Genome sequencing and molecular networking analysis of the wild fungusAnthostomella pineareveal its ability to produce a diverse range of secondary metabolites

Iacovelli, R.; He, T.; Allen, J. L.; Hackl, T.; Haslinger, K.

DOI:
10.1101/2023.10.20.563261

Genome sequencing and molecular networking analysis of the wild fungus *Anthostomella pinea* reveal its ability to produce a diverse range of secondary metabolites

R. Iacovelli[1], T. He[1], J. L. Allen[2], T. Hackl[3], and K. Haslinger[1]

[1]Department of Chemical and Pharmaceutical Biology, Groningen Research Institute of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands

[2]Biology Department, Eastern Washington University, Cheney, Washington 99004, USA

[3]Groningen Institute for Evolutionary Life Sciences, University of Groningen, 9700 CC Groningen, the Netherlands

Correspondence: k.haslinger@rug.nl

Keywords: Fungal genomics, natural products, sesquiterpenes, molecular networking, antibiotics, lichen

Abstract

Background

Filamentous fungi are prolific producers of bioactive molecules and enzymes with important applications in industry. Yet, the vast majority of fungal species remain undiscovered or uncharacterized. Here we focus our attention to a wild fungal isolate that we identified as *Anthostomella pinea*. The fungus belongs to a complex polyphyletic genus in the family of *Xylariaceae*, which is known to comprise endophytic and pathogenic fungi that produce a plethora of interesting secondary metabolites. Despite that, *Anthostomella* is largely understudied and only two species have been fully sequenced and characterized at a genomic level.

Results

In this work, we used long-read sequencing to obtain the complete 53.7 Mb genome sequence including the full mitochondrial DNA. We performed extensive structural and functional annotation of coding sequences, including genes encoding enzymes with potential applications in biotechnology. Among others, we found that the genome of *A. pinea* encodes 91 biosynthetic gene clusters, more than 600 CAZymes, and 164 P450s. Furthermore, untargeted metabolomics and molecular networking analysis of the cultivation extracts revealed a rich secondary metabolism, and in particular an abundance of sesquiterpenoids and sesquiterpene lactones. We also identified the polyketide antibiotic xanthoepocin, to which we attribute the anti–Gram-positive effect of the extracts that we observed in antibacterial plate assays.

Conclusions

Taken together, our results provide a first glimpse into the potential of *Anthstomella pinea* to provide new bioactive molecules and biocatalysts and will facilitate future research into these valuable metabolites.

Background

Fungi are ubiquitous organisms and major components of all ecosystems on Earth, although they more commonly colonize soil and plant tissues. To thrive and adapt to environmental stressors (e.g., extreme temperatures, low water and nutrient availabilities, predators), they can adopt different lifestyles ranging from parasites and opportunistic pathogens to specialized biomass decomposers and mutualistic symbionts [1–3]. The latter include endophytic fungi, that is, fungi that can live inside plant tissues throughout their life cycle and establish a beneficial or neutral relationship with their host, thereby not causing any detrimental effect or disease [4,5]. In exchange for nutrients and a habitat, endophytic symbionts offer several benefits to their host: for example, they can promote growth or protect from biotic (predators, pathogens) or abiotic stressors (UV radiation) [1,6,7]. Generally, they exert these functions through the production of small bioactive molecules called secondary metabolites (SMs), which include polyketides, alkaloids, peptides, terpenoids, and phenolic compounds [8,9]. Often SMs possess biological activities that translate into pharmaceutical applications, such as antimicrobial, antioxidant and anticancer activities [8]. Among the most encountered endophytic fungi are members of the family *Xylariaceae*, which also comprises saprotrophic, pathogenic, and endolichenic (that live inside lichen tissues) fungi [10–12]. Endophytic fungi, and in particular *Xylariaceae*, have been extensively studied in the last decades to discover and exploit new pharmaceutically relevant compounds [4,7,10,13,14]. Despite this, the vast majority of the *Xylariaceae* species studied so far belong to the eponymous genus *Xylaria* and a few others, while many are yet to be characterized [10].

In this work, we isolated a filamentous fungus from sections of dry lichen thalli collected in Cheney, Washington, USA. We identified the fungus as *Anthostomella pinea*, previously described as an epiphytic fungus of pine trees [15]. *Anthostomella* is a large and complex polyphyletic genus that comprises more than 400 species [16] within the *Xylariaceae* family,

66 though little to nothing is known about their genetic makeup and metabolome. To the best of our
67 knowledge, only a few publications report the discovery and characterization of metabolites from
68 *Anthostomella* fungi [17–19], and complete genomic data is currently available for only two species
69 [20,21]. Therefore, we set out to perform whole genome sequencing and untargeted metabolomics
70 analyses to shed light on the ability of the fungus to produce biotechnologically relevant secondary
71 metabolites and enzymes that can be used as biocatalysts in industrial applications. To do so, we
72 employed state-of-the-art long read sequencing technologies that allowed us to obtain a high-
73 quality annotated genome of the fungus. We then used high-resolution tandem mass-
74 spectrometry coupled to molecular networking analysis to analyze the intra- and extra-cellular
75 metabolome. With these approaches, we revealed that *A. pinea* is a prolific producer of
76 sesquiterpenoids and sesquiterpene lactones, and that its genome encodes an abundance of
77 biotechnologically relevant enzymes such as carbohydrate-active enzymes (CAZymes),
78 cytochrome P450s (CYP), and unspecific peroxygenases (UPOs). We also detected antimicrobial
79 activity of the extracts, which we putatively attribute to the anti–Gram-positive compound
80 xathoepocin, produced in small amounts by the fungus. Lastly, based on the genomics data, we
81 hypothesize the biosynthetic route of this compound, which could prove useful in future efforts to
82 reconstitute the pathway in a heterologous host for large scale production and engineering of
83 structural variants.

84

Results and Discussion

Isolation of *Anthostomella pinea* F5 from sections of lichen thallus and morphological characterization

Lichenized fungi, those that form obligate mutualistic symbioses with a photosynthetic partner, are known for their rich SM profiles [22] and equally rich endophytic fungal communities [23]. The wolf lichens, *Letharia*, are a genus of lichens that have received substantial recent research attention due to their unique fungal associates [24,25] and genomic architecture [26]. All wolf lichens produce certain SMs in great abundance, as is evidenced by their bright yellow-green color. Thus, we chose to investigate the biosynthetic potential of one of the most common and widespread wolf lichens in western North America, *Letharia lupina*, and its associates. To obtain axenic fungal cultures, we first collected fresh specimen from a *Pinus ponderosa* growing by the Cheney Wetlands Trail in Cheney, Washington, USA (47.483392, -117.553087), and shipped them at room temperature to the research facilities in Groningen, The Netherlands. Next, we incubated surface-sterilized sections of the lichen thallus in sterile water, based on a modified version of the Yamamoto method [27]. After 4 weeks, filamentous growth was observed on the lichen sections and ten isolates (F1-F10) were excised and transferred to MYA plates: three showed robust growth of white mycelium (F5, F6, and F7), while the other seven showed much poorer growth and dark-green mycelium. From the initial morphological observations and preliminary barcoding with the ITS region, we confirmed that isolates F1-4 and F8-10 were most likely the same species but because of poor ITS alignments and slow growth they were not further investigated (data not shown). Similarly, we determined F5, F6, and F7 to be the same species, and we therefore continued only with isolate F5.

For further morphological characterization we inoculated F5 on different nutrient media, including MEA, PDA, YES, CYA, and DG18, which features a low water activity and is used to grow and characterize xerophilic fungi [28]. Isolate F5 showed robust growth in all the tested media

but CYA, where it stopped growing approximately after 14 days (Fig. 1A). In all cases, it produced white vegetative mycelium which appeared less compact in DG18 and YES, likely attributable to stressful conditions (i.e., low water activity, and poorer availability of nutrients, respectively). Furthermore, F5 secreted yellow extrolites into the medium, which is particularly visible on the reverse side of the plates. On MEA, PDA, and DG18, the isolate showed a strong orange pigmentation at the inoculum site and around the ageing zone, visible both on the observe and reverse sides, but much more prominent on the latter (Fig. 1A). Lastly, extensive guttation was observed on the surface of the colonies on PDA. These observations are a strong indication that the isolate is able to produce and secrete secondary metabolites [29,30] in laboratory conditions.

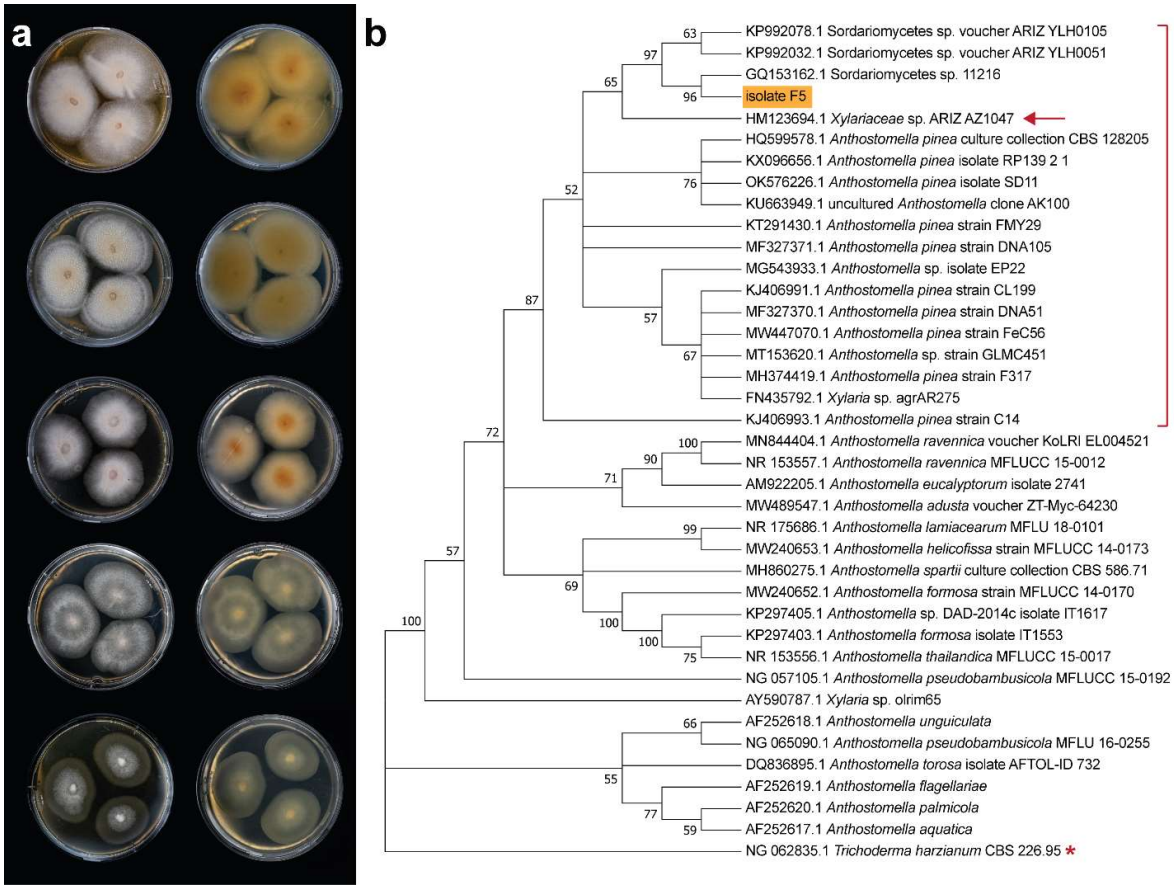


Figure 1. (a) *Morphological characterization of F5 isolate via cultivation on solid medium. From top to bottom: MEA, PDA, DG18, YES, and CYA; obverse (left) and reverse (right). The cultures were incubated for 28 days at 20 °C.* ***(b)*** *Maximum-likelihood phylogenetic tree based on ITS sequences retrieved from the NCBI database. The hypothesized cluster of A. pinea strains is highlighted by the red bracket. Isolate F5 is highlighted in orange, next to isolate AZ1047 which showed almost identical morphology (arrow). T. harzanium (starred) was used as an outgroup to root the tree.*

Next, we extracted genomic DNA from the fungal colonies to amplify the Internally Transcribed Spacer (ITS), a universal barcode region for phylogenetic analysis and taxonomic assignment [31]. We analyzed the manually curated 661 bp-long sequence by blastn with standard settings to search for hits within the kingdom of Fungi. The best 25 hits (ID > 95%) are shown in Table S1. The highest ranked hit for an accepted species was *Anthostomella pinea* strain CBS128205, an ex-type strain isolated from pine needles (Fungal Planet 53, 23 December 2010). Thus, we compared the 25 best hits and all the ITS sequences of other *Anthostomella* species available on NCBI (18 in total) by constructing a maximum-likelihood phylogenetic tree using *Trichoderma harzanium*, a phylogenetically unrelated Sordariomycete, as an outgroup (Fig. 1B). We observed that F5 clustered mainly with different strains of *A. pinea* and with several unidentified Sordariomycete species. Interestingly, F5 showed very similar morphology to isolate AZ1047 which was itself identified as *A. pinea* (Fig S1) [32,33] and for which barcode sequences of the *rbp2* and *act* genes were also available. Sequence alignments with the respective barcodes from F5 (obtained from whole genome sequencing) showed near-perfect matches (Fig S2 and S3), confirming F5 to be an isolate of *A. pinea* closely related to isolate AZ1047. Based on the phylogenetic tree, we hypothesize that the three not further classified Sordariomycete isolates that cluster together with F5 are themselves ascribable to *Anthostomella pinea* (Fig. 1B). Since no additional barcodes are available for these isolates, we cannot confirm this hypothesis.

Whole genome assembly and annotation

Once we determined isolate F5 to be a species of *Anthostomella*, we decided to obtain its full genomic sequence. For that, we extracted high-molecular-weight genomic DNA (Table S2) and subsequently performed long-read sequencing with the new, highly accurate (Q20) chemistry of Oxford Nanopore Technologies (ONT). We generated 1.95 million raw reads (6.12 Gbp) which we filtered to remove all the sequences below 2,000 bp, achieving a final number of 0.8 million

reads for a total of 5.17 Gbp. Despite the reduction in sequence information, both the mean read quality and the N50 value increased: from 15.9 to 17.3 and from 6,949 to 8,291, respectively (Table S3). We assembled the filtered reads into the complete genome of *A. pinea* F5, with a total size of 53.7 Mbp (~80 × coverage). We polished the assembly with Racon and Medaka to produce the final draft of the genome, with a contig N50 value of 5,428,629 and a 97.9% completeness score (BUSCO) [34] (Table 1 and Table S4), indicating that we obtained a highly contiguous and complete genome. Lastly, we used a subset of 10,000 reads to assemble the mitochondrial genome as well, which we fully annotated with the web-based tool GeSeq [35] (Figure 2).

Table 1. Assembly statistics and genome annotations.

Coverage	~80×
# contigs	19
Largest contig	8,924,063
Total length	53,727,737
N50	5,428,629
N90	2,115,983
L50	4
L90	10
GC (%)	51.85
BUSCO (ascomycota)	98.1% (C)
BUSCO (sordariomycetes)	97.9% (C)
Protein coding genes	14,734
rRNA genes	48
tRNA genes	199
Proteins with TM segments	2923
Proteins with SP	2109
Proteins with predicted PFAM family/domain	11,040
CAZymes	608
P450s	164
UPOs	6

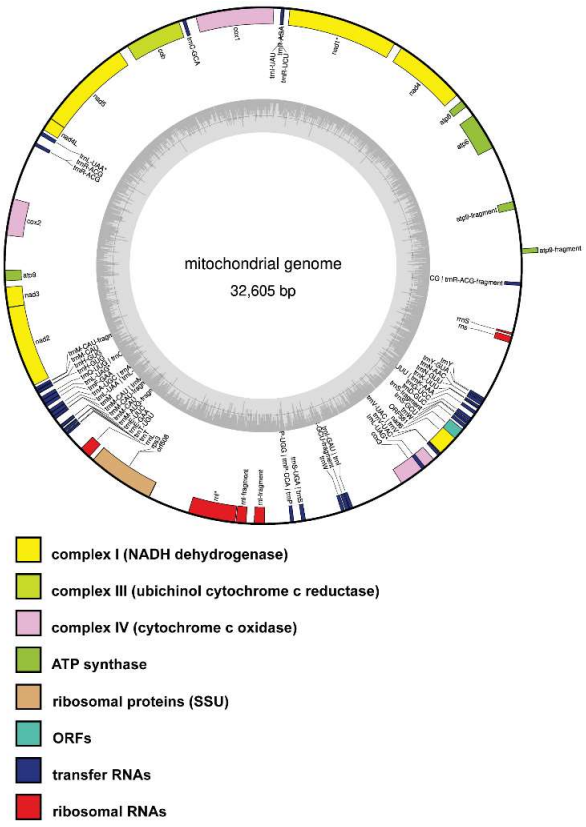


Figure 2. *Complete structural and functional annotation of the assembled mitochondrial genome from A. pinea F5.*

We then proceeded with structural and functional annotation of the genome, which revealed 14,734 protein coding genes, of which 11,040 could be assigned to PFAM domains/families (Additional file 2). Furthermore, we identified 247 RNA genes, 48 for rRNA and 199 for tRNA, respectively (Table 1). Gene Ontology (GO) analysis revealed that the top 50 terms grouped in the following categories: biological processes (24 %), molecular functions (50 %), and cellular components (26 %). Particularly interesting was the abundance of biotechnologically relevant genes: 171 genes were annotated as “secondary metabolite biosynthetic process”, 229 as “monooxygenase activity”, and 221 as “carbohydrate metabolic process” (Table S5 and Fig 3). Thus, we set out to investigate this type of genes more deeply with specific bioinformatic tools.

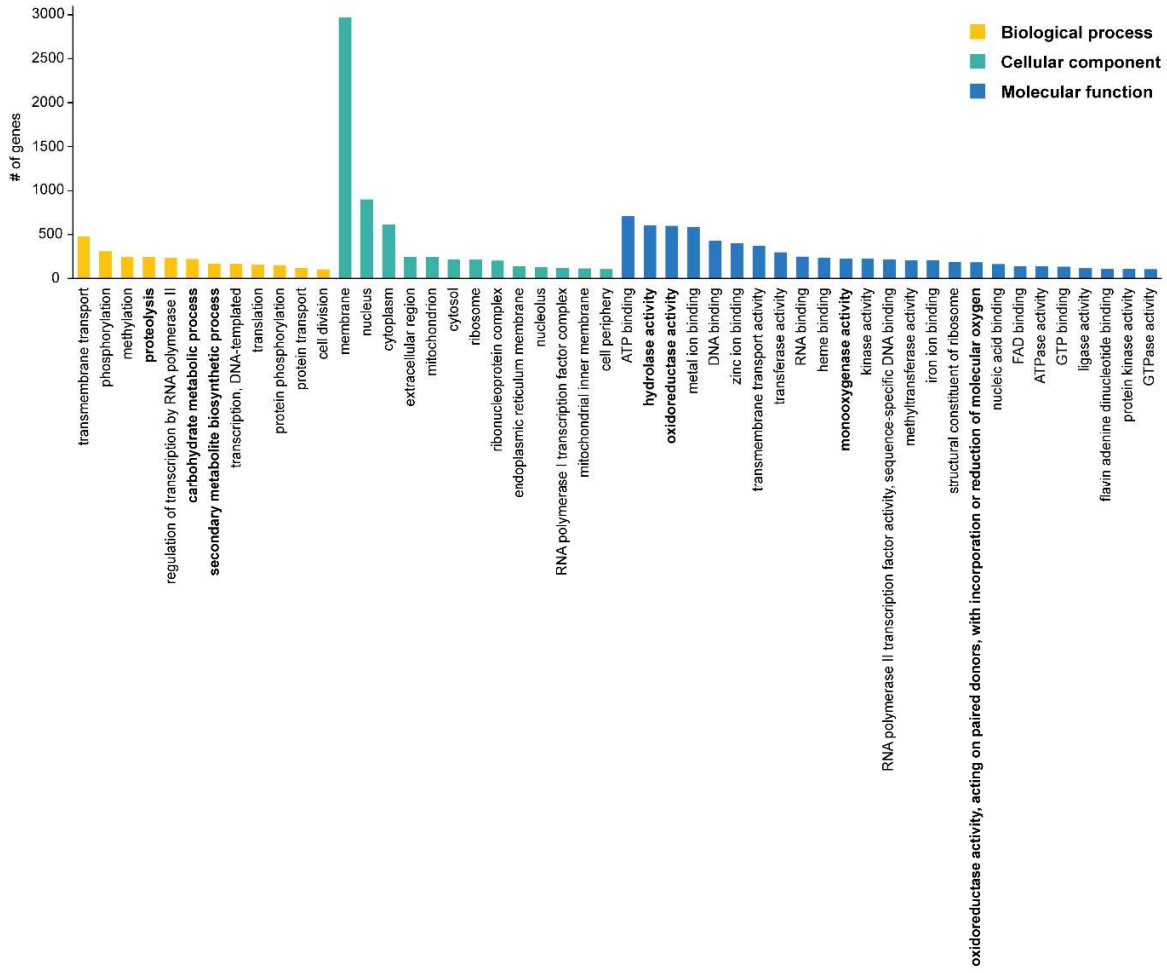


Figure 3. Gene Ontology (GO) analysis of the genome of A. pinea F5 (top 50 terms).

The genome of *A. pinea* F5 encodes a variety of biotechnologically-relevant enzymes

First, we investigated the CAZyme families in the "carbohydrate metabolic process" to assess the ability of *A. pinea* F5 to degrade lignocellulosic material [36]. We used three different tools, HMMMER, DIAMOND, and dbCAN_sub, incorporated into the webserver dbCAN2 [37]. In total, the genome of *A. pinea* F5 encodes 610 CAZymes. Among the six major CAZymes families [38,39] (glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AA), and carbohydrate-binding modules (CBMs), the GH and AA families were most abundant with 281 and 157 members, respectively (Additional file 2). We then examined the substrate specificity of the CAZymes as predicted by dbCAN_sub [37]. In total, *A. pinea* F5 possesses 146 plant biomass-degrading enzymes, with a particularly high number of enzymes active on xylan (35) and cellulose (49) (Table 2). Interestingly, we also found nine CAZymes that are predicted to be active on lichenin (Additional file 2), a complex glucan that occurs in the cell wall of certain lichen-forming fungi when they are associated with their photosynthetic partner [40]. This might suggest that *A. pinea* F5 is indeed able to colonize lichen tissues and grows as an endo/epilichenic fungus.

Table 2. Substrate specificity of CAZymes with predicted plant biomass-degrading activity.

Substrate	Count
Lignin	25
Cellulose	49
Hemicellulose	2
Pectin	20
Xylan	35
Inulin	1
Xyloglucan	7
Galactomannan	7
Total	**146**

Next, we investigated the abundance of cytochromes P450, an important class of enzymes that play essential roles in fungal metabolism. P450s naturally show very broad substrate scopes, and are capable of performing a wide range of reactions involved in the biosynthesis of bioactive molecules as well as in the degradation of pollutants and xenobiotics, and are therefore considered powerful biocatalysts [41,42]. The bioinformatic analysis revealed that the genome of *A. pinea* F5 encodes 164 P450s distributed across 42 families (Additional file 2). The 4 most represented families (≥10 counts) are CYP3 (21), CYP52 (19), CYP570 (12), and CYP65 (10). Interestingly, at least two of these families (CYP3 and CYP65) are involved in the degradation of xenobiotics [43,44], which may suggest that *A. pinea* can act as "detoxifier" for its environment. For other lichens and lichen-associated fungi it is well established that they fulfill this important ecological role by hyperaccumulating xenobiotics, while remaining unharmed [45,46].

Lastly, we looked at another group of enzymes that has recently made the headlines in the fields of biocatalysis and green chemistry: unspecific peroxygenases (UPOs). So far, UPOs have only been found in the fungal kingdom [47,48] and have attracted considerable attention given their simplicity, stability, and broad range of reactions that they can catalyze. These include, among others, epoxidation, aromatic hydroxylation, ether cleavage, and sulfoxidation, demonstrating the versatility and biocatalytic potential of such enzymes [49]. To search for UPOs in the genome of *A. pinea* F5 we performed two pHMMER analyses using two well characterized UPOs as queries: *Aae*UPO from *Agrocybe aegerita* [50], prototype enzyme of family II (long UPOs); and *Hsp*UPO from *Hypoxylon* sp. EC38, member of family I (short UPO) of which the crystal structure was recently published [51]. Both searches produced the same seven putative UPO sequences with a E-value between 0.0016 and 1.30×10^{-77} (Additional file 2). To determine whether these genes might actually encode for UPOs, we performed a multiple sequence alignment of the seven predicted proteins and the two query sequences. All but one showed the typical motifs "PCP" and "ExD" [52] confirming that these proteins are most likely UPOs (Figure

4). Based on length, we can classify three of them as short UPOs (g075280, g022190, g110720) and three as long UPOs (g137960, g143780, g025740).

Overall, by applying tailored bioinformatic approaches we show that the genome of *A. pinea* contains a wide range of enzymes that might have interesting biotechnological applications, including biocatalysis and bioremediation.

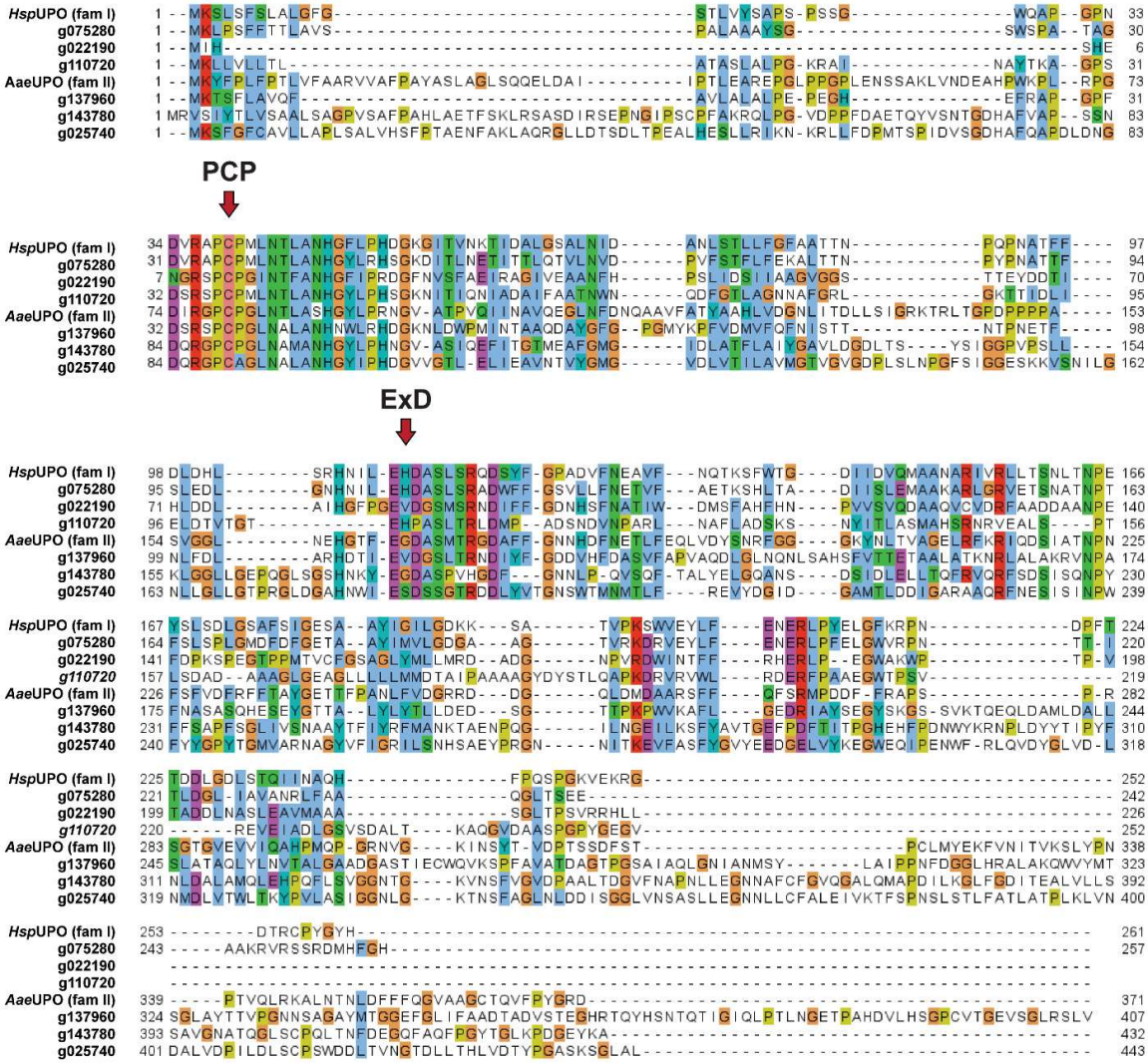


Figure 4. Multiple sequence alignment analysis of putative UPOs from A. pinea F5 and prototype UPOs. *The sequences of family I and family II prototype UPOs were retrieved from the Uniprot database. The conserved motifs "PCP" and "ExD" are highlighted by red arrows. The entry of g137950 was removed from the alignment since it lacked the conserved motifs.*

fungiSMASH predicts an abundance of secondary metabolite biosynthetic gene clusters in the genome of *A. pinea* F5

Fungi are recognized to be prolific producers of secondary metabolites, bioactive molecules with crucial ecological roles and with important applications in medicine and industry [8]. To investigate the secondary metabolism of *A. pinea* F5, we searched its genome for BGCs with the web-based tool fungiSMASH 7 [53]. The software predicted 91 clusters, grouped in the following categories: 23 NRPS (nonribosomal peptide synthetases), 23 fungal RiPP (ribosomally synthesized and post-translationally modified peptides), 22 PKS (polyketide synthases), 14 terpene, 2 indole, and 7 mixed BGCs. Out of these, 71 show no similarities to any known cluster, indicating that they might be involved in unknown biosynthetic pathways (Additional file 3). Among the ones with known similarities, a terpene BGC is predicted to produce the anticancer compound clavaric acid [54], while another terpene BGC and a PKS BGC are predicted to produce the pigments monascorubrin [55] and aurofusarin [56], respectively. These might explain the yellow-to-orange coloration that we observed when performing morphological studies (Fig. 1). Furthermore, *A. pinea* F5 possesses several more BGCs compared to the average for Pezizomycotina (42.8) and the family *Xylariaceae* (71.2) [57]. Such an abundance of BGCs encoded in the genome suggests that this species is likely able to produce a large number of secondary metabolites.

A. pinea F5 is a prolific producer of sesquiterpenoids and sesquiterpene lactones

Despite the large abundance of BGCs, the biosynthesis of secondary metabolites is often regulated and triggered only under specific environmental stimuli [8]. Therefore, we next investigated whether *A. pinea* F5 is actually able to produce secondary metabolites by means of untargeted metabolomics. For that, we grew the fungus in different solid and liquid media for 28 days, and subsequently extracted both extra- and intra-cellular metabolites for LC-MS analysis

(Figure S5). The metabolite profile looked almost identical for the different solid media, while the liquid medium showed some differences. To gain further information about the extracts, we processed the raw MS data with MZmine [58] and performed a feature-based molecular networking (FBMN) analysis using the Global Natural Products Social Molecular Networking platform (GNPS) [59,60]. The network was manually curated by deleting multiple nodes with identical mass to charge ratio (m/z) and near-identical RT, artifacts generated during pre-processing with MZmine, and keeping only the node that showed the highest signal. Nodes present in the blank extracts (sterile media) were also deleted. We colored the nodes by occurrence in the extracts of the different media and adjusted the node size according to the m/z, with bigger nodes representing higher m/z. This resulted in a polished network (Additional file 4) consisting of 299 nodes connected by 422 edges, shown in Fig. 5a. By using the additional GNPS tool MolNet Enhancer [61], three of the clusters were annotated as different types of sesquiterpenoids (STs) and sesquiterpenoid lactones (STLs): namely, guaianes (including dioxanes)(I), germacranolides (II), and eudesmanolides (III) (Fig. 5a) (Additional file 5). Guaianes are bicyclic compounds characterized by a 7-membered ring fused to a 5-membered ring. Compounds from II and III are lactones, with the main difference being that germacranolides have a 10-membered ring fused to the lactone function, while in eudesmanolides the 10-membered ring is fused in the middle resulting in two 6-membered rings [62].

Simultaneously with the FBMN workflow, we also ran an MS/MS spectral library search to identify nodes within the network by direct match with spectral data available on the platform. With that, we were able to annotate several nodes within the ST(L)s groups, as well as a few additional nodes in the network (Fig. 5a and 5b) (Table 3).

260

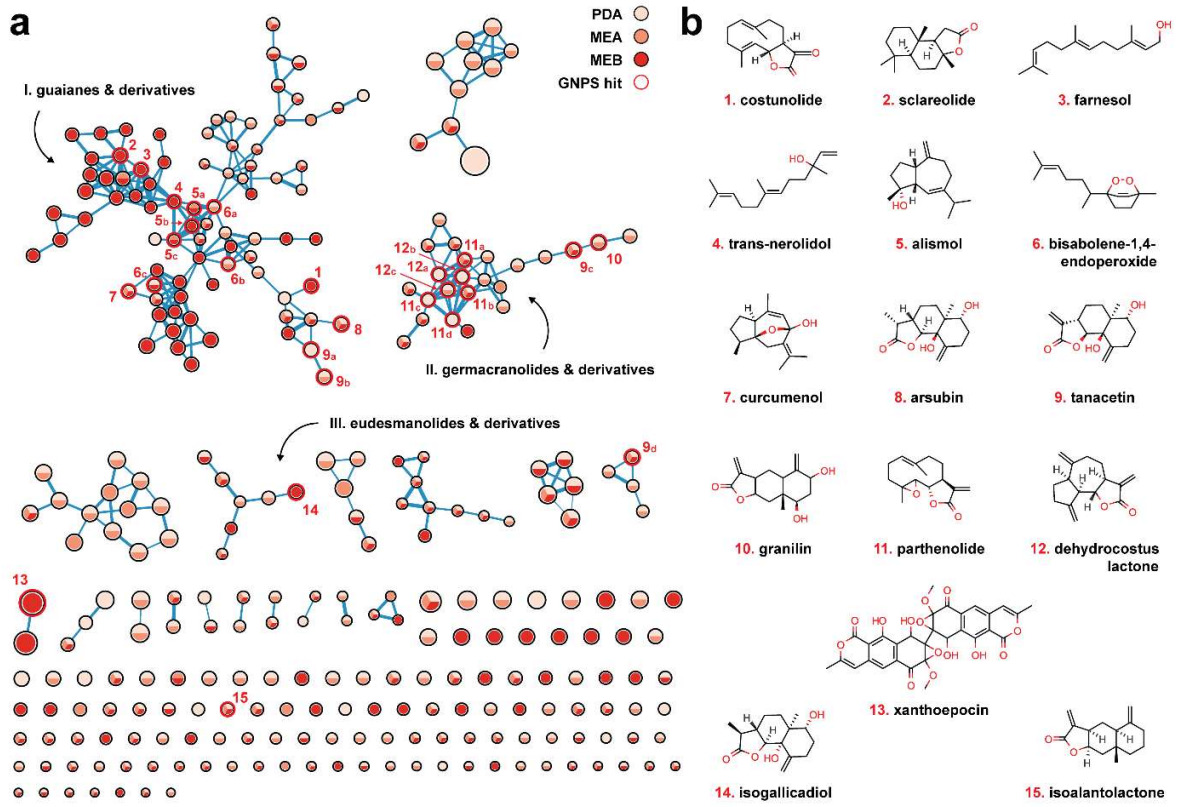


Figure 5. ***(a)*** *Molecular network of extracted secondary metabolites from cultures of A. pinea F5. Nodes are colored based on their occurrence in the extracts from the different media and sized based on the m/z ratio. Hits from the MS/MS GNPS library search are highlighted in red and numbered.* ***(b)*** *Chemical structures of the metabolites identified via library search, numbered correspondingly.*

Table 3. Spectral matches from GNPS MS/MS library search. ST = sesquiterpene; STL = sesquiterpene lactone; PK = polyketide.

Hit #	Compound name	m/z	Adduct	Error (ppm)	Type
1	Costunolide	233.1535	[M+H]	-2.57	Germacranolide STL
2	Sclareolide	251.2005	[M+H]	-2.39	Eudesmanolide STL
3	Farnesol	223.2054	[M+H]	-3.58	Linear ST
4	Trans-nerolidol	205.1949	[M-H_2O+H]	-3.41	Linear ST
5a,b,c	Alismol	203.1798	[M-H_2O+H]	-0.98	Guaiane ST
6a	Bisabolene-1,4-endoperoxide	201.1638	[M-2H_2O+H]	-2.49	Bisabolane ST
6b	Bisabolene-1,4-endoperoxide	219.1732	[M-H_2O+H]	-3.19	Bisabolane ST
6c	Bisabolene-1,4-endoperoxide	237.1851	[M+H]	-1.69	Bisabolane ST
7	Curcumenol	235.1695	[M+H]	-1.28	Guaiane ST
8	Arsubin	267.1590	[M+H]	-2.26	Eudesmanolide STL
9a,b,c,d	Tanacetin	265.1435	[M+H]	-1.89	Eudesmanolide STL
10	Granilin	265.1436	[M+H]	-1.51	Eudesmanolide STL
11a,b	Parthenolide	231.1382	[M-H_2O+H]	-1.30	Germacranolide STL
11c,d	Parthenolide	231.1377	[M-H_2O+H]	-3.46	Germacranolide STL
12a	Dehydrocostus lactone	231.1376	[M+H]	-3.89	Guaianolide STL
12b,c	Dehydrocostus lactone	231.1382	[M+H]	-1.30	Guaianolide STL
13	Xanthoepocin	607.1082	[M+H]	-0.99	Bis-napthopyrone PK
14	Isogallicadiol	267.1593	[M+H]	-1.22	Eudesmanolide STL
15	Isoalantolactone	233.1533	[M+H]	-3.86	Eudesmanolide STL

Surprisingly, the library search revealed that the distinction between the three clusters is not as marked as predicted by MolNet Enhancer, given that several hits found in cluster I (**1**, **2**, **7**, **8**) showed the structure of compounds from II and III, and in cluster II at least one hit (**12**) has a typical guaiane-like structure, though fused to a lactone (guaianolide). An important thing to note is that some of the hits are identified multiple times in the network and library search (**5**, **6**, **9**, **11**, **12**), likely due to the presence of several closely related isomers and derivatives that fragment at the MS^1 level already. This results in different nodes showing identical m/z values and the same MS/MS fragmentation pattern at different RTs (Additional file 7). Because we can't be certain which of the hits is the actual molecule that was identified through spectral match, we cannot propagate the annotation to related unknown nodes and identify them unambiguously.

Interestingly, other nodes in cluster 1 matched to compounds that do not show neither guaiane-type nor eudesmanolide/germacranolide-type structures. These include farnesol and trans-nerolidol (**3**, **4**), linear sesquiterpenoids likely derived from the hydration of farnesyl cation and its isomer nerolidyl cation, respectively. Lastly, bisabolene-1,4-endoperoxide (**5**) was also matched to three nodes in this cluster, a compound so far only observed in plants that has potential anti-tumor activity [63,64]. Another node in cluster I was matched to curcumenol (**6**), a bioactive molecule isolated from the edible rhizome of *Curcuma zedoaria* (white turmeric) which shows potent anti-inflammatory properties [65]. Both compound **5** and **6** have an endoperoxide bridge within their chemical structure, a feature that is typical of many other bioactive compounds [66], including the well-known antimalaria drug artemisinin [67].

Because of the abundance of STs and STLs in the extracts of *A. pinea* F5, we took a closer look at the terpenoid BGCs encoded in its genome. We performed blastP [68] analyses on the terpene cyclases or synthases in each of the 14 terpene BGCs to gain insight into their putative function (Additional file 8). We observed that eight of these enzymes are closely related to ST synthases. Together with the presence of various tailoring enzymes in the respective BGCs (particularly regions 1.3, 12.3 and 14.6), this confirms that *A. pinea* can produce different ST scaffolds and further diversify them into a broad range of products (Fig. 6). Although we cannot conclusively assign all the library hits to the respective matched nodes, it is evident that *A. pinea* produces a wide variety of sesquiterpenoids with different scaffolds and various functional groups. Most of the compounds matched through library search are bioactive and potentially interesting for pharmaceutical applications [69]. Given that the majority of the nodes in the network are unidentified, we can speculate that the fungus is likely producing completely new compounds, which remain to be explored in the future.

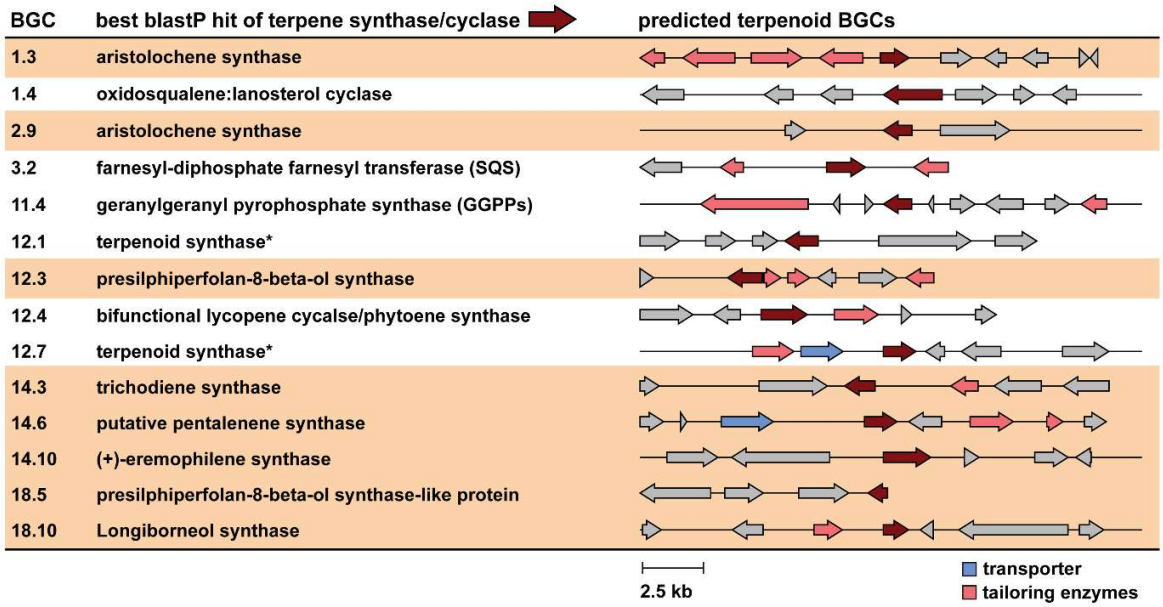


Figure 6. Analysis of terpenoid BGCs as predicted by fungiSMASH v7.0. *Clusters highlighted in tan brown correspond to putative STs and STLs biosynthetic gene clusters, based on blastP analysis of their core terpene synthase/cyclase genes.*

A. pinea F5 produces the antimicrobial compound xanthoepocin in liquid medium

Besides sesquiterpenoids and sesquiterpene lactones, we identified one other node through spectral library search—node **13**, only found in extracts from liquid MEB medium (Fig. 5a and 5b). The feature was matched to xanthoepocin, a polyketide antibiotic originally isolated from *Penicillium simplicissimum* and commonly found in other *Penicillium* species [70,71]. We also found a related node which showed a negative mass difference of 18.011, consistent with loss of a water molecule from xanthoepocin which happens already at MS^1 level (Additional files 4 and 6, Fig 5a). Xanthoepocin has a homodimeric structure composed of two napthopyrone scaffolds, typically associated with pigmentation and sporulation in fungi and with a wide range of biological activities [72–74]. Xanthoepocin in particular has been shown to possess potent antibiotic activity against Gram-positive bacteria, including resistant strains of *E. faecium* and MRSA [71]. Informed by these findings, we performed antimicrobial plate assays with the same extracts that we used for the LC-MS analysis. Indeed, we observed antibiotic activity against the indicator Gram-positive

310 species *Micrococcus luteus* for the MEB extract, albeit only when concentrated 10x (Fig. 7). We
311 also performed agar overlay experiments with 28-days old colonies of *A. pinea* on PDA and MEA

Figure 7. Antibacterial assays of the extracts from A. pinea F5. *Top left: PDA extracts; top right: MEA extracts; bottom: MEB extracts. The assays were carried out on LB media overlayed with cultures of indicator strain M. luteus, grown overnight at 30 °C. Numbers correspond to: 1 - 50% MeOH + FA 0.1%; 2 - Ampicillin 20 µg; 3 - medium extract; 4 - medium extract 10x; 5 - fungal culture extract; 6 - fungal culture extract 10x. Detected antibacterial activity (from extract of A. pinea grown on MEB) is highlighted by a red arrow.*

312 and we observed complete growth inhibition of *M. luteus* under both conditions (Fig. S6). This
313 antibacterial activity may be caused by yet another differentially produced secondary metabolite.
314 However, it is also possible that the fungus produces xanthoepocin on solid media as well, yet we
315 were not able to detect it via LC-MS analysis. This could be caused by a low extraction efficiency,
316 low starting concentration of the compound, and further degradation of the compound with
317 exposure to light [71]. Because of the scarcity of newly discovered antibiotics and its potency
318 against multi-drug resistant bacteria, xanthoepocin has recently regained attention for the potential
319 development of photodynamic antimicrobial therapies [71]. Therefore, we decided to delve deeper
320 into the genomics data to attempt to identify the BGC involved in its biosynthesis.

321

Xanthoepocin is likely biosynthesized by a polyketide BGC that includes a fungal laccase

As discussed in the section above, napthopyrones are widespread across fungi, and the biosynthetic route for many of them is already known [56,72,73]. Generally, the core polyketide structure is synthesized by large, iterative, non-reducing PKS enzymes and further modified by tailoring enzymes (e.g., monooxygenases, or dehydrogenases) to yield the monomeric structure [56]. In the case of bis-naphthopyrones a final coupling step is required to achieve the dimeric structure, which is typically performed by fungal laccases, multicopper oxidases that catalyze intramolecular phenolic coupling via generation of radical intermediates [75,76]. Thus, we analyzed the data from the fungiSMASH prediction searching for a BGC that would encode the necessary enzymes, and successfully identified region 18.3 as the putative xanthoepocin BGC (Fig. 8a, Additional file 3). The cluster encodes all the enzymes predicted to be involved in the biosynthetic route: a NR-PKS (*xaeA*), a flavin-containing monooxygenase (FMO, *xaeO*), an O-methyltransferase (OMT, *xaeM*), an alcohol dehydrogenase (ADH, *xaeD*), and a laccase (*xaeL*). A drug-resistance transporter is present in the region as well (*xaeT*). As a further confirmation, fungiSMASH predicted region 18.3 to be similar to the aurofusarin BGC [56], as well as the BGCs of ustilaginoidins [77] and viriditoxin [78], which are all bis-napthopyrone compounds.

Based on the structural features of xanthoepocin and the genes that we identified in the BGC, we propose a step-by-step biosynthetic route (Fig. 8b). The synthesis of the heptaketide scaffold and its subsequent lactonization catalyzed by the PKS are the first steps of the pathway, while the intramolecular coupling is the last. Aside from those, the succession of the remaining steps is only hypothesized, and it could occur in a different order (amber box). Intermediates **1** and **2** are already described in literature [79] as precursors of another naphtopyrone-derived fungal natural product, cercosporin [79,80]—while compound **4** has been prepared synthetically

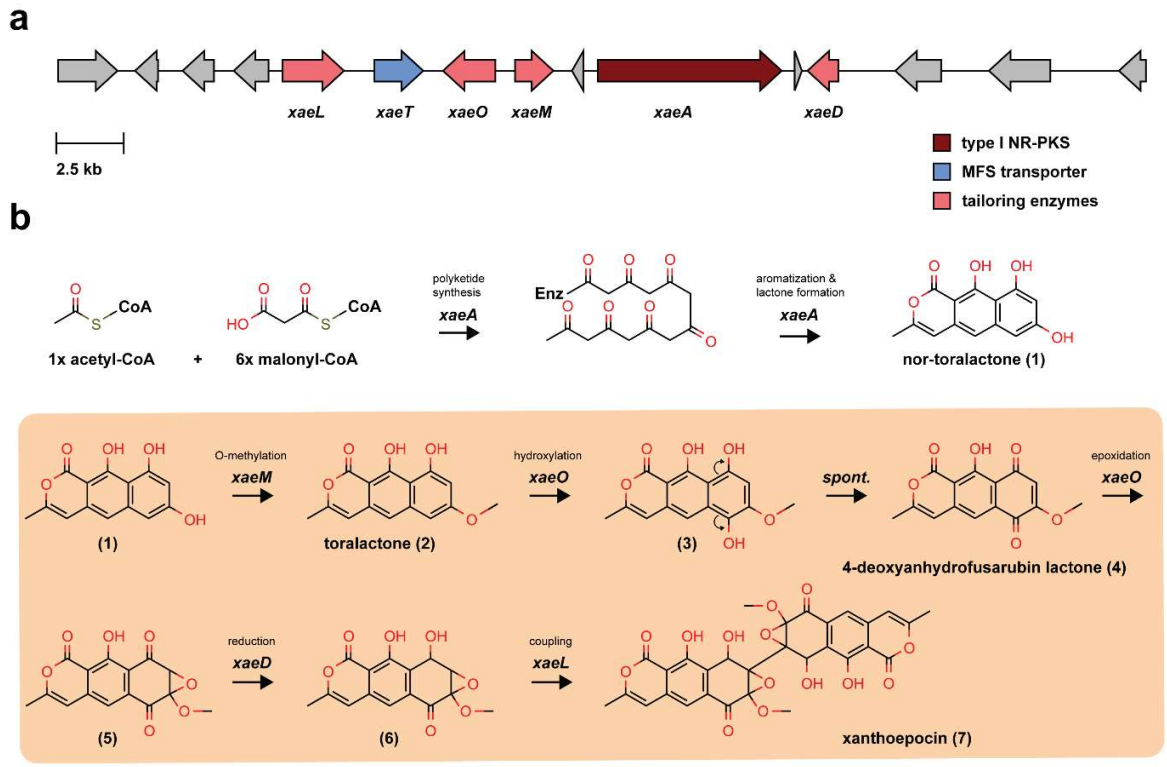


Figure 8. (a) *Putative BGC (ID 18.3) of xanthoepocin as predicted by fungiSMASH v7.0.* ***(b)*** *Proposed biosynthetic pathway of xanthoepocin based on the function of the BGC genes and the chemical structure of the final product. While the biosynthesis of the polyketide scaffold surely occurs as the first step, the order of the remaining steps (brown box) remains to be determined.*

345 [81]. Indeed, bis-napthopyrone precursors and intermediates often show very similar structures
346 that are differentiated during the late steps of the biosynthetic pathways via functionalization,
347 oxidation/reduction, chain shortening, and coupling reactions [74]. The identification of the
348 xanthoepocin BGC should be validated experimentally (e.g., via gene knockout or heterologous
349 expression) as it could prime efforts to reconstitute the pathway in a model host. This would
350 provide access to higher amounts of compound for biological testing, as well as potentially new
351 variants with engineered desirable properties.

352

Conclusions

Fungi are extremely appealing organisms for the discovery of new metabolites and biocatalysts to be used for industrial applications. To date, about 156,000 extant species have been described [82] out of an estimated 2-11 million total species [83]—and only a few thousands genomes have been fully sequenced and are currently available [20,21].

In this work, we report a high-quality whole genome assembly of a wild fungal isolate that we identified as *Anthostomella pinea* based on genetic barcoding. Following structural and functional genome annotation, we used an extensive bioinformatic pipeline to search for enzymes with potential biotechnological applications. First, we identified more than 600 CAZymes, an important group of enzymes with diverse applications in industry [84–87]. We also predicted 164 P450s—powerful biocatalysts that can be used for the regio- and stereospecific oxidation of hydrocarbons [42,88]—and six UPOs, members of a unique family of fungal enzymes that can catalyze a wide range of oxyfunctionalization reactions and that are extremely attractive for their robustness and versatility [47,49,52].

We then explored secondary metabolism in *A. pinea* by combining bioinformatic predictions of biosynthetic gene clusters and molecular networking analyses [53,59,60]. Our investigation revealed that the fungus possesses a rich biosynthetic machinery and can produce a wide variety of small molecules, particularly sesquiterpenoids and sesquiterpene lactones. Through spectral library matches, we putatively identified 14 sesquiterpenes and sesquiterpene lactones and one bis-napthopyrone polyketide with antibiotic activity which is currently being re-evaluated for biological testing [71]. For the latter, we also identified the putative BGC and proposed a step-by-step biosynthetic mechanism based on its chemical structure and the predicted activities of the BGC enzymes.

Overall, our compressive analysis showcases the rich biosynthetic capacity of the fungus *A. pinea*, a species thus far underexplored. Our findings pave the way to explore targeted approaches for the discovery and production of bioactive molecules, such as genome editing and heterologous expression of genes and BGCs of interest. Lastly, this work may prompt further investigation into the genomic and metabolic diversity of phylogenetically related fungi.

Materials and Methods

Sampling of lichen thalli

One large thallus of *Letharia lupina* was collected along the Cheney Wetland Trail, Cheney, Spokane County, Washington, USA (47.483392, -117.553087). The lichen was growing on a *Pinus ponderosa* branch along the edge of a pond and appeared robust and healthy (that is, lacking large galls, perithecia, or discoloration that would indicate parasite infection or tissue necrosis). Collection was completed with a sterile nitrile glove and the thallus was placed in a paper bag. After drying in the bag at room temperature on the lab bench in the original collection bag for ~48 hours it was shipped to Groningen, The Netherlands.

Fungal isolation and preliminary barcoding

The specimen of *A. pinea* was isolated from section of dry thalli of the lichen *Letharia lupina* using a modified version of the Yamamoto method [27]. Briefly, a sample of dry thallus was washed three times with sterile ddH_2O under a biosafety cabinet, after which it was damp-dried with sterile cheesecloth and cut into ~1 cm long sections using a metal scalpel. These sections were incubated in sterile ddH_2O for 4 weeks in the dark at room temperature (18-23 ˚C) before being inspected under a stereoscopic microscope. When filamentous growth was observed at

either end of one of the sections, this was excised and placed on Malt-Yeast Extract Agar (MYA) plates (malt exctract 20 g/L; yeast extract 2 g/L; microagar 15 g/L). In total, 10 isolates were obtained from as many growth points. The plates with the isolates were incubated at room temperature in the dark for up to six weeks to allow for sufficient growth of all the isolates.

To obtain template DNA for preliminary ITS barcoding, small sections of mycelia of approximately 4 mm^2 were scraped from the growing edges of the isolates and placed in 20 µL of DNA dilution buffer (Tris-HCl 10 mM; NaCl 10 mM, EDTA 1 mM, pH = 7.5) in 0,2 mL PCR tubes. The suspensions were then incubated at 95 ˚C for 10 minutes immediately followed by 2 minutes of incubation on ice, after which they were centrifuged at max speed for 1 minute to pellet the mycelia. The supernatant was transferred to clean tubes for storage, and 1 µL was used for direct colony PCR with with the ITSF1 (5′-CTTGGTCATTTAGAGGAAGTAA-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) primers. The PCR reaction was performed in a thermal cycler with 2x Plant Phire Master Mix (ThermoFisher Scientific, Waltham, MA, United States) with the following program: 5 min at 98 °C; 34 cycles of 10 s at 98 °C, 10 s at 60 °C, 40 s at 72 °C; 5 min at 72 °C. Two microliters of the PCR product were taken for 1% agarose gel electrophoresis analysis to confirm successful amplification of the ITS region. The PCR product was then purified using the QIAquick PCR Purification Kit (Qiagen, Venlo, the Netherlands) and sent to Macrogen Europe (Amsterdam, the Netherlands) for Sanger sequencing. The resulting sequences were analyzed with the Basic Local Alignment Search Tool (BLAST) [89] against the nucleotide collection of the National Center for Biotechnology Information (NCBI) to identify the best match for the fungal isolate based on %ID, coverage, and E-value. The original plates with isolates F5-7 (putatively identified as *A. pinea*) were stored at 4 ˚C and used for inoculation of fresh medium.

ITS-based species assignment

For species assignment, we extracted genomic DNA from 40 mg of mycelium freshly scraped from an agar plate, using the Nucleospin Microbial DNA kit (Bioké, Leiden, the Netherlands) in combination with NucleoSpin Bead Tubes Type C (Bioké, Leiden, the Netherlands) for tissue disruption in a MM 301 vibratory mill (Retsch GmbH, Haan, Germany). The extraction was carried out according to manufacturer's instructions except for the disruption time, where 2 cycles of 1 min each were used. Next, a PCR reaction was performed with 1 µL of extracted gDNA and the primers ITSF1 and ITS4, in a thermal cycler with a 2 × Q5 PCR master mix (New England Biolabs, Ipswich, MA, USA). The following program was used: 1 min at 98 °C, 30 cycles of 10 s at 98 °C, 15 s at 55 °C, 20 s at 72 °C, 5 min at 72 °C. The PCR product was visualized on agarose gel, purified, sequenced, and analyzed with BLAST as describe above. The 25 best hits and the ITS sequences of other 18 *Anthostomella* strains obtained from the NCBI nucleotide database were compared by constructing a maximum-likelihood phylogenetic tree using MEGA 11 [90]. First, the sequences were aligned with the MUSCLE algorithm with default settings [91], then the tree was built using standard settings and number of bootstrap replications of 100.

Morphological characterization

For morphological characterization, *A. pinea* was inoculated on malt extract agar (MEA: malt etraxct 30 g/L; peptone 5 g/L; microagar 15 g/L), potato dextrose agar (PDA, 39 g/L)), yeast extract sucrose agar (YES: yeast extract 4 g/L; sucrose 20 g/L; KH_2PO_4 1 g/L; $MgSO_4 \cdot 7H_2O$ 0,5 g/L; microagar 15 g/L), Czapek yeast autolysate agar (CYA: sucrose 30 g/L; yeast extract 5 g/L; $NaNO_3$ 3 g/L; K_2HPO_4 1 g/L; KCl 0,5 g/L; $MgSO_4 \cdot 7H_2O$ 0,5 g/L; $FeSO_4 \cdot 7H_2O$ 0,01 g/L; microagar 15 g/L), and dichloran-glycerol medium (DG18: peptone 5 g/L; glucose 10 g/L; KH_2PO_4 1 g/L; $MgSO_4 \cdot 7H_2O$ 0,5 g/L; glycerol 180 g/L; dichloran 0,002 g/L). The plates were incubated at 20 ˚C and the

448 growth of *A. pinea* was monitored regularly until the colonies stopped growing after 4 weeks, at
449 which points the plates were photographed using a Nikon D7500 camera (Nikon, Minato City,
450 Tokyo, Japan) coupled to a Sigma 17-50mm F/2.8 EX DC OS (Sigma Corporation, Kawasaki,
451 Kanagawa, Japan). The colonies were incubated for further 8 weeks to assess whether any
452 additional morphological change would occur, but none was observed. For morphological
453 comparisons (Fig. S1), *A. pinea* CBS128205 was purchased from the strain collection of the
454 Westerdijk Institute of Fungal Biodiversity (Utrecht, the Netherlands). Both isolate F5 and *A. pinea*
455 CBS128205 were grown on MEA for 4 weeks at 20 ˚C, before being imaged as described above.

456

457 **Extraction of HMW gDNA**

458 The fungus was grown in 25 mL malt extract broth (MEB: malt etraxct 30 g/L; peptone 5 g/L)
459 at 20 ˚C in the dark in static conditions for 28 days. The mycelium was then collected by filtration
460 with sterile Miracloth (MilliporeSigma, Burlington, MA, USA), washed with ~20 mL sterile Milli-Q
461 water, snap-frozen in liquid N_2, and lyophilized overnight using a Lyovapor L-200 (Buchi AG,
462 Flawil, Switzerland). The dried biomass was finally stored at −20 °C.

463 DNA extraction was carried out with 37 mg of freeze-dried mycelium using the Nucleospin
464 Microbial DNA kit in combination with 3 mm tungsten carbide beads (Qiagen) for tissue disruption
465 in a MM 301 vibratory mill, with significant modifications to the manufacturer's protocol. Briefly,
466 disruption time was reduced to two cycles of 5 s to minimize DNA shearing, with 30 s pause in
467 between to allow the sample to cool down; each vortexing step was replaced by gentle flicking or
468 mixing by inversion; washing steps with buffer BW was repeated four times, followed by four
469 washes with buffer B5; elution was carried out in 60 µL of pre-warmed EB buffer (~55 ˚C), and the
470 eluate was then re-applied to the column, incubated for one additional minute and re-eluted to
471 maximize DNA recovery. The integrity of the purified DNA was assessed by agarose gel
472 electrophoresis, while quality control prior sequencing was performed using NanoDrop ND-1000

(ThermoFisher, Waltham, MA, USA), Qubit 3.0 (Invitrogen, Waltham, MA, USA), and the Qubit dsDNA HS Assay Kit (Invitrogen).

Library preparation and sequencing

For long-read sequencing, the genomic DNA was prepared using the ligation sequencing kit SQK-LSK112 (Oxford Nanopore Technologies, Oxford, United Kingdom) according to the manufacturer's guidelines. Briefly, genomic DNA (1020 ng) was subjected to end repair, 5' phosphorylation and dA-tailing by NEBNext FFPE DNA Repair and NEBNext Ultra II End prep modules (New England Biolabs) and purified with AMPure XP (Beckman Coulter, Pasadena, CA, USA) magnetic beads. The sequencing adaptors were ligated using the NEB Quick T4 DNA Ligase (New England Biolabs) in combination with ligation buffer (LNB) from the SQK-LSK112 kit. The library was finally cleaned up using the long fragment buffer (LFB) and purified with AMPure XP magnetic beads. For sequencing, 12 µL of library (~400 ng) were loaded into a primed FLO-MIN112 (ID: FAT29688) flow cell on a MinION device for a 42-hour run. Data acquisition was carried out with MinKNOW software v22.05.5 (Oxford Nanopore Technologies).

Read processing and whole genome assembly

The raw reads were basecalled using Guppy v6.0.7 (Oxford Nanopore Technologies) in GPU mode using the dna_ r10. 4_ e8.1_sup.cfg model (super accuracy). The basecalled reads were subsequently filtered to a minimum length of 2 kb and trimmed by 20 nt at both ends using NanoFilt v2.8.0 [92]. NanoPlot v1.40.0 [92] was used to evaluate the filtered reads. Assembly was performed using Flye v2.9-b1778, using the parameters *--nano-hq and --read-error 0.03*, optimized for Q20+ chemistry and super accurate basecalling. The quality of the genome assembly was evaluated using QUAST v5.1.0rc1 [93] and Bandage v0.8.1 [94] (Fig. S4). The draft

497 assembly was subsequently polished in two rounds. First, Racon v1.4.10 [95] was used with the
498 following settings: *-m 8 -x -6 -g -8 -w 500*, optimized for combined use with Medaka. Second,
499 Medaka 1.6.0 was used on the polished version with the r104_e81_sup_g5015 model. The
500 completeness of assemblies was evaluated using BUSCO v5.3.2 (ascomycota_odb10 and
501 sordariomycetes_odb10 datasets) [34]. To assemble the mitochondrial DNA, we subsampled the
502 filtered reads to generate a set of 10,000 reads and then used Flye v2.9-b1778 as described
503 above. The only circular contig obtained was analyzed with BLAST against the NCBI nucleotide
504 collection to confirm its organellar origin. Lastly, the mitochondrial genome was annotated using
505 GeSeq [35] using default settings.

506 The complete sequencing data and genome assembly for this study have been deposited
507 in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB67537.

508

509 **Structural and functional genome annotation**

510 Genome annotation was carried out on the polished assembly using the online platform
511 Genome Sequence Annotation Server v6.0, which provides a pipeline for whole genome structural
512 and functional annotation [96]. Standard settings were used unless otherwise mentioned. In brief,
513 low complexity regions and repeats were masked using RepeatModeler v2.0.3 and RepeatMasker
514 v4.1.1 [97], setting the DNA source to 'Fungi'. The newly generated masked consensus sequence
515 was used for *ab initio* gene prediction using the following tools: (I) Augustus v3.4.0 [98], selecting
516 *Fusarium graminearum* as a trained organism; (II) GeneMarkES v4.48 [99]. For homology-based
517 prediction, the NCBI reference transcript and protein databases for Fungi were searched, using
518 (III) blastn v2.12.0 [89] and (IV) DIAMOND v2.0.11[100], respectively. To generate the final
519 consensus gene model EvidenceModeler v1.1.1 [101] was used on the above-mentioned
520 predictions, weighted as follows: (I)–five, (II)-five, (III)-ten, (IV)-ten. For prediction of proteins with
521 PFAM domains, the Pfam module within GenSAS was used with the following parameters: E-

522 value sequence: 1; E-value Domain: 10. The tRNA and rRNA were predicted using tRNA scan-
523 SE v2.0.11 ("check for pseudogenes = OFF") [102] and barrnap v0.9 [103], respectively.
524 Prediction of proteins with signal peptides and/or transmembrane domains was carried out on the
525 web server Phobius [104]. Gene Ontology (GO) annotation was performed using the webtool
526 PANNZER2 [105].

527

528 **Prediction of secondary metabolites BGCs, CAZymes, P450s, and UPOs**

529 Secondary metabolite biosynthetic gene clusters were identified using the fungal suite of
530 antiSMASH web server v7.0 with the default settings [53]. To annotate CAZymes, the web server
531 dbCAN2 was used [37]. The integrated HMMER, DIAMOND, and HMMER-dbCAN-sub tools were
532 used on the total proteome of *A. pinea*. The three outputs were automatically combined, and
533 CAZymes predicted only by 1/3 of the tools were removed to improve the annotation accuracy.
534 The substrate specificity of the final hits was extracted from the individual results of HMMER-
535 dbCAN-sub.

536 For the prediction of P450s, the BLAST tool of biocatnet CYPED 6.0 [106] was used, with
537 E-value cutoff set at 1.0×10^{-10}. To identify putative UPOs, the query sequences of the prototype
538 enzymes AaeUPO [50] and HspUPO [51] were retrieved from the Uniprot database [107], and
539 submitted to phmmer (HMMER v3.3.2) to search against the total proteome of *A. pinea*. The cutoff
540 was set at E-value of 0.01. The multiple sequence alignment analysis between the putative UPOs
541 from *A. pinea* and the prototype AaeUPO and HspUPO was performed with MEGA 11 [90] using
542 the MUSCLE algorithm [91] with default settings, and visualized in Jalview [108].

543

Extraction of secondary metabolites and HRMS-MS^2 analysis

Fungal mycelium was transferred from storage plates to PDA or MEA plates. For inoculation in liquid MEB, mycelium of *A. pinea* was scraped from a storage plate, coarsely ground into 1 mL of MEB in an 1.5 mL microcentrifuge tube with a pipette tip, and then transferred to 25 mL of medium. The plates and flasks were incubated at 20 °C for 28 days alongside empty PDA, MEA plates and MEB-containing flasks to be used as controls. For extraction of SMs, the whole agar pads (agar and mycelium) were cut into pieces and transferred to 50 mL polypropylene tubes, then extracted twice with 25 mL of 9:1 ethyl acetate-methanol (v/v) with 0.1% formic acid and sonicated in a sonication bath for one hour for each extraction. For the liquid cultivations, the entire content of the flasks was collected in 50 mL PP tubes and extracted as above. Prior extraction, all samples were spiked with 10 µL caffeine standard solution with a concentration of 10 mg/mL to validate the extraction procedure. The dried extracts were resuspended in 1 mL of 1:1 MeOH-MilliQ water (v/v) supplemented with 0.1% formic acid (FA), filtered with 0.45 µm PTFE filters, and stored at -20 ˚C until further analysis.

HR-LC-MS/MS analysis was performed with a Shimadzu Nexera X2 high performance liquid chromatography (HPLC) system with binary LC20ADXR coupled to a Q Exactive Plus hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). A Kinetex EVO C18 reversed-phase column was applied for HPLC separations (100 mm × 2.1 mm I.D., 2.6 µm, 100 Å particles, Phenomenex, Torrance, CA, USA), which was maintained at 50 °C. The mobile phase consisted of a gradient of solution A (0.1% formic acid in MilliQ water) and solution B (0.1% formic acid in acetonitrile). A linear gradient was used: 0–3 min 5% B, 3–51 min linear increase to 90% B, 51–55 min held at 90% B, 55–55.01 min decrease to 5% B, and 55.01–60 min held at 5% B. The injection volume was 2 µL, and the flow was set to 0.25 mL·min−1. MS and MS/MS analyses were performed with electrospray ionization (ESI) in positive mode at a spray voltage of 3.5 kV, and sheath and auxiliary gas flow set at 60 and 11, respectively. The ion transfer

tube temperature was 300 °C. Spectra were acquired in data-dependent mode with a survey scan at m/z 100–1500 at a resolution of 70,000, followed by MS/MS fragmentation of the top 5 precursor ions at a resolution of 17,500. A normalized collision energy of 30 was used for fragmentation, and fragmented precursor ions were dynamically excluded for 10 s.

Data processing and molecular networking analysis

The raw MS/MS data files were converted to mzML format using the format conversion utility MSConvert from the ProteoWizard suite [109]. Binary encoding precision was set at 32-bit and zlib compression was set as off. Data was centroided on both MS^1 and MS^2 level using the peak picking filter (algorithm set to "Vendor"). The data files were then pre-processed with MZMine v2.53 [58] to generate a .csv feature list and MS/MS .mgf file. The converted MS/MS data, feature list, and the .mgf files were subsequently uploaded to GNPS [59] using the WinSCP tool [110].

A molecular network was generated using the FBMN workflow from the GNPS website [60], version release 28.2. The precursor ion mass tolerance and MS/MS fragment ion tolerance were set to 0.01 Da and 0.02 Da, respectively. Edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Further, edges between two nodes were kept in the network if each of the nodes appeared in the other's respective top 10 most similar nodes (task ID: gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9d0fe25c59bc4efc876c9a638807a05d, generated on 6 March 2023) (Additional file 4). All mass spectrometry data have been deposited on GNPS under the accession number MassIVE ID: MSV000093120. The molecular network was visualized and curated in Cytoscape version 3.9.1 [111]. Briefly, for nodes with identical m/z ratio and near-identical RT (≤ 0.1 min), only the node with the highest signal intensity and peak area was kept. Nodes occurring in the blanks (PDA, MEA, MEB extracts) were considered background and omitted from the polished network. The network was submitted to the GNPS tool MolNetEnhancer [61] version release 22 to annotate compound families, with default settings (task ID:

gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ffb0040546004a119b34be5dc2869aaa, generated on 6 March 2023) (Additional file 5).

The spectra in the network were then searched against the GNPS spectral libraries. Matches were kept with a score above 0.7 and at least six matched peaks. The data was also analyzed by the GNPS molecular library search V2 pipeline, version release 28 [112]. The precursor ion mass tolerance and fragment ion tolerance were set to 0.01 and 0.02 Da, respectively. The minimal matched peaks were set to eight, and the score threshold was set to 0.7 (task ID: gnps.ucsd.edu/ProteoSAFe/status.jsp?task=14dff366901b437394e4d0feae71ff5d, generated on 7 July 2023). Several matched annotations in the library search mode were manually added to the molecular network (additional files 4 and 6). Lastly, the MS/MS data (.mgf file) were processed with the bioinformatic tool Sirius v5.7.1 for compound annotation [113]. The predictions of the chemical formula for each node are included in the molecular network (additional file 4).

Antimicrobial plate assays

For the whole-colony plate assays, *A. pinea* was inoculated in the center of MEA or PDA plates and grown at 20 ˚C for 28 days in the dark. *M. luteus* ATCC 10240 was grown overnight at 30 ˚C, 200 rpm, in 5mL 2x yeast extract tryptone medium (2x YT: tryptone 16 g/L, yeast extract 10 g/L, NaCl 5 g/L). The overnight culture was then diluted 1:100 in warm (~60 ˚C) 50 mL 2x YT supplemented with 0.6% agar, mixed well and immediately overlayed on the agar plates (three mL each) with the colonies of *A. pinea* and corresponding controls. The plates were incubated overnight and inspected for antibacterial activity.

To assay the antimicrobial activity of the extracts, *M. luteus* ATCC 10240 was grown overnight and diluted 1:100 in 2x YT + 0.6% agar as described above. Three mL were immediately plated on LB plates (peptone 10 g/L, NaCl 10 g/L, yeast extract 5 g/L, microagar 15 g/L) and left to dry for ~15 minutes. Six sterile diffusion disks (10mm ø, MilliporeSigma, Burlington, MA, USA)

were then gently positioned on the surface of the agar. Each disk was soaked with either 10 µL of fungal extract, fungal extract 10x, medium extract, medium extract 10x, 50% MeOH + FA 0.1% (blank solvent), and ampicillin 2 mg/mL (20 µg) as positive control. The plates were incubated overnight and inspected for antibacterial activity.

References

1. Redkar A, Sabale M, Zuccaro A, Di Pietro A. Determinants of endophytic and pathogenic lifestyle in root colonizing fungi. Curr Opin Plant Biol [Internet]. 2022 Jun;67:102226. https://doi.org/10.1016/j.pbi.2022.102226

2. Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. Fungal lifestyle reflected in serine protease repertoire. Sci Rep [Internet]. 2017 Aug 22;7(1):9147. http://dx.doi.org/10.1038/s41598-017-09644-w

3. Boddy L, Hiscox J. Fungal Ecology: Principles and Mechanisms of Colonization and Competition by Saprotrophic Fungi. In: The Fungal Kingdom [Internet]. Washington, DC, USA: ASM Press; 2017. p. 293–308. http://doi.wiley.com/10.1128/9781555819583.ch13

4. Wen J, Okyere SK, Wang S, Wang J, Xie L, Ran Y, et al. Endophytic Fungi: An Effective Alternative Source of Plant-Derived Bioactive Compounds for Pharmacological Studies. J Fungi [Internet]. 2022 Feb 20;8(2):205. https://www.mdpi.com/2309-608X/8/2/205

5. Sagita R, Quax WJ, Haslinger K. Current State and Future Directions of Genetics and Genomics of Endophytic Fungi for Bioprospecting Efforts. Front Bioeng Biotechnol [Internet]. 2021 Mar 15;9(March). https://www.frontiersin.org/articles/10.3389/fbioe.2021.649906/full

6. Galindo-Solís JM, Fernández FJ. Endophytic Fungal Terpenoids: Natural Role and Bioactivities. Microorganisms [Internet]. 2022 Feb 1;10(2):339. https://www.mdpi.com/2076-2607/10/2/339

7. Manganyi MC, Ateba CN. Untapped Potentials of Endophytic Fungi: A Review of Novel Bioactive Compounds with Biological Applications. Microorganisms [Internet]. 2020 Dec 6;8(12):1934. https://www.mdpi.com/2076-2607/8/12/1934

8. Keller NP. Fungal secondary metabolism: regulation, function and drug discovery. Nat Rev Microbiol [Internet]. 2019 Mar 10;17(3):167–80. http://dx.doi.org/10.1038/s41579-018-0121-1

9. Pusztahelyi T, Holb IJ, PÃ³csi I. Secondary metabolites in fungus-plant interactions. Front Plant Sci [Internet]. 2015 Aug 6;6(AUG):1–23. http://journal.frontiersin.org/Article/10.3389/fpls.2015.00573/abstract

650 10. Becker K, Stadler M. Recent progress in biodiversity research on the Xylariales and their secondary
651 metabolism. J Antibiot (Tokyo) [Internet]. 2021 Jan 23;74(1):1–23. http://dx.doi.org/10.1038/s41429-
652 020-00376-0

653 11. Cañón ERP, de Albuquerque MP, Alves RP, Pereira AB, Victoria F de C. Morphological and
654 Molecular Characterization of Three Endolichenic Isolates of Xylaria (Xylariaceae), from Cladonia
655 curta Ahti & Marcelli (Cladoniaceae). Plants [Internet]. 2019 Oct 8;8(10):399.
656 https://www.mdpi.com/2223-7747/8/10/399

657 12. Tang AMC, Jeewon R, Hyde KD. A re-evaluation of the evolutionary relationships within the
658 Xylariaceae based on ribosomal and protein-coding gene sequences. Fungal Divers. 2009;34:127–
659 55.

660 13. Amirzakariya BZ, Shakeri A. Bioactive terpenoids derived from plant endophytic fungi: An updated
661 review (2011–2020). Phytochemistry [Internet]. 2022 May;197(February):113130.
662 https://doi.org/10.1016/j.phytochem.2022.113130

663 14. Gupta S, Chaturvedi P, Kulkarni MG, Van Staden J. A critical review on exploiting the pharmaceutical
664 potential of plant endophytic fungi. Biotechnol Adv [Internet]. 2020 Mar;39(May 2019):107462.
665 https://doi.org/10.1016/j.biotechadv.2019.107462

666 15. Crous PW, Groenewald JZ. Anthostomella pinea. Fungal Planet 53. Persoonia. 2010;2010(24):126–
667 7.

668 16. The Index Fungorum. http://indexfungorum.org. Accessed on 16 October 2023.

669 17. Izumikawa M, Itoh M, Kawahara T, Sakata N, Tsuchida T, Mizukami T, et al. A highly oxygenated
670 ergostane—MBJ-0005—from Anthostomella eucalyptorum f25427. J Antibiot (Tokyo) [Internet].
671 2014 Dec 25;67(12):843–5. https://www.nature.com/articles/ja201482

672 18. Daranagama DA, Camporesi E, Jeewon R, Liu X, Stadler M, Lumyong S, et al. Taxonomic
673 Rearrangement of Anthostomella (Xylariaceae) Based on a Multigene Phylogeny and Morphology.
674 Cryptogam Mycol [Internet]. 2016 Dec;37(4):509–38.
675 http://www.bioone.org/doi/10.7872/crym/v37.iss4.2016.509

19. Anderson JR, Edwards RL, Whalley AJS. Metabolites of the higher fungi. Part 22. 2-Butyl-3-methylsuccinic acid and 2-hexylidene-3-methylsuccinic acid from xylariaceous fungi. J Chem Soc Perkin Trans 1 [Internet]. 1985;34:1481. http://xlink.rsc.org/?DOI=p19850001481

20. NCBI Genome database. https://www.ncbi.nlm.nih.gov/genome. Accessed on 16 October 2023.

21. JGI Mycocosm database. https://mycocosm.jgi.doe.gov/mycocosm/home. Accessed on 16 October 2023.

22. Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M. Lichen Metabolites: An Overview of Some Secondary Metabolites and Their Biological Potential. In: Reference Series in Phytochemistry [Internet]. 2020. p. 175–209. http://link.springer.com/10.1007/978-3-319-96397-6_57

23. Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, et al. A Phylogenetic Estimation of Trophic Transition Networks for Ascomycetous Fungi: Are Lichens Cradles of Symbiotrophic Fungal Diversification? Syst Biol [Internet]. 2009 Jun 1;58(3):283–97. https://academic.oup.com/sysbio/article/58/3/283/1610503

24. Tuovinen V, Ekman S, Thor G, Vanderpool D, Spribille T, Johannesson H. Two Basidiomycete Fungi in the Cortex of Wolf Lichens. Curr Biol [Internet]. 2019 Feb;29(3):476-483.e5. https://linkinghub.elsevier.com/retrieve/pii/S0960982218316543

25. Jenkins B, Richards TA. Symbiosis: Wolf Lichens Harbour a Choir of Fungi. Curr Biol [Internet]. 2019;29(3):R88–90. https://doi.org/10.1016/j.cub.2018.12.034

26. McKenzie SK, Walston RF, Allen JL. Complete, high-quality genomes from long-read metagenomic sequencing of two wolf lichen thalli reveals enigmatic genome architecture. Genomics [Internet]. 2020;112(5):3150–6. https://doi.org/10.1016/j.ygeno.2020.06.006

27. Yamamoto Y, Miura Y, Higuchi M, Kinoshita Y, Yoshimura I. Using Lichen Tissue Cultures in Modern Biology. Bryologist [Internet]. 1993;96(3):384. https://www.jstor.org/stable/3243868?origin=crossref

28. Tapia de Daza MS, Beuchat LR. Suitability of modified dichloran glycerol (DG18) agar for enumerating unstressed and stressed xerophilic molds. Food Microbiol [Internet]. 1992

Dec;9(4):319–33. https://linkinghub.elsevier.com/retrieve/pii/074000209280040B

29. Krain A, Siupka P. Fungal Guttation, a Source of Bioactive Compounds, and Its Ecological Role—A Review. Biomolecules [Internet]. 2021 Aug 25;11(9):1270. https://www.mdpi.com/2218-273X/11/9/1270

30. Kalra R, Conlan XA, Goel M. Fungi as a Potential Source of Pigments: Harnessing Filamentous Fungi. Front Chem [Internet]. 2020 May 8;8. https://www.frontiersin.org/article/10.3389/fchem.2020.00369/full

31. Raja HA, Miller AN, Pearce CJ, Oberlies NH. Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. J Nat Prod [Internet]. 2017 Mar 24;80(3):756–70. https://pubs.acs.org/doi/10.1021/acs.jnatprod.6b01085

32. U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. Am J Bot [Internet]. 2012 May;99(5):898–914. https://onlinelibrary.wiley.com/doi/10.3732/ajb.1100459

33. U'Ren JM, Miadlikowska J, Zimmerman NB, Lutzoni F, Stajich JE, Arnold AE. Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomycota). Mol Phylogenet Evol [Internet]. 2016 May;98:210–32. http://dx.doi.org/10.1016/j.ympev.2016.02.010

34. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics [Internet]. 2015 Oct 1;31(19):3210–2. https://academic.oup.com/bioinformatics/article/31/19/3210/211866

35. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res [Internet]. 2017 Jul 3;45(W1):W6–11. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx391

36. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res [Internet]. 2022 Jan 7;50(D1):D571–

728 7. https://academic.oup.com/nar/article/50/D1/D571/6445960

729 37. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated
730 carbohydrate-active enzyme annotation. Nucleic Acids Res [Internet]. 2018 Jul 2;46(W1):W95–101.
731 https://academic.oup.com/nar/article/46/W1/W95/4996582

732 38. Wu HY, Mortensen UH, Chang FR, Tsai H. Whole genome sequence characterization of Aspergillus
733 terreus ATCC 20541 and genome comparison of the fungi A. terreus. Sci Rep [Internet]. 2023 Jan
734 5;13(1):194. https://www.nature.com/articles/s41598-022-27311-7

735 39. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, et al. Comparative
736 genomics reveals high biological diversity and specific adaptations in the industrially and medically
737 important fungal genus Aspergillus. Genome Biol [Internet]. 2017 Dec 14;18(1):28.
738 https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1151-0

739 40. Clavaud C, Aimanianda V, Latge JP. Organization of Fungal, Oomycete and Lichen (1,3)-β-Glucans.
740 In: Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides
741 [Internet]. Elsevier; 2009. p. 387–424.
742 https://linkinghub.elsevier.com/retrieve/pii/B978012373971100011X

743 41. Zhang X, Guo J, Cheng F, Li S. Cytochrome P450 enzymes in fungal natural product biosynthesis.
744 Nat Prod Rep [Internet]. 2021;38(6):1072–99. http://xlink.rsc.org/?DOI=D1NP00004G

745 42. O'Reilly E, Köhler V, Flitsch SL, Turner NJ. Cytochromes P450 as useful biocatalysts: addressing
746 the limitations. Chem Commun [Internet]. 2011;47(9):2490. http://xlink.rsc.org/?DOI=c0cc03165h

747 43. Sang H, Hulvey JP, Green R, Xu H, Im J, Chang T, et al. A Xenobiotic Detoxification Pathway through
748 Transcriptional Regulation in Filamentous Fungi. Turgeon BG, editor. MBio [Internet]. 2018 Sep
749 5;9(4). https://journals.asm.org/doi/10.1128/mBio.00457-18

750 44. Burns K, Helsby NA. Cytochrome P450 in GtoPdb v.2023.1. IUPHAR/BPS Guid to Pharmacol CITE
751 [Internet]. 2023 Apr 26;2023(1). http://journals.ed.ac.uk/gtopdb-cite/article/view/8761

752 45. Ahmadjian V. Lichens are more important than you think. Bioscience [Internet]. 1995 Mar
753 1;45(3):124–124. https://academic.oup.com/bioscience/article-

lookup/doi/10.1093/bioscience/45.3.124

46. Tripathi AH, Mehrotra S, Kumari A, Bajpai R, Joshi Y, Joshi P, et al. Lichens as bioremediation agents—A review. In: Synergistic Approaches for Bioremediation of Environmental Pollutants : Recent Advances and Challenges [Internet]. Elsevier; 2022. p. 289–312. https://linkinghub.elsevier.com/retrieve/pii/B9780323918602000154

47. Hofrichter M, Kellner H, Pecyna MJ, Ullrich R. Fungal Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In 2015. p. 341–68. http://link.springer.com/10.1007/978-3-319-16009-2_13

48. Hofrichter M, Kellner H, Herzog R, Karich A, Liers C, Scheibner K, et al. Fungal Peroxygenases: A Phylogenetically Old Superfamily of Heme Enzymes with Promiscuity for Oxygen Transfer Reactions. In 2020. p. 369–403. http://link.springer.com/10.1007/978-3-030-29541-7_14

49. Monterrey DT, Menés-Rubio A, Keser M, Gonzalez-Perez D, Alcalde M. Unspecific peroxygenases: The pot of gold at the end of the oxyfunctionalization rainbow? Curr Opin Green Sustain Chem [Internet]. 2023 Jun;41:100786. https://linkinghub.elsevier.com/retrieve/pii/S2452223623000354

50. Ullrich R, Nüske J, Scheibner K, Spantzel J, Hofrichter M. Novel Haloperoxidase from the Agaric Basidiomycete Agrocybe aegerita Oxidizes Aryl Alcohols and Aldehydes. Appl Environ Microbiol [Internet]. 2004 Aug;70(8):4575–81. https://journals.asm.org/doi/10.1128/AEM.70.8.4575-4581.2004

51. Rotilio L, Swoboda A, Ebner K, Rinnofner C, Glieder A, Kroutil W, et al. Structural and Biochemical Studies Enlighten the Unspecific Peroxygenase from Hypoxylon sp. EC38 as an Efficient Oxidative Biocatalyst. ACS Catal [Internet]. 2021 Sep 17;11(18):11511–25. https://pubs.acs.org/doi/10.1021/acscatal.1c03065

52. Faiza M, Huang S, Lan D, Wang Y. New insights on unspecific peroxygenases: superfamily reclassification and evolution. BMC Evol Biol [Internet]. 2019 Dec 13;19(1):76. https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-019-1394-3

53. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and

improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res [Internet]. 2023 Jul 5;51(W1):W46–50. https://academic.oup.com/nar/article/51/W1/W46/7151336

54. Godio RP, Fouces R, Martín JF. A Squalene Epoxidase Is Involved in Biosynthesis of Both the Antitumor Compound Clavaric Acid and Sterols in the Basidiomycete H. sublateritium. Chem Biol [Internet]. 2007 Dec;14(12):1334–46. https://linkinghub.elsevier.com/retrieve/pii/S1074552107004024

55. Woo PCY, Lam CW, Tam EWT, Lee KC, Yung KKY, Leung CKF, et al. The biosynthetic pathway for a thousand-year-old natural food colorant and citrinin in Penicillium marneffei. Sci Rep [Internet]. 2014 Oct 22;4(1):6728. https://www.nature.com/articles/srep06728

56. Frandsen RJN, Nielsen NJ, Maolanon N, Sorensen JC, Olsson S, Nielsen J, et al. The biosynthetic pathway for aurofusarin in Fusarium graminearum reveals a close link between the naphthoquinones and naphthopyrones. Mol Microbiol [Internet]. 2006 Aug;61(4):1069–80. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2006.05295.x

57. Franco MEE, Wisecaver JH, Arnold AE, Ju Y, Slot JC, Ahrendt S, et al. Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. New Phytol [Internet]. 2022 Feb 7;233(3):1317–30. https://onlinelibrary.wiley.com/doi/10.1111/nph.17873

58. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics [Internet]. 2010 Dec 23;11(1):395. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-395

59. Wang M, Carver JJ, Phelan V V, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol [Internet]. 2016 Aug 9;34(8):828–37. http://www.ncbi.nlm.nih.gov/pubmed/27504778

60. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. Nat Methods [Internet]. 2020 Sep 24;17(9):905–8.

https://www.nature.com/articles/s41592-020-0933-6

61. Ernst M, Kang K Bin, Caraballo-Rodríguez AM, Nothias LF, Wandy J, Chen C, et al. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. Metabolites [Internet]. 2019 Jul 16;9(7):144. https://www.mdpi.com/2218-1989/9/7/144

62. Fraga BM. Natural sesquiterpenoids. Nat Prod Rep [Internet]. 2013;30(9):1226. http://xlink.rsc.org/?DOI=c3np70047j

63. NISHIKAWA K, ABURAI N, YAMADA K, KOSHINO H, TSUCHIYA E, KIMURA K ichi. The Bisabolane Sesquiterpenoid Endoperoxide, 3,6-Epidioxy-1,10-bisaboladiene, Isolated from Cacalia delphiniifolia Inhibits the Growth of Human Cancer Cells and Induces Apoptosis. Biosci Biotechnol Biochem [Internet]. 2008 Sep 23;72(9):2463–6. https://academic.oup.com/bbb/article/72/9/2463-2466/5941358

64. Kimura K ichi, Sakamoto Y, Fujisawa N, Uesugi S, Aburai N, Kawada M, et al. Cleavage mechanism and anti-tumor activity of 3,6-epidioxy-1,10-bisaboladiene isolated from edible wild plants. Bioorg Med Chem [Internet]. 2012 Jun;20(12):3887–97. https://linkinghub.elsevier.com/retrieve/pii/S0968089612003008

65. Lo JY, Kamarudin MNA, Hamdi OAA, Awang K, Kadir HA. Curcumenol isolated from Curcuma zedoaria suppresses Akt-mediated NF-κB activation and p38 MAPK signaling pathway in LPS-stimulated BV-2 microglial cells. Food Funct [Internet]. 2015;6(11):3550–9. http://xlink.rsc.org/?DOI=C5FO00607D

66. Tamez-Fernández JF, Melchor-Martínez EM, Ibarra-Rivera TR, Rivas-Galindo VM. Plant-derived endoperoxides: structure, occurrence, and bioactivity. Phytochem Rev [Internet]. 2020 Aug 29;19(4):827–64. https://link.springer.com/10.1007/s11101-020-09687-4

67. Krishna S, Bustamante L, Haynes RK, Staines HM. Artemisinins: their growing importance in medicine. Trends Pharmacol Sci [Internet]. 2008 Oct;29(10):520–7. https://linkinghub.elsevier.com/retrieve/pii/S0165614708001648

68. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture

and applications. BMC Bioinformatics [Internet]. 2009 Dec 15;10(1):421. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-421

69. Moujir L, Callies O, Sousa PMC, Sharopov F, Seca AML. Applications of Sesquiterpene Lactones: A Review of Some Potential Success Cases. Appl Sci [Internet]. 2020 Apr 25;10(9):3001. https://www.mdpi.com/2076-3417/10/9/3001

70. IGARASHI Y, KUWAMORI Y, TAKAGI K, ANDO T, FUDOU R, FURUMAI T, et al. Xanthoepocin, a New Antibiotic from Penicillium simplicissimum IFO5762. J Antibiot (Tokyo) [Internet]. 2000;53(9):928–33. http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/53.928?from=CrossRef

71. Vrabl P, Siewert B, Winkler J, Schöbel H, Schinagl CW, Knabl L, et al. Xanthoepocin, a photolabile antibiotic of Penicillium ochrochloron CBS 123823 with high activity against multiresistant gram-positive bacteria. Microb Cell Fact [Internet]. 2022 Dec 4;21(1):1. https://doi.org/10.1186/s12934-021-01718-9

72. Cary JW, Harris-Coward PY, Ehrlich KC, Di Mavungu JD, Malysheva S V., De Saeger S, et al. Functional characterization of a veA-dependent polyketide synthase gene in Aspergillus flavus necessary for the synthesis of asparasone, a sclerotium-specific pigment. Fungal Genet Biol [Internet]. 2014;64:25–35. http://dx.doi.org/10.1016/j.fgb.2014.01.001

73. Hu Y, Hao X, Lou J, Zhang P, Pan J, Zhu X. A PKS gene, pks-1, is involved in chaetoglobosin biosynthesis, pigmentation and sporulation in Chaetomium globosum. Sci China Life Sci [Internet]. 2012 Dec 12;55(12):1100–8. http://link.springer.com/10.1007/s11427-012-4409-5

74. Lu S, Tian J, Sun W, Meng J, Wang X, Fu X, et al. Bis-naphtho-γ-pyrones from Fungi and Their Bioactivities. Molecules [Internet]. 2014 May 30;19(6):7169–88. http://www.mdpi.com/1420-3049/19/6/7169

75. Fürtges L, Obermaier S, Thiele W, Foegen S, Müller M. Diversity in Fungal Intermolecular Phenol Coupling of Polyketides: Regioselective Laccase-Based Systems. ChemBioChem [Internet]. 2019 Aug 8;20(15):1928–32. https://onlinelibrary.wiley.com/doi/10.1002/cbic.201900041

76. Hüttel W, Müller M. Regio- and stereoselective intermolecular phenol coupling enzymes in

858 secondary metabolite biosynthesis. Nat Prod Rep [Internet]. 2021;38(5):1011–43.
859 http://xlink.rsc.org/?DOI=D0NP00010H

860 77. Xu D, Yin R, Zhou Z, Gu G, Zhao S, Xu JR, et al. Elucidation of ustilaginoidin biosynthesis reveals
861 a previously unrecognised class of ene-reductases. Chem Sci [Internet]. 2021;12(44):14883–92.
862 http://xlink.rsc.org/?DOI=D1SC02666F

863 78. Urquhart AS, Hu J, Chooi YH, Idnurm A. The fungal gene cluster for biosynthesis of the antibacterial
864 agent viriditoxin. Fungal Biol Biotechnol [Internet]. 2019 Dec 1;6(1):9.
865 https://fungalbiolbiotech.biomedcentral.com/articles/10.1186/s40694-019-0072-y

866 79. Newman AG, Townsend CA. Molecular Characterization of the Cercosporin Biosynthetic Pathway
867 in the Fungal Plant Pathogen Cercospora nicotianae. J Am Chem Soc [Internet]. 2016 Mar
868 30;138(12):4219–28. https://pubs.acs.org/doi/10.1021/jacs.6b00633

869 80. Newman AG, Vagstad AL, Belecki K, Scheerer JR, Townsend CA. Analysis of the cercosporin
870 polyketide synthase CTB1 reveals a new fungal thioesterase function. Chem Commun [Internet].
871 2012;48(96):11772. http://xlink.rsc.org/?DOI=c2cc36010a

872 81. Barbier M, Devys M, Parisot D. A Simple Synthesis of 4-Deoxyanhydrofusarubin Lactone. Synth
873 Commun [Internet]. 1993 Mar 1;23(5):651–6.
874 https://www.tandfonline.com/doi/full/10.1080/00397919308009823

875 82. The Species Fungorum. https://www.speciesfungorum.org. Accessed on 16 October 2023.

876 83. Phukhamsakda C, Nilsson RH, Bhunjun CS, de Farias ARG, Sun YR, Wijesinghe SN, et al. The
877 numbers of fungi: contributions from traditional taxonomic studies and challenges of metabarcoding.
878 Fungal Divers [Internet]. 2022 May 28;114(1):327–86. https://link.springer.com/10.1007/s13225-
879 022-00502-3

880 84. Bandi CK, Agrawal A, Chundawat SP. Carbohydrate-Active enZyme (CAZyme) enabled
881 glycoengineering for a sweeter future. Curr Opin Biotechnol [Internet]. 2020 Dec;66:283–91.
882 https://linkinghub.elsevier.com/retrieve/pii/S0958166920301300

883 85. Pallister E, Gray CJ, Flitsch SL. Enzyme promiscuity of carbohydrate active enzymes and their

884 applications in biocatalysis. Curr Opin Struct Biol [Internet]. 2020 Dec;65:184–92.
885 https://linkinghub.elsevier.com/retrieve/pii/S0959440X20301275

886 86. Mhiri S, Bouanane-Darenfed A, Jemli S, Neifar S, Ameri R, Mezghani M, et al. A thermophilic and
887 thermostable xylanase from Caldicoprobacter algeriensis: Recombinant expression,
888 characterization and application in paper biobleaching. Int J Biol Macromol [Internet]. 2020
889 Dec;164:808–17. https://linkinghub.elsevier.com/retrieve/pii/S0141813020339210

890 87. Karuppiah V, Zhixiang L, Liu H, Vallikkannu M, Chen J. Co-culture of Vel1-overexpressed
891 Trichoderma asperellum and Bacillus amyloliquefaciens: An eco-friendly strategy to hydrolyze the
892 lignocellulose biomass in soil to enrich the soil fertility, plant growth and disease resistance. Microb
893 Cell Fact [Internet]. 2021 Dec 2;20(1):57.
894 https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-021-01540-3

895 88. Urlacher VB, Girhard M. Cytochrome P450 Monooxygenases in Biotechnology and Synthetic
896 Biology. Trends Biotechnol [Internet]. 2019 Aug 1;37(8):882–97.
897 https://linkinghub.elsevier.com/retrieve/pii/S0167779919300010

898 89. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol
899 [Internet]. 1990 Oct;215(3):403–10. https://linkinghub.elsevier.com/retrieve/pii/S0022283605803602

900 90. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11.
901 Mol Biol Evol [Internet]. 2021;38(7):3022–7.
902 https://academic.oup.com/mbe/article/38/7/3022/6248099

903 91. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space
904 complexity. BMC Bioinformatics. 2004;5:1–19.

905 92. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and
906 processing long-read sequencing data. Bioinformatics [Internet]. 2018;34(15):2666–9.
907 https://academic.oup.com/bioinformatics/article/34/15/2666/4934939

908 93. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome
909 assemblies. Bioinformatics [Internet]. 2013 Apr 15;29(8):1072–5.

https://academic.oup.com/bioinformatics/article/29/8/1072/228832

94. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics [Internet]. 2015 Oct 15;31(20):3350–2. https://academic.oup.com/bioinformatics/article/31/20/3350/196114

95. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res [Internet]. 2017 May;27(5):737–46. http://genome.cshlp.org/lookup/doi/10.1101/gr.214270.116

96. Humann JL, Lee T, Ficklin S, Main D. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. In 2019. p. 29–51. http://link.springer.com/10.1007/978-1-4939-9173-0_3

97. Smit AFA., Hubley R, Green P. RepeatMasker Open-4.0. https://www.repeatmasker.org.

98. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res [Internet]. 2005 Jul 1;33(Web Server):W465–7. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki458

99. Borodovsky M, Lomsadze A. Eukaryotic Gene Prediction Using GeneMark.hmm-E and GeneMark-ES. Curr Protoc Bioinforma [Internet]. 2011 Sep 7;35(1). https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0406s35

100. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods [Internet]. 2015 Jan 17;12(1):59–60. https://www.nature.com/articles/nmeth.3176

101. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol [Internet]. 2008;9(1):R7. http://genomebiology.biomedcentral.com/articles/10.1186/gb-2008-9-1-r7

102. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In 2019. p. 1–14. http://link.springer.com/10.1007/978-1-4939-9173-0_1

103. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res [Internet].

2012 Nov 27;41(D1):D590–6. http://academic.oup.com/nar/article/41/D1/D590/1069277/The-SILVA-ribosomal-RNA-gene-database-project

104. Käll L, Krogh A, Sonnhammer EL. A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol [Internet]. 2004 May;338(5):1027–36. https://linkinghub.elsevier.com/retrieve/pii/S0022283604002943

105. Törönen P, Holm L. PANNZER —A practical tool for protein function prediction. Protein Sci [Internet]. 2022 Jan 14;31(1):118–28. https://onlinelibrary.wiley.com/doi/10.1002/pro.4193

106. Buchholz PCF, Vogel C, Reusch W, Pohl M, Rother D, Spieß AC, et al. BioCatNet: A Database System for the Integration of Enzyme Sequences and Biocatalytic Experiments. ChemBioChem [Internet]. 2016 Nov 3;17(21):2093–8. https://onlinelibrary.wiley.com/doi/10.1002/cbic.201600462

107. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res [Internet]. 2023 Jan 6;51(D1):D523–31. https://academic.oup.com/nar/article/51/D1/D523/6835362

108. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics [Internet]. 2009 May 1;25(9):1189–91. https://academic.oup.com/bioinformatics/article/25/9/1189/203460

109. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol [Internet]. 2012 Oct 10;30(10):918–20. https://www.nature.com/articles/nbt.2377

110. WinSCP FTP client. https://winscp.net/eng/index.php. Accessed on 15 November 2022.

111. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res [Internet]. 2003 Nov;13(11):2498–504. http://genome.cshlp.org/lookup/doi/10.1101/gr.1239303

112. Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, et al. Mass spectrometry searches using MASST. Nat Biotechnol [Internet]. 2020 Jan 1;38(1):23–6. https://www.nature.com/articles/s41587-019-0375-9

962 113. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik A V., Meusel M, et al. SIRIUS 4: a rapid
963 tool for turning tandem mass spectra into metabolite structure information. Nat Methods [Internet].
964 2019 Apr 18;16(4):299–302. https://www.nature.com/articles/s41592-019-0344-8

965

966

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None declared.

Data availability

All data generated or analyzed during this study are included in the main article and additional files. The sequencing data and genome assembly for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the accession number PRJEB67537. The mass spectrometry data have been deposited on GNPS under the accession number MassIVE ID: MSV000093120.

Funding

This project was supported by the Federation of Biochemical Societies through the FEBS Excellence award awarded to KH. Support for contributions from JLA was provided by NSF DEB #2115191.

990 **Author contributions**

991 RI conceived the study with input from KH and JLA; JLA collected lichen specimen; RI, THe
992 and THackl performed genome sequencing and analysis; RI performed all fungal cultivations and
993 metabolomics studies; KH acquired funding and supervised data collection and analysis, RI and
994 KH wrote the manuscript with input from all authors. All authors have read the final version of the
995 manuscript and agreed to its publication.

996

997 **Acknowledgements**

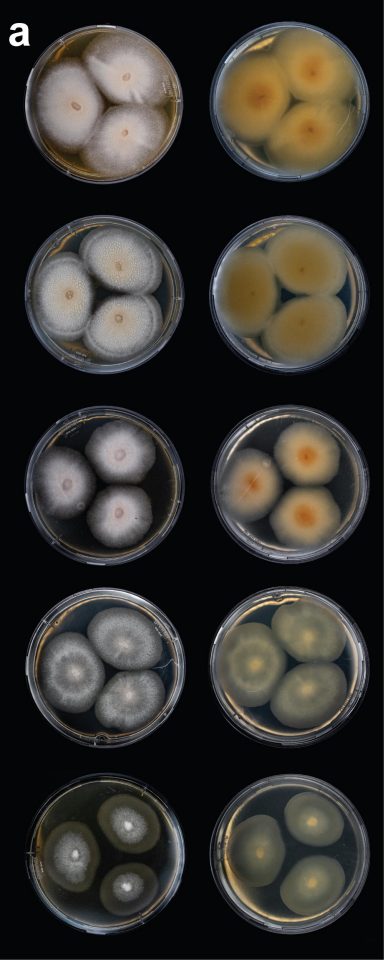
998 The authors are grateful to Xiao Li for discussions on the molecular networking analysis,
999 and the Interfaculty Mass Spectrometry Center of the University of Groningen and the University
1000 Medical Center Groningen for their services in high resolution tandem mass spectrometry.

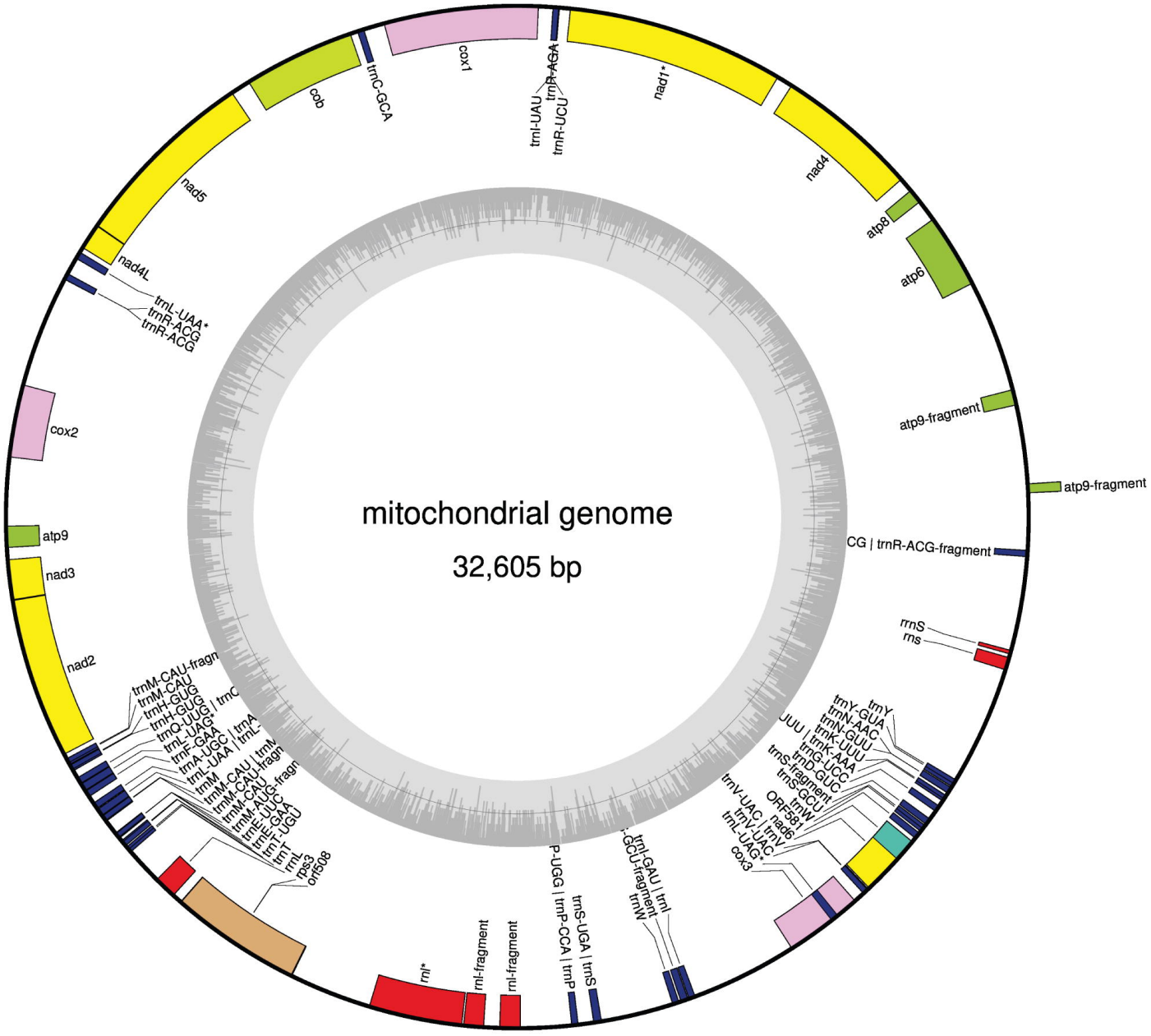
1001

1002 ## Additional files

1003 **Additional file 1:** tables S1-S5, and figures S1-S6. **Additional file 2:** structural and
1004 functional annotation of the genome of *A. pinea* F5. **Additional file 3:** BGC prediction by
1005 fungiSMASH7 on the genome of *A. pinea* F5. **Additional file 4:** molecular network of *A. pinea* F5
1006 extracts and annotation by Sirius (Cytoscape session file). **Additional file 5**: molecular network
1007 of *A. pinea* F5 extracts annotated with MolNetEnhancer (Cytoscape session file). **Additional file**
1008 **6:** list of hits from GNPS MS/MS library search of *A. pinea* extracts. **Additional file 7:** Extracted
1009 ion chromatograms, MS and MS/MS spectra of annotated nodes from the molecular network.
1010 **Additional file 8:** results of blastP search on core synthase/cyclase genes from predicted terpene
1011 BGCs in the genome of *A. pinea* F5.

1012


a
b
KP992078.1 Sordariomycetes sp. voucher ARIZ YLH0105
KP992032.1 Sordariomycetes sp. voucher ARIZ YLH0051
GQ153162.1 Sordariomycetes sp. 11216
isolate F5
HM123694.1 Xylariaceae sp. ARIZ AZ1047
HQ599578.1 Anthostomella pinea culture collection CBS 128205
KX096656.1 Anthostomella pinea isolate RP139 2 1
OK576226.1 Anthostomella pinea isolate SD11
KU663949.1 uncultured Anthostomella clone AK100
KT291430.1 Anthostomella pinea strain FMY29
MF327371.1 Anthostomella pinea strain DNA105
MG543933.1 Anthostomella sp. isolate EP22
KJ406991.1 Anthostomella pinea strain CL199
MF327370.1 Anthostomella pinea strain DNA51
MW447070.1 Anthostomella pinea strain FeC56
MT153620.1 Anthostomella sp. strain GLMC451
MH374419.1 Anthostomella pinea strain F317
FN435792.1 Xylaria sp. agrAR275
KJ406993.1 Anthostomella pinea strain C14
MN844404.1 Anthostomella ravennica voucher KoLRI EL004521
NR 153557.1 Anthostomella ravennica MFLUCC 15-0012
AM922205.1 Anthostomella eucalyptorum isolate 2741
MW489547.1 Anthostomella adusta voucher ZT-Myc-64230
NR 175686.1 Anthostomella lamiacearum MFLU 18-0101
MW240653.1 Anthostomella helicofissa strain MFLUCC 14-0173
MH860275.1 Anthostomella spartii culture collection CBS 586.71
MW240652.1 Anthostomella formosa strain MFLUCC 14-0170
KP297405.1 Anthostomella sp. DAD-2014c isolate IT1617
KP297403.1 Anthostomella formosa isolate IT1553
NR 153556.1 Anthostomella thailandica MFLUCC 15-0017
NG 057105.1 Anthostomella pseudobambusicola MFLUCC 15-0192
AY590787.1 Xylaria sp. olrim65
AF252618.1 Anthostomella unguiculata
NG 065090.1 Anthostomella pseudobambusicola MFLU 16-0255
DQ836895.1 Anthostomella torosa isolate AFTOL-ID 732
AF252619.1 Anthostomella flagellariae
AF252620.1 Anthostomella palmicola
AF252617.1 Anthostomella aquatica
NG 062835.1 Trichoderma harzianum CBS 226.95
63
97
65
96
52
76
87
57
67
72
100
90
71
99
57
69
100
100
75
100
66
55
77
59



complex I (NADH dehydrogenase)

complex III (ubichinol cytochrome c reductase)

complex IV (cytochrome c oxidase)

ATP synthase

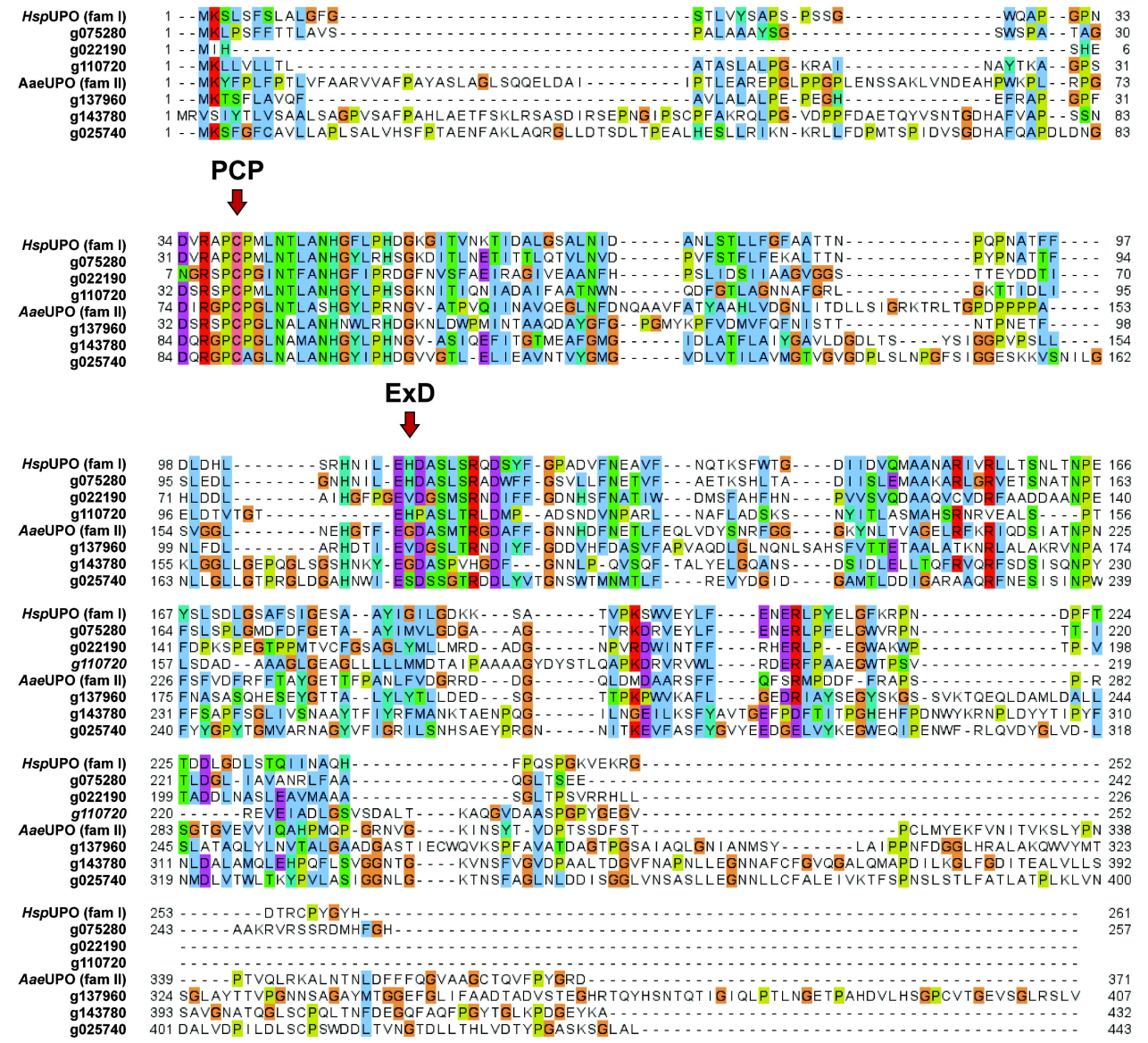
ribosomal proteins (SSU)

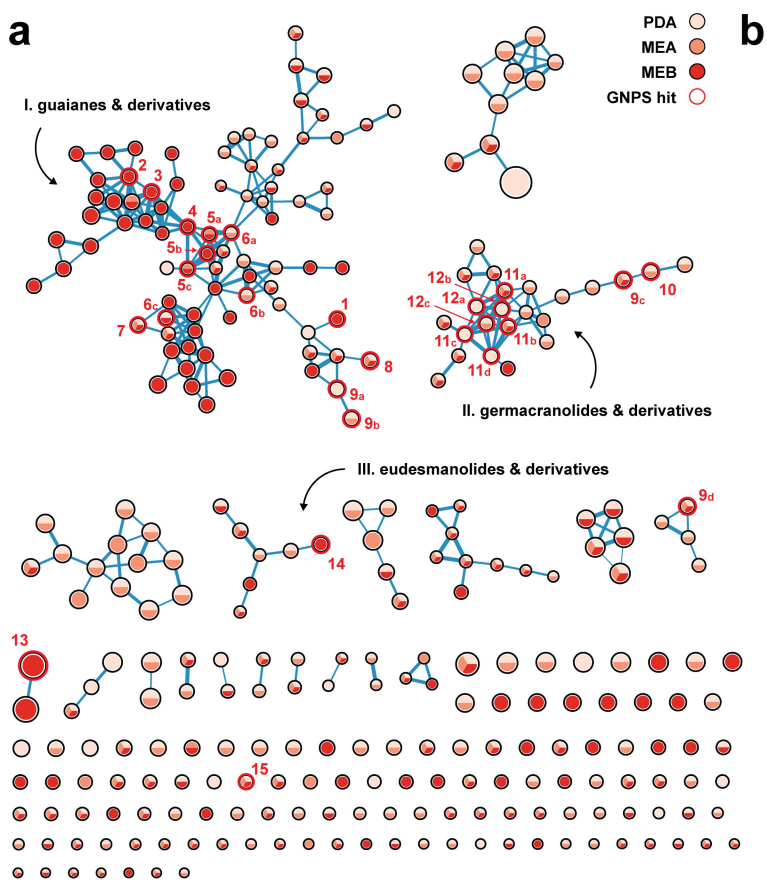
ORFs

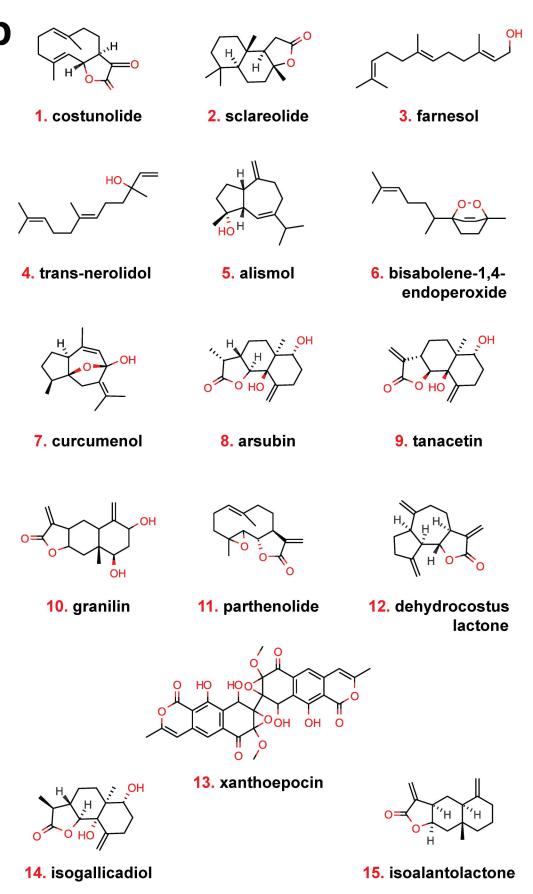
transfer RNAs

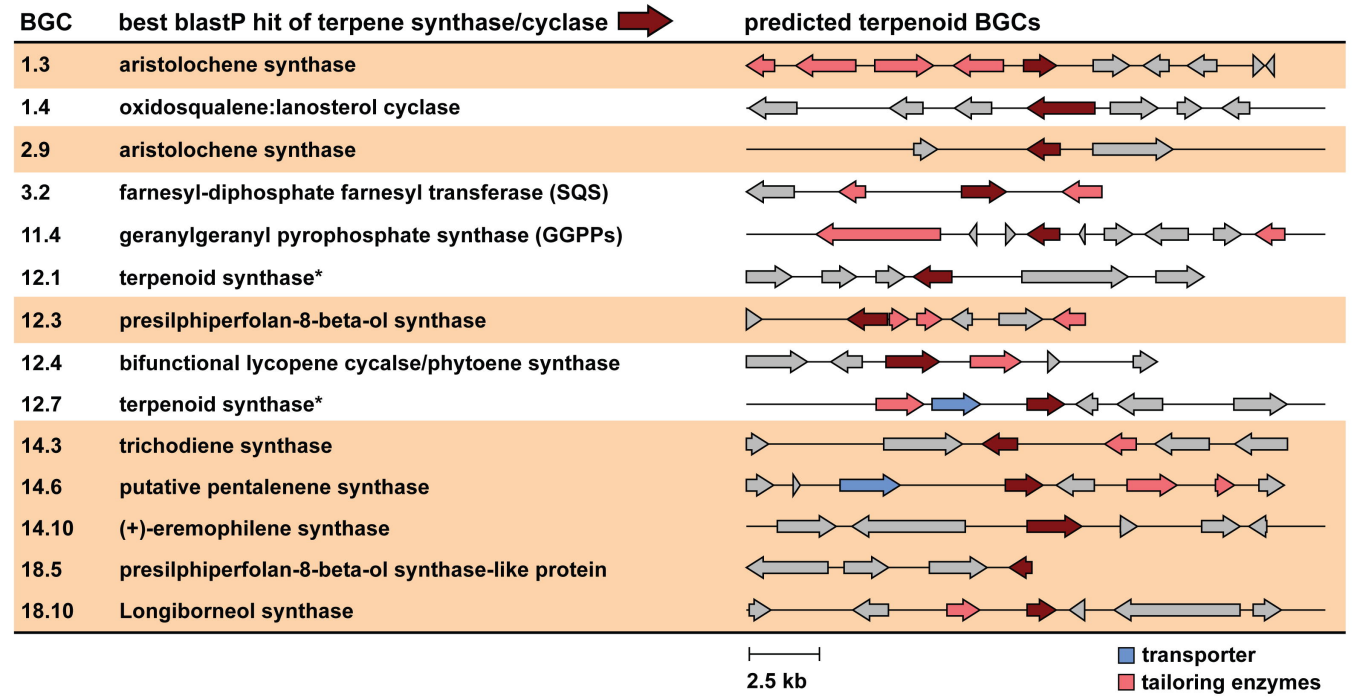
ribosomal RNAs

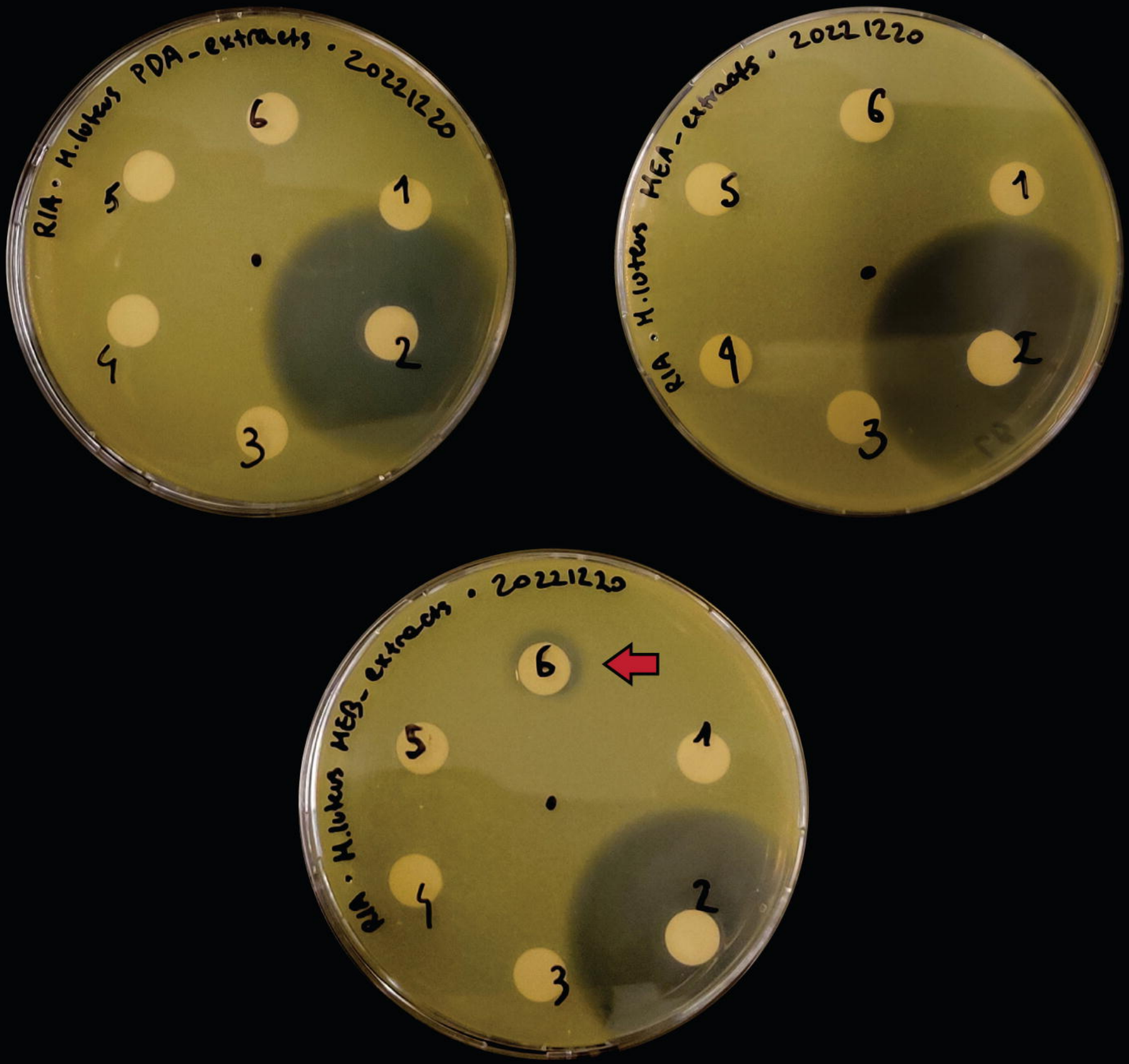

# of genes
0
500
1000
1500
2000
2500
3000
transmembrane transport
phosphorylation
methylation
proteolysis
regulation of transcription by RNA polymerase II
carbohydrate metabolic process
secondary metabolite biosynthetic process
transcription, DNA-templated
translation
protein phosphorylation
protein transport
cell division
membrane
nucleus
cytoplasm
extracellular region
mitochondrion
cytosol
ribosome
ribonucleoprotein complex
endoplasmic reticulum membrane
nucleolus
RNA polymerase I transcription factor complex
mitochondrial inner membrane
cell periphery
ATP binding
hydrolase activity
oxidoreductase activity
metal ion binding
DNA binding
zinc ion binding
transmembrane transport activity
transferase activity
RNA binding
heme binding
monooxygenase activity
kinase activity
RNA polymerase II transcription factor activity, sequence-specific DNA binding
methyltransferase activity
iron ion binding
structural constituent of ribosome
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
nucleic acid binding
FAD binding
ATPase activity
GTP binding
ligase activity
flavin adenine dinucleotide binding
protein kinase activity
GTPase activity
Biological process
Cellular component
Molecular function

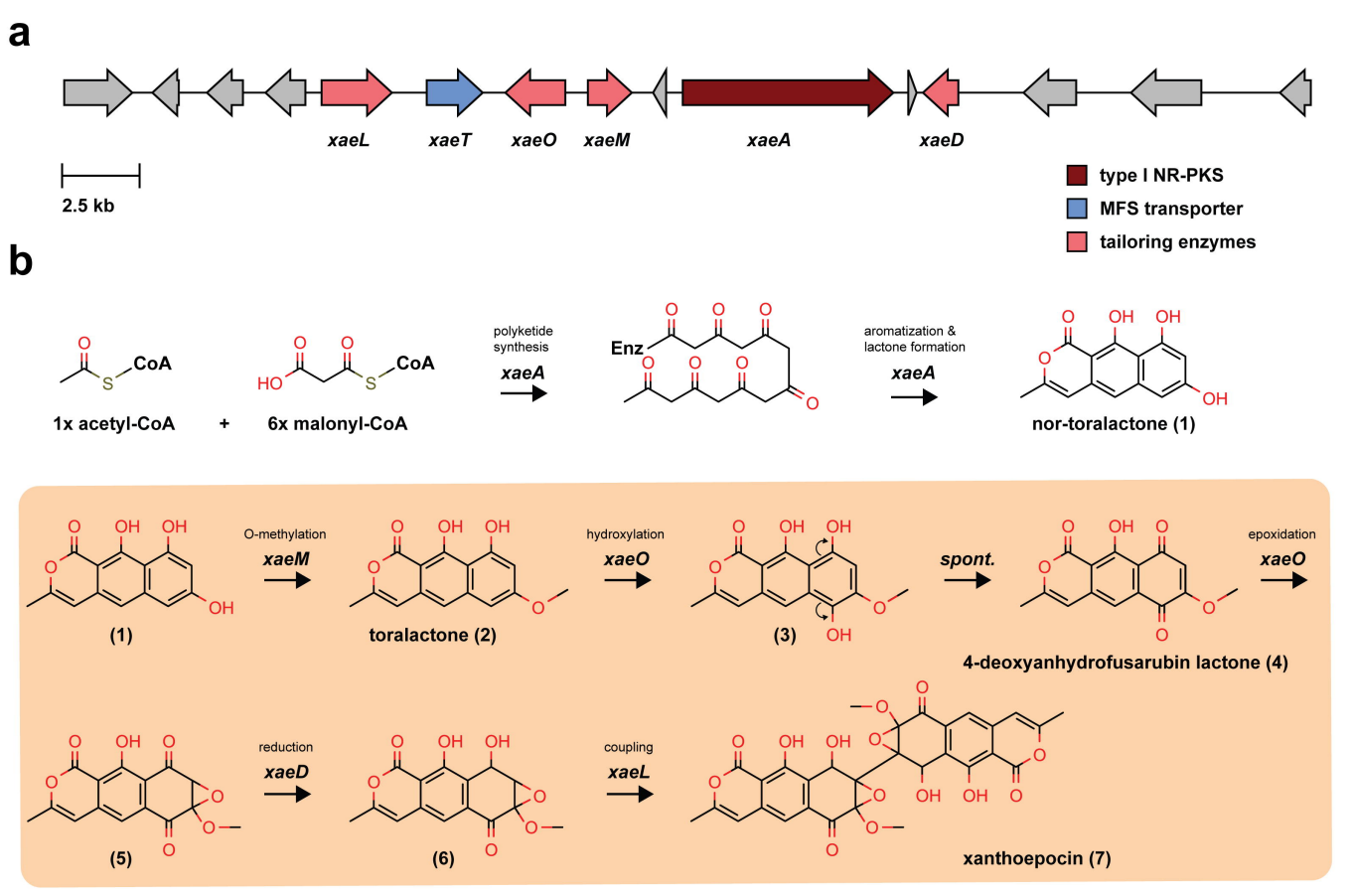

HspUPO (fam I)
g075280
g022190
g110720
AaeUPO (fam II)
g137960
g143780
g025740
PCP
ExD


a
PDA
MEA
MEB
GNPS hit
I. guaianes & derivatives
II. germacranolides & derivatives
III. eudesmanolides & derivatives
1
2
3
4
5a
5b
5c
6a
6b
6c
7
8
9a
9b
9c
9d
10
11a
11b
11c
11d
12a
12b
12c
13
14
15


b
1. costunolide
2. sclareolide
3. farnesol
4. trans-nerolidol
5. alismol
6. bisabolene-1,4-endoperoxide
7. curcumenol
8. arsubin
9. tanacetin
10. granilin
11. parthenolide
12. dehydrocostus lactone
13. xanthoepocin
14. isogallicadiol
15. isoalantolactone


BGC
best blastP hit of terpene synthase/cyclase
predicted terpenoid BGCs
1.3
aristolochene synthase
1.4
oxidosqualene:lanosterol cyclase
2.9
aristolochene synthase
3.2
farnesyl-diphosphate farnesyl transferase (SQS)
11.4
geranylgeranyl pyrophosphate synthase (GGPPs)
12.1
terpenoid synthase*
12.3
presilphiperfolan-8-beta-ol synthase
12.4
bifunctional lycopene cycalse/phytoene synthase
12.7
terpenoid synthase*
14.3
trichodiene synthase
14.6
putative pentalenene synthase
14.10
(+)-eremophilene synthase
18.5
presilphiperfolan-8-beta-ol synthase-like protein
18.10
Longiborneol synthase
2.5 kb
transporter
tailoring enzymes


RIA · H.lutens PDA-extracts · 20221220
6
5
1
4
2
3
RIA · H.lutens MEA-extracts · 20221220
6
5
1
4
2
3
RIA · H.lutens MEG-extracts · 20221220
6
5
1
4
2
3


a
xaeL
xaeT
xaeO
xaeM
xaeA
xaeD
2.5 kb
type I NR-PKS
MFS transporter
tailoring enzymes
b
1x acetyl-CoA
+
6x malonyl-CoA
polyketide synthesis
xaeA
Enz
aromatization & lactone formation
xaeA
nor-toralactone (1)
O-methylation
xaeM
toralactone (2)
hydroxylation
xaeO
(3)
spont.
epoxidation
xaeO
4-deoxyanhydrofusarubin lactone (4)
(1)
reduction
xaeD
(5)
(6)
coupling
xaeL
xanthoepocin (7)