Penicillium coffeae, a novel hyperparasite of wheat stripe rust

Jack Wess

May 2021

A thesis submitted for the degree of Bachelor of Science

(Honours)

Research School of Biology, The Australian National University

Supervised by Prof. John Rathjen

Word count 10,055

Declaration of original work

This thesis is my original work, except where otherwise indicated.

Jack Wess

27 May 2021

Table of Contents

Acknowledgements

I am most grateful for all the support and encouragement I have received from people up and down the hallway of the Plant-Microbes Interactions floor.

A special thanks to John, Alex and Yiheng, and of course the entire Rathjen Lab, for making this year so enjoyable. I have learnt a lot, not just in science or research but in attempting difficult things and persevering even when things do not go according to plan.

Thank you to my examiners David Jones, Jana Sperschneider and Ashley Jones, your encouragement and guidance have been most helpful over the course of this project.

Thank you to my Mum and Dad and Laura and Andrew, thank you for the endless encouragement and support and for always reminding me to keep going.

Finally, thank you to Sambasivan Periyannan, for discovering the hyperparasite in the first place, I wouldn't be here without you.

Abstract

Fungal hyperparasites attack other fungi as sources of nutrients. This phenomenon is of fundamental biological interest but also has potential applications via development of novel biocontrol agents. In this thesis I investigate the genomics of a novel *Pencillium* isolate which parasitises the spores of the wheat stripe rust pathogen *Puccinia striiformis* f. sp. *tritici* (*Pst*). *Pst* is a significant threat to wheat production on all continents where the crop is grown. Traditional methods of control struggle to remain effective against the rapidly evolving fungus. With conventional methods of control becoming less attractive, interest in the use of fungal hyperparasites as a form of biocontrol is growing.

Reports on rust hyperparasites in the scientific literature have largely been limited to descriptive studies. The fortuitous discovery in this laboratory of a *Penicillium* fungus growing on *Pst* spores opened an opportunity for a more in-depth investigation of the organism. This thesis describes the sequencing and assembly of the genome of the novel hyperparasite and its use in taxonomic placement and comparative transcriptomic analysis. The genome was assembled using a hybrid approach based on whole-genome sequencing using long- and short-reads to generate a highly complete chromosome level assembly. Although a phylogenomic (whole-genome) approach to identify the novel *Penicillium* to species level was not successful due to a lack of comparable datasets, a DNA-barcoding approach identified the *Penicillium* as a novel isolate of the plant endophyte *Penicillium coffeae.* The genome was annotated using Illumina RNA-seq data derived from cultures of *Penicillium* grown on synthetic media or on *Pst* spores on wheat leaves. The combined transcriptomes were used as an input to annotate 12,181 genes, which is the first gene set described for *P. coffeae*. The gene set was found to be highly complete by a BUSCO analysis which searches for highly conserved, near-universal single-copy ortholog genes that are expected to exist in the gene set. A comparative transcriptomic analysis of the *P. coffeae* isolate grown on synthetic media or on *Pst* spores revealed the upregulation of several genes with possible antifungal and hyperparasitic functions, including an efflux pump belonging to a biosynthetic gene cluster.

In summary, in this thesis I have generated a chromosome-level genome resource for a wheat stripe rust hyperparasite that I identified as an isolate of the plant endophyte *Penicillium coffeae.* I have annotated the genome for gene content and generated two transcriptomes which can be used for identification of genes that underly the facultative pathogenic lifestyle of the fungus. These valuable resources will facilitate a deeper investigation into the *P. coffeae* genes that are expressed

or upregulated during hyperparasitism to identify the unknown mechanisms underlying the colonisation of rust spores.

List of Abbreviations

benA	Beta-tubulin
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
BUSCO	Benchmarking Universal Single-Copy Ortholog
CaM	Calmodulin
CAZY	Carbohydrate-active enzymes
cv.	Cultivar
Eggnog	Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups
EVM	Evidence modeler
Fungal RiPP	Fungal ribosomally synthesised and post-translationally modified peptide product
Gb	Gigabases (1,000,000,000 base pairs)
GO-terms	Gene-ontology terms
ITS	Internal transcribed spacer
Kb	Kilobase (1,000 base pairs)
Mb	Megabase (1,000,000 base pairs)
NCBI	National Center for Biotechnology Information
NRPS	Non-ribosomal peptide synthetase (biosynthetic gene cluster)
PDA	Potato dextrose agar
Pfam	Protein families database
PGT	*Puccinia graminis* forma speciales *tritici*
PST	*Puccinia striiformis* forma speciales *tritici*
PT	*Puccinia triticina*
RNA-seq	RNA sequencing
RPB2	DNA-directed RNA polymerase II subunit
SEM	Scanning electron microscope
SM	Secondary Metabolite
SP	Signal peptide
T1PKS	Type 1 polyketide synthase (biosynthetic gene cluster)

1 Introduction

Wheat is one of the most important agricultural crops worldwide, supplying approximately 20% of the human caloric intake (Hartmann *et al*., 2017). Diseases caused by fungal plant pathogens are currently some of the greatest contributors to annual wheat yield losses and are a major threat to global food security (Godfray *et al.,* 2016). Rust diseases are amongst the most important fungal diseases in wheat causing USD $4.3-5.0 billion in losses annually (Figueroa *et al*., 2018). There are three agriculturally significant wheat rust diseases; stripe, stem and leaf rust, caused by the pathogens *Puccinia striiformis* forma speciales (special form; f. sp.) *tritici* (*Pst), P. graminis* f. sp. *tritici* (*Pgt*) and *P. triticina* (*Pt*) respectively (Singh *et al.,* 2016). Wheat stripe rust is currently the most destructive disease in wheat and exists on all continents where wheat is grown (Schwessinger 2017).

The *Pst* fungus is an obligate biotroph meaning that it must infect live host tissue to complete its lifecycle (Cheng *et al.* 2017). Infection with *Pst* begins when wind or rain deposits a mature urediniospore on the leaf surface. The urediniospore germinates and establishes an infection inside the leaf, culminating in the production of new urediniospores, which erupt through the leaf surface in the form of a pustule. These urediniospore are then dispersed to begin another infection cycle. *Pst* can complete its life cycle within 11 days (Garnica *et al.,* 2014), allowing for many infection cycles throughout the growing season. Stripe rust is a major interest of the Rathjen Laboratory and was the starting point of the current investigation.

Traditional methods of fungal control include the selective breeding of wheat cultivars with resistance traits, and the application of fungicides. Development of resistant cultivars is slow, partially due to a need to bring yield and grain quality up to standard in the new lines (Summers & Brown 2013). Newly developed cultivars might maintain resistance for only a few years before they are overcome by the rapidly evolving fungus (Draz 2019). This has led to an increased relianc on chemical fungicides, the high selection pressure of these fungicides has led to the evolution of resistant fungi (Estep *et al*., 2015). Fungicides are also coming under increased scrutiny due to their proven harmfulness both to human health and the environment and regulations regarding their use are tightening (European Food Safety *et al.,* 2018). Whilst these measures act to protect the health

of the environment, they come at the cost of food security – as we remain heavily reliant on chemical controls to maintain yields. Alternative approaches are needed that maintain effectiveness whilst impacting minimally on the environment.

With conventional methods of resistance becoming less effective, interest in the development of biocontrol agents to control fungal pathogens is increasing. A biocontrol involves the intentional release of a biological agent (organism or virus) into the environment as a means of controlling a pest or disease (Barratt *et al.,* 2018). A class of organisms well suited to this role are fungal hyperparasites. A hyperparasite is a parasitic organism whose host is also itself a parasite (Parratt & Laine 2016). Several species of *Trichoderma* are used successfully as biocontrol products, for example Trichodex™ which is an emerging alternative to chemical fungicides (Topolovec-Pintarić 2019). Pathogenesis mechanisms employed by hyperparasitic fungi include the production of toxic secondary metabolites (SMs), and direct invasion of the host via the action of cell-wall degrading enzymes (Howell 2003). Production of SMs is often coordinated by biosynthetic gene clusters (BGCs), groups of adjacent genes associated with production of a specific secondary metabolite (Rokas *et al.,* 2020). One such SM is 6-n-pentyl-2H-pyran-2-one (6PAP), produced by *Trichoderma,* and required for successful hyperparasitism (Jeleń *et al.,* 2014). Expression level of the 6PAP-BGC was used to select *Trichoderma* species most suitable for use as biocontrol agents, enabling its commercialisation. Interest is also on Carbohydrate Active enZYmes (CAZYs), groups of carbohydrate hydrolysis enzymes, with diverse functions. Recently, Eitzen *et al.,* (2021) discovered a novel fungal species that secretes a GH25 lysozyme CAZY that can control an oomycete pathogen of *Arabidopsis thaliana.*

Genetic analysis of novel fungal isolates

Genetic analysis is growing to be the predominant approach to fungal taxonomy. The internal transcribed spacer (*ITS*) hypervariable region is commonly analysed to establish fungal identity. The *ITS* is a region located between rRNA subunit genes in all fungal genomes and is useful for taxonomy due to its high variability between species (Martin & Rygiewicz 2005). The *ITS* is amplified and sequenced with the use of specific PCR primers and then compared against *ITS* regions of other known fungi, commonly obtained from a genomic database such as the National

Center for Biotechnology Information (NCBI) GenBank (Fajarningsih 2016). Advancements in second- and third-generation DNA shotgun sequencing techniques have made whole genome assembly a rapid and cheap option for genetic analysis and identification of fungi (Feau *et al.*, 2018).

Isolation of a novel *Penicillium* that hyperparasitises *Pst* spores

During an experiment testing the effects of high relative humidity on stripe rust-infected wheat plants, Dr Sambasivan Periyannan of this laboratory discovered what appeared to be white fluffy growth atop rust pustules (Figure 1.1).

Figure 1.1 Two *Pst* infected wheat plants, both kept in a high relative humidity environment, on the left the normal infection phenotype showing bright orange pustules of *Pst* spores which have burst through the leaf surface. On the right, white fluffy tissue has grown over the pustules, when isolated this tissue was found to be *Penicillium x* ANU01.

Yiheng Hu of this laboratory isolated samples of the white tissue and under a microscope observed the presence of a filamentous fungus, with a signature paintbrush shape, amongst the circular rust spores (Figure 1.2).

Figure 1.2 Microscopic images of tissue harvested from infected wheat leaves such as those shown in Figure 1.1. Arrows point to the paintbrush-shaped fruiting bodies, characteristic of filamentous fungi (Henry 2019). Amongst the filamentous fungus are the orange circular *Pst* spores, and the top right image shows tissue from an uninfected pustule of *Pst.*

The first step in investigating the fungus was to confirm that the original white fluffy phenotype was the result of this filamentous fungus, Yiheng Hu isolated the filamentous fungus, grew it axenically on nutrient agar then using it, inoculated stripe rust infected plants. Under high humidity

conditions, the filamentous fungus again grew on the pustules. Importantly, the filamentous fungus would not produce the same appearance on wheat that was not infected with stripe rust, indicating it is reliant on the rust spores for growth, and is acting as a hyperparasite. Yiheng Hu conducted a preliminary ITS-analysis of the hyperparasitic fungus, preliminarily classifying it as a species of *Penicillium*. *Penicillium* are filamentous fungi from the phylum *Ascomycota*. They are haploid, containing a single nucleus and 4-7 chromosomes (Färber & Geisen 2000). *Penicillium* species are ubiquitous in the environment and the genus is expansive containing some 400+ species (ICPA 2018). The novel *Penicillium* hyperparasite was arbitrarily designated *Penicillium x*, until a species level identification could be determined for the isolate.

How might this novel *Penicillium* hyperparasitise *Pst*?

Fungi that hyperparasitise *Pst* remain under-researched. Novel isolates of *Alternaria alternata, Cladosporium cladosporioides* and *Simplicillium obclavatum* have been documented infecting *Pst* (Zhan et al. 2014; Zheng et al. 2017; Wang et al. 2020). These studies used electron scanning microscopy to observe hyperparasite hyphae invading the *Pst* spores, a significant finding in learning about the invasion mechanisms of rust hyperparasites. More recently however, Wilson et al. (2020) discovered an isolate of *Penicillium brevicompactum* that controlled *Puccinia triticina* and *Puccinia hordei* without colonising pustules, presumably antagonising the spores via secretion of an antifungal secondary metabolite. The presence of *Penicillium x* tissue covering the rust pustules suggests that it is physically invading the *Pst* spores, however it is possible that *Penicillium x* deploys both antifungal metabolites and direct invasion during hyperparasitism. Currently, little is known about the genetic basis of mechanisms deployed by hyperparasites of *Pst* in their antagonism. This project aimed to investigate the specific genetic mechanisms behind the antagonism of *Pst* by a newly discovered *Penicillium.*

Identifying the novel hyperparasite: A genetic approach

To better understand the hyperparasite, it was important that I identify it to species level. There are very few *Penicillium* species that have been identified as hyperparasites of rust fungi (Wilson et al. 2020). The preliminary *ITS*-based identification by Mr Hu did not indicate affinity to any known

species. Thus, it was hypothesized that the hyperparasite may be a novel *Penicillium* species. The first aim of this project was therefore:

1. Confirmation of the hyperparasite as a species of *Penicillium* and taxonomic placement within the genus.

I initially attempted to complete this aim using a whole genome approach to identification. The first chapter of this paper will describe how I sequenced and assembled the genome of the novel hyperparasite. The second chapter will describe how I then used genetic information from the assembly to identify the novel hyperparasite to species level.

Investigating the genetic mechanisms of hyperparasitism: functional annotation and comparative transcriptomics

The mechanisms behind pathogenicity in hyperparasitic fungi are largely unknown. No prior research has been undertaken to identify the molecular mechanisms by which *Penicillium* hyperparasites antagonise rust fungi. I hypothesised that similar to Trichoderma, the *Penicillium* utilises one or more BGC-regulated secondary metabolite(s) in its antagonism. Therefore, the second aim of the project was:

2. Identification of the transcriptional program used by the hyperparasite during its hyperparasitic growth on *Pst* spores.

The third chapter of this thesis will describe how I generated transcriptome profiles of *Penicillium x* growing axenically or pathogenically respectively and used these data to annotate the newly assembled hyperparasite genome. I then compared the transcriptome profiles to identify genes associated with hyperparasitism. This gave a preliminary list of fungal genes with likely roles in the parasitic lifestyle.

2 Highly contiguous assembly of the *Penicillium x* genome using long- and short-read DNA sequences

2.1 Background

Two aims of this project were to identify the hyperparasite and determine potential genetic mechanisms underlying its hyperparasitic lifestyle. Both aims are facilitated by assembly of a high-quality genome for *Penicillium x.* The genome is the complete set of information in an organism's DNA, organised into chromosomes (Goldman & Landweber 2016). Most chromosomes are linear and flanked by highly repetitive telomeric regions, while mitochondrial chromosomes are significantly smaller and circular (Chen et al. 2019). The genome contributes to all traits of the organism and in turn, much of its physiology and traits can be understood better through knowledge of its genome sequence (Stein 2001). This chapter will describe my sequencing and assembly of the genome of the hyperparasitic *Penicillium*.

Genome assembly projects have become increasingly accessible and accurate with the advent of second- and third-generation sequencing technologies (Pop *et al.*, 2004). This allows whole-genome sequencing directly from DNA preparations, rather than requiring expensive and laborious cloning steps. Second-generation DNA sequencing technologies such as the Illumina platform generate short (50-300 bp), accurate DNA reads (Tan *et al.* 2019). Third-generation technologies such as Oxford Nanopore's MinION can produce ultralong (~15+ kb) reads from single-molecule DNA strands but are far more error-prone than second generation methods (Fu *et al.,* 2019). Use of these technologies together in a hybrid strategy allows rapid and cheap access to moderate size eukaryote genome assemblies (Goodwin *et al.,* 2015).

Sequence reads are built into overlapping contiguous regions called contigs via a computational process. These contigs can be very large, and in some cases can span entire chromosomes from telomere to telomere. Hybrid genome assemblers use both long and short reads in contig assembly (Wallberg *et al.*, 2019). Long reads solve common problems in assembly by spanning repetitive regions of the genome which is where short-read assemblies fail. Such contigs are large but with many sequence inaccuracies due to the error rate inherent in the technology (Fu *et al.,* 2019). However, the sequences can be corrected (or "polished") using high read-depth highly accurate short reads (Zimin & Salzberg 2020). Hybrid assembly is the method used in this paper.

The quality of a genome assembly is assessed by the so-called 3Cs, completeness, contiguity and correctness (Thrash *et al.*, 2020). Completeness refers to the completeness of the gene catalogue expected in the organism (Veeckman *et al.*, 2016). It is assessed by comparing the presence of expected single copy orthologs to those that actually appear in the assembly. Benchmarking Universal Single-Copy Orthologs (BUSCOs) are highly conserved, near-universal single-copy genes expected to be present in all species within a certain group, for example all fungal species within the order Eurotiales should possess 4191 characteristic BUSCO genes (Simão *et al.*, 2015). BUSCOs are used in assessing genome and transcriptome completeness, and their use in phylogenomics will be described in a later chapter (Waterhouse *et al.*, 2018). Contiguity refers to the number and size of the contigs that make up the assembly (Thrash *et al.*, 2020). One common metric for assessing this is the N50 value, which is a length-weighted median obtained by ranking all contigs from largest to smallest, adding their sizes successively, then identifying the size of the contig containing the base corresponding to 50% of the genome length (Earl *et al.*, 2011). A chromosome level assembly refers to an assembly in which each expected chromosome is represented by a singular contig with flanking telomeric regions (Shingate *et al.*, 2020). Fungi also possess small mitochondrial chromosomes which are most commonly circular and do not possess telomeres (Chen *et al.*, 2019). Correctness is defined as the accuracy of the assembly compared to a gold standard reference genome (Yang *et al.*, 2019). As the assembly presented in this Chapter was constructed *de novo*, meaning without reference, correctness is an inapplicable measure of assembly accuracy. Another factor controlling the level of confidence in the identity of each called base is the coverage or read depth. This is the number of times a given nucleotide base has been sequenced (appears in independent reads) (Sims *et al.*, 2014). Increased genome coverage usually means increased likelihood that a given base has been called correctly (Ou *et al.*, 2020).

A high-quality genome assembly is instrumental to downstream analyses. In this Chapter I describe how I used a hybrid assembly approach using long- and short-read sequencing to assemble the genome of *Penicillium x.* Reads were assembled into eight contigs, consisting of 7 chromosomes and one predicted mitochondrial chromosome. Achieving this high-quality assembly was an important first step in achieving the aims of this thesis.

2.2 Materials and Methods

Fungal and plant materials

Penicillium x isolate ANU01 was initially discovered by Dr Sambasivam Periyannan of this laboratory in the Controlled Environment Facility at ANU and used for all parts of this project. The fungus was grown from a spore stock on PDA plates (25g/L potato dextrose broth, 15g/L agar) in 22°C growth rooms under 12-hour day/night cycles.

Bioinformatics materials

Details of the software used in this methods section, as well as links to their respective Github pages are available at the Github for this project:

https://github.com/Jack-WessANU/Genomics_of_a_Pst_hyperparasite

DNA extraction

Tissue was harvested from four axenic growth plates of *Penicillium x*, each plate containing one colony approximately 6.5 cm in diameter. Mycelial and spore tissue was harvested via scraping into four 2 ml Eppendorf tubes (Eppendorf) containing two sterile ball bearings, and snap frozen in liquid nitrogen. Tissue was disrupted using the TissueLyser LT (QIAGEN) for four rounds of 30 s at 25 Hz, refreezing the samples between each round of disruption. After homogenisation, the samples were washed twice in 0.35 M sorbitol solution (Jones and Schwessinger 2020). After this, DNA was extracted from solution using a slightly modified version of a high MW DNA harvest protocol (Jones et al. 2020). All solutions and buffers used were made as per the protocol following the 'small' prep size procedure. Briefly, tissue was homogenized as described above, then cells were lysed, and protein and RNA degraded by addition of DNA extraction buffer, RNAse A and proteinase K, with incubation in a 55°C water bath for 60 min. Cell debris was then removed by addition of 5 M potassium acetate and centrifugation at 5,000 rcf for 5 minutes at 4°C. DNA was recovered from solution by addition of 2% Sera-Mag beads and specialised binding buffer. A magnetic rack was used to pull the beads to one side of the Eppendorf tube, and beads were washed twice in 70% ethanol. DNA was then eluted from beads using 52 µl of 10 mM Tris-HCl pH 8.0 and stored in 1.5 ml DNA LoBind® tubes (Eppendorf). Two µl of eluted DNA was quantified with a Qubit® 2.0 Fluorometer (Broad Range) (Invitrogen) following the manufacturer's instructions.

DNA purity was assessed using 2 µl eluted DNA on both Quickdrop™ Micro-Volume Spectrophotometer (SpectraMax®) and Nanodrop™ spectrophotometers (ThermoFisher). DNA integrity and strand length were assessed on a 1% agarose gel, following the protocol of Green & Sambrook (2019). Initial attempts to follow this entire protocol proved unsuccessful, either producing low DNA yields or impure samples. Small modifications made to improve yield included increasing the temperature of the water bath (step 10; cell lysis) from 55°C to 60°C, increasing volumes of Proteinase K (20mg/mL stock) and RNAse A (20mg/mL stock) from 4 µl and 8 µl to 6 µl and 9 µl respectively (step 9; cell lysis) and allowing one hour at room temperature for the DNA to elute from beads (step 28; DNA elution). The modified protocol eventually produced a sufficient quantity of DNA (15 µg), however the DNA was impure, so the decision was made to clean and size-select the samples on a 15 kb PippinHT™ HPE-7510 (Sage Science) gel filtration device. Due to constraints on the input volume of the Pippin Prep, only ~9.5 µg of DNA could be loaded. DNA was run on the Pippin Prep with the assistance of Dr Ashley Jones of the Schwessinger laboratory following manufacturer guidelines. The purified DNA was checked for purity on a Nanodrop spectrophotometer which showed the sample was highly pure and could be used for library preparation.

Library preparation was completed with the assistance of Dr Ashley Jones of the Schwessinger laboratory. DNA libraries were prepared using the 1D-ligation SQK-LSK110 kit (Oxford Nanopore Technologies) following manufacturer's instructions. Briefly, nicks in DNA strands were repaired using specialised FFPE DNA Repair Buffer and dsDNA ends were prepared for adapter ligation with a specialised End-prep enzyme mix and reaction buffer. The DNA was then cleaned and specialised adapters (attached to motor proteins) were ligated to the ends of DNA strands using T4 DNA ligase. The samples were incubated at room temperature before washing to produce a highly pure library-prepared DNA sample. The concentration of library-prepared DNA was checked using a Qubit® 2.0 Fluorometer (Broad Range) (Invitrogen). A total of 612 ng of library-prepared DNA was available for sequencing. Approximately 400 ng of library DNA was loaded on to a newly opened Nanopore FLO-MIN106D 9.4.1 Flowcell (Oxford Nanopore Technologies) and sequenced on a MinION™ Mk1B device (Oxford Nanopore Technologies). Sequencing was pre-programmed to run until four gigabases of data had been collected. Guppy (v4.2.2) (Oxford Nanopore Technologies) was used to base call nucleotide sequences from the MinION output. Adapters were trimmed from the FASTQ reads using the Porechop (v0.2.4)

software from (Wick 2018) software and default parameters. The NanoPlot (v1.36.2) software (De Coster *et al.,* 2018) was used to check read distribution and statistics regarding length and quality of reads.

The sequence reads were filtered into four smaller subsets based on the parameters of minimum read length (kb) and minimum read quality (Q) using the Nanofilt (v2.7.1) software (De Coster et al. 2018). Exact code for the Nanofilt run, including exact parameters, is presented in the GitHub page for this project. Assemblies were then built using the *10kb8Q* and *10kb12Q* Nanofilt subsets and the assembly tools Canu (v2.1.1) from Koren *et al.,* (2017) and Flye (v2.8.2) from Kolmogorov et al. (2019). Exact parameters used in all four assemblies can be found on the project's GitHub. General statistics on the preliminary assemblies were provided by the Quast (v5.0.2) software from Gurevich *et al.,* (2013) run with default parameters. Telomeric regions were identified in assemblies using the FindTelomeres.py script developed by (Sperschneider 2019). Finally, completeness of assemblies was assessed with the BUSCO (v4.1.4) software (Simão *et al.*, 2015). Specific parameters used to generate BUSCO analysis for all four preliminary assemblies is available on the project's Github. The main parameter used was '-l Eurotiales_odb10 –m genome'.

Short read processing and Polishing

Paired-end Illumina sequence reads (150 bp) from *Penicillium x* DNA were generated by Mr Yiheng Hu of this laboratory prior to the start of this project. Paired-end reads were trimmed of adapters using the TrimGalore (v0.6.6) software (Martin 2011) and parameters '-q 20 –nextera –paired'. Quality of Illumina reads was checked with the FASTQC software (Andrews 2010). All assemblies were polished using the same procedure. First a custom for-loop bash script was used to run 10 iterations of the Medaka (v1.2.1) long read polisher (ONT 2020) on each assembly. The for-loop bash script is presented on the projects Github page. Genomes were then polished twice with the Illumina paired end reads and the POLCA software developed by Zimin & Salzberg (2020). The quality of these assemblies was then checked using the Quast (v5.0.2) software from Gurevich *et al.,* (2013), the FindTelomeres script developed by (Sperschneider 2019)and a BUSCO (v4.1.4) analysis, using identical parameters to the BUSCO analyses of unpolished assemblies ('-l Eurotiales_odb10 –m genome'). The exact parameters used in assessing these assemblies is

presented in the Github page for this project. The circular representation of the chromosome level assembly of *Penicillium x* was generated with the Circos software, developed by (Krzywinski *et al.*, 2009).

2.3 Results

Isolation and purification of high molecular weight (HMW) DNA from *Penicillium x*

Isolation of high purity high molecular weight (HMW) DNA is essential to obtain long reads. The procedure described in the materials and methods above, is the result of a lengthy optimisation process described in the methods. I harvested spore and mycelial tissue from four large (~6.5cm diameter) colonies of *Penicillium* grown axenically. The tissue was snap frozen and then disrupted using a bead shaker (Qiagen TissueLyser). The pulverised material was washed in a sorbitol solution. DNA extraction buffer, Proteinase K and RNAse A were then added to the tissue and the samples were incubated at 60°C for 60 minutes. Enzyme quantities were modified to optimise the DNA extraction yield (data not shown). DNA was captured from solution using DNA-binding Sera-Mag beads and cleaned with a series of ethanol washes. Bound DNA was eluted with 10 mM Tris-HCl pH 8.0, and total yield quantified using a Qubit Broad Range assay. The size and integrity of the DNA was checked on a 1% agarose gel; data not shown. Absorbances were measured using a Quickdrop spectrophotometer. An average A260/280 of 1.7845 and A260/230 of 0.98725 suggested that the eluted DNA was not sufficiently pure. To solve this and to remove low MW fragments, I added an additional step of gel purification using a 15 kb size-selective PippinHT™. Approximately 9.5 µg of DNA was run on the Pippin Prep. Approximately 85% of DNA was lost in the process leaving ~1.5 µg highly pure DNA for library preparation. For preparation of 1D-ligation libraries for Nanopore sequencing specialised sequencing adapters attached to motor proteins are ligated to the prepared ends of DNA strands. These motor proteins facilitate the movement of the DNA strand through the Nanopore protein pore (Caldwell & Spies 2017). After library preparation, 612 ng of DNA remained. Approximately 400 ng of library prepared DNA was sequenced on a MinION device.

Long-read sequencing using a MinION device

Sequencing of highly pure HMW DNA was conducted on a MinION Mk1B device, using a FLO-MIN106D 9.4.1 flowcell (Oxford Nanopore Technologies). The sequencing run was pre-programmed to run until four gigabases (Gb) of data had been collected. The Guppy (v4.2.2) script was used to basecall nucleotide sequences from the MinION output. Statistics on sequencing output were determined using the Nanoplot software. A total of 575,122 reads equal to 4,066,222,961 bases were sequenced, with a mean read length of ~7 kbp and mean read quality of 11.8Q. A phred quality score of 10 means a 1 in 10 probability that any given base was called incorrectly (Ewing & Green 1998). A mean read quality of 11.8Q indicates a read set of sufficient quality to proceed with processing and assembly. I used the Nanofilt software to subset reads based on minimum read length in kilobases (kb) and minimum phred-score (Q). Read subsets were created, trimmed of sequencing-adapters using the Porechop software and then analysed again for size and quality distribution using the Nanoplot software, data not shown. The read subsets *10kb8Q* (minimum read length 10 kb; minimum phred score = 8) and *10kb12Q* were eventually chosen for assembly. The *10kb12Q* is a high-stringency dataset intended to test if more accurate reads would result in a more complete assembly. The *10kb8Q* subset contained 130,683 reads (total bases: 2,389,489,479), while *10kb12Q* contained 85,871 reads (total bases: 1,552,580,861).

Assembly of long-reads and polishing using short-read data Illumina data

Genome assemblies for each read subset were generated with the long-read assembly tools Canu and Flye for a total of four assemblies. Each assembler took a fastq file containing multiple Nanopore reads and associated quality scores, assembled them into large contigs and output them in a single fasta file (fastq file without the quality information) referred to as the assembly.A detailed description of the preliminary assemblies is presented in Table 4.2. Contiguity of the assemblies was assessed using the QUAST software and the FindTelomeres script. For both assemblers, the larger read sets with lower Q scores (see "Coverage") resulted in more contiguous assemblies. Canu outperformed Flye by generating assemblies in a fewer number of contigs. Genome sizes for all assemblies were similar at around 31 Mb. Canu also produced the longest contig of around 7.65 Mb and had higher N50 and N75 values than the Flye assemblies. All assemblies detected a similar GC content of around 48.4%. The largest number of telomeres were identified in Canu assemblies; 7 forward and 6 reverse, suggesting 7 linear chromosomes. Canu

but not Flye predicted a single circular contig 55,939 bp in size, likely to be the mitochondrial chromosome. There was little to choose between Canu assemblies based on these statistics, except for one fewer contig in the *10kb8Q* assembly.

The completeness of each genome assembly was assessed via a BUSCO analysis using the 'Eurotiales_odb10' BUSCO gene set as reference. This tells us how many of the 4191 BUSCO genes expected to exist in all Eurotiales (order which *Penicillium* belongs to) are present in the genome assemblies that I generated. A detailed description of the results of the BUSCO analysis is presented in Table 2.1. Overall, Flye significantly outperformed Canu. Flye assemblies contained more complete (including single copy and duplicated) BUSCO genes and fewer fragmented and missing BUSCOs. Thus, preliminary Flye assemblies were significantly more complete than preliminary Canu assemblies. Overall, Canu outperformed Flye in contiguity, and Flye outperfomed Canu in completeness. Though slightly less complete, I decided to proceed with the *10kb8Q* Canu assembly, due to its assembly of the genome in one fewer contig than the *10kb12Q* Canu.

Table 2.1. Summary statistics for unpolished preliminary assemblies of the *Penicillium x* genome.

	Canu		Flye	
Measurement	**10kb8Q**	**10kb12Q**	**10kb8Q**	**10kb12Q**
Coverage	79x	39x	76x	49x
# Total contigs	8	9	15	16
# Circular contigs	1	1	0	0
Total length (bp)	30,941,400	31,059,644	30,895,086	30,932,441
Largest contig (bp)	7,658,982	7,659,564	5,725,045	5,741,314
N50	4,563,323	4,563,703	3,566,461	3,567,672
N75	3,595,238	3,594,243	3,107,033	3,107,241
GC (%)	48.80	48.72	48.83	48.82
# Telomeres	7 forward 6 reverse	7 forward 6 reverse	4 forward 7 reverse	4 forward 6 reverse
BUSCO - Eurotiales				
# Complete BUSCOs (C)	3,899 (93.0%)	3,920 (93.6%)	3,988 (95.2%)	3,958 (94.4%)
# Complete and single-copy BUSCOs (S)	3,881 (92.6%)	3,904 (93.2%)	3,969 (94.7%)	3,941 (94.0%)
# Complete and duplicated BUSCOs (D)	18 (0.4%)	16 (0.4%)	19 (0.5%)	17 (0.4%)
# Fragmented BUSCOs (F)	87 (2.1%)	88 (2.1%)	56 (1.3%)	76 (1.8%)
# Missing BUSCOs (M)	205 (4.9%)	183 (4.3%)	147 (3.5%)	157 (3.8%)
# Total BUSCO groups searched	4191	4191	4191	4191

Polishing of initial Canu genome assemblies with short-read Illumina data

Genome polishing resolves misassembled regions or regions of ambiguity in the genome assembly by aligning more accurate data such as short Illumina sequence reads (Zimin & Salzberg 2020). This can greatly improve completeness whilst also slightly improving contiguity (Murigneux *et al.* 2020). The *10kb8Q* Canu assembly was first polished with Nanopore long reads, using 10 iterations of the Medaka (v1.2.1) polishing software. I then polished the assembly twice using Illumina short reads generated by Yiheng Hu of this laboratory from *Penicillium x* DNA. Illumina reads were trimmed of adapters and low-quality reads using the TrimGalore software, assessed with FastQC.

A total of 31,281,129 forward and reverse paired-end reads had been sequenced, with an average read length of 150 bp and an average phred-score of 34. The read set was of sufficiently high quality for assembly polishing (Chen *et al.,* 2021). Short-read polishing of *10kb8Q* was done using the POLCA software. Two iterations of POLCA resolved 17 nucleotide substitution errors, 781 insertion/deletion errors and returned a consensus quality score of 99.999%. A third round of POLCA polishing resulted in 4 substitution changes and no change to BUSCO-assessed completeness. Therefore, further polishing was not conducted. The *10kb12Q* Canu assembly was polished similarly (data not shown), however, the issue of contiguity in this assembly was not resolved. BUSCO-assessed completeness showed that the *10kb8Q* and *10kb12Q* assemblies were equally complete (data not shown for *10kb12Q*). Initially, Flye assemblies were not polished, as I assumed that their poor performance in contiguity indicated poor quality assemblies. However, I reconsidered this when I reflected on their initial BUSCO-assessed completeness. Therefore, I wished to see if Flye assemblies could outperform those of Canu once polished. A BUSCO analysis performed after polishing found that polished Flye assemblies were equally complete compared to polished Canu assemblies, however polishing could not resolve their low contiguity (data not shown). I therefore retained the *10kb8Q* Canu assembly for all subsequent analyses in this thesis.

A summary of the final genome assembly is presented in Table 2.2. Comparing the polished *10kb8Q* assembly statistics presented in Table 2.2 to the description of the unpolished *10kb8Q* Canu assembly in Table 2.1 shows that many improvements were made by polishing. No changes were made to the number of contigs, or the identification of the circular contig. The GC percentage remained unchanged, as did the number of forward and reverse telomeres. The presence of 7 linear contigs as well as 7 forward and 6 reverse telomeres indicates we possess a chromosome-level assembly, with each contig representing a single chromosome. Though one telomere remained missing even after polishing, I expect it can be identified via manual search of the incomplete end of the contig for telomeric repeats. Polishing increased the completeness of the assembly, increasing the number of complete *Eurotiales* BUSCOs by 4.4%, while reducing the number of fragmented and missing BUSCOs. A circular visualisation of the predicted *Penicillium x* chromosomes and their respective sizes are presented in Figure 2.1.

Table 2.2. Summary statistics for the final polished assembly of the *Penicillium x* genome

Measurement	**10kb8Q**
Sequencing technology	Nanopore MinION
Assembler	Canu (v2.1.1)
Long read polishing tool	Medaka (v1.2.1) (10 iterations)
Short read polishing tool	POLCA (2 iterations)
Coverage	79x
# Total contigs	8
# Circular contigs	1
Total length (bp)	30965311
Largest contig (bp)	7664923
N50	4566672
N75	3598224
GC (%)	48.82
# Telomeres - Contiguity	7 forward
	6 reverse
BUSCO - Eurotiales	
# Complete BUSCOs (C)	4085 (97.4%)
# Complete and single-copy BUSCOs (S)	4067 (97.0%)
# Complete and duplicated BUSCOs (D)	18 (0.4%)
# Fragmented BUSCOs (F)	18 (0.4%)
# Missing BUSCOs (M)	88 (2.2%)
# Total BUSCO groups searched	4191

Fig 2.1. *Penicillium x* karyotype (contigs) visualised using the Circos software. Chromosome lengths: PX1- 7,664,923 bp, PX2 – 6,380,533 bp, PX3 – 4,566,672 bp, PX4 – 4,080,968 bp, PX5 – 3,598,224 bp, PX6 – 3,108,267 bp, PX7 – 1,509,785 bp and PXmt (circular possible mitochondrial contig) - 55,939 bp.

2.4 Summary

In conclusion, I used a hybrid approach to assemble a highly complete and chromosome level genome assembly for the *Penicillium x* hyperparasite. I used the Canu assembly software to assemble large contigs from Nanopore long-reads, then polished with Nanopore reads and additional Illumina short-reads. The 30.9Mb assembly consisted of seven linear chromosomes and one circular, possibly mitochondrial contig. This high-quality genomic resource was a key component in the downstream analysis of this project and addressing the two primary aims described in the following chapters.

3 Species level identification of the novel *Penicillium* hyperparasite

3.1 Background

A preliminary phylogenetic analysis performed by Yiheng Hu in this laboratory classified our novel hyperparasitic fungus into the genus *Penicillium. Penicillium* is a large and important genus both historically and economically. Visagie *et al.,* (2014) lists 354 distinct species and many have been described since (ICPA 2018). Alexander Fleming famously discovered a strain of *Penicillium* in 1928 capable of inhibiting bacterial growth through production of a bioactive compound that he named penicillin (Gaynes 2017). This fortuitous discovery dramatically changed medicine and the process of drug discovery. Since then, fungal-derived compounds, many of which have been sourced from the *Penicillium* genus, have proven useful in a variety of human needs. Fleming's discovery of penicillin shifted the focus of medicine from identifying medicinal compounds through *de novo* synthesis of new molecules to identifying medical compounds in microorganisms (Aly *et al.,* 2011).

Investigating the chemical and genetic elements of fungal diversity is important. Fungi are hugely diverse and potentially a rich source of bioactive materials (Beattie et al. 2011). Identifying unknown fungi to the species level is instrumental in applied scientific research and identification of potential bioactive compounds (Raja *et al.,* 2017). Understanding the phylogenetic placement of a fungus gives valuable insight into its genetic composition and affinities. Traditional methods of fungal taxonomy are based largely on the morphological characteristics of fungi, their fruiting bodies and spores (Romanelli *et al.,* 2010). This requires considerable training, expertise and experience, is labour intensive and often leads to species misidentification (Ehgartner *et al.,* 2017). Improvements in DNA sequencing technology have allowed identification of fungi using genetic rather than morphological characteristics (Ko Ko *et al.*, 2011) providing more accurate estimates of genetic relatedness. In DNA-based classification approaches, relatedness between organisms in the phylogeny are determined via pairwise similarity between genetic markers or even whole genome sequences (Delsuc *et al.,* 2005).

Two common approaches to genetic identification are DNA barcoding and phylogenomics. DNA barcoding involves the use of short, standardized gene sequences called marker genes to

differentiate organisms (Hebert & Gregory 2005). Nucleotide sequences in marker genes are compared via alignment and phylogenetic trees are built based on similarity between the sequences. For fungal identification, the most utilised marker gene is the *Internal Transcribed Spacer* or *ITS* region due to its presence and high variability in all fungal species (Schoch *et al.,* 2012). *ITS*-based identification is often complemented with analysis of marker genes such as *calmodulin* (*CaM*), *beta-tubulin* (*benA*) and *DNA-directed RNA polymerase II subunit* (*RPB2*). These genes are also ubiquitous and polymorphic within fungal genomes, making them good points of comparison (Dizkirici & Kalmer 2019). DNA barcoding is used widely due to its simplicity and broad applicability. In addition, the genomes of most fungi have not been sequenced in their entirety, but marker gene sequence data is more widely available *(Shen et al.,* 2021). This makes DNA barcoding a rapid and accurate approach for fungal identification. However, its accuracy and reliability are surpassed by the second major approach, phylogenomics (Kuramae *et al.,* 2006). Instead of using 1-4 marker genes to differentiate species, phylogenomics takes advantage of the entire genome to compare large datasets. This can be via whole-genome alignments which are computationally intensive, or by examining a subset of conserved genes such as BUSCOs (Waterhouse *et al.,* 2018). BUSCOs are single copy gene orthologs, highly conserved between species. Implementation of phylogenomics strategies is held back by data availability. While advances in DNA sequencing technology have made genome sequencing and assembly more accessible, high quality complete genome sequences are a rarity. For many applications DNA barcoding remains a superior classification method due to much greater data availability (Xu 2016).

In this chapter I describe how I used genetic approaches to identify the novel *Penicillium* hyperparasite to species level. I begin by describing the results of a phylogenomic approach using the newly-assembled genome. However, this approach was limited by the data available for comparison, hence I changed to a more focussed approach using widely-available marker gene data. I was finally able to confirm the proposed phylogenetic placement via a isolate-intensive analysis.

3.2 Materials and Methods

Phylogenomic tree

The first step towards achieving a phylogenomic analysis was to generate BUSCO gene sets for *Penicillium x* and the 21 previously identified *Penicillium* species. This would produce BUSCO gene sets in gene (nucleotide) and protein (amino acid) sequence form. A *Penicillium coffeae x* ANU01 BUSCO gene set had been generated using the BUSCO (v4.1.4) software from Simao et al. (2015) and parameters '-lineage Eurotiales_odb10 –m genome' after genome assembly and polishing. I selected 28 of these BUSCO proteins (4 per linear chromosome) and utilised a BLASTp to compare them against the NCBI non-redundant protein sequences (nr) database (Altschul *et al.*, 1990). From the results, I counted the top ten *Penicillium* results from each search, and from these composed a list of the 21 most frequently returned species. I then downloaded genome sequences from the NCBI genome database for each of the 21 *Penicillium* species, accession codes are available from the Github page for this project. BUSCO gene sets were then generated for each species, using identical parameters to those used for the *Penicillium x* BUSCO gene/protein set '-lineage Eurotiales_odb10 –m genome'. I then took BUSCO output lists and using the Linux 'cat' command generated a file for each species, containing only the names (protein-id's) of the BUSCOs in their respective species' output. I then generated a core protein-id list using the 'comm 12' command, this protein-id list contained the name of all BUSCO proteins present in the BUSCO protein set of all species. More detailed code used in these processes can be found on the Github page for this project. To extract species-specific versions of 250 BUSCO proteins from each species BUSCO output, a bash script containing a for-loop was used. This generated a directory containing 250 BUSCO protein-id subdirectories each containing 22 species specific BUSCO amino-acid sequences for that protein-id. All 5,500 BUSCO proteins were then imported to Geneious Prime® (v2021.1.1) (Geneious Biologics). I then used the default 'Geneious alignment' to generate 250 protein-protein alignments (each species against each other for each protein-id) then concatenated all 250 into one superalignment. This alignment was then used to build a 22-species tree using a Jukes-Cantor Neighbor-Joining method in Geneious Prime® (v2021.1.1). A Jukes-Cantor Neighbor-Joining consensus tree with Bootstrap resampling method and 100 replicates was then built from that tree.

ITS-only tree

ITS-sequences for 417 *Penicillium* species were imported manually to the Geneious Prime® software from the NCBI database using NCBI accession codes gleaned from (ICPA 2018). ICPA (2018) is an updated version of the *Penicillium* species listed in the Visagie et al. (2014) paper. The *ITS* sequence of *Penicillium x* was extracted by first building a custom blast database from the *Penicillium x* genome in linux. The database was built in linux using the blast+ command line options. Due to its relative closeness in the previous phylogenomic tree, the *ITS*-sequence from *Penicillium coffeae* was used as a query sequence to extract the *Penicillium x ITS*. The output *Penicillium x* sequence was then imported directly to Geneious Prime®. An alignment was then generated using all imported *ITS* sequences and the default 'Geneious alignment' algorithm. From this alignment, a phylogenetic tree was built using the Tamura-Nei Neighbor-Joining method. A consensus tree was built from this tree using the Bootstrap resampling method and 100 replicates.

Multiple marker gene tree

The custom *Penicillium x* genome custom blastn database was again used to extract *Penicillium x CaM*, *benA* and *RPB2* sequences. *Penicillium coffeae* versions of all 3 genes were used as queries for extraction, as described in the *ITS* extraction. Exact code is presented in (Github). Marker gene sequences for *Penicillium x* were then imported to the Geneious Prime® (v2021.1.1) software. *CaM*, b*enA* and *RPB2* nucleotide sequences were imported manually to Geneious Prime® for 389, 416 and 350 *Penicillium* species respectively. As with the *ITS* sequence, accession codes for gene sequences were gleaned from (ICPA 2018). Firstly, separate alignments were built for each gene group using the gene versions and the default 'Geneious alignment'. Individual gene trees were then built using Tamura-Nei Neighbor-Joining methods and from these trees, consensus trees were built with the Bootstrap resampling method and 100 replicates. To build a tree that contained all 4 marker genes, marker gene alignments were concatenated together (only for species which possessed all 4, otherwise skews results) and built into a tree using the Tamura-Nei Neighbor-Joining method. A consensus tree was built from this tree with the Bootstrap resampling method and 100 replicates.

Isolate inclusive tree

To build an *ITS* marker gene tree for all *Penicillium* isolates within the clade *Charlesia, ITS* sequences were imported manually to Geneious Prime® from NCBI for all isolates of species:

c*hermesinum, phoeniceum, fellutanum, coffeae, indicum* and *charlesii* (total of 141 sequences). The total 142 isolates (including *Penicillium x*) were built into an alignment using the default 'Geneious alignment' algorithm which was built into a tree using the Tamura-Nei Neighbour-Joining method. A consensus tree was then built from this tree using the Bootstrap resampling method and 100 replicates.

3.3 Results

A phylogeny based on alignment of 250 single copy ortholog genes shows that *P. x* fails to classify with whole-genome sequenced species

I initially used a phylogenomic approach to classify *Penicillium x*, using BUSCO proteins as markers to differentiate species. All organisms in the order Eurotiales, are expected to possess 4,191 near-universal BUSCO genes/proteins in their genome (Simão *et al.*, 2015). To select appropriate species for this analysis, I firstly selected 28 *Penicillium x* BUSCO proteins (four per chromosome) and conducted a BLASTp search of each against the NCBI non-redundant protein sequences (nr) database (Agarwala et al. 2018). From the results, I counted the top ten *Penicillium* results from each search, and from these composed a list of the 21 most frequently returned *Penicillium* species over all searches. I then generated BUSCO gene/protein sets for each of these 21 species using the BUSCO (v4.1.4) software (Table 3.1). Of the 4191 possible BUSCOs, 3785 were present in all of the 22 species (including *Penicillium x*), which I then extracted as a core protein-id set.

I next built a phylogeny from the first 250 core proteins in the 3785 BUSCO gene set. Core protein-ids were used to extract amino-acid sequences of the 250 core proteins from each *Penicillium* BUSCO set. These amino-acid sequences were then concatenated into 22 species specific super-alignments in the Geneious program. These super-alignments were then aligned to one another using the default Geneious alignment algorithm. A 22-species phylogenetic tree was built from this alignment using the Jukes-Cantor Neighbor-Joining method (Figure 3.1). The identities between *Penicillium x* and each species over the 250 proteins are shown in Table 3.1. From the alignments, the closest species to *Penicillium x* was *P. brasilianum* with 69.173% similarity while the least closely related species was *P. digitatum* with 66.356% similarity. *Penicillium x* is positioned on a

branch in a clade containing five other *Penicillium* species. However, *Penicillium x* is not placed closely to these species, instead appearing as if placed in its own subclade. The length of the branch upon which *Penicillium x* sits also suggests its significant genetic distance from even the most closely related species in its clade. The tree was strongly supported in a bootstrap analysis. The low sequence similarity of even the most closely related species to *Penicillium x* and the behaviour of *Penicillium x* in the phylogeny suggests that it is acting as a novel species in this analysis. However, this high-resolution analysis is restricted because it only contains 21 complete genome sequences. This is insufficiently deep to confidently assign *Penicillium x* to a new species.

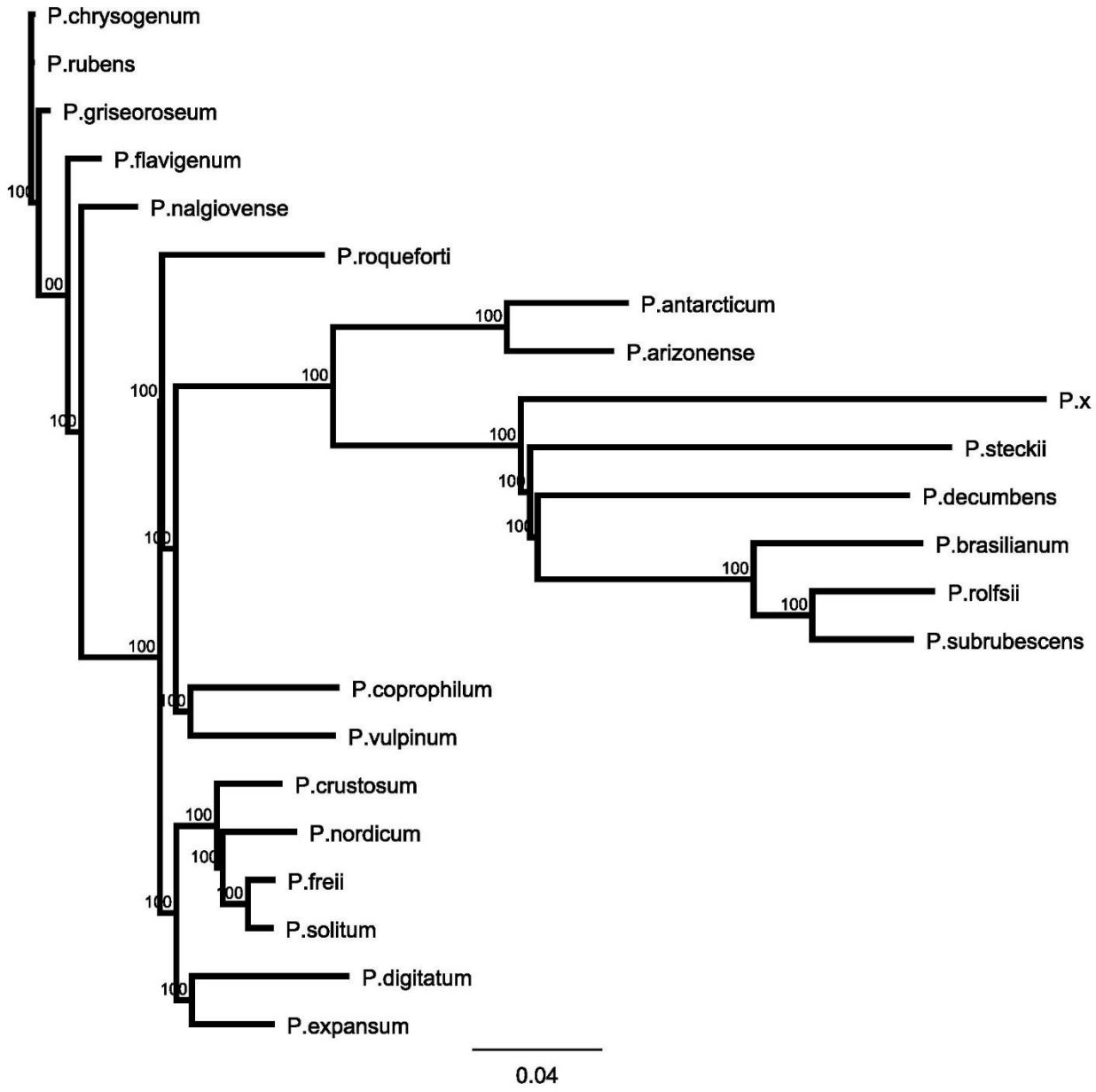

Figure 3.1. A rooted phylogenetic tree containing *Penicillium x* and 21 other *Penicillium* species. The tree was built from Geneious superalignments of 250 BUSCO proteins, generated from a BUSCO analysis with mode genome and lineage Eurotiales. The tree was constructed using the Jukes-Cantor Neighbor-Joining method in Geneious. Numbers at nodes denote bootstrap support values. Tree branch lengths are determined by the number of amino-acid substitutions per site.

Table 3.1. Summary of *Penicillium* species and BUSCOs used for phylogenomic analysis.

Penicillium species	No. of complete BUSCOs identified	Similarity of 250 selected BUSCOs to those of *P. x*
P. antarcticum	4177	66.674
P. arizonense	4180	66.966
P. brasilianum	4172	69.173
P. chrysogenum	4185	66.764
P. coprophilum	4182	66.598
P. crustosum	4181	67.239
P. decumbens	4147	68.779
P. digitatum	4150	66.356
P. expansum	4187	66.769
P. flavigenum	4189	66.964
P. freii	4162	66.892
P. griseoroseum	4184	66.913
P. nalgiovense	4181	67.102
P. nordicum	4176	66.775
P. rolfsii	4172	68.921
P. roqueforti	4181	66.645
P. rubens	4187	67.140
P. solitum	4182	66.801
P. steckii	4158	68.276
P. subrubescens	4176	69.404
P. vulpinum	4183	66.862

A phylogeny built from ITS sequences shows that *Penicillium x* is closely related to *P. coffeae* and *P. phoeniceum*

To expand the species set against which *Penicillium x* could be classified, I used alignments based on the highly polymorphic *ITS* region to build a phylogenetic tree. Due to its closeness in the phylogenomic approach above, the *P. brasilianum ITS* region was used to extract that of *Penicillium x* in a custom BLAST search. The sequences were 554 bp in length with an identity of 92%. I then downloaded *ITS* sequences for 417 *Penicillium* species from NCBI and aligned them using the default 'Geneious alignment'. These sequences were 724 bp in length on average. A tree built on the alignment of these sequences (data not shown) placed *Penicillium x* close to *P. coffeae* within the *Charlesia* clade. To investigate this further, I used the *P. coffeae ITS* to extract the complete (1076 bp) *ITS* sequence from *Penicillium x* and rebuilt the tree as previously. The

phylogenetic tree built from this alignment using the Tamura-Nei Neighbor-Joining method placed *Penicillium x* in the clade *Charlesia* close to *P. coffeae* and *P. phoeniceum* (Figure 3.2) but was not fully supported by bootstrap analysis. The tree presented in Figure 3.2, does not have bootstraps labelled, however *Penicillium x* was placed on the node with *P. coffeae* in 62/100 bootstraps, a relatively lowly supported tree. Exact *ITS* identities of *Penicillium x* to other *Penicillium* spp. in the clade *Charlesia* are shown in Table 3.2. Importantly, no discernible difference between the trees was evident when they were built with alternative methods such as Jukes-Cantor Neighbor-Joining, data not shown.

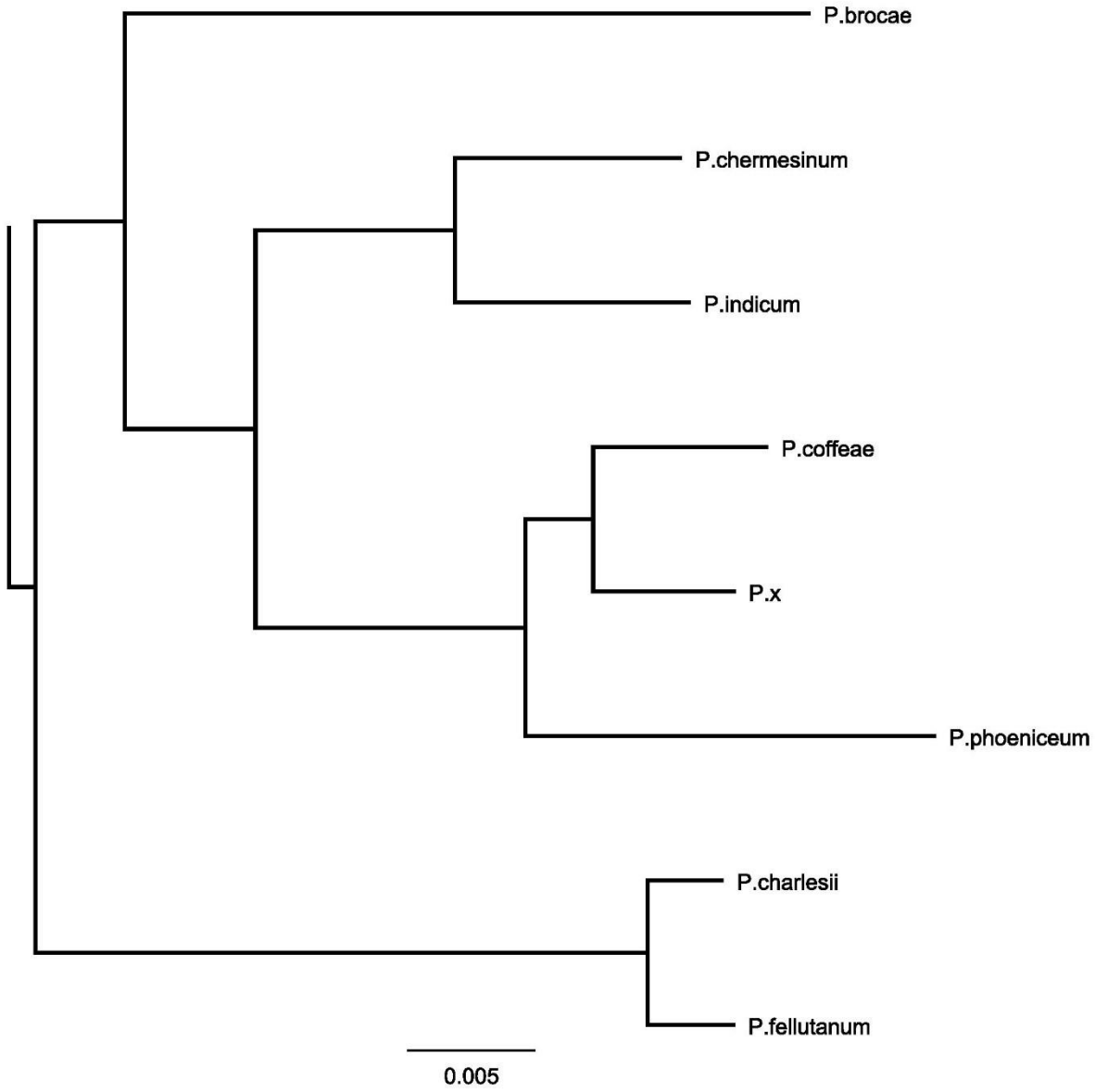

Figure 3.2. A node within a larger *Penicillium* spp. phylogenetic tree, containing *Penicillium x* and 7 other *Penicillium* species belonging to the clade *Charlesia*. The tree was built from a Geneious alignment of *Penicillium* spp. *ITS* sequences that were downloaded from the NCBI nr database. The tree was constructed using the Tamura-Nei Neighbor-Joining method in Geneious. Tree branch lengths are determined by number of substitutions per site. *Penicillium x* was placed adjacently to *P. coffeae* in 62/100 bootraps.

Expanding the phylogeny to include three additional marker genes supports the relatedness of *Penicillium x* to *P. coffeae* and *P. phoeniceum*

ITS-based identification is often complemented with other secondary fungal marker genes for more accurate and reliable phylogenies (Dizkirici & Kalmer 2019). I queried the *ITS* placement in Figure 3.2 by including sequences of the marker genes *benA, CaM* and *RPB2* into the alignment matrix. The nucleotide sequences were downloaded from Genbank for each marker gene for all *Penicillium* species available (Agarwala et al. 2018). These included 417 species for *benA*, 318 for *CaM* and 315 for *RPB2*. I then built superalignments from concatenated *ITS-benA-CaM-RPB2* sequences for 312 *Penicillium* species and constructed a phylogeny using the Tamura-Nei Neighbor Joining method and then generated a consensus tree with 100 bootstraps, which is shown in Figure 3.3. Exact values of sequence similarities between *Penicillium x* and all other members of the clade *Charlesia* are shown in Table 3.2. Inspection of Figure 3.3 revealed that the relationships found in the previous Figure were broadly similar but are now supported by robust bootstrap values. Importantly, these results supported the placement of *Penicillium x* within the clade *Charlesia*. This clarifies the data in Table 3.2 which show *Penicillium x* to be closely related to *P. phoeniceum*. The closeness between and *Penicillium x* and *P. coffeae* in this phylogeny was such that I hypothesised *Penicillium x* may be a novel isolate of *P. coffeae*. However, a final analysis was required to distinguish whether *Penicillium x* is a novel *P. coffeae* isolate or is a closely related species.

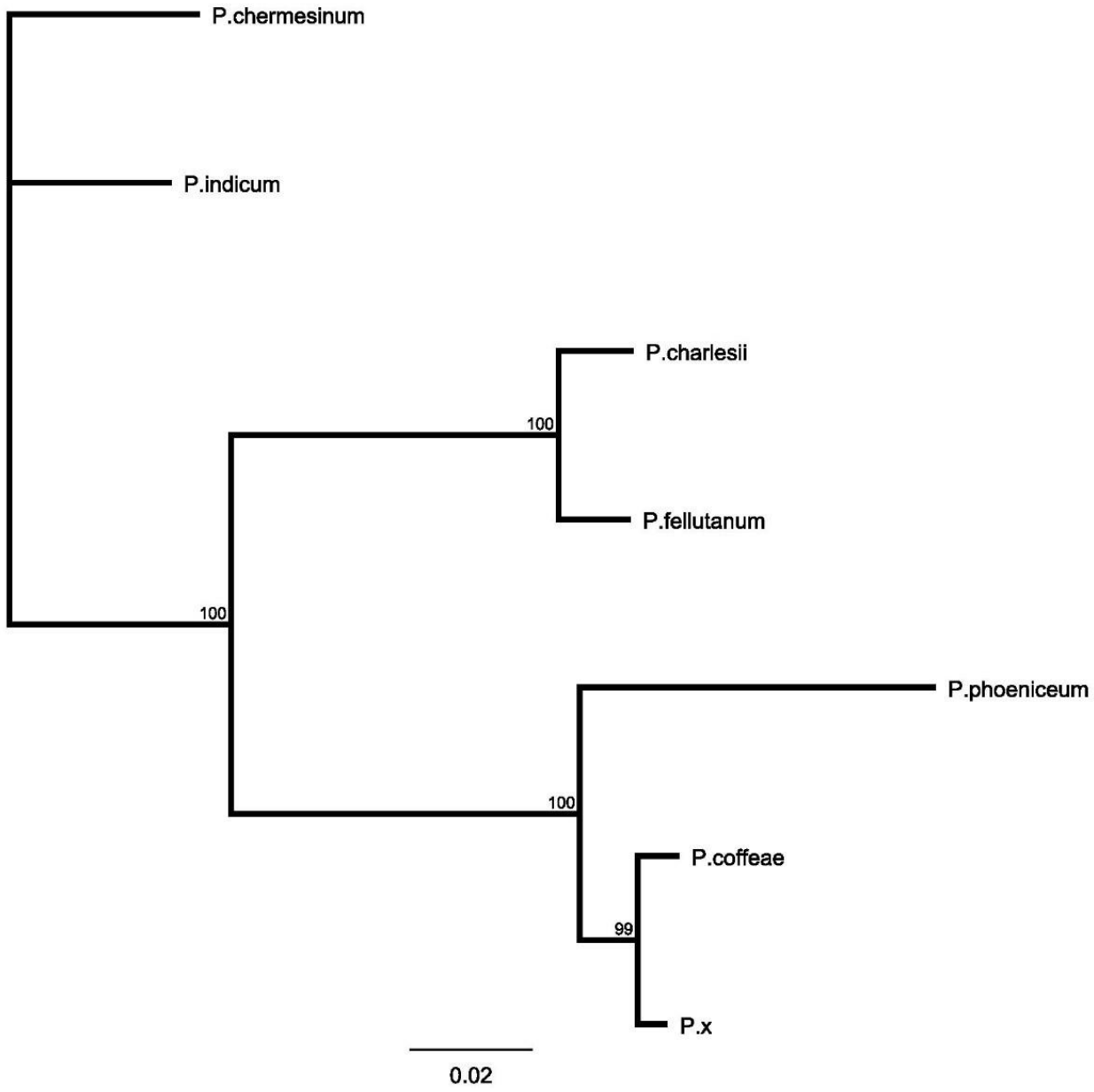

Figure 3.3. A node within a larger rooted phylogenetic tree containing *Penicillium x* and species within the clade *Charlesia*. This tree was built from Geneious superalignments of *Penicillium* spp. *ITS*, *benA*, *CaM* and *RPB2* sequences downloaded from the NCBI Genbank database. The tree was constructed using the Tamura-Nei Neighbor-Joining method in Geneious. Numbers on nodes denote confidence rating output of bootstrapping statistical analysis. Tree branch lengths are determined by the number of substitutions per site.

Table 3.2. Summary of marker gene sequence similarities for *Penicillium* species in the clade *Charlesia* with respect to *Penicillium x* orthologs.

	Sequence similarity to *P. x* marker region (%)				
Charlesia species	***ITS***	***benA***	***CaM***	***RPB2***	**Marker gene combination**
P. charlesii	89.833	80.769	71.973	85.371	83.011
P. chermesinum	91.251	77.273	71.724	84.899	81.01
P. coffeae	94.542	98.1	96.793	98.433	96.656
P. fellutanum	89.442	79.515	71.995	85.162	83.032
P. indicum	91.783	76.271	72.099	86.242	80.839
P. phoeniceum	97.804	91.706	87.41	92.617	89.082

Including all known *Charlesia* isolates in marker gene phylogeny confirms *Penicillium x as* a isolate of the plant endophyte *Penicillium coffeae*

To determine whether *Penicillium x* was an isolate of *P. coffeae*, I expanded the analysis to include all known isolates of *Penicillium* within the clade *Charlesia* for which *ITS* sequences were available. All NCBI available *ITS* sequences were downloaded for a total of 148 isolates, including 14 *P. coffeae* isolates. All *ITS* sequences were aligned in Geneious and a phylogeny produced using the Tamura-Nei Neighbor-Joining method. A node of this phylogeny that illustrates placement of *Penicillium x*. is presented in Figure 3.4. A 100-bootstrap consensus tree was built from this, data not shown, in which *P. x* was placed on a node with *P. coffeae* with only 51 bootstraps. Observation of this Figure supports the following observations. There are three major clades within the node, representing *P. chermisinum*., *P. coffeae* and *P. phoeniceum* respectively, with a possible bifurcation in the *P. coffeae* clade. There are some aberrant placements of individual isolates within this scheme, for example *P. coffeae* NIHHS317 with the *P. phoeniceum* clade, and *P. fellutanum* CBS118477 and *P. coffeae* DTO 273-A7 within the *P. chermesinum* clade. Within the *P. coffeae* clade comprising 14 isolates, *Penicillium x* was placed closer to the isolate MA-314 than MA-314 was to any other *P. coffeae* isolate. As was apparent with the original *ITS* analysis displayed in Figure 3.2, the analysis was not well supported by bootstraps when built into a consensus tree, (data not shown). Previously, including three further marker genes resolved this ambiguity (Figure 3.3). However, this analysis was not possible as marker gene sequences were not available for all isolates. Though it would be desirable to complement this analysis with further data, the placement of *Penicillium x* within this clade still strongly supports its classification as an isolate of *P. coffeae*.

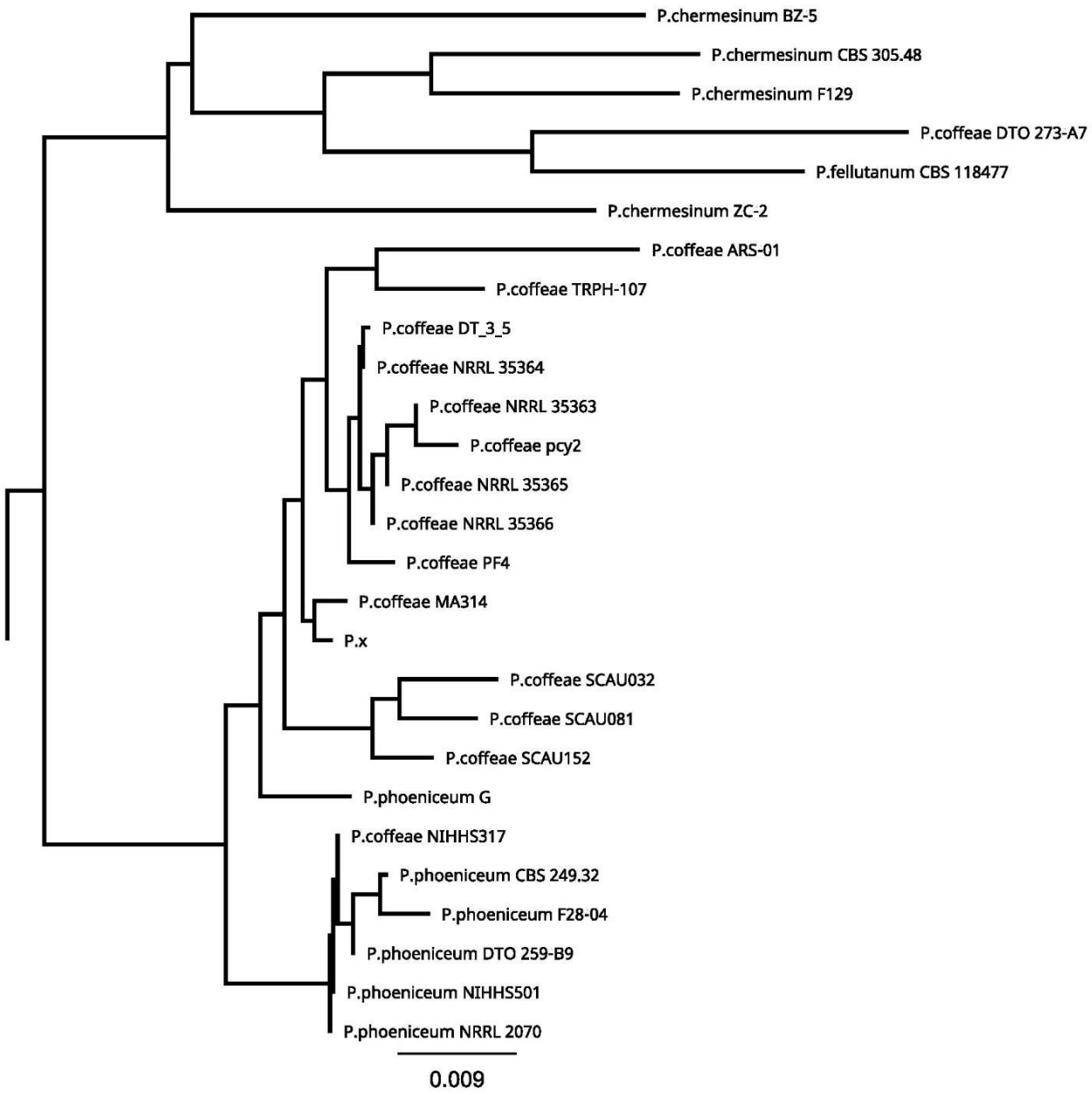

Figure 3.4. A node of a larger rooted phylogenetic tree of all available isolates of *Penicillium* in the clade *Charlesia* including *Penicillium x*. The tree was built from Geneious alignments of ITS sequences downloaded from the NCBI database. The tree was constructed using the Tamura-Nei Neighbor-Joining method in Geneious. Tree branch lengths are determined by the number of substitutions per site. Bootstrap values not shown, however not strongly supported.

3.4 Summary

Using genetic approaches, I achieved the first aim of this project, to confirm the hyperparasite as a species of *Penicillium* and to determine its taxonomic placement within the genus. An initial phylogenomic approach failed to classify the novel hyperparasite due to a lack of full genome data available for *Penicillium* species. Switching to a marker gene approach, I was able to use the *ITS* and three other polymorphic gene markers to place the hyperparasite in the clade *Charlesia*. Finally, I used an *ITS* approach including all available *Penicillium* isolates in the clade *Charlesia*, which determined that the novel hyperparasite is likely an isolate of the plant endophyte *Penicillium coffeae*. Based on this information, the novel *Penicillium* isolate was arbitrarily designated *Penicillium coffeae* ANU01 and will be referred to as such in the remainder of this thesis.

4 Differential gene expression during growth of *Penicillium coffeae* ANU01 on synthetic media and *Pst* spores

4.1 Background

The second primary aim of this project was to investigate the genetic programs used by *Penicillium coffeae* ANU01 in its hyperparasitism of *Pst* spores. A major hypothesis for this aim was that one or more biosynthetic gene clusters were important to the hyperparasitism. Although the study of fungal hyperparasitism is growing, most research in this area is focussed on infection phenomena and taxonomy rather than the molecular biology of infection. During this project several papers were published on newly discovered hyperparasites of rust fungi, however none identified potential mechanisms of hyperparasitism.

Fungal hyperparasitism is often broken down into two major strategies; parasitism with antibiosis, and direct parasitism (Benhamou *et al.*, 1999). Antibiosis involves the production of antifungal secondary metabolites to inhibit or kill the targeted host, without directly entering it. For example, several *Trichoderma* species use antibiosis to disable the plant pathogen *Botrytis cinerea* before feeding on it (Jeleń *et al.* 2014). Direct parasitism commonly involves the physical invasion of the host via hyphae and can be complemented by production of lytic enzymes or antifungal metabolites (Huang & Kokko 1987).

Little is known of the methods of parasitism used by the few known rust hyperparasites. A recent paper by Wilson et al. (2020) postulated that an isolate of *Penicillium brevicompactum* that antagonises *Puccinia triticina* and *Puccinia hordei* does so via antibiosis, as little colonisation of rust pustules was observed. Alternatively, Zheng et al. (2017) and Zhan et al. (2014) describe the colonisation and what appears to be direct hyphal penetration of *Pst* spores by the fungi *Alternaria alternata* and *Cladosporium cladosporioides* respectively. Casual observation of *Penicillium coffeae* ANU01 growth on *Pst* pustules *in planta* reveals extensive colonisation of rust pustules (Fig 1.1). One might expect that *Penicillium coffeae* ANU01 could act similarly to *P. brevicompactum*, by producing a SM or enzyme to inhibit rust spore germination without

colonising it. However, it is entirely possible that *Penicillium coffeae* ANU01 deploys both antibiosis and direct invasion in its hyperparasitism

A powerful method for identifying mechanisms underlying a developmental or pathogenic state of an organism is the use of functional genome annotation and comparative transcriptomics (Mahmood et al. 2020). Functional annotation of the genome is a preliminary step in understanding the genetic tools it possesses, while comparative transcriptomics helps to identify which tool(s) it uses for a given task. In functional annotation, genes are first predicted in the genome, either *de novo* or with the help of RNAseq data (Keilwagen *et al.,* 2018). The genes are then compared to specific gene and protein databases to find evidence of the gene's function based on homology (Han *et al.,* 2016). Comparative transcriptomic analysis compares gene expression between biological states (Zhang *et al.,* 2020). This can provide important insights into the genetic mechanisms used by an organism in a novel or cryptic process.

In this Chapter, I describe the generation of a highly complete gene set for the *Penicillium* hyperparasite and its functional annotation. I will describe how I used the newly annotated genome in a comparative transcriptomic analysis to compare gene expression of the *Penicillium* when grown on synthetic media and on *Pst* spores. Finally, I will report how I combined the results of the functional annotation and comparative transcriptomic analysis to generate a list of candidate genes presumed to be important in hyperparasitism. Genes involved in carbohydrate hydrolysis (CAZYs) and BGCs were present in the most upregulated genes.

4.2 Materials and methods

Penicillium coffeae isolate ANU01 was initially discovered by Dr Sambasivam Periyannan of this laboratory in the Controlled Environment Facility at ANU and used for all parts of this project. The fungus was grown from a spore stock on PDA plates (25g/L potato dextrose broth, 15g/L agar) in 22°C growth rooms under 12-hour day/night cycles.

The wheat (*Triticum aestivum*) cultivar (cv.) *Morocco* was used for this aspect of the project. All wheat used was grown in controlled environment chambers under a 16hr/8hr day/night cycle, with a light intensity of 100 µmolm-2 s-1.

Growing and infecting plants for *in planta* RNA cultivation

Twelve pots each containing 15 uniform wheat cv. *Morocco* seedlings were treated with 200 ml (over all pots) 0.5% (v/v) maleic hydroxide at 6 days post planting. At 8 days post planting, approximately 0.5 g of *Pst*-104E spores were suspended in Novec^{TM}7100 (3M) and applied evenly to seedling leaves via direct application with a paintbrush. Seedlings were moved to a plastic container and sprayed with autoclaved H_2O to create 100% humidity. The container was then closed and placed into a large autoclave bag to ensure high humidity and put in an 8°C cabinet for 48 h to promote *Pst* germination and infection of the leaves. After 48 h, plants were removed from the container, moved to a new growth chamber and kept under a 16 h/8 h day/night cycle, with a light intensity of 100 µmol m^{-2} s^{-1}. After approximately 14 days post infection (14 dpi), all plants were sprayed with a *Penicillium coffeae* ANU01 spore suspension made as follows: Three PDA plates containing mature *P. coffeae* ANU01 colonies (~4.5 cm diameter) were flooded with sterile H_2O and a sterile razor blade used to scrape spore and mycelial tissue from the colony surfaces. Fungal debris was filtered from the spore suspension using Miracloth (Merck Millipore) and the filtrate transferred to a sterile spray bottle. This spore suspension was then sprayed evenly onto the stripe rust pustules on the infected wheat leaves. The sprayed pots were covered individually with transparent plastic sleeves to create a humid atmosphere necessary for hyperparasitism (Zhan *et al.*, 2014).Fluffy white tissue indicative of *P. coffeae* ANU01 growth was observed on the surface of rust spores at 8 dpi. For sampling, the pots were split into three groups of four to create three biological replicates. I photographed and documented those leaves that showed signs of *Penicillium* growth and harvested *Penicillium coffeae* ANU01 tissue 24 h later (9 dpi) as described below.

Harvesting fungal issue for RNA extraction:

For *P. coffeae* ANU01 grown *in planta* samples, spore and mycelial tissues of *P. coffeae* ANU01 were scraped from the surface of *leaves* (it was difficult to scrape gently enough to not pick up the *Pst* pustules under the *Penicillium, so* a large quantity of *Pst* was also incidentally collected) into 2 ml Eppendorf tubes (Eppendorf) using a sterile razorblade (Swann-Morton). The tissue was immediately snap frozen in liquid nitrogen, and after three days of collection, harvests from each replicate were combined into three fresh Eppendorf tubes (one per replicate), each containing 2 clean steel ball-bearings. For *in vitro* samples, spore and mycelial tissues were scraped from the surface of mature (~4.5cm diameter) *P. coffeae* ANU01 colonies cultured on PDA into 2 ml Eppendorf tubes containing two clean steel ball-bearings. Three plates each containing a single *P. coffeae* ANU01 colony were used for the three biological replicates. Tubes were immediately snap frozen in liquid nitrogen. All samples were kept at –80ºC until ready for processing.

RNA extraction

All 6 tissue samples were processed identically. Tissue was disrupted using the TissueLyser LT (QIAGEN) for four rounds of 30 seconds at 25 Hz, refreezing the chamber containing the samples in liquid nitrogen between each round of disruption. Samples were then processed using the RNeasy® Plant Mini Kit (QIAGEN) following the manufacturer's protocol. RNA samples were eluted in 32 µl RNAse free water supplied with the kit. Concentration and absorbance purity ratios of eluted RNA were assessed using a Quickdrop™ Micro-Volume Spectrophotometer (SpectraMax®). Intactness of the RNA was assessed by running a sample on a 1% agarose gel, as per (Green & Sambrook 2019). Half of the volume of each sample was transferred to an individual RNA Stabilization Tube (GeneWiz). All 6 samples were then posted to GeneWiz (Suzhou, China) for library preparation and Illumina sequencing. The 20 million reads per sample option was chosen prior to sending samples.

Genome annotation using the funannotate software.

Illumina RNAseq reads were downloaded from the GeneWiz server and stored as gzipped fastq files on a local computer. Reads were trimmed of adapters using the Trimmomatic (v 0.39) software (Bolger *et al.,* 2014). With the assistance of Dr Benjamin Schwessinger (Research School of

Biology, ANU) the *P. coffeae* ANU01 genome was functionally annotated using the funannotate (v1.5.3) software (Funannotate 2020). Inputs included the genome fasta file and mRNA read files for all replicates and. This generated a variety of outputs. The mRNA-transcripts.fasta file contains all predicted transcripts for the genome. This file was then used as an index in the next step, a comparative transcriptomic analysis.

Comparative transcriptomic analysis using Salmon and Limma

The Salmon (v.1.4.0) software from Patro *et al.,* (2017) was used to quantify transcript expression in RNAseq samples. Firstly, the salmon index function was used to generate a salmon index file from the funannotate transcripts 'Penicillium_X.mRNA.transcripts.fa' file. The 'salmon quant' command with parameters '-l A –validateMappings' was then used to align and quantify trimmed RNAseq reads against the index. Salmon outputs were imported into an R-studio session for comparative transcriptomic (differential gene expression) analysis using the Limma package (Ritchie *et al.,* 2015). All code used to generate the differential gene expression Venn diagram (Figure 4.1) and comparative transcriptomic analysis (Volcano plot Figure 4.2) are presented in the github page for this project.

4.3 Results

Purification of mRNA from *P. coffeae* ANU01 grown axenically and on *Pst*-infected wheat leaves

The developmental and pathogenic states of fungi are determined by its transcriptome. To study this, I generated transcriptome profiles of the fungus growing axenically or on *Pst*-infected wheat leaves. Differentially expressed genes, especially those upregulated during hyperparasitism, should be related to pathogenesis. I purified mRNA from *Penicillium coffeae* ANU01 grown axenically on synthetic media (*in vitro*) and on *Pst* that was sporulating on wheat (*in planta*). Tissue was harvested from *in vitro* replicates when *Penicillium* colonies were large (~4.5 cm diameter), and 24 h after white fluffy tissue appeared on *Pst* pustules for *in planta* growth. Each condition was represented by three biological replicates. *Penicillium* tissue was harvested from wheat leaves by scraping with a sterile razor blade and placing the biological material into a clean Eppendorf tube.

I estimate that only ~40-50% of the total tissue harvested was *Penicillium* due to the difficulty of separating *Penicillium* and *Pst* tissue when scraping. I extracted RNA from all 6 using a RNeasy plant mini-kit. The concentration and purity of eluted RNA was determined using the Quickdrop spectrophotometer. The mean concentration of *in planta* samples (654 ng/µl) was significantly higher than that of *in vitro* grown samples (128.8 ng/µl). Mean A260/230 values of 1.923 (*in vitro*) and 2.57 (*in planta*) and mean A260/280 values of 2.155 (*in vitro*) and 2.298 (*in planta*) suggested significant impurities in both samples. However, the RNA was highly intact in all samples when run on a 1% agarose gel, data not shown. I sent all RNA samples to GeneWiz (Suzhou, China) for library preparation and Illumina RNA-sequencing (RNA-seq). All samples returned high quality read data. A mean raw read count of 31,342,137 for *in vitro* samples and 32,524,399 for *in planta* samples indicated two large datasets for analysis. The quality score of raw reads indicated that over 97% of reads were above Q20 (phred score 20) in all 6 samples, indicating high quality. Thus, I successfully generated substantial RNA-seq data for each growth type of *Penicillium coffeae* ANU01.

Genome annotation with the funannotate pipeline

I annotated the *P. coffeae* ANU01 genome using the funannotate pipeline with the help of with Dr Benjamin Schwessinger (RSB, ANU). The funannotate pipeline makes use of both genome and RNA-seq data to generate a high accuracy functional annotation of the genome. Briefly, funannotate passes the input data through a series of gene predictors, then passes the output of those predictors to the Evidence Modeler (EVM) software. The EVM then generates a consensus gene set by ascribing different weights to each input and determining the most correct gene models from each. A total of 64,456 gene models were passed to the EVM modeler producing a final gene set of 12,181 genes, coding for 12,732 mRNA transcripts and 118 tRNA transcripts. Using funannotate, the completeness of the *Penicillium coffeae* ANU01 gene set was assessed using the BUSCO Software, and the Dikarya gene set (Simão *et al.*, 2015). A total of 1,276 (97.2%) complete BUSCOs were identified of the 1,312 total BUSCO genes searched for. A complete breakdown of BUSCO analysis of completeness of the predicted gene set is shown in Table 4.1. This indicates a highly complete gene set, especially for a novel isolate (Veeckman *et al.*, 2016).

Functional annotation of the *Penicillium coffeae* ANU01 gene set

Functional annotation uses homology to known genes, proteins and conserved domains to predict the function of members of a large gene set (Griesemer *et al.,* 2018). The funannotate pipeline uses a variety of functional annotation software for this purpose. A summary of the software included in this pipeline and their respective output annotation counts are presented in Table 4.1. Briefly, eggNOG from (Huerta-Cepas et al. 2019) produced the most annotations, ascribing a homolog to 11,666 (91.6%) of the 12,732 predicted proteins. Gene ontology (GO) terms assign a known function to a predicted gene product based on homology to genes of known function (e.g. carbohydrate metabolism) (Hill *et al.,* 2008). A total of 7,684 GO-terms were given to the gene set. MEROPS gene annotation identifies peptidases, enzymes that breakdown proteins, and are believed to be involved in microbe-microbe interactions (Schönbichler *et al.,* 2020). Many MEROPS and CAZY enzymes were identified in the genome annotation, 388 and 426 respectively. A total of 973 proteins were identified to contain a secretion peptide (SP) indicating potential extracellular secretion.

Table 4.1. funannotate functional annotation for the *Penicillium coffeae* ANU01 genome

Genome	**10kb8QCanu**
Genome assembly size	30.96 Mbp
Final gene set	
Genes	12,181
Transcripts (mRNA)	12,732
Transcripts (tRNA)	118
Gene set completeness (BUSCO analysis)	97.2%
Total BUSCO groups searched	1312
Complete BUSCOs	1276 (97.2%)
Complete and single-copy BUSCOs	1235 (94.1%)
Complete and duplicated BUSCOs	41 (3.1%)
Fragmented BUSCOs	30 (2.3%)
Missing BUSCOs	6 (0.5%)
Functional annotation	
GO_terms	7,684
Interproscan	10,240
Eggnog	11,666
Pfam	9,207
Secreted products	973
Cazyme products (carbohydrate-active enzymes)	426
Merops products (peptidases)	388

A major hypothesis of this project was the importance of BGCs in the antagonism of *Pst* spores by *Penicillium coffeae* ANU01. Funannotate antiSMASH predicted 43 BGCs. The exact breakdown of BGCs were as follows; 13 T1PKS, 10 NRPS, 8 terpenes, 8 NRPS-like, 2 indoles, 1 betalactone and 1 fungal RiPP. It is not possible to describe all BGCs given space limitations; instead, I will describe the BGCs identified as upregulated in the comparative transcriptomic analysis below.

Differential expression of *Penicillium coffeae* ANU01 genes during hyperparasitic versus axenic growth

I first conducted a preliminary analysis to determine whether there was a significant difference in the gene expression profiles of *Penicillium coffeae* ANU01 grown *in vitro* and *in planta.* To do this, I separately aligned mRNA reads from each condition against the *Penicillium coffeae* ANU01 transcriptome and quantified gene expression using the Salmon software. Average gene expression value was then determined from the three replicates for each condition. I decided to define a gene as expressed if if present in the sample at an abundance of >100 transcripts per million, normalised to gene transcript length (length-scaledTPM). The results of expressed gene counts for each condition are presented in Figure 4.1 as a Venn diagram. The figure shows that 990 genes were only expressed *in vitro* while 439 were only expressed *in planta*, 9,124 genes were expressed in both conditions.

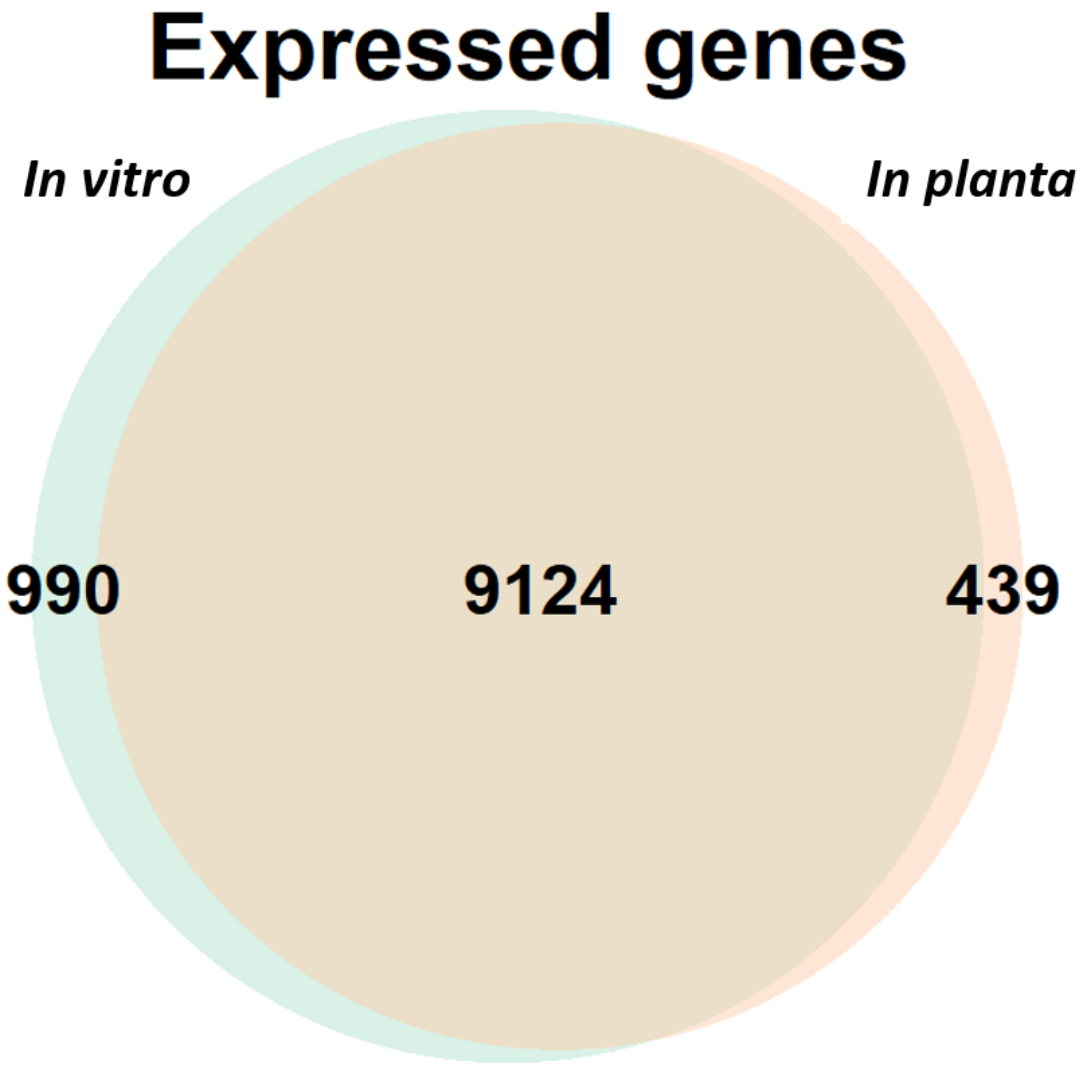

Figure 4.1 A Venn diagram to display the number of genes expressed (>100 length-scaledTPM) only on synthetic media (*in vitro*), only on the surface of *Pst* spores (*in planta*) and expressed in both conditions (centre).

To identify the specific *Penicillium coffeae* ANU01 genes upregulated during hyperparasitism, I conducted a comparative transcriptomic analysis by comparing gene expression when grown on *Pst* spores as a substrate relative to synthetic media. Unlike the previous analysis, the level of gene expression was considered in this analysis, and a gene was considered either upregulated or downregulated based on its degree of expression *in planta* relative to *in vitro.* Quantified transcript counts from Salmon were loaded into the Limma package in R, which calculated relative gene

expression values. The output of the comparative transcriptomic analysis is shown in the form of a Volcano plot in Figure 4.2. A gene was defined as differentially expressed if it exhibited an adjusted P-value < 0.05, being considered upregulated if it also possessed a log fold change value >1 and downregulated if it possessed a log fold change of < -1. A total of 1,146 genes were upregulated and 2,313 genes were downregulated in the *in planta* condition. This can be seen in Figure 4.2, where every dot-point represents a gene and its position representing its relative up or down regulation. Dot points above the cut-off on the right represent upregulated genes *in planta,* while those on the left above the cut-off represent genes that were downregulated *in planta*.

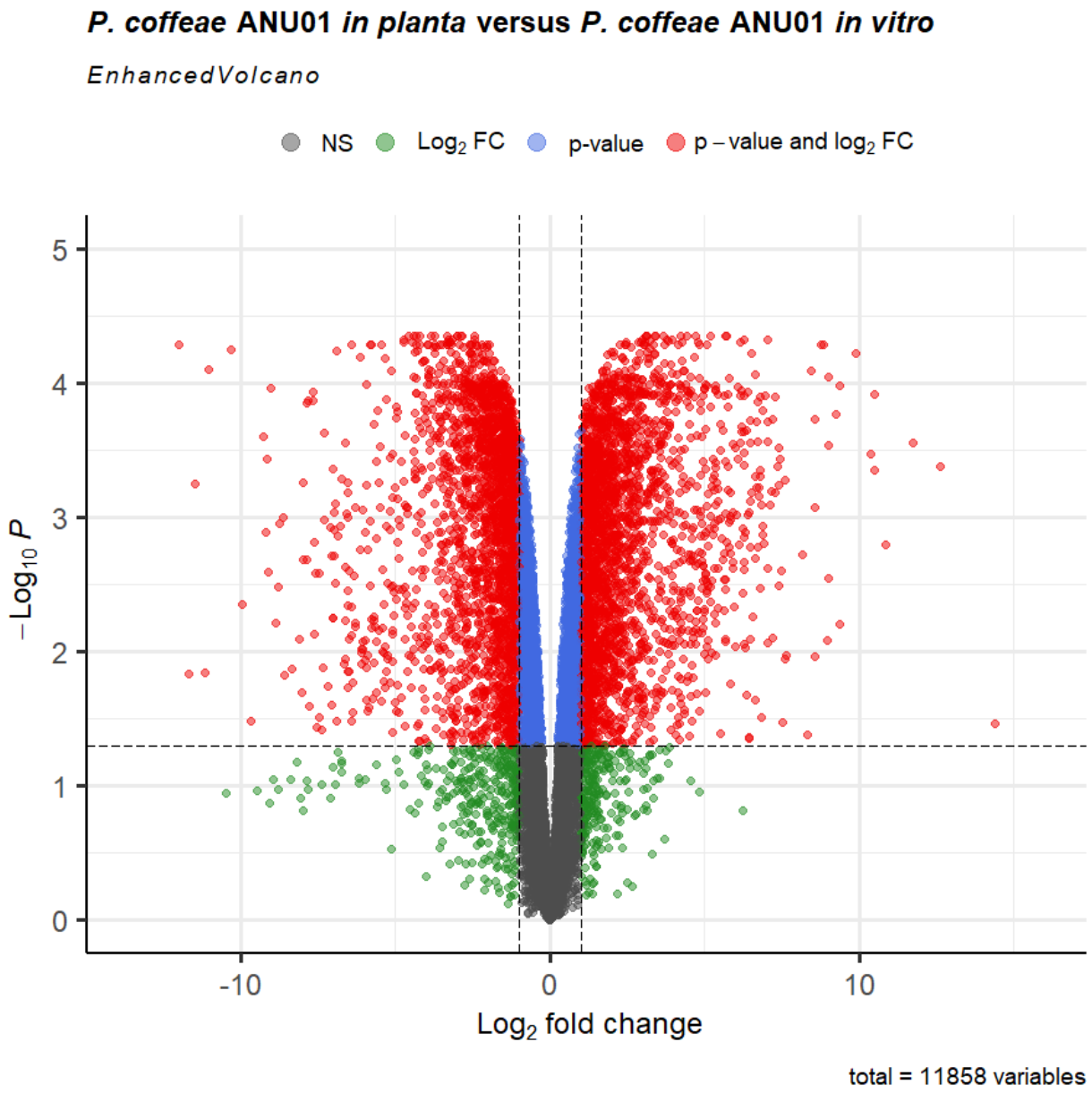

Figure 4.2 Volcano plot of differentially expressed genes in in planta vs in vitro samples. Dashed line on y axis indicates adjusted p value of 0.05, Dashed lines on x axis indicate log fold change of -1 and 1 respectively. Red dots represent genes considered differentially expressed, downregulated on the left and upregulated on the right.

Potential antifungal properties of genes highly upregulated during hyperparasitic growth

Combining the results of functional annotation and comparative transcriptomics allowed identification of several candidate genes that may underlie hyperparasitism. I compiled the 30 most upregulated genes *in planta* in Table 4.2. Examination of Table 4.2 allows a preliminary understanding of the types of genes presumed to be involved in hyperparasitism. Eleven suggest enzymatic function, of which four are CAZY enzymes, four are transmembrane proteins and ten contain secretion factors. I selected five genes that I propose as candidate genes of importance in hyperparasistism that I would choose to investigate further if time permitted:

1. *PenX_003989-T2* – the most upregulated gene, which encodes an efflux pump involved in transmembrane transport, and component of a T1PKS biosynthetic gene cluster of unknown function.
2. *PenX_011559-T1* – the fourth most upregulated gene which encodes a GH75 CAZY chitosanase enzyme containing a secretion peptide. Chitin and chitosan are integral components of the *Pst* spore coat suggesting potential substrates (El Gueddari et al. 2002).
3. *PenX_003238-T1* – This gene encodes a secreted GH43 CAZY enzyme that may mobilise carbohydrate substrates in *Pst* spores.
4. *PenX_007638-T1* – Encodes a sugar transporter-like protein, however when compared against the nr protein database using blastp, was found to likely be *STL1P*, a glycerol transporter. This is a logical candidate for uptake of glycerol derived from metabolism of the fat bodies that are abundant in *Pst* spores Staples & Wynn (1965).
5. *PenX_011498-T1* – The third most upregulated gene encoding a protein with no annotation other than a secretion peptide. Because this protein is not annotated, it offers the potential for novel discoveries when investigating its role in parasitism.

A list of the 30 most downregulated genes is presented in Table 4.3 The majority of downregulated genes appear to be involved in energy metabolism, as PDA is likely a richer nutrient source than *Pst* spores.

Table 4.2. The 30 most upregulated genes (and their functional annotation) in *Penicillium coffeae* ANU01 grown on its putative host *Pst*, compared to *Penicillium coffeae* ANU01 grown on synthetic media

Ranking	Gene_ID	Log2Fold-change	Predicted product	GO_function	GO_process	GO_localisation	CAZy	Secreted
1	PenX_003989-T2	14.38194	Efflux pump dep3 variant 2	Transmembrane transporter activity	Transmembrane transport	-	-	-
2	PenX_004159-T1	12.63861	Hypothetical protein	-	-	-	-	-
3	PenX_011498-T1	11.72882	Hypothetical protein	-	-	-	-	SignalP(1-34)
4	PenX_011559-T1	10.84043	Hypothetical protein	Chitosanase activity	-	-	GH75	SignalP(1-16)
5	PenX_010067-T1	10.50319	Hypothetical protein	-	-	-	-	SignalP(1-24)
6	PenX_007731-T1	10.48417	Peroxiredoxin prx5	Transmembrane transporter	Transmembrane transport	-	-	-
7	PenX_007985-T1	10.37541	Hypothetical protein	-	-	-	-	SignalP(1-18)
8	PenX_007733-T1	9.898196	Short-chain dehydrogenase/reductase prx7	Oxidoreductase activity	Oxidation-reduction process	-	-	-
9	PenX_000861-T1	9.386359	Hypothetical protein	-	-	-	-	SignalP(1-22)
10	PenX_004154-T1	9.381142	Hypothetical protein	-	-	-	GH64	SignalP(1-22)
11	PenX_001079-T1	9.251989	Hypothetical protein	Oxidoreductase activity	-	-	-	-
12	PenX_005744-T1	9.009765	Hypothetical protein	Catalytic activity	-	-	-	-
13	PenX_009423-T1	9.005029	Hypothetical protein	Heme binding, iron ion binding, oxidoreductase activity, acting on paired donors, with incorporation or	Oxidation-reduction process	-	-	-

				reduction of molecular oxygen				
14	PenX_003238-T1	8.993831	Hypothetical protein	Hydrolase activity, hydrolyzing O-glycosyl compounds	Carbohydrate metabolic process	-	GH43	SignalP(1-18)
15	PenX_007679-T1	8.978283	Hypothetical protein	-	-	-	-	-
16	PenX_007638-T1	8.859755	Sugar transporter-like protein	Transmembrane transporter	Transmembrane transport	Membrane, integral component of membrane	-	-
17	PenX_008810-T1	8.778884	Hypothetical protein	-	-	-	-	-
18	PenX_010483-T1	8.586574	Hypothetical protein	-	-	-	-	SignalP(1-22)
19	PenX_001184-T1	8.580307	L-galactonate dehydratase	-	Cellular amino acid catabolic process	-	-	-
20	PenX_002250-T1	8.559438	Hypothetical protein	Oxidoreductase activity	Oxidation-reduction process	-	-	-
21	PenX_007858-T1	8.451312	Hypothetical protein	Oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor, oxidoreductase activity, NA	Oxidation-reduction process, glycerol-3-phosphate catabolic process	-	-	-
22	PenX_005039-T1	8.318457	Hypothetical protein	Branched-chain-amino-acid transaminase activity, catalytic activity	Branched-chain amino acid metabolic process	-	-	-
23	PenX_007734-T1	8.163883	Hypothetical protein	-	-	-	-	-
24	PenX_002547-T1	7.627807	Hypothetical protein	-	-	-	-	SignalP(1-22)

25	PenX_004190-T1	7.612452	Hypothetical protein	Transmembrane transporter activity	Transmembrane transport	-	-	-
26	PenX_008747-T1	7.611155	Hypothetical protein	-	-	-	-	-
27	PenX_003990-T1	7.509055	Hypothetical protein	-	-	-	-	-
28	PenX_000819-T1	7.471395	Hypothetical protein	-	-	-	-	SignalP(1-16)
29	PenX_007003-T1	7.457716	Hypothetical protein	Oxidoreductase activity, acting on the CH-CH group of donors, flavin adenine dinucleotide binding	Oxidation-reduction process	-	-	-
30	PenX_002046-T1	7.42009	Hypothetical protein (MEROPS)	Hydrolase activity	-	-	-	-

Table 4.3. The 30 most downregulated genes (and their functional annotation) in *Penicillium coffeae* ANU01 grown on its putative host *Pst*, relative to *Penicillium coffeae* ANU01 grown on synthetic media

Ranking	Gene_ID	Log2Fold-change	Predicted product	GO_function	GO_process	GO_localisation	CAZy	Secreted
1	PenX_003787-T3	-12.0222	Hypothetical protein	ATP binding, ATPase-coupled transmembrane transporter activity	Transmembrane transport	Membrane, integral component of membrane	-	-
2	PenX_009290-T1	-11.6918	Hypothetical protein	-	-		-	-
3	PenX_003787-T2	-11.4834	Hypothetical protein	ATP binding, ATPase activity, ATPase-coupled transmembrane transporter activity	Transmembrane transport	Membrane, integral component of membrane	-	-
4	PenX_004364-T1	-11.1758	Hypothetical protein	Fatty acid synthase activity, 3-oxoacyl-[acyl-carrier-protein] reductase, magnesium ion binding, holo-[acyl-carrier-protein] synthase activity, transferase activity, transferring acyl groups	Long-chain fatty acid biosynthetic process, oxidation-reduction process, fatty acid biosynthetic process	Fatty acid synthase complex	-	-
5	PenX_004447-T2	-11.0353	TATA-binding protein-associated factor mot1, variant 2	Nucleosome-dependent ATPase activity, ATP binding	-	-	-	-
6	PenX_009292-T1	-10.4618	Hypothetical protein	-	-	-	-	-
7	PenX_002683-T1	-10.3021	Hypothetical protein	Zinc ion binding, DNA-binding transcription factor activity, RNA polymerase II-specific	Regulation of transcription, DNA-templated	-	-	-
8	PenX_010255-T2	-9.96811	Hypothetical protein	-	-	-	-	-
9	PenX_010933-T1	-9.6661	Hypothetical protein	-	-	-	-	-
10	PenX_006580-T2	-9.46246	Multidrug resistance protein, variant 2	ATP binding, ATPase-coupled transmembrane transporter activity	Transmembrane transport	Membrane, integral component of membrane	-	-

11	PenX_003795-T1	-9.2616	Ammonium transporter Amt1	Ammonium transmembrane transporter activity	Ammonium transmembrane transport	Membrane, integral component of membrane	-	-
12	PenX_002131-T1	-9.20946	Hypothetical protein	-	-	-	-	SignalP(1-21)
13	PenX_009229-T1	-9.15785	Pre-mRNA-splicing factor ini1	-	mRNA splicing, cia spliceosome	-	-	-
14	PenX_003014-T1	-9.0953	Hypothetical protein	-	-	-	-	-
15	PenX_003217-T1	-9.08489	Hypothetical protein	-	-	-	-	SignalP(1-16)
16	PenX_009531-T2	-9.0535	Hypothetical protein	Transmembrane transporter activity	Transmembrane transport	-	-	-
17	PenX_009131-T1	-8.94594	Hypothetical protein	-	-	-	-	-
18	PenX_000204-T1	-8.88116	Hypothetical protein	-	-	-	-	-
19	PenX_009131-T1	-8.80599	Hypothetical protein	-	-	-	-	-
20	PenX_009115-T1	-8.77755	Hypothetical protein	-	-	-	-	SignalP(1-18)
21	PenX_009288-T1	-8.73814	Hypothetical protein	-	-	-	-	-
22	PenX_009287-T1	-8.64984	Hypothetical protein	-	-	-	-	-
23	PenX_010932-T1	-8.58638	Hypothetical protein	-	-	-	-	-
24	PenX_002135-T1	-8.37691	Hypothetical protein	-	-	-	-	SignalP(1-18)
25	PenX_004799-T1	-8.34389	Hypothetical protein	Flavin adenine dinucleotide binding, oxidoreductase activity, acting on CH-OH group of donors	Oxidation-reduction process	-	AA3	-
26	PenX_009405-T1	-8.186	Glycerol channel	Channel activity	Transmembrane transport	Membrane	-	-
27	PenX_007735-T1	-8.11532	Hypothetical protein	Zinc ion binding, DNA-binding	Transcription, DNA templated	-	-	-
28	PenX_007795-T1	-8.06532	Hypothetical protein	-	-	-	-	SignalP(1-22)
29	PenX_007411-T1	-8.0112	Hypothetical protein	Heme binding, iron ion binding, monooxygenase activity,	Oxidation-reduction process	-	-	-

				oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen				
30	PenX_000125-T1	-7.98589	Hypothetical protein	Zinc ion binding, DNA binding	Transcription, DNA templated	-	-	-

4.4 Summary

I used the funannotate bioinformatic pipeline to achieve a highly complete and highly functionally annotated gene set for *Penicillium coffeae* ANU01. I then used a comparative transcriptomic analysis to compare the gene expression profiles of the *Penicillium* grown on synthetic media and on its putative host *Pst* on wheat leaves. The comparative transcriptomic analysis showed gene expression varies markedly between conditions and allowed the identification of several highly upregulated genes of hyperparasitic function for future analysis beyond Honours.

5 Final Discussion

This thesis describes the species level identification and genetic investigation of a novel *Penicillium* hyperparasite of the wheat stripe rust fungus, *Pst.* A chromosome-level genome was assembled using both long- and short-read sequencing technologies, in a hybrid assembly approach. Identification of the hyperparasite as a likely novel isolate of the plant endophyte *Penicillium coffeae,* was made possible by genetic elements gleaned from the new assembly. Functional annotation of the genome and a comparative transcriptomic analysis, comparing the novel hyperparasite grown on synthetic media and its putative host, *Pst,* allowed preliminary identification of genetic mechanisms involved in the hyperparasitism. Several candidate genes, including an efflux pump component of a T1PKS biosynthetic gene cluster and several, potentially secreted enzymes, were identified and can now be explored for their specific effect on rust spores and importance to hyperparasitism. The results of this thesis open avenues for further investigation of a possible control method of a significant plant pathogen, and one specifically relevant to the Rathjen Laboratory.

5.1 Assembling the genome to assemble knowledge

While my *Penicillium coffeae* ANU01 genome assembly outperformed others regarding contiguity, comparative BUSCO analysis suggested that its completeness is lower than other *Penicillium* sequences. As described in Chapter 3, I generated BUSCO gene sets for 21 previously sequenced *Penicillium* species. All 21 assemblies outperformed the one generated in this study regarding completeness, with the lowest being 99.0% complete and the highest 99.99% (Table 3.1). While the 97.4% completeness is still regarded as highly complete (Veeckman *et al.*, 2016), poor completeness relative to the publicly available assemblies suggests there is room for improvement in our assembly.

The use of alternate assembly software may help identify the missing BUSCOs (completeness) and the missing telomere. As the quantity of DNA used to generate this assembly was quite large, improvements can likely be found through modifying the computational approach. The list of de novo genome assembly tools is expansive and growing constantly (Chen *et al.*, 2020). The

MaSuRCA hybrid assembler incorporates both long and short reads in a unified hybrid assembly process. MaSuRCA proved capable of assembling the very large and highly repetitive genome of a progenitor of bread wheat (Zimin *et al.,* 2017*)*. This could be advantageous in resolving the issue of the missing telomere in my assembly. Telomeric regions are highly repetitive and therefore are often difficult to fully assemble and identify (Kim *et al.,* 2021). If I were to repeat this analysis, I would attempt to re-assemble the genome using the MaSuRCA assembly software.

In Chapter 3, I show using a phylogenetic analysis that the novel hyperparasite is very likely an isolate of *Penicillium coffeae*, making this the first assembly of the *Penicillium coffeae* genome. Previous studies of the traits of *Penicillium coffeae* isolates have been purely observational. The availability of a highly accurate genome resource will be useful in uncovering genetic elements behind these traits. Van Den Berg et al. (2008) identified several genes with roles in penicillin production by investigating the genome of *P. chrysogenum* isolate W54-1255. This *Penicillium coffeae* ANU01 genome sequence could serve a similar role.

5.2 Taxonomic placement of the novel *Penicillium* hyperparasite

Whole-genome sequencing provides a wealth of data for taxonomic studies, and within this, BUSCO proteins have been used successfully as input data for phylogenomics Shen et al. (2016). However, the success of BUSCO-based phylogenomics is highly dependent on the availability of genome assemblies. In this study I found that despite the increased discriminatory power provided by BUSCO analysis, the resulting phylogeny was less able to provide a precise classification of *Penicillium coffeae* ANU01 than subsequent marker gene-based approaches. This was due to the disparity in available data. Of the 400+ *Penicillium* species that have been identified, only a small fraction (21) of these have publicly available genome data (Agarwala *et al.*, 2018). Additionally, none of these were closely related to *Penicillium coffeae* ANU01. The results of this analysis highlight the fact that genome sequencing has yet to catch up to the diversity of classified species. Therefore, care must be taken when performing phylogenetic classification to strike an appropriate balance between the discriminative power of the classification method and the availability of data to which a potentially novel species can be compared.

Literature describing *Penicillium coffeae* is scarce. It was first identified by Peterson *et al.*, (2005) in Hawaii as an endophyte on coffee (*Coffeae arabica)*. Confirmation that it was a new species was achieved via a phylogeny based on DNA barcode markers. Interestingly, Peterson *et al.,* (2005) emphasised the difficulty and historical ambiguity of species phylogenetics in the clade *Charlesia* and highlighted that multilocus barcoding removed some of these ambiguities that arose from morphological identification. My research mirrors this view as the initial *ITS* derived result was solidified by the inclusion of secondary marker gene data. The presence of inaccurate DNA barcoding sequences in genetic databases was also discussed by Peterson *et al.*, (2005). This is supported by my own phylogenic analysis (Figure 3.4) in which several isolates were shown to fall into clades outside of their prior species designation.

The data presented in this thesis represent the first evidence of *Penicillium coffeae* acting as a hyperparasite. However, there is some evidence of its pathogenic potential against other plant pathogenic fungi. In my phylogenetic analysis, *P. coffeae* ANU01 was placed closest to the *P. coffeae* isolate MA-314 (Fig 3.4). Cao *et al.,* (2019) isolated three metabolites from *P. coffeae* MA-314 that exhibited "potent activity" against *Fusarium oxysporum* f. sp. *momordicae nov. F.* and *Colletotrichum gloeosporioides*. Unfortunately, I was not able to test *P. coffeae* ANU01 against these fungi due to time constraints. However, one of the three antifungal metabolites identified in *P. coffeae* MA-314 belongs to the beta-lactone family, of which penicillin is a member. Genome mining of the *P. coffeae* ANU01 genome identified a beta-lactone biosynthetic gene cluster which may be the same cluster responsible for production of the antifungal beta-lactone identified in MA-314. Verification of this would require either whole genome sequencing of the isolate they possess or chemical identification of this metabolite in *Penicillium coffeae* ANU01.

5.3 Putting the puzzle together

In Chapter 4, I utilised RNA-seq data to functionally annotate the *Penicillium coffeae* ANU01 genome and conduct a comparative transcriptomic analysis. A highly complete gene set of 12,181 genes and 12,850 transcripts was generated, including 43 biosynthetic gene clusters encoding genes involved in biosynthesis of potentially antimicrobial compounds. Of these, one was shown to be

represented in the top 30 genes most upregulated *in planta* relative to *in vitro* growth, alongside 29 genes not associated with BGCs. These proteins are of presumed hyperparasitic importance.

A limitation of my comparative transcriptomics analysis was harvesting mRNA at a single time point, 24 hours after *Penicillium* was already visibly growing on the *Pst* pustules. I predict that gene expression during early stages of infection of pustules by *Penicillium* would be different to those expressed during the late stage sampled in this study. I predict fungal recognition and deployment of antifungal metabolites may be more prominent in this early growth stage, whereas nutrient acquisition is playing a greater role at the late time point we sampled for our analysis. To examine what is occurring transcriptionally during these early events I would harvest tissue 12-24 hours before observable *Penicillium* growth. This may also give information into potential mechanisms of host identification.

The predicted functions of the most highly upregulated genes allow me to speculate on the events that occur during the *Penicillium-Pst* interaction. The presence of *PenX_003989-T2*, a member of a T1PKS BGC, as the top upregulated gene, supports my hypothesis that a BGC is important in hyperparasitism. However, the prevalence of other, non BGC-associated genes suggests that infection is not solely governed by secondary metabolite deployment. Based on the observable similarities between growth on *Pst* pustules by *P. coffeae* ANU01 and the *Pst* hyperparasite identified by Zheng *et al.,* (2017), I predict that infection may proceed in a similar manner. Upon recognition of the *Pst* spore, the *Penicillium* hyphae would make physical contact with the spore. Upon contact, *P. coffeae* ANU01 would begin bombarding the spore with an antifungal metabolite or small molecule, secreted in abundance by the highly upregulated T1PKS efflux pump. I predict it does so with the goal of killing the *Pst* spore before it can germinate, and consume its stored nutrients. After inhibiting growth, I predict the *Penicillium* would use the chitosanase CAZY enzyme to degrade the spore coat and facilitate hyphal invasion. The GH43 CAZY would then break down carbohydrates, while an unannotated lipase breaks down the abundant fat bodies in the *Pst* spore. As a likely glycerol transporter, the *ST1LP*, would then be responsible for moving the products of the lipid metabolism (glycerol) into the *Penicillium.* This sequence of events is highly speculative, and would require much additional work to confirm.

5.4 Future directions: Building on what we now know

In this project several *Penicillium* genes of presumed hyperparasitic importance were identified as highly upregulated during the hyperparsitic event. To confirm the involvement of these genes in hyperparasitism, it is important that I knock them out and measure the effect of their absence on the ability of *P. coffeae* ANU01 to hyperparasitise *Pst* spores. In order to do so it would be necessary that I establish a CRISPR/Cas9 – mediated genome editing system in *Penicillium coffeae* ANU01. Wang et al. (2021) constructed a CRISPR/Cas9 genome editing system in *Penicilliuum oxalicum*, and used the system to generate gene-deletion mutants. By following this same method to generate knockout mutants for the five candidate genes described in this thesis, I would gain valuable insight into their importance in hyperparasitism.

This project did not discover whether a physical interaction occurs between *Penicillium coffeae* ANU01 and its host *Pst.* Physical characterisation of the hyperparasite-*Pst* interaction would provide more insight into the infection strategy of *P. coffeae* ANU01. Electron Scanning Microscopy (ESM) has been used previously to observe the hyphal invasion of *Pst* spores by three different fungal hyperparasites (Wang *et al.*, 2020; Zhan *et al.*, 2014; Zheng *et al.*, 2017). I would use an ESM approach to investigate the infection strategy of *P. coffeae* ANU01 on its host *Pst* to determine whether *P. coffeae* ANU01 directly invades rust spores with its hyphae or uses a strategy that does not require direct invasion such as the *P. brevicompactum* isolate identified by Wilson et al. (2020). I would also pair this with the aforementioned knockout isolates to measure the effect each gene deletion has on *P. coffeae* ANU01 infection, giving an idea as to the contribution of each gene to the overall infection strategy.

The method by which *P. coffeae* ANU01 detects and recognises *Pst* spores and is drawn to feed on them is also currently unknown. Pst spores are abundant with nutrients, in the form of lipids, carbohydrates, fats (Staples & Wynn 1965). However, it is unknown how rust hyperparasites recognise them as a valuable nutrient source and then move to parasitise them. To test if colonisation Pst spores is driven by a chemoattractant we could be to culture *P. coffeae* ANU01 and *Pst* on opposite ends of a nutrient agar plate and observe whether the *Penicillium* grows more

quickly towards the *Pst* than it grows alone on a control plate. This could also be tested with a split plate, where only the air can reach the other side but no physical contact inside the media occurs. This should be done with both germinating and non-germinating *Pst* spores, to see whether *P. coffeae* ANU01 displays a preference, and therefore give more information into its hyperparasitic function and its use as a biocontrol, regarding optimal time to treat plants.

There is a lot to be learned about *Penicillium coffeae* ANU01 before it can be implemented successfully as a biocontrol. Two factors that will contribute greatly to its suitability are host range and tolerance of a range of environmental conditions. The aim of characterising the *P. coffeae* ANU01 host range in this study was limited to stripe and stem rust. This leaves room for host range classification to be expanded to rusts in other species such as Myrtle rust (*Puccinia psidii*) which infects Eucalypt species, Roux *et al.*, (2013), and non-rust fungal pathogens of wheat, such as the fungal necrotroph *Zymoseptoria tritici* (Torriani *et al.*, 2015). A major limitation of hyperparasites as biocontrols is their requisite for high relative humidity in order to antagonise their hosts. All of the experimental work conducted in this thesis was completed in wheat grown in a controlled environment, and the effectiveness of *Penicillium coffeae* ANU01 as a hyperparasite and therefore biocontrol, may differ greatly between growth chamber and field conditions. To determine the effectiveness of *Penicillium coffeae* ANU01 as a biocontrol, it is important that the results of these proposed experiments be confirmed in field trials.

5.5 Conclusions

There is now a high-quality genomic resource for the newly classified wheat rust hyperparasite *Penicillium coffeae* ANU01 available to the Rathjen Laboratory. This highly annotated genome enables characterisation of genes important in the hyperparasitic interaction. This project has already made use of this resource to identify several genes highly upregulated during hyperparasitism. Several of these genes possess predicted functions we would expect of genes that facilitate hyperparasitism. However, this analysis of gene function is preliminary and serves primarily as a demonstration of the type of information that this genome assembly and associated transcriptomic data can provide. To further understand the hyperparasitic process of this fungus much work needs to be done to characterise the function of genes involved. It also remains to be determined if *P. coffeae* ANU01 would be a suitable biocontrol for deployment against pathogenic fungi. Testing of both its host range and resistance to the range of climatic conditions present in an agricultural setting will be required before potential use in this capacity.

References

Agarwala, Richa, Tanya Barrett, Jeff Beck, Dennis A Benson, Colleen Bollin, Evan Bolton, Devon Bourexis, et al. 2018. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 46 (D1): D8–13. https://doi.org/10.1093/nar/gkx1095.

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/s0022-2836(05)80360-2.

Aly, Amal H, Abdessamad Debbab, and Peter Proksch. 2011. "Fifty Years of Drug Discovery from Fungi." *Fungal Diversity* 50 (1): 3–19. https://doi.org/10.1007/s13225-011-0116-y.

Barratt, B I P, V C Moran, F Bigler, and J C Van Lenteren. 2018. "The Status of Biological Control and Recommendations for Improving Uptake for the Future." *BioControl* 63 (1): 155–67. https://doi.org/10.1007/s10526-017-9831-y.

Beattie, Andrew J, Mark Hay, Bill Magnusson, Rocky De Nys, James Smeathers, and Julian F V Vincent. 2011. "Ecology and Bioprospecting." *Austral Ecology* 36 (3): 341–56. https://doi.org/10.1111/j.1442-9993.2010.02170.x.

Benhamou, Nicole, Patrice Rey, Karine Picard, and Yves Tirilly. 1999. "Ultrastructural and Cytochemical Aspects of the Interaction Between the Mycoparasite Pythium Oligandrum and Soilborne Plant Pathogens." *Phytopathology*® 89 (6): 506–17. https://doi.org/10.1094/phyto.1999.89.6.506.

Berg, Marco A Van Den, Richard Albang, Kaj Albermann, Jonathan H Badger, Jean-Marc Daran, Arnold J M Driessen, Carlos Garcia-Estrada, et al. 2008. "Genome Sequencing and Analysis of the Filamentous Fungus Penicillium Chrysogenum." *Nature Biotechnology* 26 (10): 1161–68. https://doi.org/10.1038/nbt.1498.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Caldwell, Colleen C, and Maria Spies. 2017. "Helicase SPRNTing through the Nanopore." *Proceedings of the National Academy of Sciences* 114 (45): 11809–11. https://doi.org/10.1073/pnas.1716866114.

Cao, Jin, Xiao-Ming Li, Xin Li, Hong-Lei Li, Ling-Hong Meng, and Bin-Gui Wang. 2019. “New Lactone and Isocoumarin Derivatives from the Marine Mangrove-Derived Endophytic Fungus Penicillium Coffeae MA-314.” *Phytochemistry Letters* 32: 1–5. https://doi.org/10.1016/j.phytol.2019.04.018.

Chen, Cheng, Qiang Li, Rongtao Fu, Jian Wang, Chuan Xiong, Zhonghan Fan, Rongping Hu, Hong Zhang, and Daihua Lu. 2019. “Characterization of the Mitochondrial Genome of the Pathogenic Fungus Scytalidium Auriculariicola (Leotiomycetes) and Insights into Its Phylogenetics.” *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-53941-5.

Chen, Zhao, David L Erickson, and Jianghong Meng. 2020. “Benchmarking Hybrid Assembly Approaches for Genomic Analyses of Bacterial Pathogens Using Illumina and Oxford Nanopore Sequencing.” *BMC Genomics* 21 (1). https://doi.org/10.1186/s12864-020-07041-8.

Chen, Zhao, David L Erickson, and Jianghong Meng. 2021. “Polishing the Oxford Nanopore Long-Read Assemblies of Bacterial Pathogens with Illumina Short Reads to Improve Genomic Analyses.” *Genomics* 113 (3): 1366–77. https://doi.org/10.1016/j.ygeno.2021.03.018.

Cheng, Yulin, Kuan Wu, Juanni Yao, Shumin Li, Xiaojie Wang, Lili Huang, and Zhensheng Kang. 2017. “PSTha5a23, a Candidate Effector from the Obligate Biotrophic Pathogen Puccinia Striiformis f. Sp. Tritici , Is Involved in Plant Defense Suppression and Rust Pathogenicity.” *Environmental Microbiology* 19 (5): 1717–29. https://doi.org/10.1111/1462-2920.13610.

Coster, Wouter De, Svenn D’Hert, Darrin T Schultz, Marc Cruts, and Christine Van Broeckhoven. 2018. “NanoPack: Visualizing and Processing Long-Read Sequencing Data.” *Bioinformatics* 34 (15): 2666–69. https://doi.org/10.1093/bioinformatics/bty149.

Delsuc, Frédéric, Henner Brinkmann, and Hervé Philippe. 2005. “Phylogenomics and the Reconstruction of the Tree of Life.” *Nature Reviews Genetics* 6 (5): 361–75. https://doi.org/10.1038/nrg1603.

Dizkirici, Ayten, and Aysenur Kalmer. 2019. “Utility of Various Molecular Markers in Fungal Identification and Phylogeny.” *Nova Hedwigia* 109 (1–2): 187–224. https://doi.org/10.1127/nova_hedwigia/2019/0528.

Draz, I. 2019. “Common Ancestry of Egyptian Puccinia Striiformis Population along with Effective and Ineffective Resistance Genes.” *Asian Journal of Biological Sciences* 12 (2): 217–21.

Earl, D, K Bradnam, J St. John, A Darling, D Lin, J Fass, H O K Yu, et al. 2011. “Assemblathon 1: A Competitive Assessment of de Novo Short Read Assembly Methods.” *Genome Research* 21 (12): 2224–41. https://doi.org/10.1101/gr.126599.111.

Ehgartner, Daniela, Christoph Herwig, and Jens Fricke. 2017. “Morphological Analysis of the Filamentous Fungus Penicillium Chrysogenum Using Flow Cytometry—the Fast Alternative to Microscopic Image Analysis.” *Applied Microbiology and Biotechnology* 101 (20): 7675–88. https://doi.org/10.1007/s00253-017-8475-2.

Eitzen, Katharina, Priyamedha Sengupta, Samuel Kroll, Eric Kemen, and Gunther Doehlemann. 2021. “A Fungal Member of the Arabidopsis Thaliana Phyllosphere Antagonizes Albugo Laibachii via a GH25 Lysozyme.” *ELife* 10. https://doi.org/10.7554/elife.65306.

Estep, L K, S F F Torriani, M Zala, N P Anderson, M D Flowers, B A McDonald, C C Mundt, and P C Brunner. 2015. “Emergence and Early Evolution of Fungicide Resistance in North American Populations OfZymoseptoria Tritici.” *Plant Pathology* 64 (4): 961–71. https://doi.org/10.1111/ppa.12314.

European Food Safety, Authority, Maria Arena, Domenica Auteri, Stefania Barmaz, Giulia Bellisai, Alba Brancato, Daniela Brocca, et al. 2018. “Peer Review of the Pesticide Risk Assessment of the Active Substance Chlorothalonil.” *EFSA Journal* 16 (1). https://doi.org/10.2903/j.efsa.2018.5126.

Ewing, Brent, and Phil Green. 1998. “Base-Calling of Automated Sequencer Traces UsingPhred.II. Error Probabilities.” *Genome Research* 8 (3): 186–94. https://doi.org/10.1101/gr.8.3.186.

Fajarningsih, Nurrahmi Dewi. 2016. “Internal Transcribed Spacer (ITS) as Dna Barcoding to Identify Fungal Species: A Review.” *Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology* 11 (2): 37. https://doi.org/10.15578/squalen.v11i2.213.

Färber, P, and R Geisen. 2000. “Karyotype of Penicillium Nalgiovense and Assignment of the Penicillin Biosynthetic Genes to Chromosome IV.” *International Journal of Food*

Microbiology 58 (1–2): 59–63. https://doi.org/10.1016/s0168-1605(00)00279-8.

Feau, Nicolas, Stéphanie Beauseigle, Marie-Josée Bergeron, Guillaume J Bilodeau, Inanc Birol, Sandra Cervantes-Arango, Braham Dhillon, et al. 2018. "Genome-Enhanced Detection and Identification (GEDI) of Plant Pathogens." *PeerJ* 6: e4392. https://doi.org/10.7717/peerj.4392.

Figueroa, Melania, Kim E Hammond-Kosack, and Peter S Solomon. 2018. "A Review of Wheat Diseases-a Field Perspective." *Molecular Plant Pathology* 19 (6): 1523–36. https://doi.org/10.1111/mpp.12618.

Fu, Shuhua, Anqi Wang, and Kin Fai Au. 2019. "A Comparative Evaluation of Hybrid Error Correction Methods for Error-Prone Long Reads." *Genome Biology* 20 (1). https://doi.org/10.1186/s13059-018-1605-z.

funannotate. 2020. "No Title." 2020. https://github.com/nextgenusfs/funannotate.

Garnica, Diana P, Adnane Nemri, Narayana M Upadhyaya, John P Rathjen, and Peter N Dodds. 2014. "The Ins and Outs of Rust Haustoria." *PLoS Pathogens* 10 (9): e1004329. https://doi.org/10.1371/journal.ppat.1004329.

Gaynes, Robert. 2017. "The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use." *Emerging Infectious Diseases* 23 (5): 849–53. https://doi.org/10.3201/eid2305.161556.

Godfray, H Charles J, Daniel Mason-D'Croz, and Sherman Robinson. 2016. "Food System Consequences of a Fungal Disease Epidemic in a Major Crop." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1709): 20150467. https://doi.org/10.1098/rstb.2015.0467.

Goldman, Aaron David, and Laura F Landweber. 2016. "What Is a Genome?" *PLOS Genetics* 12 (7): e1006181. https://doi.org/10.1371/journal.pgen.1006181.

Goodwin, Sara, James Gurtowski, Scott Ethe-Sayers, Panchajanya Deshpande, Michael C Schatz, and W Richard McCombie. 2015. "Oxford Nanopore Sequencing, Hybrid Error Correction, and de Novo Assembly of a Eukaryotic Genome." *Genome Research* 25 (11): 1750–56. https://doi.org/10.1101/gr.191395.115.

Green, Michael R, and Joseph Sambrook. 2019. “Agarose Gel Electrophoresis.” *Cold Spring Harbor Protocols* 2019 (1): pdb.prot100404. https://doi.org/10.1101/pdb.prot100404.

Griesemer, Marc, Jeffrey A Kimbrel, Carol E Zhou, Ali Navid, and Patrik D’Haeseleer. 2018. “Combining Multiple Functional Annotation Tools Increases Coverage of Metabolic Annotation.” *BMC Genomics* 19 (1). https://doi.org/10.1186/s12864-018-5221-9.

Gueddari, Nour Eddine El, Una Rauchhaus, Bruno M Moerschbacher, and Holger B Deising. 2002. “Developmentally Regulated Conversion of Surface-Exposed Chitin to Chitosan in Cell Walls of Plant Pathogenic Fungi.” *New Phytologist* 156 (1): 103–12. https://doi.org/10.1046/j.1469-8137.2002.00487.x.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. “QUAST: Quality Assessment Tool for Genome Assemblies.” *Bioinformatics* 29 (8): 1072–75. https://doi.org/10.1093/bioinformatics/btt086.

Han, Xiaolong, Alolika Chakrabortti, Jindong Zhu, Zhao-Xun Liang, and Jinming Li. 2016. “Sequencing and Functional Annotation of the Whole Genome of the Filamentous Fungus Aspergillus Westerdijkiae.” *BMC Genomics* 17 (1). https://doi.org/10.1186/s12864-016-2974-x.

Hartmann, Fanny E, Andrea Sánchez-Vallet, Bruce A McDonald, and Daniel Croll. 2017. “A Fungal Wheat Pathogen Evolved Host Specialization by Extensive Chromosomal Rearrangements.” *The ISME Journal* 11 (5): 1189–1204. https://doi.org/10.1038/ismej.2016.196.

Hebert, Paul D N, and T Ryan Gregory. 2005. “The Promise of DNA Barcoding for Taxonomy.” *Systematic Biology* 54 (5): 852–59. https://doi.org/10.1080/10635150500354886.

Henry, Ronnie. 2019. “Etymologia: Penicillin.” *Emerging Infectious Diseases* 25 (1): 62. https://doi.org/10.3201/eid2501.et2501.

Hill, David P, Barry Smith, Monica S Mcandrews-Hill, and Judith A Blake. 2008. “Gene Ontology Annotations: What They Mean and Where They Come From.” *BMC Bioinformatics* 9 (Suppl 5): S2. https://doi.org/10.1186/1471-2105-9-s5-s2.

Howell, C R. 2003. “Mechanisms Employed by Trichoderma Species in the Biological Control of Plant Diseases: The History and Evolution of Current Concepts.” *Plant Disease* 87 (1): 4–

10. https://doi.org/10.1094/pdis.2003.87.1.4.

Huang, H C, and E G Kokko. 1987. “Ultrastructure of Hyperparasitism of *Coniothyrium Minitans* on Sclerotia of *Sclerotinia Sclerotiorum*.” *Canadian Journal of Botany* 65 (12): 2483–89. https://doi.org/10.1139/b87-337.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, et al. 2019. “EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14. https://doi.org/10.1093/nar/gky1085.

ICPA. 2018. “No Title.” 2018. https://www.aspergilluspenicillium.org/icpa.

Jeleń, Henryk, Lidia Błaszczyk, Jerzy Chełkowski, Katarzyna Rogowicz, and Judyta Strakowska. 2014. “Formation of 6-n-Pentyl-2H-Pyran-2-One (6-PAP) and Other Volatiles by Different Trichoderma Species.” *Mycological Progress* 13 (3): 589–600. https://doi.org/10.1007/s11557-013-0942-2.

Jones, Ashley, and Benjamin Schwessinger. 2020. “Sorbitol Washing Complex Homogenate for Improved DNA Extractions v1 (Protocols.Io.Beuvjew6).” *Protocols.Io*. https://doi.org/10.17504/protocols.io.beuvjew6.

Jones, Ashley, Cynthia Torkel, David Stanley, Jamila Nasim, Justin Borevitz, and Benjamin Schwessinger. 2020. “Scalable High-Molecular Weight DNA Extraction for Long-Read Sequencing v1 (Protocols.Io.Bnjhmcj6).” *Protocols.Io*. https://doi.org/10.17504/protocols.io.bnjhmcj6.

Keilwagen, Jens, Frank Hartung, Michael Paulini, Sven O Twardziok, and Jan Grau. 2018. “Combining RNA-Seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi.” *BMC Bioinformatics* 19 (1). https://doi.org/10.1186/s12859-018-2203-5.

Kim, Eunkyeong, Jun Kim, Chuna Kim, and Junho Lee. 2021. “Long-Read Sequencing and de Novo Genome Assemblies Reveal Complex Chromosome End Structures Caused by Telomere Dysfunction at the Single Nucleotide Level.” *Nucleic Acids Research* 49 (6): 3338–53. https://doi.org/10.1093/nar/gkab141.

Ko Ko, Thida Win, Steven L Stephenson, Ali H Bahkali, and Kevin D Hyde. 2011. “From

Morphology to Molecular Biology: Can We Use Sequence Data to Identify Fungal Endophytes?" *Fungal Diversity* 50 (1): 113–20. https://doi.org/10.1007/s13225-011-0130-0.

Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37 (5): 540–46. https://doi.org/10.1038/s41587-019-0072-8.

Koren, Sergey, Brian P Walenz, Konstantin Berlin, Jason R Miller, Nicholas H Bergman, and Adam M Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptivek-Mer Weighting and Repeat Separation." *Genome Research* 27 (5): 722–36. https://doi.org/10.1101/gr.215087.116.

Krzywinski, M, J Schein, I Birol, J Connors, R Gascoyne, D Horsman, S J Jones, and M A Marra. 2009. "Circos: An Information Aesthetic for Comparative Genomics." *Genome Research* 19 (9): 1639–45. https://doi.org/10.1101/gr.092759.109.

Kuramae, Eiko E, Vincent Robert, Berend Snel, Michael WeiãŸ, and Teun Boekhout. 2006. "Phylogenomics Reveal a Robust Fungal Tree of Life." *FEMS Yeast Research* 6 (8): 1213–20. https://doi.org/10.1111/j.1567-1364.2006.00119.x.

Mahmood, Khalid, Jihad Orabi, Peter Skov Kristensen, Pernille Sarup, Lise Nistrup Jørgensen, and Ahmed Jahoor. 2020. "De Novo Transcriptome Assembly, Functional Annotation, and Expression Profiling of Rye (Secale Cereale L.) Hybrids Inoculated with Ergot (Claviceps Purpurea)." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-70406-2.

Martin, Kendall J, and Paul T Rygiewicz. 2005. "Fungal-Specific PCR Primers Developed for Analysis of the ITS Region of Environmental DNA Extracts." *BMC Microbiology* 5 (1): 28. https://doi.org/10.1186/1471-2180-5-28.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17 (1): 10. https://doi.org/10.14806/ej.17.1.200.

Murigneux, Valentine, Subash Kumar Rai, Agnelo Furtado, Timothy J C Bruxner, Wei Tian, Ivon Harliwong, Hanmin Wei, et al. 2020. "Comparison of Long-Read Methods for Sequencing and Assembly of a Plant Genome." *GigaScience* 9 (12). https://doi.org/10.1093/gigascience/giaa146.

ONT. 2020. "No Title." 2020. https://github.com/nanoporetech/medaka.

Ou, Shujun, Jianing Liu, Kapeel M Chougule, Arkarachai Fungtammasan, Arun S Seetharam, Joshua C Stein, Victor Llaca, et al. 2020. “Effect of Sequence Depth and Length in Long-Read Assembly of the Maize Inbred NC358.” *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-16037-7.

Parratt, Steven R, and Anna-Liisa Laine. 2016. “The Role of Hyperparasitism in Microbial Pathogen Ecology and Evolution.” *The ISME Journal* 10 (8): 1815–22. https://doi.org/10.1038/ismej.2015.247.

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” *Nature Methods* 14 (4): 417–19. https://doi.org/10.1038/nmeth.4197.

Peterson, Stephen W, Fernando E Vega, Francisco Posada, and Chifumi Nagai. 2005. “*Penicillium Coffeae,* a New Endophytic Species Isolated from a Coffee Plant and Its Phylogenetic Relationship to *P. Fellutanum*, *P. Thiersii* and *P. Brocae* Based on Parsimony Analysis of Multilocus DNA Sequences.” *Mycologia* 97 (3): 659–66. https://doi.org/10.1080/15572536.2006.11832796.

Pop, M; Phillippy, A; Delcher AL; Salzberg SL. 2004. “Comparative Genome Assembly.” *Briefings in Bioinformatics* 5 (3): 237–48. https://doi.org/10.1093/bib/5.3.237.

Raja, Huzefa A, Andrew N Miller, Cedric J Pearce, and Nicholas H Oberlies. 2017. “Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community.” *Journal of Natural Products* 80 (3): 756–70. https://doi.org/10.1021/acs.jnatprod.6b01085.

Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47–e47. https://doi.org/10.1093/nar/gkv007.

Rokas, Antonis, Matthew E Mead, Jacob L Steenwyk, Huzefa A Raja, and Nicholas H Oberlies. 2020. “Biosynthetic Gene Clusters and the Evolution of Fungal Chemodiversity.” *Natural Product Reports* 37 (7): 868–78. https://doi.org/10.1039/c9np00045c.

Romanelli, Anna M, Deanna A Sutton, Elizabeth H Thompson, Michael G Rinaldi, and Brian L

Wickes. 2010. “Sequence-Based Identification of Filamentous Basidiomycetous Fungi from Clinical Specimens: A Cautionary Note.” *Journal of Clinical Microbiology* 48 (3): 741–52. https://doi.org/10.1128/jcm.01948-09.

Roux, Jolanda, Izette Greyling, Teresa A Coutinho, Marcel Verleur, and Michael J Wingfeld. 2013. “The Myrtle Rust Pathogen, Puccinia Psidii, Discovered in Africa.” *IMA Fungus* 4 (1): 155–59. https://doi.org/10.5598/imafungus.2013.04.01.14.

Schoch, C L, K A Seifert, S Huhndorf, V Robert, J L Spouge, C A Levesque, W Chen, et al. 2012. “Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi.” *Proceedings of the National Academy of Sciences* 109 (16): 6241–46. https://doi.org/10.1073/pnas.1117018109.

Schönbichler, Anna, Sara M Díaz-Moreno, Vaibhav Srivastava, and Lauren Sara Mckee. 2020. “Exploring the Potential for Fungal Antagonism and Cell Wall Attack by Bacillus Subtilis Natto.” *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.00521.

Schwessinger, Benjamin. 2017. “Fundamental Wheat Stripe Rust Research in the 21 St Century.” *New Phytologist* 213 (4): 1625–31. https://doi.org/10.1111/nph.14159.

Shen, Xing-Xing, Xiaofan Zhou, Jacek Kominek, Cletus P Kurtzman, Chris Todd Hittinger, and Antonis Rokas. 2016. “Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data.” *G3 Genes|Genomes|Genetics* 6 (12): 3927–39. https://doi.org/10.1534/g3.116.034744.

Shen, Youming, Jiyun Nie, Lixue Kuang, Jianyi Zhang, and Haifei Li. 2021. “DNA Sequencing, Genomes and Genetic Markers of Microbes on Fruits and Vegetables.” *Microbial Biotechnology* 14 (2): 323–62. https://doi.org/10.1111/1751-7915.13560.

Shingate, Prashant, Vydianathan Ravi, Aravind Prasad, Boon-Hui Tay, Kritika M Garg, Balaji Chattopadhyay, Laura-Marie Yap, Frank E Rheindt, and Byrappa Venkatesh. 2020. “Chromosome-Level Assembly of the Horseshoe Crab Genome Provides Insights into Its Genome Evolution.” *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-16180-1.

Simão, Felipe A, Robert M Waterhouse, Panagiotis Ioannidis, Evgenia V Kriventseva, and Evgeny M Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation

Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Sims, David, Ian Sudbery, Nicholas E Ilott, Andreas Heger, and Chris P Ponting. 2014. "Sequencing Depth and Coverage: Key Considerations in Genomic Analyses." *Nature Reviews Genetics* 15 (2): 121–32. https://doi.org/10.1038/nrg3642.

Singh, Ravi P, Pawan K Singh, Jessica Rutkoski, David P Hodson, Xinyao He, Lise N Jørgensen, Mogens S Hovmøller, and Julio Huerta-Espino. 2016. "Disease Impact on Wheat Yield Potential and Prospects of Genetic Control." *Annual Review of Phytopathology* 54 (1): 303–22. https://doi.org/10.1146/annurev-phyto-080615-095835.

Sperschneider, J. 2019. "No Title." 2019. https://github.com/JanaSperschneider/FindTelomeres.

Staples, Richard C, and Willard K Wynn. 1965. "The Physiology of Uredospores of the Rust Fungi." *The Botanical Review* 31 (4): 537–64. https://doi.org/10.1007/BF02858608.

Stein, Lincoln. 2001. "Genome Annotation: From Sequence to Biology." *Nature Reviews Genetics* 2 (7): 493–503. https://doi.org/10.1038/35080529.

Summers, R W, and J K M Brown. 2013. "Constraints on Breeding for Disease Resistance in Commercially Competitive Wheat Cultivars." *Plant Pathology* 62: 115–21. https://doi.org/10.1111/ppa.12165.

Tan, Ge, Lennart Opitz, Ralph Schlapbach, and Hubert Rehrauer. 2019. "Long Fragments Achieve Lower Base Quality in Illumina Paired-End Sequencing." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-39076-7.

Thrash, Adam, Federico Hoffmann, and Andy Perkins. 2020. "Toward a More Holistic Method of Genome Assembly Assessment." *BMC Bioinformatics* 21 (S4). https://doi.org/10.1186/s12859-020-3382-4.

Topolovec-Pintarić, Snježana. 2019. "Trichoderma: Invisible Partner for Visible Impact on Agriculture." In *Trichoderma - The Most Widely Used Fungicide*. IntechOpen. https://doi.org/10.5772/intechopen.83363.

Torriani, Stefano F F, James P E Melichar, Colin Mills, Naomi Pain, Helge Sierotzki, and Mikaël Courbot. 2015. "Zymoseptoria Tritici: A Major Threat to Wheat Production, Integrated

Approaches to Control." *Fungal Genetics and Biology* 79: 8–12. https://doi.org/10.1016/j.fgb.2015.04.010.

Veeckman, Elisabeth, Tom Ruttink, and Klaas Vandepoele. 2016. "Are We There Yet? Reliably Estimating the Completeness of Plant Genome Sequences." *The Plant Cell* 28 (8): 1759–68. https://doi.org/10.1105/tpc.16.00349.

Visagie, C M, J Houbraken, J C Frisvad, S B Hong, C H W Klaassen, G Perrone, K A Seifert, J Varga, T Yaguchi, and R A Samson. 2014. "Identification and Nomenclature of the Genus Penicillium." *Studies in Mycology* 78: 343–71. https://doi.org/10.1016/j.simyco.2014.09.001.

Wallberg, Andreas, Ignas Bunikis, Olga Vinnere Pettersson, Mai-Britt Mosbech, Anna K Childers, Jay D Evans, Alexander S Mikheyev, Hugh M Robertson, Gene E Robinson, and Matthew T Webster. 2019. "A Hybrid de Novo Genome Assembly of the Honeybee, Apis Mellifera, with Chromosome-Length Scaffolds." *BMC Genomics* 20 (1): 275. https://doi.org/10.1186/s12864-019-5642-0.

Wang, Ning, Xin Fan, Shan Zhang, Bo Liu, Mengying He, Xianming Chen, Chunlei Tang, Zhensheng Kang, and Xiaojie Wang. 2020. "Identification of a Hyperparasitic Simplicillium Obclavatum Strain Affecting the Infection Dynamics of Puccinia Striiformis f. Sp. Tritici on Wheat." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.01277.

Waterhouse, Robert M, Mathieu Seppey, Felipe A Simão, Mosè Manni, Panagiotis Ioannidis, Guennadi Klioutchnikov, Evgenia V Kriventseva, and Evgeny M Zdobnov. 2018. "BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics." *Molecular Biology and Evolution* 35 (3): 543–48. https://doi.org/10.1093/molbev/msx319.

Wick, RR. 2018. "No Title." 2018. https://github.com/rrwick/Porechop.

Wilson, A, W S Cuddy, R F Park, G F S Harm, M J Priest, J Bailey, and M C Moffitt. 2020. "Investigating Hyperparasites as Potential Biological Control Agents of Rust Pathogens on Cereal Crops." *Australasian Plant Pathology* 49 (3): 231–38. https://doi.org/10.1007/s13313-020-00695-8.

Xu, Jianping. 2016. "Fungal DNA Barcoding." *Genome* 59 (11): 913–32. https://doi.org/10.1139/gen-2016-0046.

Yang, Li-An, Yu-Jung Chang, Shu-Hwa Chen, Chung-Yen Lin, and Jan-Ming Ho. 2019. "SQUAT: A Sequencing Quality Assessment Tool for Data Quality Assessments of Genome Assemblies." *BMC Genomics* 19 (S9). https://doi.org/10.1186/s12864-019-5445-3.

Zhan, Gangming, Yuan Tian, Fuping Wang, Xianming Chen, Jun Guo, Min Jiao, Lili Huang, and Zhensheng Kang. 2014. "A Novel Fungal Hyperparasite of Puccinia Striiformis f. Sp. Tritici, the Causal Agent of Wheat Stripe Rust." *PLoS ONE* 9 (11): e111484. https://doi.org/10.1371/journal.pone.0111484.

Zhang, Ke, Baoqi Huang, Kai Yuan, Xiaojun Ji, Ping Song, Qingqing Ding, and Yuwen Wang. 2020. "Comparative Transcriptomics Analysis of the Responses of the Filamentous Fungus Glarea Lozoyensis to Different Carbon Sources." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.00190.

Zheng, Li, Jie Zhao, Xiaofei Liang, Gangming Zhan, Shuchang Jiang, and Zhensheng Kang. 2017. "Identification of a Novel Alternaria Alternata Strain Able to Hyperparasitize Puccinia Striiformis f. Sp. Tritici, the Causal Agent of Wheat Stripe Rust." *Frontiers in Microbiology* 8. https://doi.org/10.3389/fmicb.2017.00071.

Zimin, Aleksey V, Daniela Puiu, Ming-Cheng Luo, Tingting Zhu, Sergey Koren, Guillaume Marçais, James A Yorke, Jan Dvořák, and Steven L Salzberg. 2017. "Hybrid Assembly of the Large and Highly Repetitive Genome of Aegilops Tauschii, a Progenitor of Bread Wheat, with the MaSuRCA Mega-Reads Algorithm." *Genome Research* 27 (5): 787–92. https://doi.org/10.1101/gr.213405.116.

Zimin, Aleksey V, and Steven L Salzberg. 2020. "The Genome Polishing Tool POLCA Makes Fast and Accurate Corrections in Genome Assemblies." *PLOS Computational Biology* 16 (6): e1007981. https://doi.org/10.1371/journal.pcbi.1007981.