Discovery of fungal onoceroid triterpenoids through domainless enzyme-targeted global genome mining

Jia Tang[a] and Yudai Matsuda*[a]

[a]Department of Chemistry, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China

*Corresponding author: ymatsuda@cityu.edu.hk

ABSTRACT

Genomics-guided methodologies have been widely adopted and proven beneficial for the discovery of natural products. However, a major challenge in the field of genome mining is determining how to selectively extract biosynthetic gene clusters (BGCs) for untapped natural products from a substantial number of available genome sequences. In this study, we developed a fungal genome mining pipeline that extracts BGCs encoding enzymes that lack a detectable protein domain and are not recognized as biosynthetic proteins by existing bioinformatic tools. We searched for BGCs encoding a homologue of Pyr4-family terpene cyclases, which are representative examples of domainless enzymes, in approximately 2,000 fungal genomes and discovered several BGCs with unique features. The subsequent characterization of these BGCs led to the discovery of the first fungal onoceroid triterpenoids and unprecedented onoceroid synthases. Our genome mining pipeline has broad applicability in fungal genome mining and can serve as a beneficial tool for accessing diverse, unexploited natural products.

INTRODUCTION

Recent years have witnessed an exponential increase in data accumulation across all fields, ushering us into the era of big data. Research on naturally occurring organic compounds, typically referred to as natural products, is no exception to this global trend.[1] Traditional methods used to isolate natural products often lead to the rediscovery of known compounds. Thus, scientists are now urged to discover or synthesize novel natural products or their analogs using alternative strategies. Leveraging big data presents a promising solution to this problem. The rapid accumulation of microbial genome sequence data has revealed that microbial genomes harbor more natural product biosynthetic gene clusters (BGCs) than one can predict based on the number of metabolites produced under standard laboratory cultivation conditions.[2] Thus, the activation and characterization of these silent or orphan BGCs present in microbial genomes can lead to the discovery of new natural products. A method known as genome mining has been demonstrated as useful for discovering new metabolites for over a decade.[3] However, a major challenge in the field of genome mining is determining how to prioritize BGCs identified in available genome sequence data for the efficient discovery of untapped natural products.

Fungi are prolific producers of natural products with diverse molecular structures and biological activities, establishing them as promising sources of unexplored metabolites. To facilitate genome mining in fungi, several bioinformatics tools, such as antiSMASH,[4] DeepBGC,[5] and TOUCAN,[6] can be employed to detect BGCs in a specific fungal genome. Among the available tools, antiSMASH has gained the most popularity, possibly because of its user-friendly online platform and versatile functions. However, some challenges have been encountered while using antiSMASH for BGC detection and genome mining. First, antiSMASH classifies many genes as "other genes," even when they encode a biosynthetic enzyme. For example, when antiSMASH (version 7.0.0) was employed to analyze the BGC of the fungal meroterpenoid novofumigatonin,[7] a hybrid molecule of polyketide and terpenoid origin, it categorized 6 out of 13 genes as "other genes" (Figure S1). This may be because some of these genes lacked a detectable Pfam domain.[8] Moreover, one of these six overlooked genes produces the terpene cyclase NvfL, which has widely distributed homologues in natural product pathways.[9] Thus, although antiSMASH should theoretically categorize this BGC as "T1PKS, terpene" (type I polyketide synthase + terpene synthase), it recognized only T1PKS. In addition, these domainless enzymes included the α-ketoglutarate-dependent dioxygenase NvfI and the methyltransferase NvfJ, indicating that antiSMASH failed to recognize various biosynthetic proteins. Furthermore, although antiSMASH and the other aforementioned tools can be used to extract all possible BGCs from a genome sequence, they do not facilitate the prioritization of the detected BGCs. Thus, users need to select BGCs based on the criteria set for subsequent wet lab experiments. In this study, we aimed to overcome these challenges by developing a widely applicable genome mining methodology for the rapid discovery of novel natural products and obtaining a new class of metabolites.

RESULTS

Development of a fungal genome mining pipeline

Recent studies on the biosynthesis of fungal natural products have identified many biosynthetic enzymes lacking a detectable Pfam domain, such as NvfI and NvfL, which are not recognized by antiSMASH. Because these domainless biosynthetic enzymes often drive diverse chemical reactions, we hypothesized that a genome mining strategy focusing on BGCs encoding the weak homologues of such enzymes would offer a promising approach to obtaining untapped metabolites. To identify biosynthetic proteins without detectable Pfam domains, we first collected known fungal biosynthetic proteins. Although these proteins can be obtained from the UniProtKB/Swiss-Prot or MIBiG database,[10] these databases contain gene or protein sequences that can be mispredicted and are not manually corrected. Thus, we created our own fungal BGC database by manually collecting, reviewing, correcting where necessary, and systematically annotating approximately 700 fungal BGCs (Data S1). We named this database the FunBGCs database.

Next, we extracted all protein sequences from the created database and conducted a Pfam domain search on the resulting 5,070 proteins. This led to the identification of 520 Pfam domains. However, 572 proteins lacked a detectable Pfam domain. These domainless proteins were divided into groups based on their sequence similarity, and their hidden Markov model (HMM) profiles were generated for HMMER[11] analysis. The representative members of these protein groups are the terpene cyclase Pyr4,[12] the Diels–Alderase Fsa2,[13] the hetero Diels–Alderase AsR5,[14] the epoxide hydrolase CtvD,[15] and the isomerase Trt14.[16] Furthermore, several additional HMM profiles were created for the domains of polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs). The identified Pfam domains, the manually added protein families or domains, and a few additional protein domains from the SMART[17] and TIGRFAMs[18] databases (Tables S1 and S2) were combined for use in biosynthetic protein detection. Moreover, a DIAMOND[19] database was constructed with all of the extracted biosynthetic proteins.

BGCs are extracted as follows (Figure 1A). Initially, genes whose products belong to a protein family of interest are identified from a given fungal genome. Subsequently, genes in the flanking regions of each identified gene are examined to determine if they encode a core protein (e.g., PKSs, NRPSs, terpene synthases or cyclases, and prenyltransferases; refer to Table S3 for more details) or another type of biosynthetic protein using the previously created in-house HMM library and DIAMOND database. If a target gene is found to be colocalized with a core protein gene, then the genomic region is extracted as a BGC. To determine whether a gene in flanking regions is clustered with the target, the following factors are considered: (i) whether the gene encodes a biosynthetic protein, (ii) whether the gene is duplicated in the genome (to include a possible self-resistance enzyme gene[20]), (iii) whether the gene is biosynthetically related to nearby genes, (iv) the size of the gene product, and (v) the distance between a given gene and the next one (refer to Supporting Information for the detailed method). In addition, the extracted BGCs are visualized using a web browser, which displays general information on each BGC and each gene (Figure 1B). The fungal genome mining pipeline developed in this study was named FunBGCeX (Fungal BGC eXtractor).
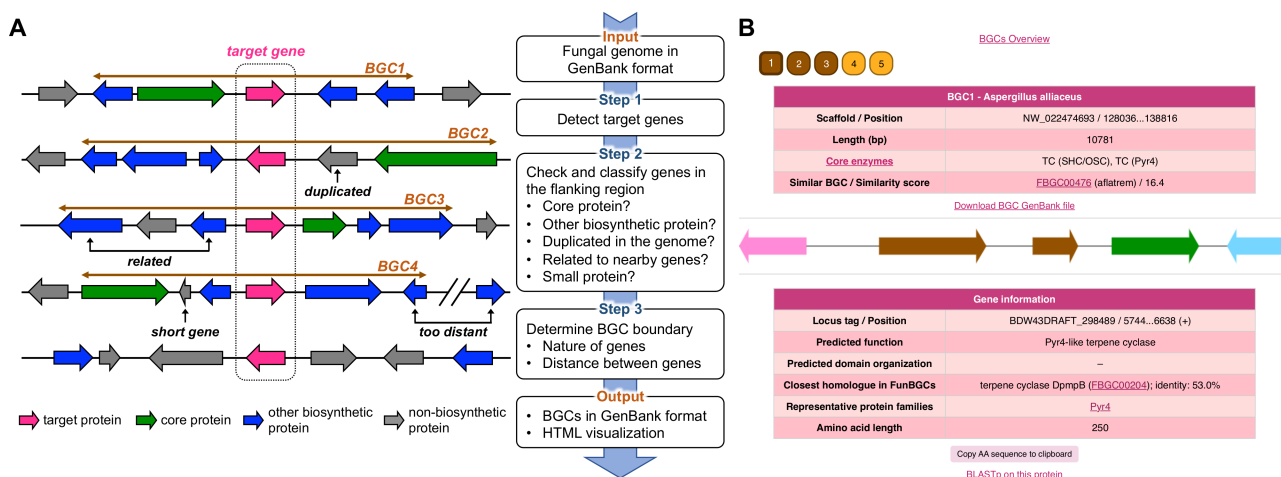
Figure 1. Fungal genome mining pipeline. (A) General workflow to extract BGCs with a target gene from a given fungal genome. (B) Example of output from the genome mining pipeline. For this panel, BGCs encoding a Pyr4 homologue were extracted from the fungus *Aspergillus alliaceus* CBS 536.65, yielding five BGCs. The top table provides general information on a BGC. A schematic representation of BGC is displayed in the middle. On clicking each gene, information on that gene is provided in the bottom table. The BGC shown here is identical to the *alli* cluster (Figure 2C).

Discovery of unusual fungal triterpenoid BGCs

To examine whether FunBGCeX allows for the efficient discovery of BGCs that potentially synthesize a new class of natural products, we extracted BGCs that encoded a homologue of Pyr4 (Figure 2A),[12, 21] which is a noncanonical transmembrane terpene cyclase and does not have a conserved domain, as inferred from the National Center for Biotechnology Information (NCBI) Conserved Domain Search[22] (Figure S2). Thus, such BGCs were extracted from 1,990 annotated fungal reference genomes downloaded from the NCBI database using an HMM profile created from Pyr4 and its homologues.[9, 23] This process resulted in the identification of 1,050 BGCs (including those with a single gene; Data S2). More than half (690) of the identified BGCs did not contain an additional core protein gene, whereas 182 BGCs possessed at least one PKS gene and 131 BGCs were found to encode a PaxC-like prenyltransferase, which is specifically required for the biosynthesis of indole sesquiterpenoids and diterpenoids.[24] Thus, except for standalone *pyr4*-like genes, the majority of the *pyr4*-like genes were clustered with either a PKS gene or a *paxC*-like gene. This observation is consistent with the fact that, in fungi, Pyr4-family terpene cyclases have been found only in the biosynthetic pathways of polyketide–terpenoid hybrids and indole terpenoids.[25] Intriguingly, 19 BGCs encode one or two protein(s) homologous to squalene–hopene cyclases (SHCs) or oxidosqualene cyclases (OSCs), which are involved in the biosynthesis of triterpenoids. Although a few of these BGCs were removed after manual review for further investigation, most of them were found to encode Pyr4-like proteins that form a new clade when subjected to phylogenetic analysis with known Pyr4-family terpene cyclases (Figure 2B). During our study, we observed that *Aspergillus fumigatus* A1163, a non-reference strain of *A. fumigatus*, harbored a similar BGC, which was reserved for further investigation. To date, no metabolic pathway involving both a Pyr4-family terpene cyclase and an SHC/OSC-

like enzyme has been identified. Furthermore, the nature of these BGCs suggests that their products are triterpenoids whose biosynthesis requires two cyclization events. However, to the best of our knowledge, none of the known fungal metabolites appear to be synthesized by these BGCs. Therefore, we speculate that these BGCs are responsible for the biosynthesis of untapped triterpenoid molecules.
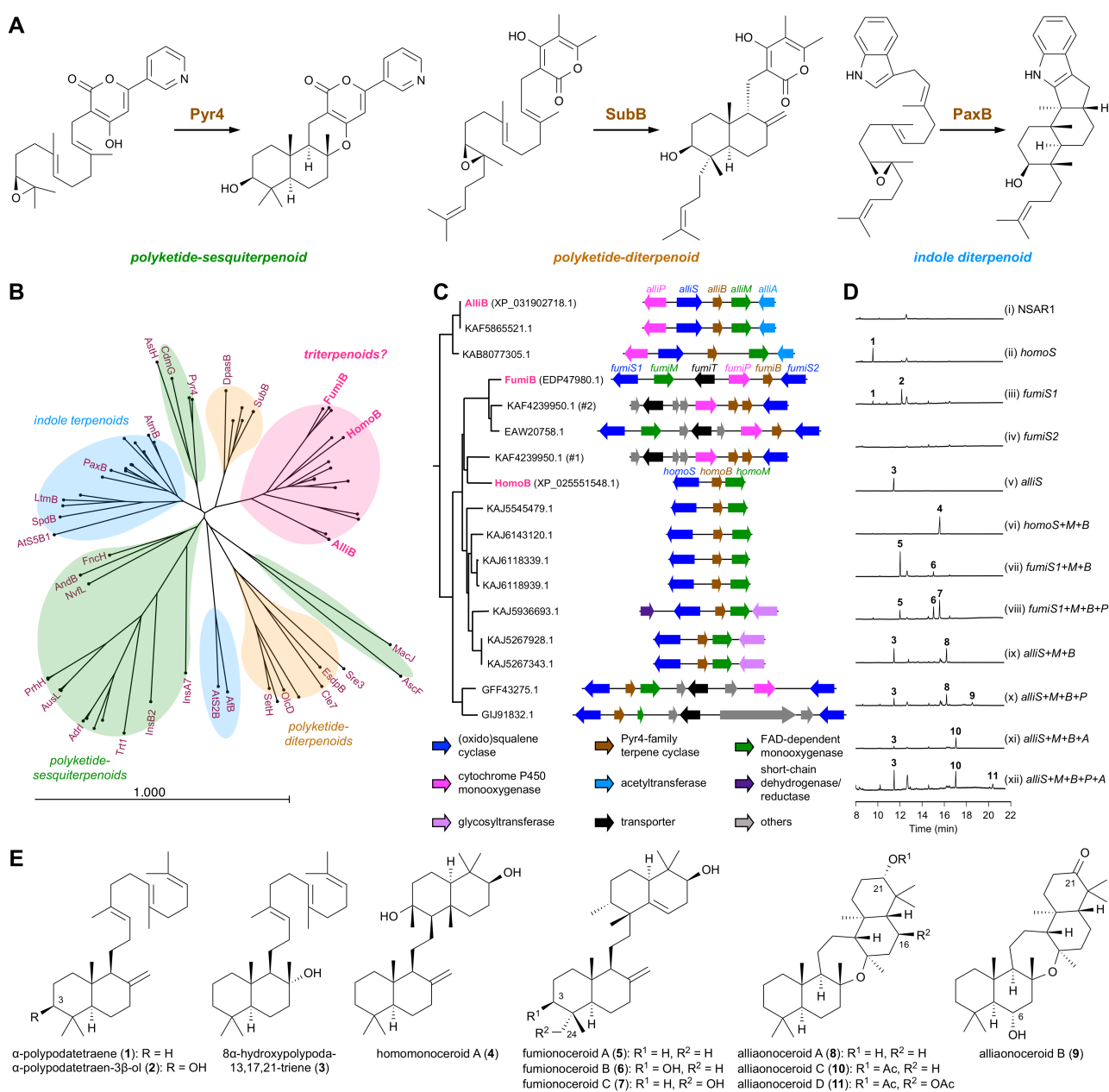

Figure 2. Discovery and characterization of fungal onoceroid BGCs. (A) Reactions catalyzed by the selected members of Pyr4-family terpene cyclases. (B) Phylogenetic analysis of known Pyr4 homologues and those clustered with an (oxido)squalene cyclase identified in this study. Because of space limitations, only selected enzymes were labeled. The gene and protein sequences of several identified Pyr4 homologues were manually revised upon creating the phylogenetic tree (refer to Table S6 for their protein sequences). (C) Fungal onoceroid BGCs. (D) GC–MS chromatograms of metabolites obtained from *A. oryzae* transformants. (E) Structures of triterpenoids obtained in this study.

Characterization of selected triterpenoid BGCs

To examine whether these BGCs yield new triterpenoids, three of them were selected for experimental investigation. The three BGCs are from *Aspergillus homomorphus* CBS 101889, *A. fumigatus* A1163 (CBS 144.89), and *Aspergillus alliaceus* CBS 536.65, which were designated as *homo*, *fumi*, and *alli* clusters, respectively (Figure 2C and Tables S4 and S5). The *homo* cluster is the simplest of the three and encodes the squalene or oxidosqualene cyclase HomoS, the Pyr4-family terpene cyclase HomoB, and the FAD-dependent monooxygenase (FMO) HomoM, the last of which is homologous to epoxidases involved in fungal meroterpenoid pathways.[25] In addition, the homologues of the three enzymes are conserved in the other two BGCs, although the *fumi* cluster contains two *homoS* homologues, *fumiS1* and *fumiS2*. On the basis of the predicted functions of Homo enzymes, the product of the *homo* cluster could be synthesized as follows. First, HomoS cyclizes squalene or oxidosqualene to yield a cyclized product. Then, HomoM epoxidizes one of the olefinic double bonds present in the cyclized product, and HomoB finally performs a second round of cyclization to produce the end pathway product. The other two BGCs would employ a similar biosynthetic mechanism; however, they each encode one or two tailoring enzymes predicted to modify the triterpenoid scaffold. The *fumi* cluster contains a cytochrome P450 monooxygenase gene *fumiP*, whereas the *alli* cluster encodes the P450 AlliP and the acetyltransferase AlliA. Collectively, all of the three BGCs are expected to yield different triterpenoid species.

To analyze the functions of the three BGCs and identify the metabolites produced, we performed heterologous expression experiments in *Aspergillus oryzae* NSAR1,[26] which has been widely utilized for the characterization of orphan BGCs.[21a, 27] We first individually expressed the four putative triterpene synthase genes (*homoS*, *fumiS1*, *fumiS2*, and *alliS*) in *A. oryzae* and then analyzed metabolites from the resulting transformants using gas chromatography–mass spectrometry (GC–MS). The findings revealed that all of the enzymes, except FumiS2, yielded specific metabolite(s) (Figure 2D, traces i to v). The *homoS*-transformed strain produced a major product **1** with m/z 410 [M]$^+$, which was not present in the host strain. After isolation, metabolite **1** was identified as bicyclic triterpene α-polypodatetraene based on the comparison of its nuclear magnetic resonance (NMR) spectra and specific rotation with reported data[28] (Figure 2E). Compound **1** was also obtained from the *A. oryzae* transformant harboring *fumiS1*; however, the transformant produced another product **2**, which was identified to be α-polypodatetraen-3β-ol,[29] the C-3 hydroxy analog of **1** (Figure 2E). The metabolic profile of the *A. oryzae* strain with *alliS* differed from those of the other two. A new metabolite **3** was detected as a major peak and was identified as 8α-hydroxypolypoda-13,17,21-triene[30] (Figure 2E).

We sought to identify the downstream metabolites of the three pathways and first focused on the end product of the *homo* pathway. The co-expression of the epoxidase gene *homoM* and the *pyr4* homologue *homoB* with *homoS* resulted in the formation of a new metabolite **4** with m/z 426 [M]$^+$ (Figure. 2D, trace vi). Analysis of the NMR data of **4** revealed the presence of two additional ring systems, whose absolute structure was established using the modified Mosher's method[31] (Figures 2E and S3). Thus, compound **4**, hereby named homomonoceroid A, is synthesized through the cyclization of squalene from both termini and is classified as an onoceroid.

Next, the co-expression system of *fumiS1*, *fumiM*, and *fumiB* was constructed, which yielded two additional metabolites, **5** and **6** (Figure 2D, trace vii). The major product **5**, designated as fumionoceroid A, was determined to be another onoceroid; its structure was established through NMR and X-ray crystallographic analyses (Figures 2E and S4; CCDC: 2279439). Compound **5** was also found to be a tetracyclic onoceroid as **4**, and both **4** and **5** appear to be derived from **1**. However, the newly formed bicyclic system of **5** differs from that of **4**. The other product **6**, fumionoceroid B, was determined to be the C-3 hydroxy form of **5** and appears to be synthesized in the same manner as **5**, using oxidosqualene as a starting material (Figure 2E). The absolute structure of **6** was deduced based on that of **5**. Subsequently, the P450 gene *fumiP* was introduced, yielding an additional product **7** (Figure 2D, trace viii). NMR and X-ray crystallographic analyses revealed that **7**, named fumionoceroid C, was a hydroxylated form of **5** at C-24 (Figures 2E and S4; CCDC: 2279440).

We then identified triterpenoids synthesized by the *alli* cluster. When *alliS*, *alliM*, and *alliB* were introduced into *A. oryzae*, the transformant with the three genes produced a new metabolite **8** (Figure 2D, trace ix). In contrast to the products obtained from the other BGCs, **8** was identified as a pentacyclic onoceroid and named alliaonoceroid A (Figure 2E). Although **8** has never been isolated from natural sources, it was previously synthesized as a predicted biosynthetic precursor of cupacinoxepin,[32] which had been isolated from the Ecuadorian plant *Cupania cinerea*.[33] We next turned our attention to the tailoring steps, which should involve the P450 AlliP and the acetyltransferase AlliA. We created four gene expression systems with either *alliP* or *alliA* to determine the enzyme required first in the biosynthetic process. Consequently, both the transformants yielded a new product. The introduction of *alliP* and *alliA* led to the formation of **9** and **10**, respectively (Figure 2D, traces x and xi), which were characterized to be the C-21 keto and C-6α hydroxy analog of **8** and the *O*-acetylated form of **8**, respectively, through NMR and X-ray crystallographic analyses (Figures 2E and S4; CCDC: 2279441 and 2279442). Finally, the transformant harboring all five *alli* genes produced **11** (Fig. 2A, trace xii), which was determined to possess two acetoxy groups at the C-16β and C-21α positions (Figures 2E and S4; CCDC: 2279443). Compounds **9**–**11** were designated as alliaonoceroids B–D, respectively.

DISCUSSION

In this study, we developed a fungal genome mining pipeline, FunBGCeX, which was based on a manually curated fungal BGC database and custom-made HMM profiles. We demonstrated that FunBGCeX could effectively identify BGCs for previously undiscovered natural products. The onoceroid BGCs characterized in this study could also be identified using antiSMASH; however, the BGCs extracted by antiSMASH contained substantially more outside genes than those obtained using FunBGCeX (Figure S5). In addition, antiSMASH could not detect the presence of *pyr4*-like genes in these BGCs. Therefore, the onoceroid BGCs would not have been effectively discovered using antiSMASH or other existing fungal genome mining tools. Although we focused only on BGCs with a *pyr4*-like terpene cyclase gene in this study, FunBGCeX can be readily applied to extract BGCs that encode a homologue of user-selected proteins. Thus, FunBGCeX can facilitate and accelerate the discovery of unexploited natural products, particularly those synthesized with the involvement of

a domainless enzyme.

In our global fungal genome mining study, we discovered three types of onoceroids, which are triterpenoids synthesized from squalene or oxidosqualene through cyclization at both ends of the prenyl chain. Onoceroids have been isolated from bacteria,[34] ferns,[35] higher plants,[36] and animals.[37] However, to the best of our knowledge, onoceroids have never been isolated from fungi. Thus, the present study provides the first examples of fungal onoceroids and their biosynthetic pathways. In terms of the biosynthesis of onoceroids, BmeTC from the bacterium *Bacillus megaterium* was the first onoceroid synthase characterized in 2013, and this enzyme solely transforms squalene into onoceroids.[34] In the fern *Lycopodium clavatum*, a pair of homologous enzymes (LCC and LCD or LCE) convert oxidosqualene into onoceroid species.[38] These known enzymes for onoceroid biosynthesis are all homologous to SHCs and OSCs. In contrast, the biosynthesis of fungal onoceroids requires two families of terpene cyclases (i.e., SHC/OSC-like enzyme and Pyr4-family terpene cyclase), introducing an unprecedented biosynthetic mechanism of onoceroids. This is the first study to demonstrate the involvement of Pyr4-family terpene cyclases in the biosynthesis of "pure" (not mero-) terpenoids. The majority of fungal triterpenoids are synthesized through the cyclization of oxidosqualene,[39] and hexaprenyl pyrophosphate is known to serve as the precursor of a few fungal triterpenoids.[40] However, the fungal onoceroid pathways identified in this study represent rare examples in which squalene is directly cyclized to produce fungal triterpenoids.

The biosynthetic pathways of the fungal onoceroids discovered in this study can be proposed as follows (Scheme 1). The biosynthesis of homomonoceroid A (**4**) begins with squalene being cyclized into α-polypodatetraene (**1**) by the squalene cyclase HomoS through the carbocationic intermediate **12**. Subsequently, the FMO HomoM epoxidizes the terminal olefin of **1** to yield epoxide **13**. The Pyr4-family terpene cyclase HomoB then protonates epoxide **13** to initiate the second round of cyclization, and the resulting tetracyclic carbocationic species **14** is neutralized by a water attack, yielding **4**. In addition, the bicyclic intermediate **13** is involved in the biosynthesis of fumionoceroid C (**7**), where FumiB accepts **13** to produce a differently cyclized product fumionoceroid A (**5**). FumiB first cyclizes **13** into the carbocationic intermediate **15**. Instead of being quenched by water, the reaction concludes with a 1,2-hydride shift, 1,2-methyl shift, and deprotonation from C-19. Subsequently, the P450 FumiP hydroxylates **5** at C-24 to complete the biosynthesis. Oxidosqualene can also be accepted by Fumi enzymes, except for FumiP, to produce the C-3 hydroxy analog of **5**, fumionoceroid B (**6**), through α-polypodatetraen-3β-ol (**2**) (Figure S6). Meanwhile, the biosynthetic pathway leading to alliaonoceroid D (**11**) branches from the other two pathways in the first step. AlliS transforms squalene into 8α-hydroxypolypoda-13,17,21-triene (**3**) instead of **1**. After the epoxidation of **3** by AlliM to yield **16**, AlliB cyclizes **16** into alliaonoceroid A (**8**). The cyclization mode adopted by AlliB is similar to that adopted by HomoB, but the AlliB-catalyzed reaction uses the C-8 hydroxy group instead of a water molecule to produce the pentacyclic onoceroid **8**. Compound **8** then undergoes acetylation at the C-21 hydroxy group; this reaction is catalyzed by the acetyltransferase AlliA, yielding alliaonoceroid C (**10**). The P450 AlliP then installs a hydroxy group at C-16 to produce **17**, which is again acetylated by AlliA to yield the end product **11**. In the absence of AlliA, AlliP might cause the hydroxylation of **3** at C-6 to yield a shunt pathway product **18**, which is then oxidized by an

endogenous enzyme of *A. oryzae* to give alliaonoceroid B (**9**) (Figure S6). The three characterized fungal onoceroid pathways suggest that fungi synthesize diverse onoceroids due to various factors affecting structural diversification, such as the (i) cyclization mode of (oxido)squalene, (ii) the reaction mechanisms of Pyr4-family terpene cyclases, and (iii) tailoring reactions. The characterization of other onoceroid BGCs not examined in this study or the construction of artificial pathways with onoceroid biosynthetic genes from different pathways can further expand the molecular diversity of fungal onoceroids.
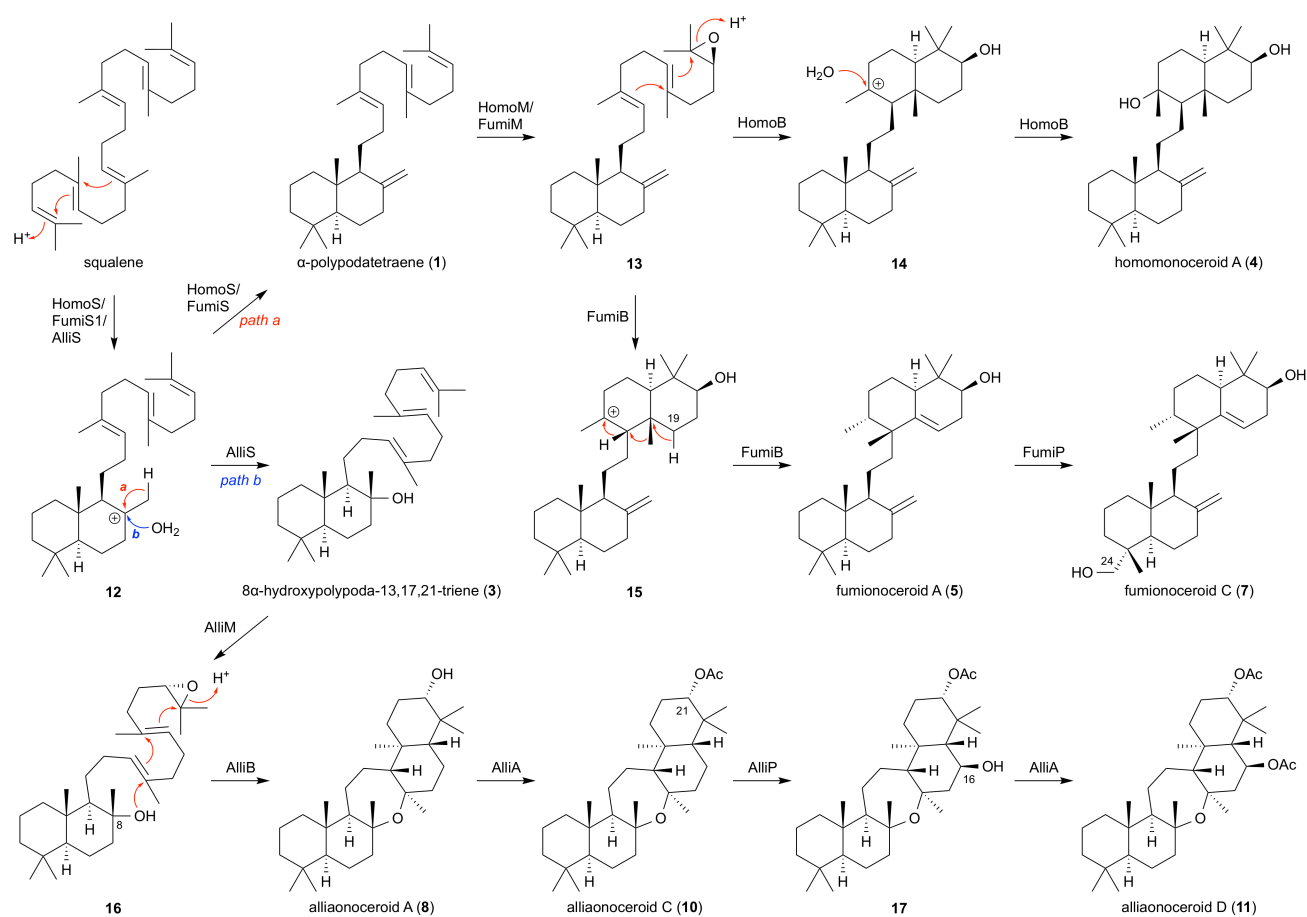
Scheme 1. Proposed biosynthetic pathways of fungal onoceroids.

In conclusion, we successfully isolated several previously unreported natural products using our genome mining pipeline, FunBGCeX. We believe that this tool can be widely applied to the discovery of natural products with novel scaffolds. Currently, FunBGCeX mainly targets "known–unknown" BGCs,[41] which produce unknown natural products synthesized by the known classes of core enzymes. A recent study has highlighted the genome mining-driven discovery of "unknown–unknown" natural products.[42] We are now enhancing the genome mining platform by incorporating additional functions to readily extract BGCs encoding self-resistance enzymes[20] or BGCs that lack a known core protein. This would facilitate the discovery of unexploited bioactive or unknown–unknown natural products from fungi.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

Supporting Information. Supplementary materials and methods, Tables S1–S9, Figures S1–S65, and supplementary references.

Data S1. Summary of the fungal BGCs manually curated in this study.

Data S2. Summary of the BGCs extracted in this study.

Data and code availability

All original codes for FunBGCeX are deposited at Zenodo under the DOI 10.5281/zenodo.8126797 and are also available at https://github.com/ydmatsd/funbgcex. The manually curated fungal BGCs and custom-made HMM profiles are deposited at Zenodo under the DOI 10.5281/zenodo.8126803. The online version of the FunBGCs database can currently be accessed at http://staffweb1.cityu.edu.hk/ymatsuda/funbgcs/funbgcs.html. The crystallographic data of compounds **5**, **7**, **9**, **10**, and **11** are deposited at the CCDC under the database identifiers 2279439, 2279440, 2279441, 2279442, and 2279443, respectively. Other data are provided in supporting information.

REFERENCES

(1) (a) Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappelini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H., A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16*, 60-68; (b) Chevrette, M. G.; Gavrilidou, A.; Mantri, S.; Selem-Mojica, N.; Ziemert, N.; Barona-Gómez, F., The confluence of big data and evolutionary genome mining for the discovery of natural products. *Nat. Prod. Rep.* **2021**, *38*, 2024-2040.

(2) (a) Nett, M.; Ikeda, H.; Moore, B. S., Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **2009**, *26*, 1362-1384; (b) Brakhage, A. A., Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **2013**, *11*, 21-32.

(3) (a) Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015**, *13*, 509; (b) Covington, B. C.; Xu, F.; Seyedsayamdost, M. R., A natural product chemist's guide to unlocking silent biosynthetic gene clusters. *Annu. Rev. Biochem.* **2021**, *90*, 763-788; (c) Bauman, K. D.; Butler, K. S.; Moore, B. S.; Chekan, J. R., Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **2021**, *38*, 2100-2129.

(4) Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B. R.; Metcalf, W. W.; Helfrich, E. J. N.; van Wezel, G. P.; Medema, M. H.; Weber, T., antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **2023**.

(5) Hannigan, G. D.; Prihoda, D.; Palicka, A.; Soukup, J.; Klempir, O.; Rampula, L.; Durcak, J.; Wurst, M.; Kotowski, J.; Chang, D.; Wang, R.; Piizzi, G.; Temesi, G.; Hazuda, D. J.; Woelk, C. H.; Bitton, D. A., A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **2019**, *47*, e110.

(6) Almeida, H.; Palys, S.; Tsang, A.; Diallo, A. B., TOUCAN: A framework for fungal biosynthetic gene cluster discovery. *NAR Genom. Bioinform.* **2020**, *2*.

(7) Matsuda, Y.; Bai, T.; Phippen, C. B. W.; Nødvig, C. S.; Kjærbølling, I.; Vesth, T. C.; Andersen, M. R.; Mortensen, U. H.; Gotfredsen, C. H.; Abe, I.; Larsen, T. O., Novofumigatonin biosynthesis involves a non-heme iron-dependent endoperoxide isomerase for orthoester formation. *Nat. Commun.* **2018**, *9*, 2587.

(8) Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, Gustavo A.; Sonnhammer, E. L. L.; Tosatto, S. C. E.; Paladin, L.; Raj, S.; Richardson, L. J.; Finn, R. D.; Bateman, A., Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2020**, *49*, D412-D419.

(9) Barra, L.; Abe, I., Chemistry of fungal meroterpenoid cyclases. *Nat. Prod. Rep.* **2021**, *38*, 566-585.

(10) Terlouw, B. R.; Blin, K.; Navarro-Muñoz, J. C.; Avalon, N. E.; Chevrette, M. G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, Michael J. J.; Reitz, Zachary L.; van Santen, Jeffrey A.; Selem-Mojica, N.; Tørring, T.; Zaroubi, L.; Alanjary, M.; Aleti, G.; Aguilar, C.; Al-Salihi, Suhad A. A.; Augustijn, Hannah E.; Avelar-Rivas, J A.; Avitia-Domínguez, Luis A.;

Barona-Gómez, F.; Bernaldo-Agüero, J.; Bielinski, V. A.; Biermann, F.; Booth, Thomas J.; Carrion Bravo, Victor J.; Castelo-Branco, R.; Chagas, Fernanda O.; Cruz-Morales, P.; Du, C.; Duncan, Katherine R.; Gavriilidou, A.; Gayrard, D.; Gutiérrez-García, K.; Haslinger, K.; Helfrich, Eric J. N.; van der Hooft, Justin J. J.; Jati, Afif P.; Kalkreuter, E.; Kalyvas, N.; Kang, Kyo B.; Kautsar, S.; Kim, W.; Kunjapur, Aditya M.; Li, Y.-X.; Lin, G.-M.; Loureiro, C.; Louwen, Joris J. R.; Louwen, Nico L. L.; Lund, G.; Parra, J.; Philmus, B.; Pourmohsenin, B.; Pronk, Lotte J. U.; Rego, A.; Rex, Devasahayam Arokia B.; Robinson, S.; Rosas-Becerra, L R.; Roxborough, Eve T.; Schorn, Michelle A.; Scobie, Darren J.; Singh, Kumar S.; Sokolova, N.; Tang, X.; Udwary, D.; Vigneshwari, A.; Vind, K.; Vromans, Sophie P. J. M.; Waschulin, V.; Williams, Sam E.; Winter, Jaclyn M.; Witte, Thomas E.; Xie, H.; Yang, D.; Yu, J.; Zdouc, M.; Zhong, Z.; Collemare, J.; Linington, Roger G.; Weber, T.; Medema, Marnix H., MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **2022**, *51*, D603-D610.

(11) HMMER: biosequence analysis using profile hidden Markov models. http://hmmer.org/.

(12) Itoh, T.; Tokunaga, K.; Matsuda, Y.; Fujii, I.; Abe, I.; Ebizuka, Y.; Kushiro, T., Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. *Nat. Chem.* **2010**, *2*, 858-864.

(13) Fujiyama, K.; Kato, N.; Re, S.; Kinugasa, K.; Watanabe, K.; Takita, R.; Nogawa, T.; Hino, T.; Osada, H.; Sugita, Y.; Takahashi, S.; Nagano, S., Molecular basis for two stereoselective Diels–Alderases that produce decalin skeletons. *Angew. Chem. Int. Ed.* **2021**, *60*, 22401-22410.

(14) Schor, R.; Schotte, C.; Wibberg, D.; Kalinowski, J.; Cox, R. J., Three previously unrecognised classes of biosynthetic enzymes revealed during the production of xenovulene A. *Nat. Commun.* **2018**, *9*, 1963.

(15) Lin, T.-S.; Chiang, Y.-M.; Wang, C. C. C., Biosynthetic pathway of the reduced polyketide product citreoviridin in *Aspergillus terreus* var. *aureus* revealed by heterologous expression in *Aspergillus nidulans*. *Org. Lett.* **2016**, *18*, 1366-1369.

(16) Matsuda, Y.; Iwabuchi, T.; Wakimoto, T.; Awakawa, T.; Abe, I., Uncovering the unusual D-ring construction in terretonin biosynthesis by collaboration of a multifunctional cytochrome P450 and a unique isomerase. *J. Am. Chem. Soc.* **2015**, *137*, 3393–3401.

(17) Letunic, I.; Khedkar, S.; Bork, P., SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* **2020**, *49*, D458-D460.

(18) Haft, D. H.; Selengut, J. D.; Richter, R. A.; Harkins, D.; Basu, M. K.; Beck, E., TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* **2012**, *41*, D387-D395.

(19) Buchfink, B.; Reuter, K.; Drost, H.-G., Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18*, 366-368.

(20) Yan, Y.; Liu, N.; Tang, Y., Recent developments in self-resistance gene directed natural product discovery. *Nat. Prod. Rep.* **2020**, *37*, 879-892.

(21) (a) Tsukada, K.; Shinki, S.; Kaneko, A.; Murakami, K.; Irie, K.; Murai, M.; Miyoshi, H.; Dan, S.; Kawaji, K.; Hayashi,

H.; Kodama, E. N.; Hori, A.; Salim, E.; Kuraishi, T.; Hirata, N.; Kanda, Y.; Asai, T., Synthetic biology based construction of biological activity-related library of fungal decalin-containing diterpenoid pyrones. *Nat. Commun.* **2020**, *11*, 1830; (b) Tagami, K.; Liu, C.; Minami, A.; Noike, M.; Isaka, T.; Fueki, S.; Shichijo, Y.; Toshima, H.; Gomi, K.; Dairi, T.; Oikawa, H., Reconstitution of biosynthetic machinery for indole-diterpene paxilline in *Aspergillus oryzae*. *J. Am. Chem. Soc.* **2013**, *135*, 1260-1263.

(22) Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C. J.; Lu, S.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z.; Yamashita, R. A.; Zhang, D.; Zheng, C.; Geer, L. Y.; Bryant, S. H., CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2016**, *45*, D200-D203.

(23) Tang, J.; Matsuda, Y., Dissection of the catalytic mechanisms of transmembrane terpene cyclases involved in fungal meroterpenoid biosynthesis. *Angew. Chem. Int. Ed.* **2023**, e202306046.

(24) Ozaki, T.; Minami, A.; Oikawa, H., Biosynthesis of indole diterpenes: A reconstitution approach in a heterologous host. *Nat. Prod. Rep.* **2023**, *40*, 202-213.

(25) (a) Matsuda, Y.; Abe, I., Biosynthesis of fungal meroterpenoids. *Nat. Prod. Rep.* **2016**, *33*, 26-53; (b) Matsuda, Y.; Abe, I., 1.14 - Fungal meroterpenoids. In *Comprehensive Natural Products III*, Liu, H.-W.; Begley, T. P., Eds. Elsevier: Oxford, 2020; pp 445-478.

(26) Jin, F. J.; Maruyama, J.; Juvvadi, P. R.; Arioka, M.; Kitamoto, K., Development of a novel quadruple auxotrophic host transformation system by *argB* gene disruption using *adeA* gene and exploiting adenine auxotrophy in *Aspergillus oryzae*. *FEMS Microbiol. Lett.* **2004**, *239*, 79-85.

(27) (a) Ye, Y.; Minami, A.; Mandi, A.; Liu, C.; Taniguchi, T.; Kuzuyama, T.; Monde, K.; Gomi, K.; Oikawa, H., Genome mining for sesterterpenes using bifunctional terpene synthases reveals a unified intermediate of di/sesterterpenes. *J. Am. Chem. Soc.* **2015**, *137*, 11846-11853; (b) Yuan, Y.; Cheng, S.; Bian, G.; Yan, P.; Ma, Z.; Dai, W.; Chen, R.; Fu, S.; Huang, H.; Chi, H.; Cai, Y.; Deng, Z.; Liu, T., Efficient exploration of terpenoid biosynthetic gene clusters in filamentous fungi. *Nature Catalysis* **2022**, *5*, 277-287; (c) Chen, L.; Wei, X.; Matsuda, Y., Depside bond formation by the starter-unit acyltransferase domain of a fungal polyketide synthase. *J. Am. Chem. Soc.* **2022**, *144*, 19225-19230.

(28) Kinoshita, M.; Ohtsuka, M.; Nakamura, D.; Akita, H., First synthesis of (+)-α- and (+)-γ-polypodatetraenes. *Chem. Pharm. Bull.* **2002**, *50*, 930-934.

(29) Bennett, G. J.; Harrison, L. J.; Sia, G.-L.; Sim, K.-Y., Triterpenoids, tocotrienols and xanthones from the bark of *Cratoxylum Cochinchinense*. *Phytochemistry* **1993**, *32*, 1245-1251.

(30) Arai, Y.; Hirohara, M.; Ageta, H.; Hsű, H. Y., Fern constituents: Two new triterpenoid alcohols with mono- and bi-cyclic skeletons, isolated from *Polypodiodes formosana*. *Tetrahedron Lett.* **1992**, *33*, 1325-1328.

(31) Ohtani, I.; Kusumi, T.; Kashman, Y.; Kakisawa, H., High-field FT NMR application of Mosher's method. The absolute configurations of marine terpenoids. *J. Am. Chem. Soc.* **1991**, *113*, 4092-4096.

(32) Bartels, F.; Hong, Y. J.; Ueda, D.; Weber, M.; Sato, T.; Tantillo, D. J.; Christmann, M., Bioinspired synthesis of pentacyclic onocerane triterpenoids. *Chem. Sci.* **2017**, *8*, 8285-8290.

(33) Gachet, M. S.; Kunert, O.; Kaiser, M.; Brun, R.; Zehl, M.; Keller, W.; Muñoz, R. A.; Bauer, R.; Schuehly, W., Antiparasitic compounds from *Cupania cinerea* with activities against *Plasmodium falciparum* and *Trypanosoma bruceirhodesiense*. *J. Nat. Prod.* **2011**, *74*, 559-566.

(34) Ueda, D.; Hoshino, T.; Sato, T., Cyclization of squalene from both termini: Identification of an onoceroid synthase and enzymatic synthesis of ambrein. *J. Am. Chem. Soc.* **2013**, *135*, 18335-18338.

(35) Ageta, H.; Shiojima, K.; Masuda, K., Fern constituents; onoceroid, α-onoceradiene, serratene and onoceranoxide, isolated from *Lemmaphyllum microphyllum* varieties. *Chem. Pharm. Bull. (Tokyo)* **1982**, *30*, 2272-2274.

(36) (a) Tanaka, T.; Ishibashi, M.; Fujimoto, H.; Okuyama, E.; Koyano, T.; Kowithayakorn, T.; Hayashi, M.; Komiyama, K., New onoceranoid triterpene constituents from *Lansium domesticum*. *J. Nat. Prod.* **2002**, *65*, 1709-1711; (b) Abbet, C.; Neuburger, M.; Wagner, T.; Quitschau, M.; Hamburger, M.; Potterat, O., Phyteumosides A and B: New saponins with unique triterpenoid aglycons from *Phyteuma orbiculare* L. *Org. Lett.* **2011**, *13*, 1354-1357.

(37) Ohloff, G., 15 - The fragrance of ambergris. In *Fragrance Chemistry*, Theimer, E. T., Ed. Academic Press: San Diego, 1982; pp 535-573.

(38) (a) Araki, T.; Saga, Y.; Marugami, M.; Otaka, J.; Araya, H.; Saito, K.; Yamazaki, M.; Suzuki, H.; Kushiro, T., Onocerin biosynthesis requires two highly dedicated triterpene cyclases in a fern *Lycopodium clavatum*. *ChemBioChem* **2016**, *17*, 288-290; (b) Saga, Y.; Araki, T.; Araya, H.; Saito, K.; Yamazaki, M.; Suzuki, H.; Kushiro, T., Identification of serratane synthase gene from the fern *Lycopodium clavatum*. *Org. Lett.* **2017**, *19*, 496-499.

(39) (a) Quin, M. B.; Flynn, C. M.; Schmidt-Dannert, C., Traversing the fungal terpenome. *Nat. Prod. Rep.* **2014**, *31*, 1449-1473; (b) Gao, Y.; Wang, G.; Huang, H.; Gao, H.; Yao, X.; Hu, D., Biosynthesis of fungal triterpenoids and steroids. *Chin. J. Org. Chem.* **2018**, *38*, 2335-2347; (c) Zhao, P.; Guan, M.; Tang, W.; Walayat, N.; Ding, Y.; Liu, J., Structural diversity, fermentation production, bioactivities and applications of triterpenoids from several common medicinal fungi: Recent advances and future perspectives. *Fitoterapia* **2023**, *166*, 105470.

(40) Tao, H.; Lauterbach, L.; Bian, G.; Chen, R.; Hou, A.; Mori, T.; Cheng, S.; Hu, B.; Lu, L.; Mu, X.; Li, M.; Adachi, N.; Kawasaki, M.; Moriya, T.; Senda, T.; Wang, X.; Deng, Z.; Abe, I.; Dickschat, J. S.; Liu, T., Discovery of non-squalene triterpenes. *Nature* **2022**, *606*, 414-419.

(41) Chiang, C.-Y.; Ohashi, M.; Tang, Y., Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining. *Nat. Prod. Rep.* **2023**, *40*, 89-127.

(42) Yee, D. A.; Niwa, K.; Perlatti, B.; Chen, M.; Li, Y.; Tang, Y., Genome mining for unknown–unknown natural products. *Nat. Chem. Biol.* **2023**, *19*, 633-640.