Organic LETTERS

Letter

pubs.acs.org/OrgLett

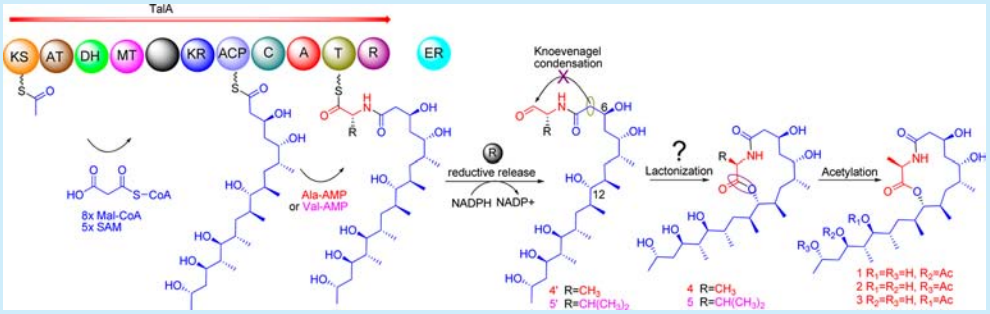
Characterization of Thermolide Biosynthetic Genes and a New Thermolide from Sister Thermophilic Fungi

Xuemei Niu,*[,†,‡] Li Chen,[†,§] Qun Yue,[†,§] Baile Wang,[‡,§] Junxian Zhang, Chunyan Zhu,[‡] Keqin Zhang,[‡] Gerald F. Bills,[†] and Zhiqiang An*[,†]

[†]Texas Therapeutics Institute, the Brown Foundation Institute of Molecular Medicine, the University of Texas Health Science Center at Houston, Houston, Texas 77054, United States

[‡]Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, 650091, P. R. China

Supporting Information



ABSTRACT: Prior chemical analysis of obligate thermophilic fungus *Talaromyces thermophilus* led to the discovery of thermolides A−F, six previously undescribed members of the lactam-bearing macrolactone class. A combination of chemical screening, genome analyses, and genetic manipulation led to the identification of the thermolide biosynthetic genes from sister thermophilic fungi *T. thermophilus* and *Thermomyces lanuginosus* and a new thermolide. The biosynthetic locus for the thermolides' mixed polyketide/amino acid structure encodes a hybrid polyketide synthase−nonribosomal peptide synthetase (PKS−NRPS). Our results reveal the first fungal hybrid iterative PKS−NRPS genes involved in the biosynthesis of bacterial-like hybrid macrolactones instead of typical fungal tetramic acids-containing metabolites. The finding provides an insight into the convergent biosynthetic end products that bridge the gap between the modular and iterative PKS−NRPS hybrids.

Macrolactones of mixed origins with polyketide and amino acid building blocks, representing a unique family of natural products endowed with a wide range of impressive biological activities, are biosynthesized mainly by bacteria, e.g., the immunosuppressant rapamycin, the tubulin polymerizing agent epothilone, the pristinamycin components of the antibiotic synercid, the ansa antibiotics including the antitubercular drugs of the rifamycin group, the heat shock protein 90 inhibitor geldanamycin, and the maytansine family antitumor agent ansamitocin.[1] These macrocyclic metabolites are assembled by multifunctional enzyme complexes known as modular polyketide synthases combined with nonribosomal peptide synthetases (PKS−NRPSs) in bacteria. Macrocyclization leading to lactones, e.g., 14-membered in erythronolide and 16-membered in epothilone, occurs at the most downstream steps of modular PKS−NRPS hybrid assembly lines during chain release by thioesterase (TE) domains.[1,2]

In fungi, such macrolactonization occurs in either iterative PKSs or NRPSs,[2] for example, the mycotoxin zearalenone from *Gibberella zeae*,[3a] the protein kinases inhibitor hypothemycin from *Hypomyces subiculosus*[3b] and the heat shock protein 90 inhibitor dehydrocurvularin from *Aspergillus terreus*,[3c] and the emericellamides from *A. nidulans*.[3d]

However, fungal iterative PKS−NRPS hybrids, where a single module of an iterative PKS is fused with a single NRPS module, usually lead to formation of metabolites containing a five-membered ring, known as tetramic acid (2-pyrrolidones). At least 13 fungal iterative PKS−NRPS hybrid gene clusters reported to date have been linked to the biosynthesis of tetramic acid-containing metabolites, including the actin-filament-destroying cytochalasins, the neuritogenic and immunosuppressive agent pseurotin A, and the mycotoxins fusarin C, cyclopiazonic acid, and flavipucine.[4] The formation of tetramic acids (2-pyrrolidones) normally occurs either directly at the most downstream step of fungal iterative PKS−NRPS hybrid assembly lines through a Dieckmann cyclization by terminal domain (D)[5] or after a reductive release by reductive domain (R), thereby generating a free linear aldehyde intermediate that immediately undergo a Knoevenagel condensation.[3] Recent studies revealed that hybrid iterative PKS−NRPS genes in the bacterium

Received: June 3, 2014
Published: July 7, 2014

dx.doi.org/10.1021/ol501595z | *Org. Lett.* 2014, 16, 3744−3747

Lysobacter enzymogene were also responsible for the biosynthesis of tetramic acid-containing macrolactams HSAF[6a,b] and Ikarugamycin.[6c,d]

To date, the linkage of iterative PKS−NRPS hybrid genes to PKS−NRPS hybrid macrolactones has not been reported.

Thermophilic fungi, growing optimally above 40 °C, represent a potential reservoir of thermostable enzymes for industrial applications and could be developed into cell factories to support the production of chemicals and materials at elevated temperatures.[7] Recently, a novel class of potent nematocides possessing a 13-membered lactam-bearing macrolactone (**1**−**6**, Figure 1)

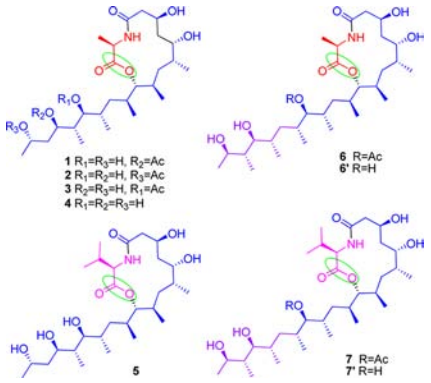


Figure 1. Structures of thermolides.

were isolated from obligate thermophilic fungus *Talaromyces thermophilus* collected from the Tengchong geothermal field of Yunnan, China.[8] Metabolites **1** and **2** exhibited potent toxicity toward a wide range of nematodes including the economically destructive root-knot nematode (*Meloidogyne incognita*) and pine-wood nematode (*Bursaphelenches siylopilus*).

Here we report that two hybrid iterative PKS−NRPS genes from sister thermophilic fungi and one new metabolite of this class were indentified through a combination of chemical screening, genome analyses, and genetic manipulation. Our results indicated these fungal hybrid iterative PKS−NRPS genes were involved in the biosynthesis of bacterial-like hybrid macrolactones instead of typical fungal tetramic acids-containing metabolites.

The genome of the type strain of *T. thermophilus* NRRL 2155 from California was recently released.[8] The strain belongs to the same specie as thermolide-producing strain *T. thermophilus* YM3−4; therefore, we predicted that *T. thermophilus* NRRL 2155 would likely produce similar thermolide-like metabolites as YM3−4. HPLC−MS analysis of organic-solvent extracts of *T. thermophilus* NRRL 2155 and YM3−4 strains grown in parallel indicated that strain NRRL 2155 also yielded thermolides A−E (**1**−**5**) and had remarkably similar chemical profiles to that of the strain YM3−4. Unexpectedly, we could not detect thermolide F (**6**) in either strain NRRL 2155 or YM3−4. We sought to examine different culture conditions and additional strains of *T. thermophilus* to find either strains with higher levels of production or strains with different product spectra among the principal thermolides. However, we could not detect thermolide F (**6**) in either *T. thermophilus* strain.

We examined the genome of *T. thermophilus* NRRL 2155 for secondary metabolite pathways.[9] Surprisingly, its genome consists of only 19.6 Mbp which is about 40−50% smaller than typical mesic species of the Trichocomaceae (e.g., *Aspergillus, Penicillium*).[9] Since the PKS moieties in the thermolides most likely originated from a reducing polyketide pathway, we scanned all the highly reducing PKS (HR-PKS) genes in the *T. thermophilus* genome along with their neighboring genes. Our attention focused on a cluster containing a predicted PKS−NRPS hybrid gene Talth1_004980_t1 (hereby named *TalA*). Most notably, the downstream flank of the cluster remains unannotated. The upstream flanking regions of *TalA* encode a hydrolase, a glutathione S-transferase, a cytochrome P450, a transporter, and three aldehyde dehydrogenases. Conserved domain analysis demonstrated that *TalA* gene encoded six catalytic domains for the PKS and four catalytic domains for the NRPS with an additional C-terminal reductase domain and lacked introns (Figure 2). Phylogenetic analysis showed that the

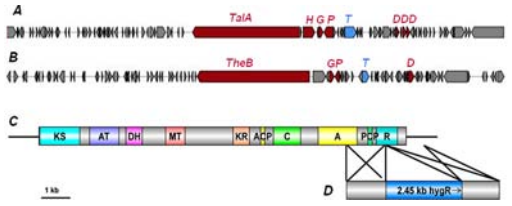


Figure 2. Thermolide biosynthetic gene clusters and genomic organization of thermolide biosynthetic genes. (A) *TalA* gene cluster spanning 51.843 kb and consisting of eight genes with their predicted functions color-coded: H, hydrolase; G,glutathione S-transferase (GST); P, cytochrome P450; D, aldehyde dehydrogenase; T, transport-related gene. (B) *TheB* gene clusters spanning 51.846 kb and consisting of five genes with their predicted functions color-coded. (C) Schematic view of the organization of thermolide biosynthetic genes encoding a polyketide synthase (KS) acyltransferase (AT), dehydratase (DH), methyltransferase (MT), β-ketoacylreductase domain (KR), acyl carrier protein (ACP), condensation (C), adenylation (A), thiolation (T), peptide carrier protein (PCP), reductase (R). (D) Representation of the genetic disruption cassette. We have not found any candidate genes that might be responsible for the acetylations.

KS domain sequence from TalA was clustered within the fungal PKS−NRPS hybrids, and formed a distinct monophyletic branch with no other closely related KS domains as of yet (Figure S1, Supporting Information).[10]

Linking the *TalA* gene with thermolide biosynthesis was challenging because no system for genetic manipulation has been previously reported for this fungus. In fact, all attempts to transform *T. thermophilus* to generate targeted gene knockouts with the protoplast method failed, even though our laboratory has successfully constructed many targeted mutants for different nematode- and insect-associated fungi.[11] Therefore, we tested *Agrobacterium tumefaciens*-mediated transformation (ATMT) as an alternative.[12]

A modified ATMT method for genetic disruption of *TalA* gene was developed using double-crossover recombination with the hygromycin-resistance gene (*hyg*) as a selection marker, followed by identification of desired Δ*TalA* mutants using diagnostic PCR. The two homologous regions were amplified from *T. thermophilus* genomic DNA using primers containing overlapping regions with the vector and the *hyg*-resistance cassette. Mutant cultures were extracted and analyzed by LC−MS (Figure 3). The gene sequences encoding for A and R domains in *TalA*, respectively, were disrupted. Metabolite analysis showed that both types of Δ*TalA* mutants were devoid of the production of compounds **1**−**5**. No intermediates, such as the free C18 linear aldehyde or carboxylic acid (Figure 5), could be detected in either of the extracts, indicating that *TalA* played a key role in the biosynthetic pathway for **4** and **5**. Since

dx.doi.org/10.1021/ol501595z | *Org. Lett.* 2014, 16, 3744−3747

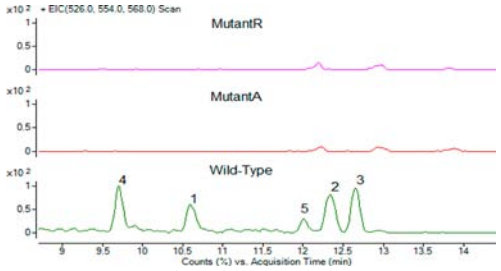


Figure 3. HPLC−MS analysis of fermentations from the wild-type and disruption mutants of *T. thermophilus*. Full-scan + mode spectra were acquired over a scan range of m/z 100−1500. When grown in YM medium, thermolides A−C (peak **1**−**3**, m/z = 568) and thermolides D−E (**4**, m/z = 526; **5**, m/z = 554) were detected in the wild-type. MutantR and MutantA: Genetic disruption of the gene sequence encoding for R and A domains in *TheB*, respectively (magenta and red lines).

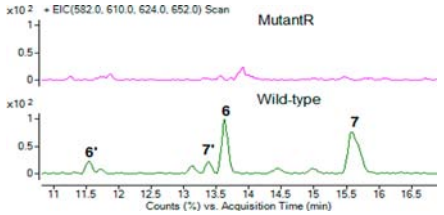


Figure 4. HPLC−MS analysis of fermentations from the wild-type (green line) and disruption mutant of *Th. lanuginosus*. Full-scan + mode spectra were acquired over a scan range of m/z 100−1500. When grown in YM medium, thermolides F and G (**6**, m/z = 624; **7**, m/z = 652) and deacetylthermolides F and G (**6′**, m/z = 582; **7′**, m/z = 610) were detected in the wild-type. MutantR: genetic disruption of the gene sequence encoding for R domain in *TheB* (magenta line).

compounds **4** and **5** differed in their amino acids, it was reasonable to attribute this flexibility to the ambivalent amino acid selection by the TalA NRPS. Such flexibility is atypical for fungal NRPSs; for example, enniatin synthetases isolated from various fungi exhibited differences in their amino acids specificities. However, this flexibility was not unexpected given our previous findings. Talathermophilins A and B from *T. thermophilus*, two terpene synthase (TPS)-NRPS metabolites with only one amino acid difference, maintained a constant content ratio in the fungal culture broth under different culture conditions.[13] Recent genome mining efforts revealed that *T. thermophilus* harbored only one TPS−NRPS gene cluster.[9] Moreover, coproduction of two related nonribosomal peptides from a bacterial NRPS is also commonplace,[14] such as occurs during the biosynthesis of aureusimines, where a second A domain incorporates both tyrosine (aureusimine A) and phenylalanine (aureusimine B/phevalin). Since the A domain selected and activated the amino acids, its relationship to other fungal NRPS gene A domains was examined by phylogenetic analysis (Figure S2, Supporting Information). Like the KS, the A domain of TalA again formed a monophyletic clade distinct from those of the other fungal and bacterial NRPSs.[14]

Our chemical and genome analysis and genetic manipulation of *T. thermophilus* revealed that *T. thermophilus* could not produce **6**. We extended our search to another thermophilic fungi collected from the same habitat as *T. thermophilus* YM3−4 on the basis of that the remarkable structural relationship of **1**−**5** to **6**. Considering the fact that **4** and **5** were coproduced in *T. thermophilus*, we postulated that the compound **6**-producing strain might also give a valine congener of **6**. Finally, thermolide F (**6**), together with the novel thermolide G (**7**), was detected and purified from the cultural broth of *Thermomyces lanuginosus* C5, which is closely related to *T. thermophilus*. Thermolide G (**7**) was assigned a molecular formula of $C_{34}H_{64}NO_9$ on the basis of a quasi-molecular ion peak at m/z 630.4583 $[M + H]^+$ in its positive HRESIMS. The 1D and 2D NMR spectra of **7** (Figures S12−17, Supporting Information) unequivocally suggested that **7** was a valine congener of **6**.

The genome of the strain *Th. lanuginosus* ATCC 200065 has also been recently released.[9] Prospective chemical analysis of extracts of the strain *Th. lanuginosus* ATCC 200065 indicated that this strain also produced **6** and **7** (Figure 4). We also detected the existence of their respective deacetyl analogues (**6′** and **7′**, Figure 4) in both *Th. lanuginosus* strains. Because of poor titers, so far, we have been unable to purify enough from the culture broths to obtain their NMR spectra.

Th. lanuginosus has a similar genome size as *T. thermophilus*[9] and contains a similar intronless hybrid PKS−NRPS gene Thela2_004538 (*TheB*). TheB was 85% identical to TalA at the protein level. The open reading frame (orf) between *TalA* and the gene putatively encoded for GST was assigned encoding a hydrolase (H, Figure 2), while the similarly located orf in *TheB* cluster remained unassigned. Conserved domain analysis of the orf indicated that its function resembles that in *TalA*. A similar method for gene disruption of *TheB* in *Th. lanuginosus* was developed (Figure 2). The sequence encoding for the R domain in *TheB* was genetically disrupted. LC−MS analysis of metabolite profiles of parallel fermentations of mutants and wild-type strains revealed that Δ*TheB* mutants also completely lacked the production of compounds **6** and **7** and their deacetyl analogues (**6′** and **7′**, Figure 4).

Thus, two genes *TalA* and *TheB* from two sister thermophilic fungal species involved in the biosynthesis of the class of bacterial-like thermolides have been identified. This is the first discovery of microbial hybrid iterative PKS−NRPS genes involved in the biosynthesis of lactam-bearing macrolactones that are common for modular PKS−NRPS genes.

The remarkable relationship of these two genes and thermolide structures suggests that these two hybrid iterative PKS−NRPS genes encode the biosynthesis of two similar PKS chains with differing lengths, a C18 carbon chain, and a C20 carbon chain, respectively. Subsequently, two structurally related amino acids (alanine and valine) are ambivalently recruited for the extension of the PKS carbon chain in the biosynthetic pathway. This is the first discovery of iterative PKS−NRPS genes responsible for the biosynthesis of two similar carbon scaffolds in fungi.

Conceivably, the linear intermediates, for example **4′** and **5′** from the R domain of TalA (Figure 5), might undergo an alternative mechanism, eventually leading to the formation of macrolactone rings. This deviation might be attributed to the position-6 hydroxyl group in the thermolides instead of a ketone for a β-ketoamide moiety, as has been proposed for all the other fungal PKS/NRPS hybrid metabolites.[7] Such a deviation might result in the adjacent methylene of the amide in the thermolides being less able to form a resonance-stabilized carbanion intermediate that could undergo intermolecular Claisen condensation with the distal end aldehyde group.[5,7] Instead, the hydroxyl group at C-12 (**1**−**7**, Figure 5) might act as an internal nucleophile to capture the distal end ketone carbon leading to lactones and yielding products with 13-membered heterocyclizations. Such a macrolactonization has normally been observed in bacterial modular PKS−NRPSs.[1−3] We were unable to detect either of the possible intermediate aldehydes or acids in

dx.doi.org/10.1021/ol501595z | *Org. Lett.* 2014, 16, 3744−3747

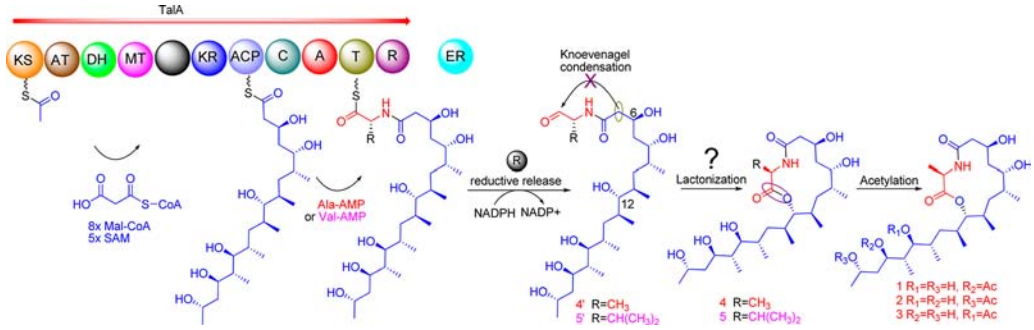


Figure 5. Architecture of the thermolide synthase TalA and model for the biosynthesis of thermolies A–E (**1**–**5**).

the wild-type strains and were unsuccessful in the heterologous expressions of *TalA* and *TheB* genes in *Aspergillus nidulans*, perhaps due to different splicing and regulation mechanisms. It seems likely that the macrolactonization was simply achieved by the nucleophilic attack from the hydroxyl group at C-12 to the enzyme-thethered thioester bond, which might be mediated by the "R" domain. Whether the orfs upstream of the *TalA* or *TheB*, for example, such as orf H (annotated as hydrolase), might contribute auxiliary functions during the biosynthesis of thermolides, warrants further investigation. Therefore, the mechansim for macrolactonization still remains unknown.

■ ASSOCIATED CONTENT

Supporting Information

Complete description of materials and methods, supporting figures (S1–S17), and the NMR and MS data of thermolide F (**7**). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: xmniu@ynu.edu.cn.
*E-mail: zhiqiang.an@uth.tmc.edu.

Author Contributions

§These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Genozymes Project, Concordia University, for access to genomes of thermophilic fungi. This work was sponsored by projects from the National Basic Research Program of China on Biological Control of Key Crop Pathogenic Nematodes (2013CB127505), National High Technology Research and Development Program of China (2011AA10A205), National Natural Science Foundation of China (U1036602), a scholarship awarded to X.N. (201207035005) from the China Scholarship Council, the University of Texas Health Science Center at Houston new faculty startup funds (to G.F.B.), and the Texas Emerging Technology Fund and the Welch Foundation (Ggrant No. AU00024 to Z.A.).

■ REFERENCES

(1) (a) Koppa, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24*, 735–749. (b) Graziani, E. I. *Nat. Prod. Rep.* **2009**, *26*, 602–609. (c) Walsh, C. T. *Science* **2004**, *303*, 1805–1810. (d) Katz, L.; Ashley, G. W. *Chem. Rev.* **2005**, *105*, 499–528. (e) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468. (f) Hertweck, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 4688–4716.

(2) (a) Xu, Y.; Zhou, T.; Zhang, S.; Xuan, L.; Zhan, J.; Molnár, I. J. *Am. Chem. Soc.* **2013**, *135*, 10783–10791. (b) Du, L.; Lou, L. *Nat. Prod. Rep.* **2010**, *27*, 255–278.

(3) (a) Gaffoor, I.; Trail, F. *Appl. Environ. Microbiol.* **2006**, *72*, 1793–1799. (b) Zhou, H.; Qiao, K.; Gao, Z.; Meehan, M. J.; Li, J. W.-H.; Zhao, X.; Dorrestein, P. C.; Vederas, J. C.; Tang, Y. *J. Am. Chem. Soc.* **2010**, *132*, 4530–4531. (c) Xu, Y.; Espinosa-Artiles, P.; Schubert, V.; Xu, Y.; Zhang, W.; Lin, M.; Gunatilaka, A. A. L.; Süssmuth, R.; Molnár, I. *Appl. Environ. Microbiol.* **2013**, *79*, 2038–2047. (d) Chiang, Y.; Szewczyk, E.; Nayak, T.; Davidson, A. D.; Sanchez, J. F.; Lo, H.; Wen-Yueh, H.; Simityan, H.; Kuo, E.; Praseuth, A.; Watanabe, K.; Oakley, B. R.; Wang, C. C. C. *Chem. Biol.* **2008**, *15*, 527–532.

(4) (a) Bradley, S. E.; Robinson, S. J.; Kelleher, N. L. *Fungal Genet. Biol.* **2011**, *48*, 49–61. (b) Boettger, D.; Hertweck, C. *ChemBioChem* **2013**, *14*, 28–42. (c) Fisch, K. M. *RSC Adv.* **2013**, *3*, 18228–18247.

(5) (a) Sims, J. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2008**, *130*, 11149–11155. (b) Liu, X.; Walsh, C. T. *Biochemistry* **2009**, *48*, 8746–8757.

(6) (a) Lou, L.; Qian, G.; Xie, Y.; Hang, J.; Chen, H.; Zaleta-Rivera, K.; Li, Y.; Shen, Y.; Dussault, P. H.; Liu, F.; Du, L. *J. Am. Chem. Soc.* **2011**, *133*, 643–645. (b) Li, Y.; Chen, H.; Ding, Y.; Xie, Y.; Wang, H.; Cerny, R. L.; Shen, Y.; Du, L. *Angew. Chem., Int. Ed.* **2014**, DOI: 10.1002/anie.201403500. (c) Antosch, J.; Schaefers, F.; Gulder, T. A. M. *Angew. Chem., Int. Ed.* **2014**, *53*, 3011–3014. (d) Zhang, G.; Zhang, W.; Zhang, Q.; Shi, T.; Ma, L.; Zhu, Y.; Li, S.; Zhang, H.; Zhao, Y.-L.; Shi, R.; Zhang, C. *Angew. Chem., Int. Ed.* **2014**, *53*, 4840–4844.

(7) Berka, R. M.; Grigoriev, I. V.; Otillar, R.; Salamov, A.; Grimwood, J.; Reid, I.; Ishmael, N.; John, T.; Darmond, C.; Moisan, M.-C; Henrissat, B.; Coutinho, P. M.; Lombard, V.; Natvig, D. O.; Lindquist, E.; Schmutz, J.; Lucas, S.; Harris, P.; Powlowski, J.; Bellemare, A.; Taylor, D.; Butler, G.; de Vries, R. P.; Allijn, I. E.; van den Brink, J.; Ushinsky, S.; Storms, R.; Powell, A. J.; Paulsen, I. T.; Elbourne, L. D. H.; Baker, S. E.; Magnuson, J.; LaBoissiere, S.; Clutterbuck, A. J.; Martinez, D.; Wogulis, M.; de Leon, A. L.; Rey, M. W.; Tsang, A. *Nat. Biotechnol.* **2011**, *29*, 922–927.

(8) Guo, J. P.; Zhu, C. Y.; Zhang, C. P.; Chu, Y. S.; Wang, Y. L.; Zhang, J. X.; Wu, D. K.; Zhang, K. Q.; Niu, X. M. *J. Am. Chem. Soc.* **2012**, *134*, 20306–20309.

(9) http://fungalgenomics.ca/wiki/Fungal_Genomes (last accessed Apr 15, 2014).

(10) Finn, R. D.; Clements, J.; Eddy, S. R. *Nucleic Acids Res.* **2011**, *39*, W29–W37.

(11) Jiang, D.; Zhu, W.; Wang, Y.; Sun, C.; Zhang, K.; Yang, J. *Biotechnol. Adv.* **2013**, *31*, 1562–1574.

(12) (a) Chen, L.; Yue, Q.; Zhang, X.; Xiang, M.; Wang, C.; Li, S.; Che, Y.; Ortiz-Lopez, F.; Bills, G.; Liu, X.; An, Z. *BMC Genomics* **2013**, *14*, 339. (b) Zhang, A.; Lu, P.; Dahl-Roshak, A. M.; Paress, P. S.; Kennedy, S.; Tkacz, J. S.; An, Z. *Mol. Gen. Genomics* **2003**, *268*, 645–655. (c) Lazo, G. R.; Stein, P. A.; Ludwig, R. A. *Nat. Biotechnol.* **1991**, *9*, 963–967.

(13) Chu, Y. S.; Niu, X. M.; Wang, Y. L.; Guo, J. P.; Pan, W. Z.; Huang, X. W.; Zhang, K. Q. *Org. Lett.* **2010**, *12*, 4356–4359.

(14) Wyatt, M. A.; Wang, W.; Roux, C. M.; Beasley, F. C.; Heinrichs, D. E.; Dunman, P. M.; Magarvey, N. A. *Science* **2010**, *329*, 294–296.