

















## IMA GENOME - F20

# A draft genome assembly of *Agroathelia rolfsii*, *Ceratobasidium papillatum*, *Pyrenopeziza brassicae*, *Neopestalotiopsis macadamiae*, *Sphaerellopsis filum* and genomic resources for *Colletotrichum spaethianum* and *Colletotrichum fructicola*

Davide D'Angelo<sup>1</sup><sup>\*</sup>, Roberto Sorrentino<sup>2\*</sup>, Tiphany Nkomo<sup>3</sup>, Xianzhi Zhou<sup>4</sup>, Niloofar Vaghefi<sup>5</sup>, Byron Sonnekus<sup>3</sup>, Tanay Bose<sup>3,6</sup>, Domenico Cerrato<sup>2</sup>, Loredana Cozzolino<sup>7</sup>, Nicky Creux<sup>8</sup>, Nunzio D'Agostino<sup>1</sup>, Gerda Fourie<sup>3</sup>, Giovanna Fusco<sup>7</sup>, Almuth Hammerbacher<sup>6</sup>, Alexander Idnurm<sup>9</sup>, Levente Kiss<sup>10,11,12</sup>, Yanping Hu<sup>13</sup>, Hongli Hu<sup>13</sup>, Ernesto Lahoz<sup>2</sup>, Jason Risteski<sup>5</sup>, Emma T. Steenkamp<sup>3</sup>, Maurizio Viscardi<sup>7</sup>, Magriet A. van der Nest<sup>14</sup>, Yuan Wu<sup>15</sup>, Hao Yu<sup>15</sup>, Jianjin Zhou<sup>16</sup>, Chinthani S. Karandeni Dewage<sup>17</sup>, Loly I. Kotta-Loizou<sup>17</sup>, Henrik U. Stotz<sup>17</sup>, Bruce D. L. Fitt<sup>17</sup>, Yong-Ju Huang<sup>17</sup>, Brenda D. Wingfield<sup>3</sup>

1 Department of Agricultural Sciences, University of Naples Federico II, piazza Carlo di Borbone 1, 80055, Portici, Naples, Italy

2 Research Centre for Cereal and Industrial Crops (CREA-CI), via Torrino 3, 81100, Caserta, Italy

3 Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0028, South Africa

4 Institute of Plant Protection, Fujian Academy of Agricultural Sciences, Wusi Road 247, Fuzhou 350003, China

5 School of Agriculture, Food and Ecosystem Sciences, Faculty of Science, The University of Melbourne, Parkville, Australia

6 Department of Zoology and Entomology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0028, South Africa

7 Istituto Zooprofilattico Sperimentale del Mezzogiorno, Via Salute 2, 80055, Portici, Naples, Italy

8 Department of Plant and Soil Science, Forestry and Agricultural Biotechnology (FABI), University of Pretoria, Pretoria 0028, South Africa

9 School of BioSciences, Faculty of Science, The University of Melbourne, Parkville, Australia

10 Centre for Crop Health, University of Southern Queensland, Toowoomba, Australia

11 Eszterházy Károly Catholic University, Eger, Hungary

12 Plant Protection Institute, Centre for Agricultural Research, HUN-REN, Budapest, Hungary

13 College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China

14 Hans Merensky Chair in Avocado Research, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0028, South Africa

15 Technology Center, Xiamen Customs, Xiamen 361026, China

16 Sanming Academy of Agricultural Sciences/Fujian Key Laboratory of Crop Genetic Improvement and Innovative Utilization for Mountain Area, Sanming, Fujian 365051, China

17 Centre for Agriculture, Food and Environmental Management Research, School of Life and Medical Sciences, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK

Corresponding author: Brenda D. Wingfield ([brenda.wingfield@up.ac.za](mailto:brenda.wingfield@up.ac.za))



Copyright: © Davide D'Angelo et al.

This is an open access article distributed under

terms of the Creative Commons Attribution

License (Attribution 4.0 International – CC BY 4.0).

\* These authors contributed equally to this work and share first authorship for the section that was submitted on the *Agroathelia rolfsii* genome.

Academic editor: Marc Stadler  
Received: 13 November 2024  
Accepted: 13 November 2024  
Published: 17 February 2025

**Citation:** D'Angelo D, Sorrentino R, Nkomo T, Zhou X, Vaghefi N, Sonnekus B, Bose T, Cerrato D, Cozzolino L, Creux N, D'Agostino N, Fourie G, Fusco G, Hammerbacher A, Idnurm A, Kiss L, Hu Y, Hu H, Lahoz E, Risteski J, Steenkamp ET, Viscardi M, van der Nest MA, Wu Y, Yu H, Zhou J, Karandeni Dewage CS, Kotta-Loizou LI, Stotz HU, Fitt BDL, Huang YJ, Wingfield BD (2025) IMA GENOME - F20 A draft genome assembly of *Agroathelia rolfsii*, *Ceratobasidium papillatum*, *Pyrenopeziza brassicae*, *Neopestalotiopsis macadamiae*, *Sphaerellopsis filum* and genomic resources for *Colletotrichum spaethianum* and *Colletotrichum fructicola*. IMA Fungus 16: e141732. <https://doi.org/10.3897/imafungus.16.141732>

## IMA GENOME-F 20A

### A draft genome assembly of *Agroathelia rolfsii*, the aetiological agent of southern blight disease on industrial hemp

#### Introduction

Hemp is an ancient plant that, over the centuries, has united generations and continents, shaping agricultural landscapes. Over time, the economies, uses and traditions of many people have been influenced by this species and its numerous applications (Clarke and Merlin 2016).

Italy, a major producer and exporter of hemp, has progressively reduced its cultivated area to the point of almost eliminating it due to the laboriousness of cultivation and processing, the diffusion of synthetic fibres, the availability of alternative and cheaper vegetable fibres such as cotton and jute, and regulations on the use of drugs (Amaducci et al. 2015).

With the approval of Law N. 242 of 2 December 2016, entitled "Provisions for the promotion of hemp cultivation and the agro-industrial supply chain," the interest of the agricultural, industrial, and entrepreneurial world towards hemp and its multifunctionality is renewed.

Hemp is susceptible to several diseases (Sorrentino et al. 2019). Among the plethora of pathogens that pose a serious threat to hemp cultivation (Punja 2021), the soil borne *Agroathelia rolfsii* (*Sclerotium rolfsii* Sacc.), the agent of southern blight, is one of the most recognized fungal pathogens in southern temperate zones and the tropics, affecting both fiber and seed cultivations (Ferri 1961a). *A. rolfsii* attacks over 500 plant species (Aycock 1966), including both monocotyledons and dicotyledons (Farr et al. 1989). Attacks on hemp have been reported in India, the United States, Japan, Greece, and Italy (Ferri 1961a; McPartland and Clarke 2000; Amaradasa et al. 2020; Mersha et al. 2020; Chatzaki et al. 2022).

Fungi of the genus *Sclerotium* are characterized by the production of sclerotia and sterile mycelia, devoid of spores. The genus *Sclerotium* includes more than 40 phytopathogenic species (Farr 2008).

Many *Sclerotium* species reproduce asexually, with sexual reproduction rare, and are recognized exclusively in their anamorphic phase (Punja and Rahe 2001). *Sclerotium rolfsii*, the best-known species of the genus, presents the teleomorph phase *Agroathelia rolfsii* (Curzi), characterized by the development of resupinate basidiocarps and hyphal filaments emerging from the germinating sclerotia (Tu and Kimbrough 1978).

A total of three *A. rolfsii* isolates, which produce symptoms of stem rot disease on groundnut, were isolated. Their genomes were sequenced and assembled, providing insights into aspects of the pathogen's pathogenicity (Iquebal et al. 2021; Yan et al. 2021). Iquebal et al. (2021) released a genome (MR10 strain) of 72.3 Mb in length and identified key virulence genes. Yan et al. (2021) compared the genomes of two different strains, namely GP3 and ZY, and revealed variations in their pathogenicity associated with a significant difference in aggressiveness on peanut plants.

Noteworthy, strains MR10 and ZY, which have genome sizes exceeding 70 Mb, are noted in the NCBI database as "atypical assemblies" due to their unusually large genome sizes. Assembling and annotating the genome of *A. rolfsii* isolated from hemp can help researchers to develop more effective control strategies to protect this crop. The identification of putative virulence candidate genes can be

A

B



**Figure 1.** **A** Hempseed field where *Agroathelia rolfsii* was discovered, showing symptoms of yellowing and desiccation **B** hemp plant collar where the pathogen creates a constriction with prolific production of both mycelium and sclerotia.

exploited to develop agrochemicals or targeted therapeutic approaches for disease control, mainly in the post-penetration phase. Here, we report the draft genome of *A. rolfsii* isolated from industrial hemp in Campania in June 2022 (Fig. 1).

### Sequenced strain

Italy: Campania: Caserta. Isolated from symptomatic plant collar of *Cannabis sativa* L. var. Codimodo, 07/2021.

### Nucleotide sequence accession number

The draft genome sequence has been deposited in NCBI GenBank under the accession number JAXRUS000000000. The version described in this paper is version JAXRUS010000000. Accession for SRA data: PRJNA1052013.

### Fungal isolation

The collars of 10 symptomatic plants were washed under running water for 10 minutes, then cut into 3 mm sections. For each sample, five sections of the collar underwent surface sterilization by immersion in 5% sodium hypochlorite for 1 minute, followed by immersion in 70% ethanol for an additional minute. The excised small pieces of tissue were then rinsed five times in sterile distilled water in a 50 ml tube, dried by blotting on sterile 3 mm Whatman paper, subdivided into smaller portions of tissue and, finally plated on potato dextrose agar (PDA) amended with 100 mg liter<sup>-1</sup> of streptomycin sulphate and stored at 24 °C ±2 °C in the dark. Plates were inspected daily, and growing colonies were sub-cultured on PDA for 10 days. In total, 50 monohyphal purified fungal isolates were obtained. Colonies of these isolates were white and produced sclerotia after 10 days. We identified the target fungal isolate from the pool by sequencing the Internal Transcribed

Spacer (ITS), as amplified using PCR. DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, USA) following the manufacturer's instructions and PCR reactions were carried out using 10 ng of DNA with the ITS1/ITS4 primers (White et al. 1990) specific for the ITS regions. Each 50 µL reaction mixture included 1X reaction buffer, 200 µM dNTPs, 500 mM of each primer, and 1 unit of Q5 Hi-Fi DNA polymerase (New England Biolabs, UK). The PCR conditions were the following: denaturation at 94 °C for 1 minute, 35 cycles of 94 °C for 30 seconds, 58 °C for 30 seconds, and 72 °C for 30 seconds, with a final extension at 72 °C for 5 minutes. Amplicons were purified using the Nucleospin Extract Kit (Macherey–Nagel, Germany) and sequenced at BMR Genomics (Padova-Italy). Fungi are maintained in PDA slant tubes and in a 20% glycerol stock solution at the phytopathology laboratory of the CREA, Cereal and Industrial Crop Research Centre, Caserta.

### DNA isolation and sequencing library preparation

DNA was purified from mycelia obtained by culturing *A. rolfsii* on PDA plates covered with cellophane films and kept in the dark at 24 °C for 7 days. Mycelia were collected, ground in liquid nitrogen and 100 mg were used to isolate the DNA using the QIAGEN® Genomic-tip kit following the manufacturer's instructions. Genomic DNA was resuspended in EB buffer (Qiagen) to avoid EDTA in the final preparation. DNA quality (260/280 ratio) was assessed using a NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE). The amount of DNA was assessed using the Qubit platform (Invitrogen, Paisley, UK). Library preparation was performed using ILMN DNA LP (M) Tagmentation (96 Samples, IPB), and IDT® for Illumina® DNA/RNA UD Indexes Set A, Tagmentation (Illumina Italy S.r.l. Milano-Italy) according to the Illumina DNA preparation guide. The quality and amplicon size of the libraries were assessed using a Tapestation 4150 (Agilent) at multiple steps during the protocol, typically after size selection and PCR amplification. Three different libraries were normalized to a working concentration pool of 12 pM using the molarity calculated from Qubit measurements adjusted for fragment size. The sequencing run was performed on Illumina MiSeq (Illumina), equipped with MiSeq Reagent Kit v3 (600 cycles) which includes: Paired-End Reagent Plate, MiSeq Flow Cell.

### Read pre-processing

Quality control on FASTQ files (paired-end 150 bp) was performed using FastQC v0.11.9 (Andrews 2020). The raw reads were subjected to a pre-processing step using Trimmomatic v0.39 (Bolger et al. 2014) with the following parameters: SLIDINGWINDOW:4:20 LEADING:20 TRAILING:20 MINLEN:75. High quality reads were used to compute the *k*-mer distribution using Kmergenie v1.7051 (Chikhi and Medvedev 2014). Genome size was estimated using GenomeScope v1.0 (Vurture et al. 2017) based on the *k*-mer distribution.

### Genome assembly

*De novo* genome assembly was performed using ABYSS v2.3.5 (Jackman et al. 2017), following the default pipeline for paired-end libraries. The *k*-mer size was set to 107 bp. Scaffolding was performed with Ragtag v2.1.0 (Alonge et al. 2022),

sorting and orienting contigs along the *A. rolfsii* reference genome (GenBank accession number [GCA\\_018343915.1](#)) using a whole-genome alignment approach.

Unique alignments less than 500 bp and scaffolds less than 1,000 bp were filtered out. Gap closure was achieved using Sealer v2.3.5 (Paulino et al. 2015), employing the initial reads and considering 150 bp-sized “flanks” of the scaffolds as pseudo-reads.

Finally, the draft genome sequence was subjected to repeat region masking using RepeatMasker v4.1.2 (Tarailo-Graovac and Chen 2009) in slow search mode with *rmblastn* as the search engine and the *Agaricomycetidae* dataset from the Dfam repository (Storer et al. 2021) release 3.3 as filtering database.

### Assembly assessment

Quality assessment of the assembly was performed with QUAST v5.2.0 (Gurevich et al. 2013). Assembly completeness was assessed by mapping reads to the final assembly using BWA v0.7.17 (Li and Durbin 2009) with default parameters. The completeness of the genome annotation was assessed based on BUSCO v5.4.4 (Simão et al. 2015) using *Agaricomycetes* as the reference dataset.

### Comparative and phylogenetic analysis

The assembled genome was compared with the *A. rolfsii* reference genome downloaded from GenBank (accession number [GCA\\_018343915.1](#)) using MUMmer v3.23 (Marçais et al. 2018), performing a whole genome alignment with a minimum cluster length of 100.

The ITS sequence from an *A. rolfsii* strain isolated from *C. sativa* (GenBank accession: [MZ242252.1](#)) was utilized as a BLASTn query to identify the corresponding ITS sequence within the assembled genome. A phylogenetic analysis was then conducted using the identified ITS sequence from the draft genome, along with sequences from 16 *A. rolfsii* strains, 5 *A. delphinii* strains, 3 *A. coffeicola* strains, and a *Rhizoctonia solani* strain used as an outgroup. Multiple sequence alignment was performed using Mafft v.7.490 with the G-INS-i method and 1,000 iterations. Finally, a phylogenetic tree was generated using IQ-TREE v2.0.7 (Minh et al. 2020) using the ModelFinder option to select the best DNA substitution model and performing 1,000 ultrafast bootstrap replicates for the consensus tree inference. The resulting tree was visualized and edited using the ITOL tool v5 (Letunic and Bork 2021).

The protein complement of *A. rolfsii* was compared with the proteomes of four other fungal species belonging to the *Agaricomycotina*. To this end, the protein sequences of *Rhizoctonia solani* ([GCA\\_016906535.1](#)), *Armillaria mellea* ([GCA\\_030407055.1](#)), *Agaricus bisporus* ([GCA\\_000300575.1](#)), and *Pleurotus ostreatus* ([GCA\\_014466165.1](#)) were downloaded from GenBank.

OrthoFinder v2.5.5 (Emms and Kelly 2018) was used with default settings to identify orthologous groups. The ggVennDiagram v1.4.9 R package (Gao et al. 2021) was used to create a Venn diagram showing the number of shared orthologous groups.

BinGO v3.0.3 software (Maere et al. 2005) was used to perform GO enrichment analysis. GO terms associated with *A. rolfsii* species-specific proteins

were used as the test set, while GO terms associated with the complete set of proteins were selected as the background set.

The enrichment analysis involved a hypergeometric test and a Benjamini and Hochberg false discovery rate (FDR) correction of the p-value. Only GO terms in the biological process and molecular function domains with an FDR-corrected p-value > 0.05 were considered statistically significant.

## Gene prediction

Gene prediction was performed by running Helixer v0.3.2 (Stiehler et al. 2021), which employs a deep neural network model trained on 128 fungal species and validated on 170 species. Then, the evidence collected from the first round of annotation was used as input for GeneMark-ES v4 (Borodovsky and Lomsadze 2011) with default parameters. Gffread v0.12.8 (Pertea and Pertea 2020) was used to extract coding sequences (CDS) and protein sequences, filtering out those with incorrect start and stop codons.

## Functional annotation

Blast2GO v3.1.9 (Gotz et al. 2008) was used to assign a functional annotation to the predicted genes. Initially, InterProScan was used to query all InterPro (Paysab-Lafosse et al. 2023) member databases, subsequently a second annotation run was performed via the EggNOG Mapper tool in BLAST2GO, searching for Cluster of Orthologous Groups (COG). The final gene ontology (GO) annotation was generated by a consensus of the results from both previous analyses. The KAAS web server (Moriya et al. 2007) was used to assign genes to KEGG molecular pathways. Carbohydrate active enzymes (CAZymes) were annotated using the dbCAN3 server (Zheng et al. 2023), to query the dbCAN, dbCANsub, and CAZY databases. A consensus on the three results led to the final list of CAZymes. Proteins putatively involved in host-pathogen interactions were identified by querying the HPIDB 3.0 database (Ammari et al. 2016) limiting the search to the model species *Arabidopsis thaliana* and *Saccharomyces cerevisiae*. Effector proteins were predicted using the Phobius web server (Käll et al. 2007), retaining only proteins with signal peptides and no transmembrane domains. Sequences with a maximum length of 500 amino acids and at least 5 cysteine residues were retained and collected into the list of high-fidelity effectors. Finally, gene sequences were submitted to the antiSMASH-fungal web server v6.0 (Blin et al. 2023) to identify gene clusters involved in the synthesis of secondary metabolites.

## Results and discussion

The sequencing run yielded 10.5 Gbp with a Q index >30. A total of 25,898,770 high-quality paired-end reads were used for genome assembly, resulting in 3,138 contigs, whose total sequence length is 44,845,534 bp with an N50 of 3,998,658 bp (Table 1).

The size of the assembled genome were larger than that estimated based on k-mer distribution (~45 Mb vs ~37 Mb) assuming a genome coverage of 86.6 X. A total of 427,196 bp (0.95%) were masked, comprising 9,616 simple repeats and 1,663 low complexity regions. The reads used in the assembly process



**Table 1.** Summary statistics on genome assembly.

Assembly features	<i>A. rolfsii</i> scaffolds
Total sequence length (bp)	44,845,534
Number of contigs	3,138
Number of contigs ( $\geq 50,000$ bp)	16
Largest contig size (bp)	5,959,244
GC content (%)	46.49
N50	3,998,658
N90	4,804
L50	5
L90	1,500
Missingness (%)	0.406

were mapped onto the assembled genome to assess assembly accuracy. Over 99.9% of the reads were successfully mapped.

A total of 14,804 genes was predicted, including 11,230 protein-coding genes and 3,574 non-coding RNAs. The BUSCO analysis indicated a completeness level of 92.5% for the *Agaricomycetes*, identifying 2,471 complete and single-copy BUSCOs, along with 208 complete but duplicated BUSCOs, and 72 fragmented BUSCOs. A total of 36,139 GO terms were assigned to 4,428 protein coding genes. A total of 4,086 proteins has been assigned to KEGG, covering 311 pathways (Suppl. material 1). Functional annotation has allowed the identification of various genes that may be involved in pathogenesis processes and could be potential targets for phenotyping studies. Within the annotated proteins, 272 putative carbohydrate-active enzymes were identified (Suppl. material 1).

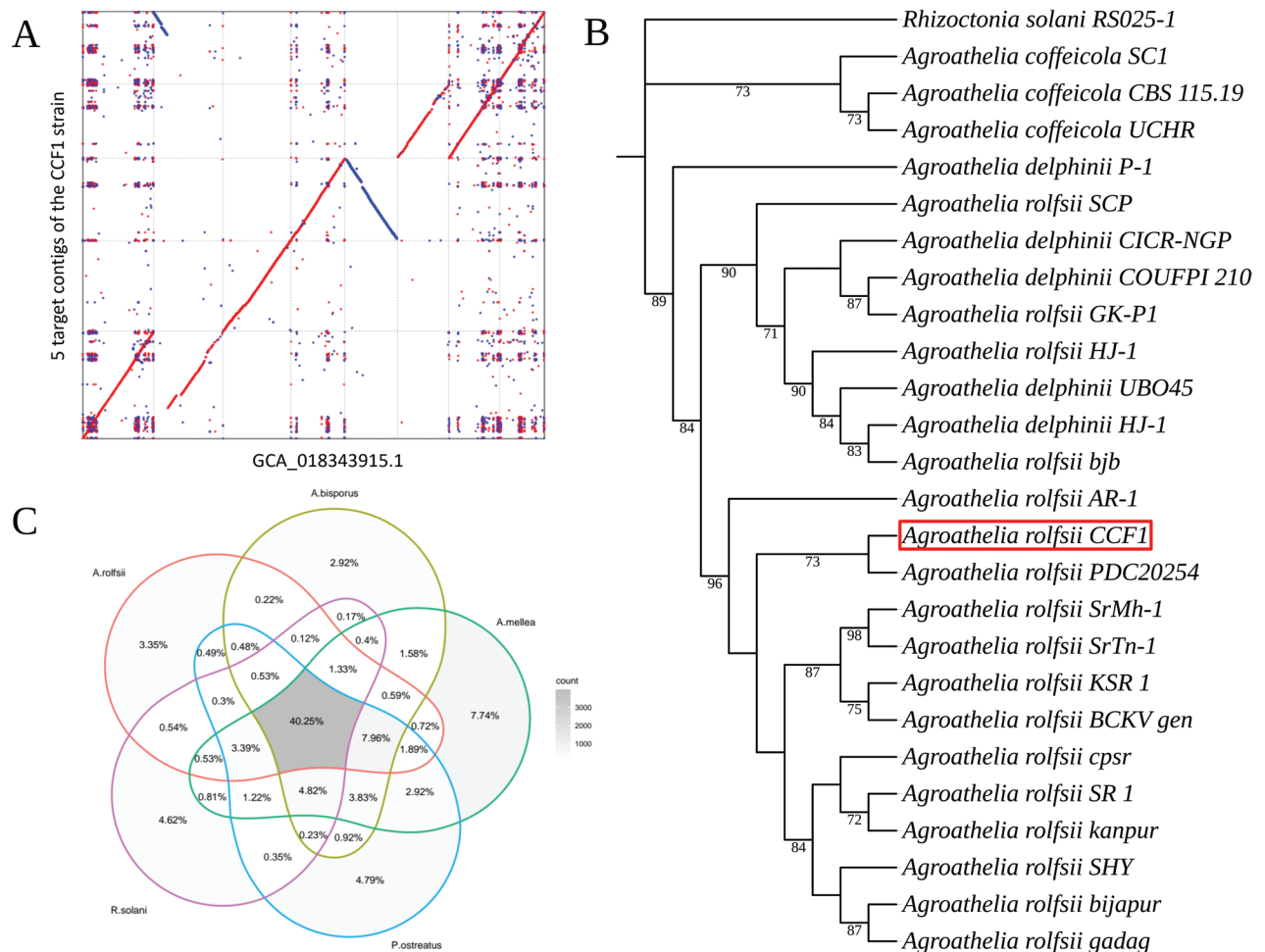
The annotation of genes involved in the biosynthesis of secondary metabolites led to the identification of 137 genes grouped into 36 clusters (Suppl. material 1).

Nine hundred sixty-two ( $N = 962$ ) proteins possibly secreted into the extracellular space were identified, as they include a signal peptide and lack transmembrane domains. Further filtering based on the number of amino acid residues and cysteine content reduced the dataset to 30 putative pathogenesis-related (PR) proteins (Suppl. material 1).

Finally, exploration of plant-pathogen interactions led to the identification of 73 protein-coding genes involved in 126 interactions, of these, 37 were classified as "association", 26 as "direct interaction", 31 as "physical association", 29 as "suppressive genetic interaction defined by inequality" and one as "colocalization" (Suppl. material 1).

Alignment between the 3,138 contigs and the 8 scaffolds of the *A. rolfsii* reference genome ([GCA\\_018343915.1](#)) resulted in 76.34% aligned bases and a total of 584,388 single nucleotide polymorphisms, indicating a high degree of similarity. Furthermore, aligning only the five largest contigs still provided good coverage of the [GCA\\_018343915.1](#) reference genome (Fig. 2A).

The substitution model "K3Pu+F" was identified as the most suitable model for tree inference using the maximum likelihood method. The ITS sequence-based tree positioned the *A. rolfsii* strain CCF1 within a major clade alongside other *A. rolfsii* strains, showing a distinct clustering pattern. In contrast, *A. delphinii* and *A. coffeicola* strains formed separate clusters (Fig. 2B).



**Figure 2.** **A** Dot plot illustrating the alignment results between the five largest contigs assembled in this study and the 8 scaffolds from the *A. rolfii* reference genome (GenBank accession: [GCA\\_018343915.1](#)) **B** Maximum-Likelihood tree constructed using Internal Transcribed Spacer (ITS) sequences, depicting the phylogenetic relationships among taxa within the *Agroathelia* genus. *Rhizoctonia solani* was used as the outgroup, with bootstrap values > 70 indicated on the branches **C** Venn diagram displaying orthogroups identified in the proteomes of *A. rolfii* and four other fungal species within the *Agaricomycotina*.

Comparison between the proteomes of *A. rolfii*, *R. solani*, *A. mellea*, *A. bisporus* and *P. ostreatus* led to the identification of 9,854 orthogroups, 3,967 (40.3%) of which were classified as core-orthogroups. A total of 10,441 (93%) *A. rolfii* protein-coding genes were assigned to 6,176 orthogroups, of which 330 (3.35%) were species-specific (Fig. 2C). The latter contained 1,916 species-specific genes, which were subjected to Gene Ontology enrichment analysis, revealing 42 overrepresented GO terms in the biological process domain including some of the most significant ones related to macromolecules and secondary metabolite biosynthetic processes, responses to environmental stresses, biosynthesis of aromatic compounds, defence responses to other organisms.

Plant pathogenic fungi pose a major threat to agriculture, causing significant yield losses. The development of new organic or traditional agrochemicals and targeted therapeutic approaches for the control of plant pathogenic fungi requires a comprehensive knowledge of their gene repertoire.





**Figure 3.** Enriched gene ontology terms in the Biological Processes domain were identified for *A. rolfsii* species-specific protein-coding genes following the OrthoFinder analysis.

In the present study, the genome of the pathogenic fungus *A. rolfsii* strain CCF1 isolated from hemp (*Cannabis sativa* L.) was assembled and annotated. The data produced will allow the scientific community to have a further resource to study genetic architecture and discover new effectors and their pathogenic mechanism. Indeed, few *A. rolfsii* genomes have been assembled so far. Although fragmented into over 3,000 contigs, the genome we released is quite complete. Indeed, the five largest contigs cover much of the genome of

another *A. rolfsii* strain sequenced using third-generation sequencing technologies (Yan et al. 2021).

In this study, a combination of self-learning and deep learning algorithms, so far rarely applied in fungal genome annotation, allowed the prediction of a considerable number of genes: 11,230 protein-coding genes and 3,574 non-coding RNAs. The quality of the annotation was satisfactory and near-complete as 92.5% of the *Agaricomycetes* BUSCOs were identified.

The phylogenetic analysis revealed the evolutionary context of *A. rolfsii* CCF1, placing it within a major clade that exclusively includes *A. rolfsii* isolates. In contrast, *A. coffeicola* and *A. delphinii* isolates were positioned in separate clades. The search for orthologs between different fungal species led to the identification of core-orthogroups, thus highlighting common and species-specific protein-coding genes.

The identification of overrepresented GO terms associated with species-specific genes provides valuable insights into the molecular mechanisms that have shaped the biology of *A. rolfsii* CCF1. The presence of these genes suggests evolutionary innovations that have contributed to its divergence and adaptation from other fungal relatives within the *Agaricomycotina*. Studying these genetic features can enhance our understanding of the evolutionary processes driving diversification and specialization among fungi. From a biological perspective, analysing species-specific proteins in *A. rolfsii* through GO enrichment provides insights into unique biological processes and molecular functions critical for its survival and adaptation. The presence of various types of interactions suggests a complex relationship between the fungal pathogen and the host plant; these could be studied to gain a deeper understanding of the molecular mechanisms of host-pathogen coevolution and adaptation. In addition, the effector identification procedure narrows the field to a small number of target genes for further study. Further studies on these genes could deepen our knowledge of *A. rolfsii*'s biology, illuminate its distinct characteristics, and aid in developing precise strategies for managing hemp diseases.

**Authors:** Davide D'Angelo, Roberto Sorrentino, Ernesto Lahoz, Domenico Cerrato, Maurizio Viscardi, Loredana Cozzolino, Giovanna Fusco, Nunzio D'Agostino\*

**\*Contact:** [nunzio.dagostino@unina.it](mailto:nunzio.dagostino@unina.it)

## IMA GENOME-F 20B

### Draft genome assembly of *Ceratobasidium papillatum* CBS 570.83

#### Introduction

*Ceratobasidium* is a genus of *Basidiomycetes* fungus that plays diverse ecological roles (Mosquera-Espinosa et al. 2013). Many *Ceratobasidium* species are plant pathogens, such as *Ceratobasidium theobromae*, *C. cereale*, *C. cornigerum*, and *C. anceps*, which cause diseases like root rot, white thread blight, vascular streak dieback, and damping-off (Schroeder and Paulitz 2012; McMahon and Purwantara 2016; De Melo et al. 2018; Lima et al. 2020). In contrast some species, such as *C. bicornis* and *C. cornigerum*, are saprophytes with bioremediation potential (Fazary et al. 2016; Harrington et al. 2011; Haas et al. 2008). *Ceratobasidium* species are considered *Rhizoctonia*-like fungi that form

symbiotic associations with orchids (Warcup and Talbot 1980; Dearnaley 2007; Decruse et al. 2018; Thixton et al. 2020).

Presently, there are 44 species within the *Ceratobasidium* genus (<https://www.mycobank.org/>). However, DNA sequences are unavailable for the majority of these species. Genomic data are only available for *C. theobromae*, the causative agent of vascular-streak dieback (VSD) in cacao (Ali et al. 2019). While other *Ceratobasidium* genomes are available on NCBI and MycoCosm, they pertain to undescribed species. To fill this gap, we sequenced, assembled, and annotated the genome of *C. papillatum*.

*Ceratobasidium papillatum* was described as a symbiont of orchids by Warcup and Talbot (1980). It has subsequently been detected in the roots of various other orchids (Freestone et al. 2021). Thus, the genome sequence of this species is a pivotal foundation for conducting mechanistic and comparative studies, encompassing both beneficial and pathogenic fungi within the *Ceratobasidium* genus.

### Sequenced strain

Australia: Toowoomba, Queensland: isolated from *Sarcochilus dilatatus* (CBS 570.83 ex type strain).

### Nucleotide sequence accession number

This Whole Genome Shotgun project for *Ceratobasidium papillatum* CBS 570.83 has been deposited at DDBJ/ENA/GenBank under the accession JAYRCO000000000.1; BioProject PRJNA1046290 and BioSample SAMN38480858. The version described in this paper is version JAYRCO010000000.

### Materials and methods

The isolate *Ceratobasidium papillatum* (CBS 570.83) was received from the Westerdijk Fungal Biodiversity Institute in the Netherlands. The isolate was revitalised on half-strength potato dextrose agar (PDA) medium (19 g PDA powder (Merck, South Africa); 7 g agar; distilled water 1 L) at 25 °C. Total genomic DNA was extracted from 14-day-old cultures following the protocol suggested by Duong et al. (2013).

The sequencing library was prepared using the MGIEasy Universal DNA library prep kit to generate 150 bp paired-end libraries. Whole genome sequencing was performed using the MGI DNBSeg G400 sequencer at the Agricultural Research Council Biotechnology Platform, South Africa (ARC-BTP) using the PE150 sequencing strategy.

The quality of the reads was assessed using FastQC v0.11.7 (<https://github.com/s-andrews/FastQC>) and was assembled using SPAdes v3.15.0 (Bankevich et al. 2012). Quast v5.0.2 (Mikheenko et al. 2018) was used to summarize the genome statistics. The completeness of the genome assembly was determined with BUSCO v5.3.2 utilizing the fungi\_odb10, Basidiomycota\_odb10 and Agaricomycetes\_odb10 lineage datasets (Manni et al. 2021b). The Funannotate v1.8.15 (<https://github.com/nextgenusfs/funannotate>) with default settings was used to annotate the assembled genome.

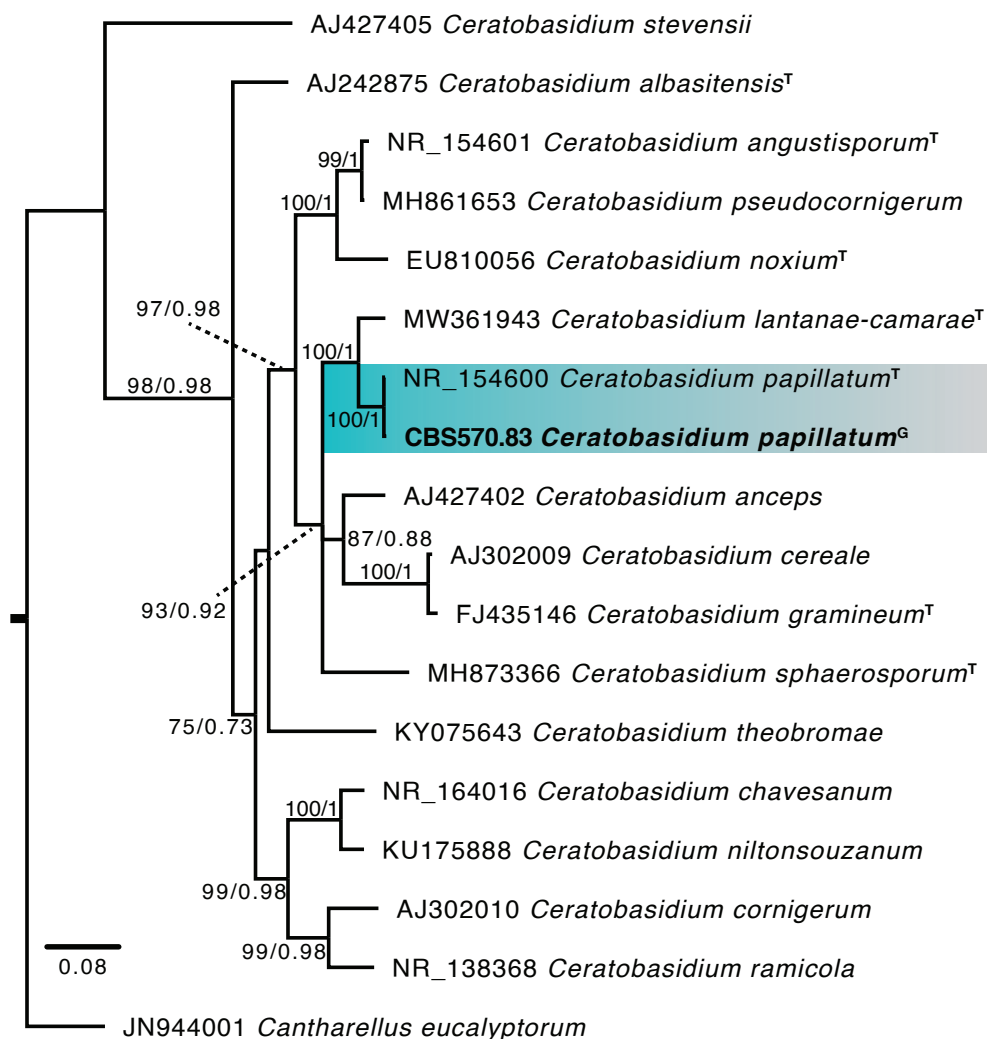
To confirm the identity of the isolate used for genome sequencing, phylogenetic analysis was conducted using maximum likelihood (ML) and Bayesian (BI) approaches. For this, a sequence dataset was prepared that included the complete internal transcribed spacer (ITS) gene region extracted from the assembled genome using CLC Genomics Workbench v23.0.5 (CLC bio, Aarhus, Denmark) and 16 *Ceratobasidium* species retrieved from NCBI GenBank, including the ITS sequence of the same *C. papillatum* isolate deposited by CBS (NR\_154600). *Cantharellus eucalyptorum* (JN944001) was used as an outgroup. This dataset was aligned using MAFFT v7.490 (Rozewicki et al. 2019) using default settings and manually adjusted using Mesquite v3.81 (<https://www.mesquiteproject.org/>). Phylogenetic analysis using ML was performed with IQ-TREE v1.2 (Trifinopoulos et al. 2016) using 1,000 bootstrap replicates. BI analysis was performed using MrBayes v3.2.7 (Ronquist et al. 2012) with four MCMC chains initiated from a random starting tree and run for 5 million generations. The stop value was set at 0.01, the temperature at 0.2 and trees were sampled every 100 generations. A burn-in of 30% of the sampled trees was discarded, and the remainder were used to generate majority rule consensus trees. The phylogenetic trees were viewed and rooted using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

## Results and discussion

The *C. papillatum* genome was 41.00 Mb in size and assembled into 3,091 contigs, of which 2,897 were longer than 1,000 bp. The N50 and N75 values were 23,435 bp and 12,862 bp, respectively, while the L50 and L75 values were 520 and 1,119, respectively. The GC content was 49.01%. Apart from *C. papillatum*, only one genome from a well-described *Ceratobasidium* species, *C. theobromae*, is available in GenBank and Mycocosm. However, genomes from several undescribed *Ceratobasidium* species are also present in these databases. The genome size of *C. papillatum* is 10 Mb larger than that of *C. theobromae*, yet 17–90 Mb smaller than those of the undescribed species. This shows a notable variation in the genome sizes of *Ceratobasidium* species. BUSCO analysis showed completeness of 93.5%, 92.5% and 79.3% using the *Fungi*, *Basidiomycota* and *Agaricomycetes* datasets, respectively. Notably, the number of predicted genes for *C. papillatum* (12,614) was found to be smaller than that of other *Ceratobasidium* species, indicative of the smaller genome size. The phylogenetic position of this isolate, using the genomic copy of the ITS region, confirms that the sequenced genome is *C. papillatum* (Fig. 4). This genome provides a valuable resource for future comparative studies and in-depth analysis of genome structure within the genus. Thus, it will address fundamental questions involving the biology and lifestyle of this group of fungi.

**Authors:** Tiphany Nkomo, Almuth Hammerbacher\*, Tanay Bose,  
Brenda D. Wingfield\*

**\*Contact:** [Brenda.Wingfield@fab.up.ac.za](mailto:Brenda.Wingfield@fab.up.ac.za);  
[Almuth.Hammerbacher@fab.up.ac.za](mailto:Almuth.Hammerbacher@fab.up.ac.za)



**Figure 4.** The maximum likelihood tree constructed using ITS sequences of 16 *Ceratobasidium* species. Type isolates and the sequence extracted from the genome of *Ceratobasidium papillatum* (CBS 570.83) are suffixed with T and G, respectively. *Cantharellus eucalyptorum* served as an outgroup. Branch support values are indicated as bootstrap values/posterior probabilities. Support values below 70 and 0.80 were deemed unreliable and, hence, were excluded from the tree.

## IMA GENOME-F 20C

### Genomic resources for *Colletotrichum spaethianum* and *Colletotrichum fruticola*, causal agents of Anthracnose in *Polygonatum cyrtonema*

#### Introduction

*Polygonatum cyrtonema* Hua. (Duohua Huangjing in Chinese; family Asparagusaceae), is an edible medicinal plant widely used in China to treat diabetes and asthma (Pharmacopeia Committee of P. R. China, 2020). The rhizomes are particularly valued as the primary components in various medicines and food. However, wild *P. cyrtonema* resources cannot meet the increasing demand for health products, such as Huangjing liquor, Huangjing tea, Huangjing noodles, etc., making artificial cultivation a necessity. Unfortunately, many



diseases also appeared as amplified with the development of artificial cultivation of *P. cyrtonema*. Anthracnose is one of the common diseases found on *P. cyrtonema* and has become a serious threat to *P. cyrtonema* production. The leaf and stem symptoms include large elliptical, irregular or long strip spots, with pale to dark brown lesions, in some case intermediate white necrosis (Zhou et al. 2017; Ma et al. 2020; Dan et al. 2023). On the fruit, dark brown necrotic lesions are usually observed. In the major growing regions of Guangze County, Nanping City (northern Fujian Province, China), anthracnose incidence reaches approximately 20% during the flowering stage. As temperature and humidity increase, anthracnose incidence in some areas can rise to over 50% during the fruit stage (Zhou et al. 2017).

*Colletotrichum* is the pathogen responsible for many plant anthracnose and among the most important plant pathogenic fungi worldwide (Dean et al. 2012). For example, *C. spaethianum* is a well-known pathogen of anthracnose in plants belonging to the genus *Polygonatum* (Tomioka et al. 2008; Liu et al. 2020; Ma et al. 2021; Utami and Hiruma 2022). Another member of this genus, *C. fructicola*, has a broad host range, including tobacco, apples, strawberries, orchids, citrus, tea-oil trees, *Paris polyphylla*, etc (Li et al. 2016; Wang et al. 2016; Hu et al. 2019; Silva-Cabral et al. 2019; Liang et al. 2020; Zhou et al. 2020). Thus far, one *C. spaethianum* (MAFF 239500) genome (Utami and Hiruma 2022) and 16 *C. fructicola* genomes have been deposited in NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/>); *C. fructicola* Nara and *C. fructicola* CGMCC3.1737, strawberry anthracnose pathogens (Gan et al. 2013; Armitage et al. 2020); isolates 15060 and 1104-7 of *C. fructicola*, causal agents of mango (Li et al. 2018) and apple (Liang et al. 2018; Liang et al. 2020) anthracnose.

Nanping is an important *P. cyrtonema*-producing region with a total planting area of approximately 1,000 ha. Thus, identifying the *Colletotrichum* species causing *P. cyrtonema* anthracnose in Nanping is a key issue. Anthracnose-diseased samples of *P. cyrtonema* were collected from four cultivation areas in Nanping City, Fujian Province: Chongren, Yushan, Heping, and Huaqiao. Previously, 76 representative *Colletotrichum* strains were isolated from diseased samples based on morphological characteristics and multilocus phylogenetic analysis. 13 species were identified, with two being novel, and then 12 were first reported in *P. cyrtonema*. The dominant causal agents of anthracnose in *P. cyrtonema* were *C. spaethianum* (31 strains, 40.79%) and *C. fructicola* (20 strains, 26.32%). *C. spaethianum* infects leaves, stems, and fruit stalks, while *C. fructicola* infects leaves, stems, and fruits. The former species is more serious leaves, while the latter is more virulent on stems. So far, the genetic differences behind these species are not clear as there are no available genome resource. Here, we report the genome of one isolate of *C. spaethianum* Y1\_DY3\_A (hereafter CsY1), and one isolate of *C. fructicola* C1\_DY2\_B (hereafter CfC1) from *P. cyrtonema*.

## Sequenced strains

*Colletotrichum spaethianum*: **China**: Fujian: Huaqiao, isolated from infected leaves of *P. cyrtonema*, 2021, X. Zhou (Y1\_DY3\_A).

*Colletotrichum fructicola*: **China**: Fujian: Chongren, isolated from infected leaves of *P. cyrtonema*, 2021, X. Zhou (C1\_DY2\_B).

## Nucleotide sequence accession number

All reported genome sequences and genes have been deposited in the Genome Warehouse (GWH, <https://ngdc.cncb.ac.cn/gwh>, accession number GWHEQUI000000000 (CfC1) and GWHEQUR000000000 (CsY1) at the National Genomics Data Center, China National Center for Bioinformation (CNCB-NGDC Members and Partners 2023). Raw sequence reads are publicly accessible at the Genome Sequence Archive (GSA, <https://ngdc.cncb.ac.cn/gsa/>, accession number CRA013791) of CNCB-NGDC under the BioProject PRJCA021822.

## Materials and methods

The isolates of *C. spaethianum* Y1\_DY3\_A (CsY1) and *C. fructicola* C1\_DY2\_B (CfC1) were deposited in the College of Plant Protection, Fujian Agriculture and Forestry University. The two strains were subcultured on potato dextrose agar (PDA) with 12 h light / 12 h dark at 25 °C for 4 days. The mycelia were collected by vacuum filtration and ground in liquid nitrogen. Genomic DNA was extracted using a rapid fungal genomic DNA isolation kit (Sangon Biotech, Shanghai, China). The purified DNAs were sent to Beijing Novogene Bioinformatics Technology Co. Ltd for library preparation and genome sequencing.

Genome size was estimated in GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) with a 21 bp *k*-mer size and a haploid model. Draft genomes were assembled in Flye version 2.4.2 (Kolmogorov et al. 2019), base errors (mainly indels and SNPs) in the draft genomes were corrected with NextPolish version 1.4.0 (Hu et al. 2020) for both PacBio long reads and Illumina short reads. Using BUSCO version 5.3.0 to evaluate genome completeness for both strains (Manni et al. 2021b). We also assessed genome completeness using the mapping rate of sequenced reads. Long and short reads were mapped to the repeat-unmasked final genome assembly in minimap2 version 2.21-r1071 (Li 2018) and BWA-MEM2 version 2.2.1 (Jung and Han 2022), respectively. Assembly quality was further estimated in Merqury version 1.3 (Rhie et al. 2020) (*k* = 17) using PacBio self-corrected reads generated with canu version 2.2 (Koren et al. 2017). Using RepeatMasker version 4.1.2 with the assistance of RepeatModeler version 2.02 (<http://www.repeatmasker.org/>) for repeat sequences *de novo* annotation.

Protein-coding genes were annotated with BRAKER version 2.1.6 (Bruna et al. 2021). EggNOG-mapper v2 (Cantalapiedra et al. 2021) assisted with whole-genome functional annotation in Pfam, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and EuKaryotic Orthologous Groups (KOG). Diamond version 2.0.11.149 (Buchfink et al. 2021) was used for fungal functional annotations. Furthermore, we deployed OrthoVenn2 (Xu et al. 2019) for comparative whole-proteome cluster analyses to understand protein distribution patterns.

## Results and discussion

We obtained 8.51 Gb (CsY1, approximately 152×,  $N_{50}$  = 12.30 kb) and 18.81 Gb (CfC1, approximately 336×,  $N_{50}$  = 17.73 kb) long reads using the PacBio RSII platform, as well as 4.52 Gb (CsY1, approximately 81×) and 8.90 Gb (CfC1, approximately 149×) 150 bp paired-end (PE150) short reads (fragment size ~350 bp) using the Illumina HiSeq 3000 platform. *K*-mer based genome size

estimation yielded expected genome sizes of 55.82 Mb for CfC1 and 52.97 Mb for CsY1. Genome assembling using PacBio long reads followed by 2-round base error correction using both of long reads and Illumina short reads resulted in the final genome assemblies of 56,767,350 bp for strain CfC1 (GC: 53.16%) and 54,491,462 bp for strain CsY1 (GC: 52.42%), representing 99.88% and 97.92% of the estimated genome sizes, respectively (Table 2).

**Table 2.** Genome assembly features of *Colletotrichum spaethianum* strain Y1\_DY3\_A (CsY1) and *Colletotrichum fructicola* strain C1\_DY2\_B (CfC1).

Features	CsY1	CfC1
PacBio long reads (Gb)	8.51	18.81
Illumina short reads (Gb)	4.52	8.90
Estimated genome size (bp)	52,970,789	55,821,760
Assembly size (bp)	54,491,462	56,767,350
Contig number	107	13
Contig N <sub>50</sub> (bp)	1,599,232	4,938,816
Contig N <sub>90</sub> (bp)	319,619	4,303,157
Average contig length (bp)	509,266	7,595,691
Maximum contig length (bp)	4,068,025	7,120,155
GC content (%)	52.42	53.16

The final CfC1 genome contains 13 contigs, with N<sub>50</sub> and N<sub>90</sub> contig lengths of 4,938,816 bp (L50:5) and 4,303,157 bp (L90:9), respectively. Its average contig length is 7,120,155 bp and its maximum contig length is 7,595,691 bp. The final CsY1 genome contains 107 contigs, with N<sub>50</sub> and N<sub>90</sub> contig lengths of 1,599,232 bp (L50:12) and 319,619 bp (L90:46), respectively. Its average and maximum contig lengths are 509,266 bp and 4,068,025 bp (Table 2).

The assembled CfC1 genome contains 747 (98.55%) complete (744 single, three duplicated) and one fragmented orthologs when compared against the fungi\_odb10 lineage (n = 758), as well as 1,666 (97.66%) complete (1,660 single, six duplicated) and two fragmented orthologs compared against the ascomycota\_odb10 lineage (n = 1,706). The assembled CsY1 genome contains 740 (97.63%) complete (733 single, seven duplicated) and one fragmented orthologs compared against the fungi\_odb10 lineage (n = 758), as well as 1,667 (97.71%) complete (1,653 single, 14 duplicated) and two fragmented orthologs compared against the ascomycota\_odb10 lineage (n = 1,706) (Table 3).

Mapping rates of long reads were 99.36% (1,276,006 / 1,284,226 reads) for CfC1 and 99.19% (1,012,391 / 1,020,624 reads) for CsY1; for short reads, they were 99.92% (59,315,969 / 59,362,650 reads, properly paired: 99.02%) for CfC1 and 99.36% (29,999,011 / 30,191,513 reads, properly paired: 97.96%) for CsY1. Assembly quality was further estimated in Merqury revealed average base error rates of  $2.09 \times 10^{-5}$  (QV = 46.81) and  $2.31 \times 10^{-4}$  (QV = 36.36) for CfC1 and CsY1, respectively, along with 99.88% and 97.92% genome completeness (Table 3).

We then *de novo* annotated and masked 1,983,026 bp (3.49% of genome size) and 4,111,949 bp (7.55%) of repeat sequences for CfC1 and CsY1, respectively. In CfC1 repeats, 1,372,796 bp (69.23%) were interspersed repeats, including short interspersed nuclear elements (SINEs, 5,009 bp), long interspersed nuclear elements (LINEs, 242,286 bp), long terminal repeats (LTRs, 241,689 bp),

**Table 3.** Completeness and accuracy of the genome assemblies of the two *Colletotrichum* species isolated from *Polygonatum cyrtonema*.

Methods		CsY1		CfC1	
BUSCO	Database	<i>Fungi</i> *	<i>Ascomycota</i> *	<i>Fungi</i>	<i>Ascomycota</i>
	Completeness	98.55%	97.66%	97.63%	97.71%
	Single copy BUSCOs	744	1,660	733	1,653
	Duplicated BUSCOs	3	6	7	14
	Fragmented BUSCOs	1	2	1	9
	Missing BUSCOs	8	34	9	30
Mapping rates	PacBio long reads	99.19%		99.36%	
	Illumina short reads	99.36% (97.96% Properly paired)		99.92% (99.02% Properly paired)	
Merquy	Average base error	2.31 × 10 <sup>-4</sup> (QV = 36.36)		2.09 × 10 <sup>-5</sup> (QV = 46.81)	
	Genome completeness	97.92%		99.88%	

\*: *Fungi* for fungi\_odb10 (n = 758), *Ascomycota* for ascomycota\_odb10 (n = 1,706).

DNA transposons (491,988 bp), and Unclassified (391,824 bp). In CsY1 repeats, 3,483,650 bp (84.72%) were interspersed, including SINEs (3,657 bp), LINEs (125,572 bp), LTRs (2,405,105 bp), DNA transposons (148,291 bp) and Unclassified (801,025 bp) (Table 4).

**Table 4.** Repetitive sequence analysis of the genome of the two *Colletotrichum* species isolated from *Polygonatum cyrtonema*.

Repetitive sequence	Features	<i>C. spaethianum</i> Y1_DY3_A		<i>C. fructicola</i> C1_DY2_B	
		Length (bp)	Percentage%	Length (bp)	Percentage%
Interspersed repeats	SINEs	3,657	0.01	5,009	0.01
	LINEs	125,572	0.23	242,286	0.43
	LTR elements	2,405,105	4.41	241,689	0.43
	DNA transposons	148,291	0.27	491,988	0.87
	Unclassified	801,025	1.47	391,824	0.69
Tandem repeats	Small RNA	14,070	0.03	8,487	0.01
	Satellites	0	0.00	14,968	0.03
	Simple repeats	556,157	1.02	531,614	0.94
	Low complexity	59,823	0.11	58,036	0.10
	Total repeats	4,111,949	7.55	1,983,026	3.49

We identified 16,928 and 15,753 protein-coding genes in CfC1 and CsY1, respectively, 11,547 (68.21%, CfC1) and 11,456 (72.72%, CsY1) genes in Pfam, 6,837 (40.39%, CfC1) and 6,778 (43.03%, CsY1) genes in GO, 5,121 (30.25%, CfC1) and 6,428 (40.80%, CsY1) in KEGG, as well as 12,378 (73.12%, CfC1) and 12,680 (80.49%, CsY1) in KOG. The Pfam annotations included a set of candidate fungal pathogenicity-related genes, with CfC1 possessing more than CsY1 (Table 5). 3,531 and 3,616 pathogen-host interaction genes, 253 and 279 carbohydrate-active enzymes, 337 and 481 membrane transport proteins, 1,299 and 885 putative secreted proteins, 677 and 395 effectors, 90 and 59 secondary metabolite biosynthetic gene in CfC1 and CsY1, respectively (Table 5).

**Table 5.** Gene functional annotation of the genome of the two *Colletotrichum* species isolated from *Polygonatum cyrtoneura*.

Annotation	<i>C. spaethianum</i> Y1_DY3_A		<i>C. fructicola</i> C1_DY2_B	
	Gene numbers	Percentage%	Gene numbers	Percentage%
Protein-coding genes	15,753	100.00	16,928	100.00
Pfam	11,456	72.72	11,547	68.21
GO	6,778	43.03	6,837	40.39
KEGG	6,428	40.80	5,121	30.25
KOG	12,680	80.49	12,378	73.12
CAZys	279	1.77	253	1.49
PHIs	3,616	22.95	3,531	20.86
Cytochrome P450 enzymes	158	1.00	279	1.65
Membrane transport proteins	481	3.05	337	1.99
Putative secreted proteins	885	5.62	1,299	7.67
Effectors	395	2.51	677	4.00
SMBGCs <sup>#</sup>	59	/	90	/

<sup>#</sup>: SMBGCs: secondary metabolite biosynthetic gene clusters.

Phylogenetic analysis confirmed the identity of the two sequenced isolates (Fig. 5). The isolates of *C. spaethianum* Y1\_DY3\_A (CsY1) and *C. fructicola* C1\_DY2\_B (CfC1) were assigned to the respective species. We compared the two sequenced genomes with those from publicly available species in the same genus (the same genomes used for annotating protein-coding genes above). The 17,410 and 15,888 proteins in CfC1 and CsY1, respectively, formed 15,618 and 12,120 clusters. These included 17,899 non-redundant clusters which contained 7,783 single-copy clusters that were present in all *Colletotrichum* species and two species-specific clusters. Strains CfC1 and CsY1 shared more gene clusters with *C. fructicola* and *C. spaethianum*, respectively (Fig. 6), reflecting their closer evolutionary relationships. In conclusion, the high-quality genomes and annotated gene resources that we provided in this study will benefit future research on the infection and pathogenicity mechanisms of *C. fructicola* and *C. spaethianum*.

**Authors:** Xianzhi Zhou\*, Yuan Wu, Yanping Hu, Hao Yu, Jianjin Zhou, Hongli Hu\*

**\*Contact:** [xianzhizhou@126.com](mailto:xianzhizhou@126.com); [huhongli7905@gmail.com](mailto:huhongli7905@gmail.com)

## IMA GENOME-F 20D

### Draft genome sequence of *Neopestalotiopsis macadamiae* from macadamia in South Africa

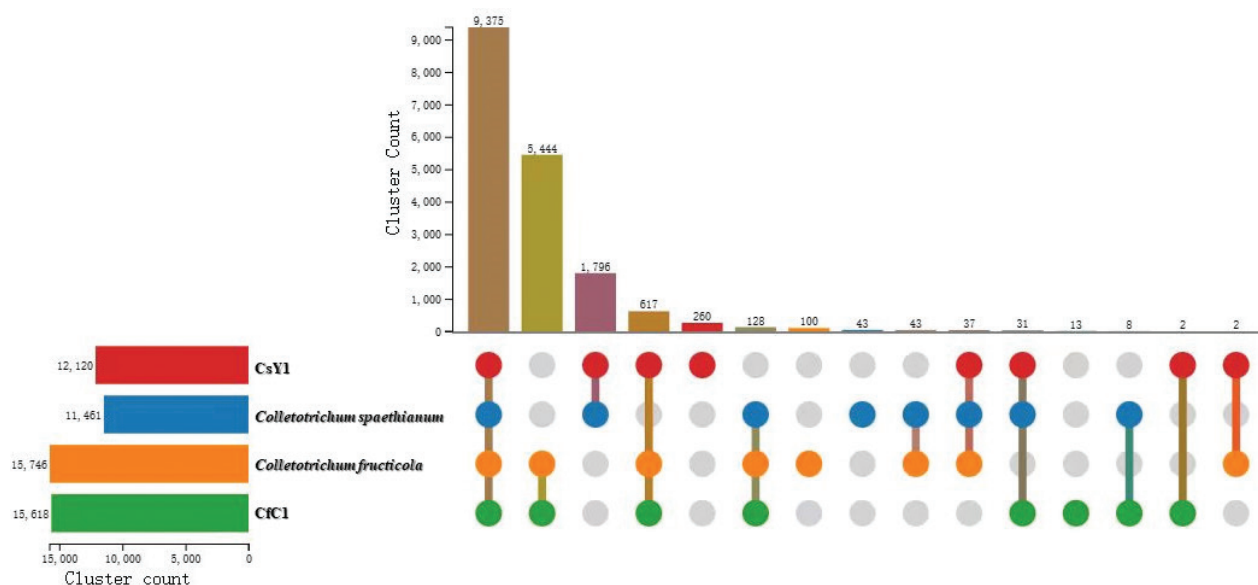
#### Introduction

*Pestalotiopsis*, a genus in the *Sordariomycetes* was reclassified in 2014 into 3 genera (*Pestalotiopsis*, *Neopestalotiopsis* and *Pseudopestalotiopsis*) based on morphology and phylogenetic analysis using the  $\beta$ -tubulin ( $\beta$ -tub), Internal Transcribed Spacer (ITS), and Translation Elongation Factor 1- $\alpha$  (tef1) gene regions (Maharachchikumbura et al. 2014). Since then, several *Neopestalotiopsis* species have been described and classified as plant pathogens and are associated with numerous diseases in various hosts, such as causing stem girdling and dieback on *Eucalyptus* (Diogo et al. 2021), grey leaf spot disease on *Mangifera indica* (mango) (Gerardo-Lugo et al. 2020), grey blight leaf disease on





**Figure 5.** Bayesian inference phylogenetic tree of *Colletotrichum* sections *gloeosporioides* and *spaethianum* indicating the two sequenced isolates (in blue). *Monilochaetes infuscans* (CBS 869.96) is used as outgroup. The tree was built using concatenated sequences of the ACT, CHS-1, GAPDH, HIS3, ITS and TUB2. Maximum likelihood bootstrap values  $\geq 50\%$ . Bayesian posterior probabilities  $\geq 0.90$  (PP /MLBS) are displayed on the phylogenetic tree.



**Figure 6.** Protein level homologous gene cluster analysis of the genome of the two *Colletotrichum* species isolated from *Polygonatum cyrtonema*.

*Camellia sinensis* (tea) (Chen et al. 2018; Wang et al. 2022) and leaf blight and crown rot on strawberry on *Fragaria × ananassa* (strawberry) (Rebollar-Alviter et al. 2020; Wu et al. 2020).

*Neopestalotiopsis macadamiae* represents an important pathogen of macadamia (*Macadamia integrifolia* and *M. tetraphylla*) grown in commercial plantations in tropical and subtropical frost-free regions in various countries around the world, including South Africa (Akinsanmi et al. 2017). This is because *N. macadamiae* causes dry flower disease (*Pestalotiopsis* blight) that results in yield loss by causing necrotic blight symptoms of the inflorescences and a dry appearance of the raceme (Akinsanmi et al. 2017). This pathogen causes yield loss and is regarded as one the most dominant and economically important threats to the macadamia industry (Akinsanmi et al. 2017).

Genome data are a valuable tool to understand the different aspects that can contribute towards developing control strategies for this pathogen. Currently, the pathogen is poorly understood and further research to understand various characteristics such as its plant host interactions, pathogenesis-related gene repertoire, and reproduction is required. *Neopestalotiopsis rosae* causes leaf blight and crown rot on strawberry and is the only publicly available genome for this important genus (Hsu et al. 2022; Han et al. 2024). The aim of this study was to generate the first draft genome assembly of a *N. macadamiae* strain, a confirmed pathogen of macadamia inflorescences through pathogenicity trials (Botha 2024), from South Africa. The pathogenic nature of several *Neopestalotiopsis* species makes it crucial to assemble high-quality reference genomes, which will provide a critical foundation for advancing research into the biology and pathogenic mechanisms of *Neopestalotiopsis* species.

## Sequenced strain

**South Africa:** Barberton: Mpumalanga, isolated from *Macadamia integrifolia* × *M. tetraphylla* inflorescences, 2022, B. Sonnekus (CMW 64591)

## Nucleotide sequence accession number

The Whole Genome Shotgun project of *Neopestalotiopsis macadamiae* CMW 64591 has been deposited at DBJ/EMBL/GenBank under the accession JBEX-AB000000000. The version described in this paper is version JBEXAB000000000.

## Methods

The culture of isolate CMW 64591 was grown on MEA (malt extract agar; Biolab, South Africa) at 25 °C and 90% relative humidity with a 12-h photoperiod for 10 days. Genomic DNA was extracted from mycelia using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, California, USA) and submitted to Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) for whole genome sequencing using PacBio® HiFi Sequel IIe Sequencer. The quality of the sequenced long reads was assessed using FastQC v0.11.7 (Babraham Bioinformatics, Babraham Institute, Cambridge, UK). Canu v2.0 (Koren et al. 2017) was used for the *de novo* assembly and SeqKit v0.10.1 (Shen et al. 2016) was used to filter contigs smaller than 500 bp (Shen et al. 2016). Minimap2 (Li 2018) and SAMtools v. 1.18 (Li et al. 2009) were used to calculate the average base coverage by mapping the reads back to the genome. Genome statistics were determined using Quast v. 5.0.2 (Gurevich et al. 2013). The genome completeness was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) in combination with the fungal data dataset (Simão et al. 2015). vRepeatModeler v2.0.1 (Flynn et al. 2020) was used to create a custom library of repeat families for the assembly to soft mask the assembly with RepeatMasker open-4.0.7 (Smit et al. 2013).

Genes were predicted in the masked assembly with Funannotate predict, which uses Augustus v3.5.0 (Stanke et al. 2006b), GeneMark-ES (Ter-Hovhannissyan et al. 2008), Glimmer v3.0.4 (Delcher et al. 2007), and SNAP (Korf 2004) as *ab initio* predictors. It also incorporates protein evidence in the form of BLAST alignments from the UniProt database (The UniProt Consortium 2023) and predicts tRNA genes with tRNAscan-SE v2.0.12 (Lowe and Eddy 1997). Functional annotations were assigned to the predicted genes by comparing them against multiple annotation databases. These included InterProScan v5.52–86.0 (Jones et al. 2014), EggNOG-mapper v2.1.11 (Cantalapiedra et al. 2021), antiSMASH v6.1.1 (Blin et al. 2021) and Phobius v1.01 (Käll et al. 2007). These annotations were merged into a single file by running the Funannotate annotate pipeline, which additionally compared predicted proteins against the dbCAN v11.0 (Yin et al. 2012) database of carbohydrate-active enzymes (CAZymes) and the MEROPS v12.0 (Rawlings et al. 2014) protease database. To confirm the identity of the isolate, the  $\beta$ -tub, ITS and tef1 gene regions were extracted from the genome assembly. These sequences were included in a maximum likelihood (ML) phylogeny with closely related *Neopestalotiopsis* species obtained from GenBank (NCBI) using *Pestalotiopsis diversiseta* MFLUCC12-0287 as an outgroup. The sequence datasets were aligned using an online version of MAFFT v. 7 (Katoh et al. 2019) and concatenated in Geneious Prime (Kearse et al. 2012). Phylogenetic analysis using ML was performed with RaxML v. 8 (Stamatakis 2014), based on the GTR substitution model with gamma-distribution rate variation and applying 1,000 bootstrap replicates. MITOS was used to identify and annotate the mitochondrial genome (Bernt et al. 2013).

## Results and discussion

The assembled genome size of *N. macadamiae* CMW 64591 was 50.05 Mb (for 0.5 kb+ scaffolds), consisting of 10 contigs and an approximated genome coverage of 22×. At 50.05 Mb, the *N. macadamiae* CMW 64591 genome is slightly smaller than the *N. rosae* ML1664 genome (53.78 Mb) (Hsu et al. 2022). However, the G/C content (50.22%) and the number of contigs (10) of *N. macadamiae* CMW 64591 is similar to the G/C content (49.88%) and the number of contigs (18) of *N. rosae* ML1664 genome.

The BUSCO analysis showed a completeness level of 99.4% with respect to the fungal dataset. Slightly more than the 98.4% BUSCO score observed for the *N. rosae* ML1664 genome (Hsu et al. 2022). In total, 454 ORFs occurred as single copies, 1 was duplicated, 0 were fragmented and 4 were missing from the *N. macadamiae* genome assembly. The Quast analysis indicated that the largest contig was 9,876,848 bp in length, a N50 of 7,068,772 bp, a N75 of 6,415,834 bp, a L50 score of 4 and L75 score 5 were observed. A total of 1.28% of the genome consisted of masked repeats. Contig 8 was identified as the mitochondrial genome of the *N. macadamiae* CMW 64591 assembly.

A total of 14,221 genes was predicted of which 14,012 were functionally annotated. This included 829 CAZymes of which the AA7 and AA3 CAZyme families, which are involved in lignin degradation in plants, were most abundant. ML analysis of three barcoding gene regions grouped the genome isolate close to *N. macadamiae* isolates obtained from Australia (Fig. 7). The phylogenetic analysis also revealed species delineation issues and an increased availability of *Neopestalotiopsis* genomes could facilitate use of more informative gene regions to use for improved phylogenetic resolution. The draft genome sequence of *N. macadamiae* generated here will facilitate future research regarding the biology and pathogenicity of this fungus. Genome data will also facilitate improved understanding of secondary metabolite production, mating strategies, and the ecological interactions between *Neopestalotiopsis* species, and their plant hosts.

**Authors:** Byron Sonnekus\*, Magriet A. van der Nest, Brenda D. Wingfield, Emma T. Steenkamp, Nicky Creux, Gerda Fourie

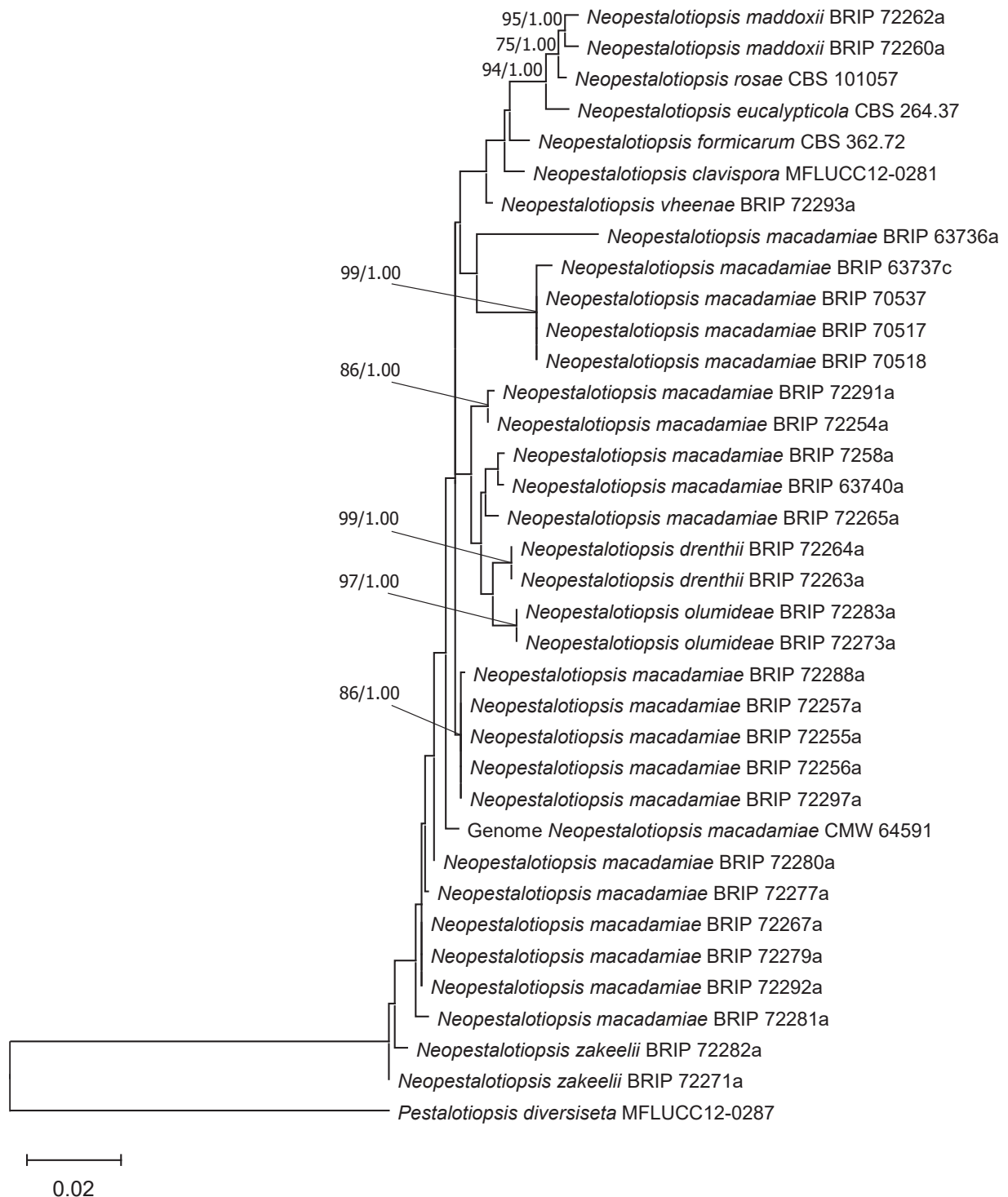
**\*Contact:** [Byron.Sonnekus@fabi.up.ac.za](mailto:Byron.Sonnekus@fabi.up.ac.za)

## IMA GENOME-F 20E

### The first draft genome assembly of *Sphaerellopsis filum*, a common mycoparasite of rust fungi

#### Introduction

The genus *Sphaerellopsis* belongs to the family *Leptosphaeriaceae* in the order *Pleosporales*. *Eudarluca*, the typified name for the sexual morph of this genus is treated as a synonym (Trakunyingcharoen et al. 2014) and *Sphaerellopsis* is preferred over *Eudarluca* (Rossman et al. 2015). Eleven *Sphaerellopsis* species names are listed as legitimate in MycoBank; amongst these, seven species are supported with DNA barcode sequences in NCBI GenBank. The type species, *S. filum*, is a common mycoparasite of many rust fungi (Yuan et al. 1998, 1999; Kiss 2001; Driessen et al. 2004; Trakunyingcharoen et al. 2014; Sun et al. 2019).



**Figure 7.** Maximum-likelihood (ML) tree topology of *Neopestalotiopsis* inferred on an alignment of the concatenated sequences of the  $\beta$ -tub, ITS and *tef1* gene regions. *Pestalotiopsis diversiseta* (MFLUCC 12-0287) was used as an outgroup taxon. The ML tree was based on the GTR substitution model with gamma-distribution rate variation. ML bootstrap support values (>75%) and Bayesian inference posterior probabilities (>90%) are displayed at the nodes.

*Sphaerellopsis anomala*, *S. macroconidialis*, *S. melampsorinearum*, and *S. paraphysata* were also described as being associated with pustules of different rust species (Nag Raj 1993; Trakunyingcharoen et al. 2014; Gómez-Zapata et al. 2024). The conidiomata of *S. hakeae* were found inside rust pustules, but also occurring separately without any association to the rust (Crous et al. 2016). Two



recently recognised species, *S. isthmospora* and *S. artemisiae*, were described as saprobes on dead plant materials (Phookamsak et al. 2019; Doilom et al. 2021).

*Sphaerellopsis filum* is commonly found in uredinia and telia of diverse rust species in the field (Yuan et al. 1998, 1999; Driessen et al. 2004; Pei et al. 2010; Gordon and Pfender 2012) and some strains have been evaluated as potential biocontrol agents of some rusts (Pei and Yuan 2005; Durgadevi et al. 2024; Gómez-Zapata et al. 2024). Despite this, little is known about the ultrastructural aspects of this specific interfungal parasitic relationship (Plachecka 2005; Gómez-Zapata et al. 2024). Experiments revealed important differences in the mycohost specialisation and virulence of different *S. filum* strains (Yuan et al. 1999; Nischwitz et al. 2005; Pei et al. 2010). Phylogenetic analyses and molecular diversity studies indicated that *S. filum* strains are genetically diverse (Lieseback and Zaspel 2004; Nischwitz et al. 2005; Bayon et al. 2006).

Strikingly, *S. filum* is closely related to *Ampelomyces* spp. (Nischwitz et al. 2005; Trakunyingcharoen et al. 2014) that are widespread mycoparasites of powdery mildews (Kiss et al. 2011; Németh et al. 2021). This appears to be a rare example of the convergent evolution of mycoparasitism in close relatives. How this evolution has occurred and its genetic basis is unknown, but comparative genomics may provide insights. Genomic resources are already available for *Ampelomyces* (Haridas et al. 2020; Huth et al. 2021). However, as of 29 August 2024, there are no genome assemblies for any *Sphaerellopsis* spp. in GenBank. The aim of this study was, therefore, to provide the first draft genome assembly for *S. filum*. This resource will serve as the starting point to decipher the unique mode of action of *S. filum* against rust fungi, and the evolution of specific mycohost-mycoparasite interactions at the molecular level.

## Sequenced strain

*Sphaerellopsis filum*: Portugal: isolated from *Puccinia hordei* on *Ornithogalum divergens*, 1951, B. d'Oliveira (CBS 235.51 = ATCC 22604).

## Nucleotide sequenced accession number

The annotated genome sequence of *Sphaerellopsis filum* CBS 235.51 has been deposited at GenBank under the accession JBFTXA000000000.

## Materials and methods

*Sphaerellopsis filum* strain CBS 235.51 was obtained from the Westerdijk Fungal Biodiversity Institute (Utrecht, The Netherlands) and was maintained on potato dextrose agar (Amyl Media, Australia) at 18 °C in darkness. For DNA extraction, the strain was cultured in 20 ml Czapek-Dox broth (BD, Bacto Laboratories, Australia) in a Falcon tube and incubated on a shaker-incubator at 150 rpm for 2 weeks at 18 °C. Harvested mycelia were flash-frozen in liquid nitrogen and lyophilized for 48 hours. Eighty milligrams of lyophilized mycelia were ground with stainless steel beads (2.8 mm diameter, Sigma–Aldrich, Australia) in a Qiagen TissueLyser II (Qiagen, Australia) at 30 Hz/s for 30 s, flash-frozen in liquid nitrogen again and stored at -80 °C until DNA extraction. DNA was extracted using a Qiagen DNeasy Plant Mini Kit (Qiagen, Australia) following a modified protocol as described by

Vaghefi et al. (2024). The internal transcribed spacers and the intervening 5.8S region were amplified using ITS1 and ITS4 primers (White et al. 1990), and submitted to Macrogen (Seoul, South Korea) for Sanger sequencing. Sequences were manually edited and consensus sequences were generated in Geneious Prime® 2024.0.5 (Biomatters Inc., New Zealand). Reference *Sphaerellopsis* sequences were downloaded from GenBank from the studies of Trakunyingcharoen et al. (2014), Crous et al. (2016), Phookamsak et al. (2019), Doilom et al. (2021), and Gómez-Zapata et al. (2024). A maximum likelihood phylogeny was constructed using RAxML v.8 (Stamatakis 2014) using a GTR GAMMA substitution model, and visualized in FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

The same DNA sample used for the ITS amplification was used for Illumina library construction and Illumina NovaSeq sequencing (150 bp paired-end [PE]) at the Australian Genome Research Facility (AGRF, Melbourne, Australia).

Total RNA was extracted from two-week-old cultures grown in Czapek-Dox broth and incubated at room temperature on a shaker at 150 rpm. Fresh fungal mycelia were flash-frozen and ground in liquid nitrogen, then used for RNA extraction with an RNeasy Plant Mini Kit (Qiagen, Australia) following manufacturer's instructions and including column-based DNase treatment. The extracted RNA was assessed via agarose gel electrophoresis and quantified using a Qubit v.4.0 fluorometer (ThermoFisher Scientific, Australia). The RNA transcriptome was sequenced by the AGRF (150 bp PE).

Raw sequence data were screened and filtered for contaminating sequences using Kraken v.2.1.2 (Wood et al. 2019) using the NCBI's RefSeq archaea, viral, and UniVec\_Core databases. Adapter removal from Illumina reads and quality trimming was conducted using BBduk from the BBmap suite v.39.03 (Bushnell 2014) (ktrim=r k=23 mink=11 hdist=1 tpe tbo qtrim=rl trimq=15). Genome assembly and error correction was conducted in SPAdes (Prjibelski et al. 2020) using read-correction and auto-*k*-mer selection options. The ITS sequence was extracted from the assembled genome by constructing a BLAST database in Geneious Prime and using the ITS sequence of the ex-neotype culture of *S. filum* (GenBank [KP170657](#)) as query against the genome. Assembly completeness was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.5.0 (Manni et al. 2021a) against the *Dothideomycetes* database (dothideomycetes\_odb10). For genome annotation, repeat content was determined using RepeatModeler v.2.0.1 (Flynn et al. 2020) and soft-masked using RepeatMasker using RepeatMasker v.4.1.0 (Smit et al. 1996–2010). Evidence-based genome annotation was conducted using BRAKER v.3 (Stanke et al. 2006a, 2008) using RNASeq data as evidence (Gabriel et al. 2021). Raw genome data were also used for genome size estimation in GenomeScope (Ranallo-Benavidez et al. 2020) Galaxy version 2.0. GC content distribution was determined using the software OcculterCut v.1.1 (Testa et al. 2016).

## Results and discussion

Using Illumina NovaSeq technology, 34,834,531 paired-end reads (10.5 Gb) were generated, 33,974,265 of which passed contamination and quality filtering (10 Gb). GenomeScope estimated a genome size of 28.9 Mb based on these DNA sequence data. *De novo* genome assembly of the *S. filum* strain included a total of 850 scaffolds larger than 250 bp and a draft genome size of 28,451,751 with 300× coverage (Table 6). Genome completeness based on

**Table 6.** Genome statistics and comparison with *Ampelomyces* species.

Strain	Host species	Sequencing technology	Assembly size (Mb)	Cov <sup>a</sup>	No. of scaffolds >500 bp	N50 (Mb)	GC content (%)	Genome Completeness (%) <sup>b</sup>	Percentage of AT-rich regions <sup>c</sup>	NCBI Accession Number	Reference
CBS 235.51	<i>Puccinia hordei</i>	Illumina paired-end NovaSeq	28.5	300×	330	375,353	49.7	95.5	10.1	JBFTXA000000000	This study
BRIP 72107	<i>Golovinomyces bolayi</i>	Illumina paired-end MiSeq and Oxford Nanopore MinION	40.4	400×	24	2,994,887	45.5	96.6	33.9	JAGTXZ000000000	Huth et al. (2021)
HMLAC 05119	Undetermined powdery mildew	Illumina paired-end and mate-pair	36.8	103×	464	258,565	46.5	96.3	26.6	VOSX000000000.1	Haridas et al. (2020)

<sup>a</sup> Genome coverage.<sup>b</sup> Assembly completeness for all the genomes was determined based on Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.5.0 (Manni et al. 2021a) against the dothideomycetes\_odb10 database.<sup>c</sup> Denotes the percentage of repeat-rich regions with reduced GC content, estimated by OcculterCut (Testa et al. 2021).

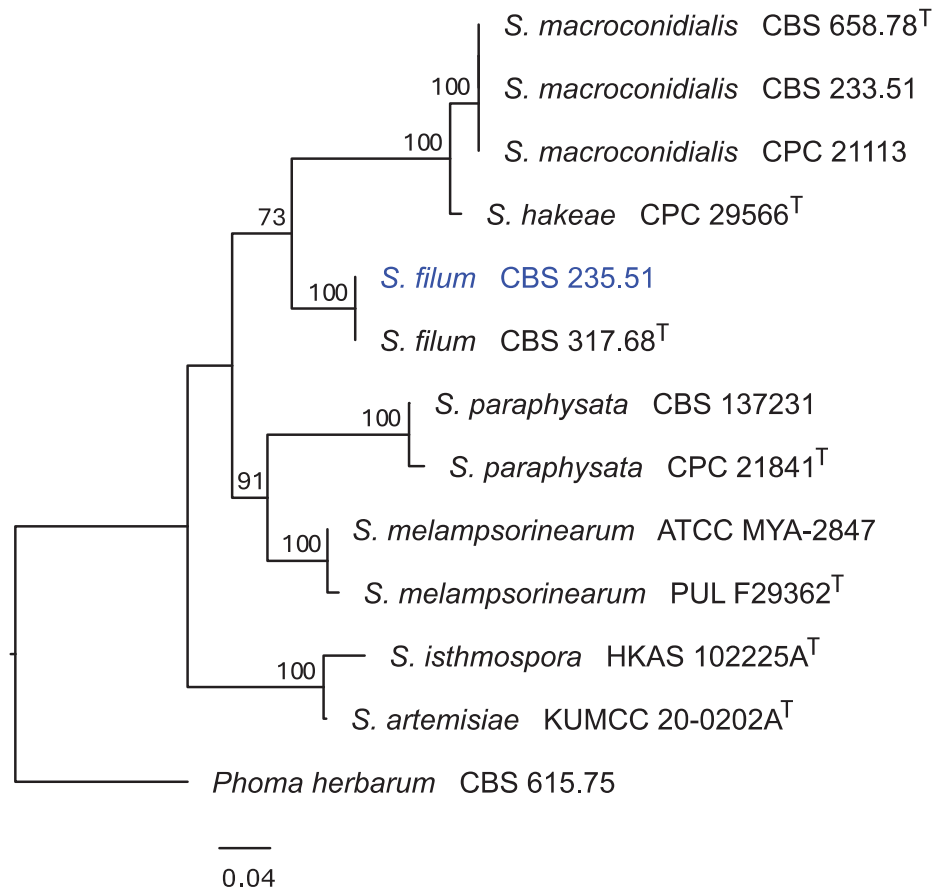
Benchmarking Universal Single-Copy Orthologs in *Dothideomycetes* was estimated to be 95.5%, corresponding to 95.4% complete and single-copy BUSCOs, 0.1% complete and duplicated BUSCOs, 0.4% fragmented and 4.1% missing BUSCOs. Transcriptome sequencing generated 30,197,914 sequences (9.1 Gb), 99% of which passed quality control. A total of 9,626 genes including 11,671 coding sequences (CDS) were predicted in the *S. filum* CBS 235.51 assembly.

The ITS sequence of CBS 235.51 produced using Sanger sequencing (GenBank accession [PP972725](#)) was identical to the ITS sequence of CBS 317.68 (GenBank [KP170657](#)), the ex-neotype culture of *S. filum*. The ITS sequence extracted from the Illumina genome sequencing of CBS 235.51 was identical to this sequence. The phylogenetic placement of strain CBS 235.51 based on the ITS sequence is provided in Fig. 8. This confirmed the identity of the genome-sequenced isolate as *S. filum*.

This is the first genome assembly of a *Sphaerellopsis* strain, which revealed a smaller genome size compared to *Ampelomyces* spp., another group of mycoparasitic fungi within *Pleosporales* (Table 6). Analysis of the assembled genome in Occultercut revealed its bipartite structure, similar to *Ampelomyces* species (Huth et al. 2021). The percentage of AT-rich regions in the assembled genome of CBS 235.51 (10%) is, however, lower compared to *Ampelomyces* species (Huth et al. 2021; Table 6). These AT-rich regions, which are also prevalent in many plant pathogenic fungi, are believed to be the result of repeat-induced point mutations, which is a defence mechanism against propagation of transposable elements in fungi (Testa et al. 2021). This could be the underlying reason for the fragmentation of the assembled genome and indicates a need for long-read sequence data to further sequence and assemble these repetitive regions. In the meantime, the availability of the first genomic resources for this important and widespread mycoparasitic species will facilitate research focussed on their evolution, ecology, and mycoparasitic activity.

**Authors:** Niloofar Vaghefi\*, Jason Risteski, Alexander Idnurm, Levente Kiss

**\*Contact:** [vaghefin@unimelb.edu.au](mailto:vaghefin@unimelb.edu.au)



**Figure 8.** Phylogenetic tree of *Sphaerellopsis* species with available sequence data, based on maximum likelihood analysis of the ITS region. Analyses were performed in Geneious Prime v.2024.0.5 using RAXML v. 8.2.11 (Stamatakis 2014) based on the GTR substitution model with gamma-distribution rate variation. Bootstrap support values (from 1,000 reiterations) greater than 70% are given at the nodes. The tree was rooted to *Phoma herbarum* CBS 615.75. The isolate sequenced in this study is shown in blue and ex-type strains are indicated with T.

## IMA GENOME-F 20F

### Genome sequence of *Pyrenopeziza brassicae*, causal organism of light leaf spot disease on *Brassica* species

#### Introduction

*Pyrenopeziza brassicae* Sutton and Rawlinson is an air-borne, extracellular fungal pathogen of *Brassica* species. It has been taxonomically accommodated within the family *Ploettnerulaceae* of the order *Helotiales* within the phylum *Ascomycota*. This pathogen was also known by its anamorph, *Cylindrosporium concentricum*, for many years; this was initially described by Greville (1823) considering the concentric ring-like pattern of the asexual sporulation. Ascospores (sexual sporulating structures) of *P. brassicae* were first observed on culture media by Thomson (1936) and Cabral (1940), and later described as immature apothecia by Hickman et al. (1955). The occurrence of *P. brassicae* apothecia under natural conditions was first reported by Staunton and Kavanagh in 1966, based on their observations in diseased vegetable brassica crops

in Ireland. The precise identification of *P. brassicae* was made by Rawlinson et al. (1978), who described it as the teleomorph of *Cylindrosporium concentricum* (Rawlinson et al. 1978; Cheah et al. 1980).

The pathogen thrives in areas with cool, wet climates and shows a wide geographic distribution, occurring in many regions of the world including the UK, continental Europe, Japan, New Zealand and North America (Boys et al. 2007; Karolewski et al. 2012; Karandeni Dewage et al. 2018; Carmody et al. 2020). In the UK, light leaf spot disease caused by *P. brassicae* is considered as one of the most economically damaging diseases of winter oilseed rape (*Brassica napus*) with yield losses potentially up to 30–50% (Rawlinson et al. 1978; AHDB 2021). The steady increase in the severity of light leaf spot epidemics since the early 2000s in England, with it replacing phoma stem canker as the main disease of winter oilseed rape there, has increased the economic importance of *P. brassicae*. There have been severe light leaf spot epidemics affecting oilseed rape crops in other parts of northern continental Europe and Poland and vegetable brassica crops in New Zealand (Boys et al. 2007; Karandeni Dewage et al. 2018; Bucur et al. 2022).

Thorough understanding of the underpinning molecular mechanisms and the genetic basis of host-pathogen interactions is a key to successful management of light leaf spot. However, there is little information available on the genes involved in *P. brassicae* pathogenicity, and no published whole genome sequence of *P. brassicae* is available at present. Also, a genome sequence can provide new tools to study the population structure of this pathogen, since there may be variations in pathogenicity between isolates from different geographic regions (Karandeni Dewage et al. 2018). According to Carmody et al. (2020), *P. brassicae* from North America is considered as a different lineage from those present in other geographic regions. The genome sequence of this pathogen is an invaluable resource to investigate these taxonomic/evolutionary identities. There will also be new possibilities for comparative genomic studies since *P. brassicae* is known to have a close evolutionary relationship with eyespot pathogens on cereals (*Oculimacula yallundae* and *O. acuformis*) and leaf blotch (also called leaf scald) pathogen *Rhynchosporium commune* (formerly known as *R. secalis*), on barley (Goodwin 2002).

Considering its economic and evolutionary importance, we present the *de novo* whole genome sequence of *P. brassicae*. This will enable new studies of the biology and genetics of this pathogen, providing knowledge to improve current light leaf spot management strategies and address evolutionary questions.

## Sequenced isolate

*P. brassicae* isolate 15WOSR64-SS1 was obtained from diseased leaves of oilseed rape (*Brassica napus*) cultivar Bristol from Hereford, UK in 2015 (Karandeni Dewage et al. 2021). The mating type of this isolate was confirmed to be *MAT1-1* by mating-type PCR (Foster et al. 2002).

## Nucleotide sequence and accession number

The whole genome sequence of *P. brassicae* (isolate 15WOSR64-SS1) has been deposited at the European Nucleotide Archive (ENA) under the accession number GCA\_958299125.



## Materials and methods

A single conidial isolate (15WOSR64-SS1) of *P. brassicae* (Karandeni Dewage et al. 2021) was grown in potato dextrose broth (Oxoid Ltd., England) at 18 °C in a shaking incubator at 120 rpm. Mycelia were harvested from three-week-old cultures and freeze-dried. Genomic DNA was extracted from freeze-dried mycelia using the CTAB DNA extraction method for high molecular weight genomic DNA (Inglis et al. 2018). Purity of the extracted DNA was tested with a Nanodrop ND-1000 spectrophotometer (Labtech International, UK) and the DNA concentration was measured using a Qubit™ 3.0 fluorometer (Fisher Scientific, UK). Quality of the extracted DNA was assessed using a genomic DNA ScreenTape assay with Agilent 4150 TapeStation (Agilent Technologies, UK) to estimate the integrity of the DNA, before sending the samples to GENEWIZ Europe (Leipzig, Germany) for sequencing.

The *P. brassicae* genome was sequenced using both the Illumina NovaSeq sequencing platform and a single PacBio Sequel SMRT cell at GENEWIZ Europe (Leipzig, Germany). The single paired-end Illumina NovaSeq library was prepared with an insert size of c. 350–450 bp and 150 bp read length. The PacBio DNA library was prepared with an insert size of 20 kbp with Blue Pippin size selection as per the manufacturer's protocol. The prepared PacBio library was sequenced on the PacBio Sequel platform with v2.0 chemistry. The hybrid *de novo* assembly was done using the MaSuRCA 3.4.0 genome assembly and analysis toolkit (Zimin et al. 2017), utilizing both Illumina and PacBio-sequenced data. Illumina reads were merged and contigs of length greater than 250 bp were used for analysis. The normalization and read correction steps made use of MaSuRCA's built-in functionalities. QUASt (Gurevich et al. 2013) was used to generate statistics for the *de novo* assembled genome. MaSuRCA's implementation of Jellyfish was utilized for clustering assemblies based on *k*-mer estimations of the expected genome size. Finally, Flye (Kolmogorov et al. 2019) was used to construct the polished contigs. Completeness of the assembled genome was assessed using the Benchmarking Universal Single-Copy Orthologs tool (BUSCO) with BUSCO v4.1.3 (Simão et al. 2015) against ascomycota-odb10 and leotiomycetes\_odb10 datasets.

To obtain *P. brassicae* RNA sequencing (RNA-seq) data, fungal mycelia harvested from two liquid cultures of the isolate 15WOSR64-SS1 (grown as described above) were snap frozen with liquid nitrogen and sent to GENEWIZ Europe (Leipzig, Germany), where total RNA was extracted using the RNeasy Plant Mini Kit according to the manufacturer's protocol. Purity of the extracted RNA was tested with a Nanodrop 2000 spectrophotometer and the concentration was measured using a Qubit™ fluorometer. Further quality assessment was done using an RNA ScreenTape assay, followed by cDNA synthesis, library preparation and sequencing using the Illumina HiSeq sequencing platform at GENEWIZ Europe. Raw Illumina reads were trimmed using skewer v0.2.2 (Jiang et al. 2014) and *de novo* transcriptome assembly was done using Trinity v2.11.0 (Grabherr et al. 2011). Transdecoder was used to identify protein coding regions.

Structural annotation of the genome was done at DEEP Seq (Queen's Medical Centre, University of Nottingham, Nottingham) using the MAKER annotation pipeline v2.31.11 (Holt and Yandell 2011) incorporating assembled transcripts and predicted peptides. SNAP (Korf 2004) and AUGUSTUS v3.4.0 (Stanke and Waack 2003) gene predictions were applied in the MAKER pipeline v2.31.11. Gene models from the closely related species, *R. commune*, genome (NCBI\_

Assembly:GCA\_900074885.1) were also included into training of gene predictors. Repeat sequences were identified via MAKER using RepeatMasker 4.1.5 (Smit et al. 2013–2015) and tRNA (transfer RNA) and snoRNA (small nucleolar RNA) genes were predicted using tRNAscan-SE v2.0.9 (Chan et al. 2021) and snoscan 0.9.1 (Lowe and Eddy 1999), respectively. Functional annotation of the genes was done using Interproscan v5.0 (Jones et al. 2014) and BlastP (Altschul et al. 1990) against UniProt/SwissProt database.

To confirm the identity of the newly sequenced genome, phylogenetic analysis was done using nucleotide sequences of the internal transcribed spacers (ITS), translation elongation factor 1 (TEF1) and the beta tubulin genes. Publicly available nucleotide sequences of related fungal species and other *P. brassicae* isolates were obtained from the National Center for Biotechnology Information (NCBI) database and compared with those extracted from *P. brassicae* isolate 15WOSR-64SS1. Multiple alignments of ITS, TEF1 and beta tubulin nucleotide sequences were produced using MUSCLE (Edgar 2004). For the phylogenetic analysis of ITS sequences, all positions with less than 20% site coverage were eliminated and the Kimura 2-parameter (K2+G) model was used. For the beta-tubulin and TEF1 sequences, the Tamura-Nei (TN93+G) model was used after eliminating all positions with less than 60% and 30% site coverage, respectively. A maximum likelihood phylogenetic tree was produced by the program MEGA 11 (Tamura et al. 2021).

## Results and discussion

The genome sequence of *P. brassicae* was assembled into 45 scaffolds larger than 1,000 bp with a total genome size estimated to be 73.24 Mb. Of the 45 scaffolds, 38 were larger than 50,000 bp and the PacBio reads provided an estimated genome coverage of 95×. The genome had an average GC content of 42.85%, N50 value of 3,083,431 bp and L50 value of 11. BUSCO analysis against the ascomycota\_odb10 dataset indicated that the assembled genome was 98.4% complete, identifying 1,684 out of 1,706 BUSCOs as present in the genome assembly (1,679 complete and single-copy, 2 complete and duplicated, 3 fragmented and 22 missing BUSCOs). BUSCO analysis against the leotiomycetes\_odb10 dataset identified 3,181 out of 3,234 BUSCOs as present in the genome assembly (3,168 complete and single-copy, 1 complete and duplicated, 13 fragmented and 53 missing BUSCOs) with the genome completeness score of 97.9%. Table 7 summarises the key statistics of the *de novo* assembled genome. A total of 25,717 genes was predicted using the MAKER annotation pipeline, of which 23,620 were protein-coding genes. There were 143 tRNA genes and 1,954 snoRNA genes predicted. The phylogenetic trees (Fig. 9) based on the internal transcribed spacers (ITS), translation elongation factor 1 (TEF1) and beta tubulin nucleotide sequences confirmed the taxonomic identity of the sequenced isolate as *P. brassicae*.

This genome assembly is the first genome sequence of *P. brassicae* available. Considering the increased importance of this pathogen in recent years in the UK and North America, this genome resource will help to gain better understanding of *P. brassicae* and guide the development of more effective and efficient control strategies for light leaf spot. Control of crop diseases plays a key role in maintaining agricultural production for food security.

**Table 7.** *Pyrenopeziza brassicae* 15WOSR-64SS1 genome assembly statistics.

Metric	Value
Number of contigs	45
Total length (bp)	73,238,495
Largest contig (bp)	4,725,742
GC content (%)	42.81
Number of contigs ( $\geq 1,000$ bp)	45
Number of contigs ( $\geq 5,000$ bp)	43
Number of contigs ( $\geq 10,000$ bp)	42
Number of contigs ( $\geq 25,000$ bp)	39
Number of contigs ( $\geq 50,000$ bp)	38
N50 (bp)	3,083,431
N75 (bp)	1,656,915
L50	11
L75	18
Number of N's per 100 kbp	0.14
Predicted gene models	25,717
BUSCO completeness (%)	
ascomycota_odb10	98.4
leotiomycetes_odb10	97.8

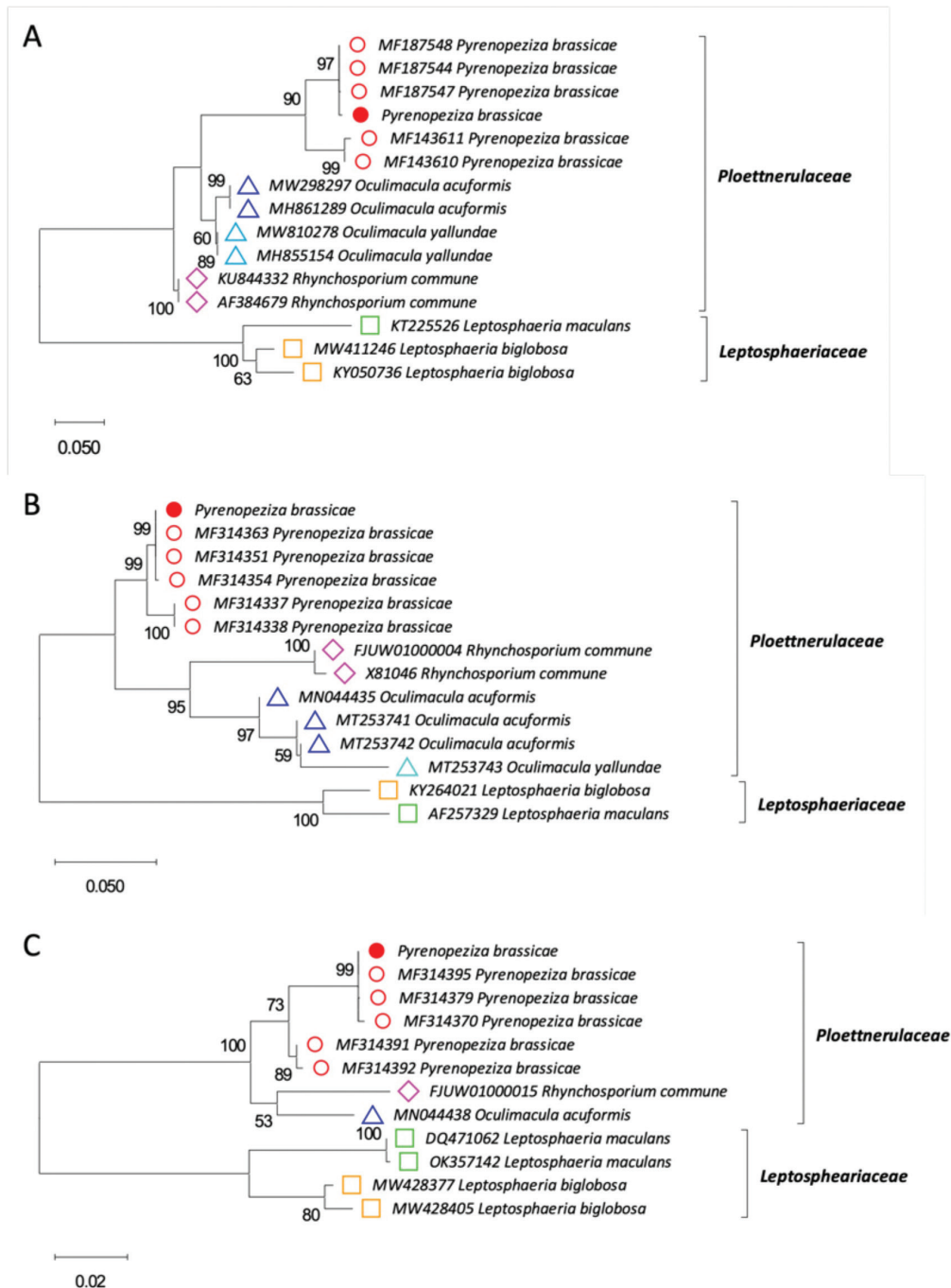
Recently, genomic resources have accelerated the research on host-pathogen interactions, making considerable improvements to crop disease control. Regarding the hosts, the genome sequences of five *Brassica* species have been made available, including *B. rapa* (Wang et al. 2011), *B. oleracea* (Liu et al. 2014), *B. napus* (Chalhoub et al. 2014), *B. juncea* and *B. nigra* (Yang et al. 2016). In contrast to the host genomic information, there is very little information about the *P. brassicae* genome. *De novo* sequencing of the *P. brassicae* genome will therefore fill this resource gap and help to understand molecular and genetic basis of its interactions with host *Brassica* species. Additionally, considering the close evolutionary relationship of *P. brassicae* to major pathogens of cereal crops, such as *R. commune* and *Oculimacula* spp., this genome will be of a wide scientific interest, as it can pave the way to new comparative genomic studies.

**Authors:** Chinthani S. Karandeni Dewage\*, Loly I. Kotta-Loizou, Henrik U. Stotz, Bruce D. L. Fitt, Yong-Ju Huang  
**\*Contact:** [c.s.karandeni-dewage@herts.ac.uk](mailto:c.s.karandeni-dewage@herts.ac.uk)

## Acknowledgements

The authors of the section on *Sphaerellopsis* would like to thank Dr Alexandros Georgios Sotiropoulos (University of Southern Queensland, Australia) for constructive discussions.

The authors of the section on *Colletotrichum* thank the College of Plant Protection, Fujian Agriculture and Forestry University for providing the laboratory and computational facility used in our research.



**Figure 9.** Phylogenetic analysis of the newly sequenced *Pyrenopeziza brassicae* and related fungi based on the nucleotide sequences of their (A) internal transcribed spacers (ITS), (B) beta-tubulin and (C) translation elongation factor 1 (TEF1). Phylogenetic analysis was done using nucleotide sequences of five related fungal species: *Rhynchosporium commune* (isolate IDs - 765.03.01, R157, UK7), *Oculimacula acuformis* (isolate IDs - 22-418, 22-443, 22-498, 22-499, CBS 495.80), *Oculimacula yallundae* (isolate IDs - 22-495, CBS 128.31, CBS 110665), *Leptosphaeria maculans* (isolate IDs - AFTOL-ID 277, Pk4), *L. biglobosa* (15PL-42, Azad4, Khal12) and of *P. brassicae* (isolate IDs - 233716, Cyc001, Cyc007, PC18, PC38). A multiple alignment of nucleotide sequences was produced using MUSCLE (Edgar 2004). For the phylogenetic analysis of ITS sequences, all positions with less than 20% site coverage were eliminated and the Kimura 2-parameter (K2+G) model was used. For the beta-tubulin and TEF1 sequences, the Tamura-Nei (TN93+G) model was used after eliminating all positions with less than 60% and 30% site coverage, respectively. A maximum likelihood phylogenetic tree was produced by the program MEGA 11 (Tamura et al. 2021). Bootstrap percentages (1,000 replicates) are shown. Different genera are indicated by different shapes and different species by different colours. Tips labelled with open shapes indicate amino acid sequences available in public databases. The red spot indicates the newly sequenced *P. brassicae*.

## Additional information

### Conflict of interest

Brenda D. Wingfield is a Senior Editor for IMA Fungus. The authors have declared that no competing interests exist.

### Ethical statement

No ethical statement was reported.

### Funding

Funding was received from the European Union's Horizon 2020 research and innovation program (RISE) under the Marie Skłodowska-Curie grant agreement No. 101008129, project acronym "Mycobiomics". The present work was supported by either Regione Campania PSR Campania 2014–2020, sottomisura 16.1 Azione 2 and Test Center authorized by Ministerial Decree 36000, 4th February 2004. This project was supported by the University of Melbourne and the Centre for Crop Health of the University of Southern Queensland, partly through Discovery Project DP210103869 funded by the Australian Research Council.

The authors of the section on *Pyrenopeziza brassicae* acknowledge funding support from UK Biotechnology and Biological Sciences Research Council (grant/award number: BBSRC/BB/P00489X/1) and Innovate UK (grant/award number: 102641).

This work was supported by grants from the Fujian Provincial Public Welfare Research Project (2021R1034002), Science and Technology Innovation Team Construction Project of FAAS (CXTD2021014-2), and the "5511" Collaborative Innovation Project of High-quality Agricultural Development and Surpassment in Fujian Province (XTCXGC2021003).

### Author contributions

D.D. and R.S. conceptualization; D.D. and R.S. wrote the original draft; R.S., M.V., L.C. and G.F. carried out laboratory work; D.D. performed bioinformatic analysis; E.L. and N.D revised and edited the manuscript; E.L. and N.D supervised the work; D.C. and E.L. funding acquisition.

XZ developed the project and obtained the funding, designed, and performed bioinformatic analysis, interpretation of data, wrote and reviewed the manuscript. YW, YH, HY and JZ performed the fungal subculture, genomic DNA extraction and purification. HH conceived and designed experiments, and performed bioinformatic analysis, interpretation of data, wrote and reviewed the manuscript. All authors read and approved the final manuscript.

### Author ORCIDs

Davide D'Angelo  <https://orcid.org/0009-0007-6542-6298>

Tiphany Nkomo  <https://orcid.org/0000-0003-3355-5080>

Xianzhi Zhou  <https://orcid.org/0000-0002-3734-075X>

Niloofar Vaghefi  <https://orcid.org/0000-0003-0430-4856>

Byron Sonnekus  <https://orcid.org/0000-0002-7228-4810>

Tanay Bose  <https://orcid.org/0000-0002-2069-042X>

Nicky Creux  <https://orcid.org/0000-0002-4179-6995>

Nunzio D'Agostino  <https://orcid.org/0000-0001-9840-3817>

Gerda Fourie  <https://orcid.org/0000-0003-2650-5448>

Almuth Hammerbacher  <https://orcid.org/0000-0002-0262-2634>



Alexander Idnurm  <https://orcid.org/0000-0001-5995-7040>  
Levente Kiss  <https://orcid.org/0000-0002-4785-4308>  
Emma T. Steenkamp  <https://orcid.org/0000-0003-0217-8219>  
Magriet A. van der Nest  <https://orcid.org/0000-0001-6914-8343>  
Chanthani S. Karandeni Dewage  <https://orcid.org/0000-0001-6621-8203>  
Brenda D. Wingfield  <https://orcid.org/0000-0002-6189-1519>

## Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information. This includes cited data base accessions.

## References

- AHDB (2021) Light leaf spot. [Available at:] <https://ahdb.org.uk/lightleafspot> [Accessed 02 January 2021]
- Akinsanmi OA, Nisa S, Jeff-Ego OS, Shivas R, G Drenth A (2017) Dry flower disease of macadamia in Australia caused by *Neopestalotiopsis macadamiae* sp. nov. and *Pestalotiopsis macadamiae* sp. nov. Plant Disease 101: 45–53. <https://doi.org/10.1094/PDIS-05-16-0630-RE>
- Ali SS, Asman A, Shao, J, Firmansyah, AP, Susilo AW, Rosmana A, McMahon P, Junaid M, Guest D, Kheng TY, L Meinhardt W, Bailey BA (2019) Draft genome sequence of fastidious pathogen *Ceratobasidium theobromae*, which causes vascular-streak dieback in *Theobroma cacao*. Fungal Biology and Biotechnology 6: 14. <https://doi.org/10.1186/s40694-019-0077-6>
- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S (2022) Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology 23(1): 258. <https://doi.org/10.1186/s13059-022-02823-7>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amaducci S, Scordia D, Liu FH, Zhang Q, Guo H, Testa G, Cosentino SL (2015) Key cultivation techniques for hemp in Europe and China. Industrial Crops and Products 68: 2–16. <https://doi.org/10.1016/j.indcrop.2014.06.041>
- Amaradasa BS, Turner A, Lowman S, Mei C (2020) First report of southern blight caused by *Sclerotium rolfsii* in industrial hemp in Southern Virginia. Plant Disease 104(5): 1563–1563. <https://doi.org/10.1094/PDIS-10-19-2157-PDN>
- Ammari MG, Gresham CR, McCarthy FM, Nanduri B (2016) HPIDB 2.0: a curated database for host–pathogen interactions. Database 2016: baw103. <https://doi.org/10.1093/database/baw103>
- Andrews S (2020) FastQC: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Armitage AD, Nellist CF, Bates HJ, Zhang L, Zou X, Gao Q-H, Harisson RJ (2020) Draft genome sequence of the strawberry anthracnose pathogen *Colletotrichum fructicola*. Microbiol Resour Announc 9(12): e01598-19. <https://doi.org/10.1128/MRA.01598-19>
- Aycock R (1966) Stem rot and other diseases caused by *Sclerotium rolfsii*: or The status of Rolfs' fungus after 70 years. Publisher, North Carolina Agricultural Experiment Station Technical Bulletin.

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5): 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bayon C, Yuan ZW, Ruiz C, Liesebach M, Pei MH (2006) Genetic diversity in the mycoparasite *Sphaerellopsis filum* inferred from AFLP analysis and ITS–5.8S sequences. *Mycological Research* 110: 1200–1206. <https://doi.org/10.1016/j.mycres.2006.06.001>
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69: 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research* 51(W1): W46–W50. <https://doi.org/10.1093/nar/gkad344>
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49: W29–W35. <https://doi.org/10.1093/nar/gkab335>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borodovsky M, Lomsadze A (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Current Protocols in Bioinformatics* 35(1): 4.6.1–4.6.10. <https://doi.org/10.1002/0471250953.bi0406s35>
- Boys EF, Roques SE, Ashby AM, Evans N, Latunde-Dada AO, Thomas JE, West JS, Fitt BDL (2007) Resistance to infection by stealth: *Brassica napus* (winter oilseed rape) and *Pyrenopeziza brassicae* (light leaf spot). *European Journal of Plant Pathology* 118: 307–321. <https://doi.org/10.1007/s10658-007-9141-9>
- Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma* 3(1): lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* 18: 366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Bucur DE, Huang YJ, Fitt BDL, Kildea S (2022) Azole fungicide sensitivity and molecular mechanisms of reduced sensitivity in Irish *Pyrenopeziza brassicae* populations. *Pest Management Science* 80(5): 2393–2404. <https://doi.org/10.1002/ps.7219>
- Bushnell B (2014) BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E). Berkeley (CA): Lawrence Berkeley National Lab (LBNL).
- Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 38: 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Carmody SM, King KM, O'camb CM, Fraaije BA, West JS, du Toit LJ (2020) A phylogenetically distinct lineage of *Pyrenopeziza brassicae* associated with chlorotic leaf spot of *Brassicaceae* in North America. *Plant Pathology* 69: 518–537. <https://doi.org/10.1111/ppa.13137>

- Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corr  a M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, B  rard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowden RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345: 950–953. <https://doi.org/10.1126/science.1253435>
- Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* 49: 9077–9096. <https://doi.org/10.1093/nar/gkab688>
- Chatzaki A, Papadaki AA, Krasagakis N, Papaisidorou G, Goumas DE, Markakis EA (2022) First report of southern blight caused by *Athelia rolfsii* on hemp in Greece. *Journal of Plant Pathology* 104(2): 871–872. <https://doi.org/10.1007/s42161-022-01072-8>
- Cheah LH, Hartill WFT, Corbin JB (1980) First report of the natural occurrence of *Pyrenopeziza brassicae* Sutton et Rawlinson, the apothecial state of *Cylindrosporium concentricum* Greville, in brassica crops in New Zealand. *New Zealand Journal of Botany* 18: 197–202. <https://doi.org/10.1080/0028825X.1980.10426917>
- Chen Y, Zeng L, Shu N, Jiang M, Wang H, Huang Y, Tong H (2018) *Pestalotiopsis*-like species causing gray blight disease on *Camellia sinensis* in China. *Plant Disease* 102: 98–106. <https://doi.org/10.1094/PDIS-05-17-0642-RE>
- Chikhi R, Medvedev P (2014) Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics* 30(1): 31–37. <https://doi.org/10.1093/bioinformatics/btt310>
- Chinese Pharmacopoeia Commission (2020) Pharmacopoeia of the People's Republic of China 2020. Beijing, The Medicine Science and Technology Press of China, 319–320.
- Clarke R, Merlin M (2016) Cannabis: evolution and ethnobotany. Univ of California Press.
- Crous PW, Wingfield MJ, Burgess TI, Hardy GE, Crane C, Barrett S, Cano-Lira JF, Leroux JJ, Thangavel R, Guarro J, Stchigel AM, Mart  n MP, Alfredo DS, Barber PA, Barreto RW, Baseia IG, Cano-Canals J, Cheewangkoon R, Ferreira RJ, Gen   J, Lechat C, Moreno G, Roets F, Shivas RG, Sousa JO, Tan YP, Wiederhold NP, Abell SE, Accioly T, Albizu JL, Alves JL, Antonioli ZI, Aplin N, Ara  jo J, Arzanlou M, Bezerra JDP, Bouchara J-P, Carlavilla JR, Castillo A, Castroagud  n VL, Ceresini PC, Claridge GF, Coelho G, Coimbra VRM, Costa LA, da cunha KC, da silva SS, Daniel R, de beer ZW, Due  as M, Edwards J, Enwistle P, Fiuza PO, Fournier J, Garc  a D, Gibertoni TB, Giraud S, Guevara-Suarez M, Gusm  o LFP, Haituk S, Heykoop M, Hirooka Y, Hofmann TA, Houbroken J, Hughes DP, Kautmanov   I, Koppel O, Koukol O, Larsson E, Latha KPD, Lee DH, Lisboa DO, Lisboa WS, L  pez-Villalba   , Maciel JLN, Manimohan P, Manj  n JL, Marincowitz S, Marney TS, Meijer M, Miller AN, Olariaga I, Paiva LM, Piepenbring M, Poveda-Molero JC, Raj KNA, Raja HA, Rougeron A, Salcedo I, Samadi R, Santos TAB, Scarlett K, Seifert KA, Shuttleworth LA, Silva GA, Silva M, Siqueira JPZ, Souza-Motta CM, Stephenson SL (2016) Fungal Planet description sheets: 469–557. *Persoonia* 37: 218–403. <https://doi.org/10.3767/003158516X694499>
- Dan Y, Liu M, Cui X, Ma W, Tang Z, et al. (2023) Identification of pathogen causing anthracnose on *Polygonatum cyrtonema* Hua. *Plant Protection* 49(2): 288–293.
- De Melo MP, Matos KS, Moreira SI, Silva FF, Concei  o GH, Nechet KL, Halfeld-Vieira BA, Beserra J  nior JEA, Ventura JA, Costa H, Furtado EL, Alves E, Ceresini PC (2018) Two

- new *Ceratobasidium* species causing white thread blight on tropical plants in Brazil. *Trop Plant Pathology* 43: 559–571. <https://doi.org/10.1007/s40858-018-0237-x>
- Dean R, van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD (2012) The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13(4): 414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x>
- Dearnaley JD (2007) Further advances in orchid mycorrhizal research. *Mycorrhiza* 17: 475–486. <https://doi.org/10.1007/s00572-007-0138-1>
- Decruse S, Neethu R, Pradeep N (2018) Seed germination and seedling growth promoted by a *Ceratobasidiaceae* clone in *Vanda thwaitesii* Hook. f., an endangered orchid species endemic to South Western Ghats, India and Sri Lanka. *South African Journal of Botany* 116: 222–229. <https://doi.org/10.1016/j.sajb.2018.04.002>
- Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679. <https://doi.org/10.1093/bioinformatics/btm009>
- Diogo E, Gonçalves CI, Silva AC, Valente C, Bragança H, Phillips AJL (2021) Five new species of *Neopestalotiopsis* associated with diseased *Eucalyptus* spp. in Portugal. *Mycological Progress* 20: 1441–1456. <https://doi.org/10.1007/s11557-021-01741-5>
- Doilom M, Hyde KD, Dong W, Liao CF, Suwannarach N, Lumyong S (2021) The plant family Asteraceae is a cache for novel fungal diversity: Novel species and genera with remarkable ascospores in *Leptosphaeriaceae*. *Frontiers in Microbiology* 12: 660261. <https://doi.org/10.3389/fmicb.2021.660261>
- Driessen SA, O'Brien PA, Hardy GEJ (2004) First record of the mycoparasite *Sphaerellopsis filum* on *Puccinia boroniae* in Australia. *Australasian Plant Pathology* 33: 463–464. <https://doi.org/10.1071/AP04048>
- Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* 50: D571–D577. <https://doi.org/10.1093/nar/gkab1045>
- Duong TA, De Beer ZW, Wingfield BD, Wingfield MJ (2013) Characterization of the mating-type genes in *Leptographium procerum* and *Leptographium profanum*. *Fungal Biology* 117: 411–421. <https://doi.org/10.1016/j.funbio.2013.04.005>
- Durgadevi, Rajendran L, Karthikeyan G, Raguchander T (2024) Investigation of rust mycoparasite *Sphaerellopsis filum* (TNAU SF1) volatilome during mono and di-partite interaction with *Puccinia arachidis*, causing Peanut rust. *European Journal of Plant Pathology* 168: 167–181. <https://doi.org/10.1007/s10658-023-02742-8>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 1–14. <https://doi.org/10.1186/s13059-019-1832-y>
- Farr DF, Bills GF (1989) *Fungi on plants and plant products in the United States*. APS Press.
- Fazary AE, Ju Y-H, Al-Shihri AS, Alfaifi MY, Alshehri MA (2016) Biodegradable siderophores: survey on their production, chelating and complexing properties. *Reviews in Inorganic Chemistry* 36: 153–181. <https://doi.org/10.1515/revic-2016-0002>
- Ferri F, Noviello C (1963) Tests on the control of *Sclerotium rolfsii* Sacc. on Hemp. *Phytopathol Mediterr* 2(2): 72–75.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020) Repeat-Modeler2 for automated genomic discovery of transposable element families. *PNAS* 117: 9451–9457. <https://doi.org/10.1073/pnas.1921046117>

- Freestone MW, Swarts ND, Reiter N, Tomlinson S, Sussmilch FC, Wright MM, Holmes GD, Phillips RD, Linde CC (2021) Continental-scale distribution and diversity of *Ceratobasidium* orchid mycorrhizal fungi in Australia. *Annals of Botany* 128: 329–343. <https://doi.org/10.1093/aob/mcab067>
- Gabriel L, Bruna T, Hoff KJ, Borodovsky M, Stanke M (2021) TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* 22: 1–12. <https://doi.org/10.1186/s12859-021-04482-0>
- Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytology* 197: 1236–1249. <https://doi.org/10.1111/nph.12085>
- Gao C-H, Yu G, Cai P (2021) ggVennDiagram: an intuitive, easy-to-use, and highly customizable R Package to generate venn diagram. *Frontiers in Genetics* 12: 706907. <https://doi.org/10.3389/fgene.2021.706907>
- Gerardo-Lugo S, Tovar-Pedraza J, Sajeewa M, Apodaca-Sanchez M, Correia K, Saucedo-Acosta CP, Camacho-Tapia M, Hyde KD, Marraiki N, Elgorban AM, Beltrán-Peña H (2020) Characterization of *Neopestalotiopsis* species associated with mango grey leaf spot disease in Sinaloa, Mexico. *Pathogens* 9: 788. <https://doi.org/10.3390/pathogens9100788>
- Gómez-Zapata PA, Díaz-Valderrama JR, Fatemi S, Ruiz-Casteo CO, Aime MC (2024) Characterization of the fungal genus *Sphaerellopsis* associated with rust fungi: species diversity, host-specificity, biogeography, and *in-vitro* mycoparasitic events of *S. macroconidialis* on the southern corn rust, *Puccinia polysora*. *IMA Fungus* 15: 18. <https://doi.org/10.1186/s43008-024-00145-w>
- Goodwin SB (2002) The barley scald pathogen *Rhynchosporium secalis* is closely related to the discomycetes *Tapesia* and *Pyrenopeziza*. *Mycological Research* 106: 645–654. <https://doi.org/10.1017/S0953756202006007>
- Gordon TC, Pfender WF (2012) Effects of the mycoparasite *Sphaerellopsis filum* on overwintering survival of stem rust in perennial ryegrass. *Plant Disease* 96: 1471–1481. <https://doi.org/10.1094/PDIS-10-11-0837-RE>
- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36(10): 3420–3435. <https://doi.org/10.1093/nar/gkn176>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652. <https://doi.org/10.1038/nbt.1883>
- Greville RK (1823) *Cylindrosporium concentricum*. *Scottish Cryptogamic Flora* 1, No. 27.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Haas H, Eisendle M, Turgeon BG (2008) Siderophores in fungal physiology and virulence. *Annual Review of Phytopathology* 46: 149–187. <https://doi.org/10.1146/annurev.phyto.45.062806.094338>
- Han H, Yoon JJ, Oh Y, Marin MV, Hugueta-Tapia J, Peres NA, Lee S (2024) Chromosome-scale genome sequence resource for two *Neopestalotiopsis* spp. isolates with



- different virulence in strawberry (*Fragaria × ananassa*). *PhytoFrontiers* 4: 422–426. <https://doi.org/10.1094/PHYTOFR-08-23-0110-A>
- Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, Bluhm BH, Cannon C, Castanera R, Culley DE, Daum C, Ezra D, González JB, Henrissat B, Kuo A, Liang C, Lipzen A, Lutzoni F, Magnuson J, Mondo SJ, Nolan M, Ohm RA, Pangilinan J, Park H-J, Ramírez L, Alfaro M, Sun H, Tritt A, Yoshinaga Y, Zwiers L-H, Turgeon BG, Goodwin SB, Spatafora JW, Crous PW, Grigoriev IV (2020) 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens. *Studies in Mycology* 96: 141–153. <https://doi.org/10.1016/j.smyco.2020.01.003>
- Harrington JM, Winkelman G, Haselwandter K, Crumbliss AL (2011) Fe (III)-complexes of the tripodal trishydroamate siderophore basidiochrome: potential biological implications. *Journal of Inorganic Biochemistry* 105: 1670–1674. <https://doi.org/10.1016/j.jinorgbio.2011.08.010>
- Hickman CJ, Schofield ER, Taylor RE (1955) Light leaf spot of Brassicae. *Plant Pathology* 4: 129–131. <https://doi.org/10.1111/j.1365-3059.1955.tb00060.x>
- Hirpara DG, Gajera HP, Hirpara HZ, Golakiya BA (2017) Antipathy of *Trichoderma* against *Sclerotium rolfsii* Sacc.: evaluation of cell wall-degrading enzymatic activities and molecular diversity analysis of antagonists. *Journal of Molecular Microbiology and Biotechnology* 27(1): 22–28. <https://doi.org/10.1159/000452997>
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12: 1-14. <https://doi.org/10.1186/1471-2105-12-491>
- Hsu S-Y, Lin Y-C, Xu Y-C, Chang H-X, Chung P-C, Ariyawansa HA (2022) high-quality genome assembly of *Neopestalotiopsis rosae* ML1664, the pathogen causing strawberry leaf blight and crown rot genome announcement. *Molecular Plant-Microbe Interactions* 35: 949–953. <https://doi.org/10.1094/MPMI-04-22-0077-A>
- Hu J, Fan J, Sun Z, Liu S (2020) NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36(7): 2253–2255. <https://doi.org/10.1093/bioinformatics/btz891>
- Hu WL, Ma YZ, Chen JZ (2019) First Report of *Citrus sinensis* anthracnose caused by *Colletotrichum fructicola* in China. *Plant Disease* 103: 1018. <https://doi.org/10.1094/PDIS-08-18-1466-PDN>
- Huth L, Ash GJ, Idnurm A, Kiss L, Vaghefi N (2021) The “bipartite” structure of the first genome of *Ampelomyces quisqualis*, a common hyperparasite and biocontrol agent of powdery mildews, may point to its evolutionary origin from plant pathogenic fungi. *Genome Biology and Evolution* 13: evab182. <https://doi.org/10.1093/gbe/evab182>
- Inglis PW, Pappas MDCR, Resende LV, Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high-quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLOS ONE* 13: e0206085. <https://doi.org/10.1371/journal.pone.0206085>
- Iqbal MA, Tomar RS, Parakhia MV, Singla D, Jaiswal S, Rathod VM, Padhiyar SM, Kumar N, Rai A, Kumar D (2017) Draft whole genome sequence of groundnut stem rot fungus *Athelia rolfsii* revealing genetic architect of its pathogenicity and virulence. *Scientific Reports* 7(1): 5299. <https://doi.org/10.1038/s41598-017-05478-8>
- Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I (2017) ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Research* 27(5): 768–777. <https://doi.org/10.1101/gr.214346.116>

- Jiang H, Lei R, Ding SW, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15: 182. <https://doi.org/10.1186/1471-2105-15-182>
- Jones P, Binns D, Chang H Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Jung Y, Han D (2022) BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinforma 38: 2404–2413. <https://doi.org/10.1093/bioinformatics/btac137>
- Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Research 35(Web Server issue): 429–432. <https://doi.org/10.1093/nar/gkm256>
- Karandeni Dewage CS, Klöppel CA, Stotz HU, Fitt BDL (2018) Host–pathogen interactions in relation to management of light leaf spot disease (caused by *Pyrenopeziza brassicae*) on *Brassica* species. Crop and Pasture Science 69: 9–19. <https://doi.org/10.1071/CP16445>
- Karandeni Dewage CS, Qi A, Stotz HU, Huang YJ, Fitt BDL (2021) Interactions in the *Brassica napus*–*Pyrenopeziza brassicae* pathosystem and sources of resistance to *P. brassicae* (light leaf spot). Plant Pathology 70: 2104–2114. <https://doi.org/10.1111/ppa.13455>
- Karolewski Z, Kaczmarek J, Jedryczka M, Cools HJ, Fraaije BA, Lucas JA, Latunde-dada AO (2012) Detection and quantification of airborne inoculum of *Pyrenopeziza brassicae* in Polish and UK winter oilseed rape crops by real-time PCR assays. Grana 51: 270–279. <https://doi.org/10.1080/00173134.2011.653401>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30: 772–780. <https://doi.org/10.1093/molbev/mst010>
- Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20: 1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kiss L (2001) The role of hyperparasites in host plant – parasitic fungi relationships. In: Jeger MJ, Spence NJ (Eds) Biotic Interactions in Plant-Pathogen Associations. CABI Publishing, Wallingford, 227–236. <https://doi.org/10.1079/9780851995120.0227>
- Kiss L, Pintye A, Kovács GM, Jankovics T, Fontaine MC, Harvey N, Xu X, Nicot PC, Bardin M, Shykoff JA, Giraud T (2011) Temporal isolation explains host-related genetic differentiation in a group of widespread mycoparasitic fungi. Molecular Ecology 20: 1492–1507. <https://doi.org/10.1111/j.1365-294X.2011.05007.x>
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37: 540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Research 27: 722–736. <https://doi.org/10.1101/gr.215087.116>
- Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5: 59. <https://doi.org/10.1186/1471-2105-5-59>

- Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49(W1): W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14): 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan, J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078. <https://doi.org/10.1093/bioinformatics/btp352>
- Li H, Zhou GY, Liu JA, Xu JP (2016) Population genetic analyses of the fungal pathogen *Colletotrichum fructicola* on tea-oil trees in China. *PLOS ONE* 11(6): e0156841. <https://doi.org/10.1371/journal.pone.0156841>
- Li Q, Bu J, Yu Z, Tang L, Huang S, Guo T, Mo J, Hisang T (2018) Draft genome sequence of an isolate of *Colletotrichum fructicola*, a causal agent of mango anthracnose. *Genome Announce* 6: e00001-18. <https://doi.org/10.1128/genomeA.00001-18>
- Liang X, Cao M, Li S, Kong Y, Sun G (2020) Highly contiguous genome resource of *Colletotrichum fructicola* generated using long-read sequencing. *Molecular Plant-Microbe Interactions* 33: 790–793. <https://doi.org/10.1094/MPMI-11-19-0316-A>
- Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, Sun G (2018) Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. *PLOS ONE* 13(4): e0196303. <https://doi.org/10.1371/journal.pone.0196303>
- Liesebach M, Zaspel I (2004) Genetic diversity of the hyperparasite *Sphaerellopsis filum* on *Melampsora* willow rusts. *Forest Pathology* 34: 293–305. <https://doi.org/10.1111/j.1439-0329.2004.00369.x>
- Lima NB, Kryvenki M, Conforto C, Serri D, Kramer R, Roca M, Vargas-Gil S (2020) First report of white thread blight caused by *Ceratobasidium niltonsouzanum* on yerba mate in Argentina. *Plant Disease* 104: 572–572. <https://doi.org/10.1094/PDIS-08-19-1603-PDN>
- Liu F, Ma ZY, Hou LW, Diao YZ, Wu WP, Damm U, Song S, Cai L (2022) Updating species diversity of *Colletotrichum*, with a phylogenomic overview. *Studies in Mycology* 101: 1–56. <https://doi.org/10.3114/sim.2022.101.01>
- Liu LP, Zhang L, Qiu PL, Wang Y, Liu YN, Li Y, Gao J, Hsiang T (2020) Leaf spot of *Polygonatum odoratum* caused by *Colletotrichum spaethianum*. *Journal of General Plant Pathology* 86: 157–161. <https://doi.org/10.1007/s10327-019-00903-4>
- Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang T-J, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park B-S, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee T-H, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications* 5: 3930. <https://doi.org/10.1038/ncomms4930>
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964. <https://doi.org/10.1093/nar/25.5.955>

- Lowe TM, Eddy SR (1999) A computational screen for methylation guide snoRNAs in yeast. *Science* 283: 1168–1171. <https://doi.org/10.1126/science.283.5405.1168>
- Ma J, Xiao X, Wang X, Guo M (2021) *Colletotrichum spaethianum* causing anthracnose on *Polygonatum cyrtoneura* in Anhui Province, China. *Plant Disease* 109(2): 509. <https://doi.org/10.1094/PDIS-04-20-0778-PDN>
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21(16): 3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Maharachchikumbura SSN, Hyde KD, Groenewald JZ, Xu J, Crous PW (2014) *Pestalotiopsis* revisited. *Studies in Mycology* 79: 121–186. <https://doi.org/10.1016/j.simyco.2014.09.005>
- Manni M, Berkeley MR, Seppey M, Simã, FA, Zdobnov EM (2021a) 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. <https://doi.org/10.1093/molbev/msab199>
- Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021b) BUSCO: Assessing genomic data quality and beyond. *Current Protocols* 1: e323. <https://doi.org/10.1002/cpz1.323>
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A (2018) MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* 14(1): e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
- McMahon P, Purwantara A (2016) Vascular streak dieback (*Ceratobasidium theobromae*): History and Biology. In: Bailey B, Meinhardt L (Eds) *Cacao Diseases*. Springer International Publishing, New York, 307–335. [https://doi.org/10.1007/978-3-319-24789-2\\_9](https://doi.org/10.1007/978-3-319-24789-2_9)
- McPartland J, Clarke R (2000) Hemp diseases and pests: management and biological control. CABI. <https://doi.org/10.1079/9780851994543.0000>
- Mersha Z, Kering M, Ren S (2020) Southern blight of hemp caused by *Athelia rolfsii* detected in Virginia. *Plant Disease* 104(5): 1562. <https://doi.org/10.1094/PDIS-10-19-2178-PDN>
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A (2018) Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34: i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5): 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35(Web Server issue): W182–185. <https://doi.org/10.1093/nar/gkm321>
- Mosquera-Espinosa AT, Bayman P, Prado GA, Gómez-Carabalí A, Otero JT (2013) The double life of *Ceratobasidium*: orchid mycorrhizal fungi and their potential for biocontrol of *Rhizoctonia solani* sheath blight of rice. *Mycologia* 105: 141–150. <https://doi.org/10.3852/12-079>
- Nag Raj TR (1993) *Coelomycetous Anamorphs with Appendage-bearing Conidia*. Waterloo, ON: Mycologue Publications.
- Németh MZ, Mizuno Y, Kobayashi H, Seress D, Shishido N, Kimura Y, Takamatsu S, Suzuki T, Takikawa Y, Kakutani K, Matsuda Y, Kiss L (2021) *Ampelomyces* strains isolated from diverse powdery mildew hosts in Japan: Their phylogeny and mycoparasitic activity, including timing and quantifying mycoparasitism of *Pseudoidium neolycopersici* on tomato. *PLOS ONE* 16(5): e0251444. <https://doi.org/10.1371/journal.pone.0251444>

- Nischwitz C, Newcombe G, Anderson CL (2005) Host specialization of the mycoparasite *Eudarluka caricis* and its evolutionary relationship to *Ampelomyces*. *Mycological Research* 109: 421–428. <https://doi.org/10.1017/S0953756205002431>
- Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I (2015) Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics* 16(1): 230. <https://doi.org/10.1186/s12859-015-0663-4>
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A (2023) InterPro in 2022. *Nucleic Acids Research* 51(D1): D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Pei MH, Ruiz C, Hunter T, Bayon C (2010) Interactions between *Melampsora larici-epitea* pathotypes and the mycoparasite *Sphaerellopsis filum* from willow rusts. *Forest Pathology* 40: 33–42. <https://doi.org/10.1111/j.1439-0329.2009.00604.x>
- Pei MH, Yuan ZW (2005) Mycoparasite *Sphaerellopsis filum* and its potential for biological control of willow rust. In: Pei MH, McCracken AR (Eds) *Rust Diseases of Willow and Poplar*. CABI Publishing International, Wallingford, UK, 243–253. <https://doi.org/10.1079/9780851999999.0243>
- Pertea G, Pertea M (2020) GFF Utilities: GffRead and GffCompare. *F1000Research* 9: 304. <https://doi.org/10.12688/f1000research.23297.2>
- Phookamsak K, Hyde KD, Jeewon R, Bhat DJ, Jones EBG, Maharachchikumbura SSN, Raspé O, Karunarathna SC, Wanasinghe DN, Hongsanan S, Doilom M, Tennakoon DS, Machado AR, Firmino AL, Ghosh A, Karunarathna A, Mešić A, Dutta AK, Thongbai B, Devadatha B, Norphanphoun C, Senwanna C, Wei D, Pem D, Ackah FK, Wang G-N, Jiang H-B, Madrid H, Lee HB, Goonasekara ID, Manawasinghe IS, Kušan I, Cano J, Gené J, Li J, Das K, Acharya K, Raj KNA, Latha KPD, Chethana KWT, He M-Q, Dueñas M, Jadan M, Martín MP, Samarakoon MC, Dayarathne MC, Raza M, Park MS, Telleria MT, Chaiwan N, Matočec N, de Silva NI, Pereira OL, Singh PN, Manimohan P, Uniyal P, Shang Q-J, Bhatt RP, Perera RH, Alvarenga RLM, Nogal-Prata S, Singh SK, Vadthananarat S, Oh S-Y, Huang S-K, Rana S, Konta S, Paloi S, Jayasiri SC, Jeon SJ, Mehmood T, Gibertoni TB, Nguyen TTT, Singh U, Thiyagaraja V, Sarma VV, Dong W, Yu X-D, Lu Y-Z, Lim YW, Chen Y, Tkalc̃ec Z, Zhang Z-F, Luo Z-L, Daranagama DA, Thambugala KM, Tibpromma S, Camporesi E, Bulgakov TS, Dissanayake AJ, Senanayake IC, Dai DQ, Tang L-Z, Khan S, Zhang H, Promputtha I, Chomnunti LCP, Zhao R-L, Lumyong S, Boonmee S, Wen T-C, Mortimer PE, Xu J (2019) Fungal diversity notes 929–1035: taxonomic and phylogenetic contributions on genera and species of fungi. *Fungal Divers* 95: 1–273. <https://doi.org/10.1007/s13225-019-00421-w>
- Plachecka A (2005) Microscopical observations of *Sphaerellopsis filum*, a parasite of *Puccinia recondita*. *Acta Agrobotanica* 58: 67–71. <https://doi.org/10.5586/aa.2005.010>
- Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes *de novo* assembler. *Current Protocols in Bioinformatics* 70: 102. <https://doi.org/10.1002/cpbi.102>
- Punja Z, Rahe J (1992) *Methods for research on soilborne phytopathogenic fungi*. American Phytopathological Society.
- Punja ZK (2021) Emerging diseases of *Cannabis sativa* and sustainable management. *Pest Management Science* 77(9): 3857–3870. <https://doi.org/10.1002/ps.6307>
- Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11: 1432. <https://doi.org/10.1038/s41467-020-14998-3>



- Rawlings ND, Waller M, Barrett, AJ, Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 42(D1): D503–D509. <https://doi.org/10.1093/nar/gkt953>
- Rawlinson CJ, Sutton BC, Muthyalu G (1978) Taxonomy and biology of *Pyrenopeziza brassicae* sp. nov. (*Cylindrosporium concentricum*), a pathogen of winter oilseed rape (*Brassica napus* ssp. *oleifera*). *Transactions of the British Mycological Society* 71: 425–439. [https://doi.org/10.1016/S0007-1536\(78\)80070-9](https://doi.org/10.1016/S0007-1536(78)80070-9)
- Rebollar-Alviter A, Silva-Rojas HV, Fuentes-Aragón D, Acosta-González U, Martínez-Ruiz M, Parra-Robles BE (2020) An emerging strawberry fungal disease associated with root rot, crown rot and leaf spot caused by *Neopestalotiopsis rosae* in Mexico. *Plant Disease* 104: 2054–2059. <https://doi.org/10.1094/PDIS-11-19-2493-SC>
- Rhie A, Walenz BP, Koren S, Phillippy AM (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* 21(1): 245. <https://doi.org/10.1186/s13059-020-02134-9>
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rossmann AY, Crous PW, Hyde KD, Hawksworth DL, Aptroot A, Bezerra JL, Bhat JD, Boehm E, Braun U, Boonmee S, Camporesi E, Chomnunti P, Dai D-Q, D'souza MJ, Dissanayake A, Jones EBG, Groenewald JZ, Hernández-Restrepo M, Hongsanan S, Jaklitsch WM, Jayawardena R, Jing LW, Kirk PM, Lawrey JD, Mapook A, McKenzie EHC, Monkai J, Phillips AJL, Phookamsak R, Raja HA, Seifert KA, Senanayake I, Slippers B, Suetrong S, Tanaka K, Taylor JE, Thambugala KM, Tian Q, Tibpromma S, Wanasinghe DN, Wijayawardene NN, Wikee S, Woudenberg JHC, Wu H-X, Yan J, Yang T, Zhang Y (2015) Recommended names for pleomorphic genera in *Dothideomycetes*. *IMA Fungus* 6: 507–523. <https://doi.org/10.5598/imafungus.2015.06.02.14>
- Rozewicki J, Li S, Amada KM, Standley DM, Katoh K (2019) MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* 47(W1): W5–W10. <https://doi.org/10.1093/nar/gkz342>
- Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, Lam KJK, Tian N, Russum S, Wang J, Medrano-Soto A (2021) The transporter classification database (TCDB): 2021 update. *Nucleic Acids Research* 49: D461–D467. <https://doi.org/10.1093/nar/gkaa1004>
- Schroeder K, Paulitz T (2012) First report of a *Ceratobasidium* sp. causing root rot on canola in Washington state. *Plant Disease* 96: 591–591. <https://doi.org/10.1094/PDIS-12-11-1038-PDN>
- Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation-Supplementary data. *PLOS ONE* 11(10). <http://shenwei356.github.io/seqkit/>
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7(1): 539. <https://doi.org/10.1038/msb.2011.75>
- Silva-Cabral JRA, da Silva JL, Soares L da SH, Costa JFO, Amorim EP da R, Lima GS de A, Assunção IP (2019) First report of *Colletotrichum fructicola* and *C. tropicale* causing anthracnose on orchids in Brazil. *Plant Disease* 103: 2672. <https://doi.org/10.1094/PDIS-03-19-0518-PDN>

- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. <http://www.repeat-masker.org>
- Smit AFA, Hubley R, Green P (2013–2015) RepeatMasker Open-4.0. <http://www.repeat-masker.org>
- Sorrentino R, Baldi GM, Battaglia V, Raimo F, Piccirillo G, Lahoz E (2021) First report of white root rot of hemp (*Cannabis sativa*) caused by *Dematophora necatrix* in Campania region (Southern Italy). *Plant Disease* 105(10): 3299. <https://doi.org/10.1094/PDIS-07-20-1521-PDN>
- Sorrentino R, Lahoz E, Battaglia V, Sorrentino MC, Cerrato D (2023) First report of seedling damping-off of industrial hemp (*Cannabis sativa*) caused by seed-transmitted *Alternaria rosae* in Italy. *Plant Disease* 107(5): 1636. <https://doi.org/10.1094/PDIS-06-22-1375-PDN>
- Sorrentino R, Pergamo R, Battaglia V, Raimo F, Cermola M, Lahoz E (2019) Characterization and pathogenicity of *Fusarium solani* causing foot rot on hemp (*Cannabis sativa* L.) in Southern Italy. *Journal of Plant Diseases and Protection* 126(6): 585–591. <https://doi.org/10.1007/s41348-019-00265-1>
- Sperschneider J, Dodds PN (2022) EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Molecular Plant-Microbe Interactions* 35: 146–156. <https://doi.org/10.1094/MPMI-08-21-0201-R>
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24(5): 637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Stanke M, Schöffmann O, Morgenstern B, Waack S (2006a) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7: 62. <https://doi.org/10.1186/1471-2105-7-62>
- Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19: ii215–ii225. <https://doi.org/10.1093/bioinformatics/btg1080>
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006b) AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* 34(Web Server issue): W435–W439. <https://doi.org/10.1093/nar/gkl200>
- Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, Denton AK (2021) Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* 36(22–23): 5291–5298. <https://doi.org/10.1093/bioinformatics/btaa1044>
- Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* 12(1): 2. <https://doi.org/10.1186/s13100-020-00230-y>
- Sun JZ, Liu XZ, McKenzie EHC, Jeewon R, Liu JK, Zhang XL, Zhao Q, Hyde KD (2019) Fungicolous fungi: terminology, diversity, distribution, evolution, and species checklist. *Fungal Diversity* 95: 337–430. <https://doi.org/10.1007/s13225-019-00422-9>
- Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* 38: 3022–3027. <https://doi.org/10.1093/molbev/msab120>

- Tarailo-Graovac M, Chen N (2004) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* 5(1): 4–10. <https://doi.org/10.1002/0471250953.bi0410s25>
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* 18: 1979–1990. <https://doi.org/10.1101/gr.081612.108>
- Testa AC, Oliver RP, Hane JK (2016) OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *GBE* 8: 2044–2064. <https://doi.org/10.1093/gbe/evw121>
- Thixton HL, Esselman EJ, Corey LL, Zettler L (2020) Further evidence of *Ceratobasidium* D.P. Rogers (*Basidiomycota*) serving as the ubiquitous fungal associate of *Platanthera leucophaea* (*Orchidaceae*) in the North American tallgrass prairie. *Botanical Studies* 61: 12. <https://doi.org/10.1186/s40529-020-00289-z>
- Tomioka K, Moriwaki J, Sato T (2008) Anthracnose of *Polygonatum falcatum* caused by *Colletotrichum dematium*. *Journal of General Plant Pathology* 74: 402–404. <https://doi.org/10.1007/s10327-008-0112-6>
- Trakunyingcharoen T, Lombard L, Groenewald JZ, Cheewangkoon R, Toanun C, Alfenas AC, Crous PW (2014) Mycoparasitic species of *Sphaerellopsis*, and allied lichenicolous and other genera. *IMA Fungus* 5: 391–414. <https://doi.org/10.5598/imafungus.2014.05.02.05>
- Trifinopoulos J, Nguyen L-T, Von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 44: W232–W235. <https://doi.org/10.1093/nar/gkw256>
- Tu CC, Kimbrough JW (1978) Systematics and phylogeny of fungi in the *Rhizoctonia* complex. *Botanical Gazette* 139(4): 454–466. <https://doi.org/10.1086/337021>
- Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, Sahu J, Iyer SV, Khamari L, De Silva N, Martinez MC, Pedro H, Yates AD, Hammond-Kosack KE (2022) PHI-base in 2022: a multi-species phenotype database for Pathogen-Host Interactions. *Nucleic Acids Research* 50: D837–D847. <https://doi.org/10.1093/nar/gkab1037>
- Utami YD, Hiruma K (2022) Genome Resource of *Colletotrichum spaethianum*, the causal agent of leaf anthracnose in *Polygonatum falcatum*. *PhytoFrontiers* 2: 152–155. <https://doi.org/10.1094/PHYTOFR-12-21-0082-A>
- Vaghefi N, Bar I, Lawley JW, Thanjavur Sambasivam P, Christie M, Ford M (2024) Population-level whole genome sequencing of *Ascochyta rabiei* identifies genomic loci associated with isolate aggressiveness. *Microbial Genomics* 10: 001326. <https://doi.org/10.1099/mgen.0.001326>
- Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14): 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Wang HC, Huang YF, Chen Q, Wang MS, Xia HQ, Shang SH, Zhang CQ (2016) Anthracnose caused by *Colletotrichum fructicola* on Tobacco (*Nicotiana tabacum*) in China. *Plant Disease* 100: 1235–1236. <https://doi.org/10.1094/PDIS-06-15-0724-PDN>
- Wang QT, Cheng YH, Zhuo L, Wang Y, Zhou H, Hou C-L (2022) *Neopestalotiopsis longiappendiculata* as the agent of grey blight disease of *Camellia* spp. *Journal of Phytopathology* 170: 770–777. <https://doi.org/10.1111/jph.13139>
- Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park B-S, Weissshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IAP, Batley J, Kim J-S, Just J, Li J,

- Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon S-J, Choi S-R, Lee T-H, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* 43: 1035–1039. <https://doi.org/10.1038/ng.919>
- Warcup JH, Talbot PHB (1980) Perfect states of Rhizoctonias associated with orchids III. *New Phytology* 86: 267–272. <https://doi.org/10.1111/j.1469-8137.1980.tb00787.x>
- Webster J, Weber R (2007) Introduction to fungi [Adobe Digital Editions version]. [Retrieved from] <http://herts.ebilib.com>
- White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications* 18: 315–322. <https://doi.org/10.1016/B978-0-12-372180-8.50042-1>
- Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with Kraken 2. *Genome Biology* 20: 257. <https://doi.org/10.1186/s13059-019-1891-0>
- Wu H Y, Tsai CY, Wu,YM, Ariyawansa HA, Chung CL, Chung P-C (2020) First report of *Neopestalotiopsis rosae* causing leaf blight and crown rot on strawberry in Taiwan. *Plant Disease* 105(2): 487–487. <https://doi.org/10.1094/PDIS-05-20-1045-PDN>
- Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, Wang Y (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 47(W1): W52–W58. <https://doi.org/10.1093/nar/gkz333>
- Yan L, Wang Z, Song W, Fan P, Kang Y, Lei Y, Wan L, Huai D, Chen Y, Wang X, Sudini H, Liao B (2021) Genome sequencing and comparative genomic analysis of highly and weakly aggressive strains of *Sclerotium rolfsii*, the causal agent of peanut stem rot. *BMC Genomics* 22(1): 276. <https://doi.org/10.1186/s12864-021-07534-0>
- Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, Hu Z, Chen S, Pental D, Ju Y, Yao P, Li X, Xie K, Zhang J, Wang J, Liu F, Ma W, Shopan J, Zheng H, Mackenzie SA, Zhang M (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics* 48: 1225–1232. <https://doi.org/10.1038/ng.3657>
- Yin Y, Mao,X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40(W1): W445–W451. <https://doi.org/10.1093/nar/gks479>
- Yuan ZW, Pei MH, Hunter T, Royle DJ (1998) *Eudarlucica caricis*, the teleomorph of the mycoparasite *Sphaerellopsis filum*, on blackberry rust *Phragmidium violaceum*. *Mycological Research* 102: 866–868. <https://doi.org/10.1017/S0953756297005832>
- Yuan ZW, Pei MH, Hunter T, Ruiz C, Royle DJ (1999) Pathogenicity to willow rust, *Melampsora epitea*, of the mycoparasite *Sphaerellopsis filum* from different sources. *Mycological Research* 103: 509–512. <https://doi.org/10.1017/S0953756298007448>
- Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y (2023) dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research* 51(W1): W115–W121. <https://doi.org/10.1093/nar/gkad328>
- Zhou X, Li M, Rao B, Chen Y, Cai C, Gao H (2020) First report of anthracnose on *Paris polyphylla* var. *chinensis* caused by *Colletotrichum fructicola* in northern Fujian, China. *Plant Disease* 104: 2728. <https://doi.org/10.1094/PDIS-04-20-0693-PDN>
- Zhou X, Su H, Chen Y, Gao H, Tang J, et al. (2017) Occurrence regularity of major diseases of *Polygonatum cyrtoneura*. *Fujian Agri Sci Technol* 48(10): 25–27.

Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the mega-reads algorithm. *Genome Research* 1: 066100. <https://doi.org/10.1101/gr.213405.116>

## Supplementary material 1

### List of protein-coding genes identified in *A. rolfsii* strain CCF1 and their annotations

Authors: Davide D'Angelo, Roberto Sorrentino, Ernesto Lahoz, Domenico Cerrato, Maurizio Viscardi, Loredana Cozzolino, Giovanna Fusco, Nunzio D'Agostino

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3897/imafungus.16.141732.suppl1>