Characterising some mechanisms of iron homeostasis in selected *Armillaria* species

by

Deborah Louisa Narh Mensah
(Student number: 18394320)

Submitted in Fulfilment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Faculty of Natural and Agricultural Sciences
Department of Biochemistry, Genetics and Microbiology
Forestry and Agricultural Biotechnology Institute
University of Pretoria, Pretoria, South Africa

May 2023

Promoter: Prof. Martin P. A. Coetzee

Co-promoter: Prof. Brenda D. Wingfield

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Declaration

I, Deborah Louisa Narh Mensah, declare that the thesis, which I hereby submit for the degree, Doctor of Philosophy in Biotechnology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Deborah L. Narh Mensah

May 2023

To my father, Solomon Narh Tettey, who never saw this adventure,

&

To my mother, Margaret Mateko Sackitey-Kamah, who saw only a glimpse.

Gone, but never forgotten!

Acknowledgments

I am grateful for the opportunity to have conducted my studies toward my PhD in the well-equipped laboratories of the Department of Biochemistry, Genetics and Microbiology and the Forestry and Agricultural Biotechnology Institute (FABI).

I would like to extend my profound gratitude to my primary supervisor, Prof. Martin P. A. Coetzee, for believing in me and fighting for me to get this opportunity even without knowing me. This PhD journey has been possible because you gave me a chance. You were there in full support for me through the professional and personal challenges which ensued throughout this journey. You also picked up my strengths and readily nominated me for extracurricular activities including serving as the Lab Manager for three years and as a Co-chair of the BGM Postgraduate Student Committee, as well as my participation in the National Science Foundation – Fungal Bioinvasions Workshop. These have enhanced my technical skills along with my managerial and other soft skills. I really appreciate all your support.

To my co-supervisor, Prof. Brenda D. Wingfield, if there were a better word to say THANK YOU, that word would have been appropriate for all the support and guidance you gave me. Our meetings were discussions on both the science and about academic life. These have contributed significantly to enhancing my professional and soft skills. I consider the lessons I learnt from our conversations as life-long treasures.

Funding is a quintessential part of any research endeavour. I thank the University of Pretoria, FABI, the Tree Protection Co-operative Programme (TPCP), and the DSI – NRF South African Research Chairs Initiative (SARChI) in Fungal Genomics (Grant number: 98353) for all the financial support. Without this financial support, commencing and completing this degree would not have been possible. I am also grateful to Mrs. Eva Muller (now retired) and Mrs. Heidi Roos for their administrative assistance during my PhD research.

Sometimes, the best of friends are strangers. As an International Student at the time herself, Dr. Agil Katumanyane understood the peculiar challenges one faces as an International Student in a country one was visiting for the first time. Even though she barely knew me, she offered and helped me to settle in. When I received one of the biggest blows of my life shortly after I started my PhD, Agil was my human angel. For this, I will be eternally grateful.

Riding towards success requires good colleagues. I thank my previous lab mates, Drs. Stephanie van Wyk, Mohammed Sayari, and Quentin Santana, who I met in the labs and have since completed their studies, for getting me familiarised with our labs and with some of the software I used during my study. I also thank my lab mates who made my three-years tenure as the Lab Manager a smooth one. I especially thank Mkhululi Maphosa, Sikelela Buthelezi and Francinah Ratsoma for the intellectual discussions. To Phrasia Mapfumo, some other friends I made at Future Africa some of whom are FABIans, and to some members of Future Africa management team, who became more of my Republic of South Africa family, I extend my deep appreciation. To everyone else in FABI, I say thank you for contributing to a conducive multicultural working environment where research, networking and collaborations thrive.

To my dear mum, you bowed out shortly after I embarked on this journey. Saying it did not negatively impact me could not have been farther from the truth. The daily lingering grief in my heart after dad's passing years ago was reinvigorated with your passing. There were times when I had doubts, but I always remembered what you said to me when I informed you about this PhD. That memory of the kind of faith you had in my abilities and the pride in your eyes when you spoke those words spurred me on through all the tough times including the COVID-19 pandemic and its related issues. As I navigated all the hills and valleys on this journey, I moved knowing that I do not come as one, but as a multitude composed of everyone who believes in me with you and dad moving right beside me. Mum, dad, we have made it! Thank you for all the gifts you gave me throughout my formative years, which have made me who I am and has enabled me to complete this journey despite everything. Thank you!

Being able to count on a solid support system no matter how far they are from you, is a blessing. I am grateful to my big sister Henrietta Darley Narh, my younger brother Samuel Korley Korda, all my other siblings, my wonderful nieces and nephews, my in-laws, and friends for all their prayers, diverse support, and encouragement. To you all: love and unlimited gratitude!

The last applause is loudest. To my dearest husband, Foster Yao Mensah, I thank you very much for your support in diverse ways. I know the journey to this achievement has been as much a sacrifice for you as it has been for me. We have been physically apart for most of the time I have been pursuing this goal. Yet, you gave me the peace of mind to be away for this long. You provided your ardent support and have been my greatest cheerleader. I do not take this for granted at all. I love and appreciate you dearly!

Preface

Plant-pathogenic organisms including fungi pose significant risks to agriculture, horticulture, and natural and plantation forests. This affects attainment of some of the 17 Sustainable Development Goals (SDGs) to transform our world. The affected goals include Goal 2 – "End hunger, achieve food security and improved nutrition and promote sustainable agriculture" and Goal 15 – "Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss". Species of *Armillaria* belong to the Physalacriaceae and have a worldwide distribution with a range of plant-pathogenic lifestyles. Current control strategies are inefficient. Hence, there is an urgent need to develop more efficient and sustainable control strategies against the plant-pathogenic members of this group. To achieve this, a deeper understanding is needed about the cellular and molecular defence strategies employed by these fungi. Therefore, the overall objective of this thesis was to increase our understanding of mechanisms employed by *Armillaria* in comparison to other species in the Physalacriaceae with regards to iron homeostasis. This was done by studying some secondary metabolite gene clusters, and investigating growth, siderophore production, and proteomic and secretomic response of *Armillaria* species to iron. A multidisciplinary approach including genetics, comparative genomics, *in vitro* bioassays, and proteomics was employed. The findings of this thesis represent the first study of some mechanisms underpinning iron homeostasis by *Armillaria* species in comparison to some other members of the Physalacriaceae.

The first chapter serves as a practical guide for genome sequencing and assembly of basidiomycetes using *Armillaria* genome projects as a case study. In this review, the status of *Armillaria* genome sequencing, assembly, and annotation in terms of samples used and technical, technological, and computational methods applied were evaluated. Considerations for decision making in future genome sequencing, assembly and annotation projects were provided. The information gathered from genomics studies of *Armillaria* species were summarized and prospects for further research were also provided. Infographic guidelines to facilitate future genome sequencing, assembly, and annotation projects on the genus *Armillaria,* which can be extended to the genomics studies of other diploid and dikaryotic fungi has also been provided in this review. Additional literature pertaining to the experimental chapters is provided in the respective chapters.

In the second chapter the aim was to identify the diversity and numbers of secondary metabolite gene clusters (SMGCs) in genomes of *Armillaria* spp. and other selected species within the Physalacriaceae with a focus on NRPS-dependent siderophore synthetase (NDSS) gene clusters. Publicly available whole genome sequences and protein sequences from eleven species in the genera *Armillaria* (6 species)*, Desarmillaria* (2 species)*,* and one species each of *Cylindrobasidium,*

Guyanagaster, and *Oudemansiella* were used to identify the SMGCs. Comparative genomics was used to characterize the NDSS gene clusters structurally and phylogenetically and to characterize the NDSS genes. Bioassays were also used to assess siderophore biosynthesis by strains belonging to selected *Armillaria* species. The results from this study provide insights into the genetic basis of secondary metabolism in general, and about NDSS gene clusters in the studied genomes. The study also provides the first record of *in vitro* siderophore biosynthesis by strains of selected *Armillaria* species. This study presents a comprehensive basis for exploring the roles of NRPS-dependent siderophore biosynthesis in these socio-economically important phytopathogens. The Chapter has been published as "**Narh Mensah, D.L.,** Wingfield, B.D., Coetzee, M.P.A. 2023. Nonribosomal peptide synthetase gene clusters and characteristics of predicted NRPS-dependent siderophore synthetases in *Armillaria* and other species in the Physalacriaceae. *Current Genetics*. **69**(1), 7-24".

In the third chapter a comparative genomics approach was used to investigate NRPS-independent siderophore (NIS) synthetase gene clusters in *Armillaria* in comparison to other species in the Physalacriaceae. The genomes studied for investigating the SMGCs were also used for this study. *In vitro* analyses were conducted to investigate effect of iron on growth, macromorphology, and siderophore biosynthesis of *Armillaria* species. Results of this study provide insights into diversity of NIS synthetase gene clusters in the genomes studied and effect of iron on various characteristics of *Armillaria* species.

The fourth chapter describes genome sequencing of an *Armillaria* isolate from Zimbabwe using long and short read sequencing and genome assembly. This study provides an invaluable genome resource of a member of the genus, *Armillaria*, from Africa. This resource could be used to gain further understanding of the biology and other characteristics of these phytopathogens and to potentially develop biotechnologically important products from these organisms. This chapter has been published in IMA Fungus as "**Narh Mensah, D.L.**, Wingfield, B.D., Maphosa, M., Duong, T.A., Coetzee, M.P.A. 2022. IMA genome-F17A Draft genome sequence of an *Armillaria* species from Zimbabwe. *IMA Fungus*, **13**(19), 1-3".

In the final chapter, the effect of iron on proteomics and secretomics of the *Armillaria* isolate whose genome was sequenced as part of my PhD research was investigated. My aim was to provide a better understanding of the mechanisms employed by *Armillaria* spp. in response to iron. A label-free and gel-free LC–MS/MS shotgun approach, coupled with bioinformatic analyses were used for this purpose. Results of this study sheds light on the kind and extent of metabolic and other changes exhibited by the fungus in response to iron and highlights potential modes of controlling the studied strain. The study also serves as a first record of the proteome and secretome of an *Armillaria* spp. from Africa.

The concluding section of the thesis provides an overview of the information that emerged from the research presented in this thesis. Also, some of the challenges experienced during the research are discussed in this section. The section then concludes by informing the reader about future perspectives for conducting research on iron homeostasis in Basidiomycota.

This thesis is written as a series of research articles. Consequently, there is some redundancy and difference in style with the references.

Table of Contents

List of Tables

List of Figures

Supplementary files of this thesis are deposited in an accessible online repository.

CHAPTER 1

A PRACTICAL APPROACH TO GENOME ASSEMBLY AND ANNOTATION OF BASIDIOMYCETES USING THE EXAMPLE OF *ARMILLARIA*

An extract of this Chapter has been submitted to BioTechniques as:

Narh Mensah, D.L., Wingfield, B.D., Coetzee, M.P.A. 2023. A practical approach to genome assembly and annotation of basidiomycetes using the example of *Armillaria*. *BioTechniques*.

Abstract

Technological advancement in genome sequencing platforms as well as genome assembly and annotation algorithms that resulted in several genomic studies have provided an opportunity to further our understanding of the biology of phytopathogens, such as those in the genus *Armillaria* (Basidiomycota, Agaricales, Physalacriaceae). Most *Armillaria* species are facultative necrotrophs that cause root- and stem-rot usually on woody plants worldwide. This weakens the plants and can kill them, resulting in significant impact on the agro-industrial and forestry sectors. Current management practices have been unsuccessful in controlling these phytopathogens due to their growth mechanisms, wide host range, and other factors that are not clearly understood. In this review, published *Armillaria* genome sequencing projects were used as an example to give an overview of the methods and tools used for fungal genome sequencing projects. Genome sequencing, assembly, and annotation in terms of samples used and technical, technological, and computational methods applied were evaluated in the review. Infographic guidelines and a database of resources to facilitate future genome sequencing, assembly, and annotation projects on the genus *Armillaria* have been provided. Considerations for decision making in future genome sequencing, assembly and annotation projects are provided. The information gathered from genomics studies of *Armillaria* species are summarized to showcase the type of research questions that such studies can answer. Prospects for further research are also provided. This practical guide can be applied to other diploid and dikaryotic fungal genomes.

Keywords:

Bioinformatics, genome assembly, evolution, next generation sequencing, phytopathogens, proteomics, multi-omics, metabolites

1. Introduction

Armillaria includes about 70 known species (Sipos *et al.*, 2018) that are distributed worldwide (reviewed in (Baumgartner *et al.*, 2011; Coetzee *et al.*, 2018)). Two species, *A. ectypa* and *A. tabescens,* that were previously placed in *Armillaria*, now reside in *Desarmillaria,* based on absence of annulus at maturity and an evolutionary trajectory separated from the rest of the *Armillaria* species (Koch *et al.*, 2017). *Armillaria* spp. are ubiquitous, have a broad host range, parasitize both herbaceous and woody plants, and survive under diverse environmental conditions (reviewed in (Heinzelmann *et al.*, 2019)). The ubiquity of *Armillaria* spp. has been attributed to their life cycles, lifestyles, and other biotic factors. The reader is referred to reviews from various authors (Baumgartner *et al.*, 2011; Coetzee *et al.*, 2018; Heinzelmann *et al.*, 2019) for further information regarding the life cycles, lifestyles, species boundaries and geographical distribution, and other properties of *Armillaria* spp.

Ecologically, *Armillaria* spp. display a range of relationships with plants and other fungi. Relationships of *Armillaria* spp. with plants are of mycorrhizal, saprophytic, and parasitic nature. For instance, *Armillaria luteo-virens* is mycorrhizal, *A. cepistipes* and *A. gallica* are mostly saprotrophic, whereas *A. luteobubalina, A. mellea, A. mexicana*, and *A. ostoyae* are parasitic (Baumgartner and Rizzo, 2002; Coetzee *et al.*, 2003; Elías-Román *et al.*, 2018; Klopfenstein *et al.*, 2014; Koch *et al.*, 2017; Lushaj *et al.*, 2010; Morrison and Pellow, 2002; Prospero *et al.*, 2004; Warwell *et al.*, 2019). The majority of *Armillaria* spp. are considered facultative necrotrophs which display parasitic and saprotrophic phases (Gregory and Rishbeth, 1991; Guillaumin *et al.*, 1993). Some other fungi such as *Entoloma abortivum*, *Wynnea* spp., *Trichoderma* spp. and *Polyporus umbellatus* are putatively mycoparasites of *Armillaria* spp. (Kikuchi and Yamaji, 2010; Koch and Herr, 2021; Liu *et al.*, 2015; Pellegrini *et al.*, 2013; Xing *et al.*, 2021). Additionally, the mycetophagous wood nematode, *Bursaphelenchus fraudulentus*, parasitizes *Armillaria* (Tomalak and Woodward, 2017).

Intra- and inter-species variations in aggressiveness, pathogenicity, and virulence among pathogenic members of *Armillaria* have been reported (Labbé *et al.*, 2017; Morrison and Pellow, 2002; Prospero *et al.*, 2004). For instance, Morrison and Pellow (2002) demonstrated that isolates of *A. mellea* are more aggressive than those of *A. ostoyae* on Douglas-fir seedlings. In addition, *A. cepistipes* is less virulent than *A. ostoyae* isolates with intraspecific variation in virulence being more predominant in *A. ostoyae* than in *A. cepistipes* in relation to virulence against Norway spruce seedlings from both low and high altitudes (Prospero *et al.*, 2004). Intraspecific variation in pathogenicity of *A. ostoyae* on maritime pine has also been reported by Labbé *et al.* (2017).

The ability of *Armillaria* spp. to exhibit different symbiosis with diverse plants and other fungi, and in various ecological niches depends on both biotic and abiotic factors. Biotic factors which influence host-specificity, pathogenicity and virulence of *Armillaria* spp. include possession of genes associated with specific growth mechanisms (e.g. rhizomorph production, branching, and other characteristics, vegetative and sexual reproduction mechanisms, etc.); expression of a large repertoire of plant cell wall degrading enzymes (PCWDE) and other enzymes; and production of secondary metabolites as well as other pathogenicity-related gene products (Bendel *et al.*, 2006; Collins *et al.*, 2013; Ferguson *et al.*, 2003; Koch *et al.*, 2017; König *et al.*, 2019; Li *et al.*, 2019; Mihail and Bruhn, 2005; Misiek *et al.*, 2011; Morrison, 2004; Morrison and Pellow, 2002; Peabody *et al.*, 2003; Sipos *et al.*, 2017). In terms of growth mechanisms, species with monopodially branched rhizomorphs such as *A. gallica* and *A. nabsnona* were shown to be less aggressive and virulent compared to species with dichotomously branched rhizomorphs such as *A. luteobubalina* and *A. limonea* against Douglas-fir and Garry oak (Morrison, 2004). These biotic factors are also either enhanced or limited based on the prevailing and/or changing environmental factors. For example, rhizomorph production by some *Armillaria* spp. is influenced by abiotic factors such as soil quality, temperature fluctuations and moisture content as has been documented for *A. luteobubalina* (Pearce and Malajczuk, 1990) and *A. mellea* (Baumgartner and Rizzo, 2001).

Due to the afore-described characteristics of *Armillaria* spp., members of the genus are of significant economic and agro-ecological importance. When they are parasitic, they cause massive devastation in horticulture, agriculture, as well as natural and managed plantation forests (Baumgartner and Rizzo, 2001; Baumgartner and Rizzo, 2002; Guillaumin *et al.*, 1993; Labbé *et al.*, 2015). However, *Armillaria* spp. enhance the forest ecosystem when they act as saprophytes (Legrand and Guillaumin, 1993). The forest ecosystem is enriched when saprobic microorganisms including *Armillaria* spp. degrade plant biomass with their extensive enzyme systems, resulting in significant contribution in releasing sequestered organic carbon into the carbon cycle (Lundell *et al.*, 2014; Sahu *et al.*, 2021). Moreover, like other saprobic Basidiomycota species, species in the genus are also edible. Hence, there is interest in domesticating species including *A. mellea* for culinary purposes (reviewed in (Ren *et al.*, 2022)). Also, the enzymes, secondary metabolites and other compounds produced by *Armillaria* spp. have the potential to be used in various industries such as the pharmaceutical industry. Hence, research on *Armillaria* spp. has focused on understanding the growth mechanisms in terms of *in situ* and *in vitro* mycelia, rhizomorph and fruit body growth, as well as pharmacological and other properties (reviewed in (Kedves *et al.*, 2021; Ren *et al.*, 2022)). Other research on these fungi has focused on mechanisms which drive their pathogenicity, and which can be exploited and/or harnessed to halt their growth (in the case of pathogenicity of socio-economically important plants) or enhance

their production of enzymes and other compounds of interest for biotechnological applications (reviewed in (Kedves *et al.*, 2021; Ren *et al.*, 2022)). Despite the research milestones made to better understand and limit or utilise these mushroom-forming fungi, there are still large knowledge gaps. Closing of these knowledge gaps is relevant as the current management practices have proven to be mostly inefficient in managing Armillaria root-rot disease (reviewed in (Heinzelmann *et al.*, 2019; Ren *et al.*, 2022)).

Armillaria species possess resources that facilitate their survival, and pathogenicity and/or virulence. To understand these mechanisms and to detect possible applications, genomics-based research have been conducted by various research groups on *Armillaria* spp. (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Heinzelmann *et al.*, 2020; Sahu *et al.*, 2021; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Wingfield *et al.*, 2022; Zhan *et al.*, 2020). These studies have given some insights into various characteristics of *Armillaria* spp., mostly from samples collected from countries in the northern hemisphere. In the present review, the extent of genomics studies on *Armillaria* species with a focus on the diversity of species and strains for which genomes have been assembled and annotated are considered. The steps, techniques and tools used for genome extraction, sequencing, assembly, evaluation, and annotation are discussed. Infographic guidelines for future *Armillaria* genome projects with some relevant resources have also been provided. Knowledge gained from these genome projects were also highlighted. We further suggest potential areas for future *Armillaria* genome projects as well as potential comparative genomics and multi-omics studies.

2. Status of genome sequencing and assembly of *Armillaria* spp.

There are presently 11 published whole genome sequences of *Armillaria* spp. in the NCBI Genome (http//www.ncbi.nlm.nih.gov/data-hub/genome/) and US DoE JGI fungal genomics resource MycoCosm databases (http//www.jgi.doe.gov/fungi/Agaricales) (Table 1). Other genomes have been uploaded to the JGI MycoCosm database but have not yet been published in journals and require permissions from the PIs for use. This review will focus on the published genomes of *Armillaria* spp. only and the knowledge gained from their genomic studies.

2.1. Starting materials and DNA extraction methods used in *Armillaria* genome projects

The starting material can influence the quantity and quality of DNA extracted for downstream processes. The quality of DNA extracted, in terms of chemical purity, structural integrity and size of the DNA, in a genome project influences the choice of sequencing technology that can be employed. These parameters also affect the quality of obtained data for downstream processes including genome assembly (Del Angel *et al.*, 2018; Henderson *et al.*, 2013). *Armillaria* genome projects have used fresh mycelia mostly obtained as liquid cultures in different media as the starting materials (Akulova

et al., 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Wingfield *et al.*, 2022; Zhan *et al.*, 2020) (Table 1). For the *A. borealis* strain AB13-TR4-IP16 genome project, fresh mycelium isolated from under the bark of infested *Abies sibirica* stems 50 cm above the soil surface was used as starting material (Akulova *et al.*, 2020). Relatively low molecular weight DNA will be sufficient for sequencing using Second-Generation Sequencing technologies (SGS; eg. Illumina, 454, SOLiD, and Ion Torrent). High molecular weight DNA in sufficient quantity and quality is required for Third-Generation Sequencing (TGS; e.g. Pacific Biosciences, PacBio and Oxford Nanopore Technologies, ONT). Typically, the minimum DNA amount required for the various sequencing technologies are ~ 3 ng, ~ 20 µg, ~ 1 µg, > 200 ng for Illumina, PacBio, ONT and BioNano respectively(Jung *et al.*, 2019). High molecular weight DNA is typically obtained from fresh material (Del Angel *et al.*, 2018). Other considerations for DNA quality requirements for *de novo* sequencing have been discussed by Del Angel *et al.* (2018).

DNA extraction for genome projects can be achieved using various protocols or DNA extraction kits. Akulova *et al.* (2020) used a modified CTAB (cetyltrimethylammonium bromide) method for extracting DNA from *A. borealis* strain AB13-TR4-IP16. The CTAB method and modifications of it have been used to extract DNA from fungal and plant samples, which characteristically contain high concentrations of impurities (RNA, polysaccharides, proteoglycans, proteins, secondary metabolites, polyphenols, and other contaminants) (Cubero *et al.*, 1999; Gawel and Jarret, 1991; Huang *et al.*, 2018; Porebski *et al.*, 1997; van Burik *et al.*, 1998). Commercial DNA extraction kits have also been used in *Armillaria* genome projects either according to the manufacturer's instructions or with some modifications (Caballero *et al.*, 2022; Collins *et al.*, 2013; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Wingfield *et al.*, 2022; Zhan *et al.*, 2020). These kits produce high quality DNA extracts but may differ in their efficiency in terms of the quality and quantity of DNA extracted in relation to the specific organism (Karstens *et al.*, 2021; Whitehouse and Hottel, 2007). Relatively high concentrations of some compounds considered to be impurities in DNA extracted for genome sequencing, including polysaccharides, proteins, secondary metabolites, and polyphenols have been reported in liquid cultures and fruiting bodies of *A. mellea* and *A. gallica* grown under different conditions and extracted with different solvents (An *et al.*, 2017; Engels *et al.*, 2011; Engels *et al.*, 2021; Lung and Chang, 2011; Zavastin *et al.*, 2015). For this reason, DNA extraction protocols or kits adopted for use in *Armillaria* genome projects should aim at reducing or eliminating these impurities to obtain high chemical purity, while maintaining the structural integrity and size of the DNA.

2.2. Choice of genome sequencing technologies

Armillaria genome projects have employed both SGS (i.e., Illumina HiSeq and Illumina MiSeq) and TGS (i.e., PacBio RS and PacBio SMRT) technologies (Table 1). These technologies have their respective strengths and limitations as reviewed by several authors (Alkan *et al.*, 2011; Athanasopoulou *et al.*, 2022; Del Angel *et al.*, 2018; Giani *et al.*, 2020; Schatz *et al.*, 2010; Slatko *et al.*, 2018; Wee *et al.*, 2019). Although the SGS technologies are relatively cheaper, generation of long contiguous DNA sequence reads during genome assembly is often impaired due to short read lengths, as well as the presence of repeat elements and polymorphisms (Alkan *et al.*, 2011; Giani *et al.*, 2020; Schatz *et al.*, 2010). Genome assemblies from reads obtained from SGS technologies also result in loss of as much as 99.1% of validated duplicated sequences (Alkan *et al.*, 2011). These challenges are largely circumvented by TGS which provide long reads for better representation of the genome being assembled, although high rates (5 – 20 %) of sequencing error of raw reads are obtained with this technology (Alkan *et al.*, 2011; Athanasopoulou *et al.*, 2022; Giani *et al.*, 2020; Wee *et al.*, 2019).

Five of the *Armillaria* genome projects utilized both SGS and TGS technologies for hybrid *de novo* genome assemblies (Table 1). These were *A. cepistipes* isolate B5, *A. fuscipes* strain CMW 2740, *A. gallica* strain 012m, *A. ostoyae* strain C18/9, and *Armillaria* African Clade B sp. strain CMW4456. Hybrid approaches were followed to minimize the effects of individual use of SGS and TGS technologies on the quality and contiguity of the assembled genomes. Four of these genome projects used the Illumina reads for error correction or polishing after the PacBio read assembly (Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Zhan *et al.*, 2020) while the reverse approach was used for assembling the *A. fuscipes* strain CMW 2740 genome (Wingfield *et al.*, 2022). This hybrid genome assembly method has been used in other fungal genome projects such as the *Fusarium equiseti* genome project (Li *et al.*, 2021) and is the most frequently used combination of techniques in bacterial RefSeq genome projects (Segerman, 2020). Another combination of genome assembly techniques is the use of Illumina and MinION reads (Minei *et al.*, 2018; Segerman, 2020). MinION is another ONT, TGS technology (Tyler *et al.*, 2018). This combination has not yet been used in *Armillaria* genome projects and might be useful for assembling good quality genomes in future.

2.3. Raw reads quality control

Quality control (QC) and reads correction pre-assembly is highly important as this can impact the quality of the final assembled genome. Methods or bioinformatic tools used for the quality control steps in the published *Armillaria* genome projects were, however, not reported. Nonetheless, various tools available for this purpose include FastQC, FastQE and Falco (FastQC Alternative Code), NanoPlot, PycoQC, LongQC, and MultiQC (Figure 1; Table 2). These tools assess quality metrics including base or sequence quality of the reads, presence of adapter sequences and GC distribution.

The choice of the tools used for raw reads QC is usually informed by the sequencing technology used and the number of genomes being assembled. Additionally, raw reads QC is sometimes integrated into genome assembly pipelines. As such, the decision of whether to perform a separate QC step should consider the genome assembly tool or pipeline to be used in the next steps.

2.4. Pre-assembly reads correction

Pre-assembly correction of raw reads is optional in genome projects because this step depends on the quality of the raw reads obtained, the purpose of the genome project, computational time, as well as the choice of genome assembly tool or workflow in the next step (Del Angel *et al.*, 2018; Yang *et al.*, 2019). Some genome assembly tools or workflows incorporate reads correction. Hence, a separate reads correction is not needed in this case.

Tools such as LoRDEC (Salmela and Rivals, 2014) and Trimmomatic (Bolger *et al.*, 2014) have been used for reads correction during the *Armillaria* genome projects (Table 1). LoRDEC is mostly intended to correct PacBio reads using Illumina reads as reference. Trimmomatic performs trimming of adapters and other Illumina-specific sequences, as well as quality and length trimming for Illumina paired-end and single-end reads (Bolger *et al.*, 2014). The order of the steps used during reads correction with Trimmomatic is important and depends on the quality of the raw sequence data. Other available tools for reads profiling and correction include Trim Reads function in CLC Genomics Workbench (QIAGEN, Aarhus), Cutadapt (Martin, 2011), Filtlong (https://github.com/rrwick/Filtlong), and Porechop (https://github.com/rrwick/Porechop) (Figure 1; Table 2). These tools also perform adapter trimming, quality trimming and length trimming for reads obtained from SGS, TGS or both. The choice of which tool(s) to use depends on the sequencing technology used, the genome assembly pipeline and other considerations, some of which are outlined in Figure 1.

2.5. Genome assembly tools, pipelines, or workflows

De novo genome assembly using various genome assembly tools has been used as opposed to reference-based assembly in *Armillaria* genome projects (Table 1). Genome assembly tools that were used include the now obsolete Celera Assembly Pipeline (Myers *et al.*, 2000), CLC Genomics Workbench v 22.0.1, SPAdes (Bankevich *et al.*, 2012), Velvet (Zerbino and Birney, 2008), MIRA (Chevreux *et al.*, 1999), ALLPATHS-LG (Gnerre *et al.*, 2011), MECAT2 (Xiao *et al.*, 2017) and Pacific Biosciences SMRT Portal software (Table 1). These tools have their respective strengths and limitations. For the specific functions and considerations for selecting and working with specific genome assembly tools, pipelines, or workflows see the links, references and workflow provided in Figure 1 and Table 2.

De novo genome assembly is impacted by computational methods used due to constant technological advancement and specific limitations of respective genome assemblers. The nature and quantity of sequence data, which relies on the features of the genome being assembled, also impacts *de novo* genome assembly (Forouzan *et al.*, 2017; Gnerre *et al.*, 2011; Gupta and Kumar, 2022). Differences in various genome assemblers in terms of memory consumption, total assembling time, and accuracy based on genome quality matrices (contig number, N50 values, and genome fractions) using prokaryotic and eukaryotic Illumina paired-end and single-end reads have been evaluated by Khan *et al.* (2018). The differences of the various assemblers as well as the unique biological characteristics of the genome being assembled requires that correct assembler versions should be selected, and various assemblers and assembly parameters should be evaluated in genome projects. Outputs of the various assemblies should be compared in each genome project to determine the optimum assembly for the specific genome. Nevertheless, knowing when to stop while chasing perfection in a genome project is vital. It is essential to consider the complexity of the genome to be assembled (i.e., genome size, amount and distribution of repeat sequences, heterozygosity, ploidy, and homogeneity of GC content), the research questions, and available resources (e.g., availability of a reference genome, project budget, and time and computational constraints) in the decision-making process (Del Angel *et al.*, 2018; Salzberg *et al.*, 2012; Wee *et al.*, 2019).

2.6. Post-assembly processing

Post-assembly processing involves sequence error correction, polishing, and haplotype phasing. Assembly errors are introduced especially when long reads are assembled. Hence, post-assembly error correction and polishing is highly recommended to improve the quality of the overall local base accuracy and resolve some mis-assemblies of the assembled genome. This can be achieved by polishing the assembly with short or long reads (Del Angel *et al.*, 2018). Tools which have been used in *Armillaria* genome projects for this purpose include NextPolish (Hu *et al.*, 2020), Pilon (Walker *et al.*, 2014), FinisherSC (Lam *et al.*, 2015), PBJelly2 (English *et al.*, 2012), RS Resequencing application (SMRT Analysis suite) (https://www.pacb.com/wp-content/uploads/SMRT-Link-Getting-Started-Guide-v4.0.0.pdf), and VelvetOptimiser (http//github.com/Slugger70/VelvetOptimiser) (Table 1). The functions and differences among these and other tools used for post-assembly processing of assembled genomes can be accessed via the links and references provided in Table 2.

Haplotype phasing is the identification of alleles that are co-located on the same chromosome. Haplotype phasing and removal of the identified heterozygous overlaps in assembled genomes is relevant to diploid or polyploid genomes and in genomes which have regions of high heterozygosity (Zhang *et al.*, 2020). This is because most algorithms in diploid-aware assemblers (e.g., Canu) will

assemble these regions as separate contigs, rather than the expected single haplotype-fused contig once a pair of allelic sequences exceeds a certain threshold of nucleotide diversity (Roach *et al.*, 2018). *Armillaria* spp. usually exist as prolonged diploid vegetative states (reviewed in (Heinzelmann *et al.*, 2019)). Hence, most sequenced genomes are diploid and warrant removal of heterozygous overlaps in assembled *Armillaria* genomes. This can be achieved by utilizing standalone tools such as Purge Haplotigs as demonstrated in the *Armillaria gallica* strain 012m genome project (Table 1). Purge Haplotigs automates the reassignment of allelic contigs and assists with manual curation of TGS-based genome assemblies specifically (Roach *et al.*, 2018). As outlined in Figure 1 and Table 2, other tools are available for haplotype phasing and differ in how they function (see provided tool websites and/or references).

2.7. Genome quality evaluation

The quality of an assembled genome must be evaluated after genome assembly. This evaluation should be conducted in terms of genome completeness, the least amount of misassemblies, and presence of contaminants. Metrics such as N50 sizes and L50 are indicative of the level of contiguity (high N50 and L50 values) of the genome assembly (Salzberg *et al.*, 2012; Yandell and Ence, 2012). N50 sizes are the contig/scaffold size above which half the genome is represented, whereas L50 is the length in bases of the shortest contig for which 50% of the genome can be contained within contigs of that size or larger. These two metrics are computed based on the lengths of assembly scaffolds or contigs (Salzberg *et al.*, 2012). Various tools can be used for genome assembly evaluation. *Armillaria* genome projects have used tools such as the now defunct CEGMA (Core Eukaryotic Genes Mapping Approach), BUSCO (Benchmarking Universal Single-Copy Orthologs), QUAST (QUality ASsessment Tool) and RS BridgeMapper protocol (SMRT Analysis Suite) to evaluate the completeness and quality of the respective genome assemblies (Table 1). BUSCO analyses of *Armillaria* genome projects have utilized different versions of BUSCO (versions 5.3.2 and 3.0.1) and different lineage-specific datasets (agaricales_odb10 lineage and basidiomycota_odb9 datasets) (Table 1). Results generated by BUSCO depends on the version of database of the lineage used for evaluation of the sequenced organism, as well as the version of the tool itself. For example, the basidiomycota_odb9 dataset contained 1,335 BUSCO groups while the basidiomycota_odb10 dataset contains 1,764 BUSCO groups. Moreover, the agaricales_odb10 and eukaryota_odb10 datasets contain 3,870 and 255 BUSCO groups respectively. BUSCO v5.4.4 is the current stable version (http//busco.ezlab.org/). Consequently, care should be taken when comparing genome completeness based on BUSCO analyses from different *Armillaria* genome projects.

In addition to evaluating genome completeness and correctness, genomes should be evaluated for absence of contaminants (i.e., non-target sequences). Presence of genome contamination can occur at

various stages of a genome project (reviewed in (Cornet and Baurain, 2022)). These contaminants can be detected and removed with various tools which function by using several parameters including detection of a highly skewed (i.e., extremely low or high) GC content, presence of an uncharacteristically high percentage of duplicated complete BUSCO genes, and results of BLAST searches of sections of the genome in the NCBI database (reviewed in (Cornet and Baurain, 2022)). BlobTools (Laetsch and Blaxter, 2017) and Anvi'o (Eren *et al.*, 2015) have been used to detect and remove genome contamination from an assembled fungal genome (Aylward *et al.*, 2022). Besides BUSCO, none of the indicated tools were used in the *Armillaria* genome projects as genome contaminants may have been absent in the respective assemblies.

In summary, genome quality evaluation should be conducted in all genome projects. Genomes that are seen to have unsatisfactory completeness and correctness metrics can be reassembled with other tools or workflows with altered assembly parameters and the best assembly in terms of these metrics selected as the final assembly. Minimal contamination can be resolved with tools such as the ones indicated in the review article by Cornet and Baurain (2022), some of which have been outlined in the present review (Figure 1 and Table 2). The purpose for the genome project can be used to select the best assembly in genome assembly quality evaluation (e.g., selecting between the assembly with the highest contiguity and the assembly with the highest percentage of complete genes). This is important as assembly quality is not well correlated with assembly contiguity (Salzberg *et al.*, 2012). Additionally, it is apparent that genome quality evaluation technologies, both in terms of computation and datasets used, are being rapidly developed and/or enhanced. Hence, research teams embarking on genome projects should keep abreast with the state-of-the-art for genome assembly and quality evaluation. Tools used for genome quality evaluation in *Armillaria* genome projects and other available tools compute different metrics which can be accessed via the information provided in Table 2. In the case of very low genome completeness and correctness as well as high contamination, it would be more prudent to sequence and assemble a new genome.

3. *Armillaria* transposable element and genome annotation

Genome annotation is required to attach biologically meaningful information to assembled genomes, which are otherwise of no significant value to biologists (Del Angel *et al.*, 2018). This process consists of annotation of groups of mobile genetic elements called TEs (transposable elements). as well as gene prediction through *ab initio* approaches with or without addition of wet-bench experimental data (Del Angel *et al.*, 2018). Wet-bench experimental data for genome annotation include Sanger sequenced ESTs (expressed sequence tags), RNA-Seq (RNA sequencing) or even long read sequenced transcripts (Del Angel *et al.*, 2018). Genome annotation also involves describing the structures and functions of the products of the predicted genes. The reader is referred to two book

chapters by de Sá *et al.* (2018) and Christoffels and van Heusden (2019) as well as the review articles by Yandell and Ence (2012) and Del Angel *et al.* (2018) for detailed steps used for genome annotation. Nevertheless, the methods and depth of *Armillaria* genome annotation have been outlined in Figure 2 and described in this section.

Gene and functional annotation of assembled genomes is preceded by identification and masking of RNA genes, detection and annotation of TEs and other repetitive elements and detection of low complexity DNA regions. TEs include Long Terminal Repeats (LTRs) and their annotation is considered a major task in any genome project (reviewed in (Del Angel *et al.*, 2018; Jung *et al.*, 2020)). RNA gene identification and masking, and detection and annotation of TEs were carried out in all the *Armillaria* genome projects excluding the *Armillaria* African Clade B sp. [CMW4456] genome project using the tools indicated in Figure 2. Tools employed in *Armillaria* genome projects for identification and masking of RNA genes in the assembled genome include RNAmmer V1.2, Rnnotator, SortMeRNA, and tRNAscan-SE (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Sipos *et al.*, 2017; Zhan *et al.*, 2020). *De novo* detection and annotation of transposable elements, other repetitive elements, and low complexity DNA regions have been achieved with tools including BLASTER, Grouper, LTRHarvest, PASTEC, Piler, Recon, RepBase, RepeatMasker, RepeatModeler, REPET and TEclass (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Zhan *et al.*, 2020). Although not clearly indicated in the *Armillaria* African Clade B sp. [CMW4456] genome project, the AUGUSTUS annotation tool on the Galaxy platform has an option called Softmasking. Repeated regions in the genome assembly are detected as lowercase letters when this option is enabled (The Galaxy Community, 2022).

Gene and functional annotation of *Armillaria* genomes has either been restricted to single *ab initio* gene prediction only or has involved full-scale annotation with or without wet-bench experimental evidence and optional manual curation of gene models (Figure 2). For prediction and annotation of genes and proteins, tools including AUGUSTUS, Blast2GO, dbCAN2 server, Exonerate, Fgenesh, GETA, PHI base, and SNAP have been employed (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Wingfield *et al.*, 2022; Zhan *et al.*, 2020). Some of these tools have been utilised in pipelines such as BRAKER2 and Maker (Sipos *et al.*, 2017).

Functional annotation of some of the *Armillaria* genomes have incorporated the use of RNA Seq data (Akulova *et al.*, 2020; Sipos *et al.*, 2017; Zhan *et al.*, 2020). Annotation with wet-bench experimental data involves transcript assembly using RNA Seq data. This has been achieved with Trinity and Rnnotator in the *Armillaria* genomes annotated by Sipos *et al.* (2017) and SortMeRNA v2.1 program in the genome annotation reported by Akulova *et al.* (2020). The tool used for RNA Seq data assembly

was not specified in the genome project by Zhan *et al.* (2020). Following the use of multiple *ab initio* predictors, genome annotation in the genome projects reported by Sipos *et al.* (2017) involved manual curation of the predicted gene models using tools or pipelines such as Gbrowse, FunyBASE, PASTEC and PEDANT system. The authors used the JGI Annotation Pipeline for annotation of the *A. gallica* isolate Ar21-2 and *A. solidipes* isolate C28-4 genomes.

Further genome annotation has been conducted in *Armillaria* genome projects to gain more insights about these fungi. For this purpose, tools such as BLASTp, GeneMark-ES and GeneMark-ET have been used for gene and protein ortholog finding (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Sipos *et al.*, 2017). Motifs and domains of proteins have been detected with two versions of InterProScan in the *A. altimontana* (NABS X) isolate 837–10, *A. gallica* strain 012m and *A. solidipes* (form *A. ostoyae*) isolate ID001 genome projects (Caballero *et al.*, 2022; Zhan *et al.*, 2020). BLASTp, SecretomeP server and SignalP were used for putative gene function predictions based on databases and/or protein evidence (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Sipos *et al.*, 2017). Only the *A. altimontana* (NABS X) isolate 837–10 and *A. solidipes* (form *A. ostoyae*) isolate ID001 genome projects predicted subcellular localizations of the genes and annotated the secondary metabolite biosynthesis gene clusters (Caballero *et al.*, 2022). These were achieved with Deeploc 1.0 and antiSMASH, respectively. Evaluation of the annotations were achieved using BUSCO in the genome projects reported by Sipos *et al.* (2017) and Caballero *et al.* (2022).

It is apparent that genome annotation in *Armillaria* genome projects has ranged from minimal annotation to in-depth annotation. In addition to the tools which have been used for annotation of *Armillaria* genomes, some other tools required for genome annotation have been outlined and can be accessed at the links and/or references provided in Table 2. The investments in terms of time, effort and evidence with a consequence on the project budget needed for genome annotation increases with increasing accuracy of the annotation, as explained by Yandell and Ence (2012). Therefore, it is relevant to consider the project budget and other resources including time and computational constraints, as well as the research questions to determine whether to only perform *in silico* gene and functional annotation or to include wet-bench experimental evidence with extensive annotations. Further considerations for genome annotation have been discussed in Christoffels and van Heusden (2019). Annotation of all the *Armillaria* genomes have been used to gain some knowledge about the biology of these fungi. Some of these have been discussed in the next section.

4. Knowledge gained from *Armillaria* genomics

Armillaria genome projects have provided draft and complete genome assemblies (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Heinzelmann *et al.*, 2020; Sipos *et al.*, 2017; Wingfield *et al.*, 2016; Wingfield *et al.*, 2022; Zhan *et al.*, 2020). Standardized QUAST and BUSCO analyses of the presently published *Armillaria* genomes were conducted for accurate comparison in this review (Table 3). This comparison shows that the *Armillaria* genomes, ranged between approximately 50.71 – 87.31 Mbp. These genome sizes are comparable among the *Armillaria* spp. and are within the reported genome sizes (approximately 28 – 175 Mb) of other Agaricales genomes (Ruiz-Duenas *et al.*, 2021). Even at the draft stage, *Armillaria* genomes are highly valuable resources with sufficiently high contiguity (N50 values greater than 5000 bp) and completeness (over 98% complete BUSCOs) and thus can be used for some comparative genomics and other studies.

Some of the *Armillaria* genomes have been extensively annotated. Transposable elements (TEs) annotation of *Armillaria* genomes by Sipos *et al.* (2017) revealed considerable variation in TEs among the different genomes and suggested that unlike other plant pathogens, TEs do not account for genome expansion in *Armillaria*. Genome expansion of rust genomes, including that of *Austropuccinia psidii*, have been attributed to TEs (Tan *et al.*, 2014; Tobias *et al.*, 2021). LTRs were the most highly represented TE family in the genome of *A. borealis* isolate AB13-TR4-IP16, with Gypsy (32) being the most abundant LTR type followed by Copia (5) (Akulova *et al.*, 2020). Other repetitive elements reported in this genome included LINE-retrotransposons and retrotransposons with a tyrosine recombinase. Similarly, the most abundant LTR type in the genome of *A. gallica* strain 012m was Gypsy (2062) followed by Copia (321) (Zhan *et al.*, 2020). This trend has been reported in other Basidiomycetes genomes (Castanera *et al.*, 2017). Although the functions of these TEs have not yet been identified in *Armillaria*, these mobile genetic elements have been speculated or shown to have various functions in genomes of other organisms. These functions include acting as mutators, contributing to genome restructuring, and participation in the evolutionary "arms race" between hosts and pathogens (reviewed in (Kidwell and Lisch, 2001)). Nevertheless, the contribution of mobile genetic elements on size variation in the mitochondrial genomes (mitogenomes) of *A. borealis, A. gallica, A. sinapina,* and *A. solidipes* was discovered through comparative mitogenomics (Kolesnikova *et al.*, 2019). It is relevant to conduct further studies on the functions of TEs in the genomes of *Armillaria* spp. to gain a better understanding of the effects of TEs on evolution, ecology, and other characteristics of these fungi.

The number of protein-coding genes identified in all the genomes ranged from 13,600 in *Armillaria* African Clade B sp. [CMW4456] to 26,261 in *A. gallica* [012m] (Akulova *et al.*, 2020; Caballero *et al.*, 2022; Collins *et al.*, 2013; Heinzelmann *et al.*, 2020; Sipos *et al.*, 2017; Wingfield *et al.*, 2016;

Wingfield *et al.*, 2022; Zhan *et al.*, 2020). The range of predicted gene contents of the *Armillaria* genomes are within the range reported by Ruiz-Duenas *et al.* (2021) for other Agaricales genomes (between 5,000 and 29,000 genes). Although the effect of gene contents (in terms of number of genes) and genome size have not been evaluated in *Armillaria*, Ruiz-Duenas *et al.* (2021) reported that these properties do not significantly correlate to the lifestyles exhibited by the Agaricales.

Information gained from *Armillaria* genomics have contributed to the body of knowledge regarding their genome evolution and architecture as well as biology. For instance, Sipos *et al.* (2017) have reported that there is a significant genome expansion in *Armillaria* spp. compared to other related fungi with a consequent effect on various properties including gene family expansion of pathogenicity-related genes. There are also records of many relatively small differences in gene content of genomes of *Armillaria* spp. which potentially contribute to differences in their ecological lifestyles (Caballero *et al.*, 2022; Heinzelmann *et al.*, 2017; Heinzelmann *et al.*, 2020; Sahu *et al.*, 2021; Sipos *et al.*, 2017). All these records point to genome plasticity of *Armillaria* spp. which can mediate adaptability to changing environments and resistance to or evasion of host defence mechanisms. These changes can occur through gene inactivation, altered gene sequence or structure, and the birth of new genes which often encode proteins involved in host-pathogen interactions as reviewed by Raffaele and Kamoun (2012).

The availability of genome sequences and consequent genome annotations has allowed for phylogenomic studies on *Armillaria* spp. In this way the position of *Armillaria* in the Physalacriaceae and existence of *Guyanagaster* and *Cylindrobasidium* as the closest relatives were confirmed using 835 conserved single copy genes (188,895 amino acid sites). Furthermore, the age of pathogenic *Armillaria* spp. was estimated at 21 million years (Myr) and their divergence from *Guyanagaster* is reckoned to be 42 Myr (Sipos *et al.*, 2017). Phylogenomics, by virtue of using large numbers of orthologous genes, provides an advantage over traditional phylogenetic methods that employ few loci for obtaining well-supported and accurate species delimitation (Pamilo and Nei, 1988). This is explained by the fact that individual gene genealogies may differ from each other and from the organismal phylogeny and therefore affects the phylogeny obtained when the analysis is based on few genes (Pamilo and Nei, 1988). However, use of more genes or loci is not always better in phylogenomics. This is because the possibility of obtaining highly supported but incorrect phylogenetic results, that are due to inconsistency resulting from systematic error from nonphylogenetic signals or model inadequacy, increases when large amounts of data are included in phylogenomic analyses (Delsuc *et al.*, 2005; Young and Gillung, 2020). To minimise this risk, the number of genes or loci used in phylogenomic analyses should ideally be proportional to the number of taxa being studied (Delsuc *et al.*, 2005; Young and Gillung, 2020). In the context of *Armillaria*,

the need for more resources and techniques for species delimitation has been highlighted in various review articles including Baumgartner *et al.* (2011), Coetzee *et al.* (2018), and Heinzelmann *et al.* (2019). The availability of genome sequences and the possibility of sequencing more genomes offer more opportunities for bridging this gap. Several genomes of *Armillaria* spp. with diverse geographical distribution will be required for further phylogenomic and/or phylogeographic studies to fully delimit species and species complexes among *Armillaria*, detect trends in horizontal gene transfer, decipher other evolutionary history of these fungi, and to use multispecies phylogenetic comparisons to infer putative functions for DNA or protein sequences.

Genomic studies have also resulted in deciphering some of the molecular basis of sexual reproduction in *Armillaria* spp. The mating (*MAT*) loci of *A. solidipes, A. ostoyae, A. cepistipes,* and *A. gallica* have been reported (Sipos *et al.*, 2017). Further insights were gained about the chromosomal locations of the *MAT-A* and *MAT-B* loci with the establishment of the only near-chromosome scale genome of an *Armillaria* spp., namely *A. ostoyae*. This research led to the discovery of the location of the *MAT* loci on pseudochromosomes LG1 and LG6, respectively (Heinzelmann *et al.*, 2020). Fungal sexual reproduction influences genetic recombination with a consequent effect on genetic diversity. These mechanisms can play a role in the evolutionary "arms race" between fungal phytopathogens and their hosts, as well as adaptation to hostile environments (Anderson *et al.*, 2010; Grandaubert *et al.*, 2019; Zhan *et al.*, 2007). Hence, more information about the *MAT* loci is relevant for understanding some of the mechanisms underlying the broad host range, host-pathogen interactions, and global distribution of *Armillaria* spp.

Discoveries made about some important characteristics of *Armillaria* spp. have benefited from a multi-disciplinary approach which incorporated genomics with other experimental investigations. For instance, through multi-omics studies including transcriptomics and proteomics, Sipos *et al.* (2017) reported that *A. cepistipes, A. ostoyae,* and *A. gallica* possess more pectinolytic gene families than lignolytic gene families compared to other white-rot fungi. The authors suggest that these gene families may enable *Armillaria* spp. to avoid competition with other white-rot fungi, enhancing their ability to colonize and thrive on some hosts. Differential regulation of the genes encoding these biotic factors were also reported in various tissues (mycelia, rhizomorphs, fruiting bodies at different developmental stages, and in different tissues of a mature fruiting body) of *A. ostoyae* with a potential influence on multicellular development as well as cap and stipe differentiation (Sipos *et al.*, 2017).

Other important discoveries have been made about *Armillaria* spp. based on multi-omics and biological investigations. Combining genomics with *in vitro* analyses of recombination rate variation, Heinzelmann *et al.* (2020) discovered 1,984 crossover events in the progeny of *A. ostoyae*, highly variable recombination rates along chromosomes, and regions close to the telomeres serving as gene-

poor recombination hotspots. Multiple glycol-degradative enzyme systems which can be useful for recalcitrant carbon compound degradation, and potent antifungal properties of *A. mellea* were also discovered based on genomic and proteomic studies (Collins *et al.*, 2013). Moreover, comparative genomics coupled with *in vitro* bioassays resulted in records of conserved putative NRPS-dependent siderophore synthetase gene clusters and *in vitro* species- and strain-independent biosynthesis of different types and quantities of siderophores by strains of various *Armillaria* spp. (Narh Mensah *et al.*, 2023). Hence, while genomic data can provide some answers, using genome data in combination with *in vitro, in situ* and *in plantae* studies including other bioinformatics studies can provide a much greater understanding of the species in this genus. This multi-disciplinary approach to investigating fungal phytopathogens has also been recommended in the review article by Aylward *et al.* (2017).

5. Suggestions for future genomics studies on the genus *Armillaria*

Armillaria genome projects have yielded 11 high quality publicly available genomes at the contig or scaffold levels on the NCBI database. These genomes have been applied in further research to gain a better understanding of these fungi. Nonetheless, there is still more to be achieved in terms of genomics and multi-omics research on *Armillaria* spp. In this section, prospects in *Armillaria* genomics are suggested.

So far, the genome of *A. ostoyae* strain SBI C18/9 is the only *Armillaria* genome which has been assembled to pseudochromosome (near chromosome) level. The pseudochromosomes were constructed based on the order of scaffolds within genetic map linkage groups of haploid progenies of *A. ostoyae* strain C15 (Heinzelmann *et al.*, 2020). The VGP (Vertebrate Genomes Project) assembly pipeline was designed for generating high-quality, near-error-free, gap-free, chromosome-level, haplotype-phased, annotated reference genome assemblies for every vertebrate species (Rhie *et al.*, 2021). This pipeline is hosted on the Galaxy platform and uses PacBio or Nanopore reads and additional techniques such as Hi-C sequencing (Rhie *et al.*, 2021; The Galaxy Community, 2022). Hi-C sequencing is a supporting technology used to improve the contiguity of genome assemblies (Del Angel *et al.*, 2018; Lieberman-Aiden *et al.*, 2009). It is expected that the VGP pipeline, will be suitable for assembling diploid, highly heterozygous *Armillaria* genomes to chromosomal level and should be explored in future. In the absence of Hi-C and other supporting data, the workflow in Figures 1 and 2 can be used as guidelines for embarking on more *Armillaria* genome projects to achieve optimum genome assemblies and annotation given available resources. Resources for accessing the tools indicated in Figures 1 and 2 as well as other resources have been provided in Table 2.

Other discoveries can be achieved with more studies based on *Armillaria* genomes. These discoveries can be facilitated by combining genomics with other experimental studies. The benefits of availability of genomes of numerous isolates of fungal phytopathogens coupled with other wet-bench experimental studies have been discussed (Aylward *et al.*, 2017; Coetzee *et al.*, 2020). In the context of *Armillaria*, embarking on assembling more high-quality genomes of several isolates belonging to various species and with a worldwide geographical sampling promises to be highly useful, especially if a multi-disciplinary approach is adopted. Such studies will help bridge the wide knowledge gap about the mechanisms used by these fungi in their evolutionary success. Additionally, the considerable variation in both intra- and inter-species aggressiveness, pathogenicity, and virulence among pathogenic members of *Armillaria* is still not clearly understood (Labbé *et al.*, 2017; Morrison and Pellow, 2002; Prospero *et al.*, 2004). The mechanisms underlying these characteristics of *Armillaria* can also be deciphered with the multi-disciplinary approach.

Some other possible research areas that can be facilitated by the multi-disciplinary approach include the following:

1. Putatively identifying genes encoding enzymes involved in secondary metabolite biosynthesis and deciphering the expression patterns of genes encoding secondary metabolites in vegetative mycelia, rhizomorphs, and fruiting bodies at different developmental stages of different *Armillaria* spp. under different growth conditions. This will help further understand the role different developmental stages and different tissues play in *Armillaria* pathogenicity.
2. Further research aimed at establishing and understanding the mechanisms of antibacterial, antifungal, and cytotoxic properties of secondary metabolites produced by *Armillaria* spp.
3. Further investigation of the molecular basis of the mating system of *Armillaria* spp. or isolates, for instance, to unravel the mechanisms of diploidization of the haploid mycelia and the signals that trigger these mechanisms.
4. Developing and/or optimizing tools for functional validation of genes in *Armillaria* for study of gene function and annotation in *Armillaria*.
5. Conducting more research on the secretion, presence, structure, and activity of ligninolytic enzymes, other PCWDE and other Carbohydrate-Active Enzymes (CAZymes) of *Armillaria* spp. *in silico, in vitro* and *in situ.*
6. Comparing the quantity and diversity of pathogenicity related gene products such as PCWDE in *Armillaria* spp. with various lifestyles *in silico, in vitro* and *in plantae* to establish the effect of enzyme machinery possessed by different *Armillaria* spp. on their pathogenicity and how

they circumvent or out-compete other sympatrically existing *Armillaria* spp. as well as microbes normally used as biocontrol agents of Armillaria root-rot disease. This information can be used for developing enhanced biocontrol methods of Armillaria root-rot.

Conclusions

Armillaria genome projects have achieved several milestones towards understanding the biology and evolution of these fungi. Advantage should be taken of the fast pace of technological advancement in genome sequencing, assembly and annotation, and the reducing cost of these technologies to increase genome projects in *Armillaria* and other Basidiomycetes. These should be coupled with other wet-bench experimental studies in a timely manner. Results from these studies can afford the scientific, agro-industrial, and pharmaceutical sectors the benefits of the knowledge and technologies that can be gained from these projects.

References

Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25(10)**,** 1329-1330. doi: 10.1093/bioinformatics/btp084.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46(W1)**,** W537-W544. doi: 10.1093/nar/gky379.

Akulova, V.S., Sharov, V.V., Aksyonova, A.I., Putintseva, Y.A., Oreshkova, N.V., Feranchuk, S.I., et al. (2020). *De novo* sequencing, assembly and functional annotation of *Armillaria borealis* genome. *BMC Genom.* 21(Suppl 7)**,** 534. doi: 10.1186/s12864-020-06964-6.

Alkan, C., Sajjadian, S., and Eichler, E.E. (2011). Limitations of next-generation genome sequence assembly. *Nat. Methods* 8(1)**,** 61-65. doi: 10.1038/nmeth.1527.

Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33(21)**,** 3387-3395. doi: 10.1093/bioinformatics/btx431.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215(3)**,** 403-410. doi: 10.1016/S0022-2836(05)80360-2.

An, S., Lu, W., Zhang, Y., Yuan, Q., and Wang, D. (2017). Pharmacological basis for use of *Armillaria mellea* polysaccharides in Alzheimer's Disease: Antiapoptosis and antioxidation. *Oxid. Med. Cell. Longev.* 2017**,** 4184562. doi: 10.1155/2017/4184562.

Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., and Singh, K.B. (2010). Plants versus pathogens: an evolutionary arms race. *Funct. Plant Biol.* 37(6)**,** 499-512. doi: 10.1071/FP09304.

Athanasopoulou, K., Boti, M.A., Adamopoulos, P.G., Skourou, P.C., and Scorilas, A. (2022). Third-Generation Sequencing: The spearhead towards the radical transformation of modern genomics. *Life* 12(1)**,** 30.

Aylward, J., Steenkamp, E.T., Dreyer, L.L., Roets, F., Wingfield, B.D., and Wingfield, M.J. (2017). A plant pathology perspective of fungal genome sequencing. *IMA Fungus* 8(1)**,** 1-15. doi: 10.5598/imafungus.2017.08.01.01.

Aylward, J., Wingfield, M.J., Roets, F., and Wingfield, B.D. (2022). A high-quality fungal genome assembly resolved from a sample accidentally contaminated by multiple taxa. *Biotechniques* 72(2)**,** 39-50. doi: 10.2144/btn-2021-0097.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5)**,** 455-477. doi: 10.1089/cmb.2012.0021.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6(1)**,** 11. doi: 10.1186/s13100-015-0041-9.

Bao, Z., and Eddy, S.R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* 12(8)**,** 1269-1276. doi: 10.1101/gr.88502.

Baumgartner, K., Coetzee, M.P.A., and Hoffmeister, D. (2011). Secrets of the subterranean pathosystem of *Armillaria*. *Mol. Plant Pathol.* 12(6)**,** 515-534. doi: 10.1111/j.1364-3703.2010.00693.x.

Baumgartner, K., and Rizzo, D.M. (2001). Ecology of *Armillaria* spp. in mixed-hardwood forests of California. *Plant Dis.* 85(9)**,** 947-951. doi: 10.1094/pdis.2001.85.9.947.

Baumgartner, K., and Rizzo, D.M. (2002). Spread of *Armillaria* root disease in a California vineyard. *Am. J. Enol. Vitic.* 53(3)**,** 197-203.

Bendel, M., Kienast, F., and Rigling, D. (2006). Genetic population structure of three *Armillaria* species at the landscape scale: a case study from Swiss *Pinus mugo* forests. *Mycol. Res.* 110(6)**,** 705-712. doi: 10.1016/j.mycres.2006.02.002.

Bendtsen, J.D., Jensen, L.J., Blom, N., von Heijne, G., and Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* 17(4)**,** 349-356. doi: 10.1093/protein/gzh037.

Bendtsen, J.D., Kiemer, L., Fausbøll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5**,** 58. doi: 10.1186/1471-2180-5-58.

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, Marnix H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49(W1)**,** W29-W35. doi: 10.1093/nar/gkab335.

Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49(D1)**,** D344-D354. doi: 10.1093/nar/gkaa977.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4)**,** 578-579. doi: 10.1093/bioinformatics/btq683.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15)**,** 2114-2120. doi: 10.1093/bioinformatics/btu170.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12(1), 59-60. doi: 10.1038/nmeth.3176.

Caballero, J.R.I., Lalande, B.M., Hanna, J.W., Klopfenstein, N.B., Kim, M.S., and Stewart, J.E. (2022). Genomic comparisons of two *Armillaria* species with different ecological behaviors and their associated soil microbial communities. *Microb. Ecol.* doi: 10.1007/s00248-022-01989-8.

Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* 48, 4 11 11-14 11 39. doi: 10.1002/0471250953.bi0411s48.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18(1), 188-196. doi: 10.1101/gr.6743907.

Castanera, R., Borgognone, A., Pisabarro, A.G., and Ramirez, L. (2017). Biology, dynamics, and applications of transposable elements in basidiomycete fungi. *Appl. Microbiol. Biotechnol.* 101(4), 1337-1350. doi: 10.1007/s00253-017-8097-8.

Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49(16), 9077-9096. doi: 10.1093/nar/gkab688.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18(2), 170-175. doi: 10.1038/s41592-020-01056-5.

Chevreux, B., Wetter, T., and Suhai, S. (Year). "Genome sequence assembly using trace signals and additional sequence information", in: *German Conference on Bioinformatics (GCB)*), 45-56.

Christoffels, A., and van Heusden, P. (2019). "Genome Annotation: Perspective From Bacterial Genomes," in *Encyclopedia of Bioinformatics and Computational Biology,* eds. S. Ranganathan, M. Gribskov, K. Nakai & C. Schönbach. (Oxford: Academic Press), 152-156.

Coetzee, M.P.A., Santana, Q.C., Steenkamp, E.T., Wingfield, B.D., and Wingfield, M.J. (2020). Fungal genomes enhance our understanding of the pathogens affecting trees cultivated in Southern Hemisphere plantations. *South. For.: a J. For. Sci.* 82(3), 215-232. doi: 10.2989/20702620.2020.1819153.

Coetzee, M.P.A., Wingfield, B.D., Roux, J., Crous, P.W., Denman, S., and Wingfield, M.J. (2003). Discovery of two northern hemisphere *Armillaria* species on Proteaceae in South Africa. *Plant Pathol.* 52(5), 604-612. doi: 10.1046/j.1365-3059.2003.00879.x.

Coetzee, M.P.A., Wingfield, B.D., and Wingfield, M.J. (2018). *Armillaria* root-rot pathogens: Species boundaries and global distribution. *Pathogens* 7(4), 83. doi: 10.3390/pathogens7040083.

Collins, C., Keane, T.M., Turner, D.J., O'Keeffe, G., Fitzpatrick, D.A., and Doyle, S. (2013). Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. *J. Proteome Res.* 12(6), 2552-2570. doi: 10.1021/pr301131t.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18), 3674-3676. doi: 10.1093/bioinformatics/bti610.

Cornet, L., and Baurain, D. (2022). Contamination detection in genomic data: more is not enough. *Genome Biol.* 23(1)**,** 60. doi: 10.1186/s13059-022-02619-9.

Cubero, O.F., Crespo, A., Fatehi, J., and Bridge, P.D. (1999). DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. *Plant Syst. Evol.* 216(3-4)**,** 243-249. doi: 10.1007/bf01084401.

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15)**,** 2666-2669. doi: 10.1093/bioinformatics/bty149.

de Sá, P.H.C.G., Guimarães, L.C., das Graças, D.A., de Oliveira Veras, A.A., Barh, D., Azevedo, V., et al. (2018). "Chapter 11 - Next-Generation Sequencing and Data Analysis: Strategies, Tools, Pipelines and Protocols," in *Omics Technologies and Bio-Engineering,* eds. D. Barh & V. Azevedo. (Oxford: Academic Press), 191-207.

de Sena Brandine, G., and Smith, A.D. (2019). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res.* 8**,** 1874. doi: 10.12688/f1000research.21142.2.

Del Angel, V.D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O.V., et al. (2018). Ten steps to get started in Genome Assembly and Annotation [version 1; peer review: 2 approved]. *F1000Res.* 7(ELIXIR)**,** 148.

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6(5)**,** 361-375. doi: 10.1038/nrg1603.

Donlin, M.J. (2009). Using the Generic Genome Browser (GBrowse). *Curr. Protoc. Bioinformatics* 9(9)**,** Unit 9. doi: 10.1002/0471250953.bi0909s28.

Dunn, N.A., Unni, D.R., Diesh, C., Munoz-Torres, M., Harris, N.L., Yao, E., et al. (2019). Apollo: Democratizing genome annotation. *PLoS Comput. Biol.* 15(2)**,** e1006790. doi: 10.1371/journal.pcbi.1006790.

Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* 21 Suppl 1(suppl_1)**,** i152-158. doi: 10.1093/bioinformatics/bti1003.

Elías-Román, R.D., Medel-Ortiz, R., Alvarado-Rosales, D., Hanna, J.W., Ross-Davis, A.L., Kim, M.S., et al. (2018). *Armillaria mexicana*, a newly described species from Mexico. *Mycologia* 110(2)**,** 347-360. doi: 10.1080/00275514.2017.1419031.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* 9(1)**,** 18. doi: 10.1186/1471-2105-9-18.

Engels, B., Heinig, U., Grothe, T., Stadler, M., and Jennewein, S. (2011). Cloning and characterization of an *Armillaria gallica* cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides. *J. Biol. Chem.* 286(9)**,** 6871-6878. doi: 10.1074/jbc.M110.165845.

Engels, B., Heinig, U., McElroy, C., Meusinger, R., Grothe, T., Stadler, M., et al. (2021). Isolation of a gene cluster from *Armillaria gallica* for the synthesis of armillyl orsellinate-type sesquiterpenoids. *Appl. Microbiol. Biotechnol.* 105(1)**,** 211-224. doi: 10.1007/s00253-020-11006-y.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7(11), e47768. doi: 10.1371/journal.pone.0047768.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30(7), 1575-1584. doi: 10.1093/nar/30.7.1575.

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319.

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32(19), 3047-3048. doi: 10.1093/bioinformatics/btw354.

Ferguson, B.A., Dreisbach, T.A., Parks, C.G., Filip, G.M., and Schmitt, C.L. (2003). Coarse-scale population structure of pathogenic *Armillaria* species in a mixed-conifer forest in the Blue Mountains of northeast Oregon. *Can. J. For. Res.* 33(4), 612-623. doi: 10.1139/x03-065.

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS One* 6(1), e16526. doi: 10.1371/journal.pone.0016526.

Forouzan, E., Maleki, M.S.M., Karkhane, A.A., and Yakhchali, B. (2017). Evaluation of nine popular de novo assemblers in microbial genome assembly. *J. Microbiol. Methods* 143, 32-37. doi: 10.1016/j.mimet.2017.09.008.

Fukasawa, Y., Ermini, L., Wang, H., Carty, K., and Cheung, M.S. (2020). LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. *G3 (Bethesda)* 10(4), 1193-1196. doi: 10.1534/g3.119.400864.

Gawel, N., and Jarret, R. (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Mol. Biol. Rep.* 9(3), 262-266.

Giani, A.M., Gallo, G.R., Gianfranceschi, L., and Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* 18, 9-19. doi: 10.1016/j.csbj.2019.11.002.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U S A* 108(4), 1513-1518. doi: 10.1073/pnas.1017351108.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29(7), 644-652. doi: 10.1038/nbt.1883.

Grandaubert, J., Dutheil, J.Y., and Stukenbrock, E.H. (2019). The genomic determinants of adaptive evolution in a fungal pathogen. *Evol. Lett.* 3(3), 299-312. doi: 10.1002/evl3.117.

Gregory, S.C., and Rishbeth, J. (1991). "Pathogenicity and Virulence," in *Agriculture Handbook (USA),* eds. C.G. Shaw III & G.A. Kile. (Washington, District of Columbia: USDA Forest Service), 76–87.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42(Database issue)**,** D699-704. doi: 10.1093/nar/gkt1183.

Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36(9)**,** 2896-2898. doi: 10.1093/bioinformatics/btaa025.

Guillaumin, J.-J., Mohammed, C., Anselmi, N., Courtecuisse, R., Gregory, S.C., Holdenrieder, O., et al. (1993). Geographical distribution and ecology of the *Armillaria* species in Western Europe. *Eur. J. For. Pathol.* 23(6-7)**,** 321-341. doi: 10.1111/j.1439-0329.1993.tb00814.x.

Gupta, A.K., and Kumar, M. (2022). Benchmarking and assessment of eight *de novo* genome assemblers on viral Next-Generation Sequencing data, including the SARS-CoV-2. *OMICS* 26(7)**,** 372-381. doi: 10.1089/omi.2022.0042.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8)**,** 1072-1075. doi: 10.1093/bioinformatics/btt086.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9(1)**,** R7. doi: 10.1186/gb-2008-9-1-r7.

Heinzelmann, R., Croll, D., Zoller, S., Sipos, G., Münsterkötter, M., Güldener, U., et al. (2017). High-density genetic mapping identifies the genetic basis of a natural colony morphology mutant in the root rot pathogen *Armillaria ostoyae*. *Fungal Genet. Biol.* 108**,** 44-54. doi: 10.1016/j.fgb.2017.08.007.

Heinzelmann, R., Dutech, C., Tsykun, T., Labbé, F., Soularue, J.-P., and Prospero, S. (2019). Latest advances and future perspectives in *Armillaria* research. *Can. J. Plant Pathol.* 41(1)**,** 1-23. doi: 10.1080/07060661.2018.1558284.

Heinzelmann, R., Rigling, D., Sipos, G., Münsterkötter, M., and Croll, D. (2020). Chromosomal assembly and analyses of genome-wide recombination rates in the forest pathogenic fungus *Armillaria ostoyae*. *Heredity* 124(6)**,** 699-713. doi: 10.1038/s41437-020-0306-z.

Henderson, G., Cox, F., Kittelmann, S., Miri, V.H., Zethof, M., Noel, S.J., et al. (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* 8(9)**,** e74787. doi: 10.1371/journal.pone.0074787.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTEC: an automatic transposable element classification tool. *PLoS One* 9(5)**,** e91929. doi: 10.1371/journal.pone.0091929.

Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods Mol. Biol.* 1962**,** 65-95. doi: 10.1007/978-1-4939-9173-0_5.

Hoff, K.J., and Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinformatics* 65(1)**,** e57. doi: 10.1002/cpbi.57.

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36(7)**,** 2253-2255. doi: 10.1093/bioinformatics/btz891.

Huang, X., Duan, N., Xu, H., Xie, T.N., Xue, Y.R., and Liu, C.H. (2018). CTAB-PEG DNA extraction from fungi with high contents of polysaccharides. *Mol. Biol. (Mosk)* 52(4), 718-726. doi: 10.1134/S0026898418040080.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34(8), 2115-2122. doi: 10.1093/molbev/msx148.

Jung, H., Ventura, T., Chung, J.S., Kim, W.J., Nam, B.H., Kong, H.J., et al. (2020). Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Comput. Biol.* 16(11), e1008325. doi: 10.1371/journal.pcbi.1008325.

Jung, H., Winefield, C., Bombarely, A., Prentis, P., and Waterhouse, P. (2019). Tools and strategies for long-read sequencing and *de novo* assembly of plant genomes. *Trends Plant Sci.* 24(8), 700-724. doi: 10.1016/j.tplants.2019.05.003.

Kahlke, T., Ralph, P.J., and Price, S. (2018). BASTA – Taxonomic classification of sequences and sequence bins using last common ancestor estimations. *Methods Ecol. Evol.* 10(1), 100-103. doi: 10.1111/2041-210x.13095.

Karstens, L., Siddiqui, N.Y., Zaza, T., Barstad, A., Amundsen, C.L., and Sysoeva, T.A. (2021). Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci. Rep.* 11(1), 6186. doi: 10.1038/s41598-021-85482-1.

Kedves, O., Shahab, D., Champramary, S., Chen, L., Indic, B., Boka, B., et al. (2021). Epidemiology, biotic interactions and biological control of Armillarioids in the Northern Hemisphere. *Pathogens* 10(1), 76. doi: 10.3390/pathogens10010076.

Khan, A.R., Pervez, M.T., Babar, M.E., Naveed, N., and Shoaib, M. (2018). A comprehensive study of *de novo* genome assemblers: Current challenges and future prospective. *Evol. Bioinform. Online* 14, 1176934318758650. doi: 10.1177/1176934318758650.

Kidwell, M.G., and Lisch, D.R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55(1), 1-24. doi: 10.1111/j.0014-3820.2001.tb01268.x.

Kikuchi, G., and Yamaji, H. (2010). Identification of *Armillaria* species associated with *Polyporus umbellatus* using ITS sequences of nuclear ribosomal DNA. *Mycoscience* 51(5), 366-372.

Klopfenstein, N.B., Hanna, J.W., Cannon, P.G., Medel-Ortiz, R., Alvarado-Rosales, D., Lorea-Hernández, F., et al. (2014). First report of the *Armillaria* Root-Disease pathogen, *Armillaria gallica*, associated with several woody hosts in three states of Mexico. *Plant Dis.* 98(9), 1280-1280. doi: 10.1094/pdis-04-14-0349-pdn.

Koch, R.A., and Herr, J.R. (2021). Transcriptomics reveals the putative mycoparasitic strategy of the mushroom *Entoloma abortivum* on species of the mushroom genus *Armillaria*. *mSystems* 6(5), e0054421. doi: 10.1128/mSystems.00544-21.

Koch, R.A., Wilson, A.W., Séné, O., Henkel, T.W., and Aime, M.C. (2017). Resolved phylogeny and biogeography of the root pathogen *Armillaria* and its gasteroid relative, *Guyanagaster*. *BMC Evol. Biol.* 17(1), 33. doi: 10.1186/s12862-017-0877-3.

Kohany, O., Gentles, A.J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinf.* 7(1), 474. doi: 10.1186/1471-2105-7-474.

Kolesnikova, A.I., Putintseva, Y.A., Simonov, E.P., Biriukov, V.V., Oreshkova, N.V., Pavlov, I.N., et al. (2019). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related *Armillaria* species. *BMC Genom.* 20(1)**,** 351. doi: 10.1186/s12864-019-5732-z.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37(5)**,** 540-546. doi: 10.1038/s41587-019-0072-8.

König, S., Romp, E., Krauth, V., Rühl, M., Dörfer, M., Liening, S., et al. (2019). Melleolides from honey mushroom inhibit 5-lipoxygenase via Cys159. *Cell Chem. Biol.* 26(1)**,** 60-70.e64. doi: 10.1016/j.chembiol.2018.10.010.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28(24)**,** 3211-3217. doi: 10.1093/bioinformatics/bts611.

Koren, S., Rhie, A., Walenz, B.P., Dilthey, A.T., Bickhart, D.M., Kingan, S.B., et al. (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* 36(12)**,** 1174-1182. doi: 10.1038/nbt.4277.

Labbé, F., Lung-Escarmant, B., Fievet, V., Soularue, J.-P., Laurent, C., Robin, C., et al. (2017). Variation in traits associated with parasitism and saprotrophism in a fungal root-rot pathogen invading intensive pine plantations. *Fungal Ecol.* 26**,** 99-108. doi: 10.1016/j.funeco.2017.01.001.

Labbé, F., Marcais, B., Dupouey, J.-L., Bélouard, T., Capdevielle, X., Piou, D., et al. (2015). Pre-existing forests as sources of pathogens? The emergence of *Armillaria ostoyae* in a recently planted pine forest. *For. Ecol. Manage.* 357**,** 248-258. doi: 10.1016/j.foreco.2015.08.028.

Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Res.* 6(1287)**,** 1287.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35(9)**,** 3100-3108. doi: 10.1093/nar/gkm160.

Lam, K.K., LaButti, K., Khalak, A., and Tse, D. (2015). FinisherSC: a repeat-aware tool for upgrading *de novo* assembly using long reads. *Bioinformatics* 31(19)**,** 3207-3209. doi: 10.1093/bioinformatics/btv280.

Leger, A., and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J. Open Source Softw.* 4(34). doi: 10.21105/joss.01236.

Legrand, P., and Guillaumin, J.-J. (1993). *Armillaria* species in the forest ecosystems of the Auvergne (Central France). *Acta Oecol.* 14**,** 389-403.

Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., et al. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102**,** 3-11. doi: 10.1016/j.ymeth.2016.02.020.

Li, H.-T., Zhou, H., Duan, R.-T., Li, H.-Y., Tang, L.-H., Yang, X.-Q., et al. (2019). Inducing secondary metabolite production by co-culture of the endophytic fungus *Phoma sp.* and the symbiotic fungus *Armillaria* sp. *J. Nat. Prod.* 82(4)**,** 1009-1013. doi: 10.1021/acs.jnatprod.8b00685.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18)**,** 3094-3100. doi: 10.1093/bioinformatics/bty191.

Li, X., Xu, S., Zhang, J., and Li, M. (2021). Assembly and annotation of whole-genome sequence of *Fusarium equiseti*. *Genomics* 113(4)**,** 2870-2876. doi: 10.1016/j.ygeno.2021.06.019.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950)**,** 289-293. doi: 10.1126/science.1181369.

Liu, M.M., Xing, Y.M., Zhang, D.W., and Guo, S.X. (2015). Transcriptome analysis of genes involved in defence response in *Polyporus umbellatus* with *Armillaria mellea* infection. *Sci. Rep.* 5**,** 16075. doi: 10.1038/srep16075.

Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* 42(15)**,** e119. doi: 10.1093/nar/gku557.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33(20)**,** 6494-6506. doi: 10.1093/nar/gki937.

Lundell, T.K., Mäkelä, M.R., de Vries, R.P., and Hildén, K.S. (2014). "Chapter Eleven - Genomics, Lifestyles and Future Prospects of Wood-Decay and Litter-Decomposing Basidiomycota," in *Advances in Botanical Research,* ed. F.M. Martin. (Oxford: Academic Press), 329-370.

Lung, M.-Y., and Chang, Y.-C. (2011). Antioxidant properties of the edible Basidiomycete *Armillaria mellea* in submerged cultures. *Int. J. Mol. Sci.* 12(10)**,** 6367-6384.

Luo, X., Kang, X., and Schonhuth, A. (2021). phasebook: haplotype-aware de novo assembly of diploid genomes from long reads. *Genome Biol.* 22(1)**,** 299. doi: 10.1186/s13059-021-02512-x.

Lupo, V., Van Vlierberghe, M., Vanderschuren, H., Kerff, F., Baurain, D., and Cornet, L. (2021). Contamination in reference sequence databases: Time for divide-and-rule tactics. *Front. Microbiol.* 12**,** 755101. doi: 10.3389/fmicb.2021.755101.

Lushaj, B.M., Woodward, S., Keča, N., and Intini, M. (2010). Distribution, ecology and host range of *Armillaria* species in Albania. *Forest Pathol.* 40(6)**,** 485-499. doi: 10.1111/j.1439-0329.2009.00624.x.

Magoč, T., and Salzberg, S.L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21)**,** 2957-2963. doi: 10.1093/bioinformatics/btr507.

Mallet, L., Bitard-Feildel, T., Cerutti, F., and Chiapello, H. (2017). PhylOligo: a package to identify contaminant or untargeted organism sequences in genome assemblies. *Bioinformatics* 33(20)**,** 3283-3285. doi: 10.1093/bioinformatics/btx396.

Manchanda, N., Portwood, J.L., Woodhouse, M.R., Seetharam, A.S., Lawrence-Dill, C.J., Andorf, C.M., et al. (2020). GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC Genom.* 21(1)**,** 193. doi: 10.1186/s12864-020-6568-2.

Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for

scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38(10)**,** 4647-4654. doi: 10.1093/molbev/msab199.

Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* 14(1)**,** e1005944. doi: 10.1371/journal.pcbi.1005944.

Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6)**,** 764-770. doi: 10.1093/bioinformatics/btr011.

Marcais, G., Yorke, J.A., and Zimin, A. (2015). QuorUM: An error corrector for Illumina reads. *PLoS One* 10(6)**,** e0130821. doi: 10.1371/journal.pone.0130821.

Marthey, S., Aguileta, G., Rodolphe, F., Gendrault, A., Giraud, T., Fournier, E., et al. (2008). FUNYBASE: a FUNgal phYlogenomic dataBASE. *BMC Bioinf.* 9**,** 456. doi: 10.1186/1471-2105-9-456.

Martin, J., Bruno, V.M., Fang, Z., Meng, X., Blow, M., Zhang, T., et al. (2010). Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genom.* 11(1)**,** 663. doi: 10.1186/1471-2164-11-663.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17(1)**,** 3. doi: 10.14806/ej.17.1.200.

Mihail, J.D., and Bruhn, J.N. (2005). Foraging behaviour of *Armillaria* rhizomorph systems. *Mycol. Res.* 109(11)**,** 1195-1207. doi: 10.1017/S0953756205003606.

Miller, J.R., Delcher, A.L., Koren, S., Venter, E., Walenz, B.P., Brownley, A., et al. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24)**,** 2818-2824. doi: 10.1093/bioinformatics/btn548.

Minei, R., Hoshina, R., and Ogura, A. (2018). *De novo* assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genom.* 19(1)**,** 700. doi: 10.1186/s12864-018-5067-1.

Misiek, M., Braesel, J., and Hoffmeister, D. (2011). Characterisation of the ArmA adenylation domain implies a more diverse secondary metabolism in the genus *Armillaria*. *Fungal Biol.* 115(8)**,** 775-781. doi: 10.1016/j.funbio.2011.06.002.

Morrison, D.J. (2004). Rhizomorph growth habit, saprophytic ability and virulence of 15 *Armillaria* species. *For. Pathol.* 34(1)**,** 15-26. doi: 10.1046/j.1439-0329.2003.00345.x.

Morrison, D.J., and Pellow, K.W. (2002). Variation in virulence among isolates of *Armillaria ostoyae*. *Forest Pathol.* 32(2)**,** 99-107. doi: 10.1046/j.1439-0329.2002.00275.x.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., et al. (2000). A whole-genome assembly of Drosophila. *Science* 287(5461)**,** 2196-2204. doi: 10.1126/science.287.5461.2196.

Nachtweide, S., and Stanke, M. (2019). Multi-Genome Annotation with AUGUSTUS. *Methods Mol. Biol.* 1962**,** 139-160. doi: 10.1007/978-1-4939-9173-0_8.

Narh Mensah, D.L., Wingfield, B.D., and Coetzee, M.P.A. (2023). Nonribosomal peptide synthetase gene clusters and characteristics of predicted NRPS-dependent siderophore synthetases in

Armillaria and other species in the Physalacriaceae. *Curr. Genet.* 69(1), 7-24. doi: 10.1007/s00294-022-01256-w.

Nielsen, H. (2017). "Predicting Secretory Proteins with SignalP," in *Protein Function Prediction: Methods and Protocols,* ed. D. Kihara. (New York, NY: Springer New York), 59-73.

Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., et al. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* 30(9)**,** 1291-1305. doi: 10.1101/gr.263566.120.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20(1)**,** 275. doi: 10.1186/s13059-019-1905-y.

Palmer, J., and Stajich, J. (2017). Funannotate: eukaryotic genome annotation pipeline. *Zenodo*. doi: 10.5281/zenodo.1134477.

Pamilo, P., and Nei, M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5(5)**,** 568-583. doi: 10.1093/oxfordjournals.molbev.a040517.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25(7)**,** 1043-1055. doi: 10.1101/gr.186072.114.

Peabody, D.C., Peabody, R.B., Tyrrell, M.G., Towle, M.J., and Johnson, E.M. (2003). Phenotypic plasticity and evolutionary potential in somatic cells of *Armillaria gallica*. *Mycol. Res.* 107(4)**,** 408-412. doi: 10.1017/S0953756203007433.

Pearce, M.H., and Malajczuk, N. (1990). Factors affecting growth of *Armillaria luteobubalina* rhizomorphs in soil. *Mycol. Res.* 94(1)**,** 38-48. doi: 10.1016/S0953-7562(09)81262-8.

Pellegrini, A., Corneo, P.E., Camin, F., Ziller, L., Tosi, S., and Pertot, I. (2013). Isotope ratio mass spectrometry identifies soil microbial biocontrol agents having trophic relations with the plant pathogen *Armillaria mellea*. *Appl. Soil Ecol.* 64**,** 142-151. doi: 10.1016/j.apsoil.2012.12.005.

Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Reporter* 15(1)**,** 8-15. doi: 10.1007/bf02772108.

Prospero, S., Holdenrieder, O., and Rigling, D. (2004). Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. *Forest Pathol.* 34(1)**,** 1-14. doi: 10.1046/j.1437-4781.2003.00339.x.

Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., et al. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* 1(2)**,** 166-175. doi: 10.1371/journal.pcbi.0010022.

Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10(6)**,** 417-430. doi: 10.1038/nrmicro2790.

Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11(1)**,** 1432. doi: 10.1038/s41467-020-14998-3.

Ren, S., Gao, Y., Li, H., Ma, H., Han, X., Yang, Z., et al. (2022). Research status and application prospects of the medicinal mushroom *Armillaria mellea*. *Appl. Biochem. Biotechnol.* doi: 10.1007/s12010-022-04240-9.

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592(7856)**,** 737-746. doi: 10.1038/s41586-021-03451-0.

Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21(1)**,** 245. doi: 10.1186/s13059-020-02134-9.

Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* 19(1)**,** 460. doi: 10.1186/s12859-018-2485-7.

Ruiz-Duenas, F.J., Barrasa, J.M., Sanchez-Garcia, M., Camarero, S., Miyauchi, S., Serrano, A., et al. (2021). Genomic analysis enlightens Agaricales lifestyle evolution and increasing peroxidase diversity. *Mol. Biol. Evol.* 38(4)**,** 1428-1446. doi: 10.1093/molbev/msaa301.

Sahu, N., Merényi, Z., Bálint, B., Kiss, B., Sipos, G., Owens, R.A., et al. (2021). Hallmarks of basidiomycete soft- and white-rot in wood-decay -omics data of two *Armillaria* species. *Microorganisms* 9(1)**,** 149.

Salamov, A.A., and Solovyev, V.V. (2000). *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* 10(4)**,** 516-522. doi: 10.1101/gr.10.4.516.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30(24)**,** 3506-3514. doi: 10.1093/bioinformatics/btu538.

Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., et al. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res.* 22(3)**,** 557-567. doi: 10.1101/gr.131383.111.

Schatz, M.C., Delcher, A.L., and Salzberg, S.L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Res.* 20(9)**,** 1165-1173. doi: 10.1101/gr.101360.109.

Segerman, B. (2020). The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. *Front. Cell Infect. Microbiol.* 10**,** 527102. doi: 10.3389/fcimb.2020.527102.

Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962**,** 227-245. doi: 10.1007/978-1-4939-9173-0_14.

Shafin, K., Pesout, T., Chang, P.C., Nattestad, M., Kolesnikov, A., Goel, S., et al. (2021). Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat. Methods* 18(11)**,** 1322-1332. doi: 10.1038/s41592-021-01299-w.

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., et al. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* 27(7)**,** 958-967. doi: 10.1016/j.cub.2017.02.031.

Sipos, G., Anderson, J.B., and Nagy, L.G. (2018). *Armillaria* *Curr. Biol.* 28(7)**,** R297-R298. doi: 10.1016/j.cub.2018.01.026.

Sipos, G., Prasanna, A.N., Walter, M.C., O'Connor, E., Bálint, B., Krizsán, K., et al. (2017). Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat. Ecol. Evol.* 1(12)**,** 1931-1941. doi: 10.1038/s41559-017-0347-8.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6(1)**,** 31. doi: 10.1186/1471-2105-6-31.

Slatko, B.E., Gardner, A.F., and Ausubel, F.M. (2018). Overview of Next-Generation Sequencing technologies. *Curr. Protoc. Mol. Biol.* 122(1)**,** e59. doi: 10.1002/cpmb.59.

Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7(Suppl 1)**,** S10 11-12. doi: 10.1186/gb-2006-7-s1-s10.

Song, L., Florea, L., and Langmead, B. (2014). Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol.* 15(11)**,** 509. doi: 10.1186/s13059-014-0509-9.

Souvorov, A., Agarwala, R., and Lipman, D.J. (2018). SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol.* 19(1)**,** 153. doi: 10.1186/s13059-018-1540-z.

Steinegger, M., and Salzberg, S.L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* 21(1)**,** 115. doi: 10.1186/s13059-020-02023-1.

Surachat, K., Taylor, T.D., Wattanamatiphot, W., Sukpisit, S., and Jeenkeawpiam, K. (2022). aTAP: automated transcriptome analysis platform for processing RNA-seq data by *de novo* assembly. *Heliyon* 8(8)**,** e10255. doi: 10.1016/j.heliyon.2022.e10255.

Tan, M.K., Collins, D., Chen, Z., Englezou, A., and Wilkins, M.R. (2014). A brief overview of the size and composition of the myrtle rust genome and its taxonomic status. *Mycology* 5(2)**,** 52-63. doi: 10.1080/21501203.2014.919967.

Tarafder, S., Islam, M., Shatabda, S., and Rahman, A. (2022). Figbird: a probabilistic method for filling gaps in genome assemblies. *Bioinformatics* 38(15)**,** 3717-3724. doi: 10.1093/bioinformatics/btac404.

Tennessen, K., Andersen, E., Clingenpeel, S., Rinke, C., Lundberg, D.S., Han, J., et al. (2016). ProDeGe: a computational protocol for fully automated decontamination of genomes. *ISME J.* 10(1)**,** 269-272. doi: 10.1038/ismej.2015.100.

The Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* doi: 10.1093/nar/gkac247.

Tobias, P.A., Schwessinger, B., Deng, C.H., Wu, C., Dong, C., Sperschneider, J., et al. (2021). *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. *G3 (Bethesda)* 11(3). doi: 10.1093/g3journal/jkaa015.

Tomalak, M., and Woodward, S. (2017). Parasitic association of the mycetophagous wood nematode, *Bursaphelenchus fraudulentus* with the honey fungus *Armillaria ostoyae*. *For. Path.* 47(3)**,** e12325. doi: 10.1111/efp.12325.

Tyler, A.D., Mataseje, L., Urfano, C.J., Schmidt, L., Antonation, K.S., Mulvey, M.R., et al. (2018). Evaluation of Oxford Nanopore's MinION Sequencing Device for microbial whole genome sequencing applications. *Sci. Rep.* 8(1)**,** 10931. doi: 10.1038/s41598-018-29334-5.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., et al. (2020). PHI-base: the pathogen-host interactions database. *Nucleic Acids Res.* 48(D1)**,** D613-D620. doi: 10.1093/nar/gkz904.

van Burik, J.A., Schreckhise, R.W., White, T.C., Bowden, R.A., and Myerson, D. (1998). Comparison of six extraction techniques for isolation of DNA from filamentous fungi. *Med. Mycol.* 36(5)**,** 299-303. doi: 10.1080/02681219880000471.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14)**,** 2202-2204. doi: 10.1093/bioinformatics/btx153.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11)**,** e112963. doi: 10.1371/journal.pone.0112963.

Walter, M.C., Rattei, T., Arnold, R., Guldener, U., Munsterkotter, M., Nenova, K., et al. (2009). PEDANT covers all complete RefSeq genomes. *Nucleic Acids Res.* 37(Database issue)**,** D408-411. doi: 10.1093/nar/gkn749.

Warwell, M.V., McDonald, G.I., Hanna, J.W., Kim, M.-S., Lalande, B.M., Stewart, J.E., et al. (2019). *Armillaria altimontana* is associated with healthy western white pine (*Pinus monticola*): Potential *in situ* biological control of the *Armillaria* root disease pathogen, *A. solidipes*. *Forests* 10(4)**,** 294.

Wee, Y., Bhyan, S.B., Liu, Y., Lu, J., Li, X., and Zhao, M. (2019). The bioinformatics tools for the genome assembly and analysis based on third-generation sequencing. *Brief. Funct. Genomics* 18(1)**,** 1-12. doi: 10.1093/bfgp/ely037.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. *Genome Res.* 27(5)**,** 757-767. doi: 10.1101/gr.214874.116.

Whitehouse, C.A., and Hottel, H.E. (2007). Comparison of five commercial DNA extraction kits for the recovery of Francisella tularensis DNA from spiked soil samples. *Mol. Cell Probes* 21(2)**,** 92-96. doi: 10.1016/j.mcp.2006.08.003.

Wingfield, B.D., Ambler, J.M., Coetzee, M., De Beer, Z.W., Duong, T.A., Joubert, F., et al. (2016). Draft genome sequences of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa, Endoconidiophora laricicola, E. polonica* and *Penicillium freii* DAOMC 242723. *IMA Fungus* 7(1)**,** 217-227. doi: 10.5598/imafungus.2016.07.01.11.

Wingfield, B.D., Berger, D.K., Coetzee, M.P.A., Duong, T.A., Martin, A., Pham, N.Q., et al. (2022). IMA genome-F17 : Draft genome sequences of an *Armillaria* species from Zimbabwe, *Ceratocystis colombiana, Elsinoe necatrix, Rosellinia necatrix,* two genomes of *Sclerotinia minor*, short-read genome assemblies and annotations of four *Pyrenophora teres* isolates from barley grass, and a long-read genome assembly of *Cercospora zeina*. *IMA Fungus* 13(1)**,** 19. doi: 10.1186/s43008-022-00104-3.

Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45(8), e57-e57. doi: 10.1093/nar/gkw1306.

Xiao, C.L., Chen, Y., Xie, S.Q., Chen, K.N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14(11), 1072-1074. doi: 10.1038/nmeth.4432.

Xing, Y.M., Li, B., Liu, L., Li, Y., Yin, S.X., Yin, S.C., et al. (2021). *Armillaria mellea* symbiosis drives metabolomic and transcriptomic changes in *Polyporus umbellatus* sclerotia. *Front. Microbiol.* 12, 792530. doi: 10.3389/fmicb.2021.792530.

Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13(5), 329-342. doi: 10.1038/nrg3174.

Yang, S.F., Lu, C.W., Yao, C.T., and Hung, C.M. (2019). To trim or not to trim: Effects of read trimming on the *de novo* genome assembly of a widespread East Asian passerine, the rufous-capped babbler (*Cyanoderma ruficeps* blyth). *Genes (Basel)* 10(10). doi: 10.3390/genes10100737.

Yao, G., Ye, L., Gao, H., Minx, P., Warren, W.C., and Weinstock, G.M. (2012). Graph accordance of next-generation sequence assemblies. *Bioinformatics* 28(1), 13-16. doi: 10.1093/bioinformatics/btr588.

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 46(W1), W71-W75. doi: 10.1093/nar/gky400.

Young, A.D., and Gillung, J.P. (2020). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Syst. Entomol.* 45(2), 225-247. doi: 10.1111/syen.12406.

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S., et al. (2011). Faster and more accurate sequence alignment with SNAP. *arXiv preprint arXiv:1111.5572*.

Zavastin, D.E., Mircea, C., Aprotosoaie, A.C., Gherman, S., Hancianu, M., and Miron, A. (2015). *Armillaria mellea*: phenolic content, in vitro antioxidant and antihyperglycemic effects. *Rev. Med. Chir. Soc. Med. Nat. Iasi.* 119(1), 273-280.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9), 847-848. doi: 10.1093/bioinformatics/17.9.847.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18(5), 821-829. doi: 10.1101/gr.074492.107.

Zhan, J., Mundt, C.C., and McDonald, B.A. (2007). Sexual reproduction facilitates the adaptation of parasites to antagonistic host environments: Evidence from empirical study in the wheat-*Mycosphaerella graminicola* system. *Int. J. Parasitol.* 37(8-9), 861-870. doi: 10.1016/j.ijpara.2007.03.003.

Zhan, M., Tian, M., Wang, W., Li, G., Lu, X., Cai, G., et al. (2020). Draft genomic sequence of *Armillaria gallica* 012m: insights into its symbiotic relationship with *Gastrodia elata*. *Braz. J. Microbiol.* 51(4), 1539-1552. doi: 10.1007/s42770-020-00317-x.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46(W1)**,** W95-W101. doi: 10.1093/nar/gky418.

Zhang, X., Wu, R., Wang, Y., Yu, J., and Tang, H. (2020). Unzipping haplotypes in diploid and polyploid genomes. *Comput. Struct. Biotechnol. J.* 18**,** 66-72. doi: 10.1016/j.csbj.2019.11.011.

Zimin, A.V., Marcais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29(21)**,** 2669-2677. doi: 10.1093/bioinformatics/btt476.

Zimin, A.V., and Salzberg, S.L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput. Biol.* 16(6)**,** e1007981. doi: 10.1371/journal.pcbi.1007981.

Tables

Table 1: Methods and tools used in sequencing and assembly of published genomes of *Armillaria* species to date (March 2023)

Species [strain or isolate]	Country of origin (lifestyle) [a]	DNA extraction methods/kits	Sequencing technology(ies) [Pre-assembly reads correction]	Sequencing technology(ies)	Genome assembly tools	Post-assembly processing	Genome completeness evaluation	NCBI accession number or genome link (Reference)
Armillaria altimontana (NABS X) [837–10]	USA (Unknown)	MoBio (Qiagen) DNeasy PowerMax Soil Kit	PacBio SMRT cells [N/A]	PacBio SMRT cells	Pacific Biosciences SMRT Portal software	N/A	QUAST, BUSCO 2.0b2 with Basidiomycota dataset	JAIWYR000000000.1 (Caballero et al. 2022)
Armillaria borealis [AB13-TR4-IP16]	Eastern Siberia (FN, weakly pathogenic, opportunistic)	Modified version of the hot-CTAB extraction at 65 °C with chloroform washing	Illumina MiSeq [N/A]	Illumina MiSeq	SPAdes v3.13.0 genome assembler	N/A	BUSCO, and RS BridgeMapper protocol	JAAGUC000000000.1 (Akulova et al. 2020)
Armillaria cepistipes [B5]	Italy (FN, weakly pathogenic)	PowerMax MOBIO DNA isolation kit	PacBio RS II platform; Illumina HiSeq 2000 [N/A]	PacBio RS II platform; Illumina HiSeq 2000	HGAP3 workflow of the SMRT Analysis suite v2.3,	PBJelly2, FinisherSC, RS Resequencing protocol, and Pilon	CEGMA and BUSCO Version 3.0.1 with the lineage specific profile library basidiomycota_odb9, and RS Resequencing protocol	FTRY00000000.1 (Sipos et al. 2017)
Armillaria fuscipes [CMW 2740]	South Africa (FN, particularly pathogenic to exotic species)	DNeasy Plant Mini Kit (Qiagen, Aarhus)	Illumina HiSeq; PacBio sequencing with 20 SMRT cells [N/A]	Illumina HiSeq; PacBio sequencing with 20 SMRT cells	Celera Assembler pipeline, CLC Genomics Workbench v. 5.5.1, and MIRA	Velvet optimizer, Graph Accordance Assembly program (GAA) and SSPACE	CEGMA and BUSCO using gene models from *Laccaria bicolor*	LWUH00000000.1 (Wingfield et al. 2016)
Armillaria gallica [012m]	China (FN, weakly or secondarily pathogenic)	DNeasy plant kit (Qiagen)	PacBio Sequel; Illumina HiSeq [LoRDEC V0.6, Trimmomatic V0.36]	PacBio Sequel; Illumina HiSeq	ALLPATHS-LG V52488, and MECAT2	Purge Haplotigs, FinisherSC V2.1, and NextPolish V1.0.21	BUSCO with the lineage-specific profile library basidiomycota_odb9	VTST00000000.1 (Zhan et al. 2020)

Armillaria gallica [Ar21-2]	USA (FN, weakly or secondarily pathogenic)	PowerMax MOBIO DNA isolation kit	Illumina MiSeq [N/A]	Illumina MiSeq	HGAP3 workflow of the SMRT Analysis suite v2.3	PBJelly2, FinisherSC, RS Resequencing protocol, and Pilon	CEGMA and BUSCO Version 3.0.1 with the lineage specific profile library basidiomycota_odb9 and RS BridgeMapper protocol	NKEW00000000.1 (Sipos et al. 2017)
Armillaria mellea [DSM 3731]	France (FN, highly pathogenic)	ZR Fungal/Bacterial DNA Kit using a modified protocol	Illumina GAII [N/A]	Illumina GAII	Velvet v0.7.53	N/A	N/A	Armme1_1_jgi link (Collins et al. 2013)
Armillaria ostoyae [C18/9]	Switzerland (pathogenic)	PowerMax MOBIO DNA isolation kit	PacBio RS II platform; Illumina HiSeq 2000 [N/A]	PacBio RS II platform; Illumina HiSeq 2000	HGAP3 workflow of the SMRT Analysis suite v2.3	PBJelly2, FinisherSC, RS Resequencing protocol, and Pilon	CEGMA and BUSCO Version 3.0.1 with the lineage specific profile library basidiomycota_odb9 and RS BridgeMapper protocol	FUEG00000000.1 (Sipos et al. 2017)
Armillaria solidipes [C28-4]	USA (FN, highly pathogenic)	PowerMax MOBIO DNA isolation kit	Illumina HiSeq 2000 [N/A]	Illumina HiSeq 2000	HGAP3 workflow of the SMRT Analysis suite v2.3	PBJelly2, FinisherSC, RS Resequencing protocol, and Pilon	CEGMA and BUSCO Version 3.0.1 with the lineage specific profile library basidiomycota_odb9, and RS BridgeMapper protocol	NKHM00000000.1 (Sipos et al. 2017)
Armillaria solidipes (form *A. ostoyae*) [ID001]	USA (FN, highly pathogenic)	MoBio DNeasy PowerMax Soil Kit	PacBio SMRT cells [N/A]	PacBio SMRT cells	Pacific Biosciences SMRT Portal software	N/A	QUAST and BUSCO 2.0b2 with Basidiomycota dataset	JAIWYQ000000000.1 (Caballero et al. 2022)
Armillaria African Clade B sp. [CMW4456]	Zimbabwe (Unknown)	Qiagen DNEasy Plant Pro Kit (50)	PacBio HiFi; Illumina HiSeq [Trimmomatic]	PacBio HiFi; Illumina HiSeq	CLC Genomics Workbench v 22.0.1	Pilon v. 1.23	QUAST v 5.0.2 and BUSCO v. 5.3.2, using the agaricales_odb10 lineage dataset in the Galaxy platform	JANDKJ000000000.1 (Wingfield et al., 2022)

[a]: Lifestyle information for *A. ostoyae* was obtained from (Labbé *et al.*, 2017; Morrison and Pellow, 2002; Prospero *et al.*, 2004). Lifestyle information for other species was summarized from (Koch *et al.*, 2017). FN = facultative necrotroph; N/A = not available

Table 2: Some resources for genome assembly and annotation

STEP	TOOLS/PLATFORMS/WORKFLOWS	WEBSITE OR SOURCE	REFERENCE(S)
GENOME ASSEMBLY AND EVALUATION			
Raw reads quality evaluation	Falco	https://github.com/smithlabcode/falco	(de Sena Brandine and Smith, 2019)
	FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	N/A
	FastQE	https://github.com/fastqe/fastqe	N/A
	LongQC	http://github.com/yfukasawa/LongQC	(Fukasawa *et al.*, 2020)
	NanoQC	http://github.com/wdecoster/nanopack	(De Coster *et al.*, 2018)
	pycoQC	http://github.com/a-slide/pycoQC	(Leger and Leonardi, 2019)
	MultiQC	http://multiqc.info/	(Ewels *et al.*, 2016)
Pre-assembly reads profiling and correction	Cutadapt	http://cutadapt.readthedocs.io/en/stable/guide.html	(Martin, 2011)
	Filtlong	https://github.com/rrwick/Filtlong	N/A
	FLASh	http://ccb.jhu.edu/software/FLASH/	(Magoč and Salzberg, 2011)
	GenomeScope2.0	https://github.com/tbenavi1/genomescope2.0	(Ranallo-Benavidez *et al.*, 2020; Vurture *et al.*, 2017)
	Jellyfish	https://github.com/alekseyzimin/Jellyfish	(Marcais and Kingsford, 2011)
	LoRDEC	http://www.atgc-montpellier.fr/lordec/	(Salmela and Rivals, 2014)
	Meryl	https://github.com/marbl/meryl	N/A
	MUMmer	https://github.com/mummer4/mummer	(Marcais *et al.*, 2018)
	Porechop	https://github.com/rrwick/Porechop	N/A
	Trimmomatic	https://github.com/timflutre/trimmomatic	(Bolger *et al.*, 2014)
	Trim Reads tool in CLC Genomics Workbench	CLC genomics website	QIAGEN, Aarhus
	QuORUM	https://github.com/alekseyzimin/Quorum	(Marcais *et al.*, 2015)
Genome assembly	CABOG	https://wgs-assembler.sourceforge.net/wiki/index.php?title=Main_Page	(Miller *et al.*, 2008)
	Canu,	https://github.com/marbl/canu	(Koren *et al.*, 2018; Nurk *et al.*, 2020)
	CLC Genomics Workbench	CLC genomics website	QIAGEN, Aarhus
	Flye	https://github.com/fenderglass/Flye	(Kolmogorov *et al.*, 2019)
	Hifiasm	https://github.com/chhylp123/hifiasm	(Cheng *et al.*, 2021)
	MaSuRCA	https://github.com/alekseyzimin/masurca	(Zimin *et al.*, 2013)
	MEGAHIT	http://github.com/voutcn/megahit	(Li *et al.*, 2016)
	MIRA	http://sourceforge.net/p/mira-assembler/wiki/Home/	(Chevreux *et al.*, 1999)
	Phasebook	https://github.com/phasebook/phasebook	(Luo *et al.*, 2021)
	Shovill	https://github.com/tseemann/shovill	N/A
	SKESA	http://github.com/ncbi/SKESA	(Souvorov *et al.*, 2018)
	SMARTdenovo	https://github.com/ruanjue/smartdenovo	N/A
	SPAdes	http://cab.spbu.ru/software/spades/	(Bankevich *et al.*, 2012)
	Supernova	https://bio.tools/supernova	(Weisenfeld *et al.*, 2017)
	Velvet	https://github.com/dzerbino/velvet	(Zerbino & Birney, 2008)

	VGP Pipeline	http://usegalaxy.eu/	(Rhie *et al.*, 2021; The Galaxy Community, 2022)
Post-assembly processing	Apollo	https://bio.tools/apollo-seq	(Dunn *et al.*, 2019)
	Figbird	https://github.com/SumitTarafder/Figbird	(Tarafder *et al.*, 2022)
	FinisherSC	https://github.com/kakitone/finishingTool	(Lam *et al.*, 2015)
	Graph Accordance Assembly (GAA) program	https://github.com/ghyao/GAA	(Yao *et al.*, 2012)
	HapDup	https://github.com/KolmogorovLab/hapdup	(Kolmogorov *et al.*, 2019; Shafin *et al.*, 2021)
	Lighter	https://github.com/mourisl/Lighter	(Song *et al.*, 2014)
	Minimap2	https://github.com/lh3/minimap2	(Li, 2018)
	NextPolish	http://github.com/Nextomics/NextPolish	(Hu *et al.*, 2020)
	PBJelly2	https://github.com/esrice/PBJelly	(English *et al.*, 2012)
	Pilon	https://github.com/broadinstitute/pilon	(Walker *et al.*, 2014)
	POLCA	https://github.com/alekseyzimin/masurca	(Zimin and Salzberg, 2020)
	Purge_Dups	http://github.com/dfguan/purge_dups	(Guan *et al.*, 2020)
	Purge Haplotigs	https://bitbucket.org/mroachawri/purge_haplotigs/src/master/	(Roach *et al.*, 2018)
	RS Resequencing protocol of the SMRT Analysis suite	http://smrt-analysis.readthedocs.io/en/latest/SMRT-Pipe-Reference-Guide-v2.3.0/	N/A
	SSPACE	https://github.com/nsoranzo/sspace_basic	(Boetzer *et al.*, 2011)
	VelvetOptimiser	http://github.com/Slugger70/VelvetOptimiser	(Zerbino and Birney, 2008)
Genome quality evaluation			
Genome completeness evaluation and miss-assemblies detection	BUSCO	http://busco.ezlab.org/	(Manni *et al.*, 2021; Seppey *et al.*, 2019)
	dnAQET	https://www.fda.gov/science-research/bioinformatics-tools/de-novo-assembly-quality-evaluation-tool-dnaqet	N/A
	GAGE	http://gage.cbcb.umd.edu/index.html	(Salzberg *et al.*, 2012)
	GenomeQC	https://github.com/HuffordLab/GenomeQC	(Manchanda *et al.*, 2020)
	Merqury	http://github.com/marbl/merqury	(Rhie *et al.*, 2020)
	QUAST	https://github.com/ablab/quast/releases/tag/quast_5.2.0	(Gurevich *et al.*, 2013)
Genome contamination detection and removal	Anvi'o	https://anvio.org/	(Eren *et al.*, 2015)
	BASTA	https://github.com/timkahlke/BASTA	(Kahlke *et al.*, 2018)
	BlobTools	https://github.com/DRL/blobtools	(Laetsch and Blaxter, 2017)
	CheckM	https://github.com/Ecogenomics/CheckM	(Parks *et al.*, 2015)
	Conterminator	https://github.com/martin-steinegger/conterminator	(Steinegger and Salzberg, 2020)
	Forty-Two	https://metacpan.org/pod/Bio::MUST::Apps::FortyTwo	(Simion *et al.*, 2017)
	Physeter	https://metacpan.org/pod/Bio::MUST::Apps::Physeter	(Lupo *et al.*, 2021)
	ProDeGE	https://bitbucket.org/berkeleylab/jgi-prodege/src/master/	(Tennessen *et al.*, 2016)
	PhylOligo	https://github.com/itsmeludo/PhylOligo	(Mallet *et al.*, 2017)
TRANSPOSABLE ELEMENTS AND GENOME ANNOTATION			

RNA gene identification and masking	aTAP	https://atap.psu.ac.th/	(Surachat *et al.*, 2022)
	FEELnc	https://github.com/tderrien/FEELnc	(Wucher *et al.*, 2017)
	RNAmmer	https://stab.st-andrews.ac.uk/wiki/index.php/Rnammer	(Lagesen *et al.*, 2007)
	SortMeRNA [a]	https://bioinfo.lifl.fr/RNA/sortmerna/	(Kopylova *et al.*, 2012)
	tRNAscan-SE	http://lowelab.ucsc.edu/tRNAscan-SE/	(Chan *et al.*, 2021)
Detection, annotation and analysis of transposable elements, other repetitive elements, and low complexity DNA regions in genomic sequences	CENSOR	https://www.girinst.org/censor/index.php	(Kohany *et al.*, 2006)
	EDTA	https://github.com/oushujun/EDTA	(Ou *et al.*, 2019)
	Grouper	https://github.com/COMBINE-lab/grouper	(Quesneville *et al.*, 2005)
	LTRharvest	https://www.zbh.uni-hamburg.de/en/forschung/gi/software/ltrharvest.html	(Ellinghaus *et al.*, 2008)
	PASTEC	https://mybiosoftware.com/tag/pastec	(Hoede *et al.*, 2014)
	PASTEClassifier	https://urgi.versailles.inra.fr/Tools/PASTEClassifier	(Hoede *et al.*, 2014)
	Piler	http://www.drive5.com/piler/	(Edgar and Myers, 2005)
	Recon	http://eddylab.org/software/recon/	(Bao and Eddy, 2002)
	RepBase-20181026	https://www.girinst.org/repbase/	(Bao *et al.*, 2015)
	RepeatMasker	https://github.com/rmhubley/RepeatMasker	N/A
	RepeatModeler	http://www.repeatmasker.org/RepeatModeler/	N/A
	REPET package	http://urgi.versailles.inra.fr/Tools/REPET	(Flutre *et al.*, 2011; Hoede *et al.*, 2014)
	TEclass	https://www.bioinformatics.uni-muenster.de/tools/teclass/index.hbi	(Abrusán *et al.*, 2009)
	TE_finder (includes BLASTER, MATCHER, GROUPER, and DUSTER)	https://github.com/urgi-anagen/TE_finder	N/A
RNA sequence assembling and annotation	Rnnotator	https://vcru.wisc.edu/simonlab/bioinformatics/programs/jgi/Rnnotator2.3Manual.pdf	(Martin *et al.*, 2010)
	Trinity	https://github.com/trinityrnaseq/trinityrnaseq/wiki	(Grabherr *et al.*, 2011)
***Ab initio* and/or evidence-based gene prediction and structural annotation**	AUGUSTUS	http://bioinf.uni-greifswald.de/augustus/	(Hoff and Stanke, 2019; Nachtweide and Stanke, 2019)
	BRAKER	https://github.com/Gaius-Augustus/BRAKER	(Hoff *et al.*, 2019)
	Deeploc	https://services.healthtech.dtu.dk/service.php?DeepLoc-2.0	(Almagro Armenteros *et al.*, 2017)
	DIAMOND	https://github.com/bbuchfink/diamond	(Buchfink *et al.*, 2015)
	eggNOG-mapper	http://eggnog-mapper.embl.de/	(Huerta-Cepas *et al.*, 2017)
	Exonerate	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate	(Slater and Birney, 2005)
	Fgenesh	https://bio.tools/fgenesh	(Salamov and Solovyev, 2000; Solovyev *et al.*, 2006)
	Funannotate [b]	https://github.com/nextgenusfs/funannotate	(Palmer and Stajich, 2017)
	FunyBASE	http://genome.jouy.inra.fr/funybase/	(Marthey *et al.*, 2008)

	GBrowse	https://github.com/GMOD/GBrowse	(Donlin, 2009)
	GeneMark-ES	http://topaz.gatech.edu/license_download.cgi	(Lomsadze *et al.*, 2005)
	GeneMark-ET	http://topaz.gatech.edu/license_download.cgi	(Lomsadze *et al.*, 2014)
	GETA	https://github.com/chenlianfu/geta	N/A
	Maker	https://www.yandell-lab.org/software/maker.html	(Campbell *et al.*, 2014; Cantarel *et al.*, 2008)
	PASA	https://bio.tools/PASA	(Haas *et al.*, 2008)
	SNAP	https://bio.tools/snap	(Zaharia *et al.*, 2011)
	TransDecoder	https://bio.tools/TransDecoder	N/A
Functional annotation	antiSMASH	https://docs.antismash.secondarymetabolites.org/	(Blin *et al.*, 2021)
	Blast2GO	https://www.biobam.com/download-blast2go/	(Conesa *et al.*, 2005)
	BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi	(Altschul *et al.*, 1990)
	dbCAN2 server	https://bcb.unl.edu/dbCAN2/	(Zhang *et al.*, 2018)
	eggNOG-mapper	http://eggnog-mapper.embl.de/	(Huerta-Cepas *et al.*, 2017)
	InterProScan	https://www.ebi.ac.uk/interpro/	(Blum *et al.*, 2021; Zdobnov and Apweiler, 2001)
	mcl algorithm	https://github.com/koteth/python_mcl	(Enright *et al.*, 2002)
	PEDANT system	https://www.hsls.pitt.edu/obrc/index.php?page=URL1100009979	(Walter *et al.*, 2009)
	PHI base	https://github.com/PHI-base	(Urban *et al.*, 2020)
	SecretomeP	https://services.healthtech.dtu.dk/service.php?SecretomeP-2.0	(Bendtsen *et al.*, 2004; Bendtsen *et al.*, 2005)
	SignalP	https://github.com/fteufel/signalp-6.0	(Nielsen, 2017)
	Trinotate	https://bio.tools/trinotate	(Grabherr *et al.*, 2011)
	WEGO	https://wego.genomics.cn/	(Ye *et al.*, 2018)
GENERAL USE PLATFORMS			
	CLC Genomics Workbench	QIAGEN CLC Genomics Workbench Portfolio	QIAGEN, Aarhus, Denmark
	Galaxy platform	https://usegalaxy.eu/	(Afgan *et al.*, 2018; The Galaxy Community, 2022)
	JGI Annotation Pipeline	https://img.jgi.doe.gov/docs/pipelineV5/	(Grigoriev *et al.*, 2014)

N/A = Not applicable

[a] = can also be used for RNA sequence annotation; [b] = can also be used for functional annotation

Table 3: QUAST and BUSCO evaluation of published *Armillaria* genomes

Assembled isolate	QUAST evaluation metrics [b]						Complete BUSCOs [c] (%, n = 3,870)
	# Contigs	Largest contig (bp)	Total length (bp)	GC (%)	N50 (bp)	# N's per 100 kbp	
A. altimontana (NABS X) [837–10]	100	5,843,527	73,739,702	47.77	1,930,169	14	98.8 (98.2, 0.6)
A. borealis [AB13-TR4-IP16]	10,402	2,136,877	56,994,024	47.76	75,348	5,501	97.9 (97.4, 0.5)
A. cepistipes [B5]	182	6,523,365	75,519,607	47.59	2,288,895	0	98.9 (97.8, 1.1)
A. fuscipes [CMW 2740]	17,117	157,180	50,706,099	47.68	6,152	982	92.6 (88.9, 3.7)
A. gallica [012m]	63	6,431,929	87,305,441	47.38	2,159,699	0	98.7 (95.9, 2.8)
A. gallica [Ar21-2]	319	4,779,317	85,336,812	47.35	1,035,263	8,165	98.6 (97.9, 0.7)
A. mellea [DSM 3731] [a]	9,867	639,705	75,786,598	46.99	26,609	16,520	97.9 (97.5, 0.4)
A. ostoyae [C18/9]	106	6,405,655	60,106,801	48.33	2,283,935	0	98.8 (98.2, 0.6)
A. solidipes [C28-4]	229	3,399,694	58,009,494	48.35	715,667	3,906	98.9 (98.3, 0.6)
*A. solidipes (*form *A. ostoyae)* [ID001]	72	4,463,803	55,735,298	48.26	2,424,439	5	98.7 (98.2, 0.5)
Armillaria African Clade B sp. [CMW4456]	840	1,463,441	54,954,005	46.53	128,967	0	98.2 (97.0, 1.2)

Unless otherwise stated, the table was created with analysis of genome statistics of Fasta sequences of published genomes at NCBI Genome (http://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=47424).

[a] = downloaded from JGI MycoCosm; [b] = QUAST analyses conducted in Galaxy using the genome assembly mode, and "Fungus use of GeneMark-ES for gene finding, Barrnap for ribosomal RNA genes prediction"; [c] = BUSCO analyses performed using BUSCO version 5.3.2 in genome mode using AUGUSTUS gene predictor with agaricales_odb10 lineage dataset. BUSCO data is presented as total complete BUSCOs (single copy BUSCOs, duplicated BUSCOs). There were less than 5% fragmented and missing BUSCOs each in all the genomes.

Figures

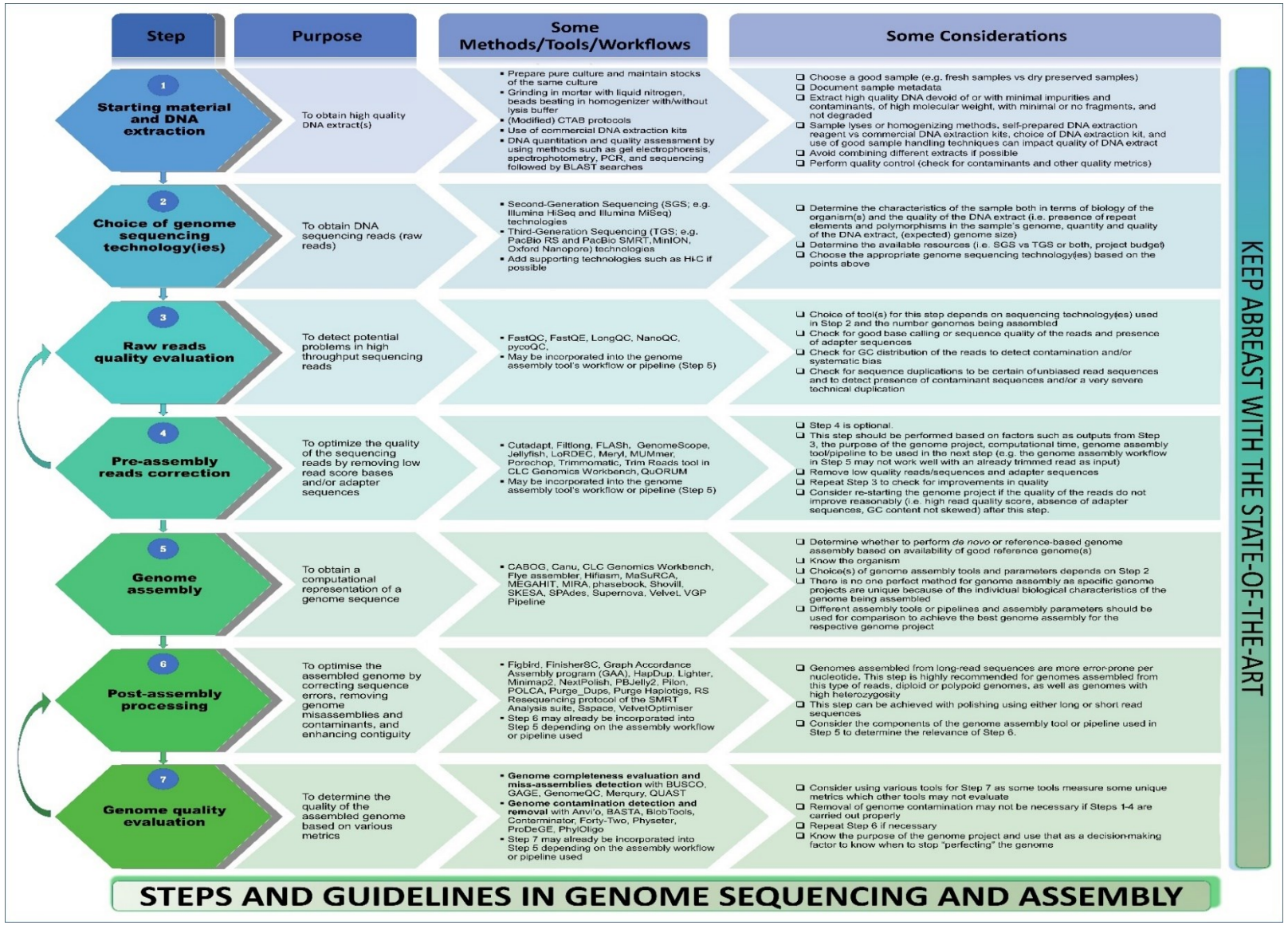


Figure 1: Infographic showing workflow, technological tools, and decision-making considerations for genome sequencing and assembly

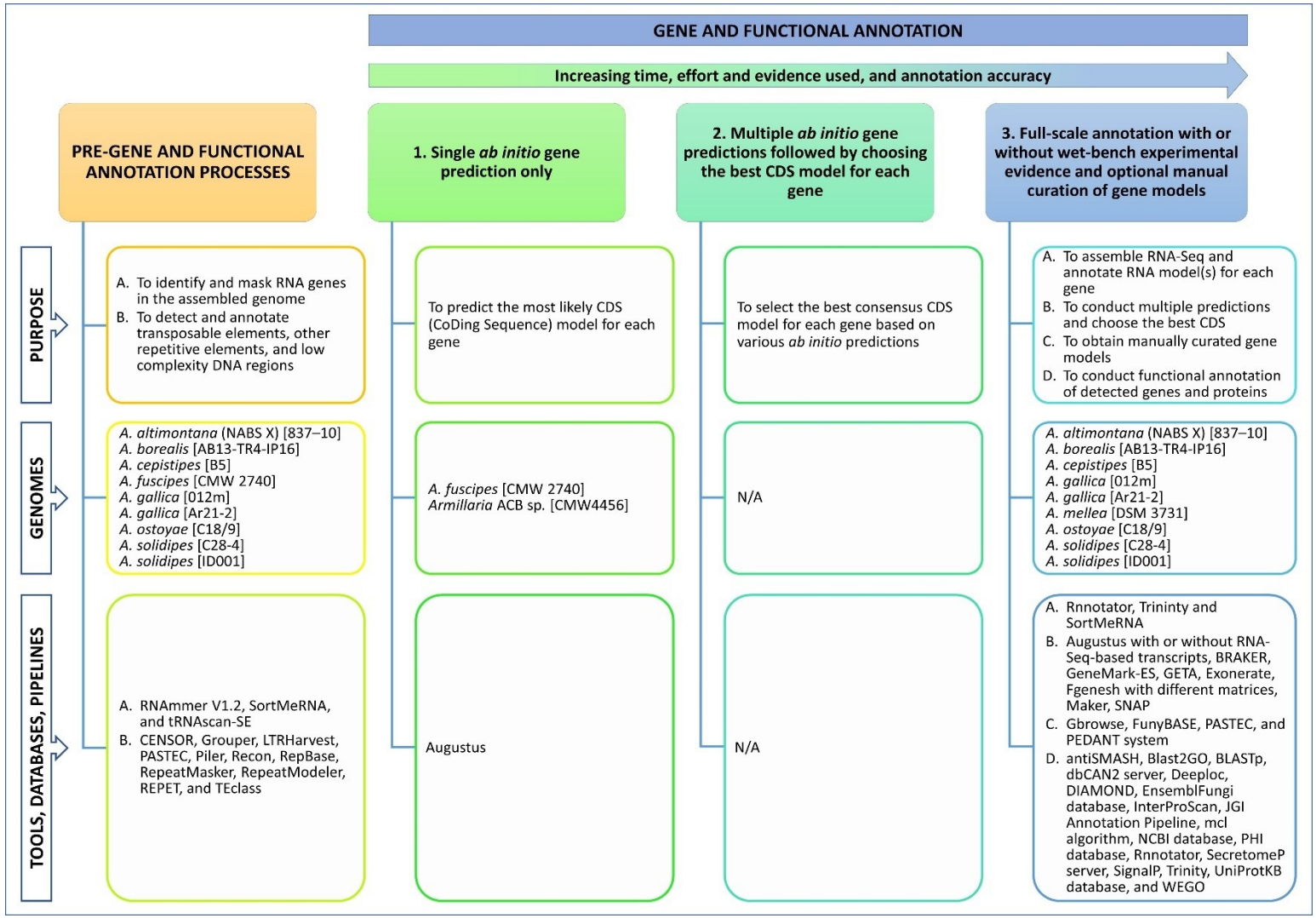


Figure 2: Steps and tools used in *Armillaria* genome annotation

Levels 1, 2, and 3 are based on the approaches to genome annotation relative to time, effort, and evidence used and the resultant increase in annotation accuracy as outlined by Yandell and Ence (2012). Level 2 can be performed with the *ab initio* prediction tools indicated in Level 3 (point B under "Tools, Databases, Pipelines"). ACB = African Clade B; N/A = not applicable

CHAPTER 2

NONRIBOSOMAL PEPTIDE SYNTHETASE GENE CLUSTERS AND CHARACTERISTICS OF PREDICTED NRPS-DEPENDENT SIDEROPHORE SYNTHETASES IN *ARMILLARIA* AND OTHER SPECIES IN THE PHYSALACRIACEAE

This chapter has been published in Current Genetics as:

Narh Mensah, D.L., Wingfield, B.D., Coetzee, M.P.A. 2023. Nonribosomal peptide synthetase gene clusters and characteristics of predicted NRPS-dependent siderophore synthetases in *Armillaria* and other species in the Physalacriaceae. *Current Genetics*. **69**(1), 7-24. https://doi.org/10.1007/s00294-022-01256-w

Abstract

Fungal secondary metabolites are often pathogenicity or virulence factors synthesized by genes contained in secondary metabolite gene clusters (SMGCs). Nonribosomal polypeptide synthetase (NRPS) clusters are SMGCs which produce peptides such as siderophores, the high affinity ferric iron chelating compounds required for iron uptake under aerobic conditions. *Armillaria* spp. are mostly facultative necrotrophs of woody plants. NRPS-dependent siderophore synthetase (NDSS) clusters of *Armillaria* spp. and selected Physalacriaceae were investigated using a comparative genomics approach. Siderophore biosynthesis by strains of selected *Armillaria* spp. was evaluated using Chrome Azurol S (CAS) and split-CAS assays. At least one NRPS cluster and other clusters were detected in the genomes studied. No correlation was observed between the number and types of SMGCs and reported pathogenicity of the species studied. The genomes contained one NDSS cluster each. All NDSSs were multi-modular with the domain architecture $(ATC)_3(TC)_2$. NDSS clusters of the *Armillaria* spp. showed a high degree of microsynteny. In the genomes of *Desarmillaria* spp. and *Guyanagaster necrorhizus,* NDSS clusters were more syntenic with NDSS clusters of *Armillaria* spp. than to those of the other Physalacriaceae species studied. Three A-domain orthologous groups were identified in the NDSSs, and atypical Stachelhaus codes were predicted for the A3 orthologous group. In vitro biosynthesis of mainly hydroxamate and some catecholate siderophores was observed. Hence, *Armillaria* spp. generally contain one highly conserved, NDSS cluster although some interspecific variations in the products of these clusters is expected. Results from this study lays the groundwork for future studies to elucidate the molecular biology of fungal phyto-pathogenicity.

Keywords: Basidiomycetes, Comparative genomics, Iron acquisition, Phytopathogenic fungi, Secondary metabolism, Siderophore

1. Introduction

Microorganisms, including fungi, are known to produce diverse low molecular weight organic compounds known as secondary metabolites (SMs). Microbial SMs typically mediate ecological interactions such as mutualism (Amin et al. 2009; Johnson et al. 2013), competition (Butaitė et al. 2017; Stubbendieck and Straight 2016), predation (Stubbendieck and Straight 2016) and pathogenicity (Oide et al. 2006) and play a critical role in microbial survival in ecological niches (Kuo et al. 2020; Stubbendieck and Straight 2016). These metabolites include terpenes or terpenoids (e.g., Melleolides) (Engels et al. 2011; König et al. 2019), polyketides (PKSs; e.g., Lovastatin) (Amnuaykanjanasin et al. 2005; Gunde-Cimerman and Cimerman 1995), and nonribosomal peptides (NRPs; e.g., epidithiodioxopiperazines and siderophores) (Brandenburger et al. 2017; Eisendle et al. 2003; Kuo et al. 2020; Le Govic et al. 2019; Schwecke et al. 2006; Welzel et al. 2005; Winterberg et al. 2010; Yuan et al. 2001). Genes involved in fungal SMs biosynthesis are located in secondary metabolite gene clusters (SMGCs) which contain genes that encode biosynthetic enzymes such as terpene cyclases, polyketide synthases, and nonribosomal peptide synthetases (NRPSs), usually named according to the SM types they produce (Brandenburger et al. 2017; Kadi and Challis 2009; Le Govic et al. 2019; Le Govic et al. 2018).

NRPSs are multifunctional enzymes which assemble the SM backbone structure by use of multiple active sites known as domains. The adenylation (A), thiolation (T; also known as peptide carrier protein, PCP), and condensation (C) domains are required for NRPS functionality. Other domains, classified as accessory domains, such as epimerization (E), formylation (F), methylation (M), reduction (R), oxidation (Ox), and thioesterase (TE) are optional and confer some structural and functional differences to the NRPs synthesized. A-domains are responsible for substrate selection and activation using ATP, whereas T- and C-domains catalyse phosphopantetheine (PP)-linkages of substrates with thioester bonds and peptide- or ester-bond formation between two adjacent linked substrates, respectively (Kadi and Challis 2009; Walsh 2008). TE domains often terminate NRP synthesis, although reductive chain termination is sometimes employed in the synthesis (Kadi and Challis 2009). NRPS are described as either monomodular if they contain only one set of the main domains or multi-modular when they contain more than one complete set of domains and may contain accessory domains (Kadi and Challis 2009; Le Govic et al. 2019). The multiplicity of the modules in an NRPS as well as the number and types of accessory domains it contains influence the complexity and structural diversity of the peptides synthesized (Gaitatzis et al. 2001).

Other enzymes in NRPS clusters function in glycosylation, acylation, halogenation, or hydroxylation for modifying the NRPS substrate or the NRP synthesized, as well as regulation for regulating production of the NRP (Brandenburger et al. 2017; Le Govic et al. 2019; Le Govic et al. 2018). NRPS

clusters may also include transporters for transporting the synthesized peptide. Other enzymes in the NRPS clusters include monooxygenases, acyltransferases, major facilitator transporters, and Zn_2Cys_6 transcription factors (Brandenburger et al. 2017; Le Govic et al. 2019; Le Govic et al. 2018). Genes encoding these proteins have been reported in NRPS-dependent siderophore synthetase (NDSS) clusters (Brandenburger et al. 2017; Le Govic et al. 2019; Le Govic et al. 2018).

Siderophores are peptides biosynthesized using one of the two main biosynthesis pathways. These are the NRPS-dependent siderophore synthetase pathway or the NRPS-independent siderophore (NIS) synthetase pathway (Brandenburger et al. 2017; Iftime et al. 2016; Kadi and Challis 2009; Kurth et al. 2016; Salwan and Sharma 2020). A third pathway, called the hybrid NRPS/NIS pathway, which utilizes a blend of both pathways may also be used (Kadi and Challis 2009; Salwan and Sharma 2020).

Siderophores are used by fungi (Brandenburger et al. 2017; Chen et al. 2019; Johnson et al. 2013; Koulman et al. 2012; Kuo et al. 2020; Mukherjee et al. 2018; Narh Mensah et al. 2018; Schrettl et al. 2007; Wallner et al. 2009; Welzel et al. 2005; Yuan et al. 2001) and other organisms (Audenaert et al. 2002; Butaitė et al. 2017; Carrano et al. 2001; Devireddy et al. 2010; Kugler et al. 2020; Maglangit et al. 2019; Schrettl et al. 2007; Taguchi et al. 2010; Wang et al. 2021) to solubilize, sequester, transport/shuttle, and store ferric iron (Fe^{3+}) especially under iron-limiting conditions, such as aerobic conditions (Schrettl et al. 2007). Siderophores are also involved in other biological functions such as structuring microbial communities and, in the case of pathogenic siderophore producers, enhancing virulence factors and suppression of host defence mechanisms (Amin et al. 2009; Gu et al. 2020; Paauw et al. 2009; Taguchi et al. 2010). Based on the structure of their Fe^{3+}-chelating functional groups, these molecules are grouped into five classes; catecholates and phenolates (collectively called aryl caps), hydroxamate, carboxylates, and mixed type, which have at least two Fe^{3+}-coordinating moieties (Carrano et al. 2001; Haas et al. 2008).

Fungal siderophores have been implicated in outcompeting or inhibition of plant pathogens, alteration of host defence mechanisms such as induced systemic resistance, as well as alteration of plant-fungal interactions (Johnson et al. 2013; Mukherjee et al. 2018). The intracellular siderophores ferricrocin and hydroxyferricrocin have been reported being required for germ tube formation, asexual sporulation, resistance to oxidative stress, catalase A activity, and virulence in *Aspergillus fumigatus* (Schrettl et al. 2007). The plant endophytic fungus, *Epichloë festucae,* produces the extracellular siderophore epichloënin A (Johnson et al. 2013). This siderophore is believed to serve as an important molecular/cellular signal for controlling fungal growth (Johnson et al. 2013). Lack of epichloënin A can alter the homeostasis of the symbiotic relationship between *E. festucae* and its host, *Lolium perenne,* from being mutually beneficial to being antagonistic (Johnson et al. 2013). *Trichoderma*

spp., including *T. virens*, produce siderophores that suppress growth of *Armillaria* spp. when they are grown together in cultures (Chen et al. 2019).

Armillaria spp. are terrestrial ubiquitous white rot basidiomycetes in the family Physalacriaceae, and typically have a broad host range (Baumgartner and Rizzo 2002; Coetzee et al. 2003; Elías-Román et al. 2018; Prospero et al. 2004; Warwell et al. 2019). *Desarmillaria* spp., that were previously treated as *Armillaria* spp., are exannulate species which include *Desarmillaria tabescens* and *D. ectypa* (Koch et al. 2017). Fungi in these genera have mycorrhizal, saprophytic, and/or pathogenic symbiotic relationships with plants and most of the species are considered to be facultative necrotrophs (Baumgartner and Rizzo 2002; Coetzee et al. 2003; Elías-Román et al. 2018; Prospero et al. 2004; Warwell et al. 2019). Other members of the Physalacriaceae that were considered in the current study include the closely related *Guyanagaster necrorhizus,* that shares a recent ancestor with *Armillaria* (Koch et al. 2017), as well as *Cylindrobasidium torrendii* and *Oudemansiella mucida*. Studies aiming to understand and eventually inhibit the predominantly pathogenic *Armillaria* species and the adverse effects they have on horticulture, agriculture, and forestry are therefore warranted.

In this study, we hypothesize that variation in siderophore production by various species and strains of *Armillaria* may influence the reported pathogenicity of these fungi. To test this hypothesis, we investigated the genomes of selected *Armillaria* spp. and other members of the Physalacriaceae to identify the types and numbers of SMGCs in the genomes *in silico.* We also determined the presence and putative characteristics of NRPS-dependent siderophore biosynthesis clusters in the genomes. Finally, we characterized the NRPS-dependent siderophore biosynthesis genes in comparison to those characterized in other organisms *in silico.* We also utilized bioassays to evaluate the production and the types of siderophores produced by strains representing selected *Armillaria* spp.

2. Materials and methods

All *in silico* studies were performed according to the flowchart presented in Fig. S1. For *in vitro* studies, all glassware was washed with 6 M HCl followed by 3 washes with doubly deionized H_2O (ddH_2O) to remove all trace metals before use (Schwyn and Neilands 1987). Media were autoclaved at 121 °C for 20 min.

2.1. *In silico* identification of secondary metabolite gene clusters

2.1.1. Genome mining of selected strains

Whole genomes of *Armillaria* spp. and other members of the Physalacriaceae (Table 1) were mined for SMGCs using previously described methods (Agger et al. 2009; Le Govic et al. 2019; Sayari et al. 2019) but with some modification. In brief, secondary metabolite gene clusters in each of the

genomes under study were predicted with the Antibiotics & Secondary Metabolite Analysis SHell (antiSMASH) using the fungal version (fungiSMASH) v. 6.0.0alpha1-56f63c8 (Blin et al. 2019b). Genome sequences uploaded into fungiSMASH for these analyses were initially annotated, with gene annotation files obtained from the Mycocosm fungal genome sequence database of JGI (https://mycocosm.jgi.doe.gov/mycocosm/home/releases?flt=Physalacriaceae) in CLC Main Workbench v21.0.4 (www.qiagenbioinformatics.com) using the "Annotate with GFF/GTF/GVF file" plugin. Relaxed detection strictness was used for these analyses. All extra features (excluding the Cluster-border prediction based on transcription factor binding sites (CASSIS) feature, which was inactive during the study) were also activated. Cluster boundaries were manually delimited up to approximately 10 Kb upstream and 30 Kb downstream of the cluster boundaries detected by fungiSMASH. These regions were selected to account for the large non-coding region in the *A. novae-zelandiae* cluster. Minimum Information about a Biosynthetic Gene (MIBiG) cluster comparison, ClusterBlast, KnownClusterBlast, Cluster Pfam Analysis, and Pfam-based GO term annotation were used for the purpose of detecting similar gene clusters. All of these are implemented in fungiSMASH. MIBiG comparison shows areas that are similar to the region of the query biosynthetic gene cluster (BGC) to the MIBiG Database (Blin et al. 2019a; Medema et al. 2015). ClusterBlast recover gene clusters from the antiSMASH database and other clusters of interest that are homologous to the query gene cluster (Medema et al. 2011). KnownClusterBlast shows clusters from the MIBiG database that are similar to the current region (Blin et al. 2019a). Various types of secondary metabolite clusters were identified in the analyzed genomes by fungiSMASH. However, since the focus of this study was on NRPS-dependent siderophore synthetase (NDSS) clusters, all further analyses were conducted only on predicted NDSS clusters.

2.1.2. Gene annotation, gene order confirmation, and generation of cluster synteny maps

The gene order and orientation were predicted using results obtained from fungiSMASH and the genome annotations obtained using CLC Main Workbench. Putative gene annotation was done by conducting a tBLASTn search using the amino acid sequences of the respective genes (as obtained from fungiSMASH) in the NCBI BLAST web-based tool using default parameters (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The search parameter was altered to include only Fungi (taxid:4751) for all the NRPS gene sequences because the sequences were longer than 4000 amino acids. The top 100 closest orthologs obtained were sorted according to percentage identity and the closest experimentally characterized orthologs were used for the putative annotation of the identified genes in the clusters. All other genes in the flanking regions beyond the cluster boundaries predicted by fungiSMASH were annotated by conducting tBLASTx searches with the annotated sequences using the BLAST tool implemented in CLC Main Workbench against the NCBI database. Other

information for the flanking genes were obtained from InterProScan. Cluster synteny maps were drawn using EasyFig v2.2.5 (Sullivan et al. 2011).

2.2. *In silico* characterization of NRPS-dependent siderophore synthetases

2.2.1. Determination of size, intron number, domain architecture and modular organization of NRPSs

Number of introns of predicted NDSS genes were observed in CLC based on the annotated genomes described in section 2.1.1 which include the use of RNAseq data. Size information of these genes were obtained from fungiSMASH. The size of the amino acid sequence was confirmed with InterProScan (https://www.ebi.ac.uk/interpro/). The domain architectures and modular organizations of identified NDSS were obtained from fungiSMASH, as well as Pfam searches in InterProScan using the amino acid sequences. Any disparities observed between the data from the two approaches were reconciled with results obtained from searches of the amino acid sequence of the respective gene against the Pfam v 33.1 web-based tool (https://pfam.xfam.org/). Attempts to use the PKS/NRPS Analysis Website (http://nrps.igs.umaryland.edu/) (Bachmann and Ravel 2009), were not successful as most of the domains were not recognized by the tool.

2.2.2. A-domain phylogenetic analyses and substrate specificity prediction

Datasets of the amino acid sequences of the A-domains of the NDSSs identified in the Physalacriaceae (obtained from fungiSMASH) were combined with the A-domain amino acid sequences of experimentally characterized NDSSs from other species, which had relatively high similarity (based on the results of the BLAST searches conducted under "Gene annotation, gene order confirmation, and generation of cluster synteny maps" to the putative NDSS genes of the presently studied members of the Physalacriaceae. The amino acid sequence of the A3 domain of *A. novae-zelandiae* was obtained from the region identified from Pfam analysis. Sequences were aligned using the MUSCLE algorithm in the MEGA X software (Stecher et al. 2020). Maximum likelihood algorithm implemented in MEGA X with the LG substitution model was used to construct a phylogenetic tree. Nodal support was obtained using 100 bootstrap replications and with the same setting used to obtain the phylogenetic tree.

NDSS A-domain specificities (i.e., amino acid codes and substrate specificities) were predicted using NRPSpredictor2 (Rausch et al. 2005; Röttig et al. 2011) implemented in fungiSMASH. NRPSpredictor2 uses both the signature sequence method (Stachelhaus codes or amino acid codes) (Stachelhaus et al. 1999) and the Support Vector Machines (SVM)-based method (Rausch et al. 2005; Röttig et al. 2011). The predicted A-domain specificities were compared to those of other

characterized NDSSs which were found to be orthologous to the NDSSs identified in the present study.

2.3. Assessment of siderophore production potential of selected *Armillaria* strains

2.3.1. Fungal strains, maintenance, and growth conditions

The fungal cultures used were strains of *A. fuscipes, A. gallica, A. mellea, A. luteobubalina* and *A. nabsnona* (Table 2). These species were selected to represent *Armillaria* species reported to have different ecology in terms of pathogenicity and virulence, and to serve as a proof-of-concept for siderophore biosynthesis by *Armillaria* species. Cultures were maintained on malt yeast extract agar (MYA) medium (15 g/L malt extract, 2 g/L yeast extract, 15 g/L agar) and incubated at 25±2 ºC in the dark for 53 days. All reagents were purchased from commercial suppliers.

2.3.2. Siderophore detection

Siderophore production and activity was detected by both the universal chrome azurol S agar (CAS) (Alexander and Zuberer 1991) and the modified chrome azurol S agar medium (split CAS/MYA) (Milagres et al. 1999) assays. These methods were used to detect siderophore production as well as generally characterize the siderophores produced by the mycelia and rhizomorphs of the studied strains of *Armillaria* species. CAS agar at pH 6.8 was prepared using the method described by Alexander and Zuberer (1991). Plates were dried overnight in a flow cabinet to ensure excess water was removed and incubation of strains followed the description above. Three replicates for each strain on each media were used in this study.

Split CAS/MYA plates were prepared as follows. CAS agar medium and MYA were separately prepared, sterilized and maintained at 50 °C in a water bath. MYA was poured first in the petri dishes and allowed to set for 1h. Half of the set MYA was aseptically cut and discarded, followed by pouring the cooled CAS media into the empty half of the plate and allowing to set. The experiment was also performed using only MYA media to investigate differences, if any, in mycelia and rhizomorph growth rates on the CAS media in comparison to the growth media.

Cultures were inoculated in the centre of the CAS and MYA plates, and in the middle of the MYA part of the split CAS/MYA plates. The inoculum consisted of 1 cm^2 culture discs sub-cultured from an actively growing strain on MYA medium. Plates were observed for change of the CAS media from blue to reddish-orange, orange, purplish-red, and purple (Arora and Verma 2017; Milagres et al. 1999). The type of siderophore produced was classified according to the colour change. Mycelia

macro-morphology was observed directly as judged by the investigators to detect any effect of culture media on culture morphology. Un-inoculated plates served as the control for this experiment.

3. Results

3.1. Identification of secondary metabolite biosynthesis gene clusters

3.1.1. Secondary metabolite gene clusters in the studied Physalacriaceae genomes

Different secondary metabolite gene clusters putatively determined as being involved in various secondary metabolite biosynthesis were identified in the genomes of the species included in this study using *in silico* analyses (Fig. 1). These SMGCs included clusters encoding NRPS, NRPS-like, siderophore (suspected NIS clusters), type 1 polyketide synthase (T1PKS), terpene, as well as hybrid clusters including NRPS-Like/Terpene, NRPS-Like/T1 PKS, and T1PKS/Terpene clusters. The genome of *O. mucida* also contained a cluster predicted to be involved in the production of indole (data not shown).

The total number of SMGCs in the *Armillaria* and *Desarmillaria* genomes ranged from 32 in *D. tabescens* to 56 in *A. borealis* (Fig. 1). The most abundant type of gene clusters in the genomes were those encoding enzymes involved in the biosynthesis ofterpenes (13 – 30 clusters), followed by NRPS-like (9 – 19 clusters), and T1PKS (1 – 8 clusters). The maximum number of NPRS, siderophore, and hybrid clusters detected were 2, 3, and 5, respectively. *Cylindrobasidium torrendii* was unique for possessing only 4 types of secondary metabolite clusters, totalling 22 in its genome (14 terpenes, 6 NRPS-like, and 1 each of NRPS and T1PKS).

No direct correlation was observed between the number and types of SMGCs and reported pathogenicity (Fig. 1). For instance, the pathogenic species, *A. novae-zelandiae* contained 48 (1 NRPS, 19 terpenes, 4 T1PKS) SMGCs. Likewise, the weakly pathogenic species, *A. nabsnona* also contained 48 SMGCs (1 NRPS, 23 terpene, 3 T1PKS). The facultative necrotrophs which are weakly pathogenic, *A. borealis* and *A. cepistipes*, contained a total of 56 (3 NRPS, 24 terpene, 8 T1PKS) and 55 (1 NRPS, 30 terpene, 3 T1PKS) SMGCs, respectively. The closely related species, *G. necrorhizus* contained a total of 30 SMGCs (1 NRPS, 13 terpene, 1 T1PKS). The pathogenicity of this species is unknown.

In all the genomes studied, the NRPS gene in only one of the different NRPS clusters identified by fungiSMASH in each species showed hits in the MIBiG database. The hit with the strongest support for all species (% identity = 27.0, % coverage = 107.1, BLAST score = 1438.0, E-value = 0.0) was with the epichloenin_A_synthetase cluster (accession number AET13875.1, which produces an extracellular siderophore identified in *Epichloë festucae*). There was no hit for the ClusterBlast nor

KnownClusterBlast. All further analysis was conducted on this cluster since NRPS-dependent siderophore synthetase clusters was the focus of this study.

3.1.2. Gene annotation and synteny comparison of NRPS-dependent siderophore synthetase clusters and flanking genes

Genome walking of the regions up- and down-stream of the clusters identified by fungiSMASH (Table S1) revealed that the predicted NDSS clusters in the *Armillaria* and *Desarmillaria* species studied ranged from approximately 72 Kb in *A. fumosa* to approximately 100 Kb in *A. novae-zelandiae*. The clusters identified in the other members of the Physalacriaceae were approximately 67, 51, and 94 Kb in *C. torrendii*, *G. necrorhizus*, and *O. mucida,* respectively.

The NDSS clusters in the *Armillaria* and *Desarmillaria* species were largely syntenic (Fig. 2, Table S2). However, the NDSS genes of *A. mellea* and *D. ectypa* showed less conserved synteny. The flanking genes located upstream of the NDSS genes contained a syntenic block comprising genes that encode RecF/RecN/SMC (involved in structural maintenance of chromosomes), choline/ethanolaminephospho-transferase, cellobiose dehydrogenase and at least one Major Facilitator Superfamily transporter (MFS) gene. The upstream region also contained genes encoding (membrane-bound) hypothetical proteins and natterin-like hypothetical proteins. These genes and the intergenic regions also showed a high degree of synteny. The genes downstream of the NDSS genes were generally syntenic for most species of *Armillaria* and *Desarmillaria* although an inversion was observed in *A. borealis* (Fig. 2). *A. mellea* was unique for possessing cytochrome P450 genes downstream of the NDSS gene. The species in the phylogenetic sister genus of *Armillaria*, *D. ectypa* and *D. tabescens,* had unique genes downstream of the NDSS genes. The genes putatively encode a member of the isochorismatase family protein, FAD-dependent oxidoreductase, cellobiose dehydrogenase, as well as (membrane-bound) hypothetical proteins (Fig. 2, Table S2). These genes were not found when a search was conducted further downstream of the cluster for the *Armillaria* species. In addition to these genes, a DNA polymerase gene as well as two copies each of protein kinase and putative transcriptional regulator genes were uniquely found in the NDSS gene cluster in *D. tabescens*. Genes flanking the NDSS genes of *C. torrendii* and *O. mucida* included cytochrome C oxidase, a Zn_2Cys_6 fungal-type domain-containing protein, alpha/beta-hydrolase, or mannitol dehydrogenase, and various (membrane-bound) hypothetical or uncharacterized proteins (data not shown).

3.2. Characteristics of identified NRPSs

3.2.1. Size, intron number, domain architecture and modular organization of NRPSs

The size in base pairs (bp) of the NDSS genes ranged from 15,689 in *C. torrendii* (Cylto1_490369) to 19,023 in *O. mucida* (Oudmuc1_1435297) (Table 3). The smallest and largest size of the translated amino acids in these genes was found in Armbor1_1740653 (4329 aa) and Oudmuc1_1435297 (5353 aa), respectively. The number of introns in the NRPS genes ranged from 45 in Cylto1_490369 to 58 in Oudmuc1_1435297 (Table 3). The protein sequences of the NDSS genes showed significant matches (query cover 80 – 97%; percentage identity 26.5 – 34.5 %; expected value 0.0) with translated nucleotide sequences of characterized NDSS genes in different fungi responsible for hydroxamate ferricrocin-type siderophore biosynthesis, as well as the uncharacterized *NRPS2* in *W. ichthyophaga* EXF-994 (Table 3). The NDSSs were all multi-modular showing three complete A-T-C modules organization ending with two T-C didomains [$(A\text{-}T\text{-}C)_3(T\text{-}C)_2$] (Table 4). A C-terminal of an A-domain was also located after the first A-domain of Guyne1_959764, as well as the NDSSs of *As. fischeri* and *F. sacchari*.

3.2.2. Phylogeny, Stachelhaus sequences, and substrate specificity of A-domains of putative NDSSs

Phylogenetic analysis of the A1 – A3 domains of the putative NDSSs placed the sequences in three main clusters with strong nodal support in the phylogenetic tree (Fig. 3). These clusters represented the different domains A1, A2 and A3, respectively. All A1-domains formed a sister group with the A2-domain cluster with 97 % bootstrap support. The A3-domain cluster formed a sister group to the A1 and A2 clusters. Within the A1 and A2 domain clusters, sequences from species in the Physalacriaceae formed respective monophyletic groups.

Sequences of all *Armillaria* species, except for those from *A. novae-zelandiae*, formed supported monophyletic groups within the A1 and A3 domain clusters. In the A1 domain cluster, the sequence of *A. novae-zelandiae* was more closely related to that of *G. necrorhizus*, a species that is in the sister genus of *Armillaria* and *Desarmillaria*. Within the A2 domain cluster, sequences from all *Armillaria* species formed a strongly supported group. In all three domain clusters, sequences from *D. ectypa* and *D. tabescens* formed groups separated from the *Armillaria* species but phylogenetically more closely related to species of *Armillaria* than to those of other Physalacriaceae. In all three domain clusters, sequences of *G. necrorhizus* were placed more closely related to *Armillaria* and *Desarmillaria* than to sequences from other species in the Physalacriaceae.

Stachelhaus codes and predicted substrate specificities for the A-domains of the predicted NDSSs of the members of the Physalacriaceae in comparison to those reported for characterized NDSSs in other fungal species gave more insights about the characteristics of the presently studied NDSSs (Fig. 3). The code of the A1 domains of the studied species was in most cases DPMMWMAINK. Those for *A. novae-zelandiae* (Armnov1_1625296), *E. festucae* strain E2368, and *E. festucae* strain F11 were DPMMDG-VFK, DLLMWCAITK, and DLLMWCAITK, respectively (Fig. 3). The code of the A2 domains of all the studied members of the Physalacriaceae were the same (DVLFIITIIK). This code was shared with the A2 domain of the NRPS of the halophilic basidiomycete, *W. ichthyophaga* EXF-994. In contrast, the A2 domain code of the basidiomycete, *U. maydis* 521 was DVQHTITVVK (Schwecke et al. 2006; Winterberg et al. 2010). A different A2 domain code (DVQHTITIVK) was shared by the ascomycetes, *As. fischeri* and *As. fumigatus* (Schwecke et al. 2006). Among these fungi, *As. nidulans* differed in that it contained a valine (val) instead of an isoleucine (iso) at the 8th position of the A2 domain code, although both amino acids have hydrophobic side chains. The A3 domains were more diverse. *C. torrendii* (Cylto1_490369) had the distinct A3 domain code, GVTALGVGIK, which only shared the last three amino acids of the A3 domain code of the *Armillaria* and *Desarmillaria* species. *O. mucida* (Oudmuc1_1435297) also had a unique A3 domain code, ESVFVIASFK.

3.3. Siderophore production and mycelial growth of selected *Armillaria* spp.

Siderophores were produced by all the strains studied in both the CAS and CAS/MYA media (Fig. 4). The CAS media on some of the plates (e.g., *A. fuscipes* strains CMW2740 and CMW3164, and *A. mellea* strain CMW31132 on CAS) showed different discoloration from blue to orange and/or reddish-orange closer to the mycelia to purple and/or purplish-red further from the mycelia and rhizomorphs, indicative of co-production of different types of (extracellular) siderophores by the strains studied. This was also the case for *A. luteobubalina* strain CMW4974 and *A. nabsnona* strain CMW6904 on CAS/MYA. The change of CAS media from blue to purplish-red for *A. fuscipes* strains CMW2740 and CMW3164 was notably more pronounced on CAS/MYA than on CAS.

Mycelia culture morphology and growth rates were altered on the CAS media for some cultures (Fig. 4). For instance, growth rate of *A. luteobubalina* strain CMW4974 was fastest on MYA followed by CAS/MYA and slowest on CAS by the end of the incubation period. The reverse trend was observed for *A. nabsnona* strain CMW6904 that showed the fastest culture growth on CAS and slowest growth on MYA. *Armillaria luteobubalina* strain CMW4974 changed from crustose mycelia on MYA to cottony mycelia on CAS. Both mycelia and rhizomorph growth of *A. luteobubalina* strain CMW4974 and *Armillaria* sp. ACB strain CMW4456 were slower when cultured on CAS and CAS/MYA compared to MYA.

4. Discussion

4.1. Genome-wide identification of secondary metabolite gene clusters

4.1.1. Secondary metabolite gene clusters in the studied genomes

In this study we explored the diversity of SMGCs in the genomes of *Armillaria* spp., *Desarmillaria* spp. and other members of the Physalacriaceae using a variety of bioinformatic tools. Different types of SMGCs namely NRPS, T1PKS, NRPS-like and terpene, as well as hybrid clusters NRPS-like/T1PKS, NRPS-like/terpene T1PKS/terpene were identified in the genomes. Clusters identified to be responsible for siderophore biosynthesis (suspected NRPS-independent siderophore biosynthesis clusters) by fungiSMASH were also identified in the genomes of the fungi studied. These types of SMGCs ranged in their numbers among the fungi.

There was no correlation in terms of cluster number and fungal lifestyle. Although this was unexpected, we postulate that the role of the secondary metabolites/metabolism in the studied species in relation to pathogenicity and/or virulence will be influenced by various biotic and abiotic factors such as the specific host and its state during host-fungal interaction, competing organisms in the environment, pH, temperature, and nutrient availability in the environment, and molecular factors such as transcription factors and other regulators. This assertion is supported by the fact that pathogenicity and/or virulence of *Armillaria* spp. and strains are in themselves affected by various factors such as intra species variation, forest management systems, host species, state of the hosts (e.g., stressed or healthy), as well as environmental factors (e.g., elevation) (Legrand et al. 1996; Mesanza et al. 2017; Prospero et al. 2004; Tsykun et al. 2012). The factors influencing the regulation and role of (plant-pathogenic) fungal SMs has been reviewed by various authors including Macheleidt et al. (2016) and Pusztahelyi et al. (2015).

The results of this research agree with previous studies on secondary metabolite genes and gene clusters in *Armillaria* species. Sipos et al. (2017) showed that 5 trichodiene genes, 1 polyprenyl synthase gene, and 6 polyketide synthase genes are expressed in the rhizomorphs of *Armillaria* spp. Engels et al. (2011) have reported on genes and gene clusters responsible for the biosynthesis of melleolides and armillyl orsellinates in *A. gallica*. In addition, various polyketides synthesized by polyketide synthases in PKS clusters have been reported to be produced by fungi (Gunde-Cimerman and Cimerman 1995). These include lovastatin identified in a white-rot basidiomycete species belonging to the genus *Pleurotus* (Gunde-Cimerman and Cimerman 1995), as well as depudecin and 19,20-epoxycytochalasin Q synthesized by the woody plant saprophytic or weakly pathogenic ascomycete species in the genus *Xylaria* (Amnuaykanjanasin et al. 2005). The low (less than 50 %) identities of the best hits in characterized NRPS gene clusters in MIBiG database, and the absence of

hits for ClusterBlast and KnownClusterBlast, suggests novelty of these NRPS gene clusters detected in the Physalacriaceae. A hit of less than 50 % in the MIBiG database was used to predict the novelty of thumolycin biosynthetic gene clusters in *Bacillus thuringiensis* (Zheng et al. 2018).

4.1.2. Synteny of predicted NDSS gene clusters and neighbouring genes

The predicted NDSS gene clusters and the neighbouring genes showed a high degree of microsynteny both among the genes and the intergenic regions despite the observed inversion and duplication events in some of the clusters. However, there were some variations in the cluster for *A. mellea, A. novae-zelandiae, D. ectypa*, and *D. tabescens.* Conserved synteny/similarity in biosynthetic gene clusters among species from the same genera has previously been reported and this is also reflected in this study (Evdokias et al. 2021; Goering et al. 2016).

Compared to the *Armillaria* spp., the genes downstream from the aryl-alcohol dehydrogenase genes differed in the *Desarmillaria* spp. *Desarmillaria tabescens* harboured a unique set of genes that included a gene which encodes an isochorismatase family protein. An ortholog of this gene was also found in the predicted NDSS gene cluster of *G. necrorhizus*. An isochorismatase and an isochorismate synthase gene have been reported to be located within an NDSS gene cluster of *Streptomyces* sp. strain ATCC 700974 (Patzer and Braun 2010). This cluster synthesizes the catechol-peptide siderophore, griseobactin (Patzer and Braun 2010). Additionally, a bifunctional isochorismate lyase/aryl carrier protein, DhbB, has been identified in an NRPS-dependent catecholic siderophore biosynthesis operon of *Bacillus subtilis*, which synthesizes itoic acid (2,3-dihydroxybenzoate (DHB)-glycine) (May et al. 2001).

Genes typical of those in NDSS gene clusters were identified in this study. These included monooxygenases, dehydrogenases, oxidoreductases, transferases, transporters, and transcription factors (Eisendle et al. 2003; Schwecke et al. 2006; Welzel et al. 2005; Winterberg et al. 2010). Contrary to our findings, the NRPS cluster in *S. apiospermum* contains an L-ornithine-N^5-monooxygenase gene just downstream of the NRPS gene, SAPIO_CDS9032. L-ornithine-N^5-monooxygenase genes have also been reported in the NDSS clusters of the ascomycetes, *Schizosaccharomyces pombe* (Schwecke et al. 2006), *As. fumigatus* (Schrettl et al. 2007), and *As. nidulans* (Eisendle et al. 2003), as well as the basidiomycetes, *C. subvermispora* (Brandenburger et al. 2017), *Omphalotus olearius* (Welzel et al. 2005), and *Ustilago maydis* (Winterberg et al. 2010). This enzyme catalyses N^5-hydroxylation of L-ornithine during hydroxamate-type siderophore biosynthesis (Eisendle et al. 2003). The absence of the L-ornithine-N^5-monooxygenase genes in the presently studied NDSS gene clusters indicates that the products of the latter will be structurally

different from those of the characterized clusters in other fungi. This assertion is also supported by the absence of a hit in the ClusterBlast and KnownClusterBlast searches.

4.2. NDSS gene characteristics

4.2.1. Genetic and structural characteristics of identified NDSS genes

The characteristics of the identified NDSS genes were largely similar among the *Armillaria* and *Desarmillaria* spp. in comparison to other members of the Physalacriaceae and the characterized NDSS genes of other fungi. Similar sizes have been reported for other NDSS genes responsible for production of intra- and extra-cellular siderophores. However, the translated NRPS gene of *O. mucida* (Oudmuc1_1435297) was larger than the sizes recorded for the NDSS genes of all the other members of the Physalacriaceae and the characterized NDSS genes from the other fungi.

The number of introns identified in this study compared favourably with those reported for other NDSS genes. The NDSS gene, *FSO1,* in *Om. olearius* has 48 introns (Welzel et al. 2005). The percentage of the translated NDSS genes that were covered by the target sequences (query covers) reported by Le Govic et al. (2018) for NDSS of *S. apiospermum* against the NDSS of *As. fumigatus* Af293 were comparable to those reported herein. The lowest query covers were recorded for the translated NDSS genes of *A. novae-zelandiae* (Armnov1_1625296) and *O. mucida* (Oudmuc1_1435297) in comparison with the characterized NDSS genes of the other fungi.

Our results indicate that the NDSS genes identified in the studied Physalacriaceae genomes encode siderophore synthetases. The $(A\text{-}T\text{-}C)_3(T\text{-}C)_2$ domain architecture, observed in this study, has been previously reported for NDSSs from *As. fumigatus*, *As. nidulans*, *E. festucae*, *F. sacchari, Om. olearius*, and *U. maydis* (Johnson et al. 2013; Munawar et al. 2013; Schwecke et al. 2006; Welzel et al. 2005). These proteins synthesize the hydroxamate-type siderophores, ferricrocin (Schrettl et al. 2008; Wallner et al. 2009), ferricrocin (Eisendle et al. 2003), epichloënin A (Johnson et al. 2013; Koulman et al. 2012), ferrirhodin (Munawar et al. 2013), ferrichrome A (Welzel et al. 2005), and ferrichrome A (Yuan et al. 2001), respectively. All genes encoding these proteins, except the gene encoding AAX49356.1 of *Om. olearius*, were orthologous to the genes for NDSSs predicted in the studied Physalacriaceae.

Results from this study provide some insights into the characteristics of the putative NDSSs of the studied Physalacriaceae. The variations in products biosynthesized by known NDSSs, irrespective of similarities in domain architectures, demonstrate that NRPSs with the same/similar architectures can synthesize different siderophores (Schwecke et al. 2006). Likewise, NDSS genes which differ in their domain architectures have been found to synthesize homologous products (Schwecke et al. 2006).

Hence, based on domain architectures, similarity of the products of the presently predicted NDSSs to that reported for the orthologous NDSSs which synthesize the characterized siderophores in other fungi cannot be fully established. It is also unclear, based on the domain architectures in comparison to characterized NDSSs, if the siderophores produced by the NDSS genes of members of the Physalacriaceae are intracellular, extracellular, or both.

4.2.2. A-domain phylogeny, Stachelhaus sequences, and substrate specificity

Phylogeny of the A-domains of NDSSs in the *Armillaria* and *Desarmillaria* species showed that the domains are highly conserved. The phylogenetic analyses also placed the A-domains into three well supported orthologous groups, which have been designated A1, A2 and A3. The Stachelhaus codes and the associated predicted substrate specificities of the A-domains were similar to the observations made using the phylogenetic analyses. The codes DPMMWMAINK, DVLFIITIIK, DVQHTITVVK, and DVQHTITIVK have been previously reported in other NDSSs (Schwecke et al. 2006).

The A1 domain code, DPMMWMAINK, has previously been predicted to activate serine (Schwecke et al. 2006). However, NRPSPredictor2 failed to allow for the prediction of a substrate specificity for the A1 domains bearing DPMMWMAINK code in the NDSSs identified in this study except for those of *D. tabescens* and *C. torrendii*. The predicted candidate substrates of the A1 domains of these species were three proteinogenic amino acids with hydrophobic side chains (val, leucine (leu), and iso), and two non-proteinogenic amino acids (L-alpha-aminobutyric acid (abu) and isovaline (iva)).

The A2 domain codes DVLFIITIIK, DVQHTITIVK, and DVQHTITVVK have previously been predicted to activate glycine (Schwecke et al. 2006). However, the results from our study suggest that DVLFIITIIK activates proline (pro) or leu, DVQHTITIVK activates pro, and DVQHTITVVK activates hydrophobic-aliphatic amino acids in general.

The A3 domain codes found in the members of the Physalacriaceae differed from the A3 domain codes previously reported for other characterized NDSSs in *As. fumigatus, As. nidulans, E. festucae, Om. olearius,* and *U. maydis* (Johnson et al. 2007; Schwecke et al. 2006; Welzel et al. 2005; Winterberg et al. 2010). The A3 domains of the NRPSs of these species activate N^5-hydroxy-L-ornithine (L-AHO), N^5-*trans*-anhydromevalonyl-N^5-hydroxyornithine (*trans*-AMHO), *N*5-*trans*-(α-methyl)-glutaconyl-N^5-hydroxy-L-ornithine (L-MGHO), and N^5-*trans*-(α-methyl)-glutaconyl-N^5-hydroxy-L-ornithine (L-MGHO), respectively (Schwecke et al. 2006; Welzel et al. 2005; Winterberg et al. 2010). It is noteworthy that L-N^5-ornithine monooxygenases have been found to be located within the clusters bearing the characterized NDSSs in the other fungi (Le Govic et al. 2018; Schwecke et al. 2006). No other gene in the characterized NDSS gene clusters in the genomes of the

other fungi has been associated with an enzyme linked to the other A-domains of the NRPSs encoded by the respective NRPS genes.

Inaccuracy of NRPS A-domain substrate predictions of various tools including NRPSPredictor2 has previously been reviewed in Vassaux et al. (2019). Nonetheless, we have demonstrated that the A1-A3 domains of the NDSSs predicted in the genomes of the studied *Armillaria* spp. and other members of the Physalacriaceae may also differ in terms of their substrate specificity compared to characterized NDSSs in other fungi.

4.3. Siderophore biosynthesis and mycelial growth of selected strains of *Armillaria* spp.

In vitro siderophore biosynthesis was observed in the studied *Armillaria* spp., irrespective of their lifestyle. Change from blue to reddish-orange, orange, purplish-red, and purple correspond to production of monohydroxamate, trihydroxamate, catecholate, and catecholate siderophores respectively, at neutral pH of CAS-assay medium/solution (Arora and Verma 2017; Milagres et al. 1999). Our results show that the *Armillaria* strains included in this study produce both hydroxamate and catecholate siderophores. These results concur with earlier studies in which co-production of different types of siderophores *in vitro* has been reported for various organisms (Kügler et al. 2020; Maglangit et al. 2019; Wang et al. 2021). To the best of our knowledge, the production of siderophores by strains of *Armillaria* spp. is first to be demonstrated in the present study. Our results serve as preliminary evidence that the NDSS clusters in *Armillaria* spp. are potentially functional and may biosynthesize siderophores. This is important as it has been established by Wiemann et al. (2013) that the mere existence of a SMGC (remnants) in the genomes of organisms such as *Fusarium* spp. does not necessarily translate to production of the associated secondary metabolites.

Different colour changes correspond to different types of siderophores while different colour intensities suggest production of different concentrations of siderophores (Milagres et al. 1999). In the present study, interspecific variation in siderophore biosynthesis was observed both in terms of nature and concentration irrespective of the assay used and the lifestyle of the *Armillaria* spp. Interspecific variation in siderophore biosynthesis has also been reported for *Trichoderma* spp. (Chen et al. 2019). Biosynthesis of different types and concentration of siderophores were also recorded depending on whether the universal and modified CAS assays were used. To the best of our knowledge, the presently observed differences in siderophore biosynthesis between these two assays has not been previously reported. The CAS agar contained casamino acids while MYA contained malt and yeast extracts. We propose that the observed variations in siderophore biosynthesis in the respective assays may be due to different substrates used by the presently reported NDSS and the suspected NRPS-independent siderophore synthetases in the genomes of *Armillaria* spp. Substrate

specificity assays combined with gene modifications and siderophore extraction and characterization will be required to assess this assertion.

Siderophore biosynthesis of *Armillaria* spp. in this study corroborates previous reports of siderophore biosynthesis by various basidiomycetes (Brandenburger et al. 2017; Milagres et al. 1999; Narh Mensah et al. 2018; Welzel et al. 2005). The clear zones at the mycelia margins of some basidiomycetes species cultured on CAS agar reported by Narh Mensah et al. (2018) was, however, not observed in the present study. In *C. subvermispora*, the NDSS, CsNPS2, catalyses biosynthesis of the trimeric hydroxamate siderophore, basidioferrin (Brandenburger et al. 2017). The NDSS cluster in *Om. olearius*, which contains the orthologous NDSS, FSO1, catalyses biosynthesis of the hydroxamate siderophore, ferrichrome A (Welzel et al. 2005).

The various strains of *Armillaria* spp. presently studied also showed differential growth rates on CAS. Variation in growth on the respective media was earlier reported for various basidiomycetes and other fungi, as well as gram-positive bacteria due to the toxicity of HDTMA and the sensitivity of the organisms to the HDTMA contained in CAS media (Schwyn and Neilands 1987). As such, the *Armillaria* strains may vary in their response to HDTMA resulting in the different growth rates. It would be interesting to investigate what drives the apparent differential response to HDTMA by *Armillaria* spp.

Conclusions

There is currently a large knowledge gap about secondary metabolism in Ascomycota and Basidiomycota. This study has contributed to bridging this knowledge gap in Basidiomycota from the genomic and biological perspective. We have demonstrated that *Armillaria* spp. generally contain one highly conserved, potentially functional NRPS-dependent siderophore synthetase gene cluster although some interspecific variation in the biosynthetic products of these clusters is expected. The study has also revealed that the identified NRPS-dependent siderophore synthetase gene clusters in the genomes of *Desarmillaria* spp. and the closely related *G. necrorhizus* are more syntenic to those of *Armillaria* spp. than to the other Physalacriaceae species. Additionally, the products of these NRPS-dependent siderophore synthetases and their clusters may be structurally and functionally different (in terms of affinity for ferric iron) from those of the characterized NRPS-dependent siderophore synthetase clusters in other fungal species. Inter-specific variation in siderophore biosynthesis in *Armillaria* spp. was also recorded. Molecular, biochemical, and biological characterization such as gene deletion/mutations, substrate specificity assays, proteomics, metabolomics, and co-culturing will be required to fully characterize the genes in these predicted clusters. Such studies will also be required to determine the function and relevance of the genes in

the clusters, investigate the biological roles of the predicted introns in the NRPS-dependent siderophore synthetase genes, determine the biosynthetic model / overall biosynthesis pathway of the clusters, characterize the products of the clusters, determine the ability of other organisms to utilize the products of the clusters as xenosiderophores, and to elucidate the role of the products in pathogenicity/virulence as well as other biological functions of these products if any. This is the first report of siderophore biosynthesis by strains of *Armillaria* spp.

References

Agger S, Lopez-Gallego F, Schmidt-Dannert C. 2009. Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. Mol Microbiol. 72(5):1181-1195.

Alexander DB, Zuberer DA. 1991. Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. Biol Fertil Soils. 12(1):39-45.

Amin SA, Green DH, Hart MC, Küpper FC, Sunda WG, Carrano CJ. 2009. Photolysis of iron–siderophore chelates promotes bacterial–algal mutualism. Proc Natl Acad Sci U S A. 106(40):17071-17076.

Amnuaykanjanasin A, Punya J, Paungmoung P, Rungrod A, Tachaleat A, Pongpattanakitshote S, Cheevadhanarak S, Tanticharoen M. 2005. Diversity of type I polyketide synthase genes in the wood-decay fungus *Xylaria* sp. Bcc 1067. FEMS Microbiol Lett. 251(1):125-136.

Arora NK, Verma M. 2017. Modified microplate method for rapid and efficient estimation of siderophore produced by bacteria. 3 Biotech. 7(6):381.

Audenaert K, Pattery T, Cornelis P, Höfte M. 2002. Induction of systemic resistance to *Botrytis cinerea* in tomato by *Pseudomonas aeruginosa* 7NSK2: Role of salicylic acid, pyochelin, and pyocyanin. Mol Plant-Microbe Interact. 15(11):1147-1156.

Bachmann BO, Ravel J. 2009. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. In: Hopwood DA, editor. Methods in Enzymology. Academic Press. p. 181-217.

Baumgartner K, Rizzo DM. 2002. Spread of *Armillaria* root disease in a California vineyard. Am J Enol Vitic. 53(3):197-203.

Blin K, Pascal Andreu V, de Los Santos ELC, Del Carratore F, Lee SY, Medema MH, Weber T. 2019a. The antiSMASH database version 2: A comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 47(D1):D625-d630.

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019b. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47(W1):W81-W87.

Brandenburger E, Gressler M, Leonhardt R, Lackner G, Habel A, Hertweck C, Brock M, Hoffmeister D. 2017. A highly conserved basidiomycete peptide synthetase produces a trimeric hydroxamate siderophore. Appl Environ Microbiol. 83(21):e01478-01417.

Butaitė E, Baumgartner M, Wyder S, Kümmerli R. 2017. Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. Nat Commun. 8(1):414.

Carrano CJ, Jordan M, Drechsel H, Schmid DG, Winkelmann G. 2001. Heterobactins: A new class of siderophores from *Rhodococcus erythropolis* IGTS8 containing both hydroxamate and catecholate donor groups. BioMetals. 14(2):119-125.

Chen L, Bóka B, Kedves O, Nagy VD, Szűcs A, Champramary S, Roszik R, Patocskai Z, Münsterkötter M, Huynh T. 2019. Towards the biological control of devastating forest pathogens from the genus *Armillaria*. Forests. 10(11):1013.

Coetzee MPA, Wingfield BD, Roux J, Crous PW, Denman S, Wingfield MJ. 2003. Discovery of two northern hemisphere *Armillaria* species on proteaceae in South Africa. Plant Pathol. 52(5):604-612.

Coetzee, M.P.A., Wingfield, B.D., Bloomer, P., Wingfield, M.J. 2005. Phylogenetic analyses of DNA sequences reveal species partitions amongst isolates of *Armillaria* from Africa. Mycol Res., 109(11), 1223-1234.

Devireddy LR, Hart DO, Goetz DH, Green MR. 2010. A mammalian siderophore synthesized by an enzyme with a bacterial homolog involved in enterobactin production Cell. 141(6):1006-1017.

Eisendle M, Oberegger H, Zadra I, Haas H. 2003. The siderophore system is essential for viability of *Aspergillus nidulans*: Functional analysis of two genes encoding L-ornithine N^5-monooxygenase (*sidA*) and a non-ribosomal peptide synthetase (*sidC*). Mol Microbiol. 49(2):359-375.

Elías-Román RD, Medel-Ortiz R, Alvarado-Rosales D, Hanna JW, Ross-Davis AL, Kim MS, Klopfenstein NB. 2018. *Armillaria mexicana*, a newly described species from Mexico. Mycologia. 110(2):347-360.

Engels B, Heinig U, Grothe T, Stadler M, Jennewein S. 2011. Cloning and characterization of an *Armillaria gallica* cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides. J Biol Chem. 286(9):6871-6878.

Evdokias G, Semper C, Mora-Ochomogo M, Di Falco M, Nguyen TTM, Savchenko A, Tsang A, Benoit-Gelber I. 2021. Identification of a novel biosynthetic gene cluster in *Aspergillus niger* using comparative genomics. J Fungi. 7(5):374.

Floudas D, Held BW, Riley R, Nagy LG, Koehler G, Ransdell AS, Younus H, Chow J, Chiniquy J, Lipzen A et al. 2015. Evolution of novel wood decay mechanisms in agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. Fungal Genet Biol. 76:78-92.

Gaitatzis N, Kunze B, Müller R. 2001. *In vitro* reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. Proc Natl Acad Sci U S A. 98(20):11136-11141.

Goering AW, McClure RA, Doroghazi JR, Albright JC, Haverland NA, Zhang Y, Ju K-S, Thomson RJ, Metcalf WW, Kelleher NL. 2016. Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. ACS Cent Sci. 2(2):99-108.

Gu S, Wei Z, Shao Z, Friman V-P, Cao K, Yang T, Kramer J, Wang X, Li M, Mei X et al. 2020. Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. Nat Microbiol. 5(8):1002-1010.

Gunde-Cimerman N, Cimerman A. 1995. *Pleurotus* fruiting bodies contain the inhibitor of 3-hydroxy-3-methylglutaryl-coenzyme a reductase—lovastatin. Exp Mycol. 19(1):1-6.

Haas H, Eisendle M, Turgeon BG. 2008. Siderophores in fungal physiology and virulence. Annu Rev Phytopathol. 46(1):149-187.

Iftime D, Kulik A, Härtner T, Rohrer S, Niedermeyer THJ, Stegmann E, Weber T, Wohlleben W. 2016. Identification and activation of novel biosynthetic gene clusters by genome mining in the kirromycin producer *Streptomyces collinus* tü 365. J Ind Microbiol Biotechnol. 43(2):277-291.

Johnson LJ, Koulman A, Christensen M, Lane GA, Fraser K, Forester N, Johnson RD, Bryan GT, Rasmussen S. 2013. An extracellular siderophore is required to maintain the mutualistic interaction of *Epichloë festucae* with *Lolium perenne*. PLoS Pathog. 9(5):e1003332.

Johnson LJ, Steringa M, Koulman A, Christensen M, Johnson RD, Voisey CR, Bryan G, Lamont I, Rasmussen S. 2007. Biosynthesis of an extracellular siderophore is essential for maintenance of mutualistic endophyte-grass symbioses. In: Popay AJ, Thom ER, editors. 6th International Symposium on Fungal Endophytes of Grasses. Dunedin, New Zealand: New Zealand Grasslands Association. p. 177-179.

Kadi N, Challis GL. 2009. Siderophore biosynthesis: A substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine. Methods in Enzymology. Academic Press. p. 431-457.

Koch RA, Wilson AW, Séné O, Henkel TW, Aime MC. 2017. Resolved phylogeny and biogeography of the root pathogen *Armillaria* and its gasteroid relative, *Guyanagaster*. BMC Evol Biol. 17(1):33.

Koch RA, Yoon GM, Aryal UK, Lail K, Amirebrahimi M, LaButti K, Lipzen A, Riley R, Barry K, Henrissat B et al. 2021. Symbiotic nitrogen fixation in the reproductive structures of a basidiomycete fungus. Curr Biol. 31(17):3905-3914.e3906.

König S, Romp E, Krauth V, Rühl M, Dörfer M, Liening S, Hofmann B, Häfner A-K, Steinhilber D, Karas M et al. 2019. Melleolides from honey mushroom inhibit 5-lipoxygenase via cys159. Cell Chem Biol. 26(1):60-70.e64.

Koulman A, Lee TV, Fraser K, Johnson L, Arcus V, Lott JS, Rasmussen S, Lane G. 2012. Identification of extracellular siderophores and a related peptide from the endophytic fungus *Epichloë festucae* in culture and endophyte-infected *Lolium perenne*. Phytochemistry. 75:128-139.

Kügler S, Cooper RE, Boessneck J, Küsel K, Wichard T. 2020. Rhizobactin b is the preferred siderophore by a novel *Pseudomonas* isolate to obtain iron from dissolved organic matter in peatlands. BioMetals. 33(6):415-433.

Kuo T-H, Yang C-T, Chang H-Y, Hsueh Y-P, Hsu C-C. 2020. Nematode-trapping fungi produce diverse metabolites during predator–prey interaction. Metabolites. 10(3):117.

Kurth C, Kage H, Nett M. 2016. Siderophores as molecular tools in medical and environmental applications. Org Biomol Chem. 14(35):8212-8227.

Le Govic Y, Papon N, Le Gal S, Bouchara JP, Vandeputte P. 2019. Non-ribosomal peptide synthetase gene clusters in the human pathogenic fungus scedosporium apiospermum. Front Microbiol. 10(2062):2062.

Le Govic Y, Papon N, Le Gal S, Lelièvre B, Bouchara J-P, Vandeputte P. 2018. Genomic organization and expression of iron metabolism genes in the emerging pathogenic mold *Scedosporium apiospermum*. Front Microbiol. 9(827).

Legrand P, Ghahari S, Guillaumin J-J. 1996. Occurrence of genets of *Armillaria* spp. in four mountain forests in Central France: The colonization strategy of *Armillaria ostoyae*. New Phytol. 133(2):321-332.

Macheleidt J, Mattern DJ, Fischer J, Netzker T, Weber J, Schroeckh V, Valiante V, Brakhage AA. 2016. Regulation and role of fungal secondary metabolites. Annual Review of Genetics. 50(1):371-392.

Maglangit F, Tong MH, Jaspars M, Kyeremeh K, Deng H. 2019. Legonoxamines a-b, two new hydroxamate siderophores from the soil bacterium, *Streptomyces* sp. Ma37. Tetrahedron Lett. 60(1):75-79.

May JJ, Wendrich TM, Marahiel MA. 2001. The dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. J Biol Chem. 276(10):7209-7217.

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39(Web Server issue):W339-346.

Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC et al. 2015. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol. 11(9):625-631.

Mesanza N, Patten CL, Iturritxa E. 2017. Distribution and characterization of *Armillaria* complex in atlantic forest ecosystems of Spain. Forests. 8(7):235.

Milagres AMF, Machuca A, Napoleão D. 1999. Detection of siderophore production from several fungi and bacteria by a modification of chrome azurol s (CAS) agar plate assay. J Microbiol Methods. 37(1):1-6.

Mukherjee PK, Hurley JF, Taylor JT, Puckhaber L, Lehner S, Druzhinina I, Schumacher R, Kenerley CM. 2018. Ferricrocin, the intracellular siderophore of *Trichoderma virens*, is involved in growth, conidiation, gliotoxin biosynthesis and induction of systemic resistance in maize. Biochem Biophys Res Commun. 505(2):606-611.

Munawar A, Marshall JW, Cox RJ, Bailey AM, Lazarus CM. 2013. Isolation and characterisation of a ferrirhodin synthetase gene from the sugarcane pathogen *Fusarium sacchari*. ChemBioChem. 14(3):388-394.

Narh Mensah DL, Duponnois R, Bourillon J, Gressent F, Prin Y. 2018. Biochemical characterization and efficacy of *Pleurotus*, *Lentinus* and *Ganoderma* parent and hybrid mushroom strains as biofertilizers of attapulgite for wheat and tomato growth. Biocatal Agric Biotechnol. 16:63-72.

Oide S, Moeder W, Krasnoff S, Gibson D, Haas H, Yoshioka K, Turgeon BG. 2006. *Nps6*, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes. Plant Cell. 18(10):2836-2853.

Paauw A, Leverstein-van Hall MA, van Kessel KP, Verhoef J, Fluit AC. 2009. Yersiniabactin reduces the respiratory oxidative stress response of innate immune cells. PLoS One. 4(12):e8240.

Patzer SI, Braun V. 2010. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of *Streptomyces* sp. Atcc 700974. J Bacteriol. 192(2):426-435.

Prospero S, Holdenrieder O, Rigling D. 2004. Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. Forest Pathol. 34(1):1-14.

Pusztahelyi T, Holb I, Pócsi I. 2015. Secondary metabolites in fungus-plant interactions. Front Plant Sci. 6:573.

Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support Vector Machines (TSVMs). Nucleic Acids Res. 33(18):5799-5808.

Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. 2011. NRPSpredictor2—a web server for predicting nrps adenylation domain specificity. Nucleic Acids Res. 39(suppl_2):W362-W367.

Ruiz-Dueñas FJ, Barrasa JM, Sánchez-García M, Camarero S, Miyauchi S, Serrano A, Linde D, Babiker R, Drula E, Ayuso-Fernández I et al. 2021. Genomic analysis enlightens Agaricales lifestyle evolution and increasing peroxidase diversity. Mol Biol Evol. 38(4):1428-1446.

Salwan R, Sharma V. 2020. Molecular and biotechnological aspects of secondary metabolites in Actinobacteria. Microbiol Res. 231:126374.

Sayari M, van der Nest MA, Steenkamp ET, Soal NC, Wilken PM, Wingfield BD. 2019. Distribution and evolution of nonribosomal peptide synthetase gene clusters in the Ceratocystidaceae. Genes. 10(5):328.

Schrettl M, Bignell E, Kragl C, Sabiha Y, Loss O, Eisendle M, Wallner A, Arst HN, Jr., Haynes K, Haas H. 2007. Distinct roles for intra- and extracellular siderophores during *Aspergillus fumigatus* infection. PLoS Pathog. 3(9):1195-1207.

Schrettl M, Kim HS, Eisendle M, Kragl C, Nierman WC, Heinekamp T, Werner ER, Jacobsen I, Illmer P, Yi H et al. 2008. Srea-mediated iron regulation in *Aspergillus fumigatus*. Mol Microbiol. 70(1):27-43.

Schwecke T, Göttling K, Durek P, Dueñas I, Käufer NF, Zock-Emmenthal S, Staub E, Neuhof T, Dieckmann R, von Döhren H. 2006. Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. ChemBioChem. 7(4):612-622.

Schwyn B, Neilands JB. 1987. Universal chemical assay for the detection and determination of siderophores. Anal Biochem. 160(1):47-56.

Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, Krizsán K, Kiss B, Hess J, Varga T, Slot J et al. 2017. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nat Ecol Evol. 1(12):1931-1941.

Stachelhaus T, Mootz HD, Marahiel MA. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 6(8):493-505.

Stecher G, Tamura K, Kumar S. 2020. Molecular evolutionary genetics analysis (mega) across computing platforms. Mol Biol Evol. 37(4):1237-1239.

Stubbendieck RM, Straight PD. 2016. Multifaceted interfaces of bacterial competition. J Bacteriol. 198(16):2145-2155.

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: A genome comparison visualizer. Bioinformatics. 27(7):1009-1010.

Taguchi F, Suzuki T, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y. 2010. The siderophore pyoverdine of *Pseudomonas syringae* pv. Tabaci 6605 is an intrinsic virulence factor in host tobacco infection. J Bacteriol. 192(1):117-126.

Tsykun T, Rigling D, Nikolaychuk V, Prospero S. 2012. Diversity and ecology of *Armillaria* species in virgin forests in the Ukrainian Carpathians. Mycol Progress. 11(2):403-414.

Vassaux A, Meunier L, Vandenbol M, Baurain D, Fickers P, Jacques P, Leclère V. 2019. Nonribosomal peptides in fungal cell factories: From genome mining to optimized heterologous production. Biotechnol Adv. 37(8):107449.

Wallner A, Blatzer M, Schrettl M, Sarg B, Lindner H, Haas H. 2009. Ferricrocin, a siderophore involved in intra- and transcellular iron distribution in *Aspergillus fumigatus*. Appl Environ Microbiol. 75(12):4194-4196.

Walsh CT. 2008. The chemical versatility of natural-product assembly lines. Acc Chem Res. 41(1):4-10.

Wang X, Zhang M, Loh B, Leptihn S, Ahmed T, Li B. 2021. A novel NRPS cluster, acquired by horizontal gene transfer from algae, regulates siderophore iron metabolism in *Burkholderia seminalis* r456. Int J Biol Macromol. 182:838-848.

Warwell MV, McDonald GI, Hanna JW, Kim M-S, Lalande BM, Stewart JE, Hudak AT, Klopfenstein NB. 2019. *Armillaria altimontana* is associated with healthy western white pine (*Pinus monticola*): Potential *in situ* biological control of the Armillaria root disease pathogen, *A. solidipes*. Forests. 10(4):294.

Welzel K, Eisfeld K, Antelo L, Anke T, Anke H. 2005. Characterization of the ferrichrome a biosynthetic gene cluster in the homobasidiomycete *Omphalotus olearius*. FEMS Microbiol Lett. 249(1):157-163.

Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus E-M, Espino JJ, Huß K, Michielse CB, Albermann S, Wagner D et al. 2013. Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. PLoS Pathog. 9(6):e1003475.

Winterberg B, Uhlmann S, Linne U, Lessing F, Marahiel MA, Eichhorn H, Kahmann R, Schirawski J. 2010. Elucidation of the complete ferrichrome a biosynthetic pathway in *Ustilago maydis*. Mol Microbiol. 75(5):1260-1271.

Yuan WM, Gentil GD, Budde AD, Leong SA. 2001. Characterization of the *Ustilago maydis* sid2 gene, encoding a multidomain peptide synthetase in the ferrichrome biosynthetic gene cluster. J Bacteriol. 183(13):4040-4051.

Zheng, D, Zeng Z, Xue B, Deng Y, Sun M, Tang Y-J, Ruan L. 2018. *Bacillus thuringiensis* produces the lipopeptide thumolycin to antagonize microbes and nematodes. Microbiol Res. 215: 22-28.

Tables

Table 1: Genome features of studied Physalacriaceae

Species	Strain	Code	Ploidy	Length (Mb)	# of Genes	Genome Source [a]
Armillaria borealis	FPL87.14	Armbor1	-	71.69	19,984	JGI
Armillaria cepistipes	B5	Armcep1	Haploid	75.50	23,460	JGI (Sipos et al. 2017)
Armillaria fumosa	CBS 122221	Armfum1	-	55.82	16,095	JGI
Armillaria mellea	ELDO17	Armme1	-	70.86	15,646	JGI
Armillaria nabsnona	CMW6904	Armnabs1	Haploid	62.72	19,015	JGI
Armillaria novae-zelandiae	2840	Armnov1	Diploid	79.33	18,551	JGI
Desarmillaria ectypa	FPL83.16	(Des)Armect1	-	40.60	12,228	JGI
Desarmillaria tabescens	CCBAS 213	(Des)Armtab1	-	74.88	19,032	JGI
Cylindrobasidium torrendii	FP15055 v1.0	Cylto1	Haploid	31.57	13,940	JGI (Floudas et al. 2015)
Guyanagaster necrorhizus	MCA 3950 v1.0	Guyne1	-	53.69	14,276	JGI and GenBank (accession: JAEACO000000000.1) (Koch et al. 2021)
Oudemansiella mucida	CBS 558.79 v1.0	Oudmuc1	-	61.73	18,562	JGI (Ruiz-Dueñas et al. 2021)

[a]: All genomes citing JGI without a reference or made public after November 2018 were used with permission from Dr. László G. Nagy

NB: All the studied genomes had RNA sequences available (JGI and respective references)

Table 2: *Armillaria* spp. lifestyle, culture codes and source information

Species	Lifestyle [a]	CMW number	Host	Original source
A. fuscipes	Facultative necrotroph/ pathogenic	CMW2740	*Pinus elliottii*	Entabeni, South Africa
		CMW3164	*Pelargonium asperum*	La Réunion
A. gallica	Facultative necrotroph/ weakly pathogenic	CMW31092	N/A	Veneto, Bellune, Italy
		CMW45397	Duff	Minnesota, USA
A. luteobubalina	Facultative necrotroph/ pathogenic	CMW4974	N/A	Australia
A. mellea	Facultative necrotroph/ highly pathogenic	CMW31132	*Ailanthus altissima*	China
A. nabsnona	Weakly pathogenic	CMW3159	*Acer macrophyllum*	Vancouver, Canada
		CMW6904	*Acer circinatum*	USA
Armillaria sp. ACB	N/A	CMW4456	*Brachystegia utilis*	Stapleford, Zimbabwe

[a] = Lifestyle information are summarized from Koch et al. (2017)

ACB = African Clade B in Coetzee et al. (2005); N/A = Not available

Table 3: Gene characteristics of putative NRPS-dependent siderophore synthetase of studied Physalacriaceae genomes and similarity to characterized NRPS-dependent siderophore synthetases

ProteinID [a]	Location in genome [b]	Strand	Size [c]	# of introns	Query cover and Percentage identity [d] (%)				
					Fer3 (4830 aa) *Ustilago maydis*	**FSN1 (4707 aa)** *Fusarium sacchari*	**NRPS2 (4339 aa)** *Wallemia ichthyophaga*	**SidC (4759 aa)** *Aspergillus fischeri*	**SidN (4690 aa)** *Epichloë festucae*
Armbor1_1740653	cds153656_169842 (68.1)	-	16,186 (4329)	52	95 (31.2)	96 (34.5)	95 (41.6)	96 (31.5)	94 (26.5)
Armcep1_12756	cds5429960_5446141 (261.7)	-	16,181 (4530)	50	96 (29.5)	97 (32.6)	96 (42.9)	97 (33.5)	95 (28.3)
Armfum1_1505943	cds4583838_4599965 (1.5)	-	16,126 (4521)	50	97 (29)	97 (31.7)	96 (42.2)	96 (32.9)	96 (26.8)
Armme1_963312	cds326940_343118 (13.1)	+	16,177 (4533)	50	96 (29.9)	97 (32.5)	96 (42.6)	96 (33.4)	95 (28.2)
Armnabs1_1027822	cds577121_593305 (22.1)	-	16,183 (4539)	50	96 (29.7)	97 (32.5)	96 (42.9)	97 (33.4)	95 (28.2)
Armnov1_1625296	cds756488_772706 (11.1)	+	16,217 (4244)	57	82 (31.3)	83 (33.1)	88 (44.3)	88 (34)	81 (27.9)
(Des)Armect1_1380706	cds3727079_3743248 (3.1)	+	16,168 (4533)	50	96 (29.3)	96 (32.5)	95 (42.3)	95 (33.1)	95 (28.2)
(Des)Armtab1_1636500	cds293769_309930 (4.1)	-	16,160 (4551)	49	96 (29.7)	96 (32.7)	96 (42.6)	96 (33.3)	95 (28.6)
Cylto1_490369	cds24790_40480 (88.1)	-	15,689 (4485)	45	95 (29.8)	95 (32.3)	95 (40.5)	95 (33.2)	97 (28.5)
Guyne1_959764	cds765973_782126 (2.1)	-	16,152 (4527)	50	97 (30)	97 (32.3)	96 (42.1)	97 (32.8)	96 (28.1)
Oudmuc1_1435297	cds135154_154178 (46.1)	+	19,023 (5353)	58	81 (30.1)	81 (32.6)	80 (41.3)	81 (33.6)	80 (28.8)

[a]: Genome code followed by ProteinId from JGI.

[b]: Presented as coding sequence number (Scaffold number followed by the Secondary metabolite cluster number located on the scaffold, after the decimal point, as determined by fungiSMASH)

[c]: Presented as number of base pairs (bp) of DNA sequence (number of amino acids (aa) in translated protein sequence)

[d]: Matches of FerrichromeA synthetase (Fer3), Ferrirhodin synthetase (FSN1), Nonribosomal peptide synthetase 2 (NRPS2), Ferricrocin synthetase (SidC), and Epichloënin A synthetase (SidN) with accession numbers XM_011389001.1, JN997437.1, XM_009271380.1, XM_001264036.1, and JN132403.1 respectively, obtained from tBLASTn analyses. The species within which proteins are found are indicated. Sizes of the respective proteins are presented in brackets and the organisms in which the proteins are reported are placed below the codes. The NRPS2 protein has not been characterized but is the closest match (in terms of percentage identity) in all the genomes studied. Data presented as Query cover (Percentage identity).

Table 4: Comparison of modular organization of NRPS genes of Physalacriaceae to NRPS genes of other fungi

Protein ID a	Modular organization b		Reference(s)
	fungiSMASH	InterProScan and Pfam	
Armbor1_1740653			This study
Armcep1_12756			This study
Armfum1_1505943			This study
Armme1_963312			This study
Armnabs1_1027822			This study
Armnov1_1625296			This study
(Des)Armect1_1380706			This study
(Des)Armtab1_1636500			This study
Cylto1_490369			This study
Guyne1_959764			This study
Oudmuc1_1435297			This study
EAW22140 (SidC)	N/A		This study
EAL91050 (SidC)	N/A		Schwecke et al. (2006)
AAP56239 (SidC)	N/A		Schwecke et al. (2006)
AET13875.1 (SidN)	N/A		Johnson et al. (2007), This study
AET13879.1 (SidN)	N/A		Johnson et al. (2007), This study
AFD36451.1 (FSN1)	N/A		Munawar et al. (2013), This study
AAX49356.1 (Fso1)	N/A		(Schwecke et al., 2006; Welzel et al., 2005)
KIS71541.1 (Fer3)	N/A		(Schwecke et al., 2006; Winterberg et al., 2010), This study
EOQ99429.1 (NRPS2)	N/A		This study

[a]: Protein identity for *Armillaria* spp. and other Physalacriaceae species. NCBI accession numbers for protein sequences of NRPS of other species are provided. The names of the genes are placed in brackets. EAL91050, AAP56239, AET13875.1, AFD36451.1, AAX49356.1, and KIS71541.1 are found within the genomes of *As. fumigatus* Af293, *As. nidulans*, *E.*

festucae strain E2368, *F. sacchari*, *Om. olearius*, and *U. maydis* 521, respectively. These proteins synthesize the siderophores, Ferricrocin, Ferricrocin, Epichloënin A, Ferrirhodin, Ferrichrome A, and Ferrichrome A, respectively. The secondary metabolites synthesized by EAW22140 of *As. fischeri* NRRL 181, AET13879.1 of *E. festucae* strain F11, and EOQ99429.1 of *W. ichthyophaga* EXF-994, have not yet been characterized.

[b]: Modular organization is domain architecture based on fungiSMASH and Pfam searches for all Physalacriaceae species and on Pfam searches for all other NRPSs. A = adenylation domain (also known as AMP-binding enzyme), T = thiolation domain (also known as Phosphopantetheine attachment site; PCP), C = condensation domain. For fungiSMASH-based modular organization, all blank spaces indicate absence of a domain. For InterProScan and Pfam-based modular organization, all domains marked by solid fill with white font colour, and solid fill with black font colour had significant and insignificant matches in the Pfam database, respectively. Domains marked with texture fill with black font colour were AMP-binding enzyme C-terminal domain. All the NRPS fall under the Type II structural type according to the classification system described by Schwecke et al. (2006). N/A = not applicable.

Figures

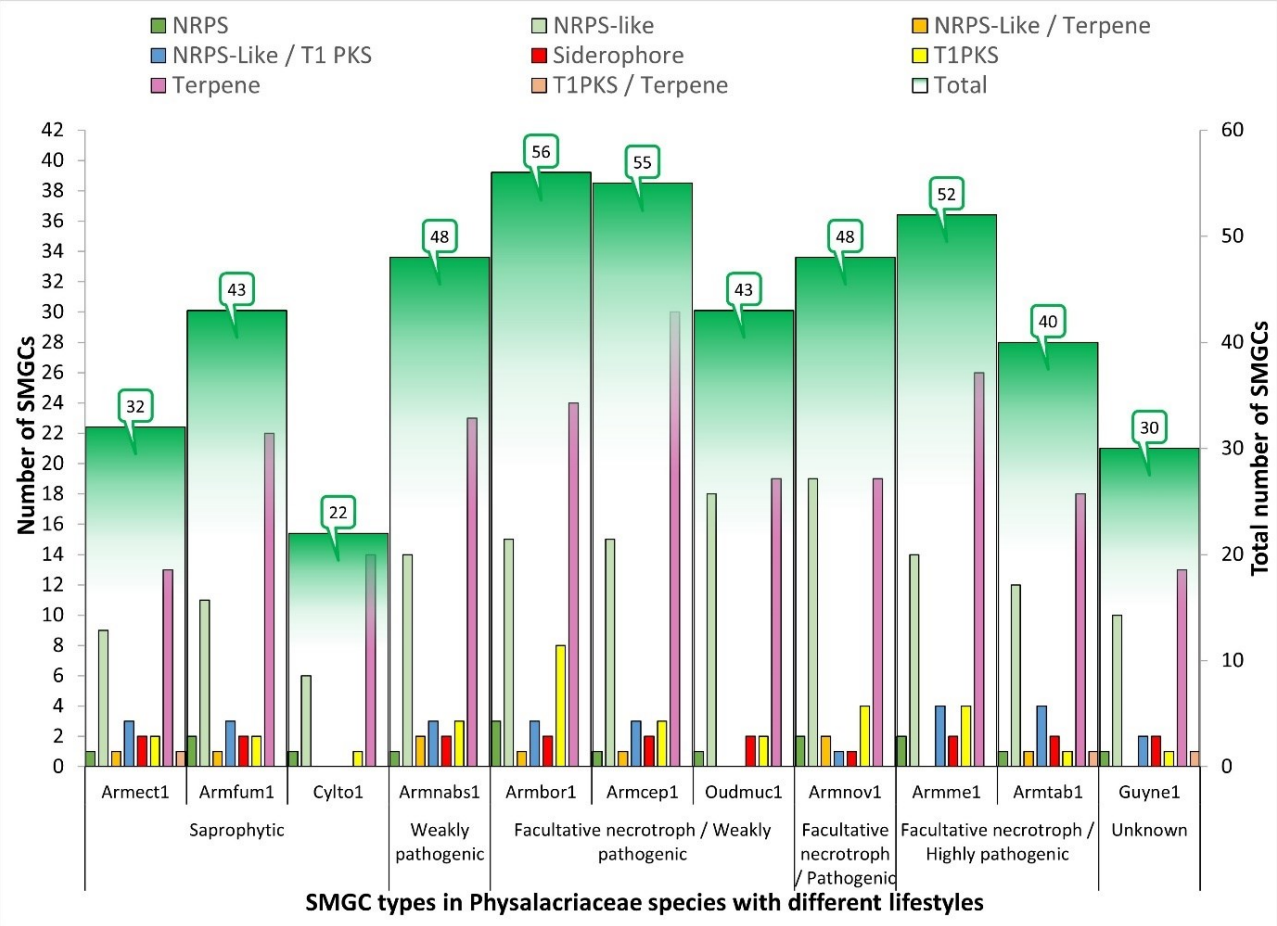


Fig. 1 SMGC diversity and distribution in genomes of the studied *Armillaria* spp. and other Physalacriaceae with different lifestyles.

Different classes or types of SMGCs, as predicted by fungiSMASH are presented by different colours. NRPS = Nonribosomal polypeptide synthetase, T1PKS = Type1 polyketide synthetase, Siderophores = suspected NIS synthetase clusters. NRPS-Like / Terpene, NRPS-Like / T1PKS, T1PKS/Terpene are predicted hybrid clusters. All genomes studied contained at least 22 SMGCs that may produce diverse compounds. Oudmuc1 contained 1 SMGC predicted to produce indole (Data not shown). Lifestyle information are summarized from Koch et al. (2017).

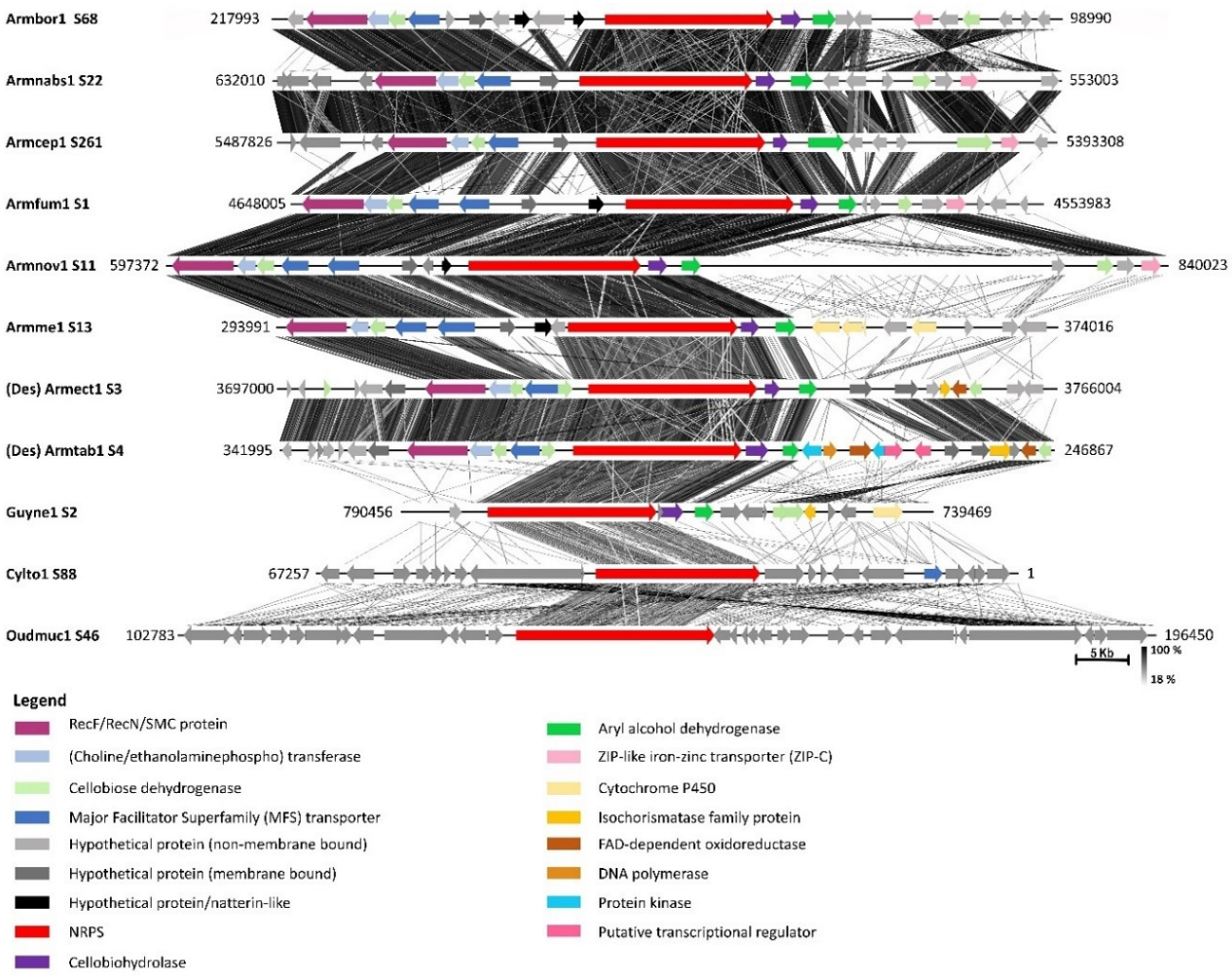


Fig. 2 NDSS gene cluster synteny map in annotated genomes of Physalacriaceae.
Numbers following the species code are the scaffolds on which the clusters are located. Different colours and orientation of arrows represent different putative proteins as determined by tBLASTn searches and direction of transcription respectively. Orthologous genes are identically coloured. All genes coloured grey in the clusters of Cylto1 and Oudmuc1 are not orthologous to the genes in the clusters of the other members of the Physalacriaceae. Darker shades of lines between clusters represent higher amino acid similarity between the different clusters based on tBLASTx. Numbers at the ends of the clusters are the locations on the scaffolds.

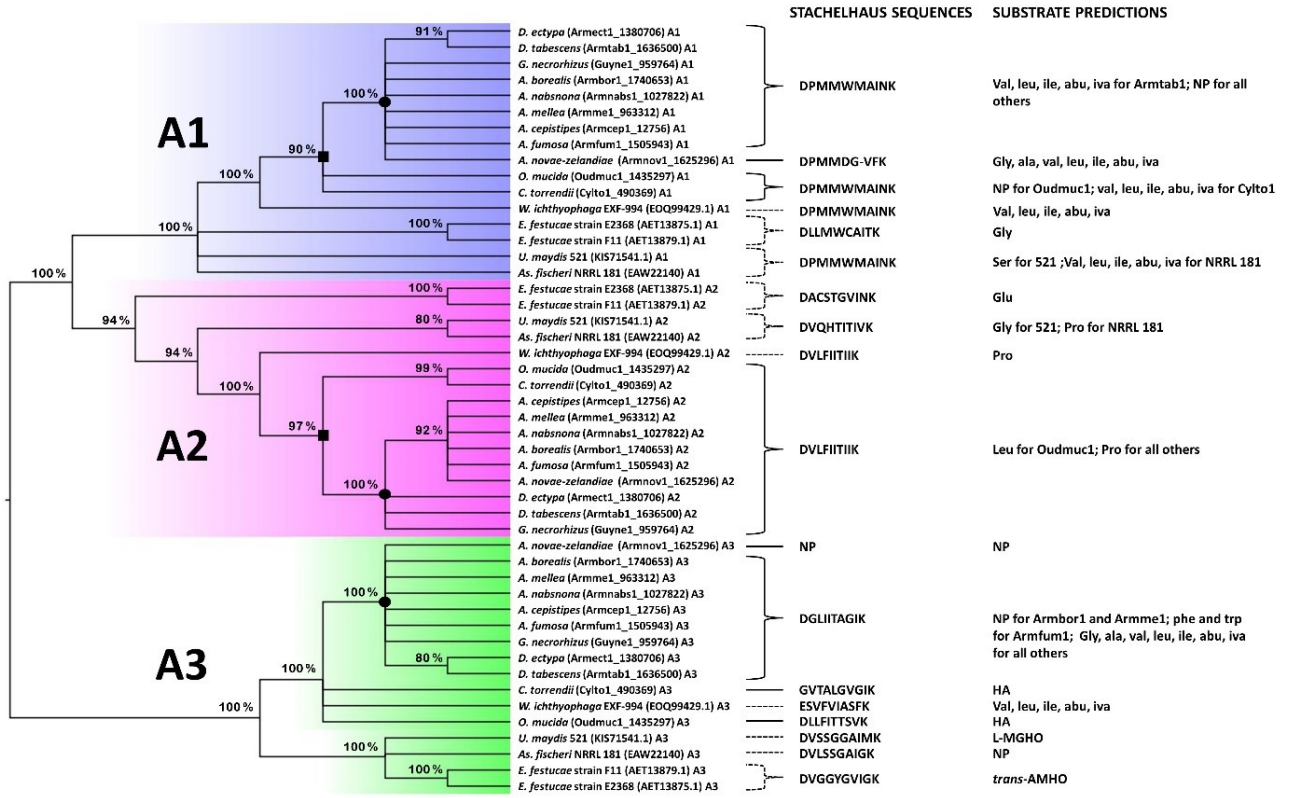


Fig. 3 Phylogenetic tree and substrate predictions of A1 – A3 domains of putative NDSSs in *Armillaria* and other Physalacriaceae genomes in comparison with *NRPS2* of *W. ichthyophaga* and other characterized NDSSs in other fungi.

Bootstrap support values greater than 70 % are shown above each node. Species names of putative NDSSs in *Armillaria* and other Physalacriaceae genomes are shown followed by ProteinIDs in brackets and the A-domain numbers. All other NDSSs are presented as species (and strain) followed by NCBI accession numbers in bracket and the A-domain numbers. Coloured areas are A1, A2, and A3 domains of NDSSs respectively. Nodes marked with black squares highlight the nodes which group the Physalacriaceae. Nodes marked with black ovals highlight the *Armillaria* spp., *Desarmillaria* spp. and *G. necrorhizus* clade of each A-domain. Solid brackets and lines refer to Physalacriaceae species. Dashes refer to other fungal species. All Stachelhaus sequences and (predicted) substrates of all characterized NRPS-dependent siderophore synthetases from other fungi except that of *As. fischeri* NRRL 181 were obtained from various authors (Johnson et al. 2007; Schwecke et al. 2006; Welzel et al. 2005; Winterberg et al. 2010). NP = None predicted, HA = hydrophobic-aliphatic amino acids, L-MGO = N^5-*trans*-(α-methyl)-glutaconyl-N^5-hydroxy-L-ornithine, and *trans*-AMHO = N^5-*trans*-anhydromevalonyl-N^5-hydroxyornithine.

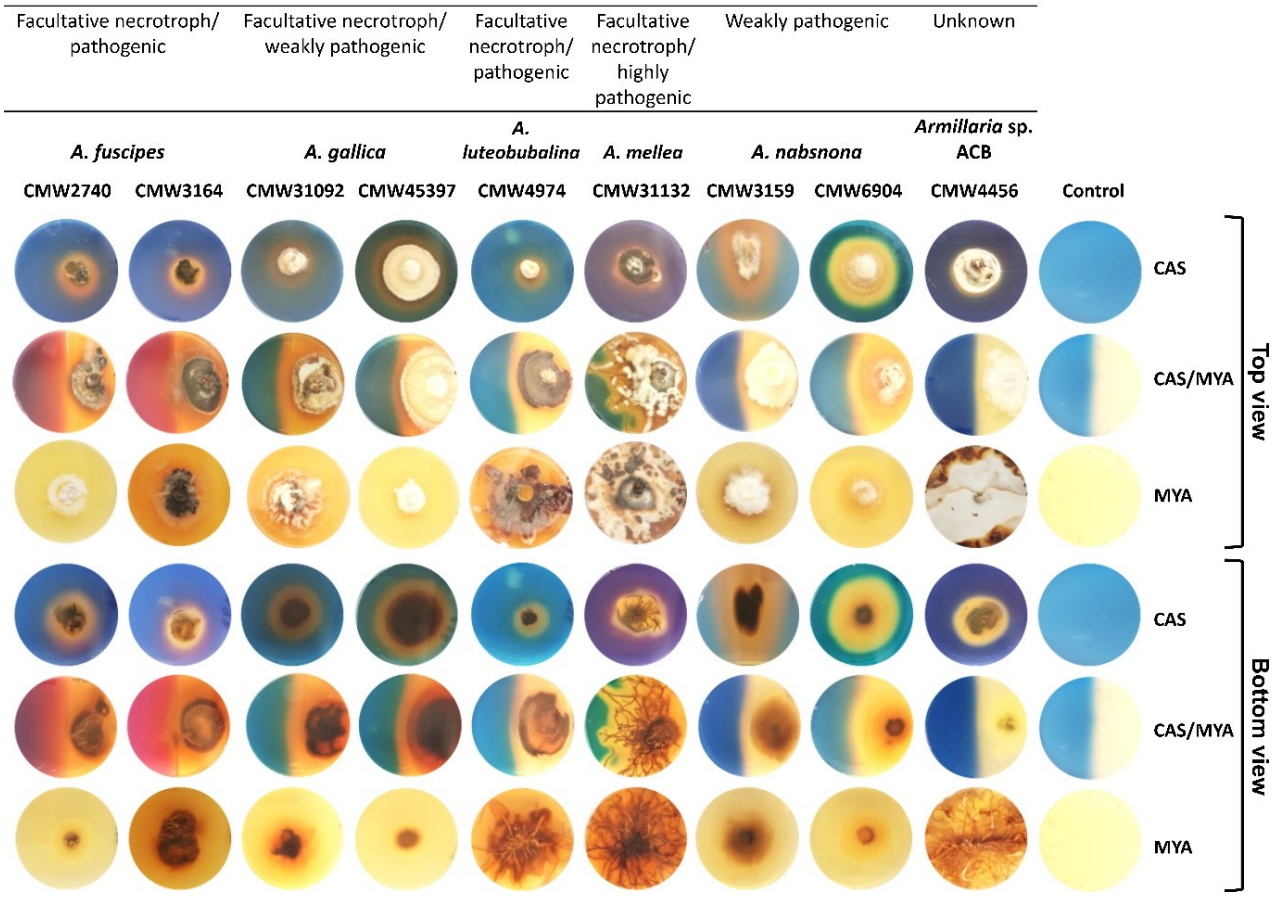


Fig. 4 Representative plates showing siderophore biosynthesis and mycelia and rhizomorph growth of various species and strains of *Armillaria*.

Top and bottom views of representative plates of *Armillaria* spp. on CAS, CAS/MYA, and MYA on 45th day of incubation. Note the colour change of the blue CAS media to purplish-red (e.g., CMW3164 on CAS/MYA), purple (e.g., CMW4456 on CAS), reddish-orange (e.g., CMW3159 on CAS), and/or orange (e.g., CMW6904 on CAS) mostly beyond the margins of the mycelia and/or rhizomorphs indicative of production of different types of (extracellular) siderophores. Control plates are the respective uninoculated media. ACB = African Clade B.

Supplementary information

The online publication of this Chapter contains the supplementary materials available at https://doi.org/10.1007/s00294-022-01256-w and can be found below:

Table S1: Positions of NRPS clusters as predicted by fungiSMASH

Species	Scaffold number	Position (nucleotide)
A. borealis	68	133,657 – 189,842
A. cepistipes	261	5,409,961 – 5,466,141
A. fumosa	1	4,563,839 – 4,619,965
A. mellea	13	306,941 – 363,118
A. nabsnona	22	557,122 – 613,305
A. novae-zelandiae	11	736,489 – 792,706
D. ectypa	3	3,707,080 – 3,763,248
D. tabescens	4	273,770 – 329,930
G. necrorhizus	2	745,974 – 802,126
C. torrendii	88	4,791 – 60,480
O. mucida	46	99,521 – 174,178

Table S2: Information on putative gene annotation of genes flanking NRPS-dependent siderophore synthetase gene based on InterProScan and tBLASTn searches

Gene	Size (aa)	PFAM domain(s)	Gene ontology (GO)	MB	Selected ortholog accession number(s)
Aryl alcohol dehydrogenase	388-855	PF00248	GO:0047834	N	XM_007386508.1, XM_007306800.1, XM_001830525.1
Cellobiohydrolase	429-454	PF00840	GO:0004553, GO:0005975	Y	MK313723.1, XM_007269301.1, LC034188.1
Cellobiose dehydrogenase	236-348	PF16010 PF13561, PF01738	- GO:0016491, GO:0055114, GO:0016787	Y, N	XM_018327719.1, XM_031221745.1
(Choline/Ethanolaminephospho) transferase	146-419	PF01066	GO:0008654, GO:0016020, GO:0016780	Y, N	XM_008034254.1, XM_001886787.1
Cytochrome P450	104-532	PF00067, PF02936	GO:0005506, GO:0016705, GO:0020037, GO:0055114, GO:0006123, GO:0005751	Y	XM_008042239.1, XM_001835070.1, XM_027753339.1, XM_007334100
DNA polymerase	158	PF00136	GO:0000166, GO:0003677	N	XM_001831569.2, XM_008040078.1
FAD dependent oxidoreductase	409	PF01266	GO:0016491, GO:0055114	N	XM_007299194.1, XM_025697947.1
Hypothetical protein (non-MB)	146-614	PF12937, PF00651	GO:0005515	N	XM_001883462.1, XM_001873677.1, XM_001879861.1,
Hypothetical protein (MB)	180-392	PF12051	-	Y	XM_008039196.1, XM_007330054.1, XM_001887233.1
Hypothetical protein/natterin-like	288-352	PF03318	-	N	XM_001831713.2, XM_001883450.1,
Isochorismatase family protein	302	PF13532	GO:0006281, GO:0006307, GO:0016491, GO:0035552, GO:0051213	N	XM_001271399.1
MFS	506-578	PF07690	GO:0022857, GO:0055085	Y	XM_032084181.1, XM_008046394.1, XM_033819763.1
NRPS	4244-5353	PF00501, PF13193, PF00550, PF00668	GO:0031177, GO:0003824	N	XM_011389001.1, JN997437.1, XM_009271380.1, XM_001264036.1, JN132403.1
Protein kinase	249-380	PF00069	GO:0004672, GO:0005524, GO:0006468	N	XM_007782407.1, XM_033600130.1, XM_031224444.1
Putative transcriptional regulator	391	-	-	N	LR728655.1
RecF/RecN/SMC protein	1575-1577	PF02463, PF06470	GO:0005515, GO:0005524, GO:0005694, GO:0051276	N	XM_022012713.1, XM_025742723.1, XM_019170948.1
ZIP-C	449-476	PF02535	GO:0016020, GO:0030001, GO:0046873, GO:0055085	Y	XM_001887139.1, XM_007365303.1, XM_007305919.1, XM_008046135.1

aa = amino acids; MB = membrane bound; Y = yes; N = no

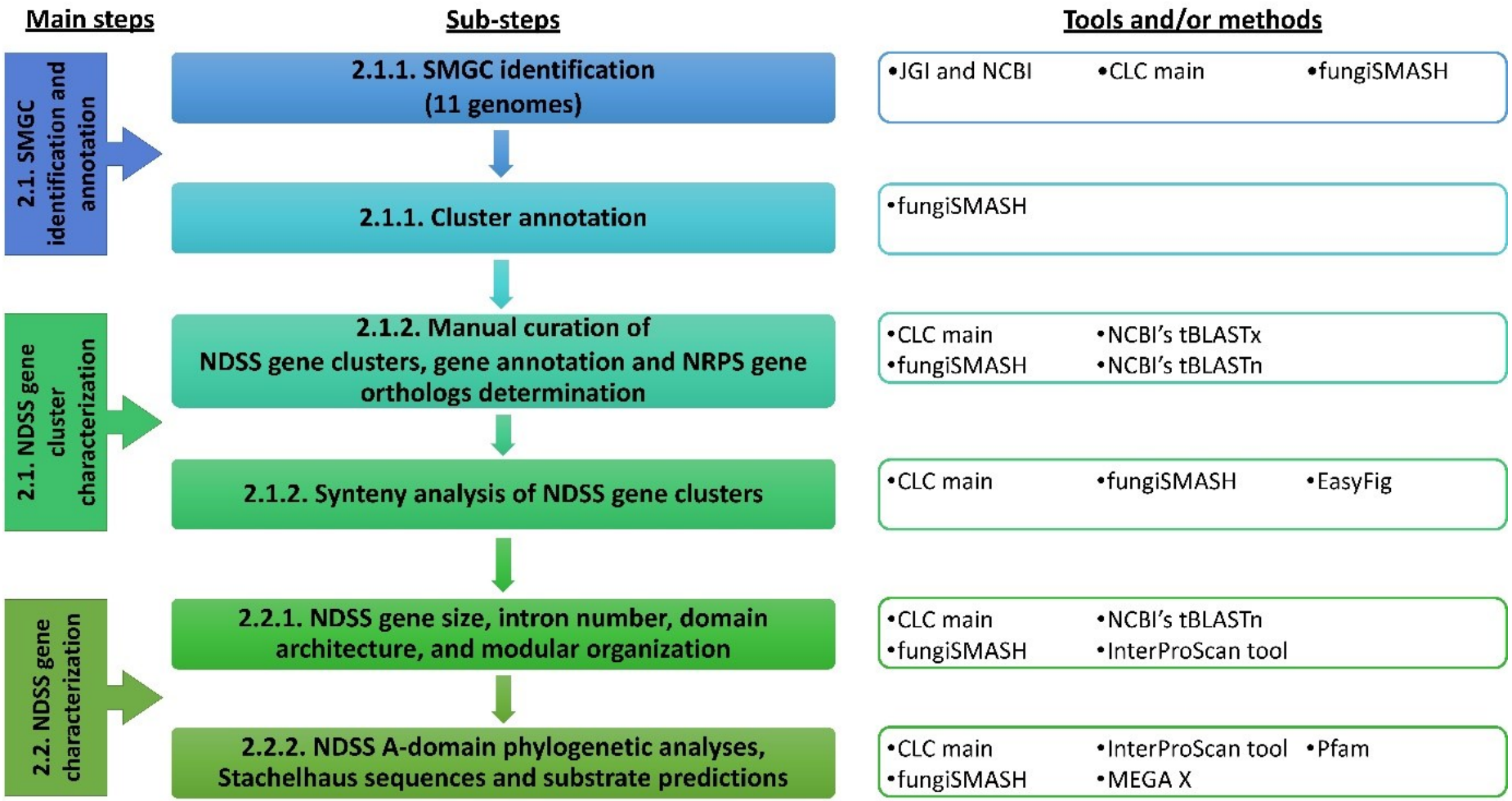


Fig. S1 Flowchart of *in silico* analyses methods and bioinformatic tools used in the study.

Acknowledgments

We are grateful to Dr. László G. Nagy and his team at Biological Research Centre, Synthetic and Systems Biology Unit, 6726 Szeged, Hungary for granting us permission to use the genome and RNA sequence data of *A. borealis* strain FPL87.14, *A. fumosa* strain CBS 122221, *A. mellea* strain ELDO17, *A. nabsnona* strain CMW6904, *A. novae-zelandiae* strain 2840, *D. ectypa* strain FPL83.16 and *D. tabescens* strain CCBAS 213. We also thank all the funding bodies for financial assistance to carry out this work.

Data availability

Publicly available genome and RNA sequences were analysed in this study. These data can be found at https://mycocosm.jgi.doe.gov/mycocosm/species-tree/tree;l2higw?organism=physalacriaceae. Unpublished genome and RNA sequence data obtained from JGI were used with permission from Dr. Laszlo G. Nagy. Data of bioassays are presented in this manuscript and can be assessed in the manuscript.

CHAPTER 3

TWO DISTINCT NRPS-INDEPENDENT SIDEROPHORE SYNTHETASE GENE CLUSTERS IDENTIFIED IN *ARMILLARIA* AND OTHER SPECIES IN THE PHYSALACRIACEAE

This chapter has been submitted to Genes, Genomes, Genetics as:

Narh Mensah, D.L., Wingfield, B.D., Coetzee, M.P.A. 2023. Two distinct NRPS-independent siderophore synthetase gene clusters identified in *Armillaria* and other species in the Physalacriaceae. *Genes, Genomes, Genetics*

Abstract

Siderophores are important for ferric iron solubilization, sequestration, transportation, and storage, especially under iron-limiting conditions such as aerobic conditions at high pH. Siderophores are mainly produced by non-ribosomal peptide synthetases (NRPS)-dependent siderophore pathway, NRPS-independent siderophore (NIS) synthetase pathway or the hybrid NRPS/NIS pathway. Outcompeting or inhibition of plant pathogens, alteration of host defence mechanisms, and alteration of plant-fungal interactions have been associated with fungal siderophores. To understand these mechanisms in fungi, studies have been conducted on siderophore biosynthesis by ascomycetes with limited focus on the basidiomycetes. *Armillaria* includes several species that are pathogens of woody plants and trees important to agriculture, horticulture, and forestry. The aim of this study was to investigate the presence of NIS synthetase gene cluster(s) in genomes of *Armillaria* species using a comparative genomics approach. Iron-dependent growth and siderophore biosynthesis in strains of selected *Armillaria* spp. was also evaluated *in vitro.* Two distinct NIS synthetase gene clusters were identified in all the genomes. All NIS synthetase genes identified putatively encode Type A' NIS synthetases, most of which have IucA_IucC and FhuF-like transporter domains at their N- and C-terminals respectively. The effect of iron on culture growth varied among the strains studied. Bioassays using the CAS assay on selected *Armillaria* spp. revealed *in vitro* siderophore biosynthesis by all strains irrespective of added $FeCl_3$ concentration. This study highlights some of the tools that *Armillaria* species allocate to iron homeostasis. The information generated from this study may in future aid in developing molecular based methods to control these phytopathogens.

Keywords: *Armillaria*; secondary metabolite gene clusters; ferric iron uptake; NIS synthetase; phytopathogens

1. Introduction

All eukaryotes, and most prokaryotes, utilize iron as an essential micronutrient for various functions. However, iron is usually not bioavailable for use by these organisms under aerobic conditions at neutral pH (Chipperfield and Ratledge, 2000). Hence, organisms have developed various mechanisms for ferric and ferrous iron uptake. One of the important mechanisms of ferric iron (Fe^{3+}) uptake, storage and transport used by various organisms is through synthesis and/or use of the low molecular weight (usually <1 kDa) secondary/specialized metabolites, known as siderophores (Ghosh *et al.*, 2020).

Biosynthesis and secretion of siderophores occur mainly through the non-ribosomal peptide synthetases (NRPS)-dependent siderophore pathway or NRPS-independent siderophore (NIS) synthetase pathway (Brandenburger *et al.*, 2017; Kadi and Challis, 2009). A third pathway for siderophore biosynthesis and secretion is the hybrid NRPS/NIS pathway, which utilizes biosynthesis enzymes belonging to both of the main pathways (Lee *et al.*, 2007). Genes, including the main biosynthetic genes (the backbone genes), involved in these pathways are contained in biosynthetic gene clusters (BGCs) or operons.

The main biosynthetic genes in the clusters code for enzymes that are responsible for the actual biosynthesis of the respective secondary metabolite. NRPS, the protein encoded by the backbone gene in the NRPS-dependent pathway, employs discrete functional units with specific essential catalytic domains (Brandenburger *et al.*, 2017). As the name suggests, NIS synthetases are the proteins encoded by the backbone biosynthetic gene of the NIS synthetase pathway. Some siderophores such as petrobactin (formerly anthrachelin) are biosynthesized via the hybrid NRPS/NIS pathway. This pathway is composed of BGCs or operons which have genes encoding both an NRPS-dependent siderophore synthetase and one or more NIS synthetase as the backbone genes (Lee *et al.*, 2007; Nusca *et al.*, 2012). Other genes in the BGCs involved in all of these pathways function in substrate modification, cluster regulation, and product transport (Brandenburger *et al.*, 2017; Dixon *et al.*, 2012; Iftime *et al.*, 2016; Jarmusch *et al.*, 2021; Le Govic *et al.*, 2018).

Unlike NRPSs, NIS synthetases act individually and use free intermediates as substrates (Oves-Costales *et al.*, 2009). NIS synthetases are classified into three major superfamilies: type A, B and C NIS synthetases, based on their substrate specificity for polyamines or amino alcohols as well as the substrate activated for condensation, (Oves-Costales *et al.*, 2009). Type A synthetases utilize citric acid, Type B synthetases utilize α-ketoglutaric acid, and type C synthetases utilize monoamide/monoester derivatives of citric acid or monohydroxamate derivatives of succinic acid as their respective preferred carboxylic acid substrates (Oves-Costales *et al.*, 2009). The type A NIS

synthetases condense their substrates with various amines and alcohols, while the type B NIS synthetases condense their substrates with only amines. Type C NIS synthetases condense simple mono-amides or amines with a citryl or succinyl intermediate. A sub-category of the Type A NIS synthetases known as Type A' NIS synthetases specifically catalyse condensation of citric acid with amines (Kadi and Challis, 2009). Additionally, a sub-category of the Type C NIS synthetases are known as Type C' NIS synthetases (Carroll and Moore, 2018). These enzymes mainly condense more complex citryl or succinyl intermediates but may also perform macro-cyclization to fully form a siderophore (Carroll and Moore, 2018).

Genomes of some organisms contain different BGCs or operons belonging to the two main siderophore biosynthesis pathways (Carmichael *et al.*, 2019; Cheung *et al.*, 2009; Cotton *et al.*, 2009; Franza *et al.*, 2005; Mahe *et al.*, 1995). Alternatively, the clusters could belong to one of the two main biosynthesis pathways and one hybrid NRPS/NIS pathway as reported in *Bacillus* spp. (Koppisch *et al.*, 2005; Oves-Costales *et al.*, 2007; Wilson *et al.*, 2006). These BGCs or operons synthesize and transport different siderophores which have various functions such as growth promotion, growth characteristics (e.g. conidiation, dimorphism transition and pigmented microsclerotium formation), pathogenicity, virulence and resistance to oxidative stress (Franza *et al.*, 2005; Koppisch *et al.*, 2005; Li *et al.*, 2016). The different siderophores are synthesized sequentially (Franza *et al.*, 2005; Koppisch *et al.*, 2005; Li *et al.*, 2016).

Studies investigating the molecular and biochemical bases of siderophore biosynthesis, secretion and uptake have advanced our understanding of these molecules. Minimal research that focus attention on the role of siderophores in fungi have been conducted (Carroll and Moore, 2018). Among fungi, studies in this area have largely been focused on the Ascomycota (Haas *et al.*, 2008). The first ever characterized fungal NIS synthetase, Rfs, which is involved in rhizoferrin (polyhydroxycarboxylamide siderophore) biosynthesis in the opportunistic human pathogen, *Rhizopus delemar*, was reported by Carroll *et al.* (2017). Rhizoferrin biosynthesis in *R. delemar* requires only Rfs (Carroll *et al.*, 2017; Ramakrishnan *et al.*, 2019). In contrast, rhizoferrin biosynthesis by the bacterial animal pathogen, *Francisella tularensis,* requires an NIS synthetase (FslA) and a pyridoxal phosphate-dependent decarboxylase (FslC) (Ramakrishnan *et al.*, 2019).

Considering the importance of iron homeostasis for diverse functions in various organisms, the wide knowledge gap in siderophore biosynthesis, uptake and utilization mechanisms in the Basidiomycota needs to be narrowed. Recently we reported on secondary metabolite gene clusters (SMGCs) in the genomes of various *Armillaria* spp. and other members of the family Physalacriaceae, with an emphasis on NRPS-dependent siderophore synthetase gene clusters (Narh Mensah *et al.*, 2023). This

study revealed that the genomes investigated, except for *Cylindrobasidium torrendii*, contained SMGCs classified as siderophore gene clusters (SGCs) (Narh Mensah *et al.*, 2023).

The aim of the present study was to further investigate the SGCs in the genomes of *Armillaria* spp. and other members of the Physalacriaceae. These SGCs were earlier suggested by Narh Mensah *et al.* (2023) to produce siderophores through the NIS pathway, but this was not studied further. For this purpose, the SGCs were explored using a similar comparative genomics approach to the study reported by Narh Mensah *et al.* (2023). Bioassays were also conducted to assess the iron-dependent growth and siderophore biosynthesis by strains of *Armillaria* spp. This knowledge is essential for gaining further understanding of the genomic basis of the molecular and cellular mechanisms underlying iron homeostasis in *Armillaria* spp. aimed at developing more effective control strategies of these phytopathogens in future.

2. Materials and methods

2.1. *In silico* identification, annotation, and characterization of NIS synthetase gene clusters

2.1.1. Cluster identification, annotation, and synteny analyses

Source information of the genomes studied are presented in Table S1. Genome mining, genome walking, and gene annotation of selected members of the Physalacriaceae followed that described by Narh Mensah *et al.* (2023) with some modification. Cluster boundaries were predicted with Cluster Assignment by Islands of Sites (CASSIS) algorithm (Wolf *et al.*, 2016) implemented in Antibiotics & Secondary Metabolite Analysis SHell (antiSMASH) using the fungal version (fungiSMASH) v 6.1.1 (Blin *et al.*, 2021). Manual identification of NIS synthetase gene clusters, which may have been missed by fungiSMASH in the respective genomes, was conducted by BLAST searches using CLC Main Workbench v21.0.4 (www.qiagenbioinformatics.com). Nucleotide sequences of NIS synthetase genes identified in the clusters of some of the genomes were used as search terms for this purpose. Gene prediction of the manually detected gene cluster was achieved with AUGUSTUS v. 3.3.3 (Stanke *et al.*, 2008) (https://bioinf.uni-greifswald.de/augustus/submission.php). Flanking genes, their annotations and characteristics were determined and gene cluster synteny maps were generated as previously described (Narh Mensah *et al.*, 2023).

2.1.2. Comparison to other BGCs

Similarity of the detected SGCs to SGCs in other genomes were also investigated. For this purpose Minimum Information about a Biosynthetic Gene cluster (MIBiG) cluster comparison (Blin *et al.*, 2021; Medema *et al.*, 2015), ClusterBlast (Medema *et al.*, 2011), KnownClusterBlast (Blin *et al.*, 2019), Cluster Pfam Analysis (Blin *et al.*, 2021; Medema *et al.*, 2015), and Pfam-based GO term

annotation (Blin *et al.*, 2021; Medema *et al.*, 2015) were used. All these bioinformatic tools are implemented in fungiSMASH.

2.1.3. Phylogenetic analyses of NIS synthetases

A database of amino acid sequences of identified NIS synthetase genes in the genomes studied were generated. Additional amino acid sequences of known NIS synthetase genes available on GenBank were also added to the dataset. The final dataset (forty-five amino acid sequences) consisted of nineteen putative NIS synthetases of *Armillaria* and other members of the Physalacriaceae, six orthologous proteins (selected based on high Query Cover and Percentage Identity, and low Expected value based on BLASTp searches) from GenBank, and twenty validated or proposed NIS synthetases as reported by Carroll and Moore (2018) and Oves-Costales *et al.* (2009). The sequences were used in a phylogenetic analysis to predict to which type(s) the detected NIS synthetases in the Physalacriaceae belonged.

Multiple sequence alignments were conducted in MAFFT (Katoh *et al.*, 2019) and phylogenies of the dataset were inferred using Molecular Evolutionary Genetics Analysis (MEGA) version X software (Stecher *et al.*, 2020). A phylogenetic tree was inferred by using the Maximum Likelihood method. The LG + G + I substitution model was identified, by a model test implemented in MEGAX, as the best substitution model based on the Bayesian Information Criterion (BIC). The initial tree(s) for the heuristic search were obtained automatically by applying Neighbour Joining (Saitou and Nei, 1987) algorithm to a matrix of pairwise distances, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.5511)). All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 428 positions in the final dataset. Nodal support was determined using Bootstrap analysis with 1000 replications (Felsenstein, 1985). The tree was exported as a Newick file and visualized using FigTree version 1.4.4 (Rambaut, 2018).

2.1.4. Determination of domain architecture, size, and exon number of NIS synthetase genes

Characteristics of the NIS synthetase genes (domain architectures, protein size in amino acids, and exon number) were obtained from fungiSMASH and confirmed using the InterPro web-based tool (Blum *et al.*, 2021) (https://www.ebi.ac.uk/interpro/) based on results retrieved from the Pfam database (Mistry *et al.*, 2021).

2.2. Cultivation and assessment of siderophore biosynthesis potential of selected *Armillaria* strains

2.2.1. Growth medium preparation

To avoid iron contamination, all experiments were performed using glassware washed with HCl (6 M) and rinsed three times with ddH_20 (Cox, 1994; Schwyn and Neilands, 1987). Sterile plastic Petri dishes were used in the growth studies.

Strains were maintained on malt yeast extract agar (MYA) medium (15 g/L malt extract, 2 g/L yeast extract, 15 g/L agar) in 6.5 cm disposable Petri dishes. Medium without iron (PDP-) were prepared with Potato Dextrose (24 g/L) supplemented with peptone (2 g/L). Solid medium (PDPA-) was prepared with addition of agarose (10 g/L) instead of agar (Giuliano Garisto Donzelli *et al.*, 2015). For iron replete conditions, PDP- and PDPA- were supplemented with 100 µM $FeC1_3.6H_2O$ (hereon referred to as PDP+ and PDPA+, respectively). To obtain different concentrations of added iron in PDP-, $FeC1_3.6H_2O$ was added to the desired concentration. All cultures were incubated at 25±2 ºC in the dark.

2.2.2. Iron-dependent culture growth on solid media

Nine strains representing six *Armillaria* spp. (Table 1) were evaluated in this experiment. The solid media (PDPA- and PDPA+) was used to investigate the iron-dependent culture growth rate and macromorphology of the respective strains. Each plate was inoculated with one 5mm diameter culture disc of actively growing culture of the respective strains in the centre of the media. Plates were incubated for 6 weeks. Perpendicular lines intersecting at the centre of the inoculum were drawn at the bottom of each Petri dish. To avoid long disruptions in the incubation period, marks were made at the tip of the mycelia/rhizomorph in the same perpendicular angle at the end of each week. Both the top and bottom views of the plates were photographed at the end of the incubation period. Radial lengths per week were measured with the ImageJ software (https://imagej.nih.gov/ij/download.html) from the centre of the inoculum to the mark made at the tip of the culture by using the Straight-line tool. A scale was set using the known size of the Petri dish in cm.

2.2.3. Evaluation of iron-dependent siderophore biosynthesis

Firstly, siderophore biosynthesis under iron deplete and iron replete conditions were evaluated for all the strains of *Armillaria* species included in this study (Table 1). The respective cultures (1 cm^2 culture disks) were aseptically inoculated into 50 ml falcon tubes containing 15 ml PDP- or PDP+. The cultures were incubated for 24 hours. Secondly, the iron-dependent repressibility of siderophore

biosynthesis was also investigated. For this purpose, 1 cm^2 culture disk of *A. fuscipes* strain CMW2740 and *A. mellea* strain CMW31132 were separately inoculated in 15 ml PDB amended with $FeCl_3$ at different concentrations (0, 20, 40, 60, 80, 100, 150, 200 µM) in 50 ml falcon tubes and incubated for 24 hours. These strains were selected to represent facultative necrotrophs/pathogenic and facultative necrotrophs/highly pathogenic species respectively (Table 1). Additionally, these two strains (*A. fuscipes* strain CMW2740 and *A. mellea* strain CMW31132) appeared to produce siderophores differently as presented in Figure 4 of Chapter 2 of this thesis.

For both experiments, siderophore biosynthesis was detected using the modified CAS assay solution with the microtiter method according to Alexander and Zuberer (1991). Cultures were swirled to mix and aliquots (100 µL) of supernatants were mixed with 100 µL of CAS assay solution in 96-well plates (TPP® Tissue Culture Test Plate 96F, Switzerland). Absorbance readings were done in a SpectraMax Plus 384 Spectrophotometer (Labotec, South Africa) at 630 nm after 20 mins. Siderophore production was quantified in percentage siderophore units (psu) using the formula:

$$\text{Siderophore production (psu)} = \frac{(Ar - As) \times 100}{Ar}$$

where A_r = absorbance of reference (CAS solution and un-inoculated broth (control)), and A_s = absorbance of sample (CAS solution and cell-free supernatant of sample) (Payne, 1994).

2.2.4. Statistical analysis

Growth experiments were performed in biological triplicates. For iron-dependent siderophore biosynthesis and iron-repressibility of siderophore biosynthesis experiments, we repeated both experiments twice independently and with three biological replicates per treatment. All data were analysed using Microsoft Excel. The mean and standard errors of all reads for the growth experiment were determined at the fourth week of incubation. This was reported as the radial culture length at week 4. Statistical significance of the radial culture lengths between treatments for each strain was evaluated using t-test. One-tailed paired two sample for means were calculated for comparison of all strains studied excluding strains CMW3159 and CMW4456. Due to the irregular and rhizomorphic growth patterns of these two strains, the one-tailed two-sample assuming equal variances t-test was performed for CMW3159 and CMW4456. A 95 % probability level was used in all analyses. Growth rates were calculated by the linear regression obtained by the fourth week of incubation. Iron-repressibility of siderophore biosynthesis were means and standard errors for each concentration of iron.

3. Results

3.1. Comparative genomics analyses

3.1.1. Identified NIS synthetase gene clusters and cluster annotation

Two different NIS synthetase gene clusters (NIS Cluster 1 and NIS Cluster 2) were found in the genomes studied. These BGCs showed conserved microsynteny with retention of gene content, order, and orientation. Some gene loss and/or duplication events were, however, observed (Figures 1 and 2).

NIS Cluster 1 (Figure 1, Table S2) included genes which putatively encode ATP-binding cassette (ABC) transporters, S-adenosyl-L methionine-dependent methyltransferase, IucA/IucC family domain containing protein (NIS synthetase gene), C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1, and cytochrome P450. The putative ABC transporter genes found in all the genomes studied were duplicated. The cytochrome P450 genes in NIS Cluster 1 were duplicated in some of the *Armillaria* and *Desarmillaria* gene clusters (Armbor1 S7, (Des)Armtab1 S1 and (Des)Armect1 S5). There appeared to be remnants of the duplicated cytochrome P450 in other gene clusters of some *Armillaria* spp. and *G. necrorhizus* (Armnabs1 S9, Armme1 S3, Armnov1 S5, and Guyne1 S3), while Oudmuc1 S1 lacked this gene.

Unlike NIS Cluster 1, NIS Cluster 2 and neighbouring genes included genes which putatively encode a ferric reductase-like transmembrane component domain containing protein, glycoside hydrolase family 61 and 95 proteins, ubiquitin-conjugating enzyme/RWD-like protein, endoglucanase V-like protein, soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) domain-containing protein, IucA/IucC family domain containing protein (NIS, biosynthetic backbone gene), WH1-domain-containing protein, glutamyl-tRNA synthetase, and Heat shock factor (HSF) type DNA-binding-domain containing protein (Figure 2, Table S2). All the NIS synthetase gene clusters contained hypothetical/expressed proteins (Figure 1 and Figure 2). The expected location of the second NIS synthetase gene of *A. novae-zelandiae* had high similarity (>70 %) with the NIS synthetase genes in *A. fumosa* (Armfum1 S10) and *D. ectypa* ((Des)Armect1 S11) (Figure 2).

3.1.2. Similarity of CASSIS-determined cluster boundaries to other BGCs

The cluster boundaries of NIS Cluster 1 as predicted by the CASSIS algorithm included the genes encoding IucA/IucC family domain containing protein (NIS synthetase, biosynthetic backbone gene), C2 domain-containing protein / MRP-like transporter and Frag/DRAM/Sfk1 in the genomes of most of the *Armillaria* and *Desarmillaria* spp. (Table 2). NIS Cluster 1 of *A. nabsnona* (Armnabs1 S9) and *D. ectypa* ((Des)Armect1 S5) also contained a gene encoding cytochrome P450, whereas that of *A.*

cepistipes (Armcep1 S262) lacked the gene encoding C2 domain-containing protein / MRP-like transporter. NIS Cluster 1 in the genome of *G. necrorhizus* (Guyne1 S3) appeared to have a remnant of the gene encoding C2 domain-containing protein / MRP-like transporter. This gene has putatively been annotated as a hypothetical protein in Guyne1 S3. The cluster boundary for NIS Cluster 1 of *Oudemansiella mucida* (Oudmuc1 S1) included genes which encode IucA/IucC family domain containing protein, WD repeat containing protein 8, and Major Facilitator Superfamily (MSF) transporter (Table 2).

Genes encoding SNARE domain-containing protein, IucA/IucC family domain containing protein, WH1-domain-containing protein, and glutamyl-tRNA synthetase were predicted as the boundaries of NIS Cluster 2 by CASSIS algorithm in *A. borealis* (Armbor1 S46), *A. nabsnona* (Armnabs1 S10), and *A. cepistipes* (Armcep1 S270). The genes in the cluster boundaries of all NIS Cluster 2 in the other studied genomes, except that of *O. mucida* (Oudmuc1 S66), encodes all the indicated proteins excluding glutamyl-tRNA synthetase. The cluster boundary for Oudmuc1 S66 consisted of genes which encode small nucleolar RNA-associated protein 3, ribosomal protein S12, SNARE domain-containing protein, IucA/IucC, and three hypothetical proteins (Table 2).

No MIBiG comparison hit was recorded for NIS Cluster 1 of all the studied genomes apart from NIS Cluster 1 in the genomes of *D. tabescens* ((Des) Armtab1 S1) and *O. mucida* (Oudmuc1 S1), in which there were hits with no similarity to the biosynthetic backbone gene, IucA/IucC (Table 2). MIBiG comparison of NIS Cluster 2 in the studied genomes revealed that a region, BGC0000944.1 (location, 2257888-2263981), responsible for staphyloferrin A biosynthesis in *Staphylococcus aureus* subsp. *aureus* NCTC 8325 (Cotton *et al.*, 2009) had a similarity score of 0.13 in the MIBiG database (Table 2). This excluded NIS Cluster 2 of *A. cepistipes* (Armcep1 S270).

Results from ClusterBlast showed NIS Cluster 1 in all the genomes excluding that of Oudmuc1 S1 had 100 % gene similarity (in terms of the proteins encoded by the genes) within the BGC of *Agaricus bisporus* var. *bisporus* H97, (NW_006267366; location, 1141250-1156888) (Table 2). No similar region was recorded for NIS Cluster 2 in all the genomes. BLASTp hits of the genes in NIS Cluster 2 of *A. borealis* to this region is shown in Table S3. No KnownClusterBlast hits were recorded for all the BGCs identified in this study.

3.1.3. Phylogeny of putative NIS synthetase genes

The phylogenetic analysis revealed that all the NIS synthetase genes (*IucA/IucC* genes) in both NIS Clusters 1 and 2 identified in the genomes studied encode Type A' NIS synthetases although the NIS synthetases of *O. mucida* branch separately from those of the other members of the Physalacriaceae in the respective gene clusters (Figure 3; Table 3). Nevertheless, the NIS synthetases from the

respective gene clusters formed two distinct highly supported phylogenetic clades (Figure 3; 100 % bootstrap value). The NIS synthetases in NIS Cluster 1 of the Physalacriaceae (C1, Figure 3) grouped with some uncharacterized orthologs in the Basidiomycota (O1, Figure 3; Table 3) and with the only characterized fungal NIS synthetase from *R. delemar* RA 99-880 (Rfs, Figure 3; Table 3).

3.1.4. Domain architectures and other characteristics of putative NIS synthetase genes

The domain architectures of the putative NIS synthetases were similar in almost all the genomes studied and were comparable to the domain architectures of the known NIS synthetase genes included in the analysis. The NIS synthetase gene of *A. cepistipes* S262 (Armcep1_13598) contained IucA_IucC (PF04183.14) domain at the N-terminal, Fhuf (PF06276.14) domain in the middle, and C2 domain at the C-terminal (top insert, Figure 3). The NIS synthetases in all the other *Armillaria* spp. as well as the *Desarmillaria* spp. and *O. mucida* had IucA_IucC and Fhuf domains at the N- and C-terminals, respectively (bottom insert, Figure 3). This architecture has also been recorded in the NIS synthetases of *R. delemar* RA 99-880, Rfs (Carroll *et al.*, 2017) and the uncharacterized Type A' NIS synthetases of all the orthologs which were identified. All the other known NIS synthetases included in the phylogenetic analyses in Section 3.1.3 also had this domain architecture.

The NIS synthetases differed in size and number of exons. The sizes of the NIS synthetases from NIS Cluster 1 were 497 – 1132 amino acids (Table 3) with 6 – 10 exons (data not shown). NIS synthetase of NIS Cluster 2 were 483 – 559 amino acids (Table 3) with 1 – 3 exons (data not shown).

Gene prediction of the NIS synthetase gene in the manually detected NIS Cluster 2 of *A. novae-zelandiae* with AUGUSTUS v. 3.3.3 identified two copies. The predicted genes were 510 and 515 amino acids with 2 or 3 exons respectively. Both genes were orthologous to IucC domain family containing proteins of *Mucidula mucida* and *Gymnopilus junonius* (accession numbers KAF8906867.1 and KAF8899909.1) with query covers and percentage identities greater than 50 %. Both predictions also had the IucA_IucC and Fhuf domains at the N- and C-terminals respectively based on InterPro predictions (results not shown).

3.2. Iron-dependent growth and siderophore biosynthesis

In vitro experiments were conducted to investigate the effect of iron on growth and siderophore biosynthesis by *Armillaria* species to gain some understanding of why members of this genus dedicate several genetic tools to siderophore biosynthesis and transport.

3.2.1. Iron-dependent growth

Culture radial growth and growth rates calculated at week 4 of incubation, and culture macromorphology by the end of incubation (week 6 of incubation) differed among the different strains studied (Figure 4). The lowest culture radii and growth rates were recorded for *A. fuscipes* strain CMW2740. The highest radii and growth rates were recorded for *Armillaria* spp. from African Clade B (CMW4456) (Figure 4). For both strains, the cultures grown on PDPA+ had a significantly longer radius than the cultures grown on PDPA- ($p<0.05$) based on a t-test using a 95 % probability level. Between the *A. gallica* strains studied, cultures of CMW31092 showed significantly lower radii and growth rates when cultured on PDPA- than on PDPA+ ($p<0.05$; Figure 4). The reverse was true for *A. gallica* strain CMW45397. R^2 values recorded for the linear regression obtained for all the strains were greater than 0.9 (results not shown), indicating a linear growth of all strains under the experimental conditions.

Culture macromorphology was generally similar for the strains on the respective media (Figure 4). *Armillaria nabsnona* strain CMW3159, however, showed a denser culture macromorphology when cultured on PDPA- compared to PDPA+ (Figure 4). On PDPA+ cultures of *Armillaria* spp. from African Clade B (CMW4456) showed whitish aerial mycelia on the rhizomorphs. All strains secreted brownish exudates which diffused into the media and/or present as liquid on the cultures on both media (Figure 4). Exudate secretion occurred at different degrees among the strains as shown by the change of the yellowish growth media to varying shades of brown and the presence of brown liquid on the cultures. Exudate secretion was generally more pronounced on PDPA+ as exemplified by *A. fuscipes* strains CMW2740 and CMW3164.

3.2.2. Iron-dependent siderophore biosynthesis

All strains studied produced siderophores both under iron deplete (PDP-) and iron replete (PDP+) conditions as shown by a colour change of the blue CAS media to different colours (Data not shown). The iron-repressibility of siderophore biosynthesis fluctuated for both strains. The best fits for the curves obtained followed polynomial trendlines with orders of 6 for both *A. fuscipes* strain CMW2740 ($R^2 = 0.973$) and *A. mellea* strain CMW31132 ($R^2 = 0.9999$) (Figure 5). Strain CMW31132 generally produced more siderophores than strain CMW2740 at the same concentration of added $FeCl_3$ (Figure 5).

4. Discussion

4.1. Genome analyses reveal two distinct NIS synthetase gene clusters and NIS synthetase genes in the Physalacriaceae

4.1.1. NIS synthetase gene clusters in the Physalacriaceae

NIS synthetase gene clusters contain gene(s) which encode one or more types of NIS synthetase only (Bull *et al.*, 1996; Carroll *et al.*, 2017; Cheung *et al.*, 2009; Deng *et al.*, 2006; Lynch *et al.*, 2001; Moon *et al.*, 2004). These clusters may also contain other genes involved in biosynthesis, transport, and regulation (Bull *et al.*, 1996; Carroll *et al.*, 2017; Cheung *et al.*, 2009; Deng *et al.*, 2006; Lynch *et al.*, 2001; Moon *et al.*, 2004). The presence of only the NIS-type backbone gene (*IucA/IucC*) in all the identified gene clusters in this study showed that they are NIS synthetase gene clusters and not hybrid NRPS/NIS gene clusters.

Results from this study provides evidence that the products of the identified gene clusters are essential for iron homeostasis in species within the investigated genera. This assertion is based on the largely conserved microsynteny in the genes and intergenic regions of both NIS Clusters 1 and 2 and their neighbouring genes in the genomes of *Armillaria, Desarmillaria,* and *Guyanagaster* despite the observed putative gene loss and/or duplication events. Species from the same and/or different genera have previously been reported to show conserved synteny as well as some gene loss and/or duplication or other forms of gene modification events in biosynthetic gene clusters (Evdokias *et al.*, 2021; Goering *et al.*, 2016; Tralamazza *et al.*, 2019). Gene recruitment, duplication, repurposing, and other modification events in BGCs have been shown to impact secondary metabolism. This effect occurs through loss of gene (cluster) function or diversification of structures of the associated secondary metabolites as exemplified by ergot alkaloids biosynthesis by members of the fungal family, Clavicipitaceae (Florea *et al.*, 2017). Additionally, *in vitro* lifestyle- and species-independent differential siderophore biosynthesis has been reported in *Armillaria* (Narh Mensah *et al.*, 2023). Hence, the biological implications of the presently recorded putative NIS synthetase gene cluster modifications in the Physalacriaceae need to be investigated in future.

Various organisms have been reported to employ different tools for siderophore biosynthesis. The present study has shown that in addition to the previously reported NRPS-dependent siderophore synthetase gene cluster in genomes of *Armillaria* and other species within the Physalacriaceae (Narh Mensah *et al.*, 2023), the genomes of these fungi also contain two distinct NIS synthetase gene clusters. The ability to differentially regulate and synthesize different siderophores with different biological functions using two of the three siderophore biosynthesis pathways has been reported in *Erwinia chrysanthemi* (Franza *et al.*, 2005) and some other bacteria (Cheung *et al.*, 2009; Cotton *et*

al., 2009; Koppisch *et al.*, 2005). We predict that the two distinct NIS synthetase gene clusters in the genomes of the respective Physalacriaceae will be regulated differently and will synthesize different siderophores which may be involved in different or overlapping functions under different conditions. This is based on the observed differences in both the NIS synthetase genes and the flanking genes in the two gene clusters within each of the studied genomes. Genes such as HSF type DNA-binding-domain containing protein in NIS Cluster 2 remain latent and are activated under stress conditions and in response to developmental signals to induce transcription of heat shock genes (reviewed in Pirkkala *et al.* (2001)). Heat shock proteins have been reported to be highly abundantly expressed by organisms in response to high metal concentrations (Dias *et al.*, 2019; Khatiwada *et al.*, 2020; Okay *et al.*, 2020).

The genomes of organisms other than fungi have been shown to contain more than one NIS synthetase gene cluster or operon with one or more type(s) of NIS synthetase gene(s) in each cluster (Carmichael *et al.*, 2019; Cotton *et al.*, 2009; Franza *et al.*, 2005). To the best of our knowledge, this is the first report of different NIS synthetase gene clusters in fungal genomes.

4.1.2. Cluster boundaries and similarities to known NIS synthetase gene clusters

Genes in the cluster boundaries of both NIS Clusters 1 and 2 of the Physalacriaceae predicted by CASSIS are typical of NIS synthetase gene clusters. These genes encode the NIS biosynthesis backbone gene (*IucA/IucC* in both NIS Clusters 1 and 2) and the transporter (C2 domain-containing protein / MRP-like transporter in NIS Cluster 1). Other proteins, such as ABC transporters, Cytochrome P450, Frag/DRAM/Sfk1, SNARE domain-containing protein and Glutamyl-tRNA synthetase, encoded by genes identified in NIS Clusters 1 and 2 in the genomes of species of *Armillaria, Desarmillaria* and *Guyanagaster* during this study may function in substrate/product modification, transport and regulation. This is supported by previous studies that showed that petrobactin acquisition in *B. anthracis* is facilitated by multiple ABC transporters (Dixon *et al.*, 2012). Furthermore, different authors showed that NIS synthetase gene clusters for desferrioxamine biosynthesis by *Streptomyces* spp. contain genes which encode cytochrome P450 monooxygenases (Iftime *et al.*, 2016; Jarmusch *et al.*, 2021). Additionally, the ferric reductase-like transmembrane component domain containing protein gene found in NIS Cluster 2 in all the studied genomes, excluding that of *O. mucida*, may be involved in iron homeostasis in these fungi. The participation of ferric reductases in iron homeostasis has been documented in other organisms including bacteria and some fungal pathogens (Cain and Smith, 2021). The putative modification of the gene encoding C2 domain-containing protein / MRP-like transporter in NIS Cluster 1 in the genome of *G. necrorhizus* and its biological implications needs to be investigated.

The gene clusters identified for *O. mucida* in this study differed from NIS Clusters 1 and 2 present in the other species. The gene clusters for *O. mucida*, however, contained genes which encode the NIS synthetase (*IucA/IucC* in both gene clusters) and other genes putatively involved in regulation and/or transport. Genes which encode transporters and regulators have also been reported in various NIS synthetase gene clusters of other organisms (Cheung *et al.*, 2009; Iftime *et al.*, 2016; Moon *et al.*, 2004).

The two NIS synthetase gene clusters detected in the studied genomes code for proteins which are likely to biosynthesize different siderophores by the respective species. This is evidenced by the fact that the genes located in the cluster boundaries of both gene clusters in the studied genomes are different. In addition, there is no significant similarity score in the MIBiG database for the NIS Cluster 1 as opposed to NIS Cluster 2 and vice versa for the ClusterBlast hits. Results from this study suggest that the product of NIS Cluster 1 in the genomes of *Armillaria, Desarmillaria* and *Guyanagaster* will be similar to the product which may be synthesized by the NW_006267366 BGC of *A. bisporus* var. *bisporus* H97. Conversely, based on MIBiG comparison, the product expressed by the NIS synthetase gene of NIS Cluster 2 of all the studied genomes, excluding that of *A. cepistipes* (Armcep1 S270), will be similar to the product expressed by the gene in the BGC, BGC0000944.1: 0-6093, although the similarity score is low. This cluster synthesizes staphyloferrin A in *S. aureus* subsp. *aureus* NCTC 8325. Presence of a set of identical genes in a siderophore BGC in the genome of an organism does not necessarily translate to biosynthesis of the same siderophore as shown in putrebactin and alcaligin BGCs in *Shewanella* spp. and *Bordetella* spp., respectively (Kadi *et al.*, 2008). This fact, as well as the no or low similarity scores presently recorded for the MIBiG comparisons and the fact that there were no KnownClusterBlast hits for NIS Clusters 1 and 2 in all the studied genomes suggest that the identified NIS synthetase gene clusters will synthesize novel siderophores.

4.1.3. Phylogenetic analysis of putative NIS synthetases reveal that the genes in both gene clusters are Type A' NIS synthetases

All the putative NIS synthetases in the Physalacriaceae were identified as Type A' NIS synthetases. The identity of the NIS synthetases was based on their clustering with representatives of Type A' NIS synthetases. The representative synthetases were FslA, LbtA, Rfs, SfaB, and SfaD which synthesize rhizoferrin (Sullivan *et al.*, 2006), legiobactin (Allard *et al.*, 2006), rhizoferrin (Carroll *et al.*, 2017), staphyloferrin A (Cotton *et al.*, 2009) and staphylofferin A (Cotton *et al.*, 2009) respectively.

Within the Type A' NIS synthetase group, the NIS synthetases of NIS Cluster 1 formed one phylogenetic group while NIS synthetases in NIS Cluster 2 formed another distinct phylogenetic group. Based on the phylogenetic analysis, the NIS synthetases encoded by the NIS synthetase genes

in NIS Cluster 1 of all the genomes studied is expected to be more similar to the only characterized fungal NIS synthetase, Rfs, from *R. delemar* (Carroll *et al.*, 2017). These results further support our notion that the two distinct NIS synthetase gene clusters in the Physalacriaceae may synthesize different siderophores.

4.1.4. Characteristics of the putative NIS synthetases are comparable to characterized NIS synthetases of other organisms

IucA_IucC and ferric iron reductase FhuF-like transporter domains at the N- and C-terminals is a conserved domain architecture in several characterized NIS synthetases (Carroll *et al.*, 2017). This characteristic feature of known NIS synthetases was also recorded in the putative NIS synthetases encoded by the studied Physalacriaceae excluding that of *A. cepistipes* S262 (Armcep1_13598). The IucA_IucC domain is responsible for biosynthesis of the siderophore whereas Fhuf may be involved in transport (Cain and Smith, 2021; Carroll *et al.*, 2017; Matzanke *et al.*, 2004; Müller *et al.*, 1998).

The sizes of the NIS synthetases in the Physalacriaceae determined in this study were generally similar to those of known NIS synthetases. Some of the NIS synthetases were smaller (Oudmuc1_1223755 [S66] and Armtab1_1497269 [S1]) or larger (Armcep1_13598 [S262]) than the known NIS synthetases. The larger size of Armcep1_13598 [S262] is explained by the fact that this gene appeared to be a fused gene consisting of the NIS synthetase gene and the C2 domain-containing protein / MRP-like transporter gene located immediately downstream of the NIS synthetase gene in NIS Cluster 1 of the other genomes studied, excluding that of *O. mucida* S1 (Oudmuc1_1247625). This kind of fusion of two genes has been documented in other siderophore BGCs or operons. For instance, Carmichael *et al.* (2019) showed that the homologs of the *IucA* and *IucB* genes in the woodybactin BGC of *Shewanella woodyi* MS32 is fused. In terms of exons, the gene encoding Rfs in *R. delemar* RA 99-880 contains 6 exons (Carroll *et al.*, 2017) and was comparable to the number of exons in the *IucA/IucC* genes of NIS Cluster 1 in this study. Further research is required to determine the biological implications of the putative NIS synthetase gene modification for *A. cepistipes*.

4.2. Iron-dependent growth and siderophore biosynthesis

Narh Mensah *et al.* (2023) recently reported one conserved NRPS-dependent siderophore synthetase gene cluster in *Armillaria* and other members of the Physalacriaceae and biosynthesis of different types of siderophores by strains of different *Armillaria* spp. In the current study we identified two putative distinct NIS synthetase gene clusters in the genomes studied. Together these studies suggest that *Armillaria* spp. and other members of the Physalacriaceae have a strong need for iron homeostasis. Research discussed in this section sought to gain insight into the effect of iron on growth and siderophore biosynthesis on species of *Armillaria*.

4.2.1. Iron-dependent growth and macromorphology of Armillaria species varied on the various media

Growth of various microorganisms have been shown to generally increase with increasing iron concentrations in the growth medium (Deng *et al.*, 2006; Giuliano Garisto Donzelli *et al.*, 2015; Manwar *et al.*, 2004). The effect of iron on both growth rate and extent of growth of the *Armillaria* spp. in the present study, irrespective of the species, is inconsistent with the reported iron-dependent growth of other organisms. This suggests that *Armillaria* spp. may differ in their requirements for iron and that this variation in iron requirement by *Armillaria* spp. is not species-specific.

Various macromorphological characteristics have been reported to be affected by iron concentration in other organisms. This includes the report that morphological characteristics such as number and biomass of microsclerotia as well as melanin production of the fungus, *Nomuraea rileyi,* increase with increasing added iron in the growth medium (Li *et al.*, 2016). *Armillaria* spp. in this study generally retained culture macromorphology under the two growth conditions, although the observed brownish exudates were more predominant on the iron replete PDPA (PDPA+). The composition of the exudates synthesized by the species included in this study are not known. Knowing the composition would provide a better insight into the effect of iron on biosynthetic properties of species of *Armillaria*.

4.2.2. Siderophore biosynthesis occurred irrespective of concentration of added $FeCl_3$

In this study all strains synthesized siderophores with no addition of $FeC1_3$ and at 100 µM added $FeC1_3$. This concentration is much higher relative to the gram-negative bacteria *P. putida* strain B 10 (Alexander and Zuberer, 1991) and *P. aeruginosa* (Manwar *et al.*, 2004) that synthesize siderophores at 0 – 40 µM but not at 50 µM added $FeC1_3.6H_2O$. This finding further supports our assertion that *Armillaria* spp. generally differ in their iron requirements.

Siderophore biosynthesis is usually inversely proportional to the concentration of iron in the growth medium (Alexander and Zuberer, 1991; Fekete *et al.*, 1989; Li *et al.*, 2016; Manwar *et al.*, 2004). For instance, the basidiomycetes *Coriolus versicolor* and *Gloeophyllum trabeum* show decreasing siderophore biosynthesis with increasing added $FeC1_3$ concentration *in vitro* at 30 days of incubation (Fekete *et al.*, 1989). This trend was not observed in the present study as shown by the polynomial trendlines recorded for siderophore biosynthesis by both strains. We propose that the observed $FeC1_3$ concentration-independent siderophore production by the studied *Armillaria* spp. may be due to biosynthesis of different siderophores at different concentrations of added $FeC1_3$. This phenomenon has been reported in *E. chrysanthemi* (Franza *et al.*, 2005).

Conclusions

In this study NIS synthetase gene clusters and NIS synthetase genes in genomes of *Armillaria* species were identified, characterized, and compared to other species in the Physalacriaceae. This is the first report of two distinct NIS synthetase gene clusters in fungal genomes. Our results suggest that the NIS synthetase gene clusters may synthesize different siderophores and that *Armillaria* species have unique requirements for iron.

Our findings demonstrate that more research is required on siderophore biosynthesis and utilization in the Basidiomycota. Studies such as gene modification/knock out combined with biochemical characterization (e.g., enzymology, proteomics, and metabolomics) will be needed to fully understand these gene clusters, to characterize the genes and siderophores synthesized, and to determine the biosynthetic models of these BGCs. Other biological studies will be required to determine the role of the siderophores in growth, pathogenicity and/or virulence, and other potential functions of siderophores in *Armillaria* spp. The knowledge generated from this study, and the suggested studies will unlock avenues for controlling fungal pathogens, such as those belonging to *Armillaria*, and may result in discovery of new biotechnologically useful products.

References

Alexander, D.B., and Zuberer, D.A. (1991). Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. *Biol. Fert. Soils* 12(1)**,** 39-45. doi: 10.1007/bf00369386.

Allard, K.A., Viswanathan, V.K., and Cianciotto, N.P. (2006). lbtA and lbtB are required for production of the *Legionella pneumophila* siderophore legiobactin. *J. Bacteriol.* 188(4)**,** 1351-1363. doi: 10.1128/JB.188.4.1351-1363.2006.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17)**,** 3389-3402. doi: 10.1093/nar/25.17.3389.

Blin, K., Kim, H.U., Medema, M.H., and Weber, T. (2019). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform* 20(4)**,** 1103-1113. doi: 10.1093/bib/bbx146.

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, Marnix H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49(W1)**,** W29-W35. doi: 10.1093/nar/gkab335.

Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49(D1)**,** D344-D354. doi: 10.1093/nar/gkaa977.

Brandenburger, E., Gressler, M., Leonhardt, R., Lackner, G., Habel, A., Hertweck, C., et al. (2017). A highly conserved basidiomycete peptide synthetase produces a trimeric hydroxamate siderophore. *Appl. Environ. Microbiol.* 83(21)**,** e01478-01417. doi: 10.1128/aem.01478-17.

Bull, C.T., Carnegie, S.R., and Loper, J.E. (1996). Pathogenicity of mutants of *Erwinia carotovora* subsp. *carotovora* deficient in aerobactin and catecholate siderophore production. *Phytopathology* 86(3)**,** 260-266.

Cain, T.J., and Smith, A.T. (2021). Ferric iron reductases and their contribution to unicellular ferrous iron uptake. *J. Inorg. Biochem.* 218**,** 111407. doi: 10.1016/j.jinorgbio.2021.111407.

Carmichael, J.R., Zhou, H., and Butler, A. (2019). A suite of asymmetric citrate siderophores isolated from a marine *Shewanella* species. *J. Inorg. Biochem.* 198**,** 110736. doi: 10.1016/j.jinorgbio.2019.110736.

Carroll, C.S., Grieve, C.L., Murugathasan, I., Bennet, A.J., Czekster, C.M., Liu, H., et al. (2017). The rhizoferrin biosynthetic gene in the fungal pathogen *Rhizopus delemar* is a novel member of the NIS gene family. *Int. J. Biochem. Cell Biol.* 89**,** 136-146. doi: 10.1016/j.biocel.2017.06.005.

Carroll, C.S., and Moore, M.M. (2018). Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. *Crit. Rev. Biochem. Mol. Biol.* 53(4)**,** 356-381. doi: 10.1080/10409238.2018.1476449.

Challis, G.L. (2005). A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem* 6(4)**,** 601-611. doi: 10.1002/cbic.200400283.

Cheung, J., Beasley, F.C., Liu, S., Lajoie, G.A., and Heinrichs, D.E. (2009). Molecular characterization of staphyloferrin B biosynthesis in *Staphylococcus aureus*. *Mol. Microbiol.* 74(3)**,** 594-608. doi: 10.1111/j.1365-2958.2009.06880.x.

Cotton, J.L., Tao, J., and Balibar, C.J. (2009). Identification and characterization of the *Staphylococcus aureus* gene cluster coding for staphyloferrin A. *Biochemistry* 48(5)**,** 1025-1035. doi: 10.1021/bi801844c.

Cox, C.D. (1994). "[24] Deferration of laboratory media and assays for ferric and ferrous ions," in *Methods in Enzymology,* eds. V.L. Clark & P.M. Bavoil.), 315-329.

Dale, S.E., Doherty-Kirby, A., Lajoie, G., and Heinrichs, D.E. (2004). Role of siderophore biosynthesis in virulence of *Staphylococcus aureus*: identification and characterization of genes involved in production of a siderophore. *Infect. Immun.* 72(1)**,** 29-37. doi: 10.1128/IAI.72.1.29-37.2004.

de Lorenzo, V., Bindereif, A., Paw, B.H., and Neilands, J.B. (1986). Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. *J. Bacteriol.* 165(2)**,** 570-578. doi: 10.1128/jb.165.2.570-578.1986.

de Lorenzo, V., and Neilands, J.B. (1986). Characterization of *iucA* and *iucC* genes of the aerobactin system of plasmid ColV-K30 in *Escherichia coli*. *J. Bacteriol.* 167(1)**,** 350-355. doi: 10.1128/jb.167.1.350-355.1986.

Deng, K., Blick, R.J., Liu, W., and Hansen, E.J. (2006). Identification of *Francisella tularensis* genes affected by iron limitation. *Infect. Immun.* 74(7)**,** 4224-4236. doi: 10.1128/IAI.01975-05.

Dias, M., Gomes de Lacerda, J.T.J., Perdigao Cota de Almeida, S., de Andrade, L.M., Oller do Nascimento, C.A., Rozas, E.E., et al. (2019). Response mechanism of mine-isolated fungus *Aspergillus niger* IOC 4687 to copper stress determined by proteomics. *Metallomics* 11(9)**,** 1558-1566. doi: 10.1039/c9mt00137a.

Dixon, S.D., Janes, B.K., Bourgis, A., Carlson, P.E., Jr., and Hanna, P.C. (2012). Multiple ABC transporters are involved in the acquisition of petrobactin in *Bacillus anthracis*. *Mol. Microbiol.* 84(2), 370-382. doi: 10.1111/j.1365-2958.2012.08028.x.

Eisendle, M., Oberegger, H., Zadra, I., and Haas, H. (2003). The siderophore system is essential for viability of *Aspergillus nidulans*: Functional analysis of two genes encoding L-ornithine N^5-monooxygenase (sidA) and a non-ribosomal peptide synthetase (sidC). *Mol. Microbiol.* 49(2), 359-375. doi: 10.1046/j.1365-2958.2003.03586.x.

Evdokias, G., Semper, C., Mora-Ochomogo, M., Di Falco, M., Nguyen, T.T.M., Savchenko, A., et al. (2021). Identification of a novel biosynthetic gene cluster in *Aspergillus niger* using comparative genomics. *J. Fungi* 7(5), 374.

Fekete, F.A., Chandhoke, V., and Jellison, J. (1989). Iron-binding compounds produced by wood-decaying basidiomycetes. *Appl. Environ. Microbiol.* 55(10), 2720-2722.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the Bootstrap. *Evolution* 39(4), 783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x.

Florea, S., Panaccione, D.G., and Schardl, C.L. (2017). Ergot alkaloids of the family Clavicipitaceae. *Phytopathology* 107(5), 504-518. doi: 10.1094/PHYTO-12-16-0435-RVW.

Floudas, D., Held, B.W., Riley, R., Nagy, L.G., Koehler, G., Ransdell, A.S., et al. (2015). Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. *Fungal Genet. Biol.* 76, 78-92. doi: 10.1016/j.fgb.2015.02.002.

Franza, T., Mahé, B., and Expert, D. (2005). *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Mol. Microbiol.* 55(1), 261-275. doi: 10.1111/j.1365-2958.2004.04383.x.

Fujita, M.J., Kimura, N., Sakai, A., Ichikawa, Y., Hanyu, T., and Otsuka, M. (2011). Cloning and heterologous expression of the vibrioferrin biosynthetic gene cluster from a marine metagenomic library. *Biosci. Biotechnol. Biochem.* 75(12), 2283-2287. doi: 10.1271/bbb.110379.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14(7), 685-695. doi: 10.1093/oxfordjournals.molbev.a025808.

Giuliano Garisto Donzelli, B., Gibson, D.M., and Krasnoff, S.B. (2015). Intracellular siderophore but not extracellular siderophore is required for full virulence in *Metarhizium robertsii*. *Fungal Genet. Biol.* 82, 56-68. doi: 10.1016/j.fgb.2015.06.008.

Glasner, J.D., Yang, C.H., Reverchon, S., Hugouvieux-Cotte-Pattat, N., Condemine, G., Bohin, J.P., et al. (2011). Genome sequence of the plant-pathogenic bacterium *Dickeya dadantii* 3937. *J. Bacteriol.* 193(8), 2076-2077. doi: 10.1128/JB.01513-10.

Goering, A.W., McClure, R.A., Doroghazi, J.R., Albright, J.C., Haverland, N.A., Zhang, Y., et al. (2016). Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. *ACS Cent. Sci.* 2(2), 99-108. doi: 10.1021/acscentsci.5b00331.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42(D1), D699-D704. doi: 10.1093/nar/gkt1183.

Haas, H., Eisendle, M., and Turgeon, B.G. (2008). Siderophores in fungal physiology and virulence. *Annu. Rev. Phytopathol.* 46(1)**,** 149-187. doi: 10.1146/annurev.phyto.45.062806.094338.

Iftime, D., Kulik, A., Härtner, T., Rohrer, S., Niedermeyer, T.H.J., Stegmann, E., et al. (2016). Identification and activation of novel biosynthetic gene clusters by genome mining in the kirromycin producer *Streptomyces collinus* Tü 365. *J. Ind. Microbiol. Biotechnol.* 43(2)**,** 277-291. doi: 10.1007/s10295-015-1685-7.

Jarmusch, S.A., Lagos-Susaeta, D., Diab, E., Salazar, O., Asenjo, J.A., Ebel, R., et al. (2021). Iron-meditated fungal starvation by lupine rhizosphere-associated and extremotolerant *Streptomyces* sp. S29 desferrioxamine production. *Mol. Omics* 17(1)**,** 95-107. doi: 10.1039/d0mo00084a.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8(3)**,** 275-282. doi: 10.1093/bioinformatics/8.3.275.

Jørgensen, S.L., Stegger, M., Kudirkiene, E., Lilje, B., Poulsen, L.L., Ronco, T., et al. (2019). Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95. *mSphere* 4(1). doi: 10.1128/mSphere.00333-18.

Kadi, N., Arbache, S., Song, L., Oves-Costales, D., and Challis, G.L. (2008). Identification of a gene cluster that directs putrebactin biosynthesis in *Shewanella* species: PubC catalyzes cyclodimerization of N-hydroxy-N-succinylputrescine. *J. Am. Chem. Soc.* 130(32)**,** 10458-10459. doi: 10.1021/ja8027263.

Kadi, N., and Challis, G.L. (2009). "Siderophore biosynthesis: A substrate specificity assay for nonribosomal peptide synthetase-independent siderophore synthetases involving trapping of acyl-adenylate intermediates with hydroxylamine," in *Methods in Enzymology*. Academic Press), 431-457.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20(4)**,** 1160-1166. doi: 10.1093/bib/bbx108.

Khatiwada, B., Hasan, M.T., Sun, A., Kamath, K.S., Mirzaei, M., Sunna, A., et al. (2020). Proteomic response of *Euglena gracilis* to heavy metal exposure – Identification of key proteins involved in heavy metal tolerance and accumulation. *Algal Res.* 45**,** 101764. doi: 10.1016/j.algal.2019.101764.

Koch, R.A., Yoon, G.M., Aryal, U.K., Lail, K., Amirebrahimi, M., LaButti, K., et al. (2021). Symbiotic nitrogen fixation in the reproductive structures of a basidiomycete fungus. *Curr Biol* 31(17)**,** 3905-3914.e3906. doi: 10.1016/j.cub.2021.06.033.

Koppisch, A.T., Browder, C.C., Moe, A.L., Shelley, J.T., Kinkel, B.A., Hersman, L.E., et al. (2005). Petrobactin is the primary siderophore synthesized by *Bacillus anthracis* str. *Sterne* under conditions of iron starvation. *Biometals* 18(6)**,** 577-585. doi: 10.1007/s10534-005-1782-6.

Le Govic, Y., Papon, N., Le Gal, S., Lelièvre, B., Bouchara, J.-P., and Vandeputte, P. (2018). Genomic organization and expression of iron metabolism genes in the emerging pathogenic mold *Scedosporium apiospermum*. *Front. Microbiol.* 9(827). doi: 10.3389/fmicb.2018.00827.

Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25(7)**,** 1307-1320. doi: 10.1093/molbev/msn067.

Lee, J.Y., Janes, B.K., Passalacqua, K.D., Pfleger, B.F., Bergman, N.H., Liu, H., et al. (2007). Biosynthetic analysis of the petrobactin siderophore pathway from *Bacillus anthracis*. *J. Bacteriol.* 189(5)**,** 1698-1710. doi: 10.1128/jb.01526-06.

Li, B., Deng, X., Kim, S.H., Buhrow, L., Tomchick, D.R., Phillips, M.A., et al. (2021). Alternative pathways utilize or circumvent putrescine for biosynthesis of putrescine-containing rhizoferrin. *J. Biol. Chem.* 296**,** 100146. doi: 10.1074/jbc.RA120.016738.

Li, Y., Wang, Z., Liu, X., Song, Z., Li, R., Shao, C., et al. (2016). Siderophore biosynthesis but not reductive iron assimilation is essential for the dimorphic fungus *Nomuraea rileyi* conidiation, dimorphism transition, resistance to oxidative stress, pigmented microsclerotium formation, and virulence. *Front. Microbiol.* 7**,** 931. doi: 10.3389/fmicb.2016.00931.

Looney, B., Miyauchi, S., Morin, E., Drula, E., Courty, P.E., Kohler, A., et al. (2022). Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. *New Phytol.* 233(5)**,** 2294-2309. doi: 10.1111/nph.17892.

Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P.O., Crosa, J.H., et al. (2001). Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *J. Bacteriol.* 183(8)**,** 2576-2585. doi: 10.1128/JB.183.8.2576-2585.2001.

Ma, L.J., Ibrahim, A.S., Skory, C., Grabherr, M.G., Burger, G., Butler, M., et al. (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 5(7)**,** e1000549. doi: 10.1371/journal.pgen.1000549.

Mahe, B., Masclaux, C., Rauscher, L., Enard, C., and Expert, D. (1995). Differential expression of two siderophore-dependent iron-acquisition pathways in *Erwinia chrysanthemi* 3937: characterization of a novel ferrisiderophore permease of the ABC transporter family. *Mol. Microbiol.* 18(1)**,** 33-43. doi: 10.1111/j.1365-2958.1995.mmi_18010033.x.

Manwar, A.V., Khandelwal, S.R., Chaudhari, B.L., Meyer, J.M., and Chincholkar, S.B. (2004). Siderophore production by a marine *Pseudomonas aeruginosa* and its antagonistic action against phytopathogenic fungi. *Appl. Biochem. Biotechnol.* 118(1-3)**,** 243-251. doi: 10.1385/abab:118:1-3:243.

Matzanke, B.F., Anemüller, S., Schünemann, V., Trautwein, A.X., and Hantke, K. (2004). FhuF, Part of a Siderophore−Reductase System. *Biochemistry* 43(5)**,** 1386-1392. doi: 10.1021/bi0357661.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39(Web Server Issue)**,** W339-346. doi: 10.1093/nar/gkr466.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., et al. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* 11(9)**,** 625-631. doi: 10.1038/nchembio.1890.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E.L.L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49(D1)**,** D412-D419. doi: 10.1093/nar/gkaa913.

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sanchez-Garcia, M., et al. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat. Commun.* 11(1)**,** 5125. doi: 10.1038/s41467-020-18795-w.

Moon, Y.-H., Tanabe, T., Funahashi, T., Shiuchi, K.-i., Nakao, H., and Yamamoto, S. (2004). Identification and characterization of two contiguous operons required for aerobactin transport and biosynthesis in *Vibrio mimicus*. *Microbiol. Immunol.* 48(5)**,** 389-398. doi: 10.1111/j.1348-0421.2004.tb03528.x.

Morin, E., Kohler, A., Baker, A.R., Foulongne-Oriol, M., Lombard, V., Nagy, L.G., et al. (2012). Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc. Natl. Acad. Sci. U S A* 109(43)**,** 17501-17506. doi: 10.1073/pnas.1206847109.

Müller, K., Matzanke, B.F., Schünemann, V., Trautwein, A.X., and Hantke, K. (1998). FhuF, an iron-regulated protein of *Escherichia coli* with a new type of [2Fe-2S] center. *Eur. J. Biochem.* 258(3)**,** 1001-1008. doi: 10.1046/j.1432-1327.1998.2581001.x.

Narh Mensah, D.L., Wingfield, B.D., and Coetzee, M.P.A. (2023). Nonribosomal peptide synthetase gene clusters and characteristics of predicted NRPS-dependent siderophore synthetases in *Armillaria* and other species in the Physalacriaceae. *Curr. Genet.* 69(1), 7-24.

Nusca, T.D., Kim, Y., Maltseva, N., Lee, J.Y., Eschenfeldt, W., Stols, L., et al. (2012). Functional and structural analysis of the siderophore synthetase AsbB through reconstitution of the petrobactin biosynthetic pathway from *Bacillus anthracis*. *J. Biol. Chem.* 287(19)**,** 16058-16072. doi: 10.1074/jbc.M112.359349.

Okay, S., Yildirim, V., Buttner, K., Becher, D., and Ozcengiz, G. (2020). Dynamic proteomic analysis of *Phanerochaete chrysosporium* under copper stress. *Ecotoxicol. Environ. Saf.* 198**,** 110694. doi: 10.1016/j.ecoenv.2020.110694.

Oves-Costales, D., Kadi, N., and Challis, G.L. (2009). The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem. Comm.* (43)**,** 6530-6541. doi: 10.1039/b913092f.

Oves-Costales, D., Kadi, N., Fogg, M.J., Song, L., Wilson, K.S., and Challis, G.L. (2007). Enzymatic logic of anthrax stealth siderophore biosynthesis: AsbA catalyzes ATP-dependent condensation of citric acid and spermidine. *J. Am. Chem. Soc.* 129(27)**,** 8416-8417. doi: 10.1021/ja072391o.

Oves-Costales, D., Kadi, N., Fogg, M.J., Song, L., Wilson, K.S., and Challis, G.L. (2008). Petrobactin biosynthesis: AsbB catalyzes condensation of spermidine with N8-citryl-spermidine and its N1-(3,4-dihydroxybenzoyl) derivative. *Chem. Commun. (Camb)* (34)**,** 4034-4036. doi: 10.1039/b809353a.

Payne, S.M. (1994). "[25] Detection, isolation, and characterization of siderophores," in *Methods in Enzymology,* eds. V.L. Clark & P.M. Bavoil. Academic Press), 329-344.

Pfleger, B.F., Lee, J.Y., Somu, R.V., Aldrich, C.C., Hanna, P.C., and Sherman, D.H. (2007). Characterization and analysis of early enzymes for petrobactin biosynthesis in *Bacillus anthracis*. *Biochemistry* 46(13)**,** 4147-4157. doi: 10.1021/bi6023995.

Pirkkala, L., Nykanen, P., and Sistonen, L. (2001). Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J.* 15(7)**,** 1118-1131. doi: 10.1096/fj00-0294rev.

Ramakrishnan, G., Perez, N.M., Carroll, C., Moore, M.M., Nakamoto, R.K., and Fox, T.E. (2019). Citryl ornithine is an intermediate in a three-step biosynthetic pathway for rhizoferrin in *Francisella*. *ACS Chem. Biol.* 14(8)**,** 1760-1766. doi: 10.1021/acschembio.9b00297.

Rambaut, A. (2018). "FigTree v1.4.4".).

Ruiz-Dueñas, F.J., Barrasa, J.M., Sánchez-García, M., Camarero, S., Miyauchi, S., Serrano, A., et al. (2021). Genomic analysis enlightens Agaricales lifestyle evolution and increasing peroxidase diversity. *Mol. Biol. Evol.* 38(4)**,** 1428-1446. doi: 10.1093/molbev/msaa301.

Sabri, M., Leveille, S., and Dozois, C.M. (2006). A SitABCD homologue from an avian pathogenic *Escherichia coli* strain mediates transport of iron and manganese and resistance to hydrogen peroxide. *Microbiology (Reading)* 152(Pt 3)**,** 745-758. doi: 10.1099/mic.0.28682-0.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4(4)**,** 406-425. doi: 10.1093/oxfordjournals.molbev.a040454.

Schmelz, S., Kadi, N., McMahon, S.A., Song, L., Oves-Costales, D., Oke, M., et al. (2009). AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat. Chem. Biol.* 5(3)**,** 174-182. doi: 10.1038/nchembio.145.

Schwyn, B., and Neilands, J.B. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* 160(1)**,** 47-56. doi: 10.1016/0003-2697(87)90612-9.

Sipos, G., Prasanna, A.N., Walter, M.C., O'Connor, E., Bálint, B., Krizsán, K., et al. (2017). Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat. Ecol. Evol.* 1(12)**,** 1931-1941. doi: 10.1038/s41559-017-0347-8.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinf.* 24(5)**,** 637-644. doi: 10.1093/bioinformatics/btn013.

Stecher, G., Tamura, K., and Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) across computing platforms. *Mol. Biol. Evol.* 37(4)**,** 1237-1239. doi: 10.1093/molbev/msz312.

Sullivan, J.T., Jeffery, E.F., Shannon, J.D., and Ramakrishnan, G. (2006). Characterization of the siderophore of *Francisella tularensis* and role of fslA in siderophore production. *J. Bacteriol.* 188(11)**,** 3785-3795. doi: 10.1128/JB.00027-06.

Tanabe, T., Funahashi, T., Nakao, H., Miyoshi, S., Shinoda, S., and Yamamoto, S. (2003). Identification and characterization of genes required for biosynthesis and transport of the siderophore vibrioferrin in *Vibrio parahaemolyticus*. *J. Bacteriol.* 185(23)**,** 6938-6949. doi: 10.1128/jb.185.23.6938-6949.2003.

Tralamazza, S.M., Rocha, L.O., Oggenfuss, U., Correa, B., and Croll, D. (2019). Complex evolutionary origins of specialized metabolite gene cluster diversity among the plant pathogenic fungi of the *Fusarium graminearum* species complex. *Genome Biol. Evol.* 11(11)**,** 3106-3122. doi: 10.1093/gbe/evz225.

Wang, Q., Liu, Q., Ma, Y., Zhou, L., and Zhang, Y. (2007). Isolation, sequencing and characterization of cluster genes involved in the biosynthesis and utilization of the siderophore of marine fish pathogen *Vibrio alginolyticus*. *Arch. Microbiol.* 188(4)**,** 433-439. doi: 10.1007/s00203-007-0261-6.

Wolf, T., Shelest, V., Nath, N., and Shelest, E. (2016). CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics* 32(8)**,** 1138-1143. doi: 10.1093/bioinformatics/btv713.

Tables

Table 1: *Armillaria* spp. lifestyle, culture codes, source information, and CAS agar reactivity

Species	Lifestyle [a]	CMW number	Host	Original source	CAS agar reactivity [b]
A. fuscipes	Facultative necrotroph/ pathogenic	CMW2740	*Pinus elliottii*	Entabeni, South Africa	Orange, purplish-red, purple
		CMW3164	*Pelargonium asperum*	La Réunion	Orange, purplish-red, purple
A. gallica	Facultative necrotroph/ weakly pathogenic	CMW31092	N/A	Veneto, Bellune, Italy	Reddish-orange
		CMW45397	Duff	Minnesota, USA	Reddish-orange
A. luteobubalina	Facultative necrotroph/ pathogenic	CMW4974	N/A	Australia	Orange, purplish-red
		CMW4977	N/A	Australia	ND
A. mellea	Facultative necrotroph/ highly pathogenic	CMW31132	*Ailanthus altissima*	China	Orange, reddish-orange, purplish-red, purple
A. nabsnona	Weakly pathogenic	CMW3159	*Acer macrophyllum*	Vancouver, Canada	Reddish-orange, purplish-red, purple
		CMW6904	*Acer circinatum*	USA	Orange, purplish-red
Armillaria sp. ACB	N/A	CMW4456	*Brachystegia utilis*	Stapleford, Zimbabwe	Orange, purple

Modified from Narh Mensah et al. (2023)

[a]: Lifestyle information are summarized from Koch et al. (2017)

[b]: Colour change as determined by universal CAS and modified CAS agar assays based on images in Narh Mensah et al. (2023)

ACB = African Clade B in Coetzee et al. (2005); N/A = Not available; ND = Not determine

Table 2: CASSIS-determined NIS synthetase gene cluster boundaries, MIBiG comparison and ClusterBlast hits

Genome [a]	NIS Cluster 1			NIS Cluster 2		
	CASSIS gene cluster [b]	**MIBiG comparison**	**ClusterBlast hit [c]**	**CASSIS gene cluster [b]**	**MIBiG comparison [d]**	**ClusterBlast hit**
A. borealis (Armbor1 S7 and S46)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1	ND	NW_006267366 (1141250-1156888)	Ubiquitin-conjugating enzyme/RWD-like protein, Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and Glutamyl-tRNA synthetase	BGC0000944.1, 0.13, staphyloferrin A	ND
A. nabsnona (Armnabs1 S9 and S22)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1, Cytochrome P450	ND	NW_006267366 (1141250-1156888)	Hypothetical protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and Glutamyl-tRNA synthetase	BGC0000944.1, 0.13, staphyloferrin A	ND
A. cepistipes (Armcep1 S262 and S270)	*IucA/IucC*, Frag/DRAM/Sfk1	ND	NW_006267366 (1141250-1156888)	SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and Glutamyl-tRNA synthetase	ND	ND
A. mellea (Armme1 S3 and S10)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1	ND	NW_006267366 (1141250-1156888)	Ubiquitin-conjugating enzyme/RWD-like protein, Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and FAD/NAD-P-binding domain-containing protein	BGC0000944.1, 0.13, staphyloferrin A	ND
A. fumosa (Armfum1 S5 and S10)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1	ND	NW_006267366 (1141250-1156888)	Ubiquitin-conjugating enzyme/RWD-like protein, Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, and WH1-domain-containing protein	BGC0000944.1, 0.13, staphyloferrin A	ND
A. novae-zelandiae (Armnov1 S5 and S22)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1	ND	NW_006267366 (1141250-1156888)	N/A	N/A	N/A
D. ectypa ((Des)Armect1 S5 and S11)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter,	ND	NW_006267366 (1141250-1156888)	Ubiquitin-conjugating enzyme/RWD-like protein, Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing	BGC0000944.1, 0.13, staphyloferrin A	ND

	Frag/DRAM/Sfk1, Cytochrome P450			protein, and FAD/NAD-P-binding domain-containing protein		
D. tabescens ((Des)Armtab1 S1 and S55)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1	ND*	NW_006267366 (1141250-1156888)	Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and FAD/NAD-P-binding domain-containing protein related to salicylate hydroxylase	BGC0000944.1, 0.13, staphyloferrin A	ND
G. necrorhizus (Guyne1S3 and S49)	*IucA/IucC*, C2 domain-containing protein / MRP-like transporter, Frag/DRAM/Sfk1, hypothetical protein	ND	NW_006267366 (1141250-1156888)	Ubiquitin-conjugating enzyme/RWD-like protein, Endoglucanase V-like protein, SNARE domain-containing protein, *IucA/IucC*, WH1-domain-containing protein, and FAD/NAD-P-binding domain-containing protein	BGC0000944.1, 0.13, staphyloferrin A	ND
O. mucida (Oudmuc1 S1 and S66)	*IucA/IucC*, WD repeat-containing protein 8, MFS general substrate transporter	ND*	ND	small nucleolar RNA-associated protein 3, ribosomal protein S12, SNARE domain-containing protein, *IucA/IucC*, and three hypothetical proteins	BGC0000944.1, 0.13, staphyloferrin A	ND

[a]: Presented as species (genome code and scaffold (S) numbers on which NIS Clusters 1 and 2 are located respectively)

[b]: Genes in cluster boundary detected by CASSIS

[c]: 100% of genes in the hit show similarity to the genes in the query sequence (i.e., the sequence of NIS synthetase gene clusters of the Physalacriaceae)

[d]: Presented as Reference of hit in MIBiG database, similarity score, compound synthesized by the hit

ND = None detected, ND* = All hits had no similarity to the biosynthetic backbone gene, *IucA/IucC*; N/A = Not applicable (cluster detected by BLAST search in CLC)

Table 3: Information about the amino acid sequences of the NIS synthetases used for the Phylogenetic analysis and other characteristics of the putative NIS synthetase genes of the Physalacriaceae

NIS type [a]	Organism	Accession [b]	Protein	No of amino acids	Siderophore synthesized	Reference(s)
Type A	*Bacillus anthracis*	WP_000679659.1	AsbA	602	petrobactin	Lee *et al.* (2007)
Type A	*Dickeya dadantii* 3937	ADM97831.1	AcsD	620	achromobactin	(Glasner *et al.*, 2011; Schmelz *et al.*, 2009)
Type A	*Escherichia coli*	AAZ29624.1	IucA	575	aerobactin	(Carroll *et al.*, 2017; de Lorenzo *et al.*, 1986; Sabri *et al.*, 2006)
Type A	*Sinorhizobium meliloti*	WP_010968203.1	RhbC	585	rhizobactin	Lynch *et al.* (2001)
Type A	*Staphylococcus aureus*	AAP82067.1	SbnE	578	staphyloferrin B	(Cheung *et al.*, 2009; Dale *et al.*, 2004)
Type A	*Vibrio alginolyticus*	ABM30202.1	PvsD	609	vibrioferrin	(Carroll *et al.*, 2017; Fujita *et al.*, 2011; Wang *et al.*, 2007)
Type A'	*Francisella tularensis*	WP_003022888.1	FslA	639	rhizoferrin	(Carroll *et al.*, 2017; Li *et al.*, 2021; Sullivan *et al.*, 2006)
Type A'	*Legionella pneumophila*	AAZ39407.1	LbtA	580	legiobactin	Allard *et al.* (2006)
Type A'	*Rhizopus delemar* RA 99-880	I1C129.1	Rfs	634	rhizoferrin	(Carroll *et al.*, 2017; Ma *et al.*, 2009)
Type A'	*Staphylococcus aureus*	WP_072519492.1	SfaB	585	staphyloferrin A	Cotton *et al.* (2009)
Type A'	*Staphylococcus aureus*	WP_001052566.1	SfaD	658	staphyloferrin A	Cotton *et al.* (2009)
Type B	*Dickeya dadantii*	WP_253940872.1	AcsA	647	achromobactin	Franza *et al.* (2005)
Type B	*Vibrio alginolyticus*	BCB59206.1	PvsB	610	vibrioferrin	Tanabe *et al.* (2003)
Type B	*Staphylococcus argenteus*	WP_000556879.1	SbnC	584	staphyloferrin B	Cheung *et al.* (2009)
Type C	*Bacillus cereus*	WP_011053144.1	AsbB	612	petrobactin	(Nusca *et al.*, 2012; Oves-Costales *et al.*, 2008)
Type C	*Dickeya chrysanthemi*	WP_040000518.1	AcsC	618	achromobactin	(Carroll *et al.*, 2017; Kadi and Challis, 2009)
Type C	*Staphylococcus aureus*	AAP82068.1	SbnF	579	staphyloferrin B	(Cheung *et al.*, 2009; Dale *et al.*, 2004)
Type C'	*Escherichia coli*	AAS66995.1	IucC	580	aerobactin	(de Lorenzo and Neilands, 1986; Jørgensen *et al.*, 2019)
Type C'	*Shewanella chilikensis*	BCV36228.1	PubC	630	putrebactin	Kadi *et al.* (2008)
Type C'	*Streptomyces scabiei*	WP_013003384.1	DesD	590	desferrioxamine	Carroll *et al.* (2017)
C1	*A. borealis*	Armbor1_1990424 [S7]	NC	621	NC	This study
C1	*A. cepistipes*	Armcep1_13598 [S262]	NC	1132	NC	This study
C1	*A. fumosa*	Armfum1_800419 [S5]	NC	613	NC	This study
C1	*A. mellea*	Armmel1_1031929 [S3]	NC	613	NC	This study
C1	*A. nabsnona*	Armnab1_838196 [S9]	NC	621	NC	This study
C1	*A. novae-zelandiae*	Armnov1_1416980 [S5]	NC	614	NC	This study
C1	*D. ectypa*	Armect1_1382614 [S5]	NC	621	NC	This study
C1	*D. tabescens*	Armtab1_1497269 [S1]	NC	497	NC	This study
C1	*G. necrorhizus*	Guyne1_882414 [S3]	NC	621	NC	This study
Oudmuc1 C1	*O. mucida*	Oudmuc1_1247625 [S1]	IucA/IucC	602	NC	This study

O1	*Agaricus bisporus* var. *bisporus* H97	XP_006459224.1	NC	624	NC	Morin *et al.* (2012); This study
O1	*Mycena galopus* ATCC 62051	KAF8207121.1	IucC	597	NC	Miyauchi *et al.* (2020); This study
O1	*Pleurotus pulmonarius*	KAF4571474.1	IucA/IucC	612	NC	This study
C2	*A. borealis*	Armbor1_2060109 [S46]	IucA/IucC	555	NC	This study
C2	*A. cepistipes*	Armcep1_18630 [S270]	IucA/IucC	543	NC	This study
C2	*A. fumosa*	Armfum1_1394615 [S10]	IucA/IucC	555	NC	This study
C2	*A. mellea*	Armmel1_1066420 [S10]	IucA/IucC	537	NC	This study
C2	*A. nabsnona*	Armnab1_70081 [S10]	IucA/IucC	555	NC	This study
C2	*D. ectypa*	Armect1_1388155 [S11]	IucA/IucC	555	NC	This study
C2	*D. tabescens*	Armtab1_1340322 [S55]	IucA/IucC	556	NC	This study
C2	*G. necrorhizus*	Guyne1_910580 [S49]	IucA/IucC	559	NC	This study
Oudmuc1 C2	*O. mucida*	Oudmuc1_1223755 [S66]	NC	483	NC	This study
O2	*Agaricus bisporus* var. *bisporus* H97	XP_006459225.1	NC	631	NC	Morin *et al.* (2012); This study
O2	*Lactarius volemus*	KAH9972089.1	IucC	632	NC	Looney *et al.* (2022); This study
O2	*Lepista nuda*	KAF9468166.1	IucA/IucC	631	NC	This study

[a]: Presented as NIS synthetase type for known proteins, cluster in which the proteins are found in the Physalacriaceae (C1, C2, Oudmuc1 C1 and Oudmuc1 C2), and orthologous proteins (O1 and O2) identified by BLASTp searches in the NCBI database. C1 = NIS synthetase genes found in NIS Cluster 1, C2 = NIS synthetase genes found in NIS Cluster 2, Oudmuc1 C1 and Oudmuc1 C2 = NIS synthetase genes of *O. mucida* found in NIS Clusters 1 and 2 respectively, O1 and O2 = genes found to be orthologous to the NIS synthetase genes of the Physalacriaceae in NIS Clusters 1 and 2 respectively.

[b]: Accession numbers provided for known proteins or identified orthologs, Genome code followed by ProteinId from JGI and [scaffold number] for the Physalacriaceae

NC = not characterized

Figures

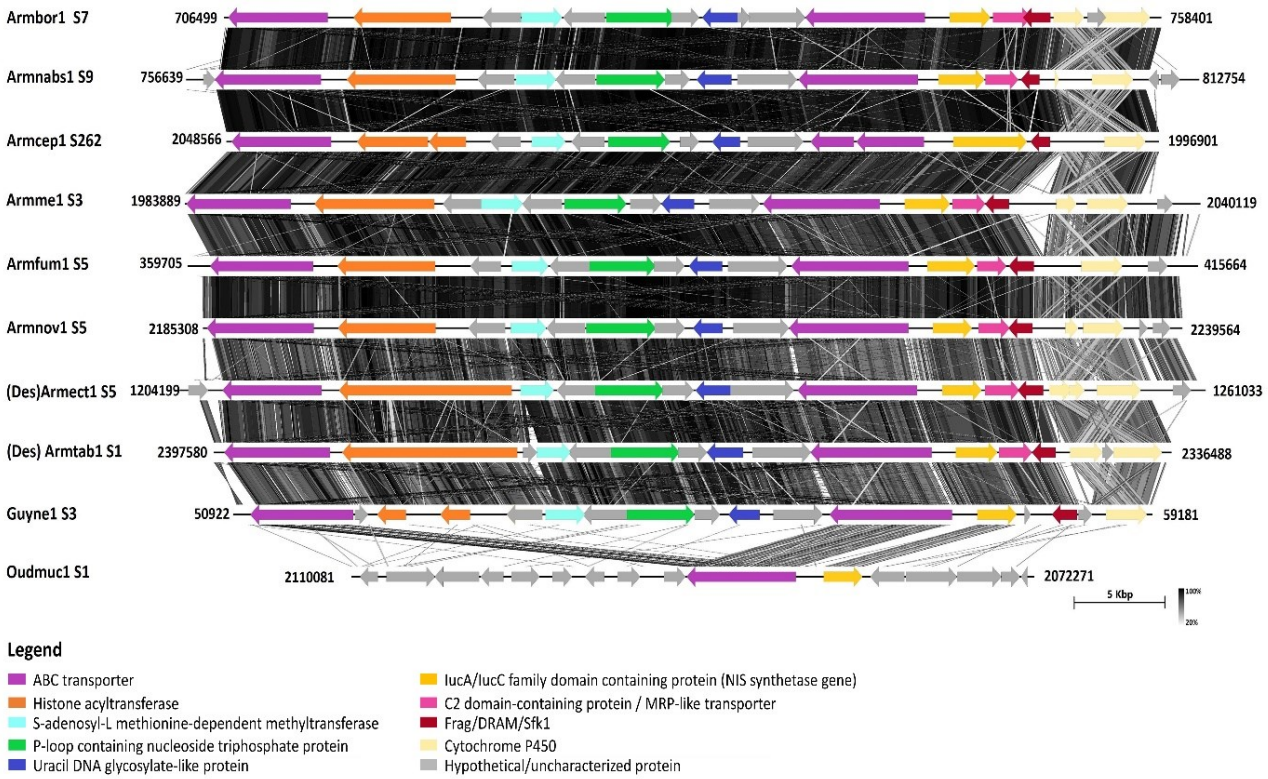


Figure 1: Synteny map of NIS Cluster 1 and neighbouring genes in annotated genomes of Physalacriaceae species

From top to bottom, NIS Cluster 1 in genomes of *A. borealis* (Armbor1), *A. nabsnona* (Armnabs1), *A. cepistipes* (Armcep1), *A. mellea* (Armme1), *A. fumosa* (Armfum1), *A. novae-zelandiae* (Armnov1), *Desarmillaria ectypa* ((Des)Armect1), *D. tabescens* ((Des)Armtab1), *Guyanagaster necrorhizus* (Guyne1), and *Oudemansiella mucida* (Oudmuc1) are presented. Numbers following the species code are the scaffolds (S) on which the clusters are located. Numbers at the ends of the clusters are the locations on the scaffolds. Different colours (different putative proteins as determined by tBLASTn searches) and orientation of arrows (direction of transcription) are shown. Orthologous genes are identically coloured. For Oudmuc1 S1, only orthologous genes are identically coloured. Darker shades of lines between clusters represent higher amino acid similarity between the respective clusters based on tBLASTx.

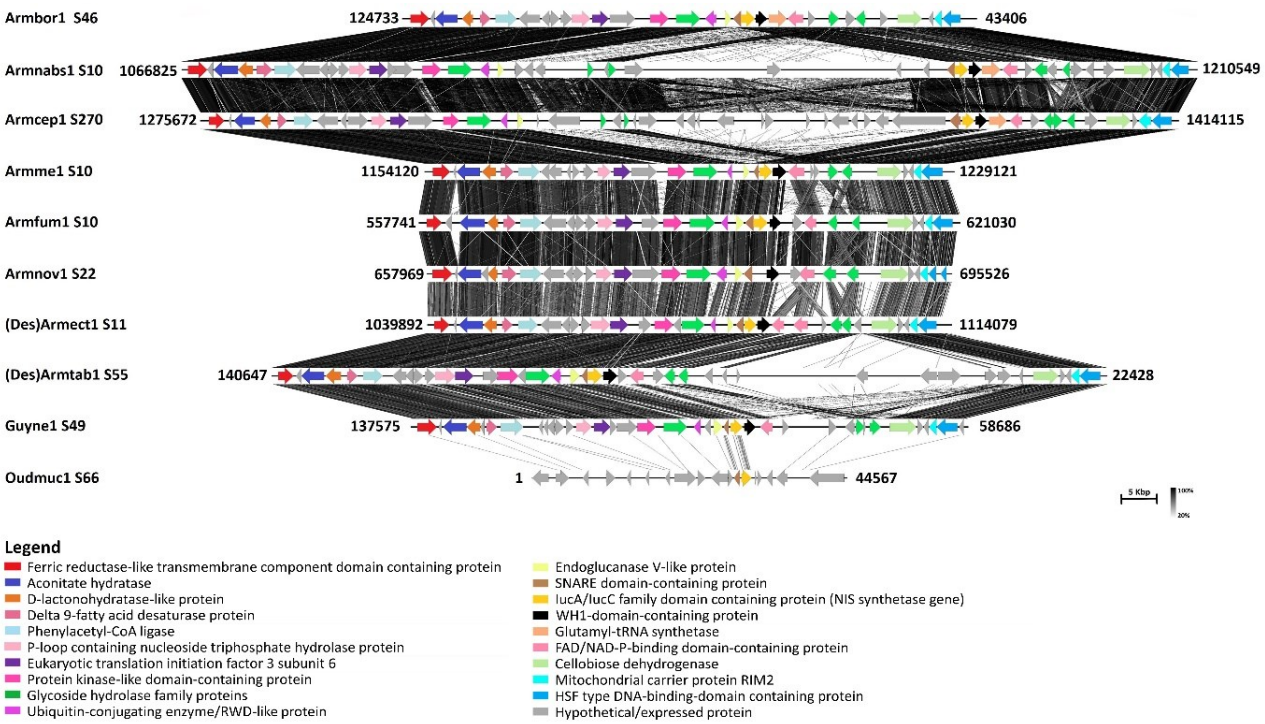


Figure 2: Synteny map of NIS Cluster 2 and neighbouring genes in annotated genomes of Physalacriaceae species

From top to bottom, NIS Cluster 2 in genomes of *A. borealis* (Armbor1), *A. nabsnona* (Armnabs1), *A. cepistipes* (Armcep1), *A. mellea* (Armme1), *A. fumosa* (Armfum1), *A. novae-zelandiae* (Armnov1), *D. ectypa* ((Des)Armect1), *D. tabescens* ((Des)Armtab1), *G. necrorhizus* (Guyne1), and *O. mucida* (Oudmuc1) are presented. Numbers following the species code are the scaffolds (S) on which the clusters are located. Numbers at the ends of the clusters are the locations on the scaffolds. Different colours (different putative proteins as determined by tBLASTn searches) and orientation of arrows (direction of transcription) are shown. Orthologous genes are identically coloured. Only the orthologous gene is identically coloured for Oudmuc1 S66. Darker shades of lines between clusters represent higher amino acid similarity between the respective clusters based on tBLASTx.

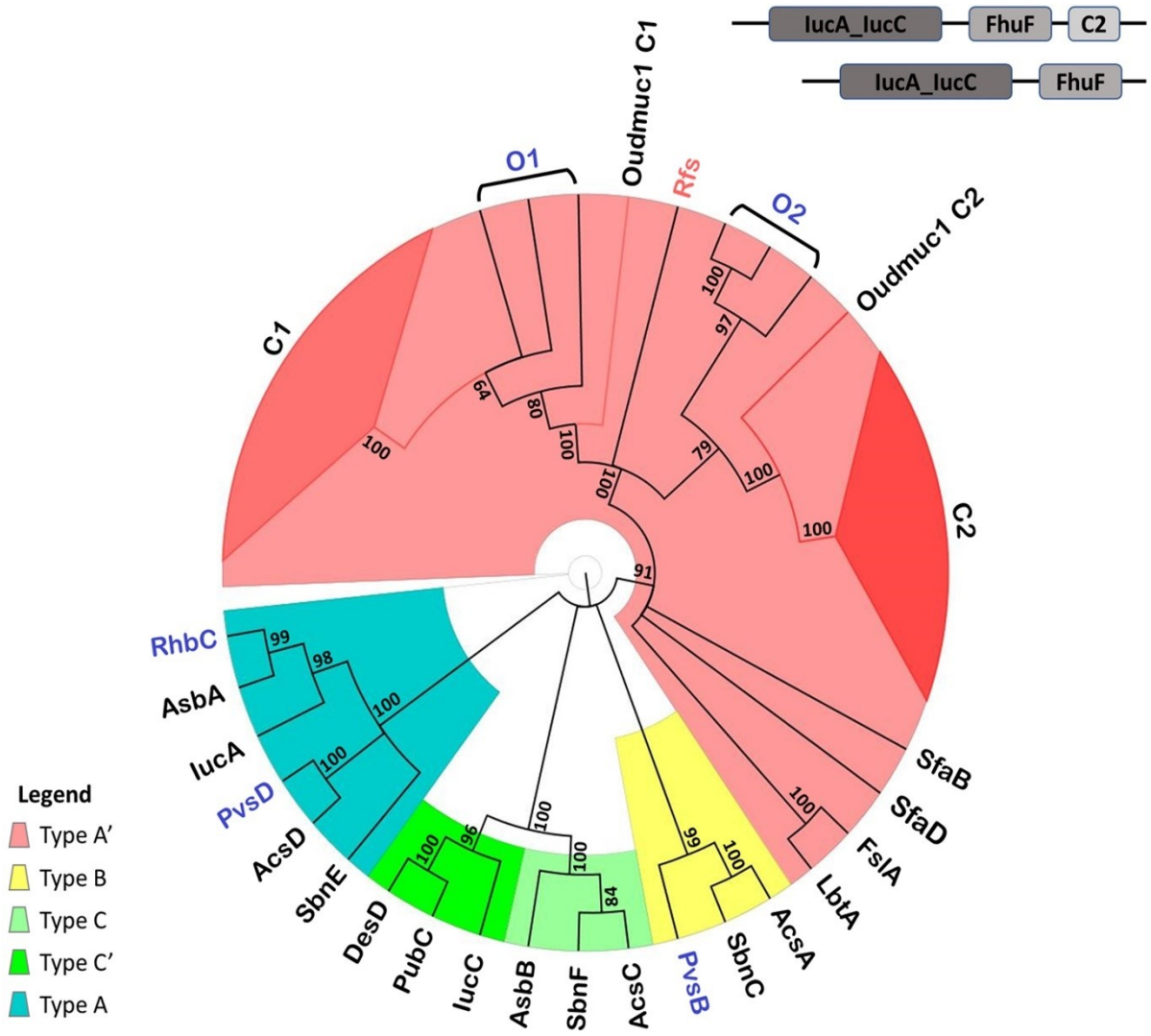


Figure 3: Cladogram showing the NIS synthetase genes in both NIS Clusters 1 and 2 grouping with known Type A' NIS synthetases and some uncharacterized orthologs in the Basidiomycota

Branches coloured red show NIS synthetases belonging to the Physalacriaceae. Subtrees of NIS synthetase genes in NIS Clusters 1 and 2 are drawn as cartoons showing C1 (deep rose) and C2 (red) respectively, all of which group together with known Type A' NIS synthetases (rose highlighted area). Oudmuc1 C1 and Oudmuc1 C2 are the NIS synthetases found in NIS Clusters 1 and 2 of *O. mucida*. O1 and O2, shown with brackets and blue texts, are orthologs obtained from BLASTp searches with protein sequences of the NIS synthetase genes in NIS Clusters 1 and 2 respectively, as well as the BLASTp hit of the similar cluster in *Agaricus bisporus* var. *bisporus* H97 obtained from fungiSMASH. The only characterized fungal NIS synthetase, Rfs, is indicated in rose text. Other protein names are indicated in black and blue text for characterized and uncharacterized NIS synthetases used in the dataset, respectively. Bootstrap values greater than 60 % are shown next to the nodes. Inserts at the top right corner are the Pfam domain architectures of all the NIS synthetases. Top insert is Pfam domain of the NIS synthetase of Armcep1 S262 (Armcep1_13598) showing the IucA_IucC, ferric iron reductase FhuF-like transporter, and C2 domains. Bottom insert is Pfam domain architectures of all other NIS synthetases in the cladogram, showing the IucA_IucC and ferric iron reductase FhuF-like transporter domains at the N- and C-terminals respectively.

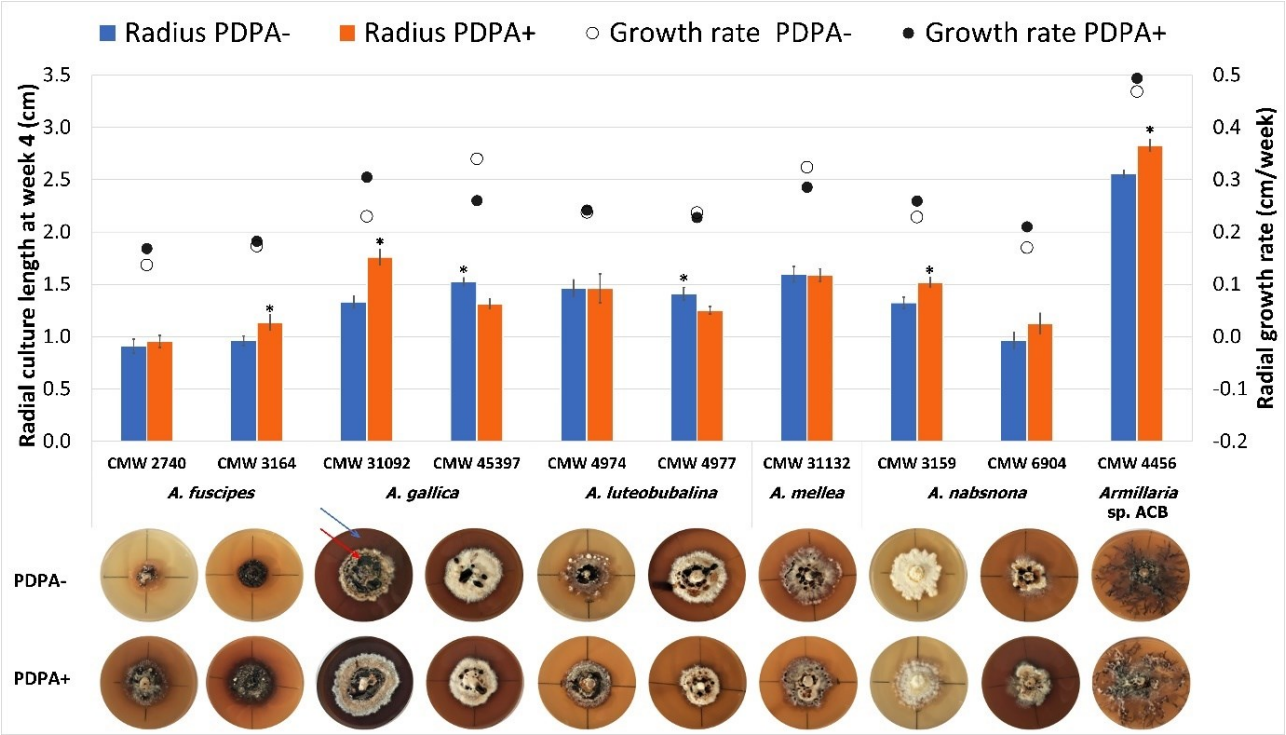


Figure 4: Iron-dependent mycelia growth on solid media.
PDPA+ and PDPA- are Potato Dextrose Peptone Agarose with and without added 100µM FeCl3 respectively Means of radial mycelia lengths are presented (n = 3). Error bars are standard errors of the means. Error bars marked with asterisks are significantly different ($p<0.05$) between treatments for the respective strains. Representative plates of strains grown on PDPA- (top row) and PDPA+ (bottom row) at end of incubation showing culture macromorphology and growth, as well as secretion of brownish exudates diffused into the media (blue arrow) or as liquid on cultures (red arrow) are also presented. Plates in each column correspond to the respective strain of *Armillaria* spp.

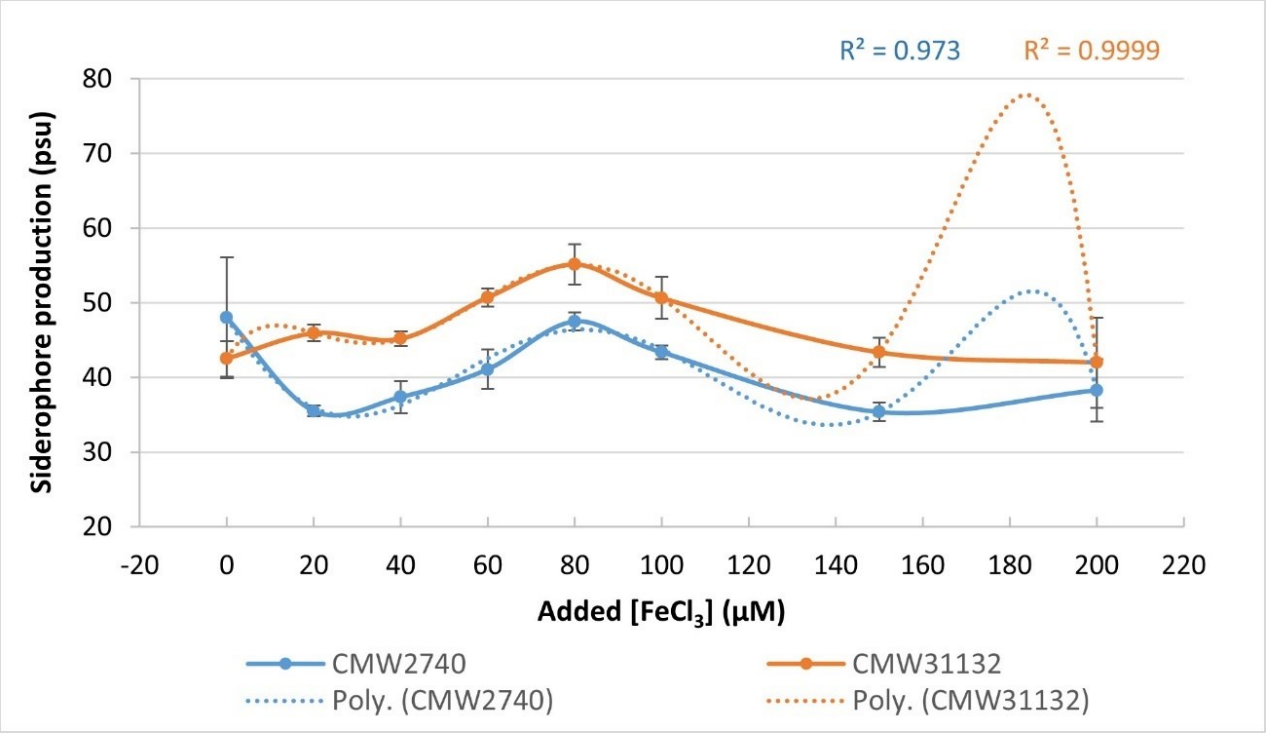


Figure 5: Iron-repressive siderophore biosynthesis by *A. fuscipes* strain CMW2740 and *A. mellea* strain CMW31132

Values presented are means of biological replicates. Error bars are standard errors of the means. R^2 values for CMW2740 (blue) and CMW31132 (brown) are indicated. Poly. = Polynomial regression with orders of 6.

Supplementary tables

Table S1: Source information for genomes analysed

Species	Strain or isolate	Genome code [a]	NCBI accession number or link to genome	Reference
Armillaria borealis	FPL87.14 v1.0	Armbor1	https://mycocosm.jgi.doe.gov/Armbor1/Armbor1.home.html	Grigoriev *et al.* (2014)
Armillaria cepistipes	B5	Armcep1	FTRY00000000.1	Sipos *et al.* (2017)
Armillaria fumosa	CBS 122221 v1.0	Armfum1	https://mycocosm.jgi.doe.gov/Armfum1/Armfum1.home.html	Grigoriev *et al.* (2014)
Armillaria mellea	ELDO17 v1.0	Armme1	https://mycocosm.jgi.doe.gov/Armmel1/Armmel1.home.html	Grigoriev *et al.* (2014)
Armillaria nabsnona	CMW6904 v1.0	Armnabs1	https://mycocosm.jgi.doe.gov/Armnab1/Armnab1.home.html	Grigoriev *et al.* (2014)
Armillaria novae-zelandiae	2840 v1.0	Armnov1	https://mycocosm.jgi.doe.gov/Armnov1/Armnov1.home.html	Grigoriev *et al.* (2014)
Desarmillaria ectypa	FPL83.16 v1.0	(Des)Armect1	https://mycocosm.jgi.doe.gov/Armect1/Armect1.home.html	Grigoriev *et al.* (2014)
Desarmillaria tabescens	CCBAS 213 v1.0	(Des)Armtab1	https://mycocosm.jgi.doe.gov/Armtab1/Armtab1.home.html	Grigoriev *et al.* (2014)
Cylindrobasidium torrendii	FP15055 v1.0	Cylto1	JYFH00000000.1	Floudas *et al.* (2015)
Guyanagaster necrorhizus	MCA 3950 v1.0	Guyne1	JAEACO000000000.1	Koch *et al.* (2021)
Oudemansiella mucida (*Mucidula mucida*)	CBS 558.79 v1.0	Oudmuc1	JADNYV000000000.1	Ruiz-Duenas *et al.* (2021)

[a]: (Des) has been included in the original genome code to indicate that these species belong to the genus *Desarmillaria.*

Table S2: Information about putative proteins coded by genes in NIS Clusters 1 and 2 of *A. borealis*

Protein annotation	Protein CDS name	Protein size [a]	Pfam accession	InterPro entry accession	Gene Ontology (GO)	Selected NCBI ortholog [b]
Genes in Armbor1_S7 cluster (NIS Cluster 1)						
ABC transporter	jgi.p|Armbor1|1899603	1400 (156.2)	PF00664, PF00005	IPR003439, IPR011527, IPR003593, IPR027417, IPR036640, IPR017871	GO:0055085, GO:0005524, GO:0140359, GO:0016887, GO:0016020	XM_043190246.1 (100, 80.71, 0.0)
Histone acyltransferase	jgi.p|Armbor1|1704428	1479 (163.6)	PF00628, PF17772, PF01853	IPR013083, IPR016181, IPR001965, IPR019787, IPR011011, IPR002717, IPR040706, IPR019786, IPR036388	GO:0006355, GO:0016573, GO:0004402	XM_037359142.1 (43, 50.23, 0.0)
S-adenosyl-L methionine-dependent methyltransferase	jgi.p|Armbor1|1778568	469 (52.3)	PF01189	IPR001678, IPR029063, IPR023267	GO:0001510, GO:0008168	XM_043185102.1 (96, 91.17, 0.0)
P-loop containing nucleoside triphosphate protein	jgi.p|Armbor1|1990336	868 (97.6)	PF13625, PF04851, PF16203	IPR001650, IPR027417, IPR032830, IPR014001, IPR032438, IPR006935, IPR001161	GO:0006367, GO:0006289, GO:0005524, GO:0016787, GO:0003677, GO:0003678	XM_041317970.1 (97, 82.72, 0.0)
Uracil DNA glycosylate-like protein	jgi.p|Armbor1|1778607	376 (41.3)	PF03167	IPR018085, IPR002043, IPR005122, IPR036895	GO:0006281, GO:0006284, GO:0016799, GO:0004844	XM_043189853.1 (100, 81.12, 0.0)
ABC transporter 1	jgi.p|Armbor1|1778635	1409 (157.0)	PF00664, PF00005	IPR003439, IPR011527, IPR003593, IPR027417, IPR036640, IPR017871	GO:0055085, GO:0005524, GO:0140359, GO:0016887, GO:0016020	XM_043180774.1 (99, 87.54, 0.0)
IucA/IucC family domain containing protein (NIS synthetase gene)	jgi.p|Armbor1|1990424	621 (69.4)	PF04183, PF06276	IPR037455, IPR007310, IPR022770	GO:0019290	XM_046234057.1 (95, 60.70, 0.0)
C2 domain-containing protein / MRP-like transporter	jgi.p|Armbor1|1303686	506 (54.8)	PF00168	IPR035892, IPR000008, IPR037791	N/A	XM_037359091.1 (35, 51.89, 1e-51)
Frag/DRAM/Sfk1	jgi.p|Armbor1|1303804	251 (28.4)	PF10277	IPR019402	N/A	XM_048024330.1 (98, 52.42, 4e-85)
Cytochrome P450	jgi.p|Armbor1|1732297	329 (37.2)	PF00067	IPR036396, IPR017972, IPR002401, IPR001128	GO:0016705, GO:0005506, GO:0020037, GO:0004497	XM_043189314.1 (76, 50.00, 3e-79)
Cytochrome P450 1	jgi.p|Armbor1|1778663	502 (56.5)	PF00067	IPR036396, IPR017972, IPR002401, IPR001128	GO:0016705, GO:0005506, GO:0020037, GO:0004497	XM_043189314.1 (93, 75.80, 0.0)
Genes in Armbor1_S46 Cluster (NIS Cluster 2)						

Ferric reductase-like transmembrane component domain containing protein	jgi.p|Armbor1|996651	623 (67.6)	PF01794, PF08022	IPR013130, IPR017927, IPR039261, IPR017938, IPR013112	GO:0016491	XM_046228015.1 (94, 54.58, 0.0)
Aconitate hydratase	jgi.p|Armbor1|1827031	784 (84.6)	PF00330, PF00694	IPR015932, IPR015928, IPR015931, IPR001030, IPR036008, IPR000573, IPR018136, IPR006248	GO:0006099, GO:0051539, GO:0003994	XM_043181510.1 (100, 96.43, 0.0)
D-lactonohydratase-like protein	jgi.p|Armbor1|1907016	388 (41.7)	PF08450	IPR011042, IPR013658	N/A	XM_043187440.1 (100, 83.13, 0.0)
Delta 9-fatty acid desaturase protein	jgi.p|Armbor1|1907015	416 (47.0)	PF00487, PF00173	IPR001199, IPR001522, IPR036400, IPR005804, IPR015876, IPR018506	GO:0006629, GO:0016717, GO:0020037	XM_001879546.1 (97, 72.35, 0.0)
Phenylacetyl-CoA ligase	jgi.p|Armbor1|1826986	578 (63.6)	PF00501, PF13193	IPR000873, IPR042099, IPR045851, IPR025110	N/A	XM_043188395.1 (97, 92.53, 0.0)
P-loop containing nucleoside triphosphate hydrolase protein	jgi.p|Armbor1|995985	611 (68.7)	PF08740, PF00004	IPR014851, IPR027417, IPR003593, IPR003959, IPR003960	GO:0016887, GO:0005524	XM_043177958.1 (91, 73.10, 0.0)
Eukaryotic translation initiation factor 3 subunit 6	jgi.p|Armbor1|2010755	606 (69.0)	PF10255	IPR019382	GO:0003743, GO:0005852, GO:0005737	XM_001834986.1 (99, 76.28, 0.0)
Protein kinase-like domain-containing protein	jgi.p|Armbor1|1826934	508 (57.8)	PF00069, PF00433	IPR011009, IPR017892, IPR000719, IPR000961, IPR008271, IPR017441	GO:0006468, GO:0004674, GO:0005524, GO:0004672	XM_043187439.1 (94, 93.14, 0.0)
Glycoside hydrolase family 95 protein	jgi.p|Armbor1|2021490	855 (93.2)	PF14498	IPR008928, IPR012341, IPR027414	GO:0005975	XM_043190676.1 (95, 82.91, 0.0)
Ubiquitin-conjugating enzyme/RWD-like protein	jgi.p|Armbor1|995512	229 (25.6)	PF00179	IPR016135, IPR000608	N/A	XM_027759635.1 (80, 64.32, 8e-79)
Endoglucanase V-like protein	jgi.p|Armbor1|1892064	183 (18.5)	N/A	IPR036908, IPR007112	N/A	XM_007382807.1 (90, 67.82, 2e-71)
SNARE domain-containing protein	jgi.p|Armbor1|995449	206 (23.1)	PF05739	IPR000727	N/A	XM_037362815.1 (90, 40.74, 2e-25)
IucA/IucC family domain containing protein (NIS synthetase gene)	jgi.p|Armbor1|2060109	555 (61.1)	PF04183, PF06276	IPR037455, IPR007310, IPR022770	GO:0019290	XM_046226752.1 (77, 39.53, 9e-106)

WH1-domain-containing protein	jgi.p|Armbor1|2038045	455 (47.3)	PF00786	IPR036936, IPR003124, IPR000697, IPR011993, IPR000095	GO:0003779	XM_050346853.1 (43, 51.60, 1e-60)
Glutamyl-tRNA synthetase	jgi.p|Armbor1|1892062	682 (76.7)	PF00749, PF03950	IPR020061, IPR020059, IPR020058, IPR020056, IPR011035, IPR014729, IPR000924, IPR036282, IPR004526, IPR001412	GO:0006418, GO:0043039, GO:0006412, GO:0006424, GO:0005524, GO:0000166, GO:0004812, GO:0004818, GO:0005737	XM_027756926.1 (98, 55.45, 0.0)
FAD/NAD-P-binding domain-containing protein	jgi.p|Armbor1|1973252	429 (46.7)	PF01494	IPR002938, IPR036188	GO:0071949	XM_043189205.1 (99, 76.54, 0.0)
Glycoside hydrolase family 61 protein	jgi.p|Armbor1|2085613	291 (30.4)	PF03443	IPR005103	N/A	XM_043181506.1 (100, 74.91, 1e-156)
Glycoside hydrolase family 61a protein	jgi.p|Armbor1|1973221	321 (32.9)	PF03443	IPR005103	N/A	XM_043181506.1 (98, 56.85, 4e-114)
Cellobiose dehydrogenase	jgi.p|Armbor1|1892058	715 (76.5)	PF16010, PF13450, PF00732, PF05199	IPR015920, IPR000172, IPR036188, IPR007867	GO:0050660, GO:0016614	XM_043181505.1 (98, 77.54, 0.0)
Mitochondrial carrier protein RIM2	jgi.p|Armbor1|1892056	317 (34.0)	PF00153, PF00153, PF00153	IPR023395, IPR018108, IPR002067	GO:0055085	XM_043183799.1 (96, 90.23, 0.0)
HSF type DNA-binding-domain containing protein	jgi.p|Armbor1|1714821	672 (74.5)	PF00447, PF00072	IPR011006, IPR000232, IPR036388, IPR027725, IPR001789	GO:0006355, GO:0000160, GO:0003700, GO:0043565	XM_048025672.1 (75, 68.08, 3e-105)

Information is presented in the order of gene appearance from left to right on the respective synteny maps (Figure 1 and Figure 2). Information about hypothetical/uncharacterized proteins are excluded from the table. All information excluding NCBI information was obtained from InterPro. The genes in the clusters reported for Armbor1 are representative of the genes in the clusters of the other genomes studied. N/A = none available.

[a]: Presented as number of amino acids (weight in kDa). Protein weight of protein coding sequences (CDS) of the respective genes obtained from annotated genomes in CLC Main Workbench was estimated using the 'Create Protein Report' tool in CLC Main Workbench to one decimal point.

[b]: NCBI ortholog information was obtained from tBLASTn searched using the CDSs of the respective genes. Data is presented as Accession (Query Cover, Percentage Identity, Expected value). Query Cover and Percentage Identity are in %.

Table S3: ClusterBlast details of Armbor1 S7 cluster showing 100 % gene similarity with NW_006267366 in *Agaricus bisporus* var. *bisporus* H97

Armbor1 S7 CASSIS cluster information (21,784 - 33,963 nt.)		BLASTP Hit NW_006267366 (1141250-1156888) genes				BLASTP Ortholog			
Putative protein name	**Location**	**Protein Id**	**Percentage coverage (Percentage identity)**	**E-value**	**NCBI Conserved Domain hits**	**Protein name [Organism]**	**Protein ID**	**Percentage coverage (Percentage identity)**	**E-value**
IucA/IucC family domain containing protein	26783 - 28963	XP_006459225.1	99 (44)	2.1e-146	COG4264, Pfam04183, NF033586	siderophore [*Pleurotus pulmonarius*]	KAF4571475.1	99 (68.42)	0.0
		XP_006459224.1	99 (57)	1.8e-198	COG4264, pfam04183, NF033583	IucC family-domain-containing protein [*Lepista nuda*]	KAF9468165.1	98 (60.29)	0.0
C2 domain-containing protein / MRP-like transporter	29243 - 30866	XP_006459223.1	50 (50)	1.3e-56	Cd08681, pfam00168, smart00239, PHA03247, COG5038	C2 domain-containing protein [*Lepista nuda*]	KAF9468164.1	36 (65.43)	7e-74
Frag/DRAM/Sfk1	31062 - 32138	XP_006459222.1	92 (63)	1.7e-78	Pfam10277, COG0474	Protein sfk1 [*Leucoagaricus* sp. SymC.cos]	KXN81326.1	89 (78.60)	2e-107

Acknowledgments

Permission granted by Dr. László G. Nagy and his team at Biological Research Centre, Synthetic and Systems Biology Unit, 6726 Szeged, Hungary to use the genome and RNA sequence data of *A. borealis* strain FPL87.14, *A. fumosa* strain CBS 122221, *A. mellea* strain ELDO17, *A. nabsnona* strain CMW6904, *A. novae-zelandiae* strain 2840, *D. ectypa* strain FPL83.16, *D. tabescens* strain CCBAS 213, and *G. necrorhizus* strain MCA 3950 v1.0 is highly appreciated. We also thank all the funding bodies for financial assistance to carry out this work.

Data availability

Publicly available genome and RNA sequences were analysed in this study. These data can be found at https://mycocosm.jgi.doe.gov/mycocosm/species-tree/tree;05h0Ue?organism=physalacriaceae. Unpublished genome and RNA sequence data obtained from JGI were used with permission from Dr. László G. Nagy. Data of bioassays are presented in this manuscript and can be assessed in the manuscript.

CHAPTER 4

DRAFT GENOME SEQUENCE OF AN *ARMILLARIA* SPECIES FROM ZIMBABWE

This chapter has been published in IMA Fungus as:

Narh Mensah, D.L., Wingfield, B.D., Maphosa, M., Duong, T.A., Coetzee, M.P.A. 2022. IMA genome-F17A Draft genome sequence of an *Armillaria* species from Zimbabwe. *IMA Fungus*, **13**(19), 1-3. https://doi.org/10.1186/s43008-022-00104-3

1. Introduction

The genus *Armillaria* includes at least 38 species, most of which are facultative necrotrophs (Gregory and Rishbeth 1991). Pathogenicity of these organisms can result in *Armillaria* root and stem rot and what is referred to as shoestring root rot (Morrison 1991). This disease can bring about massive devastation to woody plants grown for horticulture, agriculture, as well as natural and managed forests across the various continents (Baumgartner and Rizzo 2001, 2002; Guillaumin et al. 1993; Labbé et al. 2015). The saprophytic nature of some *Armillaria* spp. results in enhancement of forest ecosystems through the breakdown of woody material, resulting in carbon and mineral cycling (Baumgartner et al. 2011; Heinzelmann et al. 2019). Transition from a saprophytic to a pathogenic lifestyle, and vice versa, can occur due to intra-species variation, forest management systems, the state of the host (e.g., stressed or healthy), as well as environmental factors (e.g., elevation) (Legrand et al. 1996; Prospero et al. 2004; Tsykun et al. 2012).

Various groups have conducted omics-based research on *Armillaria* species (Akulova et al. 2020; Anderson et al. 2018; Caballero et al. 2022; Collins et al. 2017; Collins et al. 2013; Heinzelmann et al. 2020; Kolesnikova et al. 2019; Linnakoski et al. 2021; Misiek et al. 2011; Misiek and Hoffmeister 2012; Sipos et al. 2017; Sonnenbichler et al. 1997; Sun et al. 2020; Wingfield et al. 2016a, b; Zhan et al. 2020). These genomics, proteomics and metabolomics studies were done to gain insight into the molecular mechanisms and biochemical properties that drive the pathogenicity and virulence of *Armillaria* spp. This information would eventually help to develop efficient strategies for identifying these fungi, containing their spread, and minimising damage to forest ecosystems.

The previously determined nuclear and mitochondrial genomes of various *Armillaria* species are providing invaluable resources for genome-based research (Table 1). Studies using these genomes have broadened our understanding of the biology of the *Armillaria* species and the evolution of their genomes. For example, Sipos et al. (2017) showed that genome evolution in the genus was predominantly caused by gene family expansion. Kolesnikova et al. (2019) assembled the complete mitochondrial genomes of *A. borealis, A. gallica, A. sinapina,* and *A. solidipes* and found a high degree of variation in size, gene content and genomic organization among these phylogenetically closely related species. Recently, the first chromosome-level *Armillaria* genome assembly became available, revealing genome-wide recombination in the genome of *A. ostoyae* (Heinzelmann et al. 2020).

Sequenced *Armillaria* species originate primarily in the Northern Hemisphere. The genome of only one species from the Southern Hemisphere, *A. fuscipes* from South Africa, has so far been published

(Wingfield et al. 2016a, b). It is known that species from the Northern Hemisphere, Australasia together with Southern America, and Africa, respectively, reside in distinct monophyletic lineages (Coetzee et al. 2011; Koch et al. 2017). Genomes of species in these geographic locations may, therefore, have followed very different evolutionary pathways. Within the African clade, Coetzee et al. (2005) identified two lineages, referred to as *A. fuscipes* and African Group B. Here, we report the genome of an *Armillaria* isolate belonging to African Group B (sensu Coetzee et al. 2005), sequenced using both long- and short-read technologies. This genome expands the sequence resources for *Armillaria* species from the Southern Hemisphere and Africa.

Sequenced strain

Zimbabwe: Stapelford, Manicaland isolated from *Brachystegia utilis,* 2001, *E. Mwenje* (culture CMW4456; PREM 63337—dried culture).

Nucleotide sequence accession number

The Whole Genome Shotgun project of the *Armillaria* sp. genome has been deposited at DDBJ/ENA/GenBank under the accession JANDKJ000000000. The version described in this paper is version JANDKJ010000000.

2. Materials and methods

The culture of isolate CMW4456 were grown and maintained in MYA (1.5% Malt extract, 0.2% Yeast extract, 1.5% Agar) at 24 °C in the dark for 4 weeks. DNA was extracted from the harvested mycelia using the method described by Duong et al. (2013). PacBio sequencing was conducted on the Sequel IIe system using the circular consensus sequencing (CCS) mode at Inqaba Biotechnical Industries (Pty) Ltd. (Pretoria, South Africa).

For short read sequencing on the Illumina HiSeq platform, genomic DNA was extracted from cultures grown in MY broth (1.5% Malt extract, 0.2% Yeast extract) for six weeks at 24 °C in the dark. Harvested cultures were kept at − 80 °C, followed by lyophilization. DNA was extracted with the Qiagen DNEasy Plant Pro Kit (50) Cat. No. 69204 (Qiagen, Sandton, South Africa) following the manufacturer's instructions. Illumina paired-end library preparation and whole-genome sequencing was done with an insert size of 350 bp and read-length of 150 bp at Macrogen.

Trimmomatic v. 0.38 (Bolger et al. 2014) was used to trim adapter sequences and low-quality ends of the Illumina reads (ILLUMINACLIP, TruSeq3-PE.fa:2:30:10:8; LEADING, 3; TRAILING, 3; MINLEN, 30).

The PacBio HiFi reads were assembled with CLC Genomics Workbench v 22.0.1 (QIAGEN, Aarhus). The assembly was subsequently polished with the trimmed Illumina HiSeq reads, using

Pilon v. 1.23 (Walker et al. 2014). Genome completeness was evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.3.2, using the agaricales_odb10 lineage dataset (Manni et al. 2021). AUGUSTUS v. 3.4.0 (Keller et al. 2011; Stanke et al. 2008; Stanke and Waack 2003) was used to predict protein coding genes, applying gene models of the closely related species, *Coprinus cinereus*. QUAST v 5.0.2 (Gurevich et al. 2013) was used to evaluate metrics, including contig number, total length, GC content, and N50 for the genome assemblies. BUSCO, AUGUSTUS and QUAST were run using the Galaxy platform (Afgan et al. 2018; The Galaxy Community 2022) (https://usegalaxy.eu/). *Armillaria ostoyae* strain C18/9 genome (Sipos et al. 2017) with accession number FUEG00000000.1 was used as the reference genome for genome quality evaluation.

The identity of the *Armillaria* African Clade B isolate CMW4456 for which a genome was sequenced was confirmed based on phylogenetic grouping with published DNA sequences. DNA sequences from the internal transcribed spacer region (ITS) and the translation elongation factor one alpha (tef1-α) were extracted from the genome. Since few tef1-α sequences from African isolates are available in databases, the tef1-α sequence was compared to sequences on GenBank using BLASTn. The ITS sequence was included in the data matrix of Coetzee et al. (2005) and aligned using the online version of MAFFT v. 7. (Katoh et al. 2019). The TrN + G nucleotide substitution model was determined as best fitting the sequence alignment, using jModelTest and the Akaike Information Criterion (Darriba et al. 2012; Guindon and Gascuel 2003), and incorporated in the maximum likelihood analyses. A maximum likelihood phylogenetic tree was constructed using PHYML v. 3.0 (Guindon et al. 2010), applying 1000 bootstrap replicates. The tree was rooted with sequences of *A. hinnulea*.

3. Results and discussion

The 4 PacBio read length ranged between 289 and 15627 bases. The 2 × 151 bp Illumina HiSeq paired-end libraries yielded a total of 14,927,540,182 reads, amounting to 98,857,882 nucleotides. The PacBio and Illumina reads were assembled into 840 contigs with a total assembly size of 54.95 Mbp. All contigs were longer than 1000 bp, with the largest contig being 1,463,441 bp. The N50 and N75 values were 128,967 bp and 45,059 bp, respectively. The L50 and L75 values were 85 and 270 contigs, respectively. The GC content of the assembled genome was 46.53%. Genome completeness was estimated to be 98%, corresponding to 96.8% complete and single-copy BUSCOs, 1.2% complete and duplicated BUSCOs, 0.1% fragmented BUSCOs, and 1.9% missing BUSCOs (n = 3870). AUGUSTUS predicted 13,600 protein coding genes.

The genome statistics of the sequenced *Armillaria* strain correlated with that reported for the genomes of other species in the genus (Table 1). The assembly size fell within the range of 53.00–73.74 Mbp,

though the number of predicted protein coding genes (13,600 genes) was somewhat lower than the 14,473–26,261 genes reported in the assembled genomes of other species of *Armillaria.* The GC contents of the genomes of other *Armillaria* spp. (47.4–49.1%) are similar to the GC content of 46.53% reported here.

The sequenced genome of *Armillaria* sp. strain CMW4456 grouped with other strains of the African *Armillaria* Clade B from Cameroon, Zambia and Zimbabwe (Coetzee et al. 2005), confirming its identity (Fig. 1). The tef1-α sequence from the genome was identical to the tef1-α sequence of CMW4456 on GenBank (accession number DQ435617.1). This genome sequence will serve as a useful resource for investigating the biology, chemistry, and pathogenicity of *Armillaria* species from Africa in comparison to those from other continents.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, **46**(W1), W537-W544.

Akulova, V.S., Sharov, V.V., Aksyonova, A.I., Putintseva, Y.A., Oreshkova, N.V., Feranchuk, S.I., Kuzmin, D.A., Pavlov, I.N., Litovka, Y.A., Krutovsky, K.V. 2020. *De novo* sequencing, assembly and functional annotation of *Armillaria borealis* genome. *BMC Genomics*, **21**(7), 534.

Anderson, J.B., Bruhn, J.N., Kasimer, D., Wang, H., Rodrigue, N., Smith, M.L. 2018. Clonal evolution and genome stability in a 2500-year-old fungal individual. *Proc. Biol. Sci.*, **285**(1893), 20182233.

Baumgartner, K., Coetzee, M.P.A., Hoffmeister, D. 2011. Secrets of the subterranean pathosystem of *Armillaria*. *Mol. Plant Pathol.*, **12**(6), 515-534.

Baumgartner, K., Rizzo, D.M. 2001. Ecology of *Armillaria* spp. in mixed-hardwood forests of California. *Plant Dis.*, **85**(9), 947-951.

Baumgartner, K., Rizzo, D.M. 2002. Spread of *Armillaria* root disease in a California vineyard. *Am. J. Enol. Vitic.*, **53**(3), 197-203.

Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15), 2114-2120.

Caballero, J.R.I., Lalande, B.M., Hanna, J.W., Klopfenstein, N.B., Kim, M.S., Stewart, J.E. 2022. Genomic comparisons of two *Armillaria* species with different ecological behaviors and their associated soil microbial communities. *Microb. Ecol.*

Coetzee, M.P.A., Bloomer, P., Wingfield, M.J., Wingfield, B.D. 2011. Paleogene radiation of a plant pathogenic mushroom. *PLoS One*, **6**(12), e28545.

Coetzee, M.P.A., Wingfield, B.D., Bloomer, P., Wingfield, M.J. 2005. Phylogenetic analyses of DNA sequences reveal species partitions amongst isolates of *Armillaria* from Africa. *Mycol. Res.*, **109**(11), 1223-1234.

Collins, C., Hurley, R., Almutlaqah, N., O'Keeffe, G., Keane, T.M., Fitzpatrick, D.A., Owens, R.A. 2017. Proteomic characterization of *Armillaria mellea* reveals oxidative stress response mechanisms and altered secondary metabolism profiles. *Microorganisms*, **5**(3), 60.

Collins, C., Keane, T.M., Turner, D.J., O'Keeffe, G., Fitzpatrick, D.A., Doyle, S. 2013. Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. *J. Proteome Res.*, **12**(6), 2552-2570.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*, **9**(8), 772-772.

Duong, T.A., de Beer, Z.W., Wingfield, B.D., Wingfield, M.J. 2013. Characterization of the mating-type genes in *Leptographium procerum* and *Leptographium profanum*. *Fungal Biol.*, **117**(6), 411-421.

Gregory, S.C., Rishbeth, J. 1991. Pathogenicity and Virulence. in: *Agriculture Handbook (USA)*, (Eds.) C.G. Shaw III, G.A. Kile, Vol. No. 691, USDA Forest Service. Washington, District of Columbia, pp. 76–87.

Guillaumin, J.-J., Mohammed, C., Anselmi, N., Courtecuisse, R., Gregory, S.C., Holdenrieder, O., Intini, M., Lung, B., Marxmüller, H., Morrison, D., Rishbeth, J., Termorshuizen, A.J., Tirró, A., van Dam, B. 1993. Geographical distribution and ecology of the *Armillaria* species in western Europe. *Eur. J. For. Pathol.*, **23**(6-7), 321-341.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**(3), 307-21.

Guindon, S., Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**(5), 696-704.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**(8), 1072-1075.

Heinzelmann, R., Dutech, C., Tsykun, T., Labbé, F., Soularue, J.-P., Prospero, S. 2019. Latest advances and future perspectives in *Armillaria* research. *Can. J. Plant Pathol.*, **41**(1), 1-23.

Heinzelmann, R., Rigling, D., Sipos, G., Münsterkötter, M., Croll, D. 2020. Chromosomal assembly and analyses of genome-wide recombination rates in the forest pathogenic fungus *Armillaria ostoyae*. *Heredity*, **124**(6), 699-713.

Katoh, K., Rozewicki, J., Yamada, K.D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*, **20**(4), 1160-1166.

Keller, O., Kollmar, M., Stanke, M., Waack, S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, **27**(6), 757-763.

Koch, R.A., Wilson, A.W., Séné, O., Henkel, T.W., Aime, M.C. 2017. Resolved phylogeny and biogeography of the root pathogen *Armillaria* and its gasteroid relative, *Guyanagaster*. *BMC Evol. Biol.*, **17**(1), 33.

Kolesnikova, A.I., Putintseva, Y.A., Simonov, E.P., Biriukov, V.V., Oreshkova, N.V., Pavlov, I.N., Sharov, V.V., Kuzmin, D.A., Anderson, J.B., Krutovsky, K.V. 2019. Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related *Armillaria* species. *BMC Genomics*, **20**(1), 351.

Labbé, F., Marcais, B., Dupouey, J.-L., Bélouard, T., Capdevielle, X., Piou, D., Robin, C., Dutech, C. 2015. Pre-existing forests as sources of pathogens? The emergence of *Armillaria ostoyae* in a recently planted pine forest. *For. Ecol. Manage.*, **357**, 248-258.

Legrand, P., Ghahari, S., Guillaumin, J.-J. 1996. Occurrence of genets of *Armillaria* spp. in four mountain forests in Central France: The colonization strategy of *Armillaria ostoyae*. *New Phytol.*, **133**(2), 321-332.

Linnakoski, R., Sutela, S., Coetzee, M.P.A., Duong, T.A., Pavlov, I.N., Litovka, Y.A., Hantula, J., Wingfield, B.D., Vainio, E.J. 2021. *Armillaria* root rot fungi host single-stranded RNA viruses. *Sci. Rep.*, **11**(1), 7336.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.*, **38**(10), 4647-4654.

Misiek, M., Braesel, J., Hoffmeister, D. 2011. Characterisation of the ArmA adenylation domain implies a more diverse secondary metabolism in the genus *Armillaria*. *Fungal Biol.*, **115**(8), 775-781.

Misiek, M., Hoffmeister, D. 2012. Sesquiterpene aryl ester natural products in North American *Armillaria* species. *Mycol. Progress*, **11**(1), 7-15.

Morrison, D.J. 1991. Identity of *Armillaria* isolates used in studies of rhizomorph behaviour and pathogenicity. *Mycol. Res.*, **95**(12), 1437-1438.

Prospero, S., Holdenrieder, O., Rigling, D. 2004. Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. *Forest Pathol.*, **34**(1), 1-14.

Sipos, G., Prasanna, A.N., Walter, M.C., O'Connor, E., Bálint, B., Krizsán, K., Kiss, B., Hess, J., Varga, T., Slot, J., Riley, R., Bóka, B., Rigling, D., Barry, K., Lee, J., Mihaltcheva, S., LaButti, K., Lipzen, A., Waldron, R., Moloney, N.M., Sperisen, C., Kredics, L., Vágvölgyi, C., Patrignani, A., Fitzpatrick, D., Nagy, I., Doyle, S., Anderson, J.B., Grigoriev, I.V., Güldener, U., Münsterkötter, M., Nagy, L.G. 2017. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat. Ecol. Evol.*, **1**(12), 1931-1941.

Sonnenbichler, J., Guillaumin, J.-J., Peipp, H., Schwarz, D. 1997. Secondary metabolites from dual cultures of genetically different *Armillaria* isolates. *Eur. J. For. Pathol.*, **27**(4), 241-249.

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinf.*, **24**(5), 637-44.

Stanke, M., Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, **19**(suppl_2), ii215-ii225.

Sun, X., Zhang, T., Zhao, Y., Zhu, H., Cai, E. 2020. Protoilludane sesquiterpenoid aromatic esters from *Armillaria mellea* improve depressive-like behavior induced by chronic unpredictable mild stress in mice. *J. Funct. Foods*, **66**, 103799.

The Galaxy Community. 2022. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.*

Tsykun, T., Rigling, D., Nikolaychuk, V., Prospero, S. 2012. Diversity and ecology of *Armillaria* species in virgin forests in the Ukrainian Carpathians. *Mycol. Progress*, **11**(2), 403-414.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**(11), e112963.

Wingfield, B.D., Ambler, J.M., Coetzee, M., De Beer, Z.W., Duong, T.A., Joubert, F., Hammerbacher, A., McTaggart, A.R., Naidoo, K., Nguyen, H.D. 2016. Draft genome sequences of *Armillaria fuscipes, Ceratocystiopsis minuta, Ceratocystis adiposa, Endoconidiophora laricicola, E. polonica* and *Penicillium freii* DAOMC 242723. *IMA Fungus*, **7**(1), 217-227.

Zhan, M., Tian, M., Wang, W., Li, G., Lu, X., Cai, G., Yang, H., Du, G., Huang, L. 2020. Draft genomic sequence of *Armillaria gallica* 012m: insights into its symbiotic relationship with *Gastrodia elata*. *Braz. J. Microbiol.*, **51**(4), 1539-1552.

Table

Table 1: Genome information for the published *Armillaria* species in comparison to *Armillaria* African Clade B isolate CMW4456

Species	Strain	Number of scaffolds	Assembly size (Mbp)	Number of predicted protein coding genes	GC content (%)	Origin	Reference
Armillaria altimontana (NABS X)	837–10	100	73.74	19,326	47.8	USA	Caballero et al. (2022)
Armillaria borealis	AB13-TR4-IP16	44,365	66.59	21,969	N/A	Russia	Akulova et al. (2020)
Armillaria cepistipes	B5	182	75.52	23,461	47.6	Italy	Sipos et al. (2017)
Armillaria fuscipes	CMW 2740	24,403	52.98	14,515	N/A	South Africa	Wingfield et al. (2016)
Armillaria gallica	Ar21-2	319	85.34	25,704	47.5	USA	Sipos et al. (2017)
Armillaria gallica	012m	63	87.31	26,261	47.4	China	Zhan et al. (2020)
Armillaria mellea	DSM 3731	4,377	58.36	14,473	49.1	France	Collins et al. (2013)
Armillaria ostoyae	C18/9	106	60.11	22,705	48.3	Switzerland	Sipos et al. (2017)
Armillaria solidipes	C28-4	229	58.01	20,811	48.4	USA	Sipos et al. (2017)
Armillaria solidipes (form *A. ostoyae*)	ID001	72	55.74	16,357	48.3	USA	Caballero et al. (2022)
Armillaria African Clade B sp.	CMW4456	840	54.95	13,450	46.5	Zimbabwe	Described here

N/A = not available

Figure

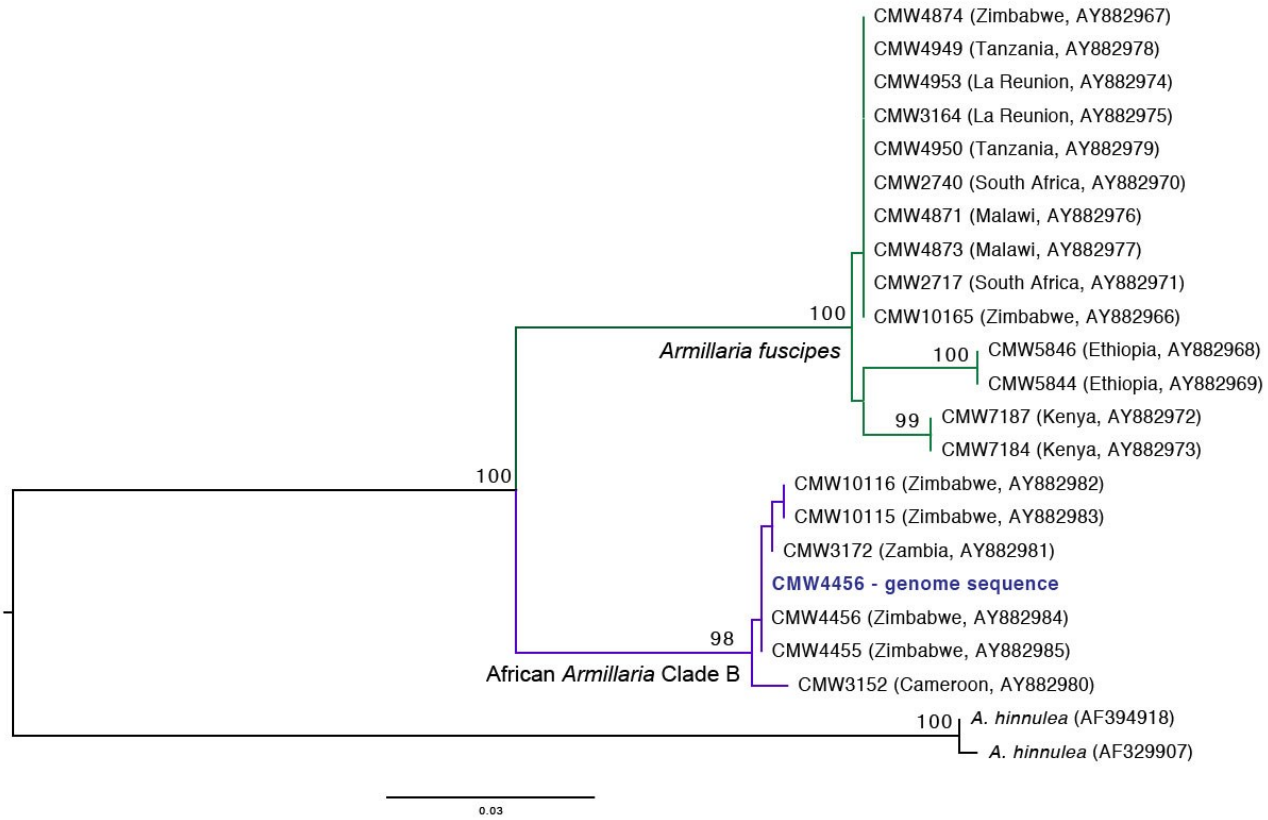


Figure 1: Maximum likelihood tree based on ITS sequence data, confirming the identity of the *Armillaria* African Clade B sp. strain CMW4456 sequenced in this study (highlighted in bold).
Bootstrap values above 80% are shown above the nodes. The scale bar represents nucleotide substitutions per site.

CHAPTER 5

COMPARATIVE PROTEOMIC AND SECRETOMIC INVESTIGATIONS OF AN AFRICAN *ARMILLARIA* SPECIES IN RESPONSE TO IRON

This chapter has been written in the format required for Journal of Proteome Research as:
Narh Mensah, D.L., Wingfield, B.D., Coetzee, M.P.A. 2023. Comparative proteomic and secretomic investigations of an African *Armillaria* species in response to iron. *Journal of Proteome Research*.

Abstract

Armillaria species have attracted considerable interest of researchers because they are widely distributed, mostly plant-pathogenic, and exhibit unique characteristics such as possession of rhizomorphs. Abiotic factors have been reported to influence intra- and inter-species variations in pathogenicity and/or virulence of these fungi. However, the mechanisms involved in these variations are not well understood. Iron is one of the important abiotic factors in various organisms because it is an indispensable element in several molecular and biological processes. Yet, iron can be toxic to organisms due to the occurrence of Fenton-like reactions when it is in excessive abundance. The aim of this study was to gain some insight into the type and extent of iron-responsive proteomic and secretomic changes in an *Armillaria* sp. supplemented with 100 μM $FeCl_3$. Our results show that this iron concentration did not elicit an oxidative stress response by the fungus. However, iron-dependent alteration of proteins involved in primary and secondary metabolism and amino acid biosynthesis were observed in the proteomes and secretomes. The proteins which increased in abundance with iron supplementation included dihydrolipoyl dehydrogenase, fructose-bisphosphate aldolase, fumarate hydratase, pyrophosphokinase, pyruvate carboxylase, 3-isopropylmalate dehydrogenase, and ribose-phosphate diphosphokinase. Proteins such as aconitate hydratase, glucose-6-phosphate 1-dehydrogenase, glucose-6-phosphate isomerase, 6-phosphogluconate dehydrogenase, and transketolase were down-expressed in the presence of added iron. The knowledge generated in this study contributes to better understanding of the mechanisms employed by *Armillaria* spp. in response to iron. The results also give insights into possible modes of inhibiting and attenuating pathogenicity and/or virulence of this strain of *Armillaria*.

Keywords: *Armillaria*, basidiomycete proteomics, fungal proteomics, iron homeostasis, secretomics

1. Introduction

Armillaria species have a worldwide geographic distribution and inhabit a wide range of ecological niches [1, 2]. Most *Armillaria* spp. are classified as facultative necrotrophs because they have both saprophytic and pathogenic phases [2]. These species threaten the health of woody plants in native and managed timber stands and agronomic plantations in areas where they were established [1, 3, 4].

The lifestyle of *Armillaria* species is influenced by various factors. Among these, growth mechanisms, as well as expression of genes responsible for production of plant cell wall degrading enzymes (PCWDEs), secondary metabolites, and other pathogenicity related gene products are factors responsible for their lifestyle [5-7]. The switch from a saprophytic to a pathogenic lifestyle and vice-versa, and host invasion by these fungi are influenced by factors such as intra-species variation, forest management systems, age and state of the hosts (e.g. diseased, stressed or healthy), elevation, as well as presence of other organisms in the rhizosphere of the host [8-13]. The mechanisms underlying these effects are not clearly understood.

Abiotic factors such as metal concentration and/or availability [14-18], drought or water activity [19, 20], carbon dioxide concentration [21], salinity [18, 22] and heat [23] elicit various proteomic responses in organisms. Among metals, iron homeostasis is vital because iron is an indispensable element in several molecular and biological processes in many organisms [24]. Additionally, animal-microbe, microbe-microbe, and plant-microbe interactions are influenced by iron uptake and transport mechanisms exhibited by the respective organisms [25-30]. Yet, abundance of internal iron is toxic to organisms due to the occurrence of Fenton-like reactions that can produce both hydroxyl radicals and higher oxidation states, which may cause biological damage [31, 32]. Consequently, investigations into the iron-dependent proteomes and/or secretomes of organisms including the opportunistic fungal human pathogen, *Aspergillus fumigatus* [16, 33-35], and other microorganisms [17, 36-38] have been conducted. These studies have highlighted various, often extensive, changes in the proteomes and/or secretomes of the studied organisms. The reported changes in these proteomes and/or secretomes have often been associated with mechanisms utilised for iron homeostasis. These mechanisms include assembly and/or synthesis of antioxidant enzymes, metal-chelating proteins or molecules, and free radical scavengers [35]. The reader is referred to the review article by Misslinger et al [35] for details about these mechanisms.

Unusual iron-dependent growth and siderophore (low affinity iron-chelating protein molecules) biosynthesis in strains of various species of *Armillaria* has previously been reported in Chapter Three of this thesis. The results suggested that *Armillaria* spp. have species-independent atypical iron requirements. In the present study, we applied the gel-free and label free LC–MS/MS bottom-up

shotgun approach. The study was aimed at gaining an insight into the type and extent of iron-responsive proteomic and secretomic changes exhibited by cultures grown in liquid medium of one isolate of *Armillaria* sp. from Zimbabwe. The knowledge generated in this study will unveil some iron-dependent mechanisms employed by *Armillaria* spp.

2. Materials and Methods

All reagents used in this study were analytical grade or equivalent and purchased from commercial suppliers.

2.1. Strain used and culture conditions

The studied isolate, *Armillaria* sp. strain CMW4456, belongs to African Group B (*sensu* Coetzee et al. 2005) and was obtained from *Brachystegia utilis* at Stapleford, Zimbabwe [39, 40]. The genome of this isolate has been sequenced by our team [41].

Plastic Petri dishes were used, and glassware were washed with HCl (6 M) followed by a 3 times rinse with ddH_20 to avoid iron contamination [42, 43]. Cultures were grown on potato dextrose peptone agarose (PDPA; 24 g/L potato dextrose, 2 g/L peptone, 10 g/L agarose) in 6.5 cm disposable Petri dishes. Agarose was used instead of agar to reduce iron contamination in the medium [44]. For iron deplete conditions in liquid medium, agarose was omitted from PDPA to obtain PDP-. Iron replete conditions were attained by supplementing PDP- with 100 µM $FeC1_3.6H_2O$ (PDP+). The liquid media (100 ml) were inoculated with 1 cm^2 actively growing culture. There were three biological replicates per treatment. All cultures were incubated at 25±2 ºC in the dark for 3 weeks. Liquid cultures were swirled to mix every week. Siderophore biosynthesis was determined in liquid cultures at the end of the incubation period using the modified chrome azurol S (CAS) assay solution with the microtiter method described by Alexander and Zuberer [45]. However, there was no significant difference in siderophore biosynthesis in cultures grown under both conditions. The fresh weight of the mycelia was also determined.

2.2. Protein extraction

Protein extraction from mycelia was performed according to previously described methods [16, 46], but with some modification. In brief, mycelia were separately harvested from liquid cultures in PDP- and PDP+ by centrifugation (10,000 g for 20 min at 4 °C) followed by decantation of the culture supernatant into fresh Eppendorf tubes on ice. Harvested mycelia were washed three times with sterile distilled water. Mycelia obtained from both treatments were suspended in 6 mL of cold lysis buffer (25 mM Tris-HCl, 6 M GdnHCl, 10 mM DTT pH 8.6) in a tube containing a ceramic bead and homogenised. The samples were then transferred to new tubes and sonicated (Cycle 6, 10 s, Power 10 %) on ice. All lysates were clarified twice by centrifugation (10,000 g for 20 min at 4 °C), after

which the resulting supernatants were brought to 15 % trichloroacetic acid (TCA). Precipitated protein was pelleted by centrifugation after 30 min of incubation in TCA, followed by three washes with ice-cold acetone. Pellets were finally resuspended in UT buffer (6 M urea, 2 M thiourea, 0.1 M Tris-HCl pH8).

Protein extraction from culture supernatants also followed previously described methods with some modifications [6]. In brief, 50 mL culture supernatants were centrifuged (10,000 g,10 min, 4 °C) and transferred to clean tubes on ice. The samples were then brought to 15 % (w/v) TCA, swirled on a rotor on ice for 30 mins to precipitate proteins, and then centrifuged. Proteins were washed three times with ice-cold acetone with centrifugation steps. Pellets were resuspended in UT buffer.

2.3. On-bead digestion and LC−MS/MS analyses

Sample digestion, peptide concentration determination, and LC-MS/MS analyses were conducted at the Proteomics Spectrometry Unit, Central Analytical Facilities of Stellenbosch University (Stellenbosch, South Africa) and described below.

2.3.1. Sample preparation for LC-MS/MS

Buffer change and sample clean-up were performed preceding digestion by resuspending all sample solutions in Tris buffer containing 5 mM tris(2-carboxyethyl)phosphine (TCEP; Fluka). The samples were then sonicated for 30 s in a cooled sonic bath followed by vortexing at high speed for 30 s. This process was repeated five times. For on-bead trypsin digestion of protein extracts, samples were re-suspended in 100 mM Tris-HCl (Sigma) containing 100 mM NaCl and 1 % SDS (Sigma) before reduction with 5 mM TCEP in 100 mM Tris buffer for 1 hour at 60 °C. Cystein residues were thiomethylated with 20 mM S-methyl methanethiosulfonate (Sigma) in 100 mM triethylammonium bicarbonate (TEAB; Thermo Scientific) for 30 min at room temperature. Samples were then diluted two-fold with binding buffer (100 mM ammonium acetate, 30 % acetonitrile, pH 4.5). The protein solution was added to MagResyn HILIC magnetic particles (Resyn Biosciences) prepared according to manufacturer's instructions and incubated overnight at 4 °C to bind. After binding, the supernatant was removed, and the magnetic particles washed twice with washing buffer (95 % acetonitrile (ACN; Romil). The magnetic particles were then suspended in 25 mM ammonium bicarbonate containing trypsin (New England Biosystems) to a final ratio of 1:50. After an 18-hour incubation at 37 °C the peptides were extracted once with 50 µL 15 % trifluoro acetic acid (TFA; Sigma). The samples were then dried and re-suspended in 30 µL ddH_20 for peptide concentration determination. The peptide concentration of each biological replicate per sample were determined using a Pierce™ Quantitative Peptide Assays & Standards (Thermo Scientific) according to manufacturer's instructions.

2.3.2. LC-MS/MS analysis

Liquid chromatography was performed on a Dionex Ultimate 3000 RSLC nano LC (Thermo Scientific; Massachusetts, USA) equipped with a C_{18} trapping column (Thermo Scientific; 5 mm × 300 µm, 5 µm; pore size 100 Å) and a CSH 25 cm × 75 µm 1.7 µm particle size C_{18} analytical column (Waters) according to the protocol described by Ellero, et al. [47] with some modification. The samples were loaded onto the trap column using loading solvent (2% acetonitrile:water; 0.1% formic acid (FA)) at a flow rate of 2 µL/min for 5 min from a temperature controlled autosampler set at 7 °C, before the sample was eluted onto the analytical column. The solvent system employed for elution was: Solvent A: 2% acetonitrile:water; 0.1% FA and Solvent B: 100% acetonitrile:water. Flow rate for elution was set to 300 nL/min and the gradient generated as follows: 5 % – 30 % solvent B over 60 min and 30 – 50 % solvent B from 60 – 80 min. Chromatography was performed at 45 °C and the outflow delivered to the mass spectrometer.

Mass spectrometry (MS) was performed using a Thermo Scientific Fusion Orbitrap Mass Spectrometer equipped with a Nanospray Flex ionization source. The sample was introduced through a stainless-steel nano-bore emitter. Data was collected in positive mode with the following parameters: spray voltage = 1.8 kV, ion transfer capillary = 275 °C. Polysiloxane ions at m/z = 445.12003 were used to internally calibrate the spectra. MS1 scans were performed in profile mode using the Orbitrap detector set at 120,000 resolution over the scan range 375-1500 with Automatic Gain Control (AGC) target set at 4 x 10^5 and maximum injection time of 50 ms. MS2 acquisitions in centroid mode were performed using monoisotopic precursor selection for ion with charges +2 – +7 with error tolerance set to +/- 10 ppm. Precursor ions were excluded from fragmentation once for a period of 60 s. Precursor ions were selected for fragmentation in higher-energy collisional dissociation (HCD) mode using the quadrupole mass analyser with HCD energy set to 30 %. Fragment ions were detected in the Orbitrap mass analyser set to 30,000 resolution. The AGC target was set to 5 x 10^4 and the maximum injection time to 100 ms.

2.4. Spectra processing and bioinformatics analyses

Data retrieved from LC-MS/MS analyses of biological replicates from mycelia obtained from PDP- and PDP+ represented the proteome under iron deplete (Mde) and iron replete (Mre) conditions respectively. Supernatants obtained from PDP- and PDP+ represented the secretome under iron deplete (Sde) and iron replete (Sre) conditions respectively. Data processing and analyses was conducted at BGI Genomics (https://www.bgi.com/global) and outlined below.

The thermo.raw files generated by the mass spectrometer were imported into MaxQuant 1.6.2.3. [47] and processed using the integrated search engine of MaxQuant, Andromeda [48]. MaxQuant was also

used to extract peak areas and calculate protein quantitation values. Quantitative analysis was performed based on the peak intensity, peak area, and LC retention time of the peptides related to the first-order mass spectrometry. A series of statistical analysis and quality control were also performed to obtain significant identification results using the following parameters: Enzyme = Trypsin, Peptide_Mass_Tolerance = 10 ppm, Fragment_Mass_Tolerance = 0.02 Da, Minimal peptide length = 7, PSM-level FDR = 0.01, Protein-level FDR = 0.01, Fixed modification = Carbamidomethyl (C), Variable modifications = Oxidation (M), Acetyl (Protein N-term), Deamidated(NQ), and Gln->pyro-Glu, and Database = 1253_all_filtered.fasta (83,019 sequences). The data was searched using UniProt Protein Database, Augustus annotation of the genome of strain CMW4456 (accession number JANDKJ000000000.1) [41], and protein sequences of *A. mellea* strain DSM 3731 (sample code Armme1_1) [6] and *A. ostoyae* strain C18/9 (sample code Armosto1) [5] obtained from JGI [50]. The database used has been deposited in a repository and can be accessed at the University of Pretoria website (https://www.up.ac.za/). Protein quantification and difference analysis were conducted on the set comparison groups (i.e., Mre vs Mde and Sre vs Sde). The multiples of differences in the proteins in each comparison group were calculated. Significance test was performed using Welch's t-test. Protein abundances with a fold change (FC) > 1.5 and p-value < 0.05 were considered as differentially expressed proteins (DEPs).

Functional annotations such as Gene Ontology (GO) [51], euKaryotic Ortholog Groups (KOG) [52], and Kyoto Encyclopedia of Genes and Genomes (KEGG) [53] for pathway analyses were automatically performed based on the identified DEPs. GO functional annotation analysis included two parts: protein2go and go2protein. For each protein, a list of identities (IDs) and all corresponding GO functions were given in protein2go results. In terms of go2protein, for the GO entries involved in the three ontology classifications (i.e., biological process, cellular component, and molecular function), the identities and the number of proteins of all the corresponding proteins were listed, and the GO entries without the corresponding proteins were omitted. KOGs are databases for orthologous classification of proteins. Each KOG entry contains a series of orthologous proteins or paralogs. The analysis compared the identified proteins with the KOG database, predicted the possible functions of these proteins and performed functional classification statistics on the identified proteins. *In vivo*, different proteins coordinate biological behaviours and pathway-based analysis can therefore help further understand their biological functions. KEGG is the main public database on pathway [53] and was used to annotate molecular networks (pathways) of the DEPs.

The draft genome sequence of CMW4456 (Chapter 4) was also investigated to detect the NRPS and NIS synthetase gene clusters reported in Chapters 2 and 3 of this thesis for other strains of *Armillaria*, and to link these to the proteins detected in the present chapter. The same bioinformatic tools, as described in the two chapters, were used for this purpose. NRPS genes, NIS synthetase genes and other genes typically present in both types of gene clusters were detected in the genome. However, some of the genes appeared fragmented and/or there were other copies of such genes in the genome. Using this information to investigate the proteomic response to iron with respect to siderophore mediated iron uptake or storage by strain CMW4456 was therefore considered unreliable and not used further in this study.

3. Results

3.1. General characteristics of the proteome and secretome under the experimental conditions

There were 2,498 – 2,509, 2,360 – 2,453, 393 – 448, 379 – 463, 327, and 555 detected proteins in samples Mre, Mde, Sre, Sde, Cre, and Cde, respectively. The samples are proteomes under iron-deplete (Mde) and iron-replete (Mre) conditions and the secretomes under iron-deplete (Sde) and iron-replete (Sre) conditions. The media only with and without iron supplementation are Cre and Cde, respectively. The molecular weights of most of the detected proteins were in the range of 30 – 40 kDa.

The entire dataset is presented as principal component analysis (PCA) in which replicate samples are reduced into their corresponding sample groups (Figure 1). The positioning of samples in the PCA depicts the variability of the samples relative to one another. Results of the PCA indicated that the combined variance between components 1 (PC1) and 2 (PC2) amounts to 75.38 % with PC1 contributing 71.11% to the variance. The PCA plot of proteins of *Armillaria* sp. strain CMW4456 demonstrated distinct differences in the proteomes and the secretomes under both conditions. The media only (Cde and Cre) also grouped separately. The results also showed good repeatability of each biological replicate in the same sample group (i.e., proteome and secretome).

The proteomes and secretomes were altered with iron supplementation. Proteomes obtained with (Mre) and without (Mde) iron supplementation revealed 167 differentially expressed proteins (DEPs). Among these, 60 proteins were down-expressed (decreased in abundance), and 107 proteins were up-expressed (increased in abundance) in Mre compared to Mde (Figure 2A). Secretomes obtained under the two conditions revealed 15 DEPs between the iron-replete (Sre) and iron deplete (Sde) conditions, 13 of which were down-expressed in Sre compared to Sde (Figure 2B). The list of up-, down-, and non-differentially expressed proteins in both the proteome and secretome can be found in Supplementary material (S1).

3.2. Qualitative variations in differentially expressed proteins

3.2.1. Gene Ontology classification of DEPs

Gene Ontology (GO) annotation and functional classification of the DEPs identified in the evaluated samples revealed that the proteome and secretome of *Armillaria* sp. strain CMW4456 were altered in response to iron supplementation. The DEPs were classified into, biological process (BP), cellular component (CC), and molecular function (MF). The numbers of DEPs under different GO classification in Mre vs Mde and Sre vs Sde comparison groups are displayed in descending order (Figure 3A and B, respectively).

The GO functional enrichment analysis of DEPs in the Mre vs Mde comparison group classified 623 DEPs divided into 29 GO terms. There were 10 BP, 12 CC, and 7 MF terms. The 3 main GO annotations in the proteomes in the BP classification were “cellular process” (11 down- and 45 up-expressed DEPs), “metabolic process” (16 down- and 40 up-expressed), and “biological regulation” (0 down- and 10 up-expressed). The conspicuous terms under CCs for this comparison group were “cell” (12 down- and 49 up-expressed DEPs), “cell part” (12 down- and 48 up-expressed DEPs), and “organelle” (8 down- and 33 up-expressed DEPs), while that classified under MFs were “catalytic activity” (36 down- and 62 up-expressed DEPs) and “binding” (20 down- and 51 up-expressed DEPs) (Figure 3A).

In the Sre vs Sde comparison group, there were fewer GO terms to which the DEPs were annotated (18 GO terms) with 8, 7, and 3 terms classified under BP, CC, and MF, respectively (Figure 3B). As found in the proteomes, GO terms in the secretomes in the BP classification included “cellular process” (6 down- and 1 up-expressed DEPs) and “metabolic process” (6 down- and 1 up-expressed DEPs). Those in the CC classification were “cell” and “cell part” (3 down- and 1 up-expressed DEPs for both functions), while the most abundant GO terms in the MF classification were “catalytic activity” (10 down- and 2 up-expressed DEPs) and “binding” (8 down- and 2 up-expressed DEPs) (Figure 3B).

3.2.2. KOG classification of differentially expressed proteins

KOG functional classification of the DEPs also revealed differences in functional annotation of the DEPs of the proteome and secretome under the experimental conditions. There were 22 functions identified among the DEPs detected in proteomes obtained under iron-replete compared to iron-deplete conditions (Mre vs Mde; Figure 4A). “General function prediction only”, belonging to the “poorly characterized” category accounted for 32 proteins out of the 193 DEPs. There was a total of 72, 35, and 46 DEPs assigned to the main KOG classifications, “metabolism”, “information storage

and processing", and "cellular processes and signalling" respectively. Under the "metabolism" category, 20, 14, and 13 DEPs were assigned to "energy production and conversion", "lipid transport and metabolism", and "amino acid transport and metabolism" respectively. Additionally, 10 each of the DEPs under this category were assigned to "secondary metabolite biosynthesis, transport and catabolism" and "carbohydrate transport and metabolism". Under the "information storage and processing" category, "translation, ribosomal structure and biogenesis" (11 DEPs) and "RNA processing and modification" (9 DEPs) were the most abundant KOG functions. Most of the DEPs (23 proteins) classified under the "cellular processing and signalling" category function in "post-translational modification, protein turnover, chaperons", followed by 9 DEPs functioning in "signal transduction mechanisms".

The 21 DEPs of secretomes obtained under iron-replete compared to iron-deplete conditions (Sre vs Sde) were annotated to 11 functions under all the 4 KOG classifications (Figure 4B). Two DEPs each were assigned to "General function prediction only" and "translation, ribosomal structure and biogenesis" under the "poorly characterized" and "information storage and processing" categories respectively. Most of these DEPs (11 proteins) were assigned to the KOG term, "metabolism", followed by 6 DEPs assigned to "cellular processing and signalling". Under the "metabolism" category, 3 DEPs each function in "carbohydrate transport and metabolism" and "amino acid transport and metabolism". Three DEPs under the "cellular processing and signalling" category function in "post-translational modification, protein turnover, chaperons".

3.2.3. KEGG pathway annotation

All pathways with DEPs can be accessed at the research data management repository of the University of Pretoria.

The DEPs were mapped to KEGG pathways to explore the biological functions of the DEPs with regards to the different sample comparison groups (Figure 5). Between Mre vs Mde (Figure 5A) and Sre vs Sde (Figure 5B) comparison groups, 8 and 4 KEGG pathways respectively were found to have significantly different enrichment patterns. The pathways significantly enriched in proteomes under iron-replete compared to the iron-deplete conditions (Mre vs Mde) had Rich Factors between 0.3 and 0.4. Eight of these DEPs function in "pentose phosphate pathway" (p-value = 0.000). All the other pathways had Rich Factors less than 0.2. These included "biosynthesis of secondary metabolites" (30 DEPs, lowest Rich Factor, p-value = 0.043), "carbon metabolism" (16 DEPs, second lowest Rich Factor, p-value = 0.006), and "citrate cycle (TCA cycle)" (6 DEPs, third lowest Rich Factor, p-value = 0.036).

In the secretome, under iron-replete compared to the iron-deplete conditions (Sre vs Sde), there were 2 DEPs each in all the 4 reported pathways, all of which were down-expressed. The pathways with the lowest and highest Rich Factors were shown to function in "glycolysis/gluconeogenesis" (Rich Factor <0.05; p-value = 0.049) and "ascorbate and aldarate metabolism" (Rich Factor>0.1; p-value<0.004), respectively.

4. Discussion

Results obtained in this study revealed proteomic and secretomic alterations of strain CMW4456 with iron supplementation. The differentially expressed proteins identified in the proteomes and secretomes were involved in primary and secondary metabolism as well as amino acid biosynthesis and growth of *Armillaria* sp. strain CMW4456.

4.1. Iron-dependent proteome and secretome profiles of strain CMW4456

Many proteins (2,360 – 2,509) were detected in the proteomes obtained under the two conditions, with and without addition of iron. This high number of proteins is comparable to that detected for other white-rot basidiomycetes. For example, the number of proteins detected using similar techniques in *Pleurotus ostreatus* proteome was 2062 proteins [54] whereas 2543 proteins were detected in mycelia of *Hericium erinaceus* using iTRAQ labelling and nano-LC-MS/MS analysis [55]. On the contrary, notably lower number of proteins (921 and 1984 proteins) have been detected in proteomes of *Armillaria mellea* using different approaches [6, 46].

Clear separation in the proteome and secretome under both growth conditions in the PCA plot reflected the significant change of protein expression between mycelia and supernatant, as was expected. The separation also revealed the iron-dependent alteration of protein expression by *Armillaria* sp. strain CMW4456. All the qualitative analyses conducted in this study also highlighted the change of the proteome and secretome of strain CMW4456 in response to iron supplementation. Proteomic responses to iron have also been reported in the fungi, *Aspergillus fumigatus* and *Paracoccidioides brasiliensis* [16, 56].

The range of molecular masses of most of the detected proteins was slightly lower than that reported for the proteome of *Aspergillus flavus* in response to water activity [19]. This disparity can be explained by factors including the fact that only the proteome was considered in the report on *As. flavus*, as well as the analysis techniques used in the respective studies. Different experimental techniques for proteomics have been shown to influence protein detection and other parameters in proteomics research [57].

4.2. Iron supplementation does not alter oxidative stress response of strain CMW4456

Supplementation of PDP medium with 100 µM $FeCl_3$ did not alter oxidative stress response by strain CMW4456. This contrasts with the proteomic and secretomic response observed for various organisms in response to metal concentrations [14, 15, 18, 22, 58-60]. Oxidative stress causes reactive oxygen species (ROS) production and can result in cellular damage when it is in high concentration (i.e., protein degradation, DNA damage, membrane peroxidation, and apoptosis) [61]. Therefore, organisms utilise various mechanisms for responding to oxidative stress to ensure correct redox balance. These mechanisms include assembly of antioxidant enzymes, metal-chelating proteins or molecules, and free radical scavengers [18].

Organisms up-express various enzymes under oxidative stress caused by abiotic factors, including exposure to high metal concentrations. These enzymes include aldehyde dehydrogenase, ATP phosphoribosyltransferase, ATP synthase subunits, glutamate–cysteine ligase, glutathione reductase, glutathione synthetase, glutathione S-transferase, heat shock proteins, NADPH cytochrome P450 reductase, NADH-ubiquinone oxidoreductase, superoxide dismutase, thioredoxin reductase, and trehalase [14, 15, 18, 22, 58-60]. However, glutathione S-transferase was identified under both iron-replete and iron-deplete conditions in *As. fumigatus* microsomal extracts [16]. Over-expression of proteins such as ATP phosphoribosyltransferase and saccharopine dehydrogenase, involved in histidine and glutamate biosynthesis respectively, is indicative of an increase in amino acid metabolism potentially for replacement of misfolded proteins caused by ROS [14, 18, 60]. Most of the implicated enzymes play a significant role in the glutathione biosynthesis pathway (reviewed in [62, 63]). Glutathione (GSH) is an important antioxidant which can sequester metals for providing tolerance to cellular components of organisms against heavy metals [64]. Additionally, accumulation of trehalase is implicative of accumulation of the disaccharide trehalose, which functions in preventing aggregation of denatured proteins and scavenging free radicals [65].

The proteins usually reported to be increased in abundance in response to oxidative stress and/or metal toxicity, including trehalase, generally showed no significant change in expression pattern between the two experimental conditions in this study. Hence, addition of iron at 100 µM to the growth medium did not alter the oxidative stress response of strain CMW4456 and is therefore evidently not a toxic concentration for this strain of *Armillaria*.

4.3. Primary metabolism by strain CMW4456 is altered in response to iron supplementation

Iron supplementation altered the expression of proteins of primary metabolic pathways including the trichloroacetic acid cycle (i.e., TCA cycle or citrate cycle), the pentose phosphate pathway (PPP) and/or glycolysis by strain CMW4456. Enzymes involved in the TCA cycle, including dihydrolipoyl dehydrogenase, dihydrolipoyllysine-residue succinyltransferase, fumarate hydratase, pyruvate dehydrogenase E1 component subunit alpha, and pyruvate carboxylase were up-expressed in the proteome of strain CMW4456 with iron supplementation. These enzymes were down-expressed during iron deprivation of the fungi *Paracoccidioides brasiliensis* [56] and *Histoplasma capsulatum* [38].

Aconitate hydratase was the only down-expressed enzyme in the TCA cycle with iron supplementation in this study. Decreased abundance of aconitate hydratase has been recorded during iron starvation of *P. brasiliensis* [56]. The authors attributed this observation to metabolic switching to pathways independent of enzymes with Fe/S clusters, such as aconitate hydratase, by *P. brasiliensis* during iron starvation [56].

Isocitrate lyase, that is involved in the glyoxylate cycle (a variation of the TCA cycle), showed no significant change in abundance in both the proteome and secretome by strain CMW4456 with iron supplementation. This enzyme produces glyoxylate and succinate using isocitrate as substrate [66]. Isocitrate lyase was decreased in abundance in the proteome of *P. brasiliensis* during iron starvation [56]. The fact that no significant change in abundance of isocitrate lyase was observed under the study condition suggests that iron supplementation of the growth medium with 100 µM $FeCl_3$ does not constitute a significant increase in iron availability for strain CMW4456.

Given that several other enzymes involved in the TCA cycle were up-expressed in the present study, the observed down-expression of aconitate hydratase suggests that only interconversion of citrate and isocitrate is limited in response to iron supplementation by strain CMW4456. Taken together, results obtained in this study suggest that the TCA cycle is favoured with iron supplementation of the growth medium with 100 µM $FeCl_3$. This iron concentration may, however, be insufficient for producing enzymes with Fe/S clusters by the fungus as aconitate hydratase expression was still repressed at this concentration. Other enzymes with Fe/S clusters such as various subunits of succinate dehydrogenase and cytochrome c showed no significant change in expression patterns with iron supplementation in both the proteome and secretome. Such enzymes are mostly involved in redox reactions but also participate in the control of gene expression, oxygen/nitrogen sensing, control of labile iron pool and DNA damage recognition and repair (reviewed in [67]). The presence of subunits of mitochondrial membrane-bound proteins such as succinate dehydrogenase [ubiquinone] flavoprotein subunit in the secretome is an interesting finding. Although this may be due to cell lyses in the liquid media used for the experiment, this finding would require further research.

Various broad-spectrum fungicides belong to the succinate dehydrogenase inhibitor (SDHI) class. Several phytopathogenic fungi including basidiomycetes such as *Ustilago maydis* have been reported to have developed resistance to these fungicides using various mechanisms (reviewed in [68-70]). Although various chemicals have been evaluated against *Armillaria* [71, 72], to the best of our knowledge, SDHI class fungicides have not been evaluated against *Armillaria*. We propose that existing SDHIs should be evaluated against *Armillaria* spp. and/or new SDHIs should be developed based on the succinate dehydrogenases produced by *Armillaria* spp. These novel fungicides may also provide alternatives to fungicides to which plant-pathogenic fungi have developed resistance.

Enhanced metabolism through the TCA cycle with iron supplementation contributes to enhanced ATP production [73]. The recorded protein expression patterns exhibited by strain CMW4456 is congruent with the observed faster culture growth of the fungus in the presence of iron and siderophore biosynthesis under both experimental conditions in this study. Faster culture growth of strain CMW4456 was recorded on solid potato dextrose peptone media supplemented with 100 µM $FeCl_3$ in Chapter Three of this thesis. Additionally, siderophore biosynthesis of this strain and other strains of *Armillaria* species in PDP with up to 200 µM added $FeCl_3$ suggested that *Armillaria* have species- and strain-independent unique requirements for iron.

Proteins such as glucose-6-phosphate 1-dehydrogenase (G6PDH), glucose-6-phosphate isomerase (G6PI), and transketolase, involved in PPP and/or glycolysis, were down-expressed in the proteome of strain CMW4566 in the presence of added iron. Transketolase was up-expressed in the fungus *Phanerochaete chrysosporium*, in response to the ROS produced by copper [15]. On the contrary, G6PI was down-expressed under copper-induced oxidative stress, while transketolase was commonly expressed by mycelia of *Penicillium chrysogenum* irrespective of copper concentration in growth media [18]. Other enzymes including fructose-bisphosphate aldolase and ribose-phosphate diphosphokinase were up-expressed by strain CMW4456 with iron supplementation. These enzymes are also involved in PPP and glycolysis and have been shown to be down-expressed in response to copper-induced oxidative stress in the fungus, *Pe. chrysogenum* [18]. Transketolase of the rice blast fungus, *Magnaporthe oryzae*, plays an essential role in facilitating host colonization of rice cells and invasive hyphal growth was curtailed in transketolase null mutants [74]. Amide compounds/derivatives have been evaluated for plant and fungal transketolase inhibition [75]. Therefore, inhibition of transketolase produced by *Armillaria* spp. using novel or existing amide compounds may provide avenues for controlling Armillaria root- and stem-rot disease.

Additionally, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), an enzyme involved in glycolysis and a target for oxidative stress, was not significantly differentially expressed in both the proteome and secretome of strain CMW4456 when iron was added. This enzyme was up-expressed by *Ph. chrysosporium* mycelia during oxidative stress induced by copper and lead treatment [15, 76]. In contrast, this enzyme was down-expressed in culture supernatants of a different strain of the same fungus under cadmium exposure [77].

Overall, reverse expression patterns of almost all the oxidative stress-related proteins involved in PPP and/or glycolysis were observed in this study compared to the reported expression patterns of these proteins by other fungi. These results suggest that iron supplementation of PDP at 100 µM for strain CMW4456 does not elicit the stress responses reported for the other fungi to various stress factors. This provides further support to our hypothesis that iron supplementation at the experimental concentration is not a stress factor to this fungus but could be insufficient for molecular and biological processes of strain CMW4456.

4.4. Amino acid biosynthesis, secondary metabolism and growth are enhanced with iron supplementation

Proteins associated with amino acid biosynthesis and secondary metabolism were differentially expressed in proteomes obtained under iron supplementation. Enzymes such as 3-isopropylmalate dehydrogenase increased in abundance in the proteome of strain CMW4456 with iron supplementation. This enzyme is involved in leucine biosynthesis and is required for biosynthesis of the mycotoxin, pneumocandin in the yeast, *Glarea lozoyensis* [78, 79]. Methods for inhibition of 3-isopropylmalate dehydrogenase have been patented as fungicides and have been successful in attenuating the virulence of various fungi including the plant-pathogenic fungus, *Magnaporthe* spp. [80]. The small subunit of acetohydroxy-acid synthase was also up-expressed by strain CMW4456 when PDP was supplemented with iron. Acetohydroxy acid synthase is required for biosynthesis of the branched-chain amino acids, valine and isoleucine [81]. Deletion of the gene which encodes this protein (*ilv2*) attenuates virulence of *Candida albicans* and elicited starvation-cidal phenotypes in both *C. albicans* and *Saccharomyces cerevisiae ilv2*Δ mutants [82]. Glutamate-5-semialdehyde dehydrogenase, an enzyme involved in arginine and proline metabolism, but which specifically functions in the proline pathway, was also up-expressed in the proteome under the experimental conditions. This enzyme has been used as an antifungal drug target [83] and should be explored for developing antifungal agents against pathogenic *Armillaria* spp.

Our findings suggest that iron limitation and antifungal substances which target proline, valine, leucine, and isoleucine biosynthesis pathways can be explored for inhibiting strain CMW4456. This is because isopropylmalate dehydrogenase, acetohydroxy-acid synthase and glutamate-5-semialdehyde dehydrogenase were increased in abundance with iron supplementation in the present study. Attenuation of pathogenicity and/or virulence of strain CMW4456 by antifungal substances which target these amino acid biosynthesis pathways should be evaluated.

Mycelia growth of strain CMW4456 was enhanced with iron supplementation at 100 µM. Culture fresh weight of 0.311 – 0.376 g in PDP- and 0.501 – 0.556 g in PDP+ were recorded. This was also observed in the proteomes obtained under the experimental conditions. This is demonstrated by the fact that cellular processes, cell, and cell part comprised most of the up-expressed DEPs of the proteome of the fungus with added iron as shown by the results of the GO annotations. For example, the growth-related protein, pyruvate dehydrogenase increased in abundance with iron supplementation. This enzyme functions in pyruvate metabolism and has been reported to contribute to long-term survival of fission yeast [84]. Additionally, microtubule binding protein was up-expressed in the proteome by strain CMW4456 with iron supplementation. This enzyme is involved in microtubule cytoskeleton organization [85]. Microtubule cytoskeleton is essential for filament integrity, polarized growth, transport of organelles and vesicles, as well as spindle assembly, and required for normal cell morphogenesis in various organisms including the human fungal pathogen, *Cryptococcus neoformans* [85], and Basidiomycota such as the plant-pathogen, *Ustilago maydis* (reviewed in [86]). A member of the microtubule's building block, tubulin alpha chain, also increased in abundance with iron supplementation. These proteins play a role in nuclear migration and positioning in filamentous fungi (reviewed in [87]). The exact effects of these enzymes on growth of strain CMW4456 and other strains of *Armillaria* spp. should be investigated to develop more effective control strategies for these phytopathogens.

Conclusions

In this study, the proteome and secretome of an *Armillaria* strain from Africa was investigated for the first time. The comparative proteomic and secretomic study conducted demonstrated that a change in protein expression by strain CMW4456 was induced with supplementation of 100 µM $FeCl_3$. The difference in protein expression is not the same as one would expect if the supplement caused an oxidative stress response by the fungus. The supplementation of medium with 100 µM $FeCl_3$ was apparently not sufficient to support all molecular and biological processes of the fungus, although fungal growth was enhanced under this condition. The result of this study gives a deeper understanding of the biology of this strain of *Armillaria* in terms of iron homeostasis using the gel-free and label-free shotgun LC-MS/MS approach. Results from this study also identify potential

targets for inhibition and attenuation of pathogenicity and/or virulence of strain CMW4456. These include limitation of iron availability as well as use of antifungal substances such as SDHIs and which target biosynthesis of proline and the branched-chain amino acids, valine, leucine, and isoleucine. Further research is required to investigate applicable efficacious and sustainable technologies and potential antifungal substances such as existing and/or novel succinate dehydrogenase and transketolase inhibitors, and antifungal substances which will target iron bioavailability to the fungus, as well as biosynthesis of the indicated amino acids for controlling these phytopathogens. The draft genome sequence of CMW4456 should also be improved to enhance identification of secondary metabolite gene clusters in the genome of this strain.

References

(1) Coetzee, M. P. A.; Wingfield, B. D.; Wingfield, M. J. *Armillaria* root-rot pathogens: Species boundaries and global distribution. *Pathogens* **2018**, *7* (4), 83. DOI: 10.3390/pathogens7040083

(2) Gregory, S. C.; Rishbeth, J. Pathogenicity and Virulence. In *Agriculture Handbook (USA)*, Shaw III, C. G., Kile, G. A. Eds.; Armillaria root disease, Vol. No. 691; USDA Forest Service, 1991; pp 76–87.

(3) Baumgartner, K.; Coetzee, M. P. A.; Hoffmeister, D. Secrets of the subterranean pathosystem of *Armillaria*. *Mol. Plant Pathol.* **2011**, *12* (6), 515-534. DOI: 10.1111/j.1364-3703.2010.00693.x

(4) Elías-Román, R. D.; Guzmán-Plazola, R. A.; Klopfenstein, N. B.; Alvarado-Rosales, D.; Calderón-Zavala, G.; Mora-Aguilera, J. A.; Kim, M.-S.; García-Espinosa, R. Incidence and phylogenetic analyses of *Armillaria* spp. associated with root disease in peach orchards in the State of Mexico, Mexico. *Forest Pathol.* **2013**, *43* (5), 390-401. DOI: 10.1111/efp.12043

(5) Sipos, G.; Prasanna, A. N.; Walter, M. C.; O'Connor, E.; Bálint, B.; Krizsán, K.; Kiss, B.; Hess, J.; Varga, T.; Slot, J.; et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. *Nat. Ecol. Evol.* **2017**, *1* (12), 1931-1941. DOI: 10.1038/s41559-017-0347-8

(6) Collins, C.; Keane, T. M.; Turner, D. J.; O'Keeffe, G.; Fitzpatrick, D. A.; Doyle, S. Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. *J. Proteome Res.* **2013**, *12* (6), 2552-2570. DOI: 10.1021/pr301131t

(7) Ross-Davis, A. L.; Stewart, J. E.; Hanna, J. W.; Kim, M. S.; Knaus, B. J.; Cronn, R.; Rai, H.; Richardson, B. A.; McDonald, G. I.; Klopfenstein, N. B.; et al. Transcriptome of an Armillaria root disease pathogen reveals candidate genes involved in host substrate utilization at the host-pathogen interface. *For. Pathol.* **2013**, *43* (6), 468-477. DOI: 10.1111/efp.12056

(8) Tsykun, T.; Rigling, D.; Nikolaychuk, V.; Prospero, S. Diversity and ecology of *Armillaria* species in virgin forests in the Ukrainian Carpathians. *Mycol. Progress* **2012**, *11* (2), 403-414. DOI: 10.1007/s11557-011-0755-0

(9) Prospero, S.; Holdenrieder, O.; Rigling, D. Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. *Forest Pathol.* **2004**, *34* (1), 1-14. DOI: 10.1046/j.1437-4781.2003.00339.x

(10) Morrison, D. J.; Pellow, K. W. Variation in virulence among isolates of *Armillaria ostoyae*. *For. Pathol.* **2002**, *32* (2), 99-107. DOI: 10.1046/j.1439-0329.2002.00275.x

(11) Chandelier, A.; Gerarts, F.; San Martin, G.; Herman, M.; Delahaye, L. Temporal evolution of collar lesions associated with ash dieback and the occurrence of *Armillaria* in Belgian forests. *For. Pathol.* **2016**, *46* (4), 289-297. DOI: 10.1111/efp.12258

(12) Bakys, R.; Vasiliauskas, A.; Ihrmark, K.; Stenlid, J.; Menkis, A.; Vasaitis, R. Root rot, associated fungi and their impact on health condition of declining *Fraxinus excelsior* stands in Lithuania. *Scandinavian J. For. Res.* **2011**, *26* (2), 128-135. DOI: 10.1080/02827581.2010.536569

(13) Kwaśna, H. Fungi in the rhizosphere of common oak and its stumps and their possible effect on infection by *Armillaria*. *Appl. Soil Ecol.* **2001**, *17* (3), 215-227. DOI: 10.1016/S0929-1393(01)00137-8

(14) Khatiwada, B.; Hasan, M. T.; Sun, A.; Kamath, K. S.; Mirzaei, M.; Sunna, A.; Nevalainen, H. Proteomic response of *Euglena gracilis* to heavy metal exposure – Identification of key proteins involved in heavy metal tolerance and accumulation. *Algal Res.* **2020**, *45*, 101764. DOI: 10.1016/j.algal.2019.101764

(15) Okay, S.; Yildirim, V.; Buttner, K.; Becher, D.; Ozcengiz, G. Dynamic proteomic analysis of *Phanerochaete chrysosporium* under copper stress. *Ecotoxicol. Environ. Saf.* **2020**, *198*, 110694. DOI: 10.1016/j.ecoenv.2020.110694

(16) Moloney, N. M.; Owens, R. A.; Meleady, P.; Henry, M.; Dolan, S. K.; Mulvihill, E.; Clynes, M.; Doyle, S. The iron-responsive microsomal proteome of *Aspergillus fumigatus*. *J. Proteomics* **2016**, *136*, 99-111. DOI: 10.1016/j.jprot.2015.12.025

(17) de Souza, A. F.; Pigosso, L. L.; Silva, L. O. S.; Galo, I. D. C.; Paccez, J. D.; KSF, E. S.; de Oliveira, M. A. P.; Pereira, M.; Soares, C. M. A. Iron deprivation modulates the exoproteome in *Paracoccidioides brasiliensis*. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 903070. DOI: 10.3389/fcimb.2022.903070

(18) Lotlikar, N.; Damare, S.; Meena, R. M.; Jayachandran, S. Variable protein expression in marine-derived filamentous fungus *Penicillium chrysogenum* in response to varying copper concentrations and salinity†. *Metallomics* **2020**, *12* (7), 1083-1093. DOI: 10.1039/c9mt00316a

(19) Zhang, F.; Zhong, H.; Han, X.; Guo, Z.; Yang, W.; Liu, Y.; Yang, K.; Zhuang, Z.; Wang, S. Proteomic profile of *Aspergillus flavus* in response to water activity. *Fungal Biol.* **2015**, *119* (2-3), 114-124. DOI: 10.1016/j.funbio.2014.11.005

(20) Dong, A.; Yang, Y.; Liu, S.; Zenda, T.; Liu, X.; Wang, Y.; Li, J.; Duan, H. Comparative proteomics analysis of two maize hybrids revealed drought-stress tolerance mechanisms. *Biotechnol. Biotechnol. Equip.* **2020**, *34* (1), 763-780. DOI: 10.1080/13102818.2020.1805015

(21) Lin, R.; Zhang, L.; Yang, X.; Li, Q.; Zhang, C.; Guo, L.; Yu, H.; Yu, H. Responses of the mushroom *Pleurotus ostreatus* under different CO_2 concentration by comparative proteomic analyses. *J. Fungi (Basel)* **2022**, *8* (7), 652. DOI: 10.3390/jof8070652

(22) Wang, Y.; Guan, Y.; Lin, W.; Yan, H.; Neng, J.; Sun, P. Quantitative proteomic profiling of fungal growth, development, and ochratoxin a production in *Aspergillus ochraceus* on high- and low-NaCl cultures. *Toxins (Basel)* **2021**, *13* (1), 51. DOI: 10.3390/toxins13010051

(23) Xu, L.; Guo, L.; Yu, H. Label-free comparative proteomics analysis revealed heat stress responsive mechanism in *Hypsizygus marmoreus*. *Front. Microbiol.* **2020**, *11*, 541967. DOI: 10.3389/fmicb.2020.541967

(24) Loper, J. E.; Buyer, J. S. Siderophores in microbial interactions on plant surfaces. *Mol. Plant-Microbe Interact.* **1991**, *4* (1), 5-13.

(25) Amin, S. A.; Green, D. H.; Hart, M. C.; Küpper, F. C.; Sunda, W. G.; Carrano, C. J. Photolysis of iron–siderophore chelates promotes bacterial–algal mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (40), 17071-17076. DOI: 10.1073/pnas.0905512106

(26) Gu, S.; Wei, Z.; Shao, Z.; Friman, V.-P.; Cao, K.; Yang, T.; Kramer, J.; Wang, X.; Li, M.; Mei, X.; et al. Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nat. Microbiol.* **2020**, *5* (8), 1002-1010. DOI: 10.1038/s41564-020-0719-8

(27) Johnson, L. J.; Koulman, A.; Christensen, M.; Lane, G. A.; Fraser, K.; Forester, N.; Johnson, R. D.; Bryan, G. T.; Rasmussen, S. An extracellular siderophore is required to maintain the mutualistic interaction of *Epichloë festucae* with *Lolium perenne*. *PLoS Pathog.* **2013**, *9* (5), e1003332. DOI: 10.1371/journal.ppat.1003332

(28) Mukherjee, P. K.; Hurley, J. F.; Taylor, J. T.; Puckhaber, L.; Lehner, S.; Druzhinina, I.; Schumacher, R.; Kenerley, C. M. Ferricrocin, the intracellular siderophore of *Trichoderma virens*, is involved in growth, conidiation, gliotoxin biosynthesis and induction of systemic resistance in maize. *Biochem. Biophys. Res. Commun.* **2018**, *505* (2), 606-611. DOI: 10.1016/j.bbrc.2018.09.170

(29) Schrettl, M.; Bignell, E.; Kragl, C.; Sabiha, Y.; Loss, O.; Eisendle, M.; Wallner, A.; Arst, H. N., Jr.; Haynes, K.; Haas, H. Distinct roles for intra- and extracellular siderophores during *Aspergillus fumigatus* infection. *PLoS Pathog.* **2007**, *3* (9), 1195-1207. DOI: 10.1371/journal.ppat.0030128

(30) Eng, T.; Herbert, R. A.; Martinez, U.; Wang, B.; Chen, J. C.; Brown, J. B.; Deutschbauer, A. M.; Bissell, M. J.; Mortimer, J. C.; Mukhopadhyay, A. Iron supplementation eliminates antagonistic interactions between root-associated bacteria. *Front. Microbiol.* **2020**, *11* (1742). DOI: 10.3389/fmicb.2020.01742

(31) Winterbourn, C. C. Toxicity of iron and hydrogen peroxide: the Fenton reaction. *Toxicol. Lett.* **1995**, *82-83*, 969-974. DOI: 10.1016/0378-4274(95)03532-x

(32) Valko, M.; Morris, H.; Cronin, M. T. Metals, toxicity and oxidative stress. *Curr. Med. Chem.* **2005**, *12* (10), 1161-1208. DOI: 10.2174/0929867053764635

(33) Mulvihill, E. D.; Moloney, N. M.; Owens, R. A.; Dolan, S. K.; Russell, L.; Doyle, S. Functional investigation of iron-responsive microsomal proteins, including MirC, in *Aspergillus fumigatus*. *Front. Microbiol.* **2017**, *8*, 418. DOI: 10.3389/fmicb.2017.00418

(34) Kurucz, V.; Kruger, T.; Antal, K.; Dietl, A. M.; Haas, H.; Pocsi, I.; Kniemeyer, O.; Emri, T. Additional oxidative stress reroutes the global response of *Aspergillus fumigatus* to iron depletion. *BMC Genom.* **2018**, *19* (1), 357. DOI: 10.1186/s12864-018-4730-x

(35) Misslinger, M.; Hortschansky, P.; Brakhage, A. A.; Haas, H. Fungal iron homeostasis with a focus on *Aspergillus fumigatus*. *Biochim. Biophys. Acta Mol. Cell Res.* **2021**, *1868* (1), 118885. DOI: 10.1016/j.bbamcr.2020.118885

(36) El-Sayed, M. T.; Ezzat, S. M.; Taha, A. S.; Ismaiel, A. A. Iron stress response and bioaccumulation potential of three fungal strains isolated from sewage-irrigated soil. *J. Appl. Microbiol.* **2022**, *132* (3), 1936-1953. DOI: 10.1111/jam.15372

(37) Roset, M. S.; Alefantis, T. G.; DelVecchio, V. G.; Briones, G. Iron-dependent reconfiguration of the proteome underlies the intracellular lifestyle of *Brucella abortus*. *Sci. Rep.* **2017**, *7* (1), 10637. DOI: 10.1038/s41598-017-11283-0

(38) Winters, M. S.; Spellman, D. S.; Chan, Q.; Gomez, F. J.; Hernandez, M.; Catron, B.; Smulian, A. G.; Neubert, T. A.; Deepe, G. S., Jr. *Histoplasma capsulatum* proteome response to decreased iron availability. *Proteome Sci.* **2008**, *6* (1), 36. DOI: 10.1186/1477-5956-6-36

(39) Coetzee, M. P. A.; Wingfield, B. D.; Bloomer, P.; Wingfield, M. J. Phylogenetic analyses of DNA sequences reveal species partitions amongst isolates of *Armillaria* from Africa. *Mycol. Res.* **2005**, *109* (11), 1223-1234. DOI: 10.1017/S095375620500393X

(40) Mwenje, E.; Wingfield, B. D.; Coetzee, M. P. A.; Wingfield, M. J. Molecular characterisation of *Armillaria* species from Zimbabwe. *Mycol. Res.* **2003**, *107* (3), 291-296. DOI: 10.1017/S0953756203007408

(41) Wingfield, B. D.; Berger, D. K.; Coetzee, M. P. A.; Duong, T. A.; Martin, A.; Pham, N. Q.; van den Berg, N.; Wilken, P. M.; Arun-Chinnappa, K. S.; Barnes, I.; et al. IMA genome-F17 : Draft genome sequences of an *Armillaria* species from Zimbabwe, *Ceratocystis colombiana, Elsinoe necatrix, Rosellinia necatrix,* two genomes of *Sclerotinia minor*, short-read genome assemblies and annotations of four *Pyrenophora teres* isolates from barley grass, and a long-read genome assembly of *Cercospora zeina*. *IMA Fungus* **2022**, *13* (1), 19. DOI: 10.1186/s43008-022-00104-3

(42) Schwyn, B.; Neilands, J. B. Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* **1987**, *160* (1), 47-56. DOI: 10.1016/0003-2697(87)90612-9

(43) Cox, C. D. [24] Deferration of laboratory media and assays for ferric and ferrous ions. In *Methods Enzymol.*, Clark, V. L., Bavoil, P. M. Eds.; Vol. 235; Academic Press Inc., 1994; pp 315-329.

(44) Giuliano Garisto Donzelli, B.; Gibson, D. M.; Krasnoff, S. B. Intracellular siderophore but not extracellular siderophore is required for full virulence in *Metarhizium robertsii*. *Fungal Genet. Biol.* **2015**, *82*, 56-68. DOI: 10.1016/j.fgb.2015.06.008

(45) Alexander, D. B.; Zuberer, D. A. Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. *Biol. Fert. Soils* **1991**, *12* (1), 39-45. DOI: 10.1007/bf00369386

(46) Collins, C.; Hurley, R.; Almutlaqah, N.; O'Keeffe, G.; Keane, T. M.; Fitzpatrick, D. A.; Owens, R. A. Proteomic characterization of *Armillaria mellea* reveals oxidative stress response mechanisms and altered secondary metabolism profiles. *Microorganisms* **2017**, *5* (3), 60. DOI: 10.3390/microorganisms5030060

(47) Ellero, A. A.; van den Bout, I.; Vlok, M.; Cromarty, A. D.; Hurrell, T. Continual proteomic divergence of HepG2 cells as a consequence of long-term spheroid culture. *Sci. Rep.* **2021**, *11* (1), 10917. DOI: 10.1038/s41598-021-89907-9

(48) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301-2319. DOI: 10.1038/nprot.2016.136

(49) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10* (4), 1794-1805. DOI: 10.1021/pr101065j

(50) Grigoriev, I. V.; Nikitin, R.; Haridas, S.; Kuo, A.; Ohm, R.; Otillar, R.; Riley, R.; Salamov, A.; Zhao, X.; Korzeniewski, F.; et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **2014**, *42* (Database issue), D699-704. DOI: 10.1093/nar/gkt1183

(51) The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47* (D1), D330-D338. DOI: 10.1093/nar/gky1055

(52) Tatusov, R. L.; Fedorova, N. D.; Jackson, J. D.; Jacobs, A. R.; Kiryutin, B.; Koonin, E. V.; Krylov, D. M.; Mazumder, R.; Mekhedov, S. L.; Nikolskaya, A. N.; et al. The COG database: an updated version includes eukaryotes. *BMC Bioinf.* **2003**, *4* (1), 41. DOI: 10.1186/1471-2105-4-41

(53) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36* (Database issue), D480-484. DOI: 10.1093/nar/gkm882

(54) Zhu, W.; Hu, J.; Chi, J.; Li, Y.; Yang, B.; Hu, W.; Chen, F.; Xu, C.; Chai, L.; Bao, Y. Label-free proteomics reveals the molecular mechanism of subculture induced strain degeneration and discovery of indicative index for degeneration in *Pleurotus ostreatus*. *Molecules* **2020**, *25* (21), 4920. DOI: 10.3390/molecules25214920

(55) Zeng, X.; Ling, H.; Yang, J.; Chen, J.; Guo, S. Proteome analysis provides insight into the regulation of bioactive metabolites in *Hericium erinaceus*. *Gene* **2018**, *666*, 108-115. DOI: 10.1016/j.gene.2018.05.020

(56) Parente, A. F.; Bailao, A. M.; Borges, C. L.; Parente, J. A.; Magalhaes, A. D.; Ricart, C. A.; Soares, C. M. Proteomic analysis reveals that iron availability alters the metabolic status of the pathogenic fungus *Paracoccidioides brasiliensis*. *PLoS One* **2011**, *6* (7), e22810. DOI: 10.1371/journal.pone.0022810

(57) Kroll, K.; Pahtz, V.; Kniemeyer, O. Elucidating the fungal stress response by proteomics. *J. Proteomics* **2014**, *97*, 151-163. DOI: 10.1016/j.jprot.2013.06.001

(58) Liu, Q.; Zhang, Y.; Wang, Y.; Wang, W.; Gu, C.; Huang, S.; Yuan, H.; Dhankher, O. P. Quantitative proteomic analysis reveals complex regulatory and metabolic response of *Iris lactea* Pall. var. *chinensis* to cadmium toxicity. *J. Hazard. Mater.* **2020**, *400*, 123165. DOI: 10.1016/j.jhazmat.2020.123165

(59) Zhong, Z.; Li, N.; Liu, L.; He, B.; Igarashi, Y.; Luo, F. Label-free differentially proteomic analysis of interspecific interaction between white-rot fungi highlights oxidative stress response and high metabolic activity. *Fungal Biol.* **2018**, *122* (8), 774-784. DOI: 10.1016/j.funbio.2018.04.005

(60) Poirier, I.; Hammann, P.; Kuhn, L.; Bertrand, M. Strategies developed by the marine bacterium *Pseudomonas fluorescens* BA3SM1 to resist metals: A proteome analysis. *Aquat. Toxicol.* **2013**, *128-129*, 215-232. DOI: 10.1016/j.aquatox.2012.12.006

(61) Zuo, L.; Zhou, T.; Pannell, B. K.; Ziegler, A. C.; Best, T. M. Biological and physiological role of reactive oxygen species - the good, the bad and the ugly. *Acta Physiol. (Oxf)* **2015**, *214* (3), 329-348. DOI: 10.1111/apha.12515

(62) Vaskova, J.; Kocan, L.; Vasko, L.; Perjesi, P. Glutathione-related enzymes and proteins: A review. *Molecules* **2023**, *28* (3), 1447. DOI: 10.3390/molecules28031447

(63) Bachhawat, A. K.; Yadav, S. The glutathione cycle: Glutathione metabolism beyond the gamma-glutamyl cycle. *IUBMB Life* **2018**, *70* (7), 585-592. DOI: 10.1002/iub.1756

(64) Pompella, A.; Visvikis, A.; Paolicchi, A.; De Tata, V.; Casini, A. F. The changing faces of glutathione, a cellular protagonist. *Biochem. Pharmacol.* **2003**, *66* (8), 1499-1503. DOI: 10.1016/s0006-2952(03)00504-5

(65) Singer, M. A.; Lindquist, S. Multiple effects of trehalose on protein folding *in vitro* and *in vivo*. *Mol. Cell* **1998**, *1* (5), 639-648. DOI: 10.1016/s1097-2765(00)80064-7

(66) Dunn, M. F.; Ramirez-Trujillo, J. A.; Hernandez-Lucas, I. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. *Microbiology (Reading)* **2009**, *155* (Pt 10), 3166-3175. DOI: 10.1099/mic.0.030858-0

(67) Brzóska, K.; Meczyńska, S.; Kruszewski, M. Iron-sulfur cluster proteins: electron transfer and beyond. *Acta Biochim. Pol.* **2006**, *53* (4), 685-691.

(68) Sang, H.; Lee, H. B. Molecular mechanisms of succinate dehydrogenase inhibitor resistance in phytopathogenic fungi. *Res. Plant Dis.* **2020**, *26* (1), 1-7. DOI: 10.5423/RPD.2020.26.1.1

(69) Anderson, J. B.; Catona, S. Genomewide mutation dynamic within a long-lived individual of *Armillaria gallica*. *Mycologia* **2014**, *106* (4), 642-648. DOI: 10.3852/13-367

(70) Hu, M.; Chen, S. Non-target site mechanisms of fungicide resistance in crop pathogens: A review. *Microorganisms* **2021**, *9* (3), 502. DOI: 10.3390/microorganisms9030502

(71) Aguin, O.; Mansilla, J. P.; Sainz, M. J. *In vitro* selection of an effective fungicide against *Armillaria mellea* and control of white root rot of grapevine in the field. *Pest Manag. Sci.* **2006**, *62* (3), 223-228. DOI: 10.1002/ps.1149

(72) Adaskaveg, J. E.; Forster, H.; Wade, L.; Thompson, D. F.; Connell, J. H. Efficacy of sodium tetrathiocarbonate and propiconazole in managing Armillaria root rot of almond on peach rootstock. *Plant Dis.* **1999**, *83* (3), 240-246. DOI: 10.1094/PDIS.1999.83.3.240

(73) Oexle, H.; Gnaiger, E.; Weiss, G. Iron-dependent changes in cellular energy metabolism: influence on citric acid cycle and oxidative phosphorylation. *Biochim. Biophys. Acta* **1999**, *1413* (3), 99-107. DOI: 10.1016/s0005-2728(99)00088-2

(74) Fernandez, J.; Marroquin-Guzman, M.; Wilson, R. A. Evidence for a transketolase-mediated metabolic checkpoint governing biotrophic growth in rice cells by the blast fungus *Magnaporthe oryzae*. *PLoS Pathog.* **2014**, *10* (9), e1004354. DOI: 10.1371/journal.ppat.1004354

(75) Huo, J.; Zhao, B.; Zhang, Z.; Xing, J.; Zhang, J.; Dong, J.; Fan, Z. Structure-based discovery and synthesis of potential transketolase inhibitors. *Molecules* **2018**, *23* (9), 2116. DOI: 10.3390/molecules23092116

(76) Yildirim, V.; Ozcan, S.; Becher, D.; Buttner, K.; Hecker, M.; Ozcengiz, G. Characterization of proteome alterations in *Phanerochaete chrysosporium* in response to lead exposure. *Proteome Sci.* **2011**, *9* (1), 12. DOI: 10.1186/1477-5956-9-12

(77) Chen, G.; Zhou, Y.; Zeng, G.; Liu, H.; Yan, M.; Chen, A.; Guan, S.; Shang, C.; Li, H.; He, J. Alteration of culture fluid proteins by cadmium induction in *Phanerochaete chrysosporium*. *J. Basic Microbiol.* **2015**, *55* (2), 141-147. DOI: 10.1002/jobm.201300398

(78) Singh, R. K.; Kefala, G.; Janowski, R.; Mueller-Dieckmann, C.; von Kries, J. P.; Weiss, M. S. The high-resolution Structure of LeuB (Rv2995c) from *Mycobacterium tuberculosis*. *J. Mol. Biol.* **2005**, *346* (1), 1-11. DOI: 10.1016/j.jmb.2004.11.059

(79) Huttel, W. Structural diversity in echinocandin biosynthesis: the impact of oxidation steps and approaches toward an evolutionary explanation. *Z. Naturforsch. C J. Biosci.* **2017**, *72* (1-2), 1-20. DOI: 10.1515/znc-2016-0156

(80) Hamer, L.; Adachi, K.; Dezwaan, T.; Lo, C.; Montenegro-Chamorro, M.; Frank, S.; Darveaux, B.; Mahanty, S.; Heiniger, R.; Skalchunes, A. Methods for the identification of inhibitors of 3-isopropylmalate dehydratase as antibiotics. United States of America 2004.

(81) Chipman, D.; Barak, Z.; Schloss, J. V. Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases. *Biochim. Biophys. Acta* **1998**, *1385* (2), 401-419. DOI: 10.1016/s0167-4838(98)00083-1

(82) Kingsbury, J. M.; McCusker, J. H. Cytocidal amino acid starvation of *Saccharomyces cerevisiae* and *Candida albicans* acetolactate synthase (ilv2Delta) mutants is influenced by the carbon source and rapamycin. *Microbiology (Reading)* **2010**, *156* (Pt 3), 929-939. DOI: 10.1099/mic.0.034348-0

(83) Jastrzebowska, K.; Gabriel, I. Inhibitors of amino acids biosynthesis as antifungal agents. *Amino Acids* **2015**, *47* (2), 227-249. DOI: 10.1007/s00726-014-1873-1

(84) Kim, J. Y.; Kim, E. J.; Lopez-Maury, L.; Bahler, J.; Roe, J. H. A metabolic strategy to enhance long-term survival by Phx1 through stationary phase-specific pyruvate decarboxylases in fission yeast. *Aging (Albany NY)* **2014**, *6* (7), 587-601. DOI: 10.18632/aging.100682

(85) Staudt, M. W.; Kruzel, E. K.; Shimizu, K.; Hull, C. M. Characterizing the role of the microtubule binding protein Bim1 in *Cryptococcus neoformans*. *Fungal Genet. Biol.* **2010**, *47* (4), 310-317. DOI: 10.1016/j.fgb.2009.12.010

(86) Banuett, F.; Quintanilla, R. H., Jr.; Reynaga-Pena, C. G. The machinery for cell polarity, cell morphogenesis, and the cytoskeleton in the Basidiomycete fungus *Ustilago maydis*-a survey of the genome sequence. *Fungal Genet. Biol.* **2008**, *45 Suppl 1* (Suppl 1), S3-S14. DOI: 10.1016/j.fgb.2008.05.012

(87) Xiang, X.; Fischer, R. Nuclear migration and positioning in filamentous fungi. *Fungal Genet. Biol.* **2004**, *41* (4), 411-419. DOI: 10.1016/j.fgb.2003.11.010

Figures

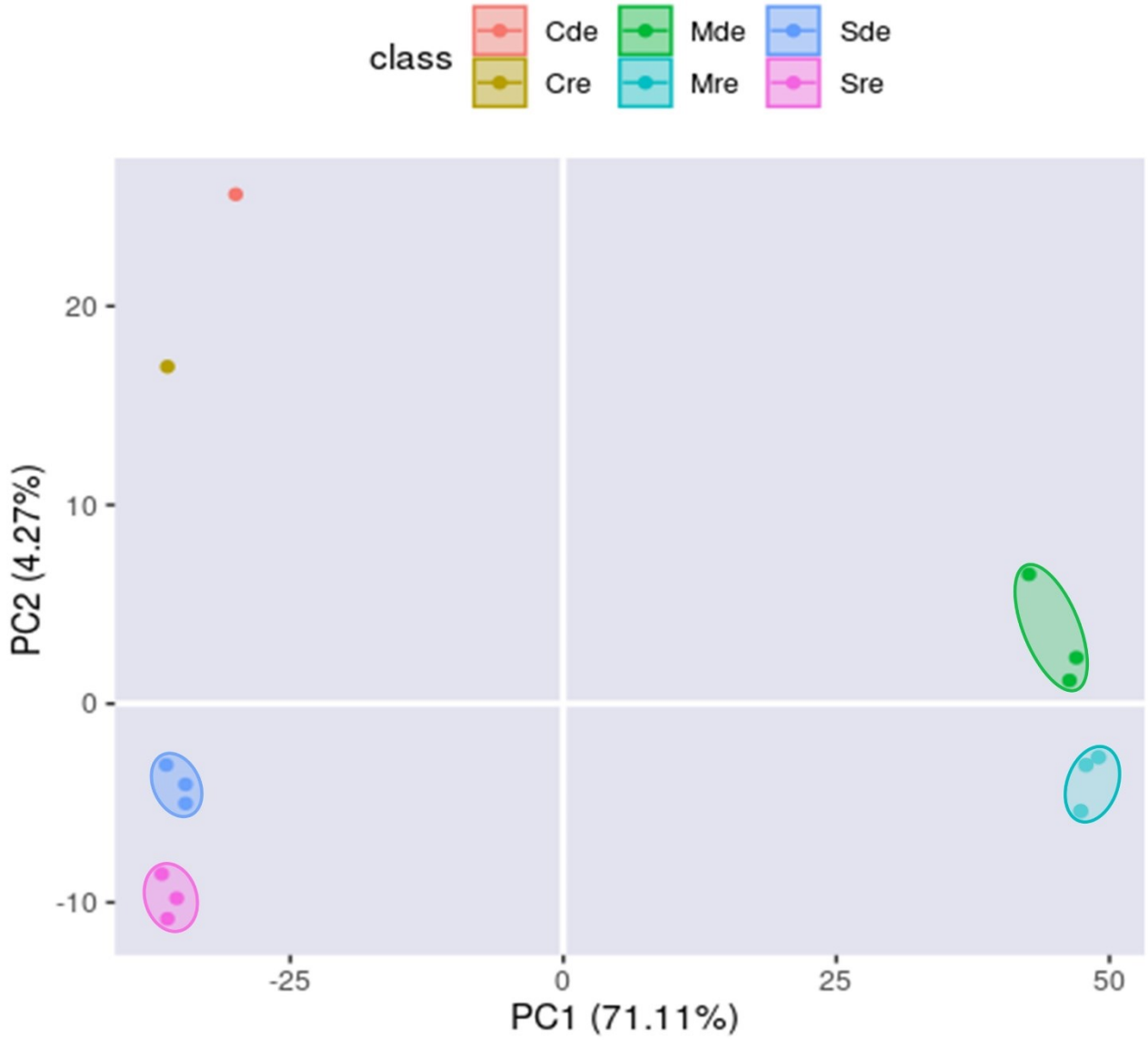


Figure 1: Score plot of principal component analysis (PCA) of samples comparing components 1 and 2
The horizontal axis is the first principal component (PC1), and the vertical axis is the second principal component (PC2). Corresponding color-coded oval shapes indicate clustering of the respective samples. Control for each treatment were uninoculated media (Cde and Cre; n = 1 each). Cde = iron-deplete medium (i.e., uninoculated medium without addition of $FeCl_3$; PDP-); Cre = iron-replete medium (i.e., uninoculated medium with added 100 µM $FeCl_3$; PDP+); Mde = iron-deplete mycelia (i.e., iron-deplete proteome); Mre = iron-replete mycelia (i.e., iron-replete proteome); Sde = iron-deplete supernatant (i.e., iron-deplete secretome); Sre = iron-replete supernatant (i.e., iron-replete secretome). n = 3 per sample group.

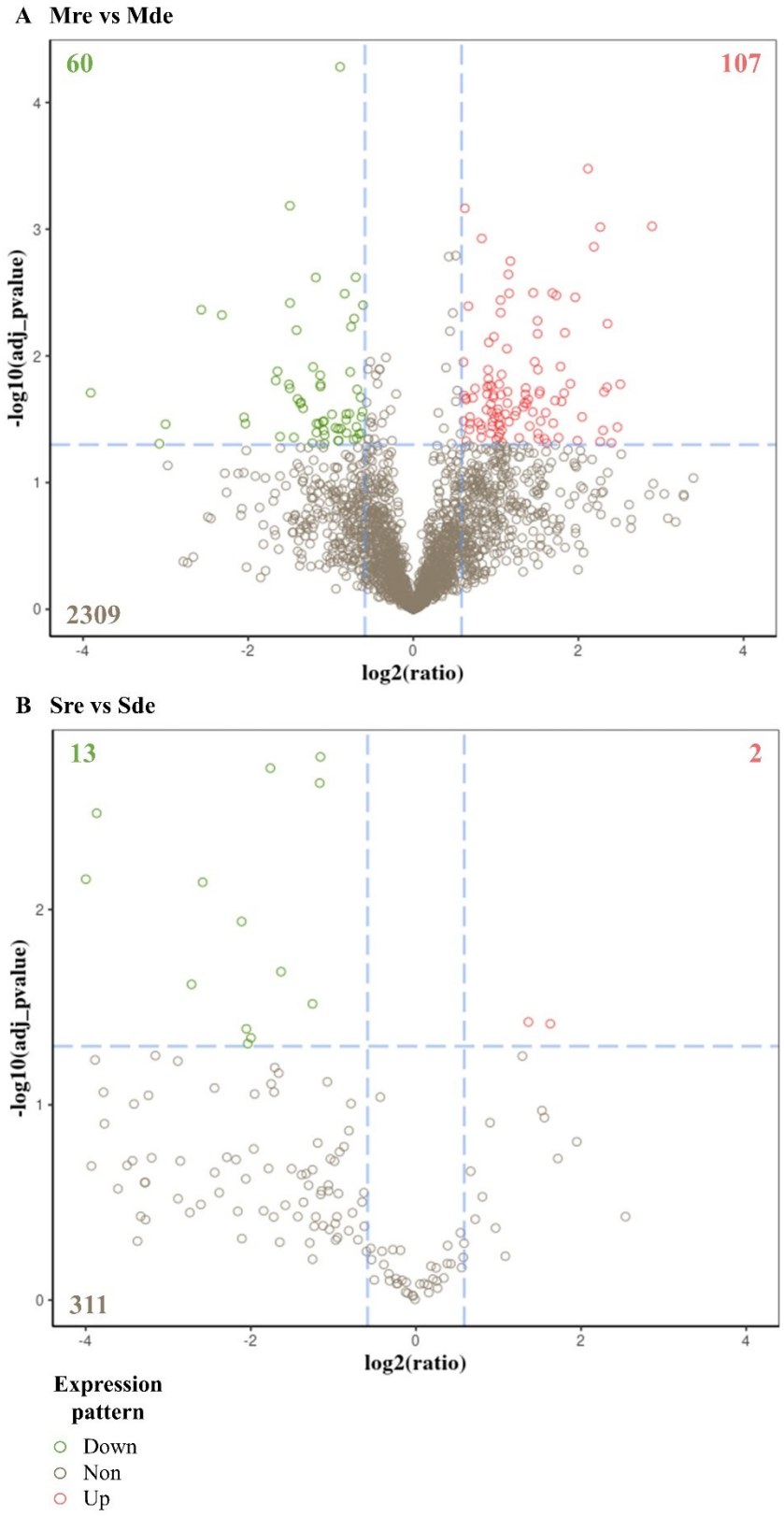


Figure 2: Volcano plots showing changes in protein expression.
(A) proteome under iron-replete compared to iron-deplete conditions (Mre vs Mde) and (B) secretome under iron-replete compared to iron-deplete conditions (Sre vs Sde). The x-axes are the protein fold change (log 2), and the y-axes are the corresponding -log10 (adjusted *p*-value). In the figures, the green circles are significantly down-expressed proteins, the red circles are significantly up-expressed proteins, and the brown circles are proteins with non-significantly altered expression. Corresponding color-coded numbers within the quadrants are the numbers of down-expressed, up-expressed, and non-significantly altered proteins respectively, under iron-replete compared to iron-deplete conditions.

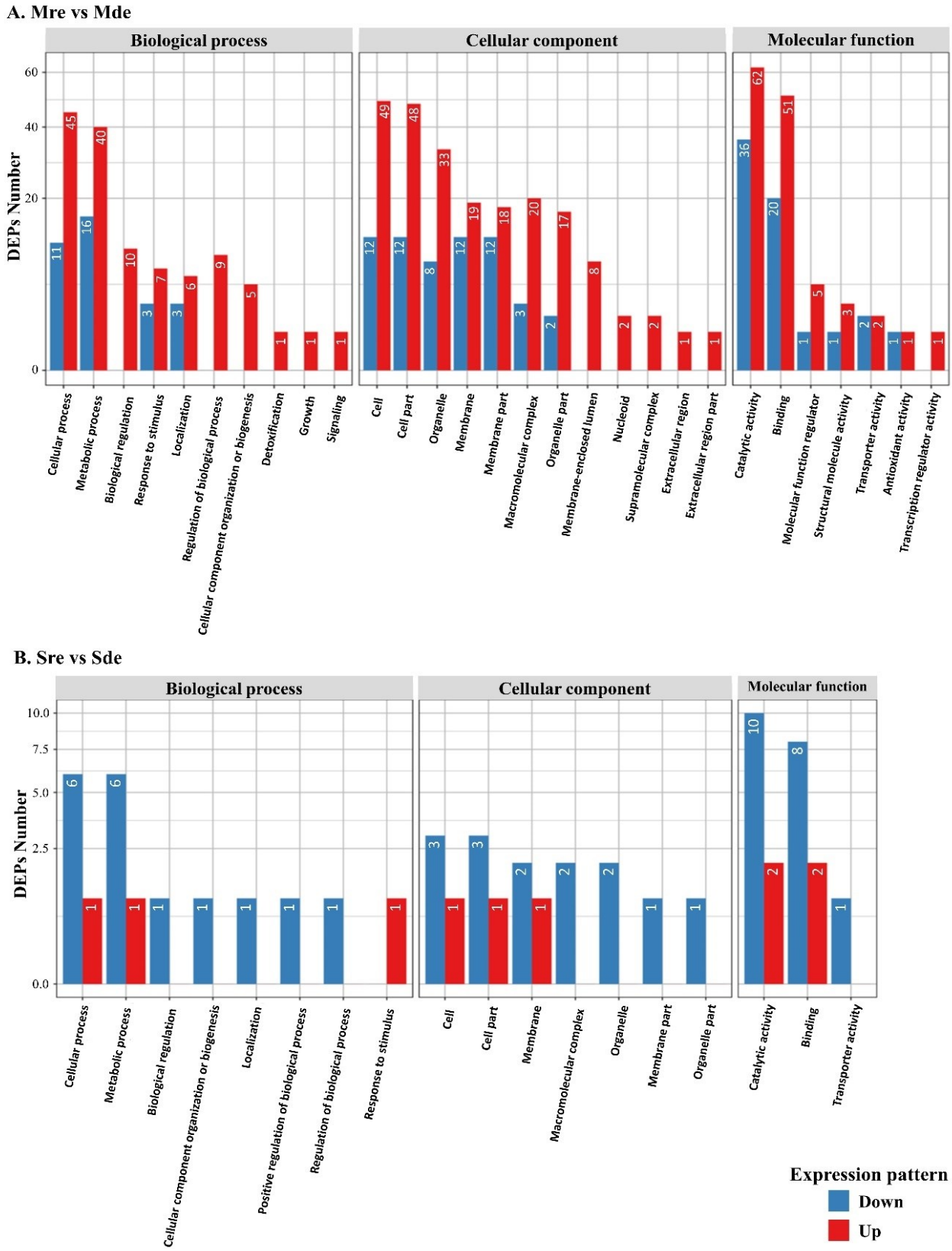
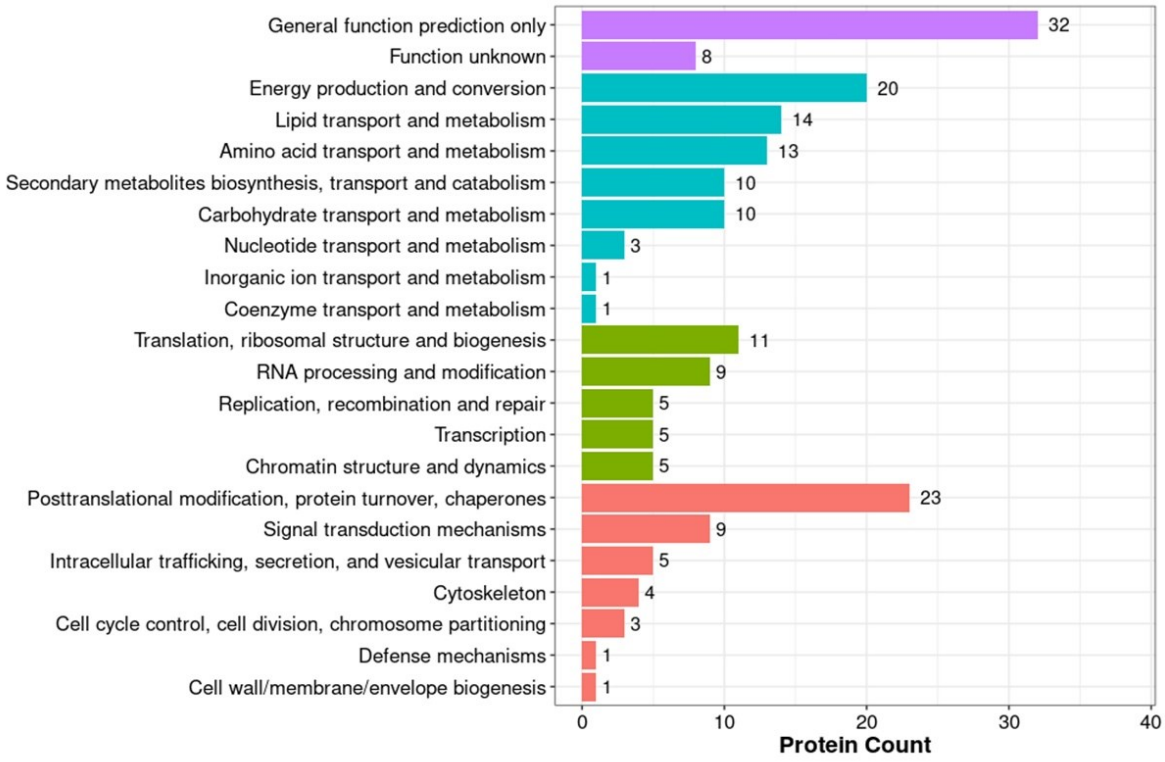


Figure 3: Significantly enriched Gene Ontology annotation and functional classification of the DEPs.
(A) proteome under iron-replete compared to iron-deplete conditions (Mre vs Mde) and (B) secretome under iron-replete compared to iron-deplete conditions (Sre vs Sde). The x-axes represent the GO annotation entries, classified into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). The y-axes represent the number of differentially up or down-expressed proteins (blue and red bars respectively). Numbers within the bars are the number of down or up-expressed proteins.

A. Mre vs Mde



B. Sre vs Sde

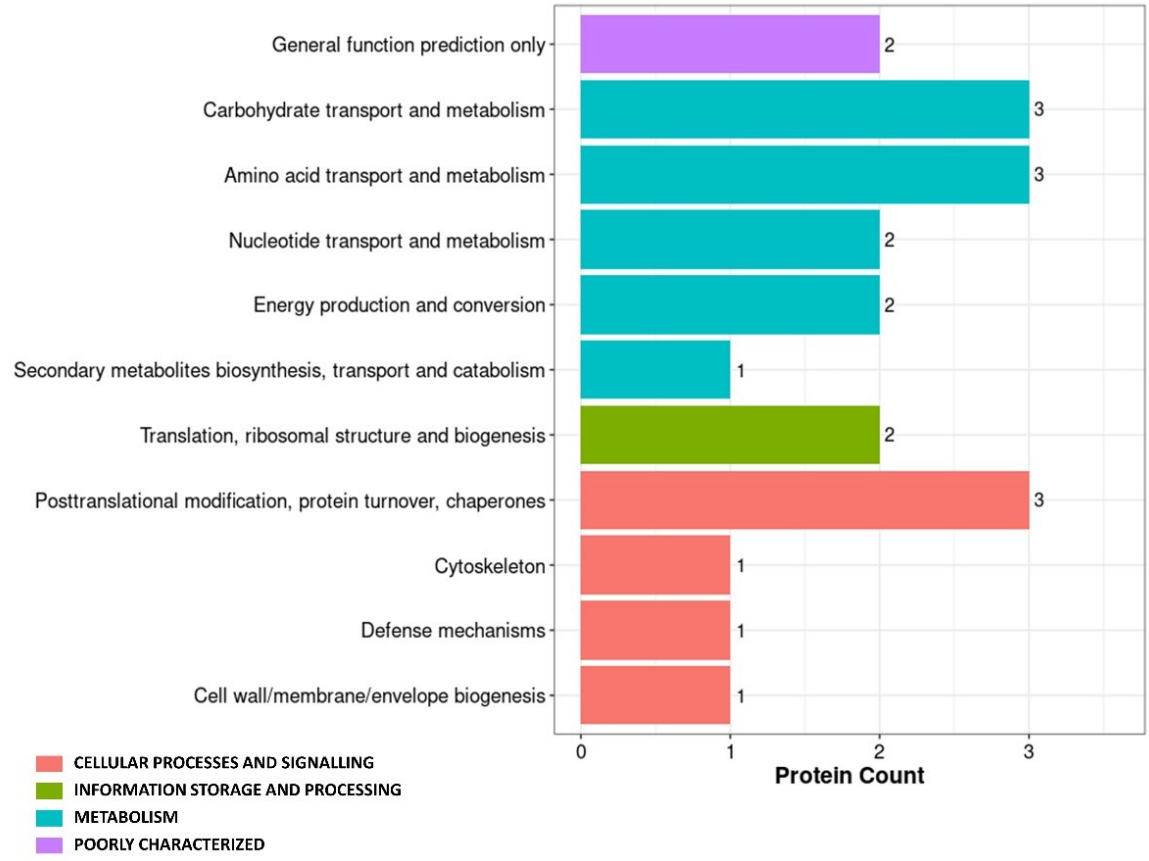


Figure 4: KOG classification of differentially expressed proteins.
(A) proteome under iron-replete compared to iron-deplete conditions (Mre vs Mde) and (B) secretome under iron-replete compared to iron-deplete conditions (Sre vs Sde). The x-axes display protein count, and the y-axes display KOG terms grouped under the terms, cellular processes and signalling, information storage and processing, metabolism, and poorly characterized. Numbers outside the bars are the number of DEPs.

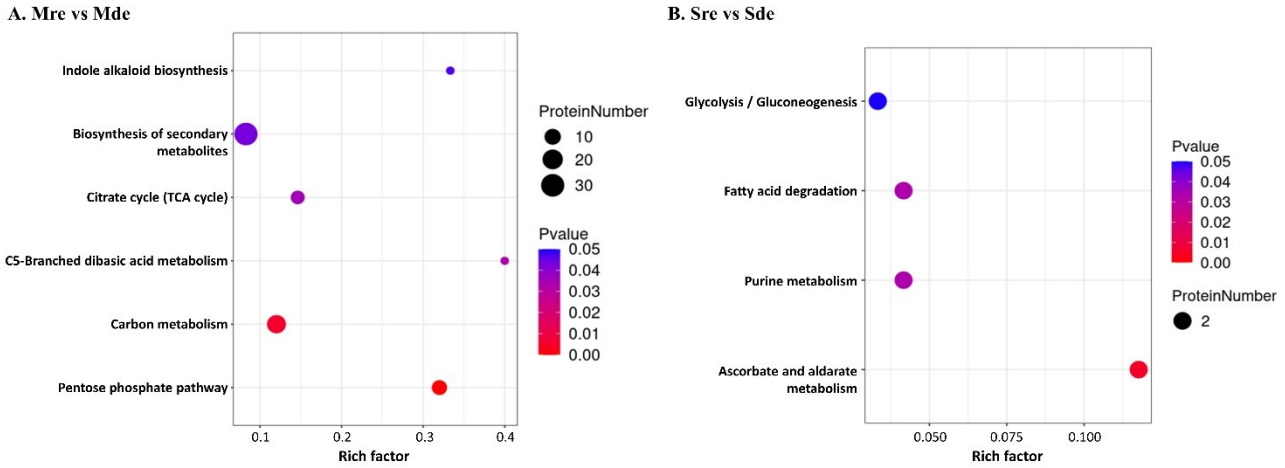


Figure 5: Bubble diagrams of KEGG pathway enrichment of DEPs showing the pathways in which the DEPs are significantly enriched.
(A) proteome under iron-replete compared to iron-deplete conditions (Mre vs Mde) and (B) secretome under iron-replete compared to iron-deplete conditions (Sre vs Sde). The Rich Factor (also known as enrichment factor; x-axes) is the ratio of the number of DEPs annotated to the pathway to all the proteins identified in the pathway. The greater the Rich Factor, the higher the degree of enrichment. The y-axes are the pathways enriched in DEPs. The bubble sizes represent the number of DEPs annotated to the pathway and the depth of the bubble colour represents the adjusted *p*-value.

Supplementary Material

Supplementary material S1

Supplementary material S1 can be found at the University of Pretoria website using the link, https://figshare.com/s/2fe55e301fe03a92ed16. The definitions of the headings can be found in the sheet named "Definitions". All proteins discussed in the manuscript have been highlighted.

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

The research presented in this thesis aimed at characterising some mechanisms utilised by selected *Armillaria* species for iron homeostasis using growth experiments, bioassays, genomics, and proteomics.

A review was conducted that considered the status of *Armillaria* genome projects and technologies that were used. The information that emerged from the review were compiled in a practical guide that includes infographics for genome sequencing, assembly, and annotation in the Basidiomycota and other diploid and dikaryotic fungi. This review revealed that even at the draft stage, *Armillaria* genomes can be used for some comparative genomics and other studies because they have sufficiently high contiguity and completeness. Knowledge gained from *Armillaria* genome projects were also highlighted and suggestions for further genomics studies were given. These included suggestions on how to achieve chromosome scale genome assembly and the relevance of adopting multi-disciplinary approaches to bridge the wide knowledge gap about the mechanisms used by these fungi in their evolutionary success. Several possible research areas which could benefit from the multi-disciplinary approach were also enumerated.

A multi-disciplinary approach was adopted to answer the following central questions in this thesis:

1. What are some of the cellular and molecular mechanisms of iron homeostasis of *Armillaria* species in comparison to other members of the Physalacriaceae?

2. In which ways does iron affect growth and other biological and molecular characteristics of *Armillaria* species?

To answer these questions, comparative genomics and bioinformatics was used to evaluate secondary metabolite gene clusters (SMGCs) in publicly available genomes of *Armillaria* and other species belonging to the Physalacriaceae. The study focused on NRPS-dependent siderophore synthetase (NDSS) gene clusters and NRPS-independent siderophore (NIS) synthetase gene clusters. Based on these studies, genomes of *Armillaria* spp. were shown to contain one highly conserved, potentially functional NDSS gene cluster. Additionally, this thesis contains the first record of two distinct NIS synthetase gene clusters in fungal genomes. These are important discoveries as fungi in the Basidiomycota are generally known to produce hydroxamate-type siderophores. These types of siderophores are usually synthesised by NDSS gene clusters. NIS synthetase gene clusters are well known in bacteria but not in fungi.

Two bioassays were used to investigate siderophore biosynthesis by selected *Armillaria* spp. qualitatively. These studies resulted in the first record of siderophore biosynthesis by *Armillaria* spp. A further study which investigated the effect of iron concentration on culture growth and siderophore

biosynthesis of selected *Armillaria* spp. also revealed that these plant-pathogenic fungi have unique requirements for iron and siderophore biosynthesis at relatively high concentrations of added iron. These studies provide the first evidence of the iron-bioaccumulation capacity of *Armillaria* spp.

A genome of an *Armillaria* sp. from Africa was generated in this study using both long- and short-read sequencing. This genome adds to the available genome data of species of *Armillaria.* This genome, in addition to other available genomes and other genomes to be sequenced in future, can be used for further studies. These studies can be used to understand the biology, chemistry, and pathogenicity and/or virulence of these socio-economically important fungi and to explore modes of inhibiting them and/or developing biotechnologically useful products. As a starting point, the predicted protein sequence of this genome was included in the database used for investigating the proteomic and secretomic alterations of cellular and molecular characteristics exhibited by the sequenced strain in response to iron supplementation. This study has provided the first proteomes and secretomes of an isolate of *Armillaria* from Africa. Additionally, this study revealed that supplementation of medium with 100 µM $FeCl_3$ enhanced growth of the fungus, although this concentration of added iron was apparently not sufficient to support all molecular and biological processes of the fungus. Potential targets for inhibition and attenuation of pathogenicity and/or virulence of the studied strain were also identified. These included limitation of iron availability, as well as developing and/or utilising antifungal substances which target biosynthesis of various amino acids and enzymes by the fungi.

Attempts were made to extract and partially characterise the siderophores produced by the studied strains using liquid cultures and high performance liquid chromatography. However, this study could not be pursued further due to the lack of a negative control as the studied strains produced siderophores up to 200 µM of added $FeCl_3$ as determined by the chrome azurol S (CAS) assay followed by spectrophotometry both before and after methanolic extraction. The wet-bench time limitations imposed during the COVID-19 pandemic also limited our ability to conduct more troubleshooting to optimise the experiments.

Nonetheless, the results of the studies conducted in this research have literally ‘opened the Pandora’s box’ in terms of iron homeostasis in *Armillaria* and in the Basidiomycota in general. More gaps concerning genomic, biological, and biochemical mechanisms of secondary metabolism and iron homeostasis in these macrofungi have been exposed. These include studies to fully characterize the genes in the predicted SMGCs in the Physalacriaceae which can be extended to studies on other Basidiomycetes, to determine the function and relevance of the genes in the SMGCs, to investigate the biological roles of the predicted introns in the NDSS genes, determine the biosynthetic model / overall biosynthesis pathways of the SMGCs, to characterize the products of the SMGCs, to

determine the ability of other organisms to utilize the products of the NDSS and NIS synthetase gene clusters as xenosiderophores, and to elucidate the role of the products in pathogenicity/virulence as well as other biological functions of the gene products, if any.

Techniques which can be used for the suggested studies include gene modification / knock out combined with biochemical characterization (e.g., enzymology, proteomics, and metabolomics). Enzymology studies can be conducted by extracting and purifying the individual gene products of the predicted genes and characterising these enzymes. Various proteomic approaches can be used to isolate and characterise various proteins which influence biosynthesis of the respective secondary metabolites when expressed under different conditions by these organisms. In terms of metabolomics, use of technologies including high performance liquid chromatography-mass spectrometry (HPLC-MS) and nuclear magnetic resonance (NMR) spectroscopy coupled with bioassays will be required to structurally and biochemically characterise the exact siderophores produced by these organisms. The purified and characterised siderophores can then be utilized for further studies such as development of some biotechnologically relevant products. Such studies can be conducted to investigate the other SMGCs predicted in the present study.

Experiments also need to be conducted on the use of various microbial communities which can reduce the iron availability for pathogenic *Armillaria* spp. to control Armillaria root- and stem-rot disease. A further research approach for controlling these phytopathogens is to investigate the efficacy of the presently available antifungal agents which inhibit biosynthesis of various amino acids by fungi, and/or developing more efficacious types of these kinds of antifungal agents.

Embarking on the suggested research will generate knowledge and technologies which can be used to limit these plant pathogens with a consequential effect on enhancing the attainment of Goal 2 – "End hunger, achieve food security and improved nutrition and promote sustainable agriculture" and Goal 15 – "Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss" with a consequent effect on Goal 1 – "End poverty in all its forms everywhere" of the 17 Sustainable Development Goals (SDGs) to transform our world.

Siderophores have also been explored extensively in the medical field for various purposes. Among others, the medical applications of siderophores include:

1. to deliver antibiotics,
2. serving as vaccines when coupled to carrier proteins,
3. for treatment of iron-related diseases when used for chelation therapy,
4. for limiting iron availability for tumour initiation, growth, and metastasis when used in cancer therapy,
5. for infection imaging when used as sensors for metal ion and pathogen detection, and
6. for acceleration of wound healing.

Research on application of siderophores produced by *Armillaria* spp. and/or their analogs in medicine and the eventual use of these technologies could potentially contribute towards achieving the above-mentioned SDGs as well as Goal 3 – "Ensure healthy lives and promote well-being for all at all ages" of the SDGs.

Clearly there is still much to be discovered about secondary metabolism in *Armillaria* spp. and fungi belonging to the Basidiomycota in general. Pursuing this research agenda offers many opportunities for human capacity development in terms of postgraduate research training. It also offers several job creation opportunities. These include post-doctoral and other research opportunities and jobs in the pharmaceutical and other industries. These will contribute to the attainment of Goal 4 – "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" and Goal 8 – "Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all" and consequently, enhance attainment of Goal 1 of the SDGs.

CV

The degree Doctor of Philosophy

Deborah Louisa **Narh Mensah**

In this thesis, **Characterising some mechanisms of iron homeostasis in selected *Armillaria* species**, the promovendus investigated mechanisms utilised by some selected *Armillaria* species for iron homeostasis using bioassays, genomics, and proteomics. A review was conducted on the status of, and technologies used for *Armillaria* genome projects. This was used to develop a practical guide for genome sequencing, assembly, and annotation of basidiomycetes and other diploid and dikaryotic fungi. One genome of an *Armillaria* isolate from Zimbabwe was sequenced. This resource was used for proteomic evaluation of the isolate in response to iron. Bioinformatics and comparative genomics analyses as well as bioassays conducted provided the first records of NRPS-dependent siderophore synthetase gene clusters and siderophore biosynthesis by *Armillaria* species. The study also serves as the first record of two NRPS-independent siderophore synthetase gene clusters in fungal genomes. The study further highlights potential novel avenues for controlling Armillaria root- and shoot-rot disease which affects diverse plants in agriculture and forestry worldwide.