Epigenetics Hot Paper

Deutsche Ausgabe: DOI: 10.1002/ange.201502452
Internationale Ausgabe: DOI: 10.1002/anie.201502452

# Epigenetic Genome Mining of an Endophytic Fungus Leads to the Pleiotropic Biosynthesis of Natural Products**

*Xu-Ming Mao, Wei Xu, Dehai Li, Wen-Bing Yin, Yit-Heng Chooi, Yong-Quan Li, Yi Tang,* and Youcai Hu**

***Abstract:*** *The small-molecule biosynthetic potential of most filamentous fungi has remained largely unexplored and represents an attractive source for the discovery of new compounds. Genome sequencing of Calcarisporium arbuscula, a mushroom-endophytic fungus, revealed 68 core genes that are involved in natural product biosynthesis. This is in sharp contrast to the predominant production of the ATPase inhibitors aurovertin B and D in the wild-type fungus. Inactivation of a histone H3 deacetylase led to pleiotropic activation and overexpression of more than 75% of the biosynthetic genes. Sampling of the overproduced compounds led to the isolation of ten compounds of which four contained new structures, including the cyclic peptides arbumycin and arbumelin, the diterpenoid arbuscullic acid A, and the meroterpenoid arbuscullic acid B. Such epigenetic modifications therefore provide a rapid and global approach to mine the chemical diversity of endophytic fungi.*

Filamentous fungi are prolific producers of bioactive natural products,[1] exemplified by the antibiotic penicillin[2] and the anti-hypercholesterolemia drug lovastatin.[3] Recent genome-sequencing efforts for many fungal species have revealed significant biosynthetic potential, as represented by a large number of cryptic and diverse biosynthetic pathways.[4] Synthetic biological efforts to activate these silent pathways in well studied fungal species, mostly in the *Penicillium* and *Aspergillus* genera, have led to the discovery of new natural products.[5] In particular, epigenetic approaches that lead to chromatin remodeling have resulted in the activation of individual gene clusters.[6]

Endophytic fungi are increasingly recognized as significant underachievers in natural product biosynthesis.[7] For example, recent genome sequencing of the entomophytic fungus *Metarhizium anisopliae* and the endophytic fungus *Beauveria bassiana* revealed that each one encodes a large number of potential secondary-metabolite gene clusters (85 and 45, respectively), significantly more than the number of compounds that have been reported for each strain.[8] Such a high biosynthetic potential is a reflection of the complex natural ecological environment,[9] which is difficult to replicate in the laboratory, resulting in most gene clusters being silent in axenic cultures. Therefore, genetically modifying endophytic fungi to activate, ideally globally, a plethora of natural-product pathways can be particularly fruitful in accessing new chemical scaffolds.

*Calcarisporium arbuscula* is an endophytic fungus living in the fruiting bodies of mushrooms.[10] Fungi in the *Calcarisporium* genus have been noted for the production of a few antimicrobial compounds and mycotoxins.[11] *C. arbuscula* predominantly produces the F1-ATPase inhibitors **1** and **2**, which are polyketides containing an unusual 2,6-dioxabicyclo-[3.2.1]octane core that is connected to an α-pyrone through a triene linker (Scheme 1).[12] Bioinformatic analysis of the draft genome sequence of *C. arbuscula* showed 68 gene clusters encoding potential natural-product biosynthesis pathways, among which 41 contain polyketide synthases (PKSs, including 4 PKS/nonribosomal peptide synthetase (NRPS) hybrids; 18 contain NRPSs, and 9 contain terpene synthases (TSs); see the Supporting Information, Tables S1–S3). This large collection of biosynthetic gene clusters is in sharp

[*] Dr. X. M. Mao,[+] Dr. W. Xu, Dr. D. Li, Dr. W. B. Yin,[$] Dr. Y. H. Chooi,[$$] Dr. Y. Tang, Dr. Y. Hu[+]
Department of Chemical and Biomolecular Engineering and Department of Chemistry and Biochemistry
University of California
Los Angeles, CA 90095 (USA)
E-mail: yitang@ucla.edu

Dr. D. Li
Key Laboratory of Marine Drugs, Chinese Ministry of Education
School of Medicine and Pharmacy, Ocean University of China
Qingdao, 266003 (China)

Dr. X. M. Mao,[+] Dr. Y. Q. Li
College of Life Sciences, Zhejiang University
Hangzhou 310058 (China)

Dr. Y. Hu[+]
State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College
Beijing 100050 (China)
E-mail: huyoucai@imm.ac.cn

[$] Current address: State Key Laboratory of Mycology
The Institute of Microbiology, Chinese Academy of Sciences
Beijing 100101 (China)

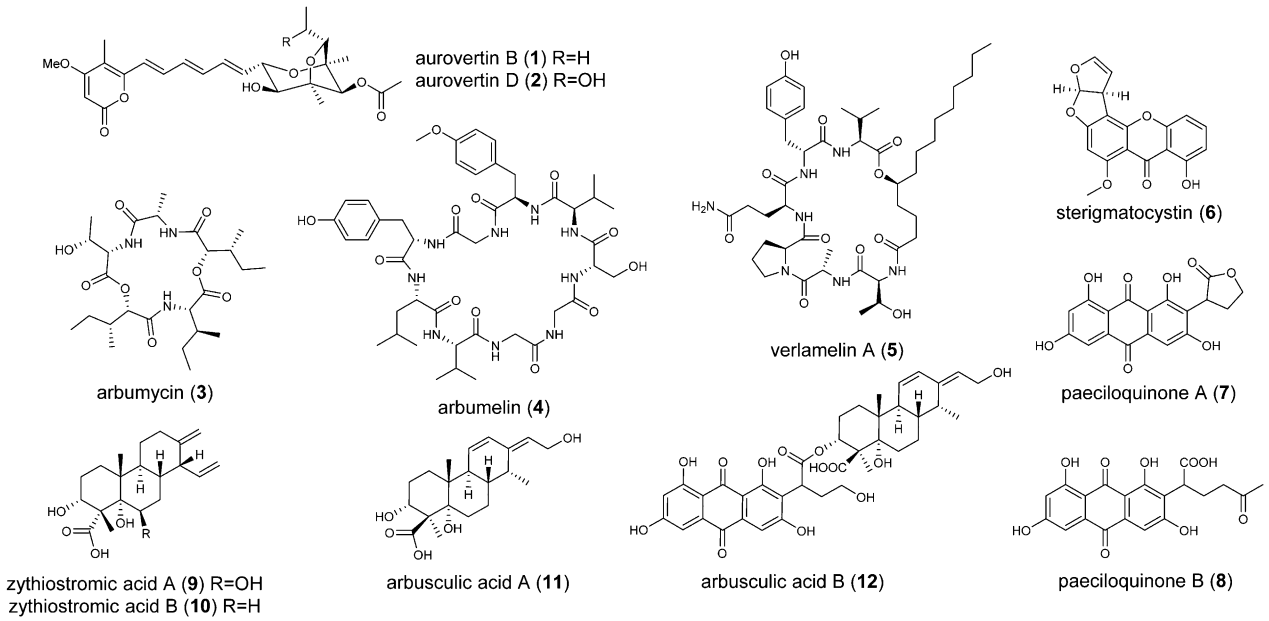
[$$] Current address: Research School of Biology
Australian National University
Canberra, ACT 0200 (Australia)

[+] These authors contributed equally to this work.

[**] This work was supported by the National Institute of Health (1DP1GM106413). Y.H. was supported by the PUMC Youth Fund (33320140175) and the State Key Laboratory Fund for Excellent Young Scientists (GTZB201401). X.M.M. was funded by a fellowship of the New Star Project from Zhejiang University. We thank Prof. David E. Cane for helpful discussions.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502452.




***Scheme 1.*** Compounds purified from *C. arbuscula*. Compounds **1** and **2** were purified from the wild type and **3**–**12** from the *ΔhdaA* mutant.

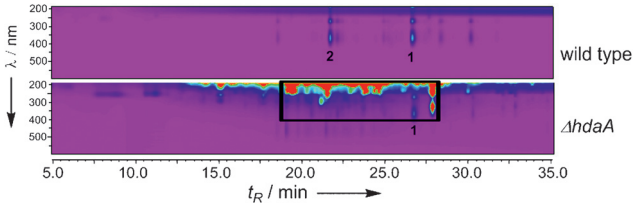



***Figure 1.*** HPLC contour plot of extracts from the wild-type and *ΔhdaA* *C. arbuscula* strains. The boxed region contains compounds **3**–**12**, which were isolated in this work.

contrast to the two predominant metabolites **1** and **2** (Figure 1).[12a] This organism therefore represents a prime resource for genome mining using globally effective approaches. However, initial attempts to culture the fungus on different media, such as MEPA, CYA, YMEG or YG, did not lead to the production of new compounds. The nongenetic strategy of adding the DNA methyltransferase inhibitor 5-azacytidine or the histone deacetylase (HDAC) inhibitor suberoyl bis(hydroxamic acid)[13] also did not change the metabolomic profile.

In filamentous fungi, many silenced gene clusters are located within the heterochromatic regions and subsequently transcriptionally repressed.[14] HDACs remove acetyl groups from the amino tails of histones and maintain the chromatin in an inaccessible state for the transcriptional machinery.[15] Keller and co-workers showed that fungal HDACs negatively regulate the production of sterigmatocystin and penicillin in *A. nidulans,* and attenuate transcription of NRPS gene clusters in *A. fumigatus.*[16] To investigate the roles of HDACs in *C. arbuscula*, we deleted *hdaA* (Figure S3), which encodes the histone H3 lysine 14 (K14) deacetylase. Removal of HdaA resulted in slower growth, shorter mycelia, and defective sporulation in *C. arbuscula* (Figure S4). Metabolite extraction and LC/MS analysis revealed the production of significantly more compounds compared to the wild type (Figure 1). Comprehensive RT-PCR analysis of all 68 core biosynthetic genes showed that whereas weak to no expression of most core genes was observed in the wild type, deletion of *hdaA* led to increased expression of 75% (31 out of 41) of the PKS genes, 78% (14 of 18) of the NRPS genes, and 78% (7 of 9) of the TS genes (Figure S5). Hence, *C. arbuscula* HdaA globally suppresses biosynthetic genes under axenic growth conditions, and its deletion leads to the pleiotropic activation of secondary metabolism.

To characterize the newly produced compounds in the *ΔhdaA* strain, the organic extracts were partitioned with *n*-hexane and methanol. Then, the methanol phase was fractionated by reverse-phase chromatography. Four fractions that were particularly rich in metabolites were selected for further purification, which led to the isolation of ten compounds, including three peptides (**3**, **4**, and **5**), three polyketides (**6**, **7**, and **8**), three diterpenes (**9**, **10**, and **11**), and one polyketide–diterpene hybrid (**12**). Among them, **3**, **4**, **11**, and **12** are new compounds (Scheme 1).

Two new cyclic peptides, namely arbumycin (**3**; 10 $mg\,L^{-1}$) and arbumelin (**4**; 0.17 $mg\,L^{-1}$), and the lipopeptide verlamelin A (**5**; 0.25 $mg\,L^{-1}$) were isolated from the *ΔhdaA* mutant. When selective ion monitoring was performed, a ten-fold decrease in the level of **3** could be detected in the wild type, and no trace of **4** or **5** could be found (Figure 3a). The structure of **3** was elucidated as a cyclic pentadepsipeptide by NMR spectroscopy (Table S11, Figures S13–S18) and X-ray crystallography (Figure S6). The asymmetric peptide is derived from three amino acids (L-Ala, L-Thr, L-Ile) and two molecules of the same hydroxy acid, (2*S*,3*R*)-2-hydroxy-3-methylpentanoic acid (**13**; Figure 2c). Most fungal depsipeptides characterized to date are symmetric and are synthesized by the iterative action of a two-module NRPS depsipeptide synthase.[17] Therefore, the asymmetric nature of

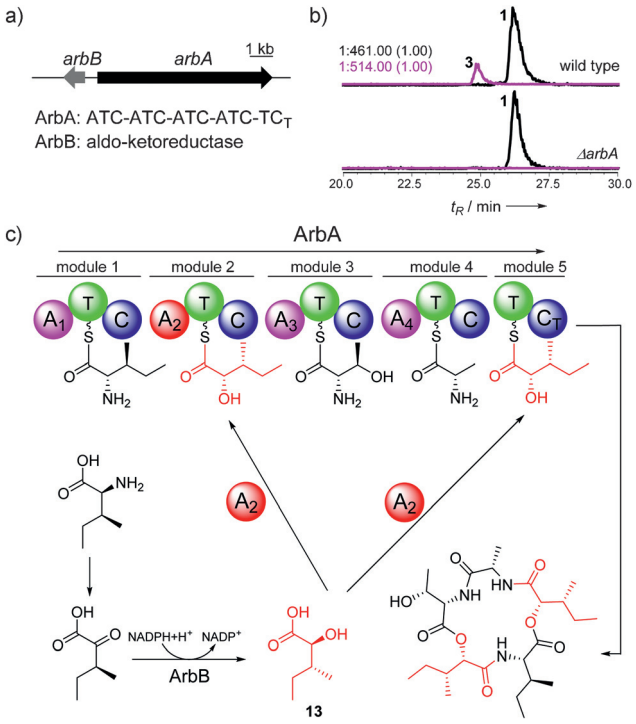



***Figure 2.*** Characterization of arbumycin (**3**) and its gene cluster in *C. arbuscula*. a) The putative gene cluster for the biosynthesis of **3**. b) Genetic verification of the gene cluster responsible for the biosynthesis of **3**. c) Proposed pathway for the biosynthesis of **3**.

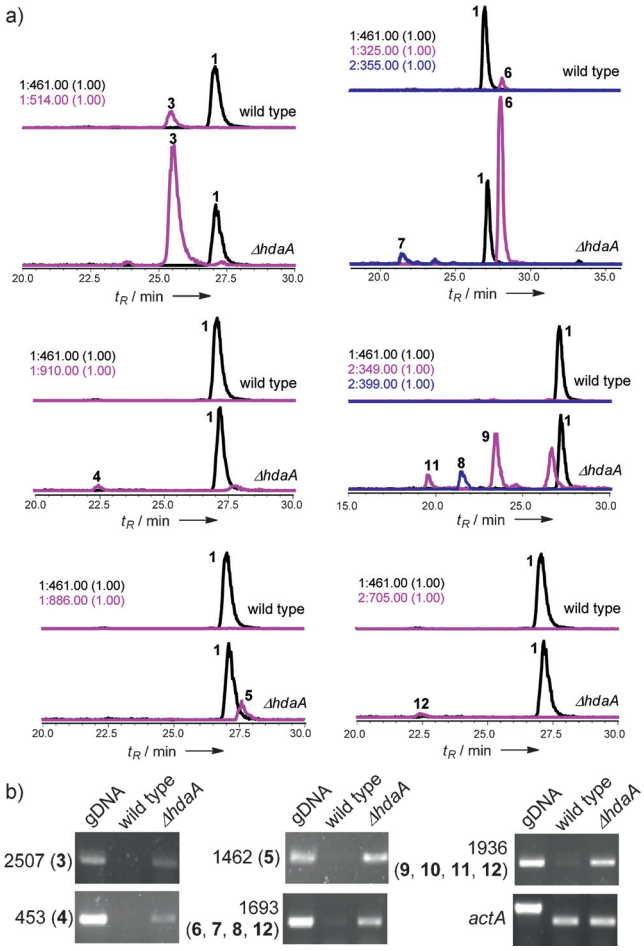



***Figure 3.*** Overproduction of compounds in the *ΔhdaA* mutant. a) Compounds **3**–**9**, **11**, and **12** are formed in higher yields in the *ΔhdaA* mutant based on EIC than compound **1**. b) RT-PCR analysis of the gene expression of selected core genes. The housekeeping *actA* gene served as the internal control. The contig numbers (with corresponding compounds in parentheses) are also shown.

**3** presents an interesting case of structure variation in this family. The 2*S*,3*R* stereochemistry in **13** was an unexpected finding from the crystal structure, and indicates possible epimerization of the 3*R* methyl group from (2*S*,3*S*)-isoleucine.

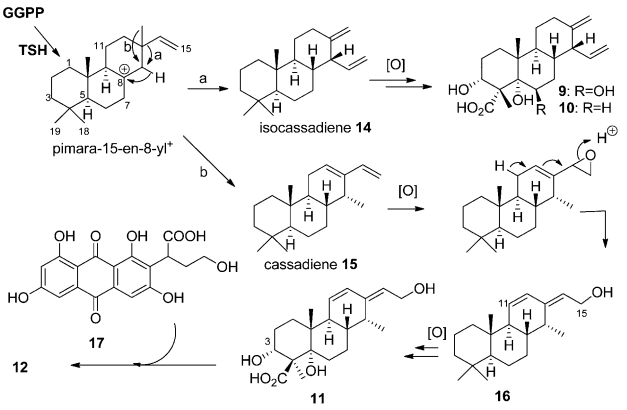
To understand the biosynthesis of **3**, we identified a potential cluster on scaffold 82 that consists of a five-module NRPS (ArbA) and an aldo-ketoreductase (ArbB) (Figure 2a). The transcription level of *arbA* is increased in the *ΔhdaA* strain (Figure 3b). Whereas ArbA contains a C-terminal condensation domain ($C_T$) consistent with formation of a cyclic product,[18] the last module (module 5) did not contain an adenylation (A) domain. This indicates that module 5 may use the same substrate (**13** is used twice in **3**) as a previous module, and that an A domain may be shared between the two modules. Sequence analysis of the ten amino acid specificity codes of the A domains revealed that A2 should be responsible for the incorporation of **13**, as the highly conserved Asp235, which anchors the amino groups of amino acids, was replaced by Ala235 whereas a threonine (Thr330) conserved in hydroxy acid activating A domains was present (Table S4).[17] Indeed, inactivation of *arbA* in wild-type *C. arbuscula* (Figure S7) completely suppressed the production of **3** (Figure 2b). The proposed biosynthesis of **3** from ArbA and ArbB is shown in Figure 2c. We suggest that ArbB may be responsible for the ketoreduction and epimerization of 2-keto-(3*S*)-pentanoic acid to yield **13**. This is highly reminiscent of the epimerizing ketoreductase (KR) domains found in bacterial type I PKSs, in which the KR domains are able to control the stereochemistry at vicinal α-methyl- and β-hydroxy-substituted carbon atoms starting from α-methyl-β-ketone substrates.[19]

Arbumelin (**4**) is a new cyclic nonapeptide containing three Gly and two D-amino acids as determined by NMR spectroscopy (Table S12 and S13, Figures S19–S27). The stereochemical configurations of the constituent amino acids were determined through hydrolysis, derivatization, and HPLC analysis (see the Supporting Information). One nine-module NRPS on contig 453 is proposed to be the candidate for the biosynthesis of **4**. Sequence analysis showed that the ten amino acid specificity code of the A domains in module 3, 7, and 8 are nearly identical and consistent with the three Gly positions in **4** (Table S5). The E domains in modules 4 and 5 should be responsible for the epimerization of the consecutive D-*O*-methyl-Tyr and D-Val units. Taken together, the linear sequence of **4** is likely to be L-Leu-L-Tyr-Gly-D-*O*-methyl-Tyr-D-Val-L-Ser-Gly-Gly-L-Val, which is macrocyclized by the terminal $C_T$ domain in the NRPS (Scheme S1).[18] Immediately adjacent to the NRPS gene is

a gene encoding a methyltransferase (Figure S8). This enzyme may be responsible for formation of the *O*-methyl-Tyr residue prior to activation by module 4. The structure of compound **5** was resolved by extensive NMR spectroscopy (Figure S27–S31) and is the same as that of the recently reported cyclic hexalipopeptide verlamelin, an antifungal compound isolated from *Lecanicillium* sp. HF627.[20] The corresponding gene cluster is located on contig 1462 with the same set of genes (such as NRPS, fatty acyl ligase) as reported for verlamelin (Figure S8).[20] The NRPS-encoding genes on contig 453 and 1462 were verified to be overexpressed in the *ΔhdaA* mutant (Figure 3b).

Deletion of *hdaA* led to a twenty-fold increase in the level of the mycotoxin sterigmatocystin (**6**; Figure 3a). The activated gene cluster on contig 1693 was readily identified based on the >70% sequence identity to that from *Aspergillus ochraceoroseus* (Figure S1).[21] Overexpression of this pathway (Figure 3b) also led to the isolation of the protein tyrosine kinase inhibitors paeciloquinone A (**7**) and B (**8**; Figures S33 and S34), which are shunt products of the biosynthetic pathway of **6** (Scheme S2).[22] Bioinformatic analysis of the remaining PKS gene clusters also revealed the potential to produce other mycotoxins, such as citrinin (contig 921; Figure S2).[23]

The *ΔhdaA* strain produced three labdane-related diterpenoid compounds (**9**–**11**) as characterized by NMR spectroscopy (Table S14, Figures S35–S44). Upon structure elucidation, **9** (0.67 mg L$^{-1}$) and **10** (0.23 mg L$^{-1}$) were confirmed to be the isocassadienes zythiostromic acid A and B, respectively.[24] Both compounds are heavily oxidized derivatives of isocassadiene **14**.[25] The new diterpene arbusculic acid A (**11**) is structurally related to **9** and **10**, and most likely derives from cassadiene **15** (Scheme 2). Compound **11** contains a dienol functional group that has not been observed in fungal diterpenoids. Because of their structural similarities, we propose that compounds **9**–**11** are synthesized from the same diterpene biosynthetic gene cluster. BLAST search against pimaradiene synthase from *A. nidulans* (AN1594)[26] and *ent*-kaurene synthase from the fungus *Phaeosphaeria* sp. L487[27] revealed a diterpene synthase with the highest identity on contig 1936 (Figure S9a). Based on structural modeling against the taxadiene synthase from *Taxus brevifolia*,[28] this *C. arbuscula* diterpene cyclase is organized in three α-helix domains, including two class I and one class II domain (Figure S10b,c).[28] This gene was highly transcribed in the mutant strain compared to the wild type (Figure 3b). Two adjacent P450 enzymes are also found in the gene cluster with sequence similarity to *ent*-kaurene oxidases (Figure S10).[27] The proposed mechanism of formation of **11** is shown in Scheme 2, in which **15** is first epoxidized at the terminal olefin followed by acid-catalyzed ring-opening hydrolysis with transposition of the diene to yield **16**. Both **15** and **16** can be subjected to multiple hydroxylations at the C3 and C5 positions as well as six-electron oxidation at C19 to yield **9** and **11**, respectively. Intriguingly, we also isolated the meroterpenoid compound arbuscullic acid B (**12**; 0.2 mg L$^{-1}$) from the *ΔhdaA* mutant (Table S15, Figures S45–S49). Compound **12** is derived from the esterification of **11** to the carboxylic acid moiety of paeciloquinone D (**17**),[22a] which is




***Scheme 2.*** Proposed pathways for the biosynthesis of **9**–**12**.

the free acid form of lactone **7**, through the C3 hydroxy group. The regioselectivity of the ester linkage was confirmed by extensive 2D NMR spectroscopy (Table S15), whereas the alternative esterification regioselectivity between **11** and **17** was not detected. This suggests that the coupling of **11** and **17** may be enzyme-catalyzed rather than result from a nonenzymatic reaction.

In summary, the deletion of a single HDAC gene in the endophytic fungus *C. arbuscula* led to the global alteration of secondary metabolism and the pleiotropic production of new natural products of different chemotypes and biosynthetic origins. Continued chemical isolation from this strain, along with additional regulatory engineering efforts, can lead to the enhanced realization of the biosynthetic potential of this and other endophytic fungi.

**Keywords:** biosynthesis · endophytic fungi · epigenetics · gene expression · natural products

**How to cite:** *Angew. Chem. Int. Ed.* **2015**, *54*, 7592–7596
*Angew. Chem.* **2015**, *127*, 7702–7706

[1] A. Schueffler, T. Anke, *Nat. Prod. Rep.* **2014**, *31*, 1425–1448.
[2] G. Ozcengiz, A. L. Demain, *Biotechnol. Adv.* **2013**, *31*, 287–311.
[3] K. C. Mulder, F. Mulinari, O. L. Franco, M. S. Soares, B. S. Magalhaes, N. S. Parachin, *Biotechnol. Adv.* **2015**, DOI: 10.1016/j.biotechadv.2015.04.001.
[4] Y. M. Chiang, K. H. Lee, J. F. Sanchez, N. P. Keller, C. C. Wang, *Nat. Prod. Commun.* **2009**, *4*, 1505–1510.
[5] a) A. A. Brakhage, V. Schroeckh, *Fungal Genet. Biol.* **2011**, *48*, 15–22; b) A. O. Zabala, W. Xu, Y. H. Chooi, Y. Tang, *Chem. Biol.* **2012**, *19*, 1049–1059.
[6] R. H. Cichewicz, *Nat. Prod. Rep.* **2010**, *27*, 11–22.
[7] S. Chandra, *Appl. Microbiol. Biotechnol.* **2012**, *95*, 47–59.
[8] D. M. Gibson, B. G. Donzelli, S. B. Krasnoff, N. O. Keyhani, *Nat. Prod. Rep.* **2014**, *31*, 1287–1305.
[9] S. Kusari, C. Hertweck, M. Spiteller, *Chem. Biol.* **2012**, *19*, 792–798.
[10] P. Watson, *Trans. Br. Mycol. Soc.* **1955**, *38*, 409–414.
[11] a) J. Silber, B. Ohlendorf, A. Labes, A. Erhard, J. F. Imhoff, *Mar. Drugs* **2013**, *11*, 3309–3323; b) J. Silber, B. Ohlendorf, A. Labes, C. Nather, J. F. Imhoff, *J. Nat. Prod.* **2013**, *76*, 1461–1467; c) P. Chrisp, P. M. Dewick, F. T. Boyle, *Z. Naturforsch. C* **1990**, *45*, 179–186; d) N. J. Yu, S. X. Guo, H. Y. Lu, *J. Asian Nat. Prod. Res.* **2002**, *4*, 179–183.

[12] a) M. D. Osselton, H. Baum, R. B. Beechey, *Biochem. Soc. Trans.* **1974**, *2*, 200 – 202; b) M. J. van Raaij, J. P. Abrahams, A. G. Leslie, J. E. Walker, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6913 – 6917.
[13] A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *J. Am. Chem. Soc.* **2011**, *133*, 10990 – 10998.
[14] a) J. M. Palmer, N. P. Keller, *Curr. Opin. Microbiol.* **2010**, *13*, 431 – 436; b) A. Gacek, J. Strauss, *Appl. Microbiol. Biotechnol.* **2012**, *95*, 1389 – 1404.
[15] M. Bulger, *J. Biol. Chem.* **2005**, *280*, 21689 – 21692.
[16] a) E. K. Shwab, J. W. Bok, M. Tribus, J. Galehr, S. Graessle, N. P. Keller, *Eukaryotic Cell* **2007**, *6*, 1656 – 1664; b) I. Lee, J. H. Oh, E. K. Shwab, T. R. Dagenais, D. Andes, N. P. Keller, *Fungal Genet. Biol.* **2009**, *46*, 782 – 790.
[17] R. Süssmuth, J. Müller, H. von Döhren, I. Molnár, *Nat. Prod. Rep.* **2011**, *28*, 99 – 124.
[18] X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh, Y. Tang, *Nat. Chem. Biol.* **2012**, *8*, 823 – 830.
[19] a) C. R. Valenzano, R. J. Lawson, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2009**, *131*, 18501 – 18511; b) Y. O. You, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2013**, *135*, 7406 – 7409; c) A. Garg, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2013**, *135*, 16324 – 16327.
[20] K. Ishidoh, H. Kinoshita, T. Nihira, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7501 – 7510.
[21] J. W. Cary, K. C. Ehrlich, S. B. Beltz, P. Harris-Coward, M. A. Klich, *Mycologia* **2009**, *101*, 352 – 362.
[22] a) F. Petersen, A. Fredenhagen, H. Mett, N. B. Lydon, R. Delmendo, H. B. Jenny, H. H. Peter, *J. Antibiot.* **1995**, *48*, 191 – 198; b) R. E. Minto, C. A. Townsend, *Chem. Rev.* **1997**, *97*, 2537 – 2556; c) C. A. Townsend, *Nat. Prod. Rep.* **2014**, *31*, 1260 – 1265.
[23] Y. P. Li, Y. Xu, Z. B. Huang, *Biotechnol. Lett.* **2012**, *34*, 131 – 136.
[24] W. A. Ayer, A. Q. Khan, *Phytochemistry* **1996**, *42*, 1647 – 1652.
[25] R. J. Peters, *Nat. Prod. Rep.* **2010**, *27*, 1521 – 1530.
[26] K. Bromann, M. Toivari, K. Viljanen, A. Vuoristo, L. Ruohonen, T. Nakari-Setala, *PloS One* **2012**, *7*, e35450.
[27] H. Kawaide, R. Imai, T. Sassa, Y. Kamiya, *J. Biol. Chem.* **1997**, *272*, 21706 – 21712.
[28] a) M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, *Nature* **2011**, *469*, 116 – 120; b) M. Köksal, H. Hu, R. M. Coates, R. J. Peters, D. W. Christianson, *Nat. Chem. Biol.* **2011**, *7*, 431 – 433.

Received: March 16, 2015
Published online: May 26, 2015