scientific **data**

Check for updates

DATA DESCRIPTOR

Genome sequence and annotation of *Periconia digitata* a hopeful biocontrol agent of phytopathogenic oomycetes

Elena Bovio [1,3 ✉], Corinne Rancurel [1,3 ✉], Aurélie Seassau[1], Marc Magliano[1], Marie Gislard[2], Anaïs Loisier[2], Claire Kuchly[2], Michel Ponchet[1], Etienne G. J. Danchin [1] & Cyril Van Ghelder[1]

The *Periconia* fungal genus belongs to the phylum Ascomycota, order Pleosporales, family Periconiaceae. *Periconia* are found in many habitats, but little is known about their ecology. Several species from this genus produce bioactive molecules. *Periconia digitata* extracts were shown to be deadly active against the pine wilt nematode. Furthermore, *P. digitata* was shown to inhibit the plant pathogenic oomycete *Phytophthora parasitica*. Because *P. digitata* has great potential as a biocontrol agent and high quality genomic resources are still lacking in the Periconiaceae family, we generated long-read genomic data for *P. digitata*. Using PacBio Hifi sequencing technology, we obtained a highly-contiguous genome assembled in 13 chromosomes and totaling ca. 39 Mb. In addition, we produced a reference transcriptome, based on 12 different culture conditions, and proteomic data to support the genome annotation. Besides representing a new reference genome within the Periconiaceae, this work will contribute to our better understanding of the Eukaryotic tree of life and opens new possibilities in terms of biotechnological applications.

Background/Introduction

Taxonomy. In 1934, the family Periconiaceae was established with *Periconia* as type genus[1]. Subsequently, species belonging to the genus *Periconia* have been assigned to the family Massarinaceae, but a recent systematic revision brought all the *Periconia* spp. in the family Periconiaceae[2]. The *Periconia* genus was established in 1791 by Tode Ex Fries[1], while the first record is dated back to a fossil preserved in the Baltic amber from the upper Eocene epoch (34–38 Mya)[3]. The genus comprises 211 species epithets in Index Fungorum (2022) with 25 of them having a new current name belonging to another genus. To date, only 27 species have been confirmed by molecular data[4].

Ecology. The habitat of *Periconia* is various, the reports start from the sea[5,6] to the Himalayas[7]. Although the genus *Periconia* is present in many habitats, little is known about its ecology. To the best of our knowledge, most species present a saprobe behaviour[1,4,8,9], while some are described as endophytes and are included in the dark septate endophytes - DSE group[10–12]. Some endophytic *Periconia* spp. were found to be facultative parasites on the invasive weed *Parthenium hysterophorus* (Asteraceae)[13]. *P. circinata* was described as the causal agent of root rot on *Sorghum* spp. ("milo disease")[14]. *Periconia* spp. were also found associated with human keratomycosis[15].

Secondary metabolites and biological activities. Many studies only focused on the ability of given species (not always identified at species level) to produce molecules of biotechnological interest. Since 1969, the genus has yielded 104 compounds belonging to terpenoids, polyketides, aromatic ketone and phenolics, some exhibiting interesting biological activities including antimicrobial (bacteria, fungi), antiviral, anti-inflammatory and cytotoxic activity[16]. The anti-oomycetes activity of a *Periconia* has been explored only once with *Periconia*

[1]Institut Sophia Agrobiotech, INRAE 1355, CNRS and Université Côte d'Azur, 400, Route des Chappes, BP 167, 06903, Sophia Antipolis Cedex, France. [2]GeT-PlaGe (genomic platform), Campus INRAE, 24 chemin de borde rouge, Auzeville CS 52627, 31326, CASTANET-TOLOSAN Cedex, France. [3]These authors contributed equally: Elena Bovio, Corinne Rancurel. ✉e-mail: ele.bovio@gmail.com; corinne.rancurel@inrae.fr
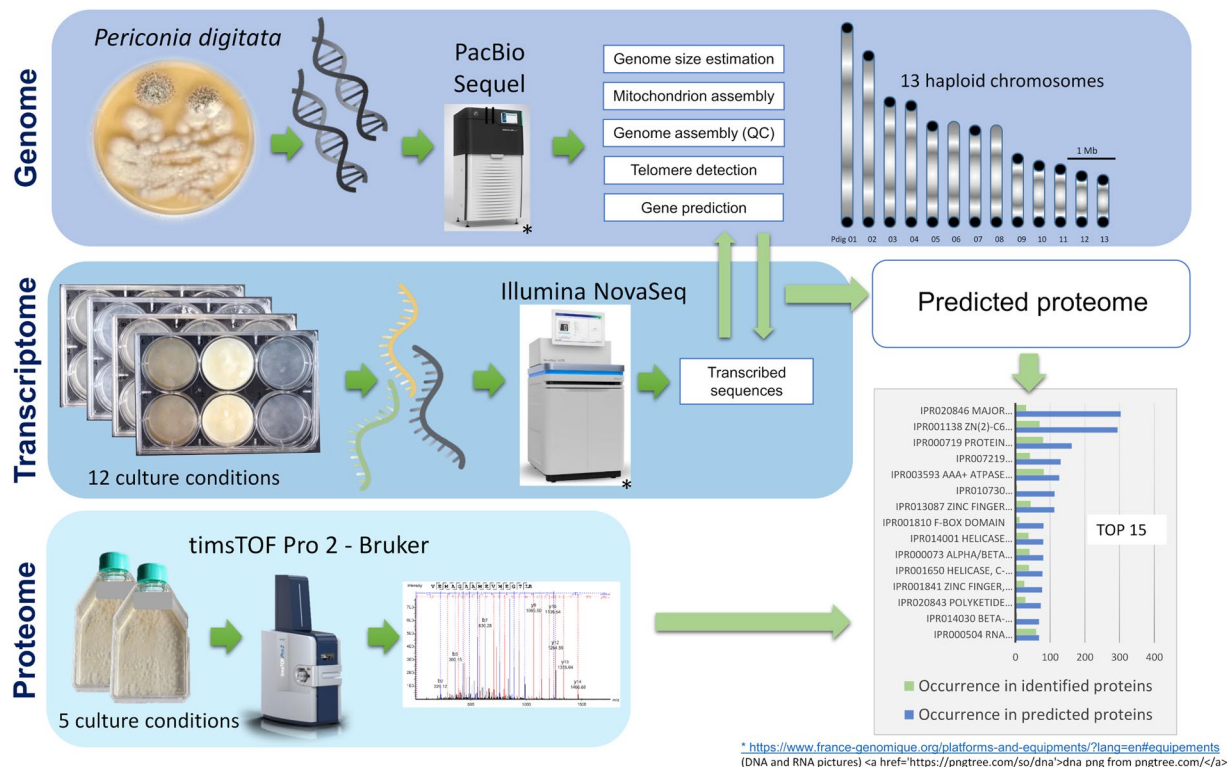
Fig. 1 Schematic overview of the study design.

digitata (strain Y3), previously misidentified as *Phoma* sp. CNCM I-4278[17,18]. *P. digitata* CNCM I-4278 was able to inhibit the growth and cyst germination of the plant pathogenic oomycete *Phytophthora parasitica* both *in vitro* and in planta, without phytotoxicity[17,18]. In addition to anti-oomycete activity, the water filtrate and/or the crude extract of *P. digitata* CNCM I-4278 also inhibited the growth of several phytopathogenic fungi[17]. In another screening of fungal culture filtrates isolated from freshwater submerged wood, a *P. digitata* strain was highly active on the fungivorous and phytophagous nematode *Bursaphelenchus xylophilus* responsible for dramatic losses in pine forests. Indeed, 70% to 80% of nematodes were killed 48 h after the treatment with *P. digitata* extracts[19].

Multi-omic resources. Owing to its high potential as a biocontrol agent for plant protection against several classes of problematic plant pathogens, we present the chromosome-scale genome assembly and annotation of *P. digitata* CNCM I-4278. To date, only 2 other genomes were available in the genus *Periconia* and the family Periconiaceae, the genomes of *Periconia macrospinosa*[20] and *Periconia sp*. R9002. However, these genome assemblies, made of hundreds of contigs/scaffolds, appear to be much fragmented compared to the genomes that can be obtained nowadays using long read technologies. Therefore, we used a PacBio HiFi sequencing technology to produce highly accurate long reads and assembled the *P. digitata* genome in 13 haploid chromosomes with a total length of ca. 39 Mb. The genome annotation was guided by the assembly of a reference transcriptome composed of the transcriptomes of 12 different culture conditions, including stresses, leading to very different mycelium phenotypes. The annotation revealed 15,520 protein-coding genes and conserved InterPro domains could be identified in 60% of the proteins. In addition, we carried out Nano-HPLC-HRMS analyses to characterize the proteome of *P. digitata* and strengthen our functional annotation. The proteomic analysis retrieved and confirmed more than one-third of the predicted proteins no matter the chosen parameter (1 or 2 unique peptides).

Overall, this work generated a high-quality genome completed with rich transcriptomic and proteomic data that will be useful to future research (Fig. 1). This study will constitute an important resource for our further understanding of the Eukaryotic tree of life and for future comparative genomics[21]. This will help to better delineate the evolution within fungi, a kingdom that is constantly revisited from a systematic point of view on the basis of molecular markers rather than on life history traits. This work will also contribute to bringing new insights into the *Pleosporales*, potentially the largest order of Dothideomycetes that account for more than 300 genera and 4,700 species[22] and contains only 109 sequenced genomes Mycocosm Portals (doe.gov)[22] (2022).

Methods

Strain identification. The strain *Phoma* sp. CNCM I-4278 was previously isolated from the rhizosphere of *Nicotiana tabacum* (cv Xanthi, Solanaceae) grown under controlled conditions[17]. The fungus was identified according to the closest similarity of its 18 S rRNA sequence (HM161743[23]) with those present in GenBank by that time and it was deposited in the National Collection of Institut Pasteur (CNCM I-4278)[17,18]. This first attempt to identify the strain was uncertain since the alignment showed more than 30 mismatches with a putative *Phoma*

sp. However, the increasing availability of fungal sequences in the database allowed a taxonomic reevaluation of the strain.

The fungus was cultivated for 2 weeks on Petri dishes containing Potato Dextrose Agar (PDA - 20 g glucose, 4 g potato extract, 15 g agar, up to 1 L Milli-Q® water) in order to obtain sufficient biomass to perform a DNA extraction. The DNA extraction protocol was optimized in our laboratory starting from previous works[24–26]. Briefly, about 100 mg of mycelium were placed in 2 mL Eppendorf tubes with two steel beads and disrupted in a MM301 tissue lyzer (Retsch GmbH, Haan, Germany). Then, a volume of 1 mL of lysis buffer CTAB (28 mM NaCl, 2 mM Tris-base, 0.4 mM Na$_2$EDTA, pH 8) plus 2% Polyvinylpyrrolidone (PVP) was added to the samples. The tubes were incubated at 65 °C for 2 h. 400 μL of chloroform:isoamyl alcohol (24:1 v/v) were then added to the samples that were vortexed and centrifuged for 5 min at 13,000 rpm. The supernatant (600 μL) was transferred to a new Eppendorf tube and 100 μL of 10 M ammonium acetate was added; the solution was gently mixed and the samples were incubated at 4 °C for 20 min. The mixture was vortexed and centrifuged for 10 min at 13,000 rpm, the supernatant (650 μL) was transferred to a new Eppendorf tube and one volume of isopropanol (kept at −20 °C) was added. The sample was gently mixed and incubated overnight at −20 °C in order to precipitate the DNA. The final pellet was collected by centrifugation for 5 min at 13,000 rpm at 4 °C. The supernatant was discarded, and the pellet was washed with 500 μL of 75% aqueous ethanol (kept at −20 °C) and recovered by centrifugation for 2 min at 13,000 rpm. The supernatant was discarded, and the pellet was dried under the airflow of a chemical hood. Then, the DNA was resuspended in 30 μL of sterile Milli-Q® water.

The DNA quality and quantity were evaluated using a NanoDrop 2000 (Thermo Scientific, Wilmington, USA). The DNA was stored at −20 °C.

The obtained DNA was used to amplify partial sequences of two genetic markers. The primer pairs ITS1/ITS4[27] and LR0R/LR5[28] were used to amplify the internal transcribed spacers and the 28 S large ribosomal subunit (nrLSU) region, respectively. The PCR reaction was performed in 25 μL final volumes and consisted of 12.5 μL GoTaq® G2 Hot Start Colorless Master Mix (2X - Promega), 1 μL of each primer (10 μM), 5 μL genomic DNA extract (10 ng/μL) and 5.5 μL nuclease-free water. PCR products were loaded on 2% agarose gel electrophoresis in 0.5X Tris-acetate-EDTA buffer. The gel was stained with ethidium bromide and the PCR products were visualized under UV light. The PCR products were sequenced (accession numbers OP329216[29] - ITS; and OP329219[30] - 28 S) and used to build a Maximum Likelihood (ML) phylogenetic tree with MEGA X[31]. The dataset used to build the tree is shown in the Table 1. Among the 47 *Periconia* species having sequences deposited in GenBank, only 27 exhibited both ITS and 28 S sequences of sufficient length (Table 1) allowing to build trees from trimmed and concatenated ITS-28S. For the sequences selection a priority has been given to strains cultured from holotype or considered"reference strains" from culture collections. The phylogenetic tree was inferred by ML method with the Tamura-Nei model of evolution[32]. The tree with the highest log likelihood (−6834,70) is shown (Fig. 2). The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0,3430)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 29,28% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

The resulting tree allowed unambiguously identifying the strain *Phoma* sp. CNCM I-4278 as *Periconia digitata* whose name was used further throughout this work.

Periconia digitata culture conditions. Unless otherwise stated, *P. digitata* was grown in 6-well plates containing 5 mL of sterile media (Table 2) that were inoculated with a spore suspension (1.5×10^4 spores/mL final concentration). For the high molecular weight DNA extraction, the fungus was grown on Potato Dextrose Broth - PDB (Potato extract 4 g/L, Dextrose 20 g/L) and incubated at 24 °C, in the dark, for 7 days, prior to DNA extraction.

In order to generate a reference transcriptome of *P. digitata*, different culture conditions were selected and several stress factors were applied to capture a variety of transcripts and generate an as comprehensive as possible transcriptome. We applied (or not) light, salt, heat, cold, oxidative and heavy metal stresses. We also introduced complex media of plant origin that induced very different phenotypes in terms of mycelium organization. The liquid culture conditions were prepared in 6-well plates with two wells for each condition in the dark. A solid culture was also prepared in Petri dishes (9 cm Ø); the plates were inoculated in three spots with 10 μL of the same spore suspension used for the liquid culture.

The 12 culture conditions, the stress factors and when they were applied are detailed in Table 2. The plates were incubated for 8 days prior to RNA extractions.

High molecular weight DNA extraction.
High molecular weight DNA extraction was performed using the MasterPure™ Complete DNA and RNA Purification Kit from Epicentre. Seven days old-mycelia of *P. digitata* were retrieved from a 6-well plate on PDB as described above and directly ground in liquid nitrogen with sterilized pestles and mortars. The resulting powder was transferred in six microtubes containing 1 μl of proteinase K and 300 μl of tissue and cell lysis solution for each tube and homogenized. All homogenisation steps were performed gently to avoid DNA fragmentation. Tubes were incubated at 65 °C for 15 minutes, then cooled down at 37 °C before adding 1 μl of 5 μg/μl RNase A and finally incubated for 30 minutes at 37 °C. Samples were left on ice for 5 minutes and 175 μl of MPC protein precipitation reagent was added to each sample and mixed. The debris were pelleted by centrifugation at 4 °C for 10 minutes at ≥10,000 × g. The supernatant was transferred to a clean microcentrifuge tube. 500 μl of isopropanol was added and tubes were inverted several times before centrifugation at 4 °C for 10 minutes. Isopropanol was removed and pellets were rinsed 2 times with 70% ethanol and left dry before solubilization with 35 μL of EB buffer (Qiagen). The six samples were pooled together and purified.

Strain	Species	Status	GenBank accession number	
			ITS	LSU
CBS 321.79	*Periconia algeriana*	RS	MH861212	MH872979
CBS 381.55	*Periconia atropurpurea var. microspora*	RS	MH857524	MH869061
CBS 685.70	*Periconia byssoides*	RS	MH859902	MH871694
MFLU 19-2784	*Periconia celtidis*	CH	MW063162	NG_079543
KUMCC 20-0266	*Periconia chimonanthi*	CH	NR_176752	NG_081512
CBS 414.50	*Periconia circinata*	RS	MH856694	MH868210
CBS 144434	*Periconia cyperacearum*	CH	MH327815	MH327851
CNCM I-4278	*Periconia digitata*	this work	OP329216	OP329219
CBS 510.77	*Periconia digitata*	RS	LC014584	AB807561
MFLUCC 17-0087	*Periconia elaeidis*	CH	MG742713	MH108552
CGMCC 3.23929	*Periconia festucae*	CH	NR_185800	OP955998
CBS 322.79	*Periconia genistae*	RS	MH861213	MH872980
MFLUCC 17-0341	*Periconia heveae*	RS	OL780490	OL782069
HHUF 29105	*Periconia homothallica*	CH	NR_153446	NG_059397
CBS 583.66	*Periconia igniaria*	RS	MH858888	MH870553
CGMCC 3.23931	*Periconia imperatae*	CH	NR_185801	OP956009
CBS 292.36	*Periconia lateralis*	RS	MH855804	MH867311
DSE2036	*Periconia macrospinosa*	G	OM337552	OM337552
CBS 146062	*Periconia neobrittanica*	CH	MN562149	MN567656
CGMCC 3.23928	*Periconia penniseti*	CH	NR_185799	OP955996
CBS 209.64	*Periconia prolifica*	CH	MH858422	MH870050
CBS 147067	*Periconia pseudobyssoides*	RS	ON811518	ON811576
H4151	*Periconia pseudobyssoides*	RS	LC014587	AB807568
HHUF 29370	*Periconia pseudodigitata*	CH	NR_153490	NG_059396
MFLU 19-1235	*Periconia salina*	CH	MN047086	MN017846
CGMCC 3.23932	*Periconia spodiopogonis*	CH	NR_185798	OP955988
KUMCC 20-0262	*Periconia thysanolaenae*	CH	NR_176751	NG_081511
CBS 122368	*Trematosphaeria pertusa*	CH	OM337545	OM337545
CBS 221.30	*Penicillium roqueforti*	N	NR_103621	NG_069624

Table 1. List of strains with their corresponding sequence accession number used to build the phylogenetic tree of *Periconia* species. RS: reference strain (holotype not available), CH: culture from holotype, G: available genome, N: neotype.

DNA was analyzed for quality and quantity controls using nanodrop, Qubit and fragment analyser. The final sample used for library preparation contained 19.2 µg of gDNA with an average fragment size of 30,145 bp.

RNA extraction for transcriptome sequencing. Mycelia from the 12 different culture/stress conditions (Table 2) were retrieved separately and directly ground in liquid nitrogen with sterilized pestles and mortars. The resulting powder for each condition was transferred into 2 to 3 microtubes, depending on the sample quantity. 800 µl of extraction buffer (CTAB 2.5%, PVPP 2%, Tris-HCL 100 mM, EDTA 25 mM, NaCl 2 M, β-mercaptoethanol 2%) was added to each tube. After 30 minutes of incubation at 65 °C, 800 µl of Chloroform/Isoamyl alcohol (CI; v/v; 24/1) was added, and after homogenization, the tubes were centrifuged (16000 g - 8 min - 4 °C). The supernatant was retrieved and the same volume of water-saturated Phenol (pH 4.5-5)/Chloroform/Isoamyl alcohol (PCI; v/v; 25/24/1) (ca. 700 µl) was added and centrifuged. A second step with CI was carried out and the supernatant was retrieved and mixed with 500 µl NaCl 5 M and 500 µl of isopropanol and stored overnight at −20 °C. After centrifugation (16000 g - 20 min - 4 °C), two cleaning steps were carried out with ethanol 70%. The pellet was dried and resuspended in 30 µl of buffer EB (Qiagen). Genomic DNA was removed from samples using the kit TURBO DNA-free (Ambion) following the supplier's instructions. Samples purity and quality were assessed with a nanodrop and a bioanalyzer. The samples with the best qualitative parameters ($1.98 < OD_{260/280} < 2.07$; $1.72 < OD_{260/230} < 2.19$; $2.5 < RIN < 5$; $3.6 µg < RNA\ Quantity < 15.8 µg$) in each condition were kept for library preparation Table 3.

DNA sequencing. Library preparation and sequencing were performed at GeT-PlaGe core facility, INRAE Toulouse according to the manufacturer's instructions "Procedure & Checklist Preparing HiFi SMRTbell Libraries using SMRTbell Express Template Prep Kit 2.0". At each step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). DNA purity was tested using a nanodrop (Thermofisher) and size distribution and degradation assessed using the Femto pulse Genomic DNA 165 kb Kit (Agilent). Purification steps were performed using AMPure PB beads (PacBio)0.15 µg of DNA was purified then sheared at 20 kb using the Megaruptor1 system (Diagenode). Using SMRTbell Express Template prep kit 2.0, a Single strand overhangs
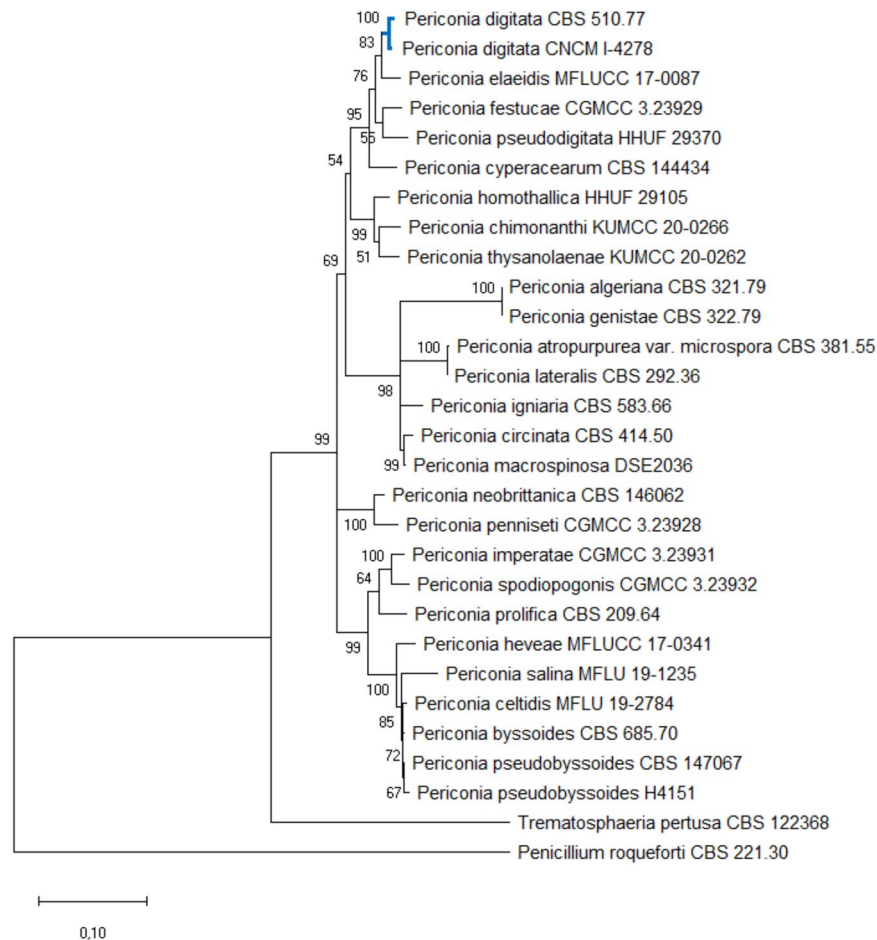
Fig. 2 Phylogenetic inference based on a combined ITS and 28 S dataset. The tree is rooted *to Trematosphaeria pertusa* and *Penicillium roqueforti*. The blue branch highlights the cluster of *P. digitata*, where our strain *P. digitata* CNCM I-4278 is positioned. The numbers indicate the percentage of trees in which the associated taxa clustered together (1000 bootstrap). Bar = expected changes per site (0.10).

removal, a DNA and END damage repair step were performed on 5 µg of sample. Then blunt hairpin adapters were ligated to the library. The library was treated with an exonuclease cocktail to digest unligated DNA fragments. A size selection step using a 9 kb cutoff was performed on the BluePippin Size Selection system (Sage Science) with "0,75% DF Marker S1 High Pass 15–20 kb" protocol. Using Binding kit 2.0 kit and sequencing kit 2.0, the primer V2 annealed and polymerase 2.0 bounded library was sequenced by diffusion loading onto 1 SMRTcell on Sequel2 instrument at 95 pM with a 2 hours pre-extension and a 30 hours movie.

RNA sequencing. RNAseq was performed at the GeT-PlaGe core facility, INRAE Toulouse. RNA-seq libraries have been prepared according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. Briefly, mRNA were selected using poly-T beads. Then, RNAs were fragmented to generate double stranded cDNA and adaptors were ligated to be sequenced. 11 cycles of PCR were applied to amplify libraries. Library quality was assessed using a Fragment Analyser and libraries were quantified by QPCR using the Kapa Library Quantification Kit. RNA-seq experiments have been performed on an Illumina NovaSeq. 6000 using a paired-end read length of 2 × 150 pb with the Illumina NovaSeq. 6000 sequencing kits.

Genome size estimation. We used Jellyfish version 2.2.6[33] count and histo commands with a k-mer size of 21 and a maximum multiplicity of 1,000,000 on the PacBio Hi-Fi genome reads to count k-mers and their multiplicity. The output of Jellyfish histo was then used as an input for GenomeScope2[34] with a ploidy level of 1 and 2 for genome size and heterozygosity level estimation.

Mitochondrion assembly. The assembly of the mitochondrial genome sequence was done with ALADIN (https://github.com/GDKO/aladin) using the mitochondrion mode from PacBio HiFi reads. The complete mitochondrial genome of the closest available relative *Phaeosphaeria nodorum* SN15 (NC_009746.1) downloaded from GeneBank was used as a reference seed sequence.

Genome assembly, QC, Contamination. PacBio Hifi reads with a highly accurate median accuracy of minimum 99.9% (Q30) were used as input for the HiCanu assembler[35]. Post-assembly quality control and

Short reference	Type of medium[1]	Medium	Composition	Stress	When the stress is applied
1	L	V8 - Vegetable Juice	Commercial medium modified: addition of 15 g/L CaCO3, centrifugation and dilution 1:5 of the supernatant with Milli-Q water	—	—
2	L	Oat milk	Commercial medium	—	—
3	L	RMI glucose vitamins B	RMI[2], 10 g/L glucose, vitamins B mix (1 mg/L nicotinic acid, 1 mg/L pyridoxine, 1 mg/L Calcium panthotenate, 1 mg/L Thiamine HCl, 0.1 mg/L Biotin)	—	—
4	L	RMI malt extract	RMI[2], 10 g/L Malt extract (Difco)	—	—
5	L	RMI glucose wheat peptone	RMI[2], 10 g/L glucose, 2 g/L wheat peptone HE1 (Organotechnie)	—	—
6	L			1 mM $CuSO_4$	7 days
7	L			200 mM NaCl	7 days
8	L			1 mM H_2O_2	7 days
9	L			light 16 h photoperiod	all over the experiment
10	L			cold stress at 4 °C	15 h before the RNA extraction
11	L			heat stress at 37 °C	15 h before the RNA extraction
12	S	PDA	Potato extract 4 g/L, Dextrose 20 g/L, Agar 15 g/L (Difco)	light 16 h photoperiod	all over the experiment

Table 2. Culture conditions used to obtain a reference transcriptome for *P. digitata*. [1](L) liquid culture in 6-well plates; (S) solid culture in Petri dishes [2]RMI (600 mg/L KH_2PO_4, 700 mg/L KNO_3, 250 mg/L $MgSO_4 \cdot 7H_2O$, 125 mg/L $K_2HPO_4 \cdot 3H_2O$, 430 mg/L $Ca(NO_3)_2 \cdot 4H_2O$, 1 mg/L H_3BO_3, 1.5 mg/L $MnSO_4H_2O$, 4 mg/L $ZnSO_4 \cdot 7H_2O$, 0.1 mg/L $Na_2MoO_4 \cdot 2H_2O$, 0.02 mg/L KI, 0.02 mg/L $CuSO_4 \cdot 5H_2O$, 0.02 mg/L $CoCl_2 \cdot 6H_2O$, 0.01 mg/L $FeSO_4 \cdot 7H_2O$, 0.015 Na_2 EDTA $\cdot 2H_2O$

taxonomic partitioning were assessed with BlobTools[36,37]. Previously quality-filtered PacBio Hifi reads were mapped back to the assembly with mimimap2[38] to estimate contigs coverage. Each contig was assigned a taxonomy affiliation based on BLAST[39,40] results against the NCBI nt database.

Telomere detection. Terminal telomeric repeats were searched using tidk software v0.1.5 (https://github.com/tolkit/telomeric-identifier). The tidk software explore module was used to search the genome for repeats from length 5 to 10. Positions of repeats are only reported if they occur sequentially in a higher number than the threshold of 5. The most represented repeat unit was AACCCT with a maximum frequency of 209. Then, this putative telomeric repeat was scanned on the contigs with the tidk search module using a window size of 150 to calculate repeat counts. This information is then used as an input for the tidk plot module to visualize positions of the putative telomeric repeats along each contig sequence.

Gene prediction. Gene models prediction was done with the fully automated pipeline EuGene-EP version 1.6.5[41]. EuGene has been configured to integrate similarities with known proteins of "ascomycota" section of UniProtKB/Swiss-Prot library (UniProt Consortium 2018[42]), with the prior exclusion of proteins that were similar to those present in RepBase[43].

The dataset of *Periconia digitata* transcribed sequences generated in this study were aligned on the genome and used by EuGene as transcription evidence. For this, we first assembled *de novo* using Trinity[44] the transcriptomes of *P. digitata* obtained from the twelve above-described conditions and for a given trinity locus we only retained the transcript returning the longest ORF. Finally, only *de novo* assembled transcripts that aligned on the genome on at least 30% of their length with at least 97% identity were retained.

The EuGene default configuration was edited to set the "preserve" parameter to 1 for all datasets, the "gmap_intron_filter" parameter to 1 and the minimum intron length to 35 bp. Finally, the Fungi specific Weight Array Method matrices were used to score the splice sites (available at this URL: http://eugene.toulouse.inra.fr/Downloads/WAM_fungi_20180126.tar.gz).

Genome and protein set completeness assessment. We used BUSCO[45] version 5.2.2 in protein and genome modes with the eukaryota odb10 dataset of 255 BUSCO groups and the fungi odb10 dataset of 758 BUSCO groups to assess the completeness of the predicted protein set as well as the genome assembly. We compared BUSCO scores to those obtained for the *Periconia macrospinosa* genome and predicted proteins[20].

Functional annotation. All predicted proteins were scanned for the presence of conserved protein domains and motifs using InterProScan v.5.51–85.0[46] with the options -iprlookup, -goterms and -pa to assign Gene Ontology (GO) terms, MetaCyc and Reactome biochemical pathways based on detection of Interpro domains.

Gene prediction and functional annotation of mitochondrion. The annotation was performed using MITOS2[47] including the ncRNA (t- and r-RNA) and the protein coding sequences with the codon usage number 4. The gene predictions were refined using the assembled Trinity transcripts aligned on the mitogenome by direct translation in ORFfinder as well as annotation with smartBlast (https://www.ncbi.nlm.nih.gov/orffinder/), and intron reconstruction with BioEdit v 7.0.5.3[48]. We also compared our results to the following Pleosporales available mitogenomes: NC_058694 (*Edenia gomezpompae*, 37 kb, 14 ORF), NC_040008 (*Coniothyrium glycines*, 98 kb, 35 ORF), NC_026869 (*Shiraia bambusicola*, 39 kb, 17 ORF) and NC_035636

(*Pithomyces chartarum*, 69 kb, 37 ORF) in addition to this of *Phaeosphaeria nodorum* (see mitochondrion assembly). The concatenated file obtained from MITOS2 and protein-coding sequences coordinates was used for genbank submission (OP787475[49]) and mitogenome drawing by OGDRAW version 1.3.1[50].

Protein extraction and sample preparation for proteomics analysis. *P. digitata* was cultivated, in 500 mL plastic Roux bottles, in 50 mL of five different sterile liquid media (RMI free of asparagine and vitamins, RMI supplemented with B vitamins, RMI plus wheat peptone, RMI plus *Citrus* pectin, RMI plus Guar gum, RMI plus malt) over 7 days in the dark at 24 °C. The media were chosen for their ability to change the strain phenotype (see RNA extraction above); two biological replicates were prepared. The mycelium was recovered by filtration on GF/C Whatman glass filter and rinsed with water. It was ground using liquid nitrogen then sequentially extracted using successive buffers starting with a Tris buffer 20 mM pH 8, 10 mM DTT (dithiothreitol), then the same buffer supplemented with 200 mM NaCl buffer, the same buffer supplemented with 8 M urea and finally the same buffer with 6 M guanidine hydrochloride (1 mL per 300 mg fresh material weight). Centrifugations (13,200 rpm, 5 min, 4 °C) were achieved after each extraction and the 4 successive supernatants were recovered. The two first buffers (alone and saline) were immediately adjusted to 8 M urea. All these fractions were incubated at 37 °C for 15 min then alkylated by iodoacetamide (41.6 mM) during 15 minutes at room temperature. After two buffer exchanges with trypsin buffer on Vivaspin 15 R (5 kDa) columns (Sartorius), the samples were digested by trypsin with 1/80 (w/w) trypsin/total protein ratio (Sequencing Grade Modified Trypsin, Promega) according to the manufacturer recommendations. The residual pellets after successive extractions were suspended in 1 mL of Tris 20 mM pH 8, 10 mM DTT, 8 M urea then incubated at 37 °C for 15 min and alkylated with 41.6 mM of iodoacetamide for 15 min at room temperature with gentle mixing. After centrifugation (see above), the pellets were rinsed/centrifuged 2 fold in the trypsin buffer. The final pellet suspension was directly digested by trypsin (1 µg/pellet) overnight under gentle agitation on a rotator mixer. After centrifugation, the supernatants were recovered. The samples were then cleaned-up on a C18 SPE cartridge (100 mg, NUCLEODUR® 100-30 C18 end-capped, Macherey Nagel) equilibrated in H2O 0.1% formic acid (FA). After washing with H2O 0.1% FA, peptides were eluted with acetonitrile (CH3CN)-0.1% FA/water-0.1% FA 40/60 and 70/30 (v/v). Both fractions were pooled.

Proteomics analysis. The samples were analyzed by nanoUHPLC-HRMS (nanoElute – timsTOF Pro, Bruker Daltonics). 5 µL of sample were injected on an Aurora column (75 µm id × 250 mm, C18, 1.6 µm, ionOpticks) with a flow rate of 200 nL/min at 50 °C. The mobile phase was a gradient of CH3CN-0.1% FA (B) in 0.1% FA-H$_2$O (A) as follows: 5% B for 1 min, 5% to 13% of B for 18 min, 13% to 19% of B for 7 min, 19% to 22% of B for 4 min, 22% to 95% of B for 3 min.

The timsTOF Prowas equipped with the CaptiveSpray nano-electrospray ion source. MS and MSMS data were acquired in a positive mode, in a PASEF (Parallel Accumulation -Serial Fragmentation) data dependent acquisition (DDA), TIMS ON mode from 100 to 1700 m/z mass range (TimsControl version 2.0.53.0). Ion mobility resolution (1/K0) was set to 0.70–1.10 V·s/cm^2 over a ramp time of 180 ms. To exclude low m/z, singly charged ions from PASEF precursor selection, a polygon filter was applied in the m/z and ion mobility space.

Analyses were processed by Peaks Studio X Pro (version 10.6, bioinformatics Solutions Inc.) MSMS raw data were processed using Peaks solution applying 3 levels of identification. MSMS spectra were matched against the *P. digitata* Y3 predicted proteome (merged core and mitochondrial genomes), including the MaxQuantcontaminant database (contaminants.fasta, MaxQuant 2.1). Parameters were set as follows: protein FDR <1%, decoy fusion method, cysteine carbamidomethylation as a fixed modification, 2 miscleavage, 3 to 5 post-translational modifications (PTM) per peptide. The MSMS spectra that did not match with these parameters were processed again, looking for other possible PTM and amino acid mutations to enrich the list of identified proteins. The obtained results were merged in a list of significant proteins[51].

Data Records

ITS and 28 S sequences that have been used to identify *P. digitata* are available in NCBI (accession numbers **OP329216**[29] and **OP329219**[30], respectively).

All Illumina and PacBio HiFi raw data used to assemble the genome and transcriptomes are publicly available at the EMBL-EBI's European Nucleotide Archive under the project number **PRJEB55037**[52]. The detailed list of raw data accession numbers are presented in the table below.

	Run Accession	Sample	Experiment
RNA of P. digitata condition 12	ERR10025227	ERS12553875	ERX9565917
RNA of P. digitata condition 11	ERR10025174	ERS12553874	ERX9565864
RNA of P. digitata condition 10	ERR10025118	ERS12553873	ERX9565808
RNA of P. digitata condition 9	ERR10025114	ERS12553872	ERX9565804
RNA of P. digitata condition 8	ERR10025101	ERS12553871	ERX9565791
RNA of P. digitata condition 7	ERR10025086	ERS12553870	ERX9565776
RNA of P. digitata condition 6	ERR10025067	ERS12553869	ERX9565757
RNA of P. digitata condition 5	ERR10025065	ERS12553868	ERX9565755
RNA of P. digitata condition 4	ERR10025044	ERS12553867	ERX9565734
RNA of P. digitata condition 3	ERR10024935	ERS12553866	ERX9565625
RNA of P. digitata condition 2	ERR10024912	ERS12553865	ERX9565602
RNA of P. digitata condition 1	ERR10020517	ERS12553864	ERX9561207
DNA of P. digitata	ERR10009562	ERS12521256	ERX9550522

doi accession	Description
https://doi.org/10.57745/FB9OVN[63]	genome assembly fasta file
https://doi.org/10.57745/BAPYY5[55]	gene models prediction and eugene statistics files
https://doi.org/10.57745/5ODV0C[64]	genes fasta file
https://doi.org/10.57745/71DS6F[65]	CDS fasta file
https://doi.org/10.57745/TV4EB7[66]	predicted proteins fasta file
https://doi.org/10.57745/GU67TE[56]	functional annotation provided by interproscan
https://doi.org/10.57745/FHNYJU[51]	proteins identified by mass spectrometry. It includes the list of proteins identified with more than 1 peptide, only 1 peptide, the Post Translational Modifications, proteins displaying a signal peptide.

Table 3. Datasets present in the Recherche Data Gouv repository.

The mass spectrometry proteomic raw data are available in the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier **PXD038112**[53] and **PXD038175**[54].

All the analyzed data are publicly available at https://entrepot.recherche.data.gouv.fr/dataverse/pdig. The detailed list of analyzed files are presented in the Table 3.

The complete mitochondrial genome assembly of *P. digitata* can be retrieved at the NCBI through the accession number **OP787475**[49].

Technical Validation

Contamination assessment. Blobtools analysis showed that all the contigs formed a dense blob at a homogenous coverage (550X) and GC content (49%) indicating no evidence for contamination (i.e., no contig deviates from this distribution) (Fig. 3). Moreover, the taxonomic affiliation analysis based on homology, using BLAST against the NCBI's nt library, showed that all contigs are of Ascomycota origin, which is consistent with the absence of evident contamination (Fig. 4).

Genome size estimation and *de novo* assembly. Based on k-mer multiplicity distribution using GenomeScope2, both 1n and 2n models converged in showing one single peak at a very high coverage of ca. 560X. The GenomeScope model fit values were slightly higher for the haploid (1n) model (92.59% - 94.61%) than for the diploid (2n) model (92.59% - 94.31%). Collectively, these results strongly suggest a haploid genome sequenced at a very high coverage. The haploid 1n model returned an estimated genome size of ca. 36 Mb with an error rate of ca. 0.43% (Fig. 5).

The HiCanu assembler yielded a genome assembly that was ca. 39 Mb long, consistent with haploid genome size estimated with k-mers. The genome was assembled in 15 contigs with a N50 value of 3 Mb and a L50 of 5 (i.e. half of the genome is present in the 5 biggest contigs) (Table 4). Among the 15 contigs, HiCanu generated two outliers of 27–28 kb (Pdig14 and Pdig15, Table 5). Both contained only rDNA repeats that can be partially aligned at the end of Pdig08, which exclusively exhibits rDNA repeats.

The repeat sequence (AACCCT)n we have identified at the terminal regions of the contigs corresponds to the reverse complement of the (TTAGGG)n telomeric repeat widely conserved in vertebrates, many other animals, plants as well as several different eukaryotes, including fungal species.

It is worthy to note that the telomeric repeats were detected at both ends of 11 out of 13 contigs. Telomeric repeats were detected at only one end in the contigs Pdig06 and Pdig08 (Table 5). Overall, we obtained a highly-contiguous genome assembly for *Periconia digitata* that is structured in 13 chromosomes in its haploid mycelium with a total length of 38,967,494 bp.

By comparison with the two other genomic resources available for the *Periconia* genus (P. macrospinosa and P. sp. R9002), the assembly of *P. digitata* greatly improves the resolution of *Periconia*'s genome structure. *P. macrospinosa* (assembly GCA_003073855.1 Perma1[20]) and *P. sp.* R9002 (assembly GCA_023627715.1 ASM2362771v1) assemblies (54.99 Mb and 44.55 Mb, respectively) are made of 2,976/1,566 and 1,000/896 contigs/scaffolds, with a N50 of 140,954 bp and 237,696 bp, respectively (Table 4).

Genome completeness assessment. BUSCO (v5.2.2) analysis at the genome level indicated that 99.6% and 99.1% of nearly-universal single copy genes from the eukaryota and fungi datasets, respectively, were retrieved in full-length. Only 0.4 and 0.7% of eukaryotic and fungal BUSCO genes were identified as duplicated, consistent with a genome assembled in a haploid state with no evidence for substantial gene duplications (Table 6). Although the genome of *P. macrospinosa* was considerably more fragmented, the gene content completeness was comparable to that of *P. digitata* according to BUSCO metrics.

Genome annotation and assessment of predicted proteins. Using the Eugene-EP pipeline, 15,815 genes were predicted including 15,520 protein-coding genes and 295 non-coding genes. Genes cover 26.3 Mb (~68%) of the genome assembly. The coding portion accounts for 51% of the assembly and spliceosomal introns were detected in 71% of protein coding-genes, with an average of 2.61 exons per gene[55].

A BUSCO analysis in protein mode, revealed that 96.9% and 97.3% of complete eukaryotic and fungal BUSCO proteins, respectively, were found in the predicted proteome dataset (Table 7). The metrics obtained on *P. macrospinosa* annotation were at least as good, suggesting again a complete gene set despite a fragmented genome assembly.
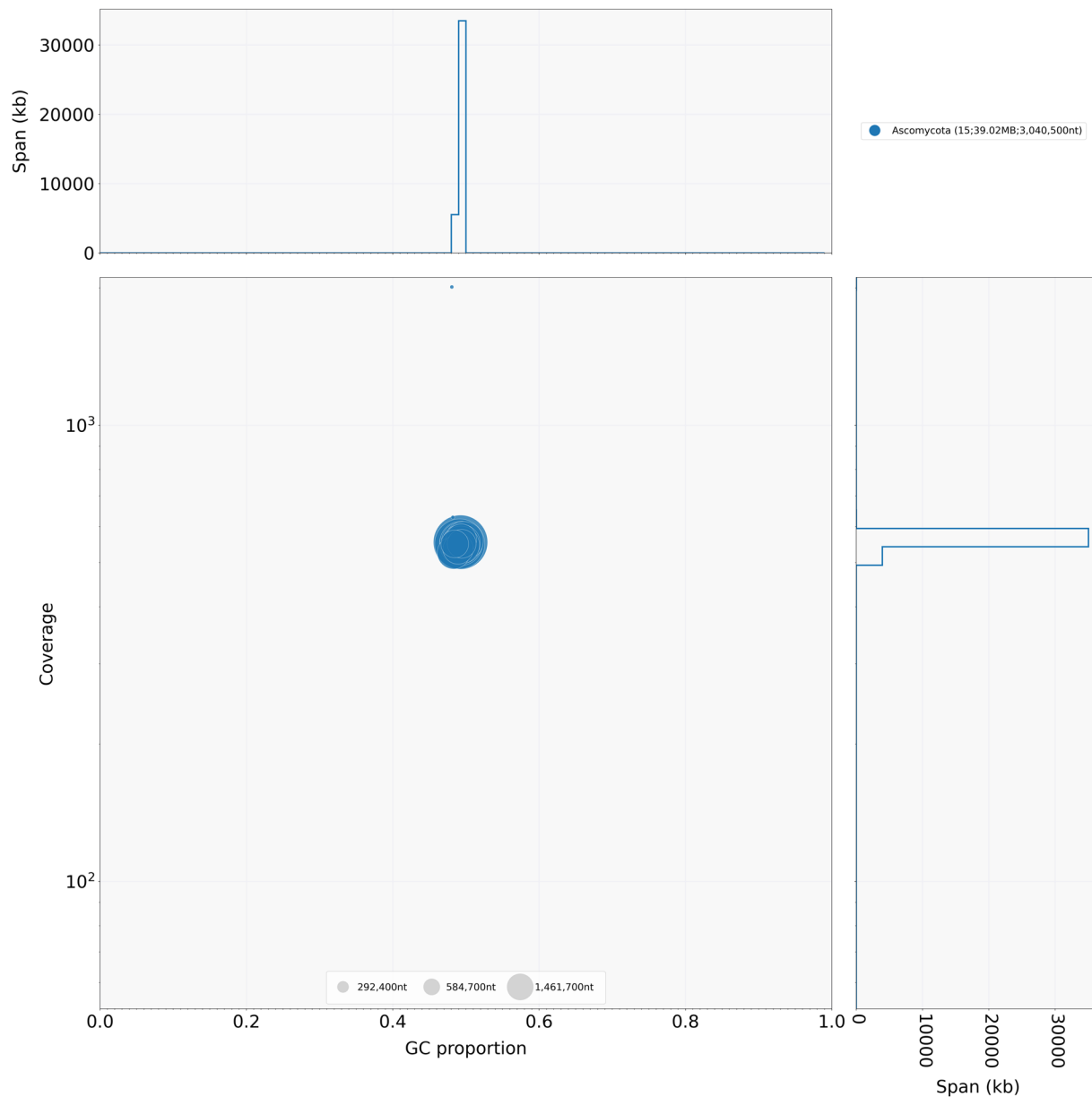
Fig. 3 BlobPlot of the genome assembly. Each circle is a contig proportionally scaled by contig length and coloured by taxonomic annotation based on BLAST similarity search results. Contigs are positioned based on the GC content (X-axis) and the coverage of PacBio reads (Y-axis).
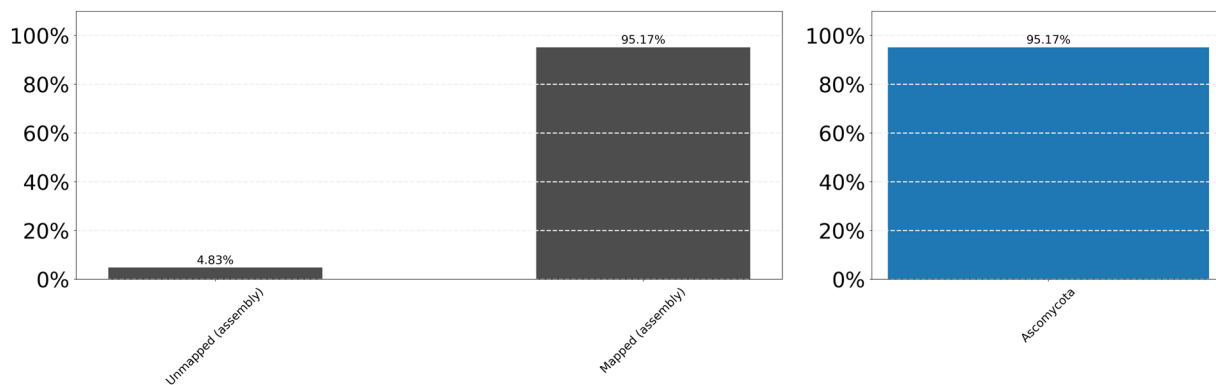

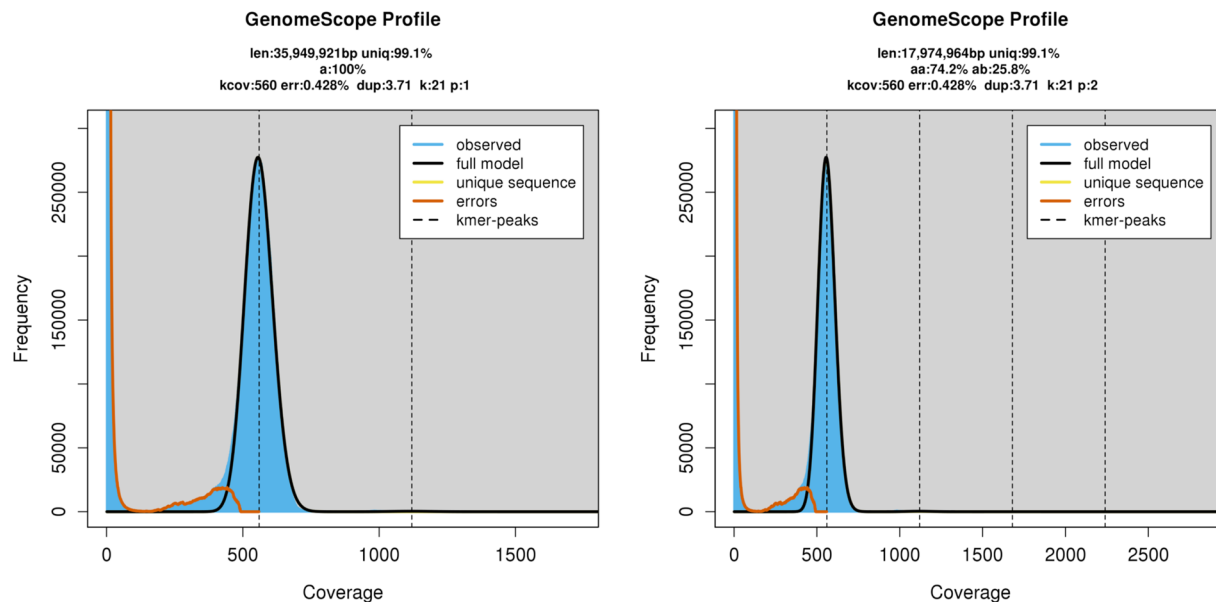**Fig. 4** ReadCovPlot. Mapped reads are shown by the taxonomic group at the rank of 'phylum'.

Fig. 5 GenomeScope2 k-mer profiles for the haploid (1n) (left) and diploid (2n) (right) models. Coverage (kcov), error rate (err.), haploid genome size estimation (len.), k-mer size (k), ploidy level (p).

Species	Number of Contigs/scaffolds	Minimal length bp)	Maximal length bp)	N50 (bp)	L50	N90 (bp)	L90	Mean (bp)	Median (bp)	Total assembly (pb)
Pdig	15/0	27,658	5,846,744	3,040,500	5	1,824,482	11	2,601,553	2,871,062	39,023,295
Pmac	2,976/1,566	1,000	1,324,167	140,954	101	18,263	478	35,116	6,289	54,992,973
R9002	1,000/896	501	1401378	331741	39	53319	140	49720	2953	44,549,138

Table 4. Metrics of *Periconia digitata*'s genome assembly (Pdig) and comparison to *P. macrospinosa* (Pmac) (assembly GCA_003073855.1 Perma1, Knapp *et al*. 2018) and *P.* R9002 (assembly GCA_023627715.1 ASM2362771v1).

ContigsID	length (bp)	GC%	Telomere detection (Tidk)
Pdig01	5,846,744	49.26	both ends
Pdig02	4,988,348	49.15	both ends
Pdig03	3,687,698	49.32	both ends
Pdig04	3,596,248	49.34	both ends
Pdig05	3,040,500	49.32	both ends
Pdig06	3,027,279	49.72	start
Pdig07	2,891,359	49.35	both ends
Pdig08	2,871,062	49.26	start
Pdig09	2,118,696	48.31	both ends
Pdig10	1,885,720	49.35	both ends
Pdig11	1,824,482	48.93	both ends
Pdig12	1,658,798	49.42	both ends
Pdig13	1,530,560	48.47	both ends
Pdig14	28,143	48.07	none
Pdig15	27,658	48.21	none

Table 5. Metrics of *Periconia digitata*'s contigs and telomere detection.

Functional annotation. Conserved InterPro domains and motifs were identified on 61.1% of the 15,520 predicted protein sequences. The 9,479 annotated proteins returned 7,713 different InterPro domains[56].

The top 15 interpro homologous superfamilies and domains contained some large gene families found in most organisms together with domains restricted to fungi. Among the detected domains, we noticed the presence of the HET (Heterokaryon incompatibility) domain (Fig. 6) which is specific to Ascomycota[57] and domains

	P. digitata		P. macrospinosa	
	eukaryota	fungi	eukaryota	fungi
Number of species	70	549	70	549
Number of BUSCO	255	758	255	758
% of complete BUSCO	99.6	99.1	99.2	99.1
% of complete and single copy BUSCO	99.2	98.4	98.8	98.4
% of complete and duplicated BUSCO	0.4	0.7	0.4	0.7
% of fragmented BUSCO	0.4	0.1	0.8	0.3
% of missing BUSCO	0	0.8	0.0	0.6

Table 6. BUSCO scores for the genome of *P. digitata* and *P. macrospinosa* using the eukaryota odb10 and the fungi odb10 datasets.

	P. digitata		P. macrospinosa	
	eukaryota	fungi	eukaryota	fungi
Number of proteomes	70	549	70	549
Number of BUSCO	255	758	255	758
% of complete BUSCO	96.9	97.3	97.6	98.5
% of complete and single copy BUSCO	96.9	96.8	97.6	97.8
% of complete and duplicated BUSCO	0.0	0.5	0.0	0.7
% of fragmented BUSCO	1.2	0.8	1.2	0.4
% of missing BUSCO	1.9	1.9	1.2	1.1

Table 7. BUSCO scores for the proteins of *P. digitata* and *P. macrospinosa* using the eukaryota odb10 and the fungi odb10 datasets.
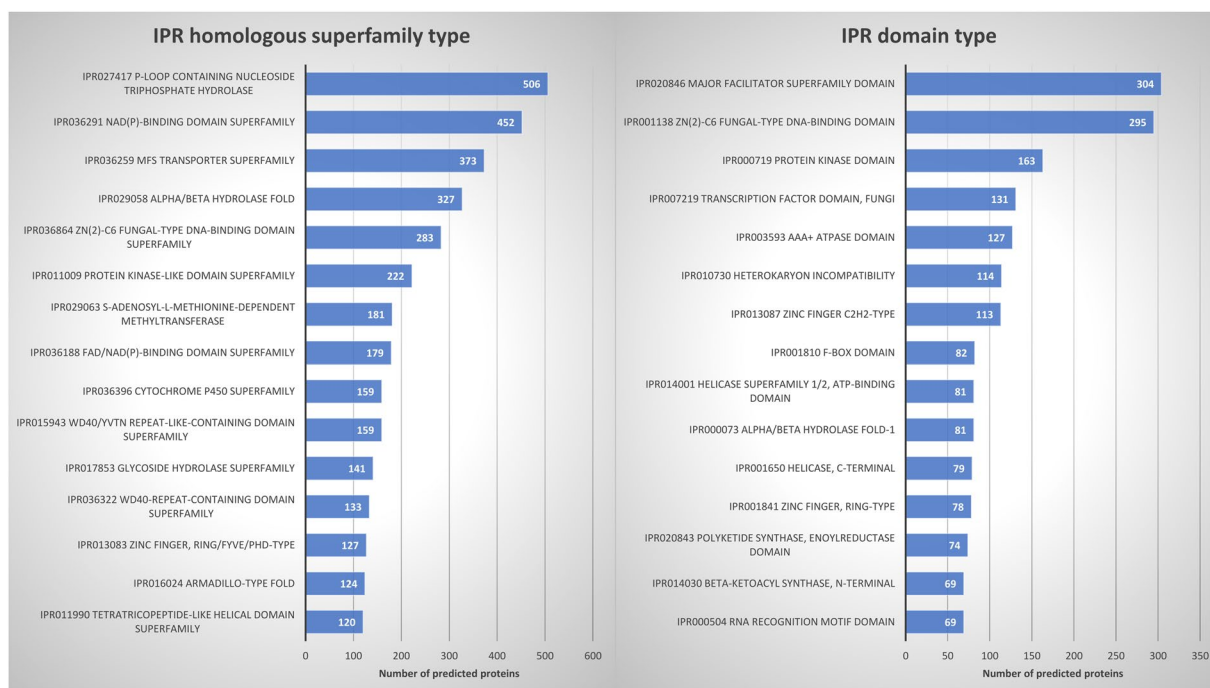
Fig. 6 Interpro functional annotation of the *P. digitata* predicted proteome. The Top 15 homologous superfamilies (**a**) and domains (**b**) are indicated.

associated with key enzymes for many secondary metabolites biosynthesis in fungi such as the beta-ketoacyl synthase or the polyketide synthase, enoylreductase domains[58].

Using SinalP (v6.0)[59], we assessed the number of predicted proteins that contained a putative signal peptide for secretion. Among the 15,520 proteins, 1,597 (10.3%) were predicted to display a signal peptide thus potentially to be addressed to extracellular space or the membrane. This value is in the upper section of the
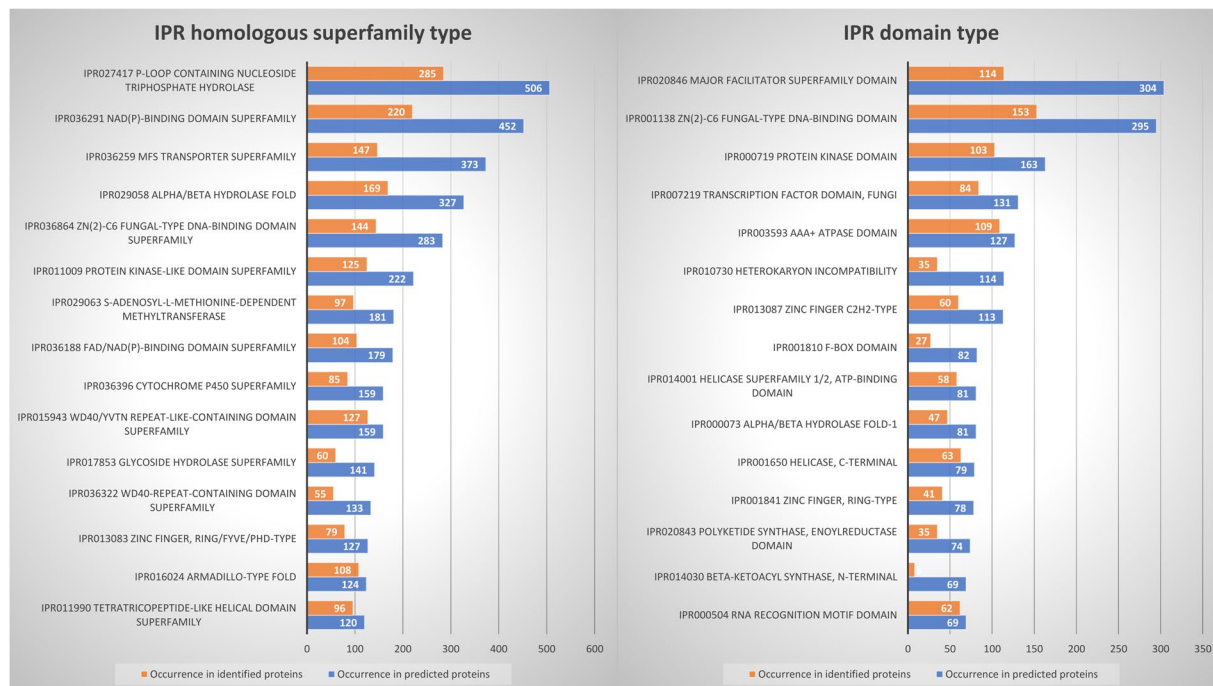
Fig. 7 Top 15 homologous superfamilies (**a**) and domains (**b**) obtained from Interpro predictions (blue) together with the corresponding homologous superfamilies and domains in proteins identified by proteomics (orange).

	Unique peptides	Unique proteins	Unique sites	Targeted residue (%)
Phosphorylation (STY)	374	344	348	Ser (74), Thr (22), Tyr (4)
Methylation(KR)	355	205	233	Lys (57), Arg (42)
Acetylation (K)	218	132	131	Lys
Dimethylation(KR)	46	30	31	Arg (71), Lys (29)
Biotinylation	43	40	41	
Iminobiotinylation	9	9	9	
Ubiquitination (diGly, LRGG)	58	28	49	Lys (98)
Thiophosphorylation	23	20	21	Ser (67, Thr (14), Tyr (19)
Myristoylation	12	12	12	Gly (75)
Sulfation	11	11	11	Ser (55), Thr (27), Tyr (18)
Aminotyrosine with sulfation	7	6	6	Tyr
Octanoyl	6	6	6	Ser (50), Thr (50)
Phosphopantetheine	1	1	1	Lys
Lipoyl	1	1	1	Ser

Table 8. Detected post translational modifications (PTM) with a A score > 20 (equivalent to pvalue < 0.01).

range observed in fungal proteomes, e.g. 8.5% in *Trichoderma asperellum*[60], 1.1% - 12% in 132 Zygomycota proteomes[61], 3% - 10% obtained in 49 fungal proteomes[62].

Proteomics support of predicted proteins. Of the 15,551 predicted proteins, 6,598 (42.4%) returned matches with at least 2 unique peptides (-10lgP > 50) at a maximum FDR of 1%. The identification reached 46.9% after inclusion of proteins identified with a single peptide (698, -10lgP > 50). This value is in line with the numbers observed among the top 15 interpro homologous superfamilies and domains identified (Fig. 7). Although 6,041 proteins (38.9% of the predicted ones) returned no InterPro annotation, 883 of them (14.6%) were identified by proteomics, suggesting they are actual proteins with no known conserved domain to date.

In addition, many post translational modifications (PTM, A score >20) were detected[51] (Table 8). Phosphorylation mainly concerned proteins predicted to be involved in transport (13), ubiquitin trafficking (13) cytoskeleton (12) and transcription (8). Acetylation and methylation mainly concerned proteins predicted to interact with nucleic acids (from histones to ribosomal proteins) or to have diverse enzymatic activities.
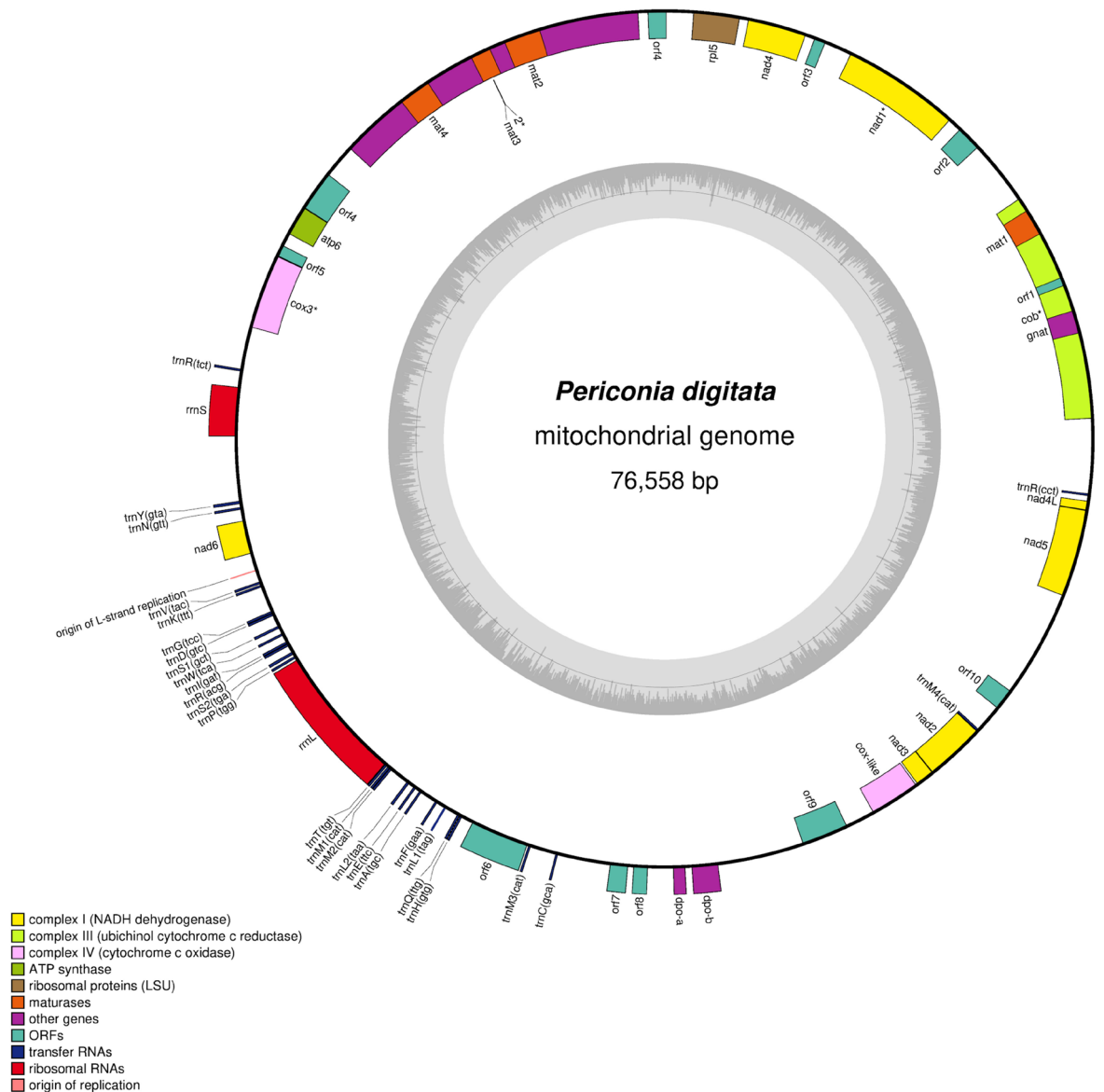
Fig. 8 Mitochondrial genome of *Periconia digitata*.

Accession	Coverage (%)	Unique peptides	Annotation
PdigMGS11	34	72	cox1&2
PdigMGS09	47	24	rpl5
PdigMGS30	21	22	nad5
PdigMGS25	36	21	hypothetical protein
PdigMGS15	35	15	hypothetical protein
PdigMGS28	15	10	nad2
PdigMGS08	15	9	nad4
PdigMGS27	31	8	nad3
PdigMGS29	31	8	hypothetical protein
PdigMGS18	6	6	cox3
PdigMGS01	12	6	cob
PdigMGS16	15	5	atp6
PdigMGS06	4	2	nad1
PdigMGS31	9	1	nad4L

Table 9. List of the 14 mitochondrial proteins identified by proteomics.

Mitochondrial genome assembly and annotation. The assembled mitochondrial genome[49] was 76,558 bp long with 27.56% of GC (Fig. 8). As for other Pleosporales mitogenomes, we retrieved a complete set of tRNA, among which some were in multiple copies (Met, Ser, Leu) as well as rnl and rns. Out of the 31 predicted proteins encoded in the mitochondrion, 14 were identified by proteomics including typical enzymes, one ribosomal protein and 3 hypothetical proteins (Table 9).

Code availability

No specific codes or scripts were used in this study. All software used is in the public domain, with parameters clearly described in the Methods section.

References

1. Hyde, K. *et al.* Mycosphere notes 169–224. *Mycosphere* **9**, 271–430 (2018).
2. Tanaka, K. *et al.* Revision of the *Massarineae* (*Pleosporales*, *Dothideomycetes*). *Studies in Mycology* **82**, 75–136 (2015).
3. Tischer, M. *et al.* New fossils of ascomycetous anamorphic fungi from Baltic amber. *Fungal Biology* 804–810 (2019).
4. Yang, E.-F. *et al.* Taxonomic Reappraisal of Periconiaceae with the Description of Three New *Periconia* Species from China. *JoF* **8**, 243 (2022).
5. Bovio, E. *et al.* The culturable mycobiota associated with three Atlantic sponges, including two new <i/> species: Thelebolus balaustiformis and T. spongiae. *Fungal Systematics and Evolution* **1**, 141–167 (2018).
6. Cantrell, S. A., Hanlin, R. T. & Emiliano, A. *Periconia variicolor* sp. nov., a new species from Puerto Rico. *Mycologia* **99**, 482–487 (2007).
7. Prasher, I. B. & Verma, R. K. *Periconia* Species New To North- Western Himalayas. 3 (2012).
8. Crous, P. W. *et al.* Fungal Planet description sheets: 785–867. *persoonia* **41**, 238–417 (2018).
9. Markovskaja, S. & Kačergius, A. Morphological and molecular characterisation of *Periconia pseudobyssoides* sp. nov. and closely related P. byssoides. *Mycol Progress* **13**, 291–302 (2014).
10. Mandyam, K., Fox, C. & Jumpponen, A. Septate endophyte colonization and host responses of grasses and forbs native to a tallgrass prairie. *Mycorrhiza* **22**, 109–119 (2012).
11. Verma, V. C., Lobkovsky, E., Gange, A. C., Singh, S. K. & Prakash, S. Piperine production by endophytic fungus *Periconia* sp. Isolated from Piper longum L. *J Antibiot* **64**, 427–431 (2011).
12. Li, J. Y. *et al.* The induction of taxol production in the endophytic fungus–*Periconia* sp from Torreya grandifolia. *Journal of Industrial Microbiology and Biotechnology* **20**, 259–264 (1998).
13. Romero, A., Carrion, G. & Rico-Gray, V. Fungal latent pathogens and endophytes from leaves of. *Fungal Diversity* 7 (2001).
14. Odvody, G., Dunkle, L. & Edmunds, L. *Phytopathology* **67**, 1485–1489 (1977).
15. Gunasekaran, R. *et al. Periconia* species - An unusual fungal pathogen causing mycotic keratitis. *Indian Journal of Medical Microbiology* **39**, 36–40 (2021).
16. Azhari, A. & Supratman, U. The Chemistry and Pharmacology of Fungal Genus *Periconia*: A Review. *Sci. Pharm.* **89**, 34 (2021).
17. Galiana, E. *et al.* Ecosystem Screening Approach for Pathogen-Associated Microorganisms Affecting Host Disease. *Appl Environ Microbiol* **77**, 6069–6075 (2011).
18. Galiana, E., Ponchet, M. & Marais, A. Treatment of plants against oomycete infection. (2011).
19. Zhu, Y. *et al.* Screening and isolation of antinematodal metabolites against Bursaphelenchus xylophilus produced by fungi. *Ann. Microbiol.* **58**, 375–380 (2008).
20. Knapp, D. G. *et al.* Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Sci Rep* **8**, 6321 (2018).
21. Blaxter, M. *et al.* Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. USA* **119**, e2115636118 (2022).
22. Taylor, T. N., Krings, M. & Taylor, E. L. Ascomycota. in *Fossil Fungi* 129–171, https://doi.org/10.1016/B978-0-12-387731-4.00008-6 (Elsevier, 2015).
23. Galiana, E. *et al.* Genbank https://identifiers.org/ncbi/insdc:HM161743 (2011).
24. Neuhauser, S., Huber, L. & Kirchmair, M. A DNA based method to detect the grapevine root-rotting fungus Roesleria subterranea in soil and root samples. 19 (2011).
25. Plassart, P. *et al.* Evaluation of the ISO Standard 11063 DNA Extraction Procedure for Assessing Soil Microbial Abundance and Community Structure. *PLoS ONE* **7**, e44279 (2012).
26. Schneegurt, M., Dore, S. & Kulpa, C. Direct Extraction of DNA from Soils for Studies in Microbial Ecology. *Current Issues in Molecular Biology* https://doi.org/10.21775/cimb.005.001 (2003).
27. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. in *PCR Protocols* 315–322, https://doi.org/10.1016/B978-0-12-372180-8.50042-1 (Elsevier, 1990).
28. Vilgalys, R. & Hester, M. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species. *J Bacteriol* **172**, 4238–4246 (1990).
29. Bovio, E., Industri, B. & Ponchet, M. Genbank https://identifiers.org/ncbi/insdc:OP329216 (2022).
30. Bovio, E., Industri, B. & Ponchet, M. Genbank https://identifiers.org/ncbi/insdc:OP329219 (2022).
31. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**, 1547–1549 (2018).
32. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* https://doi.org/10.1093/oxfordjournals.molbev.a040023 (1993).
33. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
34. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).
35. Nurk, S. *et al.* HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).
36. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. & Blaxter, M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet.* **4** (2013).
37. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Res* **6**, 1287 (2017).
38. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
39. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
40. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

41. Sallet, E., Gouzy, J. & Schiex, T. EuGene: An Automated Integrative Gene Finder for Eukaryotes and Prokaryotes. in *Gene Prediction* (ed. Kollmar, M.) vol. **1962** 97–120 (Springer New York, 2019).
42. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **46**, 2699–2699 (2018).
43. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
44. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
45. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).
46. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
47. Donath, A. *et al.* Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research* **47**, 10543–10552 (2019).
48. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* **41**, 95–98 (1999).
49. Bovio, E. *et al.* https://identifiers.org/ncbi/insdc:OP787475 (2022).
50. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**, W59–W64 (2019).
51. Ponchet, M., Seassau, A. & Magliano, M. *Periconia digitata* proteins identified by mass spectrometry. *Periconia digitata* proteins identified by mass spectrometry. *Recherche Data Gouv* https://doi.org/10.57745/FHNYJU (2023).
52. *ENA European Nucleotide Archive.* https://identifiers.org/ena.embl:PRJEB55037 (2023).
53. Ponchet, M., Seassau, A. & Magliano, M. *PRIDE.* https://identifiers.org/pride.project:PXD038112 (2023).
54. Ponchet, M., Seassau, A. & Magliano, M. *PRIDE.* https://identifiers.org/pride.project:PXD038175 (2023).
55. Rancurel, C. *Periconia digitata* structural annotation. *Recherche Data Gouv* https://doi.org/10.57745/BAPYY5 (2023).
56. Rancurel, C. & Danchin, E. J. *Periconia digitata* Interproscan annotation. *Recherche Data Gouv* https://doi.org/10.57745/GU67TE (2023).
57. Espagne, E., Balhadère, P., Penin, M.-L., Barreau, C. & Turcq, B. HET-E and HET-D Belong to a New Subfamily of WD40 Proteins Involved in Vegetative Incompatibility Specificity in the Fungus *Podospora anserina*. *Genetics* **161**, 71–81 (2002).
58. Nicolaisen, M., Frisvad, J. C. & Rossen, L. A Penicillium freii gene that is highly similar to the β-keto-acyl synthase domain of polyketide synthase genes from other fungi. *Letters in Applied Microbiology* **25**, 197–201 (1997).
59. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).
60. Zheng, F. *et al.* Whole-Genome Sequence and Comparative Analysis of Trichoderma asperellum ND-1 Reveal Its Unique Enzymatic System for Efficient Biomass Degradation. *Catalysts* **12**, 437 (2022).
61. Chang, Y. *et al.* Evolution of zygomycete secretomes and the origins of terrestrial fungal ecologies. *iScience* **25**, 104840 (2022).
62. Pellegrin, C., Morin, E., Martin, F. M. & Veneault-Fourrey, C. Comparative Analysis of Secretomes from Ectomycorrhizal Fungi with an Emphasis on Small-Secreted Proteins. *Front. Microbiol.* **6** (2015).
63. Bovio, E. *et al. Periconia digitata* genome. *Recherche Data Gouv.* https://doi.org/10.57745/FB9OVN (2023).
64. Rancurel, C. *Periconia digitata* gene fasta file. *Recherche Data Gouv* https://doi.org/10.57745/5ODV0C (2023).
65. Rancurel, C. *Periconia digitata* CDS fasta file. *Recherche Data Gouv* https://doi.org/10.57745/71DS6F (2023).
66. Rancurel, C. *Periconia digitata* protein fasta file. *Recherche Data Gouv* https://doi.org/10.57745/TV4EB7 (2023).

Acknowledgements

Author contributions

M.P., E.B., C.R. C.V.G. and E.G.J.D. conceived the study. C.R., EB., M.P., E.G.J.D. and C.V.G. wrote the manuscript with input from all authors. C.R., C.K. and E.G.J.D. designed and carried out the bioinformatic analyses and data transfer. E.B. produced the biological material. C.V.G. carried out the nucleic acid extractions. M.P., A.S. and M.M. designed and carried out the proteomic analyses. M.G. performed the PacBio HiFi library and sequencing. A.L. performed the RNA libraries preparations and Illumina sequencing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.B. or C.R.

Reprints and permissions information is available at www.nature.com/reprints.

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