

RESEARCH

Open Access



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

R. Iacovelli¹, T. He¹, J. L. Allen², T. Hackl³ and K. Haslinger^{1*}

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 *Anthostomella pinea* to provide new bioactive molecules and biocatalysts and will facilitate future research into these valuable metabolites.

Keywords Fungal genomics, Natural products, Sesquiterpenes, Molecular networking, Antibiotics, Lichen

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

*Correspondence:

K. Haslinger
k.haslinger@rug.nl

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

² Department of Biology, Eastern Washington University, Cheney, WA 99004, USA

³ Groningen Institute for Evolutionary Life Sciences, University of Groningen, 9747 AG Groningen, The Netherlands



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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, though little to nothing is known about their genetic makeup and metabolome. To the best of our knowledge, only a few publications report the discovery and characterization of metabolites from *Anthostomella* fungi [17–19], and complete genomic data is currently available for only two species [20, 21]. Therefore, we set out to perform whole genome sequencing and untargeted metabolomics analyses to shed light on the ability of the fungus to produce biotechnologically relevant secondary metabolites and enzymes that can be used as biocatalysts in industrial applications. To do so, we employed state-of-the-art long read sequencing technologies that allowed us to obtain a high-quality annotated genome of the fungus. We then used high-resolution tandem mass-spectrometry coupled to molecular networking analysis to analyze the intra- and extra-cellular metabolome. With these approaches, we revealed that *A. pinea* is a prolific producer of sesquiterpenoids and sesquiterpene lactones, and that its genome encodes an abundance of biotechnologically relevant enzymes such as carbohydrate-active enzymes (CAZymes), cytochrome P450s (CYP), and unspecific peroxygenases (UPOs). We also detected antimicrobial activity of the fungal extracts, which we putatively attribute to the anti-Gram-positive

compound xathoepocin, produced in small amounts by the fungus. Lastly, based on the genomics data, we hypothesize the biosynthetic route of this compound, which could prove useful in future efforts to reconstitute the pathway in a heterologous host for large scale production and engineering of structural variants.

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 recently received substantial 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 specimens from a *Pinus ponderosa* tree growing by the Cheney Wetlands Trail in Cheney, Washington, USA and shipped them to the research facilities in Groningen, The Netherlands. Next, we prepared surface-sterilized sections of the lichen thallus as described previously [27], and incubated them in sterile water. 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), with the mycelium appearing extremely thin in the peripheral zone. In all other cases, the colonies spread with dense cream to white vegetative mycelium and smooth, regular margins. Colony surface appeared somewhat velvety, more woolly on YES media. Aerial mycelium was little to

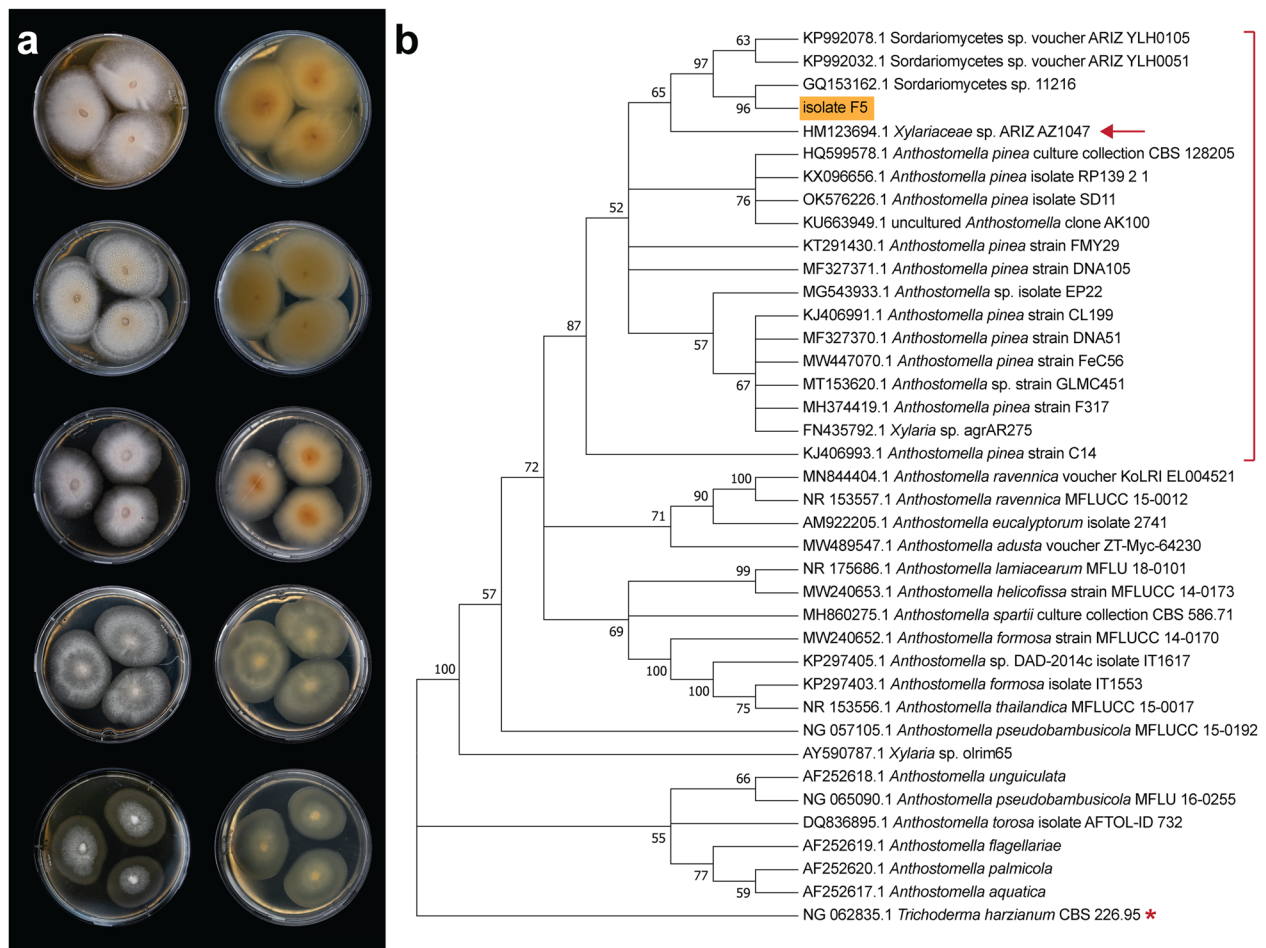


Fig. 1 **a** Morphological characterization of F5 isolate via cultivation on solid medium. From top to bottom: MEA, PDA, DG18, YES, and CYA; obverse (left) and reverse (right). The cultures were incubated for 28 days at 20 °C. **b** Maximum-likelihood phylogenetic tree based on ITS sequences retrieved 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

none on all tested media. Despite robust growth, on DG18 the colonies did not grow to full contact, while on YES plates they showed a two-ring morphology and less compact mycelium formation. Both these characteristics are 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 obverse 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.

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 Additional file 1: 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* (Additional file 1: 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 (Additional file 1: Figs. S2, 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 (Additional file 1: 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 2000 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 6949 to 8291, respectively (Additional file 1: 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 Additional file 1: 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] (Additional file 1: Fig. S4).

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 process (24%), molecular function (50%), and cellular component (26%). Particularly interesting was the abundance of biotechnologically relevant genes: 171 genes were annotated as “secondary metabolite biosynthetic process”, 229 as “monooxygenase

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

activity”, and 221 as “carbohydrate metabolic process” (Additional file 1: Table S5 and Fig. 2). Thus, we set out to investigate these types of genes more deeply with specific bioinformatic tools.

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, HMMER, 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

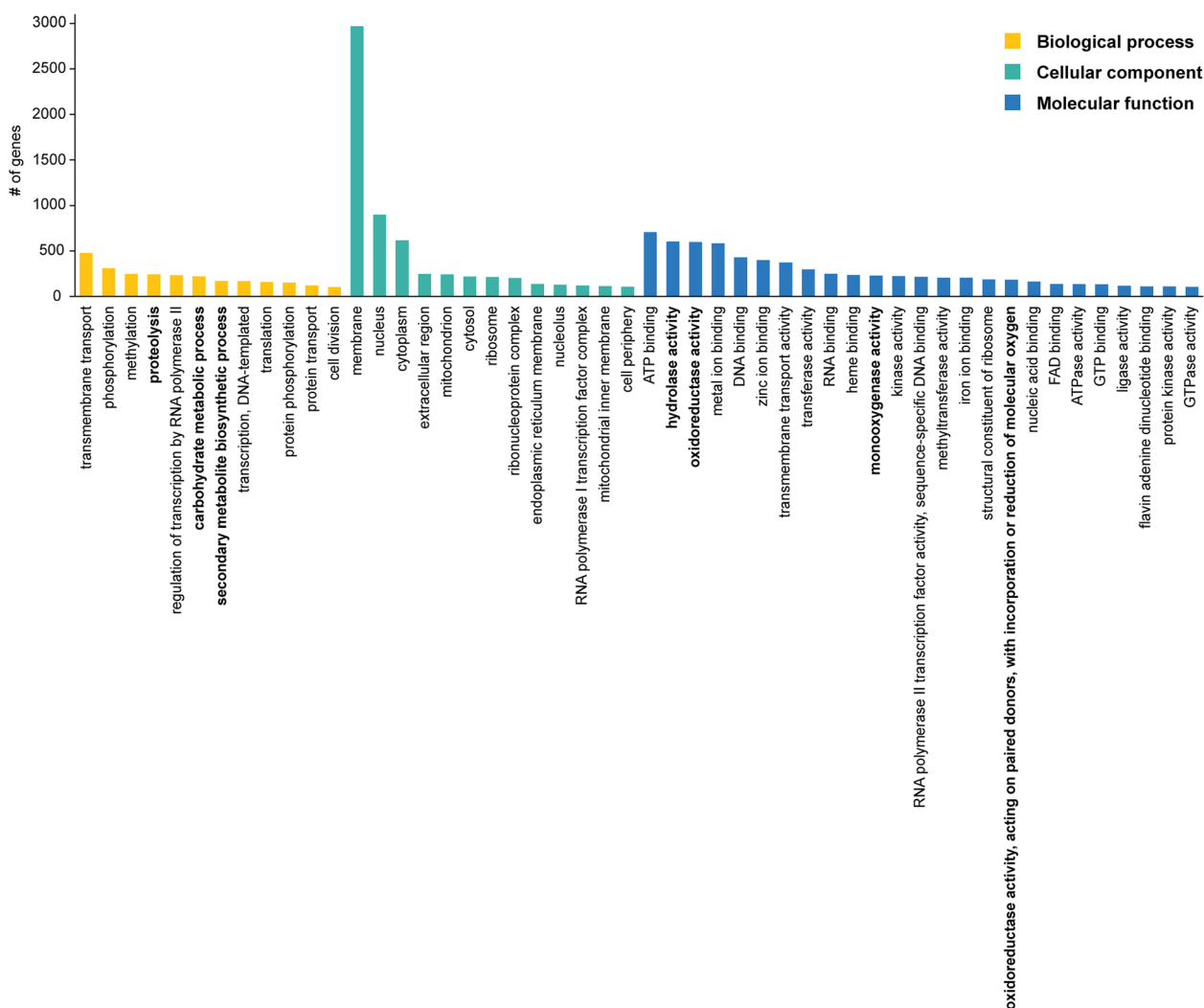


Fig. 2 Gene Ontology (GO) analysis of the genome of *A. pinea* F5 (top 50 terms). Biotechnologically relevant categories are highlighted in bold

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

able to colonize lichen tissues and grow as an endo/epilichenic fungus.

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 the 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: *AaeUPO* from *Agrocybe aegerita* [50], prototype enzyme of family II (long UPOs); and *HspUPO* 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 an 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 (Fig. 3). 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.

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 tightly 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 (Additional file 1: Fig. S6). 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 by 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 value, 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. 4a. By using the additional GNPS tool MolNet Enhancer [61], three of the clusters were annotated as different types of sesquiterpenoids (STs) and sesquiterpene lactones (STLs): namely, guaianes (including dioxanes) (I), germacranolides (II), and eudesmanolides (III) (Fig. 4a) (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

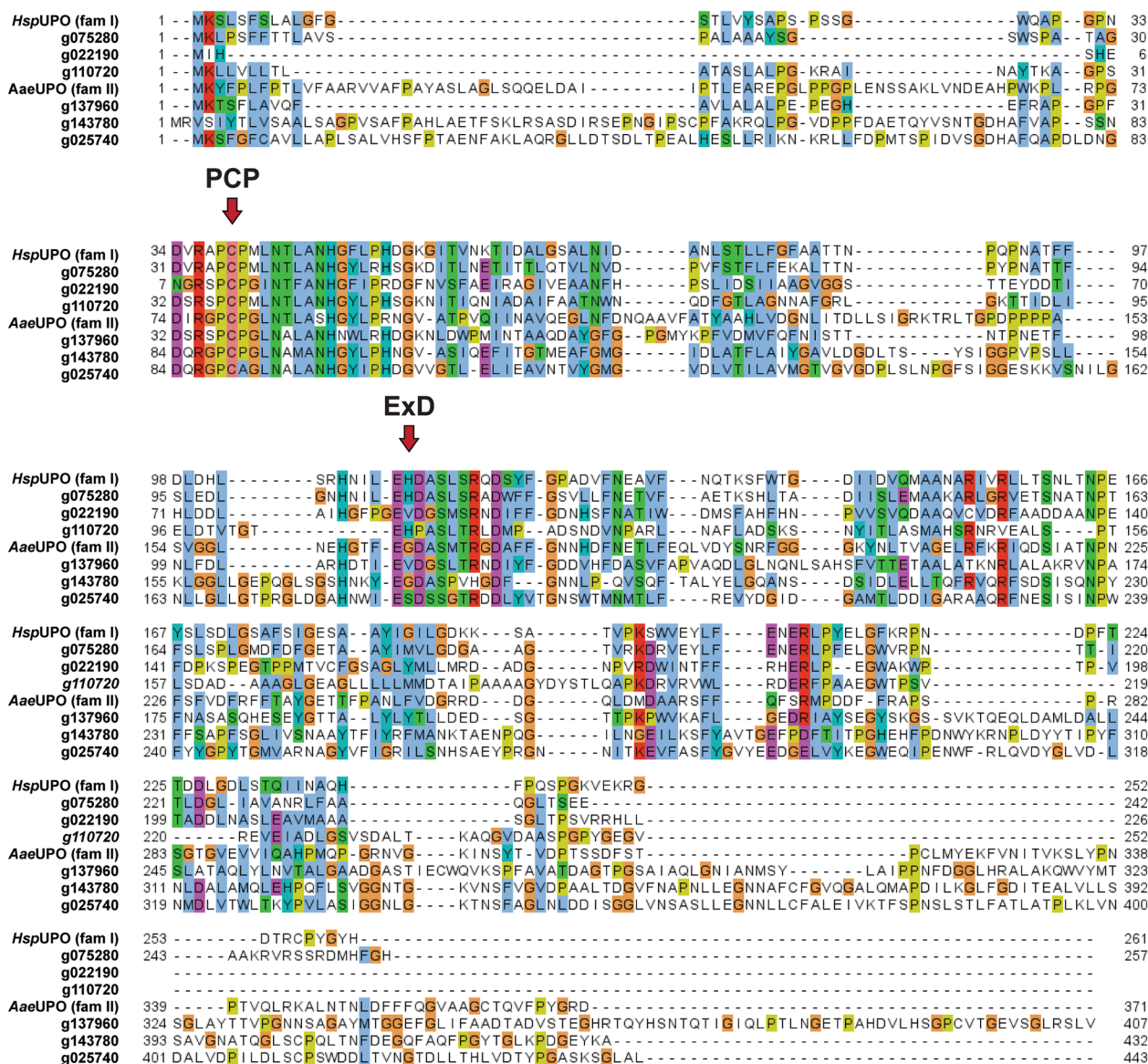


Fig. 3 Multiple sequence alignment analysis of putative UPOs from *A. pinea* F5 and prototype UPOs. The sequences of family I and family II prototype UPOs were retrieved from the Uniprot database. The conserved motifs “PCP” and “ExD” are highlighted by red arrows. The entry of g137950 was removed from the alignment since it lacked the conserved motifs

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. 4a, b) (Table 3).

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, 8, 9) 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 MS¹ level already. This results in different nodes showing identical m/z values and the same MS/MS fragmentation

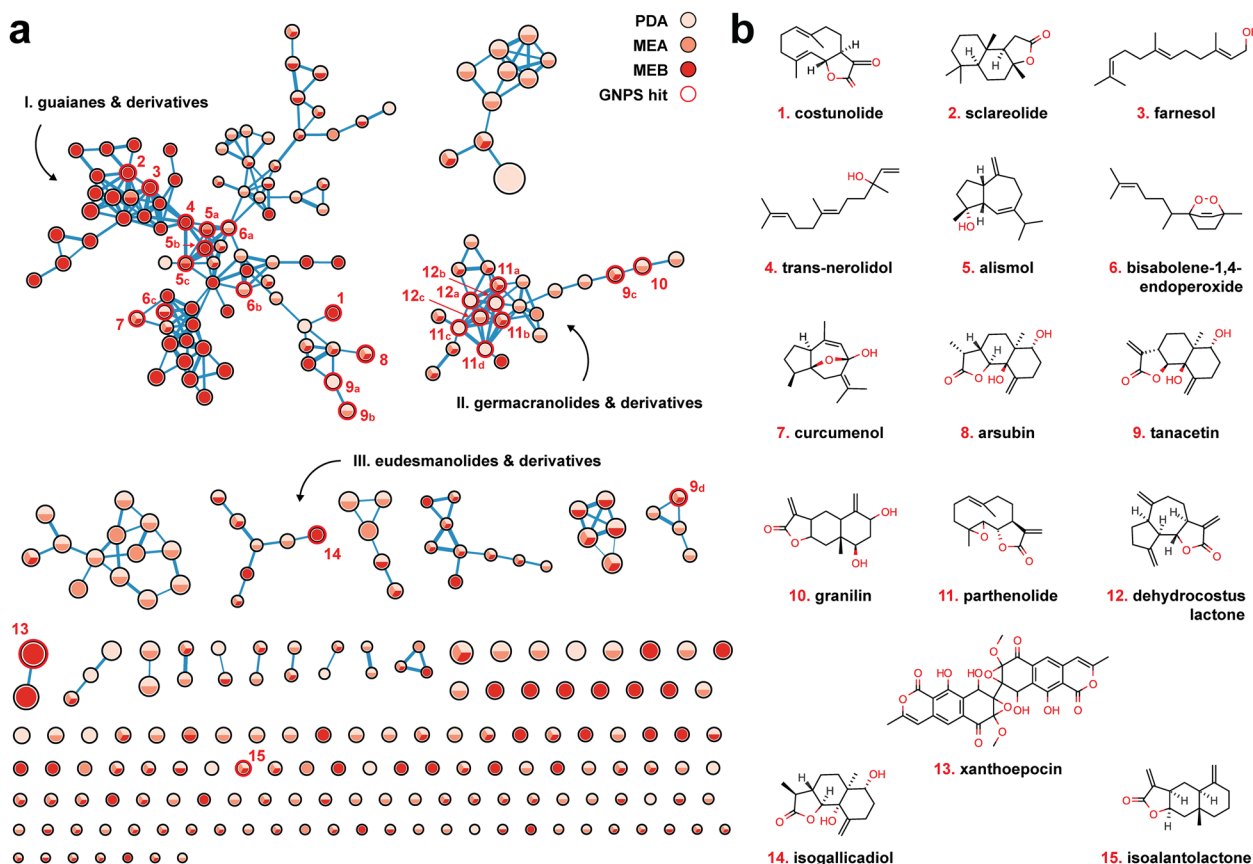


Fig. 4 **a** Molecular network of extracted secondary metabolites from cultures of *A. pinea* F5. Nodes are colored based on their occurrence in the extracts from the different media and sized based on the m/z ratio. Hits from the MS/MS GNPS library search are highlighted in red and numbered. **b** Chemical structures of the metabolites identified via library search, numbered correspondingly

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 the 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 compounds 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 terpene 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. 5). 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 core structures 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

Table 3 Spectral matches from GNPS MS/MS library search

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 ₂ O+H]	-3.41	Linear ST
5a,b,c	Alismol	203.1798	[M-H ₂ O+H]	-0.98	Guaiane ST
6a	Bisabolene-1,4-endoperoxide	201.1638	[M-2H ₂ O+H]	-2.49	Bisabolane ST
6b	Bisabolene-1,4-endoperoxide	219.1732	[M-H ₂ O+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 ₂ O+H]	-1.30	Germacranolide STL
11c,d	Parthenolide	231.1377	[M-H ₂ O+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-naphthopyrone PK
14	Isogallicadiol	267.1593	[M+H]	-1.22	Eudesmanolide STL
15	Isoalantolactone	233.1533	[M+H]	-3.86	Eudesmanolide STL

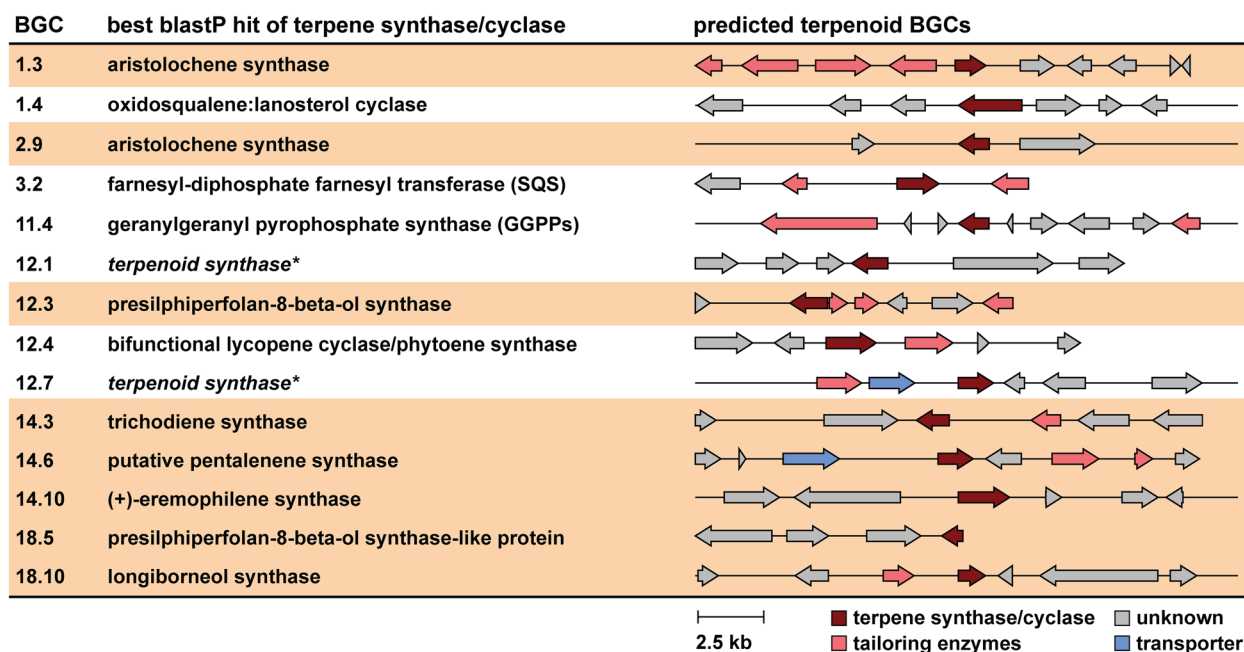


Fig. 5 Analysis of terpenoid BGCs as predicted by fungiSMASH v7.0. Clusters highlighted in tan brown correspond to putative STs and STLs biosynthetic gene clusters, based on blastP analysis of their core terpene synthase/cyclase genes. *The blastP search for the terpene synthase genes in BGCs 12.1 and 12.7 (italic) did not return any annotated hits, therefore these were indicated only as “generic” terpenoid synthases

unidentified, we can speculate that the fungus is likely producing completely new compounds, which remain to be explored in the future.

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. 4a, b). 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 MS¹ level (Additional files 4 and 6, Fig. 4a). Xanthoepocin has a homodimeric structure composed of two naphthopyrone 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 species *Micrococcus luteus* for the MEB extract, albeit only when concentrated 10x (Fig. 6). We also performed agar overlay experiments with 28 days old colonies of *A. pinea* on PDA and MEA and we observed complete growth inhibition of *M. luteus* under both conditions (Additional file 1: Fig. S7). This antibacterial activity may be caused by yet another differentially produced secondary metabolite. However, it is

also possible that the fungus produces xanthoepocin on solid media as well, yet we were not able to detect it via LC–MS analysis. This could be caused by a low extraction efficiency, low starting concentration of the compound, and further degradation of the compound with exposure to light [71]. Because of the scarcity of newly discovered antibiotics and its potency against multi-drug resistant bacteria, xanthoepocin has recently regained attention for the potential development of photodynamic antimicrobial therapies [71]. Therefore, we decided to delve deeper into the genomics data to attempt to identify the BGC involved in its biosynthesis.

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

As discussed in the section above, naphthopyrones 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. 7a, 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-naphthopyrone 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. 7b). 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 (tan box in the figure). Intermediates **1** and **2** are already described in literature [79] as precursors of another naphthopyrone-derived fungal natural product, cercosporin [79, 80]—while compound **4** has been prepared

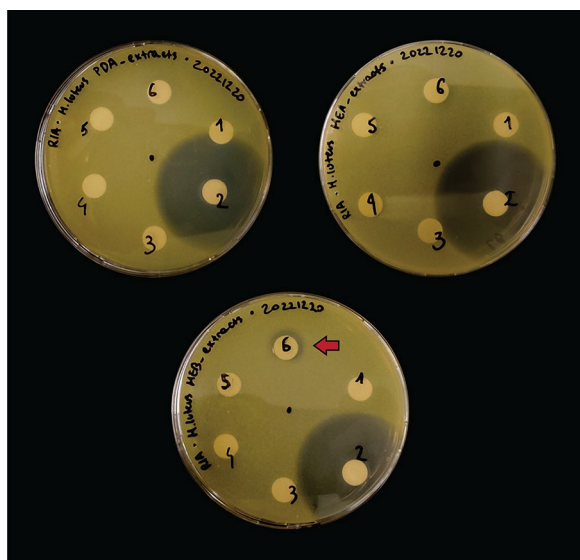


Fig. 6 Antibacterial assays of the extracts from *A. pinea* F5. Top left: PDA extracts; top right: MEA extracts; bottom: MEB extracts. The assays were carried out on LB media overlaid with cultures of indicator strain *M. luteus*, grown overnight at 30 °C. Numbers correspond to: 1—50% MeOH + FA 0.1%; 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

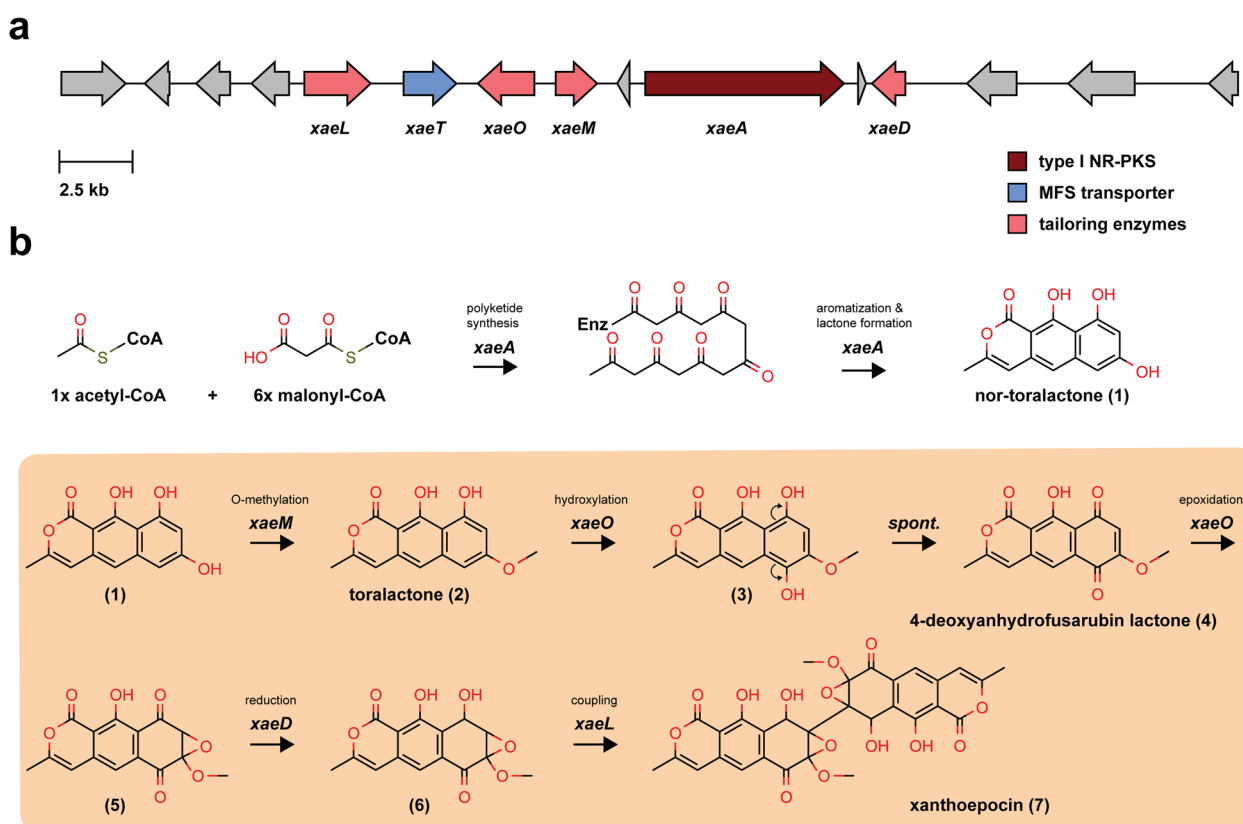


Fig. 7 **a** Putative BGC (ID 18.3) of xanthoepocin as predicted by fungiSMASH v7.0. **b** Proposed biosynthetic pathway of xanthoepocin based on the function of the BGC genes and the chemical structure of the final product. While the biosynthesis of the polyketide scaffold surely occurs as the first step, the order of the remaining steps (tan box) remains to be determined

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

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-naphthopyrone polkyetide 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 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 ddH₂O 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 ddH₂O 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 extract 20 g/L; yeast extract 2 g/L; micro-agar 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 mm² 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 min immediately followed by 2 min of incubation on ice, after which they were centrifuged at max speed for 1 min 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 ITS1 (5′-CTT GGTCATTTAGAGGAAGTAA-3′) and ITS4 (5′-TCC TCCGCTTATTGATATGC-3′) primers. The PCR reaction was performed in a thermal cycler with 2×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 onto 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 ITS1 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 described 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 extract 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; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 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; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0,5 g/L; $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 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; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0,5 g/L; glycerol 180 g/L; dichloran 0,002 g/L). The plates were incubated at 20 °C and the growth of *A. pinea* was monitored regularly until the colonies stopped growing after 4 weeks, at which point the plates were photographed using a Nikon D7500 camera (Nikon, Minato City, Tokyo, Japan) coupled to a Sigma 17-50 mm F/2.8 EX DC OS (Sigma Corporation, Kawasaki, Kanagawa, Japan). The colonies were incubated for further 8 weeks to assess whether any additional morphological change would occur, but none was observed. For morphological comparisons (Additional file 1: Fig. S1), *A. pinea* CBS128205 was purchased from the strain collection of the Westerdijk Institute of Fungal Biodiversity (Utrecht, the Netherlands). Both isolate F5 and *A. pinea* CBS128205 were grown on MEA for 4 weeks at 20 °C, before being imaged as described above.

Extraction of HMW gDNA

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

DNA extraction was carried out with 37 mg of freeze-dried mycelium using the Nucleospin Microbial DNA kit in combination with 3 mm tungsten carbide beads (Qiagen) for tissue disruption in a MM 301 vibratory mill, with significant modifications to the manufacturer's protocol. Briefly, disruption time was reduced to two cycles of 5 s to minimize DNA shearing, with a 30 s pause in between to allow the sample to cool down; each vortexing step was replaced by gentle flicking or mixing by inversion; the washing step with buffer BW was repeated four times, followed by four washes with buffer B5; elution was carried out in 60 μL of pre-warmed EB buffer

(~55 °C), and the eluate was then re-applied to the column, incubated for one additional minute and re-eluted to maximize DNA recovery. The integrity of the purified DNA was assessed by agarose gel 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 onto a primed FLO-MIN112 (ID: FAT29688) flow cell on a MinION device for a 42-h 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. S5). The draft assembly was subsequently polished in two rounds. First, Racon v1.4.10 [95] was used with the following settings: *-m 8 -x -6 -g -8 -w 500*, optimized for combined use with Medaka. Second, Medaka v1.6.0 was used on the polished version with the r104_e81_sup_g5015 model. The completeness of the assembly was evaluated using BUSCO v5.3.2 (ascmycota_odb10 and sordariomycetes_odb10 datasets) [34]. To assemble the mitochondrial DNA, we subsampled the filtered reads to generate a set of 10,000 reads and then used Flye v2.9-b1778 as described above. The

only circular contig obtained was analyzed with BLAST against the NCBI nucleotide collection to confirm its organellar origin. Lastly, the mitochondrial genome was annotated using GeSeq [35] using default settings.

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

Structural and functional genome annotation

Genome annotation was carried out on the polished assembly using the online platform Genome Sequence Annotation Server (GenSAS) v6.0, which provides a pipeline for whole genome structural and functional annotation [96]. Standard settings were used unless otherwise mentioned. In brief, low complexity regions and repeats were masked using RepeatModeler v2.0.3 and RepeatMasker v4.1.1 [97], setting the DNA source to 'Fungi'. The newly generated masked consensus sequence was used for ab initio gene prediction using the following tools: (I) Augustus v3.4.0 [98], selecting *Fusarium graminearum* as a trained organism; (II) GeneMarkES v4.48 [99]. For homology-based prediction, the NCBI reference transcript and protein databases for Fungi were searched, using (III) blastn v2.12.0 [89] and (IV) DIAMOND v2.0.11 [100], respectively. To generate the final consensus gene model, EvidenceModeler v1.1.1 [101] was used on the above-mentioned predictions, weighted as follows: (I)–five, (II)–five, (III)–ten, (IV)–ten. For prediction of proteins with PFAM domains, the Pfam module within GenSAS was used with the following parameters: E-value sequence: 1; E-value Domain: 10. The tRNA and rRNA were predicted using tRNA scan-SE v2.0.11 ("check for pseudogenes=OFF") [102] and barrnap v0.9 [103], respectively. Prediction of proteins with signal peptides and/or transmembrane domains was carried out on the web server Phobius [104]. Gene Ontology (GO) annotation was performed using the webtool PANNZER2 [105].

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

Secondary metabolite biosynthetic gene clusters were identified using the fungal suite of antiSMASH web server v7.0 with the default settings (accessed on 31 Jan 2023) [53]. To annotate CAZymes, the web server dbCAN2 (accessed on 3 Feb 2023) was used [37]. The integrated HMMER, DIAMOND, and HMMER-dbCAN-sub tools were used on the total proteome of *A. pinea*. The three outputs were automatically combined, and CAZymes predicted only by 1/3 of the tools were removed to improve the annotation accuracy. The substrate specificity of the final hits was extracted from the individual results of HMMER-dbCAN-sub.

For the prediction of P450s, the BLAST tool of biocat-net CYPED v6.0 [106] was used (accessed on 3 Feb 2023), with E-value cutoff set at 1.0×10^{-10} . To identify putative UPOs, the query sequences of the prototype enzymes *AaeUPO* [50] and *HspUPO* [51] were retrieved from the Uniprot database (accessed on 23 May 2023) [107], and submitted to phmmer (HMMER v3.3.2) to search against the total proteome of *A. pinea*. The cutoff was set at E-value of 0.01. The multiple sequence alignment analysis between the putative UPOs from *A. pinea* and the prototype *AaeUPO* and *HspUPO* was performed with MEGA 11 [90] using the MUSCLE algorithm [91] with default settings, and visualized in Jalview [108].

Extraction of secondary metabolites and HRMS-MS² 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) supplemented 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 a binary LC20ADXR pump 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. 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 MS¹ and MS² levels using the peak picking filter (algorithm set to “Vendor”). The data files were then pre-processed with MZMine v2.53 [58] to generate a feature list (.csv file) and a MS/MS spectra list (.mgf file). The converted MS/MS data, feature list, and the .mgf file 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 5 mL 2×yeast extract tryptone medium (2×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 2×YT supplemented with 0.6% agar, mixed well and immediately overlaid on the agar plates (3 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 2×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 min. Six sterile diffusion disks (10 mm ϕ , MilliporeSigma, Burlington, MA, USA) were then gently positioned on the surface of the agar. Each disk was soaked with either 10 μ L of fungal extract, fungal extract 10x, medium extract, medium extract 10x, 50% MeOH + FA 0.1% (blank solvent), and ampicillin 2 mg/mL (20 μ g) as positive control. The plates were incubated overnight and inspected for antibacterial activity.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40694-023-00170-1>.

Additional file 1: Tables S1–S5, and figures S1–S7

Additional file 2: Structural and functional annotation of the genome of *A. pinea* F5.

Additional file 3: BGC prediction by fungiSMASH7 on the genome of *A. pinea* F5

Additional file 4: Molecular network of *A. pinea* F5 extracts and annotation by Sirius (Cytoscape session file).

Additional file 5: Molecular network of *A. pinea* F5 extracts annotated with MolNetEnhancer (Cytoscape session file)

Additional file 6: List of hits from GNPS MS/MS library search of *A. pinea* extracts.

Additional file 7: Extracted ion chromatograms, MS and MS/MS spectra of annotated nodes from the molecular network.

Additional file 8: Results of blastP search on core synthase/cyclase genes from predicted terpene BGCs in the genome of *A. pinea* F5. The best annotated hits for each blastP search are highlighted in yellow.

Acknowledgements

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

Author contributions

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

Availability of data and materials

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.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None declared.

Received: 26 October 2023 Accepted: 7 December 2023

Published online: 03 January 2024

References

- Redkar A, Sabale M, Zuccaro A, Di Pietro A. Determinants of endophytic and pathogenic lifestyle in root colonizing fungi. *Curr Opin Plant Biol.* 2022;67: 102226. <https://doi.org/10.1016/j.pbi.2022.102226>.
- Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. Fungal lifestyle reflected in serine protease repertoire. *Sci Rep.* 2017;7(1):9147. <https://doi.org/10.1038/s41598-017-09644-w>.
- Boddy L, Hiscox J. Fungal ecology: principles and mechanisms of colonization and competition by saprotrophic fungi. *Microbiol Spectr.* 2016. <https://doi.org/10.1128/microbiolspec.funk-0019-2016>.
- 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.* 2022;8(2):205.
- Sagita R, Quax WJ, Haslinger K. Current state and future directions of genetics and genomics of endophytic fungi for bioprospecting efforts. *Front Bioeng Biotechnol.* 2021;15:9. <https://doi.org/10.3389/fbioe.2021.649906/full>.
- Galindo-Solís JM, Fernández FJ. Endophytic fungal terpenoids: natural role and bioactivities. *Microorganisms.* 2022;10(2):339.
- Manganyi MC, Ateba CN. Untapped potentials of endophytic fungi: a review of novel bioactive compounds with biological applications. *Microorganisms.* 2020;8(12):1934.
- Keller NP. Fungal secondary metabolism: regulation, function and drug discovery. *Nat Rev Microbiol.* 2019;17(3):167–80. <https://doi.org/10.1038/s41579-018-0121-1>.
- Pusztahelyi T, Holb IJ, Pócsi I. Secondary metabolites in fungus-plant interactions. *Front Plant Sci.* 2015;6(6):1–23. <https://doi.org/10.3389/fpls.2015.00573/abstract>.
- Becker K, Stadler M. Recent progress in biodiversity research on the Xylariales and their secondary metabolism. *J Antibiot (Tokyo).* 2021;74(1):1–23. <https://doi.org/10.1038/s41429-020-00376-0>.
- Cañón ERP, de Albuquerque MP, Alves RP, Pereira AB, de Victoria F. Morphological and molecular characterization of three endolichenic isolates of Xylaria (Xylariaceae) from *Cladonia curta* Ahti & Marcelli (Cladoniaceae). *Plants.* 2019;8(10):399.
- Tang AMC, Jeewon R, Hyde KD. A re-evaluation of the evolutionary relationships within the Xylariaceae based on ribosomal and protein-coding gene sequences. *Fungal Divers.* 2009;34:127–55.
- Amirzakariya BZ, Shakeri A. Bioactive terpenoids derived from plant endophytic fungi: an updated review (2011–2020). *Phytochemistry.* 2022;197: 113130. <https://doi.org/10.1016/j.phytochem.2022.113130>.
- Gupta S, Chaturvedi P, Kulkarni MG, Van Staden J. A critical review on exploiting the pharmaceutical potential of plant endophytic fungi. *Biotechnol Adv.* 2020;39: 107462. <https://doi.org/10.1016/j.biotechadv.2019.107462>.
- Crous PW, Groenewald JZ. *Anthostomella pinea* fungal planet 53. *Persoonia.* 2010;20(24):126–7.
- The Index Fungorum. <http://indexfungorum.org>. Accessed on 16 Oct 2023.
- Izumikawa M, Itoh M, Kawahara T, Sakata N, Tsuchida T, Mizukami T, et al. A highly oxygenated ergostane—MBJ-0005—from *Anthostomella eucalyptorum* f25427. *J Antibiot.* 2014;67(12):843–5.
- Daranagama DA, Camporesi E, Jeewon R, Liu X, Stadler M, Lumyong S, et al. Taxonomic rearrangement of *Anthostomella* (Xylariaceae) based on a multigene phylogeny and morphology. *Cryptogam Mycol.* 2016;37(4):509–38. <https://doi.org/10.7872/crym/v37.iss4.2016.509>.
- 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.* 1985;1(34):1481.
- NCBI Genome database. <https://www.ncbi.nlm.nih.gov/genome>. Accessed on 16 Oct 2023.
- JGI Mycocosm database. <https://mycocosm.jgi.doe.gov/mycocosm/home>. Accessed on 16 Oct 2023.
- 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: Mérillon JM, Ramawat K, editors. *Co-Evolution of Secondary Metabolites Reference Series in Phytochemistry*. Cham: Springer; 2020.
- 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.* 2009;58(3):283–97.
- Tuovinen V, Ekman S, Thor G, Vanderpool D, Spribille T, Johannesson H. Two Basidiomycete fungi in the cortex of wolf lichens. *Curr Biol.* 2019;29(3):476–483.e5.
- Jenkins B, Richards TA. Symbiosis: wolf lichens harbour a choir of fungi. *Curr Biol.* 2019;29(3):R88–90. <https://doi.org/10.1016/j.cub.2018.12.034>.
- 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.* 2020;112(5):3150–6. <https://doi.org/10.1016/j.ygeno.2020.06.006>.
- Yamamoto Y, Miura Y, Higuchi M, Kinoshita Y, Yoshimura I. Using lichen tissue cultures in modern biology. *Bryologist.* 1993;96(3):384–93.
- Tapia de Daza MS, Beuchat LR. Suitability of modified dichloran glycerol (DG18) agar for enumerating unstressed and stressed xerophilic molds. *Food Microbiol.* 1992;9(4):319–33.

29. Krain A, Siupka P. Fungal guttation, a source of bioactive compounds and its ecological role—a review. *Biomolecules*. 2021;11(9):1270.
30. Kalra R, Conlan XA, Goel M. Fungi as a potential source of pigments: harnessing filamentous fungi. *Front Chem*. 2020;8:8. <https://doi.org/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*. 2017;80(3):756–70. <https://doi.org/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*. 2012;99(5):898–914. <https://doi.org/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*. 2016;98:210–32. <https://doi.org/10.1016/j.mycenv.2016.02.010>.
34. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
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*. 2017;45(W1):W6–11.
36. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. 2022;50(D1):D571–7.
37. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95–101.
38. Wu HY, Mortensen UH, Chang FR, Tsai H. Whole genome sequence characterization of *Aspergillus terreus* ATCC 20541 and genome comparison of the fungi *A terreus*. *Sci Rep*. 2023;13(1):194.
39. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, et al. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol*. 2017;18(1):28. <https://doi.org/10.1186/s13059-017-1151-0>.
40. Clavaud C, Aïmanianda V, Latge JP. Organization of fungal, oomycete and Lichen (1,3)- β -glucans chemistry, biochemistry, and biology of 1–3 beta glucans and related polysaccharides. Amsterdam: Elsevier; 2009.
41. Zhang X, Guo J, Cheng F, Li S. Cytochrome P450 enzymes in fungal natural product biosynthesis. *Nat Prod Rep*. 2021;38(6):1072–99.
42. O'Reilly E, Köhler V, Flitsch SL, Turner NJ. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chem Commun*. 2011;47(9):2490.
43. Sang H, Hulvey JP, Green R, Xu H, Im J, Chang T, et al. A xenobiotic detoxification pathway through transcriptional regulation in filamentous fungi. *MBio*. 2018. <https://doi.org/10.1128/mBio.00457-18>.
44. Burns K, Helsby NA. Cytochrome P450 in IUPHAR/BPS Guid to Pharmacol CITE. *GtoPdb v20231*. 2023. <https://doi.org/10.2218/gtopdb/F242/2023.1>.
45. Ahmadjian V. Lichens are more important than you think. *Bioscience*. 1995;45(3):124–124. <https://doi.org/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 developments in applied microbiology and biotechnology synergistic approaches for bioremediation of environmental pollutants recent advances and challenges. Amsterdam: Elsevier; 2022.
47. Hofrichter M, Kellner H, Pecyna MJ, Fungal UR. Unspecific Peroxygenases: Heme-Thiolate Proteins That Combine Peroxidase and Cytochrome P450 Properties. In: Hrycak E, Bandiera S, editors. *Monoxygenase Peroxidase and Peroxygenase Properties*. Cham: Springer; 2015.
48. Hofrichter M, Kellner H, Herzog R, Karich A, Liers C, Scheibner K, et al. Fungal peroxigenases a phylogenetically old superfamily of heme enzymes with promiscuity for oxygen transfer reactions grand challenges in fungal biotechnology. *Grand Challenge Biol iotechnol*. 2020;9:369–403. https://doi.org/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*. 2023;41:100786.
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*. 2004;70(8):4575–81. <https://doi.org/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*. 2021;11(18):11511–25. <https://doi.org/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*. 2019;19(1):76. <https://doi.org/10.1186/s12862-019-1394-3>.
53. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 70: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res*. 2023;51(W1):W46–50.
54. Godio RP, Fouces R, Martín JF. A squalene epoxidase is involved in biosynthesis of both the antitumor compound clavarinic acid and sterols in the basidiomycete *H sublateritium*. *Chem Biol*. 2007;14(12):1334–46.
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*. 2014;4(1):6728.
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*. 2006;61(4):1069–80. <https://doi.org/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*. 2022;233(3):1317–30. <https://doi.org/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 Bioinform*. 2010;11(1):395. <https://doi.org/10.1186/1471-2105-11-395>.
59. Wang M, Carver JJ, Phelan VV, 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*. 2016;34(8):828–37.
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*. 2020;17(9):905–8.
61. Ernst M, Bin KK, Caraballo-Rodríguez AM, Nothias LF, Wandy J, Chen C, et al. MolNetEnhancer enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites*. 2019;9(7):144.
62. Fraga BM. Natural sesquiterpenoids. *Nat Prod Rep*. 2013;30(9):1226.
63. Nishikawa K, Aburai N, Yamada K, Koshino H, Tsuchiya E, Kimura K. The bisabolane sesquiterpenoid endoperoxide, 3,6-epidioxy-1,10-bisaboladiene, isolated from *Cacalia delphinifolia* Inhibits the growth of human cancer cells and induces apoptosis. *Biosci Biotechnol Biochem*. 2008;72(9):2463–6.
64. Kimuraichi K, 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*. 2012;20(12):3887–97.
65. Lo JY, Kamarudin MNA, Hamdi OAA, Awang K, Kadir HA. Curcumenol isolated from *Curcuma zedoaria* suppresses Akt-mediated NF- κ B activation and p38 MAPK signaling pathway in LPS-stimulated BV-2 microglial cells. *Food Funct*. 2015;6(11):3550–9.
66. Tamez-Fernández JF, Melchor-Martínez EM, Ibarra-Rivera TR, Rivas-Galindo VM. Plant-derived endoperoxides: structure, occurrence, and bioactivity. *Phytochem Rev*. 2020;19(4):827–64.
67. Krishna S, Bustamante L, Haynes RK, Staines HM. Artemisinin: their growing importance in medicine. *Trends Pharmacol Sci*. 2008;29(10):520–7.
68. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinform*. 2009;10(1):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*. 2020;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)*. 2000;53(9):928–33.
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*. 2022;21(1):1. <https://doi.org/10.1186/s12934-021-01718-9>.
72. Cary JW, Harris-Coward PY, Ehrlich KC, Di Mavungu JD, Malysheva SV, De Saeger S, et al. Functional characterization of a veA-dependent polyketide synthase gene in *Aspergillus flavus* necessary for the synthesis of asparosone, a sclerotium-specific pigment. *Fungal Genet Biol*. 2014;64:25–35. <https://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*. 2012;55(12):1100–8. <https://doi.org/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*. 2014;19(6):7169–88.
75. Fürtges L, Obermaier S, Thiele W, Foegen S, Müller M. Diversity in fungal intermolecular phenol coupling of polyketides: regioselective laccase-based systems. *ChemBioChem*. 2019;20(15):1928–32. <https://doi.org/10.1002/cbic.201900041>.
76. Hüttel W, Müller M. Regio- and stereoselective intermolecular phenol coupling enzymes in secondary metabolite biosynthesis. *Nat Prod Rep*. 2021;38(5):1011–43.
77. Xu D, Yin R, Zhou Z, Gu G, Zhao S, Xu JR, et al. Elucidation of ustilagoinidin biosynthesis reveals a previously unrecognized class of ene-reductases. *Chem Sci*. 2021;12(44):14883–92.
78. Urquhart AS, Hu J, Chooi YH, Idnurm A. The fungal gene cluster for biosynthesis of the antibacterial agent viriditoxin. *Fungal Biol Biotechnol*. 2019;6(1):9. <https://doi.org/10.1186/s40694-019-0072-y>.
79. Newman AG, Townsend CA. Molecular characterization of the cercosporin biosynthetic pathway in the fungal plant pathogen *Cercospora nicotianae*. *J Am Chem Soc*. 2016;138(12):4219–28. <https://doi.org/10.1021/jacs.6b00633>.
80. Newman AG, Vagstad AL, Belecki K, Scheerer JR, Townsend CA. Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun*. 2012;48(96):11772.
81. Barbier M, Devys M, Parisot D. A simple synthesis of 4-deoxyanhydrofusarubin lactone. *Synth Commun*. 1993;23(5):651–6. <https://doi.org/10.1080/00397919308009823>.
82. The Species Fungorum. <https://www.speciesfungorum.org>. Accessed 16 Oct 2023.
83. Phukhamsakda C, Nilsson RH, Bhunjun CS, de Farias ARG, Sun YR, Wijesinghe SN, et al. The numbers of fungi: contributions from traditional taxonomic studies and challenges of metabarcoding. *Fungal Divers*. 2022;114(1):327–86. <https://doi.org/10.1007/s13225-022-00502-3>.
84. Bandi CK, Agrawal A, Chundawat SP. Carbohydrate-Active enZyme (CAZyme) enabled glycoengineering for a sweeter future. *Curr Opin Biotechnol*. 2020;66:283–91.
85. Pallister E, Gray CJ, Flitsch SL. Enzyme promiscuity of carbohydrate active enzymes and their applications in biocatalysis. *Curr Opin Struct Biol*. 2020;65:184–92.
86. Mhiri S, Bouanane-Darenfed A, Jemli S, Neifar S, Ameri R, Mezghani M, et al. A thermophilic and thermostable xylanase from *Caldicoprobacter algeriensis*: recombinant expression, characterization and application in paper biobleaching. *Int J Biol Macromol*. 2020;116(4):808–17.
87. Karupiah V, Zhixiang L, Liu H, Vallikkannu M, Chen J. Co-culture of Ve1-overexpressed *Trichoderma asperellum* and *Bacillus amyloliquefaciens*: an eco-friendly strategy to hydrolyze the lignocellulose biomass in soil to enrich the soil fertility, plant growth and disease resistance. *Microb Cell Fact*. 2021;20(1):57. <https://doi.org/10.1186/s12934-021-01540-3>.
88. Urlacher VB, Girhard M. Cytochrome P450 monooxygenases in biotechnology and synthetic biology. *Trends Biotechnol*. 2019;37(8):882–97.
89. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
90. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7.
91. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. 2004;5(113):1–19. <https://doi.org/10.1186/1471-2105-5-113>.
92. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34(15):2666–9.
93. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–5.
94. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 2015;31(20):3350–2.
95. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;27(5):737–46. <https://doi.org/10.1101/gr.214270.116>.
96. Humann JL, Lee T, Ficklin S, Main D. Structural and functional annotation of eukaryotic genomes with GenSAS. *Methods Mol Biol*. 2019;1962:29–51. https://doi.org/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. *Nucl Acids Res*. 2005. <https://doi.org/10.1093/nar/gki458>.
99. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark hmh-E and GeneMark-ES. *Curr Protoc Bioinforma*. 2011. <https://doi.org/10.1002/0471250953.bi040635>.
100. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
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*. 2008;9(1):R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
102. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In: Kollmar M, editor. *Gene Prediction*. Berlin: Springer; 2019.
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*. 2012;41(D1):D590–6.
104. Käll L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*. 2004;338(5):1027–36.
105. Törönen P, Holm L. PANNZER — A practical tool for protein function prediction. *Protein Sci*. 2022;31(1):118–28.
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*. 2016;17(21):2093–8. <https://doi.org/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*. 2023;51(D1):D523–31.
108. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.
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*. 2012;30(10):918–20.
110. WinSCP FTP client. <https://winscp.net/eng/index.php>. Accessed on Nov 15 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*. 2003;13(11):2498–504. <https://doi.org/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*. 2020;38(1):23–6.
113. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods*. 2019;16(4):299–302.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.