE, GS-NXAE, and GS-NXAG.** Genes *atA* and *atE* were obtained by means of primers *ZB*-atA-*F* and *ZB*-atA-his6-*R* or *ZB*-atE-his6-*R* from *A. terreus* cDNA, respectively. They were then ligated to the pPICZ B vector digested with *EcoRI* and *XhoI* via seamless assembly, leading to expression plasmids pPICZ B-*atA* and pPICZ B-*atE*. The plasmids were transfected into *E. coli* TOP10 and positive transformants with correct plasmids were identified by colony PCR with primers 5AOX1 and 3AOX1 and DNA sequencing. The *GAP* promoter as an integration locus was amplified from *P. pastoris* genomic DNA with primers *ZB*-BglII-*F* and *GAP-AOX-R*. The *GAP* promoter was linearized by means of *BglII* and transfected into wild-type *P. pastoris* GS115 by electroporation. The hisidine auxotroph was used for screening of transformants for those positive for GS115-NpgA-AtX (GS-NX). The strains were then verified by genotyping PCRs (Supplementary Fig. S11–S13).

**Construction of strains GS-NXAE, GS-NXAG, GS-NXAGD.** The *GAP* promoter as an integration locus was amplified from *P. pastoris* genomic DNA with primers 3-pGGAP-F and 3-pGGAP-R. After that, it was digested with *SacI* and *SpeI* and ligated into the same sites of opened vector pAG32 to obtain vector pAG (Hyg*). The *atG* gene was cloned from *A. terreus* cDNA using primers *ZB*-atG-*F* and *ZB*-atG-*R* and inserted into pPICZ B digested with *EcoRI* and *XhoI*. The *atD* gene was cloned from *A. terreus* cDNA by means of primers pAG-*atD-F* and pAG-*atD-his6-*R* and inserted into the pAG plasmid digested with *SalI* and *BamHI*. Thus, expression plasmids pPICZ B-*atG* and pAGG-*atD* were obtained. The correct plasmids were then identified by colony PCR with primers 5AOX1 and 3AOX1 and DNA sequencing. Considering the selection marker and His tag, the *atG* containing a promoter and terminator was amplified from pPICZ B-*atG* with primers...
3.5k-AOX-atG-F and 3.5k-his6-atG-R, and next inserted into vector pPIC3.5K digested with EcoRI and BamHI, so that pPIC3.5K-atG carrying a His tag was constructed. The atG gene containing a promoter and terminator was amplified from pPICZ B-atG with primers ZB-BglII-AOX-F and TT-AOX-R, and inserted into pPICZ B-atA-GAP-atE digested with BglII, leading to expression plasmid pPICZ B-atA-GAP-atE-atG-atD. The atD gene carrying the promoter and terminator was amplified from plasmid pAGG-atD, followed by insertion into pPICZ B-atA-GAP-atE-atG digested with BglII, leading to expression plasmid pPICZ B-atA-GAP-atE-atG-atD. The correct plasmids were identified by colony PCRs and DNA sequencing. After that, plasmids pAGG-atD, pPICZ B-atA-GAP-atE-atG, and pPICZ B-atA-GAP-atE-atG-atD were linearized by AvrII digestion and transfected into strains GS-NXAE, GS-NX, and GS-NX by electroporation, separately. Zeocin (100 μg/mL) and hygromycin (750 μg/mL) served for screening of transformants for those positive for GS115-NpgA-AtX-AtE-AtD (strain GS-NXAE), GS115-NpgA-AtX-AtA-AtE-AtG (strain GS-NXAEG), or GS115-NpgA-AtX-AtA-AtE-AtG-AtD (strain GS-NXAEGDC). These strains were then verified by genotyping PCRs (Supplementary Figs S14–S16).

Construction of strains GS-NXAE-G, GS-NXAE-G, and GS-NXAE-G. The atC gene was cloned from A. terreus cDNA via primers ZB-atC-F and ZB-atC-his6-R and inserted into pPICZ B digested with EcoRI and XhoI, leading to expression plasmid pPICZ B-atC. The correct plasmid was then identified by colony PCR with primers 3AOX1 and 3AOX1 and DNA sequencing. Genes atD and atC containing a promoter and terminator were amplified from pPICZ B-atG and pAGG-atD with primers AG-AOX-F and AOX-TT-R or TT-AOX-F and AOX-TT-R, respectively, and then inserted into pAGG digested with SalI and BamHI, thus generating expression plasmid pAGG-atD-atC finally. The correct plasmids were identified by colony PCR. Next, the pAGG-atD-atC vector was linearized by AvrII digestion and transfected into strains GS-NXAE and GS-NXAEG. Hygromycin at a final concentration of 750 μg/mL was used for screening of transformants for those positive for GS115-NpgA-AtX-AtA-AtE-AtD-AtC (strain GS-NXAE-G) and GS115-NpgA-AtX-AtA-AtE-AtG-AtD-AtC (strain GS-NXAEGDC). These strains were verified by genotyping PCRs (Supplementary Figs S17 and S18).

Creation of strains GS-NXAE-G, GS-NXAE-G, and GS-NXAE-G. To eliminate the influence of biosynthetic copies on compound production levels and to test whether atG can assist the functioning of aAOX, the strain GS115 was separately integrated at the aAOX site and verified with primers 5′AOX and atG-yz-R as well as atG-xy-F and 3′AOX.

Transcriptional analysis. Total RNA was extracted according to Pichia protocols. RQI RNase-Free DNase (Promega) was employed to remove the residual DNA. Reverse transcription was conducted by means of the PrimeScript™ RT Reagent Kit (TaKaRa). Wild-type P. pastoris GS115 served as a negative control.

Protein expression and western blot analysis. Strains GS-NX and GS-NXAE carry AtA and AtE with a His tag. Plasmids pAGG-atD, pPIC3.5K-atG, and pPICZ B-atC were engineered to contain a His tag and then linearized and transfected into the wild-type P. pastoris GS115 by electroporation to obtain strains GS115-AtD-HIS6, GS115-AtG-HIS6, GS115-AtC-HIS6. These strains were then analyzed to test whether the TA biosynthetic enzymes could be correctly expressed in GS115. A total of 2 μg of plasmid DNA was added into a 2.0 mL screw cap tube with 1 g of zirconium followed by disruption in a BeadBeater (Minilys, Bertin Technologies) for 8 cycles (30 s vibration and 1 min of an ice bath in each cycle). The lysate was centrifuged (12000 × g, 5 min) and the supernatant was discarded. For western blot analysis, 20 μL of total protein samples (analyzed with the Bradford protein assay kit, Tiangen Biotech) were loaded into polyacrylamide gel wells and separated under denaturing conditions. After that, the proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane. The mouse anti-His antibody (Tiangen Biotech) and the peroxidase-conjugated goat anti-mouse immunoglobulin G (Tiangen Biotech) served as the primary antibody and secondary antibody, respectively.

Extraction and identification of TA and intermediates. After centrifugation for 5 min at 3000 × g, 50 mL of the supernatant was extracted with an equal volume of ethyl acetate. The organic phase was removed in a rotary evaporator at 40 °C, and the remainder was dissolved in 1 mL of methanol. Further analysis of the extracts was carried out by HPLC on a C18 column (Kromasil™, Sweden, 250 mm × 4.6 mm × 5 μm, 100 Å spherical silica) at a flow rate of 0.4 mL/min and detection by UV absorbance at 254 nm (intermediates) and 330 nm (TA). The gradient system was 0.1% acetic acid in H2O (solvent A) and acetonitrile (solvent B). Gradient conditions were as follows: minute 0, 15% B; minute 40, 85% B; minutes 0–55, 100% B (for terreic acid); or minute 0, 15%; minute 40, 85%; minutes 40–45, 100% B (for other compounds); or minute 0, 15%; minute 30, 85%; minutes 30–35, 100% B (for other compounds). To confirm the compounds, further analysis was performed by LC with high-resolution MS (LC-HRMS; Agilent 6230 TOF LC-MS) and NMR (Bruker-AM-400-spectro) in a freeze-dried sample dissolved in deuterated DMSO or deuterochloroform for 1H NMR, 13C NMR, HMBC, and HSQC analyses.
Data availability. All data generated or analysed during this study are included in this published article (and its Supplementary Information files). Genes re-annotated are also deposited in GenBank and the assigned accession numbers are provided in this published article.

Ethical approval and informed consent. We declare that this paper does not report any data collected from humans or animals.

References
1. Yamamoto, H., Moriyama, K., Jinnouchi, H. & Yagishita, K. Studies on terreic acid. Jpn. J. Antibiot. 33, 320–328 (1980).
2. Kawakami, Y. et al. Terreic acid, a quinone epoxide inhibitor of Bruton's tyrosine kinase. Proc. Natl. Acad. Sci. USA 96, 2227–32 (1999).
3. Byrd, J. C. et al. Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. Engl. J. Med. 369, 507–516 (2013).
4. Reed, G. & Dining, L. C. The biosynthesis of terreic acid. Chem. Commun. 16, 935–937 (1968).
5. Reed, G., Westlake, D. W. & Dining, L. C. Quinone epoxides. V1. The biosynthesis of terreic acid. Can. J. Biochem. 47, 1071–1079 (1969).
6. Boruta, T. & Rizkoje, M. J. Culture-based and sequence-based insights into biosynthesis of secondary metabolites by Aspergillus terreus ATCC 20542. J. Biotecnol. 75, 53–62 (2014).
7. Guo, C. J., Sun, W. W., Bruno, K. S. & Wang, C. C. Molecular genetic characterization of terreic acid pathway in Aspergillus terreus. Org. Lett. 16, 5250–5253 (2014).
8. Fujii, I. et al. Cloning of the polyketide synthase gene atX from Aspergillus terreus and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Mol. Gen. Genet. 253, 1–10 (1996).
9. Puel, O., Galtier, P. & Oswald, I. P. Biosynthesis and toxicological effects of patulin. Toxins 2, 613–631 (2010).
10. Gao, L. M. et al. Engineered fungal polyketide biosynthesis in Pichia pastoris: a potential excellent host for polyketide production. Microb. Cell Fact. 12, 77 (2013).
11. Huang, C., Ghavtadze, N., Chattopadhyay, B. & Gevorgyan, V. Synthesis of catechols from phenols via Pd-catalyzed silanol-directed C=O hydrogenation. J. Am. Chem. Soc. 135, 17630–17633 (2011).
12. Melha, J., Caufield, J. H. & Uetz, P. The yeast two-hybrid system: a tool for mapping protein–protein interactions. Cold Spring Harb. Protoc. 425–430 (2015).
13. Takemori, S., Hon-Nami, K., Kawahara, F. & Katagiri, M. Mechanism of the salicyl-1-monoxygenase reaction. VI. The monomeric nature of the enzyme. Biochem. Biophys. Acta 342, 137–144 (1974).
14. Peng, P., Shang, Y., Cen, K. & Wang, C. Fungal biosynthesis of the benzenoquinone oosporein to evade insect immunity. Proc. Natl. Acad. Sci. USA 112, 11365–11370 (2015).
15. Graminia, M. A., Rocha, E. M., Prade, R. A. & Martinez-Rossi, N. M. Terbinafine resistance mediated by salicylate 1-monooxygenase in Aspergillus nidulans. Antimicrob. Agents Chem. 48, 3530–3535 (2004).
16. Davison, J. et al. Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. Proc. Natl. Acad. Sci. USA 109, 7642–7647 (2012).
17. Fadah, A. A. et al. Oxidative deamortisation: the key step of sorbicilloidin biosynthesis. Chem. Sci. 5, 523–527 (2014).
18. Biju, B., Suh, J. W., Park, S. H. & Kwon, H. J. Delineating Monascus azaphilone pigment biosynthesis: oxidoertuctive modifications determine the ring cyclization pattern in azaphilone biosynthesis. RSC Adv. 4, 39405–39408 (2014).
19. He, Y. & Cox, R. J. The molecular steps of citrinin biosynthesis in fungi. Chem. Sci. 7, 2119–2127 (2016).
20. Sinini, S. P. et al. The gene PatG involved in the biosynthesis pathway of patulin, a food-borne mycotoxin, encodes a 6-methylsalicylic acid decarboxylase. Int. J. Food Microbiol. 171, 77–83 (2014).
21. Artigot, M. P. et al. Molecular cloning and functional characterization of two CYP619 cytochrome P450s involved in biosynthesis of patulin in Aspergillus clavatus. Microbiology 155, 1753–1766 (2009).
22. Cheeseman, K. et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun. 5, 2876 (2014).
23. Favre, B. & Ryder, N. S. Characterization of squaene epoxide activity from the dermatomycete Trichophyton rubrum and its inhibition by terbinafine and other antifungal agents. Antimicrob. Agents Chem. 40, 443–7 (1996).
24. Minami, A. et al. Sequential enzymatic epoxidation involved in polyether lasalocid biosynthesis[1]. J. Am. Chem. Soc. 134, 7246–7249 (2012).
25. Li, S. et al. Substrate recognition by the multifunctional cytochrome P450 MycG in mycinamicin hydroxylation and epoxidation reactions. J. Biol. Chem. 287, 37880–37890 (2012).
26. Kudo, F., Motegi, A., Mizoue, K. & Eguchi, T. Cloning and characterization of the biosynthetic gene cluster of 16-membered macrolide antibiotic FD-891: involvement of a dual functional cytochrome P450 monooxygenase catalyzing epoxidation and hydroxylation. ChemBioChem 11, 1574–1582 (2010).
27. Silva, J. C. et al. Isolation and characterization of the versicolorin B synthase gene from Aspergillus parasiticus. Expansion of the aflatoxin b1 biosynthetic gene cluster. J. Biol. Chem. 271, 13600–13608 (1996).
28. McGuire, S. M., Silva, J. C., Casillas, E. G. & Townsend, C. A. Cultivation strategies to enhance productivity of Pichia pastoris. J. Biotechnol. 98, 5301–5317 (2014).
29. Gasser, R. et al. Pichia pastoris: protein production host and model organism for biomedical research. Future Microbiol. 8, 191–208 (2013).
30. Vogl, T. & Glieder, A. Regulation of Pichia pastoris promoters and its consequences for protein production. New Biotechnol. 30, 385–404 (2013).
31. Looser, V. et al. Cultivation strategies to enhance productivity of Pichia pastoris: A review. Biotechnol. Adv. 33, 1177–1193 (2015).
32. Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast 15, 1541–1553 (1999).
33. Crege, J. M. Methods in Molecular Biology. Vol. 389. Pichia Protocols. 2nd ed. Totowa NJ: Humana Press (2007).
34. Wang, X. L. et al. Mtt transcription factor mediates methanol signaling and regulates the alcohol oxidase 1 (AOXI) promoter in Pichia pastoris. J. Biol. Chem. 291, 6245–6261 (2016).
Acknowledgements
This work was supported by Fundamental Research Funds for the Shanghai Science and Technology Innovation Action Plan (17JC1402400) and Talent Program of School of Biotechnology in East China University of Science and Technology. We thank Dr. Faliang An, East China University of Science and Technology, for the assistance in NMR analysis. The authors declare that they have no conflict of interest. We also gratefully acknowledge Prof. Hong Lu, School of Life Sciences, Fudan University, for kindly providing the yeast two-hybrid system.

Author Contributions
M.C. conceived the project and supervised the research. C.K. and M.C. are responsible for project planning and experimental design. C.K. and H.H. performed most of the experiments. Y.X., Q.X. and Y.L. participated in strain construction and product analysis. Q.P., Q.L., Q.X. and Y.L. participated in protein expression analysis. Q.L. and Q.Z. performed the yeast two-hybrid assay. M.C., C.K., H.H. and Y.Y. analyzed the results. C.K. wrote the manuscript and M.C. revised and accomplished the manuscript finally. X.S. and Y.Z. reviewed the manuscript. All authors have read and approved the final manuscript and contributed to scientific discussion.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-20514-x.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018