Using collections to explore the evolution of plant associated lifestyles in the *Ascomycota*

Rowena Hill

2022

Submitted in partial fulfillment of the
requirements of the Degree of Doctor of Philosophy

Supervised by

Professor Richard J. A. Buggs
School of Biological and Behavioural Sciences
Queen Mary University of London

and

Dr Ester Gaya
Comparative Fungal Biology
Royal Botanic Gardens, Kew

Statement of originality

I, Rowena Hill, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 21/12/2022

Details of collaboration and publications

Author contributions and, where applicable, publication details are listed below. This work was supported by the London NERC DTP (NERC Ref: NE/L002485/1); the Evolution and Education Trust; and the Pragnell Fund. I use the pronoun 'we' in sections involving collaborators to acknowledge their contribution.

Chapter 1 - Rowena Hill wrote the chapter with feedback from Ester Gaya and Richard J. A. Buggs. Parts were published in papers described below, as well as a paper published in Fungal Biology Reviews (Hill, Leitch and Gaya, 2021); **R.H.** implemented the analysis and **R.H.**, Ilia J. Leitch and E.G. designed and wrote the paper.

Chapter 2 - This chapter was published in Frontiers in Microbiology (Hill, Llewellyn et al., 2021). **R.H.** and E.G. designed the study, implemented the analysis and wrote the paper. The paper used molecular data collected by **R.H.**, Theo Llewellyn, Elizabeth Downes, Joseph Oddy and Catriona MacIntosh prior to the start of this PhD project. Simon Kallow and John B. Dickie provided the samples and contributed to the writing of the paper, and S.K. performed the tetrazolium chloride testing. Bart Panis performed the embryo rescue testing.

Chapter 3 - This chapter was published in Molecular Biology and Evolution (Hill, Buggs et al., 2022). **R.H.** designed the study, performed the molecular lab work, implemented the analysis and wrote the paper. E.G. and R.J.A.B. supervised the work, designed the analysis and wrote the paper. Dang Toan Vu provided the samples and read and approved the final manuscript.

Chapter 4 - R.H. performed the molecular lab work, designed and implemented the bioinformatics analysis and wrote the chapter. Elena Arrigoni, Miguel Bonnin and Anthony Kermode assisted culturing and E.A. also assisted DNA extraction. Quentin Levicky assisted MinION sequencing. Amy Junnonen performed flow cytometry measurements under the supervision of Sahr Mian and I.J.L. Frances Pitsillides and A.J. assisted phylogenetic analyses. Alan G. Buddie provided the samples from CABI and guidance on culturing. E.G. and R.J.A.B. supervised the work and all collaborators provided feedback on the writing.

Chapter 5 - R.H. wrote the chapter with feedback from E.G. and R.J.A.B.

Abstract

The *Ascomycota* form the largest phylum in the fungal kingdom and show a wide diversity of lifestyles, some involving beneficial or harmful associations with plants. Distinguishing between fungal endophytes – species which live asymptomatically in plant tissues – and plant pathogens is of major significance to economic and ecological issues relating to plant health. Evolutionary genomics methods can provide insight into the genetic determinants of these lifestyles, and collections can act as an invaluable source of material to enable such analyses.

As endophytes are comparatively poorly studied, comparing plant associated lifestyles in the *Ascomycota* first requires novel endophyte discovery. In this thesis, I have demonstrated the unexplored promise of Kew's Millennium Seed Bank for isolating viable fungal endophytes and, in the process, highlighted the potential issues of overlooking the seed microbiome in the seed banking practice. I then performed whole genome sequencing, assembly and annotation of novel endophytic *Fusarium* strains for a case-study exploring lifestyle evolution in the genus. The distribution of lifestyles across the phylogeny; similarity of gene repertoires; and patterns of codon optimisation suggested that *Fusarium* taxa have a shared capacity for pathogenicity/endophytism. Exploring to what extent these results are common to different lineages of the *Ascomycota* requires the generation of new genomic resources for endophytes at large. Consequently, I sequenced, assembled and annotated genomes for a further 15 endophyte strains from CABI's collections, which spanned 8 families and 5 orders and additionally represent the first assembly for the genus and/or species for 7 of the strains. Together, this thesis demonstrates the value of existing plant and fungal collections for producing material and data to explore the pathogenic-mutualistic spectrum in plant associated ascomycetes.

Acknowledgements

Without a doubt I can say that my research career has been shaped by Ester Gaya – thank you for the unwavering confidence you have shown in me over 8 (!) years of mentorship, and for making sure I got firmly hooked on fungi. I must also thank Richard Buggs for welcoming me into his research group, and the many past and present members of the Buggs Lab, particularly Laura Kelly, for broadening my horizons at the weekly lab meetings. I have been privileged to work with countless lovely people at Kew, all of whom I've been grateful to call colleagues, but special thanks to Theo Llewellyn for being a constant source of solidarity. Thank you to Alan Buddie and fellow collaborators at CABI for being so generous with their time and resources. Thanks also to my London NERC DTP cohort, especially Book Club, for the support as we all navigated our PhDs together.

Thanks to my friends, family and in-laws who I love very much, and Marcus for everything.

Contents

List of Tables

List of Figures

List of Abbreviations

ANOSIM . . . . . analysis of similarity

ANOVA . . . . . analysis of variance

AR . . . . . autocorrelated-rates

ASV . . . . . amplicon sequence variant

CAZyme . . . . . carbohydrate-active enzyme

CSEP . . . . . candidate secreted effector protein

CV . . . . . coefficient of variation

CWR . . . . . crop wild relative

dN/dS . . . . . the ratio of nonsynonymous to synonymous substitutions

EF1α . . . . . translation elongation factor 1 alpha

FBRSC . . . . . *Fusarium burgessii* species complex

FCOSC . . . . . *Fusarium concolor* species complex

FDESC . . . . . *Fusarium decemcellulare* species complex

FFSC . . . . . *Fusarium fujikuroi* species complex

FHSC . . . . . *Fusarium heterosporum* species complex

FIESC . . . . . *Fusarium incarnatum-equiseti* species complex

FLSC . . . . . *Fusarium lateritium* species complex

FNSC . . . . . *Fusarium nisikadoi* species complex

FOSC . . . . . *Fusarium oxysporum* species complex

FSAMSC . . . . . *Fusarium sambucinum* species complex

FSSC . . . . . *Fusarium solani* species complex

FSTSC . . . . . *Fusarium staphylae* species complex

FTSC . . . . . *Fusarium tricinctum* species complex

IR . . . . . independent-rates

ITS . . . . . internal transcribed spacer

LSU . . . . . nuclear ribosomal large subunit

ML . . . . . maximum likelihood

MSB . . . . . Millennium Seed Bank

NMDS . . . . . non-metric multidimensional scaling

OTU . . . . . operational taxonomic unit

PCA . . . . . principal component analysis

PCWDE . . . . . plant cell wall degrading enzyme

PERMANOVA . . . . . permutational multivariate analysis of variance

PERMDISP . . . . . permutational analysis of multivariate dispersions

PGLS . . . . . phylogenetic generalised least squares

RPB1 . . . . . RNA polymerase II largest subunit

RPB2 . . . . . RNA polymerase II second largest subunit

RSCU . . . . . relative synonymous codon usage

T-BAS . . . . . Tree-Based Alignment Selector toolkit

WGS . . . . . whole genome sequencing

Chapter 1

Introduction

Publication details

Parts of this chapter have been published in the following papers:

> **Hill, R.**, Leitch, I. J., Gaya, E. (2021). Targeting Ascomycota genomes: what and how big? *Fungal Biology Reviews* 36:52-59. DOI: 10.1016/j.fbr.2021.03.003.

R.H. implemented the analysis and **R.H.**, I.J.L. and E.G. designed and wrote the paper.

> **Hill, R.**, Llewellyn, T., Downes, E., Oddy, J., MacIntosh, C., Kallow, S., Panis, B., Dickie, J.B. and Gaya, E. (2021). Seed Banks as Incidental Fungi Banks: Fungal Endophyte Diversity in Stored Seeds of Banana Wild Relatives. *Frontiers in Microbiology* 12:643731. DOI: 10.3389/fmicb.2021.643731.

R.H. and E.G. designed the study, implemented the analysis and wrote the paper. The paper used molecular data collected by **R.H.**, T.L., E.D., J.O. and C.M. prior to the start of this PhD. S.K. and J.D. provided the samples and contributed to the writing of the paper, and S.K. performed the tetrazolium chloride testing. B.P. performed the embryo rescue testing.

> **Hill, R.**, Buggs, R.J.A., Vu, D.T., Gaya, E. (2022). Lifestyle transitions in fusarioid fungi are frequent and lack clear genomic signatures. *Molecular Biology and Evolution* 39(4):msac085. DOI: 10.1093/molbev/msac085.

R.H. designed the study, performed molecular lab work, implemented the analysis and wrote the paper. E.G. and R.J.A.B. supervised the study, designed the analysis and wrote the paper. D.T.V. provided the samples and read and approved the final manuscript.

1.1 The endophytic lifestyle

Fungi are known to have a wide range of associations with plants, from infamous plant pathogens to mutualistic symbionts. The latter are even thought to have played an essential role in plants successfully colonising the land more than 500 million years ago (Pirozynski and Malloch, 1975; Heckman et al., 2001; Taylor and Krings, 2005; Chang et al., 2015; Field and Pressel, 2018; Morris, Puttick et al., 2018; Strullu-Derrien et al., 2018). Fungal endophytes (hereafter, endophytes) are

fungi which live asymptomatically inside plant tissues, and they appear to be present in all land plants (Petrini, 1991; Stone, Bacon and White, 2000; Rodriguez, White Jr et al., 2009; Hardoim et al., 2015; Rashmi, Kushveer and Sarma, 2019; Harrison and Griffin, 2020). Endophytes are known to belong predominantly to the *Ascomycota*, the largest phylum of the *Fungi* containing ∼105,000 of the ∼155,000 described species (68%) in Species Fungorum as of December 2022 (P. Kirk, personal communication; `http://www.speciesfungorum.org/`). Besides endophytes and phytopathogens, the phylum also comprises other economically and environmentally important lifestyles such as animal mutualists and pathogens, saprotrophs and lichenised fungi, which makes the *Ascomycota* an ideal framework for exploring fungal lifestyle evolution.

Certain endophytes are known to provide benefits to the plant host, such as stress tolerance, growth promotion and disease resistance (Redman, Sheehan et al., 2002; Rodriguez, Redman and Henson, 2004; Waller et al., 2005; Bilal et al., 2018). Numerous endophytic species are additionally insect pathogens and thus deter plant pests (Vidal and Jaber, 2015; Vega, 2018), with some species even shown to transfer nitrogen from the insect they have infected and killed directly to plant hosts (Behie and Bidochka, 2014), encouraging the hope that they can be used in agriculture as potential pest and pathogen biocontrol agents. The value of this would be that they could ideally replace or reduce ecologically harmful chemical controls and aid sustainable intensification of agriculture without increased use of chemical fertilisers (Waller et al., 2005; Card et al., 2016; Le Cocq et al., 2016; Kandel et al., 2017; Bamisile et al., 2018; Vega, 2018; De Silva et al., 2019). Indeed, multiple endophytic *Trichoderma* species and *Beauveria bassiana* are already used commercially as biocontrol agents in a range of crops (Woo et al., 2014; Mascarin and Jaronski, 2016; Mawar, Manjunatha and Kumar, 2021). Additionally, endophytes can produce a suite of secondary metabolites as part of the plant–fungal interaction, providing a valuable opportunity for discovery of useful bioactive compounds such as antivirals and antibiotics (Schulz, Boyle et al., 2002; Gupta et al., 2020), amongst other diverse applications (Prescott et al., 2018).

The role of endophytes in plant health is more complicated than it first seems, however. The ultimate outcome of endophyte colonisation can be highly dependent on the context of the plant–fungal interaction, such as the status of the plant immune system and nutrient conditions (Junker, Draeger and Schulz, 2012; Lahrmann et al., 2015; Hacquard et al., 2016; Hiruma et al., 2016), as well as the presence of other endophytes within the microbiome (Redman, Dunigan and Rodriguez, 2001; Durán et al., 2018; Mesny, Miyauchi et al., 2021; Wolinska et al., 2021) and even light conditions (Álvarez-Loayza et al., 2011; Garnica et al., 2022). The transient status of endophytism for many taxa is evident from observations of endophytes becoming saprotrophs or pathogens following some change in host or abiotic conditions (Slippers and Wingfield, 2007; Arnold, Miadlikowska et al., 2009; Promputtha et al., 2010; Delaye, García-Guzmán and Heil, 2013; Swett and Gordon, 2015; Nelson et al., 2020). In some cases, however, an evolutionary transition from pathogenicity to endophytism may represent a permanent switch to obligate commensalism or mutualism (Gazis et al., 2016), and it has also been hypothesised that endophytism may be an ancestral 'waiting room' for the evolution of mycorrhizal symbiosis (Selosse, Schneider-Maunoury and Martos, 2018; Selosse, Petrolli et al., 2022). Mycorrhizal fungi form mutualistic associations with plant roots, where the fungal partner makes mineral nutrients available in exchange for carbon from the host plant (van der Heijden et al., 2015; Genre et al., 2020). Mycorrhizas that are located inside host root tissues (i.e., arbuscular, orchid and ericoid mycorrhizas) are sometimes conflated with root endophytes, however I am not including them in the definition of endophytism here as they produce specialised mycorrhizal structures for resource

exchange – e.g., arbuscules, hyphal coils and pelotons – and can influence root tissue development (van der Heijden et al., 2015; Genre et al., 2020; Selosse, Petrolli et al., 2022).

The concept that the term endophyte represents a range of functional roles within the plant host has been referred to as the 'endophytic continuum' (Saikkonen, Faeth et al., 1998; Schulz and Boyle, 2005). Of the 399 species classified as endophytes in the FUNGuild database (Nguyen et al., 2016) as of October 2022: 153 (38%) were also classified as plant pathogens; 27 as saprotrophs (7%); and 22 (6%) as other various guilds. Indeed, in phylogenetic analysis, endophytes are commonly found to be closely related to pathogens and saprotrophs, as well as endolichenic fungi, their lichen associated counterpart (Arnold, Miadlikowska et al., 2009; U'Ren, Dalling et al., 2009; U'Ren, Lutzoni, Miadlikowska and Arnold, 2010). A switch from commensal to pathogenic has been observed in some endophytes due to unfavourable environmental conditions (Slippers and Wingfield, 2007; Ribeiro et al., 2020), and there is evidence that endophytes found only in living tissues do not significantly differ in cellulolytic activity (i.e., decomposing capacity) from those found only in dead leaves (U'Ren and Arnold, 2016).

Endophytes that are apparently obligately non-pathogenic are sometimes referred to as 'true' endophytes (Mishra, Bhattacharjee and Sharma, 2021; Collinge, Jensen and Jørgensen, 2022), the most famous example in the *Ascomycota* being mutualistic grass endophytes belonging to the genus *Epichloë* (Tadych, Bergen and White Jr., 2014; Saikkonen, Young et al., 2016). While *Epichloë* species are capable of exhibiting antagonistic behaviour to their grass host at times (Schardl, 1996), Ewald (1987) asserts that it is the net effect on the host's fitness over its entire lifetime that is important in defining whether an interaction is a mutualism. Endophytes may reduce plant fitness in one regard, but improve it in another to such a degree that the interaction is net positive for the plant (Rudgers et al., 2012). Newman, Gillis and Hager (2022) argue that, in addition to looking at the interaction across the lifetime of the host, the key to categorising certain *Epichloë* species as mutualists rather than parasites is the fact that they are vertically transmitted from host parent to offspring, meaning that there is selective pressure on the endophyte to reinforce successful reproduction of the plant host.

That is not to say that endophytes which are always or sometimes transmitted horizontally cannot be mutualists, only that it is not required for their persistence (Rodriguez, White Jr et al., 2009; Newman, Gillis and Hager, 2022). Arbuscular mycorrhizal fungi are horizontally transmitted and not host-specific, yet are considered to represent a stable mutualism where cooperation is thought to be maintained exactly because both plant and fungal partners can discriminate based on the relative costs and benefits of the interaction (Kiers, Duhamel et al., 2011; Noë and Kiers, 2018; van der Heijden et al., 2015; Põlme et al., 2018; Bennett and Groten, 2022; Semchenko et al., 2022). It is generally assumed that the majority of endophytes are horizontally transmitted from other plant individuals and/or the environment based on the frequent occurrence of many endophytic species in other niches. There are few examples of experimental verification as to whether certain endophytes are vertically transmitted, horizontally transmitted, or both (e.g., Tintjer, Leuchtmann and Clay, 2008; Wiewióra, Żurek and Pañka, 2015) and vertical transmission may be more widespread than currently documented (Harrison and Griffin, 2020). Habitual testing of endophyte transmission routes would be extremely informative for investigating individual plant–endophyte interactions, however the laborious nature of doing so combined with the magnitude of estimated endophyte diversity would make it challenging to achieve on a broad scale.

Improving our understanding of the endophytic lifestyle is a pressing issue in the context of global

change. We know that fungi are impacted by human-induced global change factors such as reduced host availability, nitrogen deposition, elevated atmospheric CO_2, altered precipitation and climate warming (Boddy, 2016; Bidartondo et al., 2018; Nic Lughadha et al., 2020). Plant pathogenic fungi already represent a major threat to our ecosystems and crops (Dean et al., 2012; Fisher, Gurr et al., 2020) which will only be exacerbated by global change – Delgado-Baquerizo et al. (2020) predict that relative abundance of soilborne fungal phytopathogens will increase globally with warming. Experiments and meta-analysis show that global change factors shift fungal soil communities to be dominated by generalist species (Rillig et al., 2019; Zhou, Wang and Luo, 2020), potentially at the cost of specialised mutualist species. A review of 179 empirical studies by Kiers, Palmer et al. (2010) found that most mutualisms are degraded by global change factors, which could result in a loss of the interaction, extinction of the species or even a shift from mutualistic to pathogenic. Baldrian et al. (2022) call this hypothetical shift among fungi "likely the largest potential threat for the future functioning of natural and managed ecosystems". Aside from environmental factors, globalisation has removed geographical barriers to enable the spread of known pathogens, but also the emergence of novel pathogens (Fisher, Henk et al., 2012; Sikes et al., 2018; Fones et al., 2020). As we know that harmless endophytes of one plant species can be harmful pathogens of another, it is likely that biosecurity measures underestimate the risk of moving even asymptomatic plant materials (Crous, Groenewald et al., 2016; Burgess et al., 2016; Cleary et al., 2019). In the light of global change, endophytes represent a rich pool of fungi from which new pathogens may emerge.

1.2 Genetic features of plant–fungal interactions

Most of our understanding of plant-fungal interactions to date has been through the lens of plant pathology, as the mechanisms by which fungal pathogens infect plants to cause disease have been extensively studied (e.g., Flor, 1971; Mendgen, Hahn and Deising, 1996; Dangl and Jones, 2001; van der Does and Rep, 2017). One key aspect of pathogenesis is expression of small secreted proteins referred to as effectors, which help the fungus to subvert host detection and the plant immune response (Stergiopoulos and de Wit, 2009; de Jonge, Bolton and Thomma, 2011; Toruño, Stergiopoulos and Coaker, 2016; Franceschetti et al., 2017; Shen, Liu and Naqvi, 2018; Singh, Nair and Verma, 2021). While effectors were initially only discussed in the context of establishing disease, considering many non-pathogenic fungi also have the ability to colonise plant tissues without triggering the plant immune response, it is perhaps unsurprising that the expression of effectors is not unique to pathogens, but is in fact an essential component of broader plant-fungal interactions (Rafiqi et al., 2012; Stergiopoulos, Kourmpetis et al., 2012; Lo Presti et al., 2015; Plett and Martin, 2015; Shen, Liu and Naqvi, 2018). Even in pathogens, expression of effectors is highest in biotrophic infection stages when pathogens are keeping host cells alive, rather than the actively damaging necrotrophic stage (van der Does and Rep, 2017).

Our knowledge of effectors beyond plant pathogens is best in mycorrhizal fungi, as multiple effectors have been identified in ecto-, arbuscular and ericoid mycorrhizal taxa which alter plant host behaviour to promote the symbiosis (Kloppholz, Kuhn and Requena, 2011; Plett, Kemppainen et al., 2011; Casarrubia et al., 2018; Zeng et al., 2019; Plett, Plett et al., 2020). Although endophytes generally appear to have a comparable number of effectors as other plant associated lifestyles (e.g., Mesny, Miyauchi et al., 2021), less is known about whether there are effectors which are specialised to the endophytic lifestyle. Eaton et al. (2015) identified fourteen putative effector genes in the grass endophyte *Epichloë festucae* which were differentially expressed in a wild-type mutualistic

strain versus three plant-antagonistic mutant strains, suggesting that they may indeed be involved in maintaining a mutualistic endophytic interaction with the plant host. Redkar et al. (2022) demonstrated that deletion of 'early root colonisation' effectors in an endophytic *Fusarium* strain resulted in impaired colonisation, however the same effectors were also shown to be essential to virulence in a pathogenic *Fusarium* strain. As these effectors also have homologues in many other fungi outside of the genus *Fusarium*, it suggests that they are part of a 'core' effector machinery common to multiple plant associated lifestyles, rather than being specialised to an endophytic lifestyle.

The concept of core effectors might be seen as somewhat paradoxical, as little to no sequence similarity between species is often treated as a defining feature of phytopathogen effectors, due to rapid diversification in the evolutionary 'arms-race' with the plant host (Franceschetti et al., 2017). For instance, obligate biotrophic, host-specific rust fungi have a high proportion of species-specific effectors (Beckerson et al., 2019). In addition to the aforementioned study by Redkar et al. (2022) there are also multiple examples, however, of effectors which are conserved across families, or even the entire kingdom (de Jonge, Esse et al., 2010; Stergiopoulos, Kourmpetis et al., 2012; Hemetsberger et al., 2015; Irieda et al., 2019). This suggests that effectors can broadly be grouped into those that are common to all or many taxa and deliver functions that are fundamental to the plant–fungal interaction, while others occur in a single species or lineage and are highly specialised for their niche. Much like in phytopathogens, many effectors found in different lineages of mycorrhizal fungi are species-specific, although Plett and Martin (2015) hypothesise that these may have evolved convergently to play similar functional roles, rather than being indicators of extreme specialisation, which corresponds with the fact that many mycorrhizal fungi are not host-specific (van der Heijden et al., 2015; Põlme et al., 2018; Semchenko et al., 2022). Comparing the effector repertoires of endophytes to other plant associates may shed light on where individual taxa fall on the endophytic continuum, and reveal whether there is a distinct effector toolkit that enables the endophytic lifestyle.

Another frequently studied component of the plant–fungal interaction are carbohydrate-active enzymes (CAZymes), enzymes which build, modify or break down carbohydrates and carbohydrate-linked molecules known as glycoconjugates (Cantarel, Coutinho et al., 2009). They are classified under six classes – glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, auxiliary activities and carbohydrate-binding modules (Drula et al., 2022) – each with different catalytic machinery to target different substrates. In plant associated fungi, many CAZymes are plant cell wall degrading enzymes (PCWDEs), acting on the major plant cell wall substrates of cellulose, cutin, hemicellulose, lignin and pectin (Glass et al., 2013; Hage and Rosso, 2021). The repertoire of CAZymes in a fungus will depend on their plant host(s) – as different plants will have different cell wall makeup – and lifestyle of the fungus (Kubicek, Starr and Glass, 2014).

As CAZymes are required to break down plant matter they have often been thought of as saprotrophic or pathogenic features (Kubicek, Starr and Glass, 2014; Lebreton et al., 2021), but they are also abundant in endophytes (Zhao, Liu et al., 2013; Knapp et al., 2018; Mesny, Miyauchi et al., 2021), certain lichenised fungi (Resl et al., 2022) and ericoid mycorrhizal fungi (Martino et al., 2018). Unlike effectors, a proliferation of CAZymes is not common to all plant associated lifestyles, as one of the hallmarks of the transition to ecto- and arbuscular mycorrhizal symbiosis is a decrease in total numbers of CAZymes (Kohler et al., 2015; Peter et al., 2016; Miyauchi, Kiss et al., 2020). However, specific CAZymes play key roles in remodelling the plant cell wall to enable the establishment and maintenance of ectomycorrhizal symbiosis (Veneault-Fourrey, Commun et al., 2014; Doré et al., 2017; Marqués-Gálvez et al., 2021), highlighting that they can be implicated in mutualistic interactions as

well as pathogenic or saprotrophic.

With the influx of whole genome data, there are many new opportunities to assess the patterns of effector and CAZyme content in different fungal species and lifestyles. Identifying CAZymes from genomic data is generally done on a sequence similarity basis using 30 years of curated CAZyme sequences from the CAZy database as reference material (Drula et al., 2022). Predicting and annotating putative effector genes from genomic data requires more complex bioinformatics pipelines. These typically include screening for signal peptides (extracellular secretion signals), followed by a series of steps to filter out motifs which contradict secretion, such as the exclusion of genes encoding transmembrane domains or GPI-anchors, which would indicate that they are lodged in or anchored to the cell membrane (Sonah, Deshmukh and Bélanger, 2016; Dalio et al., 2018; Beckerson et al., 2019). Machine learning methods trained on validated effector sequences have also recently been developed, which can be used in tandem with secretion prediction as mentioned above (Sperschneider and Dodds, 2021). When genes encoding effectors are predicted computationally in these ways, they are often referred to as candidate secreted effector proteins (CSEPs). Like all computational predictions produced using bioinformatics tools, they can only provide a hypothesis that a gene encodes an effector protein, and so CSEPs that are of particular interest for further study ultimately require experimental validation to determine their function. The pathogen–host interactions database (PHI-base) collates and curates genes, including effector genes from fungi, that have been experimentally verified as being involved in pathogen–host interactions (Urban et al., 2020). While this makes PHI-base an incredibly valuable resource, the development of a similar resource for non-pathogenic microbe–host interactions would also be desirable to capture the full range of interactions that exist.

Exploring the genetic features of endophytes versus pathogens requires an approach that can account for relatedness of taxa, namely an 'evolutionary genomics' approach. Many comparisons of genomic content between lifestyles make inferences about convergent patterns without accounting for lineage evolution (e.g., Lo Presti et al., 2015; Lebreton et al., 2021) even though shared ancestry alone can often describe more of the genetic variation than lifestyle (Krijger et al., 2014; Miyauchi, Kiss et al., 2020; Mesny, Miyauchi et al., 2021). Phylogeny is the hypothetical evolutionary history of a group of organisms, usually represented visually as a branching tree of life. Nowadays, phylogeny is an integral component of biological classification, which underpins our understanding of all living things. Phylogenetics is particularly crucial for fungal classification, as morphological features alone are insufficient to deal with the prevalent cryptic speciation and phenotypic convergence in the kingdom (Crous and Groenewald, 2005; Shivas and Cai, 2012). Phylogenies are also essential tools for exploring evolutionary processes and create a foundation from which functional traits can be compared between closely or distantly related organisms. For these reasons, a robust phylogenetic framework is a prerequisite for comparing the genetic content of endophytic and phytopathogenic taxa. In order to build upon the latest genome-scale *Ascomycota* phylogenies (Choi and Kim, 2017; Shen, Steenwyk et al., 2020) and explore the gene repertoires implicated in the plant–fungal interaction, we need genome assembly data.

1.3 A summary of the available genomic data for *Ascomycota*

It is only within the last 20 years that huge leaps forward in the development of sequencing technologies have enabled whole genome sequencing (WGS) on a broad scale. This first came with the commercialisation of second generation (massively parallel, short-read) sequencing in 2005 and then,

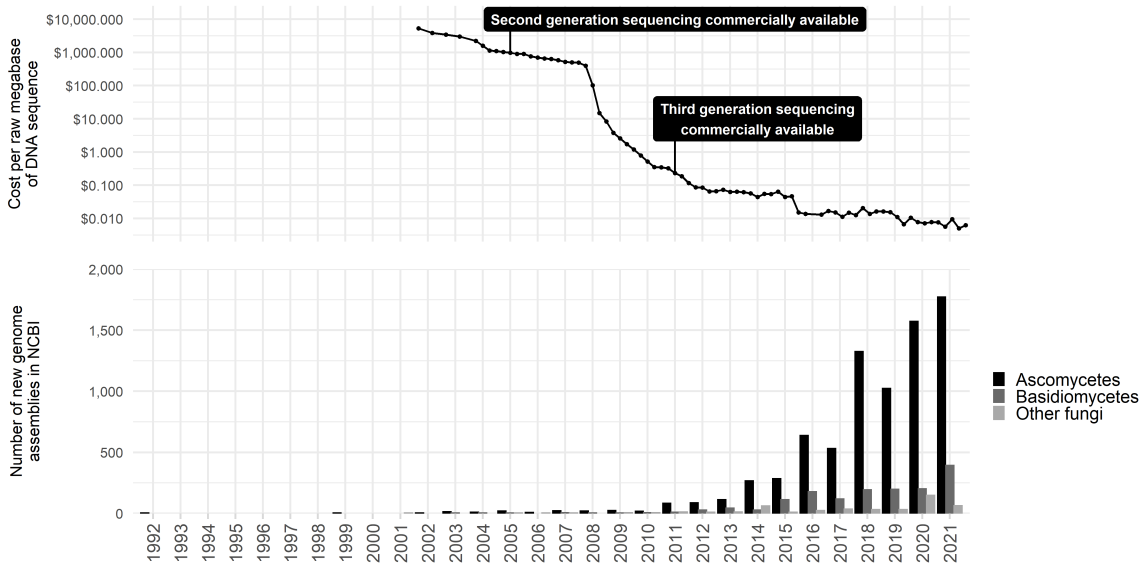


Figure 1.1: The downward trend in cost of sequencing alongside the number of fungal genome assemblies available in NCBI (`https://www.ncbi.nlm.nih.gov/`, downloaded on 28th October 2022). MycoCosm (Grigoriev et al., 2014) data is not included due to difficulty in obtaining release dates for assemblies. Sequencing costs are in US$ and were downloaded from the National Human Genome Research Institute (`https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data`) on the 28th October 2022 and visualised using ggplot2 v3.3.5 (Wickham, 2016) in R v4.1.2 (R Core Team, 2020).

only a few years later in 2011, third generation (real-time, single-molecule, long-read) sequencing (Shendure et al., 2017; Athanasopoulou et al., 2022). These advancements were accompanied by an extreme drop in the cost of sequencing (Figure 1.1), although it should be noted that cost still remains a barrier to utilising WGS in many low and middle income countries (Helmy, Awad and Mosa, 2016). Additionally, while sequencing itself may become cheaper, there are also the costs of associated resources to consider, such as storage of massive WGS datasets and computing power for genome assembly (Sboner et al., 2011; Muir et al., 2016).

Constructing a *de novo* (i.e., 'from scratch') genome assembly using raw WGS data involves the application of assembly algorithms to piece together overlapping reads into continuous sequences (Miller, Koren and Sutton, 2010; Meng et al., 2022). A huge number of assembly tools have been developed based on these algorithms – see Appendix A.1.3 for a non-exhaustive list – which can vary in accuracy and efficiency (Zhang, Chen et al., 2011; Abbas, Malluhi and Balakrishnan, 2014; Utturkar et al., 2014; Khan et al., 2018). Although *de novo* assembly is possible using only short-read data, the resulting assemblies can be highly fragmented (Paszkiewicz and Studholme, 2010; Richards, 2018), particularly due to the challenge of reconstructing repetitive regions in the genome (Miller, Koren and Sutton, 2010; Tørresen et al., 2019). As the ultimate aim of high quality genome assembly is to minimise fragmentation – i.e., for eukaryotes, to have each chromosome captured in its entirety as one continuous sequence – the introduction of long-reads that can span difficult to assemble regions has dramatically improved the ability to produce 'finished' or 'complete' assemblies (English et al., 2012; Utturkar et al., 2014; Koren and Phillippy, 2015; Jiao and Schneeberger, 2017).

Long-reads are not a cure-all for genome assembly, however. Long-reads come with higher error rates (Meng et al., 2022), which can impact downstream protein prediction (Watson and Warr,

2019). Hybrid assembly approaches using both long- and short-reads can help to maximise both contiguity and accuracy (Utturkar et al., 2014; Rice and Green, 2019), although the feasibility of using these approaches may be limited by insufficient material and/or funds to perform two rounds of sequencing. Current long-read sequencing methods rely on extraction of high molecular weight DNA, which can be difficult to produce for microbes, especially those that are challenging or impossible to isolate in culture (Tedersoo, Albertsen et al., 2021). While long-read sequencing has enabled the production of numerous high-standard ascomycete reference genomes (e.g., Faino et al., 2015; Baroncelli, Pensec et al., 2021; Voorhies et al., 2022), it is unlikely to fully supplant short-read sequencing for wider WGS projects in the near future, especially those at the population level (Jiao and Schneeberger, 2017).

The sequencing revolution has given rise to many ambitious WGS initiatives, which ultimately aim to record the full genetic code of all life (e.g., Robinson et al., 2011; Cheng et al., 2018; Lewin et al., 2018; The Darwin Tree of Life Project Consortium, 2022). The Kingdom *Fungi* is no exception and in 2011 the 1000 Fungal Genomes Project (`https://mycocosm.jgi.doe.gov/mycocosm/home/1000-fungal-genomes`) launched with plans to sequence two reference genomes for each fungal family, contributing to genome assemblies for at least 6,500 fungal strains available in NCBI (`https://www.ncbi.nlm.nih.gov/`) and MycoCosm (Grigoriev et al., 2014) as of October 2022. While most existing fungal assemblies belong to the *Ascomycota* (Figure 1.1), considerable taxonomic gaps remain in the genomic data available for the phylum – as of January 2021, 63 of the 126 orders (50%) and 2 of the 19 classes (10%) in the *Ascomycota* (*sensu* Wijayawardene et al., 2018) had no representative genome assembly (Figure 1.1). This included species-rich orders such as the *Meliolales* (2,379 spp.) and *Asterinales* (1,161 spp.), both of which are known for obligate plant associate species (Hongsanan, Li et al., 2014; Hongsanan, Tian et al., 2015). Orders missing genomic data will vary in e.g., phylogenetic position, ecological/functional diversity and species richness, but all represent significant gaps in the study of ascomycete evolution inasmuch as they indicate missing genomic data for a whole group of species, the level at which evolutionary and ecological processes occur.

Regarding lifestyles, WGS has generally been biased towards pathogenic taxa (Aylward et al., 2017), which is unsurprising considering their relevance to human interests. However, recent WGS efforts are rapidly improving genomic resources for other lifestyles – of all the assemblies which are assigned to lifestyles in the MycoCosm repository, for instance, the number of non-pathogenic plant associates and saprotrophs are catching up with phytopathogens (Figure 1.2). As of October 2022, there are 132 assemblies of ascomycete endophytes in MycoCosm, 80 of which have been published in 15 studies. The oldest of these studies dates back to 2012, showing just how recently WGS of these fungi started being tackled. Building on these endophyte genome resources is an essential first step to ensure balanced taxon sampling of different lifestyles when reconstructing ascomycete lifestyle evolution.

1.4 Producing high quality fungal genome assemblies: why size matters

Production of high quality genome assemblies is contingent on many factors such as confidence in specimen identification, avoiding contamination and isolation of sufficient DNA. One factor which can be comparatively easily addressed, and yet little focused on in fungi, is knowledge of genome size. When selecting appropriate short-read WGS protocols, determining the number of reads required to obtain sufficient coverage for a high quality genome assembly – estimated to be at least 50× for fungi

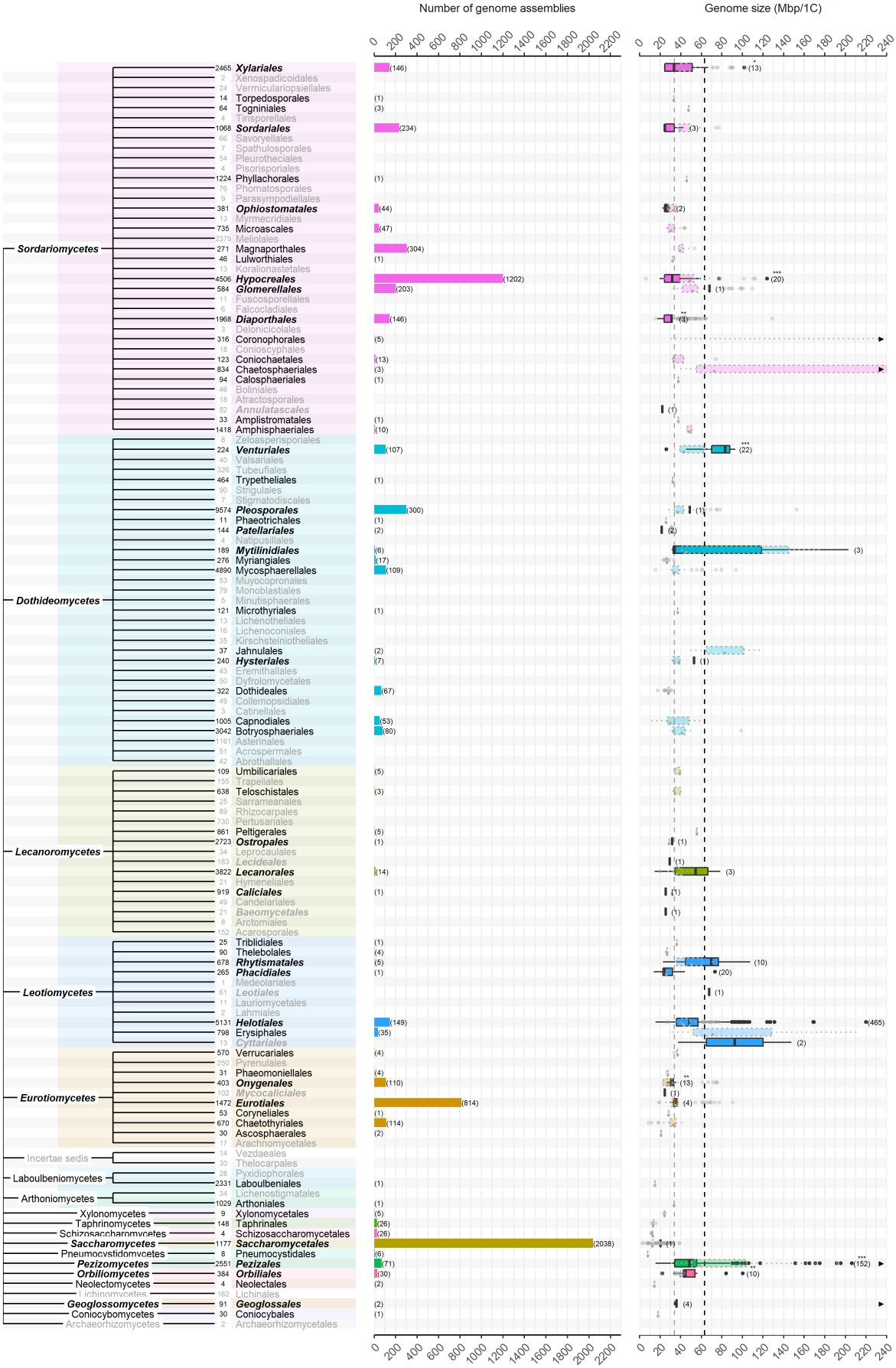

Number of genome assemblies
Genome size (Mbp/1C)
Sordariomycetes
Dothideomycetes
Lecanoromycetes
Leotiomycetes
Eurotiomycetes
Incertae sedis
Laboulbeniomycetes
Arthoniomycetes
Xylonomycetes
Taphrinomycetes
Schizosaccharomycetes
Saccharomycetes
Pneumocystidomycetes
Pezizomycetes
Orbiliomycetes
Neolectomycetes
Lichinomycetes
Geoglossomycetes
Coniocybomycetes
Archaeorhizomycetales
2465 Xylariales
2 Xenospadicoidales
24 Vermiculariopsiellales
14 Torpedosporales
64 Togniniales
4 Tirisporellales
1068 Sordariales
66 Savoryellales
7 Spathulosporales
54 Pleurotheciales
4 Pisorisporiales
1224 Phyllachorales
76 Phomatosporales
9 Parasympodiellales
381 Ophiostomatales
13 Myrmecridiales
735 Microascales
2379 Meliolales
271 Magnaporthales
46 Lulworthiales
13 Koralionastetales
4506 Hypocreales
584 Glomerellales
11 Fuscosporellales
6 Falcocladiales
1968 Diaporthales
3 Deionicicolales
316 Coronophorales
18 Conosphaerales
123 Coniochaetales
834 Chaetosphaeriales
94 Calosphaeriales
46 Boliniales
18 Atractosporales
52 Annulatascales
33 Amplistromatales
1418 Amphisphaeriales
8 Zeloasperisporiales
224 Venturiales
40 Valsariales
326 Tubeufiales
464 Trypetheliales
90 Strigulales
7 Stigmatodiscales
9574 Pleosporales
11 Phaeotrichales
144 Patellariales
4 Natipusillales
189 Mytilinidiales
276 Myriangiales
4890 Mycosphaerellales
53 Muyocopronales
79 Monoblastiales
5 Minutisphaerales
121 Microthyriales
13 Lichenotheliales
16 Lichenoconiales
35 Kirschsteiniotheliales
37 Jahnulales
240 Hysteriales
43 Eremithallales
50 Dyfrolomycetales
322 Dothideales
49 Collemopsidiales
3 Catinellales
1005 Capnodiales
3042 Botryosphaeriales
1161 Asterinales
51 Acrospermales
42 Abrothallales
109 Umbilicariales
155 Trapeliales
638 Teloschistales
25 Sarrameanales
89 Rhizocarpales
730 Pertusariales
861 Peltigerales
2723 Ostropales
34 Leprocaulales
183 Lecideales
3822 Lecanorales
21 Hymeneliales
919 Caliciales
49 Candelariales
21 Baeomycetales
8 Arctomiales
152 Acarosporales
25 Triblidiales
90 Thelebolales
678 Rhytismatales
265 Phacidiales
1 Medeolariales
61 Leotiales
11 Lauriomycetales
2 Lahmiales
5131 Helotiales
798 Erysiphales
13 Cyttariales
570 Verrucariales
250 Pyrenulales
31 Phaeomoniellales
403 Onygenales
102 Mycocaliciales
1472 Eurotiales
53 Coryneliales
670 Chaetothyriales
30 Ascosphaerales
17 Arachnomycetales
14 Vezdaeales
30 Thelocarpales
28 Pyxidiophorales
2331 Laboulbeniales
34 Lichenostigmatales
1029 Arthoniales
9 Xylonomycetales
148 Taphrinales
4 Schizosaccharomycetales
1177 Saccharomycetales
8 Pneumocystidales
2551 Pezizales
384 Orbiliales
4 Neolectales
182 Lichinales
91 Geoglossales
30 Coniocybales
2 Archaeorhizomycetales
0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200
0 20 40 60 80 100 120 140 160 180 200 220 240

Figure 1.2: A summary showing the taxonomy of the different classes and orders currently recognised in the *Ascomycota* (*sensu* Wijayawardene et al., 2018), with number of genome assemblies in NCBI and Mycocosm (central bar graph) and range of genome sizes per order (right hand boxplot) as of January 2021. Black taxon labels indicate taxa with representative genome assemblies versus grey for no genome assemblies and bold-italic labels indicate taxa with representative cytometric genome size estimates versus plain text for no cytometric genome size estimates. The number of species for each order according to Species Fungorum (`http://www.speciesfungorum.org/`) is shown to the left of taxon labels. Boxplots of 762 genome size measurements (from 504 species) made using cytometric approaches are taken from the Fungal Genome Size Database (Kullman, Tamm and Kullman, 2005) and are shown using opaque colours (the sample sizes for each order are shown in brackets). In contrast, boxplots for 6,600 genome sizes (from 3,273 strains) based on genome assemblies are given in translucent colours; here the number of samples per order are the same as the number in brackets given for number of genome assemblies shown in the central bar graph. Asterisks (*) shown above the sample size in the genome size boxplots indicate orders with significant differences in mean genome sizes between cytometric and assembly-based estimates (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The black dashed line shows the mean genome size of all *Ascomycota* species estimated using cytometric methods, whereas the grey dashed line corresponds to the mean genome size from all estimates obtained from genome assembly data. For the sake of visualisation, extreme outliers are not shown – black arrows on the far right indicate orders with genome size data exceeding the x-axis – but can be seen in Appendix A.1.2. For full methodology also see Appendix A.1.1

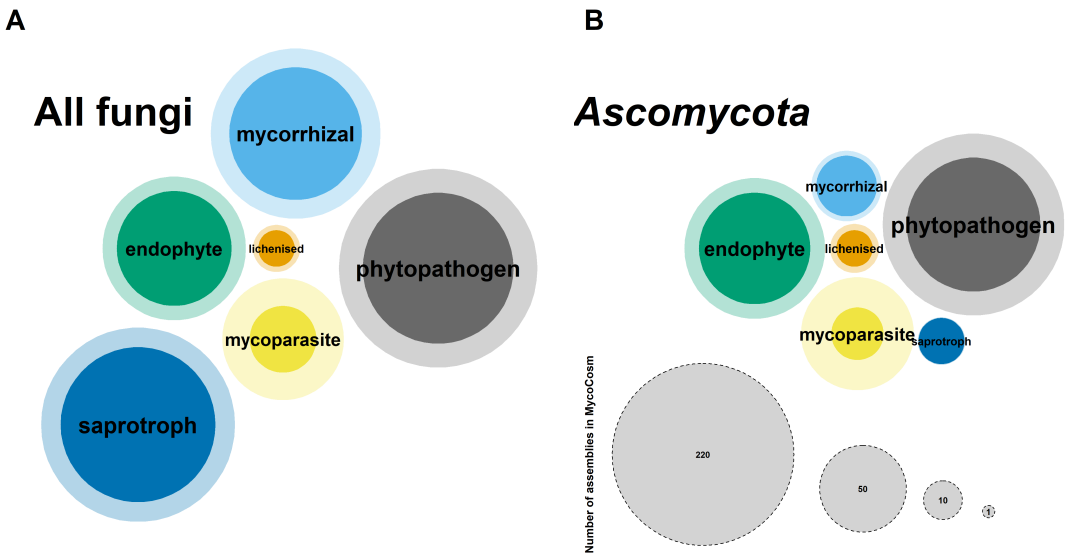


Figure 1.3: A summary of the number of genome assemblies for different fungal lifestyles available in MycoCosm (`https://mycocosm.jgi.doe.gov/mycocosm/home`), across **(A)** all fungi and **(B)** the *Ascomycota*. Data was scraped from the website on 25/10/2022 using the package rvest v1.0.2 (Wickham, 2020) in R v4.1.2 (R Core Team, 2020) and visualised using the packages packcircles v0.3.4 (Bedward, Eppstein and Menzel, 2020) and ggforce v0.3.3 (Pedersen, 2021). Darker inner circles indicate the number of published assemblies, while lighter outer circles indicate the total number of assemblies including those that have not yet been published.

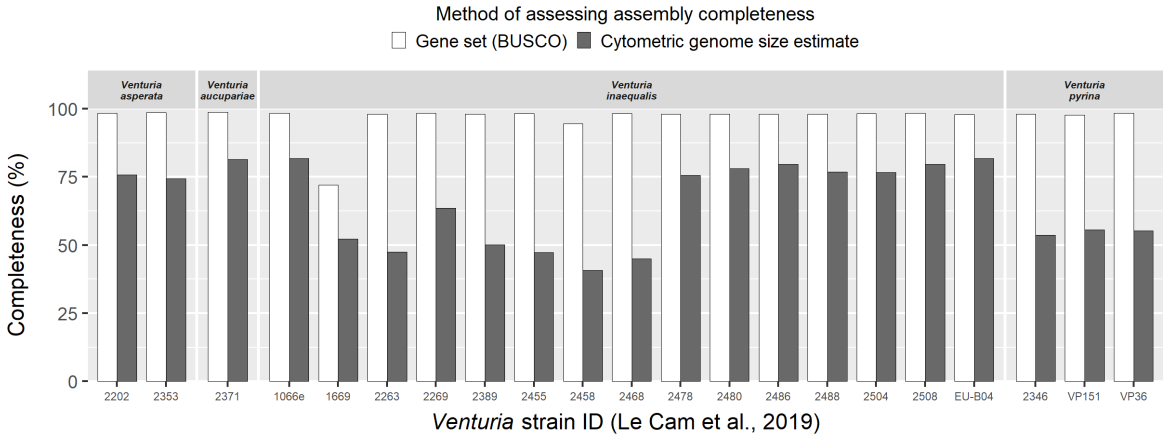


Figure 1.4: Genome assembly completeness as measured by gene set (BUSCOs) versus cytometric genome size estimation for strains of four *Venturia* species (*Venturiales*, *Dothideomycetes*) from Le Cam et al. (2019).

(Desai et al., 2013) – is conditional on reliable estimation of the species' genome size. Additionally, with an ever-increasing variety of *de novo* genome assembly tools available, prior knowledge of genome size can act as a metric to assess assembly quality (e.g., Mita et al., 2004; Yoshida et al., 2011; Kooij and Pellicer, 2020), and is already required for certain long-read assembly protocols (e.g., Ruan and Li, 2020). Conceptually, we can divide the methods used to estimate genome size into two groups – 1) those inferred bioinformatically from WGS data and 2) those estimated using cytometric methods, of which Feulgen microdensitometry and, more recently, flow cytometry are the two most widely used approaches, with the latter now being the method of choice (Bennett and Leitch, 2011; D'hondt et al., 2011; Talhinhas, Tavares et al., 2017).

Assembly quality is usually interpreted from metrics based on the number and size of contigs/scaffolds (e.g., as calculated by QUAST; Gurevich et al., 2013), as well as measures of gene set completeness (e.g., using BUSCO; Simão et al., 2015), but neither of these approaches can guarantee 'correctness' (Studholme, 2016). Indeed, a high BUSCO completeness can be reported from an assembly that is less than 50% complete according to cytometric genome size estimation (Figure 1.4). Such discrepancies between BUSCO completeness and the proportion of the whole genome that is actually sequenced and assembled, highlights the potential to miss large amounts of biologically important yet non-coding DNA sequences (e.g., regulatory regions, transcription factors, repetitive DNA). This emphasises the importance of having a robust cytometric estimate as an additional metric to evaluate assembly tool performance.

Maximising assembly quality is not trivial, as it can impact subsequent gene annotation and therefore evolutionary and functional inferences regarding gene loss/gain (Denton et al., 2014; Deutekom et al., 2019; Kooij and Pellicer, 2020). Furthermore, inadvertent collapsing of repetitive regions by assembly tools (Tørresen et al., 2019) can also compromise studies seeking to understand the biological significance of repetitive DNA (e.g., Seidl, Kramer et al., 2020). For example, genes with potential pathogenicity roles in ascomycete phytopathogens, such as those encoding effector proteins and secondary metabolites, have been found to occur in repeat-rich regions which are vulnerable to misassembly (Raffaele and Kamoun, 2012; Rao et al., 2018). There is a proliferation of highly repetitive regions in various obligate plant associates: the powdery mildews (*Erysiphales*) have a high proportion of repetitive DNA due to an abundance of retrotransposons (Spanu et al., 2010),

and similar is seen in the mycorrhizal species *Cenococcum geophilum* (Peter et al., 2016) and *Tuber melanosporum* (Veneault-Fourrey and Martin, 2011).

Cytometric genome size estimation provides a simple quality-check to help combat the significant research implications of poor assemblies. Having obtained sufficient material for WGS, flow cytometry requires relatively little extra time and effort – especially in the context of potentially expensive and complex genome sequencing and assembly pipelines – provided there is access to a flow cytometer and associated expertise in its use for fungi. Even in the absence of cultures, genome size estimations of biotrophic basidiomycetes have been obtained using flow cytometry of fungal-infected leaves (Tavares et al., 2014), highlighting that culturing is not always essential for cytometric genome size analysis. But having highlighted the importance of having a robust genome size estimate for genome assembly, why are cytometric estimates desirable?

Previous comparisons of genome sizes estimated from genome assemblies and cytometric approaches in eukaryotes at large have suggested that estimations from cytometric methods are typically, but not always, larger than those from assemblies (Bennett and Leitch, 2005b; Bennett and Leitch, 2011; Elliott and Gregory, 2015). Certainly this is borne out when comparing across all *Ascomycota*, where 762 cytometric estimations taken from the Fungal Genome Size Database (Kullman, Tamm and Kullman, 2005) gives an average genome size almost double that derived from 6,600 assembly-based estimations i.e., ∼63 Mbp/1C versus ∼34 Mbp/1C (Figure 1.2). One explanation for these results could be that WGS has historically been biased towards species with smaller genomes, skewing the average assembly-based genome size towards a lower value, whereas cytometric measurements are not size-dependent and can capture the upper extremes of genome size. Obviously, the most meaningful comparison of cytometric versus assembly-based genome sizes is between estimates for the same species, which are rarely available for both methods. In the few cases where this is possible, there is no consistent pattern. For example, for species such as *Venturia inaequalis* and *V. pyrina* (*Pleosporales*, *Dothideomycetes*), the higher estimates are reported from cytometric methods, while for *Aspergillus flavus*, *A. niger* (*Eurotiales*, *Eurotiomycetes*) and *Paracoccidioides brasiliensis* (*Onygenales*, *Eurotiomycetes*) the estimates are more consistent between methods (Figure 1.5A). We cannot, therefore, assume that genome assembly universally underestimates ascomycete genome size.

Of course, not all genome assemblies are made equal either. Choice of sequencing technology, bioinformatics tools and different computational settings/parameters can result in assemblies which vary significantly in quality (Mavromatis et al., 2012; Desai et al., 2013; Abbas, Malluhi and Balakrishnan, 2014; Khan et al., 2018) and can thus produce differing assembly sizes (Figure 1.5B). Even state of the art scaffolding approaches are sensitive to methodological choices and require validation against, for instance, cytological data (Kadota et al., 2020). Comparing the performance of multiple assembly tools on the same WGS dataset is, therefore, desirable to maximise contiguity and 'completeness'.

1.5 Capitalising on collections

The genome assembly gaps in Figure 1.2 can broadly be grouped into lineages which 1) have been recently discovered; 2) have attracted less study interest; and 3) are difficult to isolate and/or sequence. Addressing gaps from the latter group is mostly reliant on technological and computational advances. For instance, the first attempt at obtaining fungal genomes using single-cell genomics (Ahrendt et al., 2018) and the development of Hi-C methods to obtain genomes from mixed microbial samples

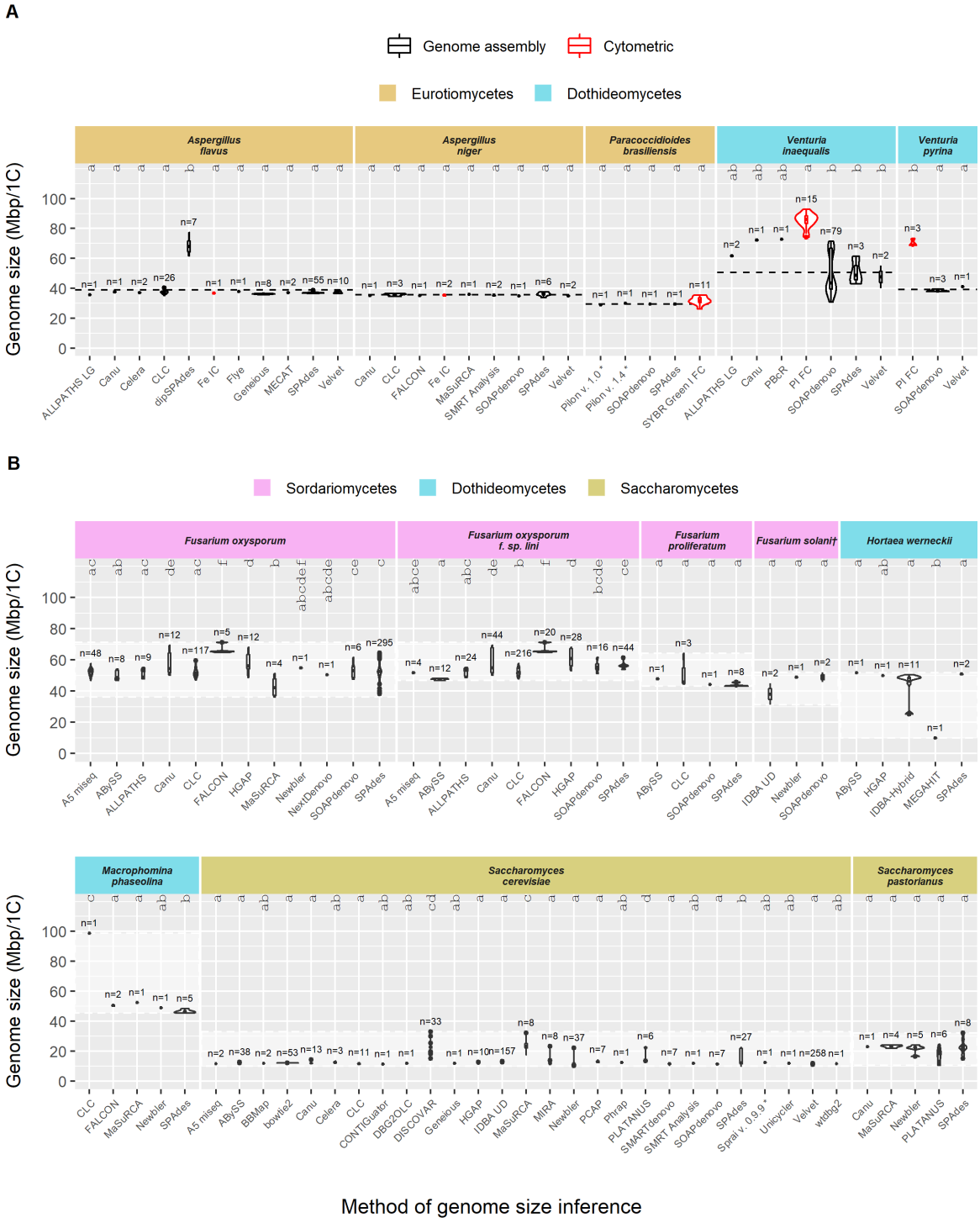


Figure 1.5: Case studies of variability in genome size estimates depending on method of inference for 13 ascomycete species as of January 2021. Sample size is indicated above points and statistically different groups according to TukeyHSD are indicated by letters at the top of the plots. Asterisks (*) beside method names mark methods that are believed to be incorrect in the NCBI genome reports. For details on the cytometric and genome assembly methods shown see Appendix A.1.3. **(A)** Species with both genome assembly– based estimates (black) and cytometric estimates (red). The dashed line indicates the mean for assembly-based estimates. **(B)** Species where the range of assembly-based genome size estimates exceeds 20 Mbp/1C. White translucent bands indicate the total range within which genome size estimates fall. (†) *Fusarium solani* was reassigned to the genus *Neocosmospora* by Sandoval-Denis, Lombard and Crous (2019).

(Burton et al., 2014; Press et al., 2017) show promise that new sequencing approaches and bioinformatics tools will be able to extend WGS widely to unculturable species. The first sequenced genome for the class *Laboulbeniomycetes* was achieved with a single-cell whole genome amplification kit on an individual thallus from the obligate cockroach pathogen *Herpomyces periplanetae* (Haelewaters et al., 2020). For taxa that are not necessarily challenging to sequence but have not been prioritised for sequencing, however, an 'easy win' for gap filling can come from exploiting existing accessions in biological collections.

Collections are an important resource for documenting biodiversity and provide curated specimens spanning time and space (Funk, 2018; Paton et al., 2020). Dedicated mycological collections such as fungaria (containing dried specimens, i.e., the fungal equivalent of plant herbaria) and culture collections act as a record of known species and their distributions, which can be used to address diverse research questions such as historical trends relating to pollution, climate and other environmental factors (Kauserud et al., 2008; Agnan, Séjalon-Delmas and Probst, 2013; Huang, Bowman et al., 2018; Andrew, Diez et al., 2018; Andrew, Büntgen et al., 2019) or even tracking fungal epidemics (Ristaino, 2020; Peck et al., 2021). One major limitation of fungarium specimens is that the age and original method of treatment for preservation can result in degraded DNA that is difficult to successfully sequence (Bainard, Klironomos and Hart, 2010; Andrew, Diez et al., 2018; Smith, Sawbridge et al., 2020; Dal Forno et al., 2022; Miller, Karakehian and Raudabaugh, 2022) – that is, if there is physically enough material to sample from in the first place, which will depend on the taxon, specimen and collector. Nonetheless, rapidly evolving sequencing technologies are making it more feasible to obtain molecular data from fungaria, even in sufficient quantities for WGS (Dentinger et al., 2016). The ongoing Fungal Tree of Life Project at the Royal Botanic Gardens, Kew (`https://www.kew.org/science/our-science/projects/fungal-tree-of-life`) is making use of Kew's fungarium – the biggest in the world with over 1.25 million accessions (Willis, 2018), many of which are type specimens – to generate molecular data for lineages that are missing from existing phylogenetic reconstructions of the *Fungi* at large (e.g., James et al., 2006; Li, Steenwyk et al., 2021). As of October 2022, more than 1,400 specimens have been sampled, almost 150 of which have had sufficient DNA for WGS, which will represent a 25% increase in the number of fungal families with a representative genome assembly (R. Woods et al., unpublished data). Although, due to aforementioned challenges, the genome assemblies produced from fungarium specimens may be less complete than those from fresh material, they can produce more than enough data for filling gaps in the fungal tree of life, for instance by using a subset of genes alongside more complete genome-scale data (e.g., Varga et al., 2019).

For fungi that can be cultured, there are culture collections such as the Westerdijk Fungal Biodiversity Institute (>100,000 strains; Vu et al., 2019), CABI (>28,000 strains; Smith, Ryan and Caine, 2022) and the Fungal Genetics Stock Center (>21,000 accessioned strains; McCluskey, Wiest and Plamann, 2010). These collections preserve living fungal strains, which are more complicated and costly to maintain than fungarium specimens in terms of time, space and human and material resources. However, the recompense for this is that viable cultures provide more opportunities for future research: sampling can be repeated multiple times for different contexts without permanently destroying the original specimen; sufficient material for high quality, long-read WGS can be readily produced; and strains can be used for experimental work. As cryopreserved cultures are both living and frozen in time, they can also be used for 'historical genomics', as demonstrated by Peck et al. (2021) who tracked genetic differences between pathogenic strains across disease outbreaks which

occurred decades ago.

Despite the demonstrable value of fungal culture collections, global repositories only contain a small proportion of the known fungal species – let alone the unknown species – and are highly regionally and taxonomically biased (Paton et al., 2020). Knowing that all plants harbour diverse fungal endophytes, which likely comprise a considerable proportion of the more than 2 million estimated species of yet undiscovered fungi (Petrini, 1991; Arnold, Maynard et al., 2000; Hawksworth and Lücking, 2017), it follows that we should turn to plant collections as secondary resources for fungal material. Dried herbarium specimens have been successfully targeted for sequencing of fungal associates, including powdery mildew pathogens (Bradshaw and Tobin, 2020; Smith, Sawbridge et al., 2020), arbuscular mycorrhizal fungi (Heberling and Burke, 2019) and taxonomically diverse endophytes (Daru et al., 2018). Seed banks are perhaps the more promising avenue for fungal endophytes as, much like culture collections versus fungaria, they have the considerable advantage of being living collections. When seeds are kept viable in cryopreservation, presumably their fungal associates are too, enabling potential isolation of live fungal strains in culture. This is particularly valuable as it facilitates the compilation of endophytic culture collections for further study, meaning seed banks effectively offer a 'two for the price of one' deal on preserving both plants and fungi. And yet, prior to this thesis, the largest seed bank in the world, Kew's Millennium Seed Bank (MSB), had not been explored for fungi, and we are ignorant about the potential impact of fungal endophytes on stored seeds.

1.6 Thesis outline

As highlighted in Section 1.1 above, there are major uncertainties associated with whether endophytism is a stable lifestyle across the *Ascomycota*. This has implications for the safety of using endophytes for agricultural biocontrol, but also more widely for the health of our ecosystems under global change. Understanding the genetic basis and evolutionary histories of plant associated lifestyles is essential to explore these issues and is reliant on genome assembly data for taxonomically and functionally diverse taxa.

In this thesis, I have taken advantage of existing collections to address the genomic data deficit for endophytes, and subsequently explored the pressing question: can we use genomic data to distinguish endophytes and plant pathogens? The thesis is structured around the following objectives:

Objective 1 - Explore Kew's MSB for novel fungal endophyte diversity.

Objective 2 - Determine to what extent we can use genome data to distinguish endophytes and plant pathogens – a case study in the genus *Fusarium*.

Objective 3 - Produce new genomic resources for a broader taxonomic range of fungal endophytes by capitalising on culture collections.

To compare closely related endophytes and phytopathogens, in the first instance we need to isolate endophyte strains. As outlined above, living plant collections such as the MSB represent an excellent potential resource for novel endophyte discovery. In Chapter 2 I have tackled **Objective 1** with a proof-of-concept study to demonstrate for the first time that viable fungal endophytes can be isolated in culture from seeds deposited in the MSB. In doing so, it was also revealed that endophyte community composition, diversity and abundance was significantly different depending on the habitat seeds had been collected from, and these differences also correlated with seed germination/viability.

The results show that there may be knock-on effects for the efficacy of seed banking if we continue to overlook microbial associates of seeds, and therefore that seed collection and storage procedures should also account for the seed microbiome.

Having isolated novel endophytes, in Chapter 3 I have sequenced, assembled and annotated genomes for a subset of strains belonging to the genus *Fusarium*, including a newly described species. Due to the variation in assembly tool performance outlined above in Section 1.4, this included a comparison of multiple tools to optimise assembly quality. I have then used *Fusarium* and closely allied genera as a case study to compare gene repertoires between different lifestyles in the group using an evolutionary genomics approach, as well as exploring patterns of selection and codon optimisation (**Objective 2**). As the gene repertoires of *Fusarium* endophytes and phytopathogens broadly resembled each other – suggesting a shared capacity for both lifestyles in the group – we question the suitability of *Fusarium* species for biocontrol. These results support the current understanding of most *Fusarium* species being prolific generalists.

Reconstructing evolutionary lifestyle histories of endophytes versus phytopathogens more broadly across the *Ascomycota* is currently hampered by a lack of genomic data. In Chapter 4 I have made use of endophyte strains deposited in CABI's culture collection to supplement the existing pool of genomic resources for endophytes (**Objective 3**). As I was able to obtain high molecular weight DNA from around half the strains, this included long-read sequencing to produce highly contiguous hybrid assemblies, once again including a comparison between multiple assembly tools. I have also demonstrated the value of cytometric genome size estimates for assessing assembly quality, as argued in Section 1.4 above. Phylogenetic analyses revealed these to be the first genome assembly for the genus and/or species for 11 of the total 15 strains, emphasising how effective collections can be for filling taxonomic gaps.

Chapter 2

Seed banks as incidental fungi banks: fungal endophyte diversity in stored banana wild relative seeds

Publication details

This chapter has been published as the following paper:

> **Hill, R.**, Llewellyn, T., Downes, E., Oddy, J., MacIntosh, C., Kallow, S., Panis, B., Dickie, J.B. and Gaya, E. (2021). Seed Banks as Incidental Fungi Banks: Fungal Endophyte Diversity in Stored Seeds of Banana Wild Relatives. *Frontiers in Microbiology* 12:643731. DOI: 10.3389/fmicb.2021.643731.

R.H. and E.G. designed the study, implemented the analysis and wrote the paper. The paper used molecular data collected by **R.H.**, T.L., E.D., J.O. and C.M. prior to the start of this PhD. S.K. and J.D. provided the samples and contributed to the writing of the paper, and S.K. performed the tetrazolium chloride testing. B.P. performed the embryo rescue testing.

2.1 Abstract

Seed banks were first established to conserve crop genetic diversity, but seed banking has more recently been extended to wild plants, particularly crop wild relatives (CWRs) (e.g., by the Millennium Seed Bank (MSB), Royal Botanic Gardens Kew). CWRs have been recognised as potential reservoirs of beneficial traits for our domesticated crops, and with mounting evidence on the importance of the microbiome to organismal health, it follows that the microbial communities of wild relatives could also be a valuable resource for crop resilience to environmental and pathogenic threats. Endophytic fungi reside asymptomatically inside all plant tissues and have been found to confer advantages to their plant host. Preserving the natural microbial diversity of plants could therefore represent an important secondary conservation role of seed banks. At the same time, species that are reported as endophytes may also be latent pathogens. We explored the potential of the MSB as an incidental fungal endophyte bank by assessing diversity of fungi inside stored seeds. Using banana CWRs in

the genus *Musa* as a case-study, we used a similarity and phylogenetics approach for classification of endophyte operational taxonomic units (OTUs) from an extended internal transcribed spacer (ITS)–nuclear ribosomal large subunit (LSU) fragment. Fungi were detected inside just under one third of the seeds, with a few genera accounting for most of the OTUs – primarily *Lasiodiplodia*, *Fusarium* and *Aspergillus* – while a large variety of rare OTUs from across the *Ascomycota* were isolated only once. *Fusarium* species were notably abundant – of significance in light of Fusarium wilt, a disease threatening global banana crops – and so we additionally sequenced the translation elongation factor 1 alpha (EF1α) marker in order to delimit species and place them in a phylogeny of the genus. Endophyte community composition, diversity and abundance was significantly different across habitats, and we explored the relationship between community differences and seed germination/viability. Our results show that there is a previously neglected invisible fungal dimension to seed banking that could well have implications for the seed collection and storage procedures, and that collections such as the MSB are indeed a novel source of potentially useful fungal strains.

2.2 Introduction

Seed banks were initially conceived in the 20$^{\text{th}}$ century as a measure to conserve crop genetic diversity (Peres, 2016), the most famous example likely being the Svalbard Global Seed Vault (Westengen, Jeppson and Guarino, 2013). The MSB, managed by the Royal Botanic Gardens Kew, is the world's largest seed bank and part of a global partnership network for seed conservation (Liu, Cossu et al., 2020). The MSB is notably directed to wild plant conservation, with one of its priorities being CWRs (Liu, Breman et al., 2018). CWRs, the close relatives of our domesticated crop species, act as an additional pool of genetic diversity to breed improvements into our crops, such as increased productivity and resilience against disease and environmental stressors (Hajjar and Hodgkin, 2007; Brozynska, Furtado and Henry, 2016). More recently, similar benefits have been equally demonstrated by inoculation of various crops with endophytes from CWRs (Murphy, Jadwiszczak et al., 2018; Murphy, Hodkinson and Doohan, 2018; Murphy, Doohan and Hodkinson, 2019). This potential role of CWR endophytes in both the health of wild plant populations and their crop counterparts brings in additional value to the MSB collections, making them not only important for plant conservation, but also plant microbiome conservation.

Considering the range of ecological roles exhibited by fungi in the endophytic lifestyle (as outlined in Chapter 1.1), there is uncertainty as to which endophytes inhabiting stored seeds are beneficial – or even essential – to the plant host, and which are potentially harmful. This uncertainty has obvious implications in seed storage protocols, which most often focus on the harmful fungi. For example, in internationally recognised reports on best-practise gene banking, mention of fungi (and bacteria) is almost exclusively in the context of avoidance, with recommendations for the use of antifungals/antibiotics on collections (FAO, 2014; Center for Plant Conservation, 2018). These recommendations overlook an essential question: what are the impacts on seed banking if we fail to preserve healthy endophyte communities? How do endophyte communities impact the success of recovered plant populations down the line? Such endophytic communities may be playing similar roles as the microbial associates of humans or animals, which we now know to be essential for normal, healthy functioning and imbalances of which cause disease (Dudek-Wicher, Junka and Bartoszewicz, 2018). While great care is taken to optimise the phylogenetic and geographical diversity and longevity of MSB seed collections, consideration of the microbial communities associated with the seeds is notably absent (Liu, Cossu et al., 2020). Considering that there are endophytes known to be implic-

ated in germination and seedling success (Tamura et al., 2008; Hubbard, Germida and Vujanovic, 2014; Li, Song et al., 2017; Shearin et al., 2018; Leroy et al., 2019), this is a significant oversight.

To explore these issues and demonstrate the value of seed banks for endophyte discovery, we focused on a case study of CWRs of banana (and plantain, *Musa* spp. L.), one of the most important crops in the world. Global production of banana is estimated to be 116 million tonnes annually, worth US$31 billion (FAO, 2020b). *Musa* taxa are tall herbaceous monocarpic monocotyledons in the family *Musaceae*, order *Zingiberales*. They are native to tropical and subtropical Asia to western Pacific regions (Govaerts and Häkkinen, 2006) with approximately 80 taxa (hereon called 'species') in the genus (Häkkinen and Väre, 2008; POWO, 2019). There are around 1,000 cultivars of edible bananas (Ruas et al., 2017; FAO, 2020a), most of which stem from two species: *Musa acuminata* Colla and *M. balbisiana* Colla (Carreel et al., 2002; Langhe et al., 2009; Perrier et al., 2011; Rouard et al., 2018; Martin, Cardi et al., 2020). In spite of this diversity, the vast majority of commercial banana plantations are clones of a single cultivar, Cavendish, which makes the crop highly susceptible to disease (Ordonez et al., 2015). In the 1970s, *Fusarium oxysporum* f. sp. *cubense* emerged to cause Fusarium Wilt of Cavendish bananas, and the predominant strain (Foc TR4) has since spread across the global tropics to most banana producing countries (Dita et al., 2018; `https://www.promusa.org`). Considering the global value of the banana crop, 85% of which is eaten locally as a major contribution to people's diets (FAO, 2020b), Foc TR4 represents a major threat to both economic and food security in banana producing countries. Stored banana CWR seeds are a precious conservation resource in light of the susceptible Cavendish banana cultivar, and so present a valuable case-study for investigating associated endophyte diversity.

While many endophytic species can be grown in culture, many more cannot, and so molecular tools are relied upon to detect much more of the true extent of endophytic diversity (e.g., Higgins et al., 2011; Parmar et al., 2018; U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019). Nonetheless, culturing is still a necessary tool, as it not only isolates strains for future study, but also provides an indication of which fungal strains are alive, which is particularly relevant when assessing post-storage endophytes. Here we used both a culture-dependent and culture-independent approach to maximise discovery of endophytic diversity from accessions belonging to six species of banana wild relatives in the genus *Musa*. By PCR-cloning individual seed DNA extractions for the culture-independent approach, we were able to assess the number of unique OTUs – a proxy for species – per seed. We made use of metadata and seed viability assessments from the MSB collections in order to explore the association of habitat, host *Musa* species, post-storage seed viability and germination rate with endophyte community composition, diversity and abundance.

2.3 Materials and methods

Isolation of strains and molecular work was completed prior to this PhD and is described in detail in the corresponding paper by Hill, Llewellyn et al. (2021). In brief, seeds from 45 *Musa* accessions (with each accession containing 50 seeds collected from between one and five plant individuals belonging to the same *Musa* species in the sampling site) were obtained from the MSB, all of which had been stored at -20°C (Supplementary Table S2.1). Seeds were surface sterilised and DNA was extracted from both axenic cultures grown from the seeds and directly from crushed seeds. An extended ITS–partial LSU fragment was amplified for PCR cloning and Sanger sequencing; sequences from 642 endophytes (235 cultures, 280 direct sequences and 127 clones) were deposited in GenBank under

accession numbers MW298868-MW299510. Additionally, EF1α was amplified and sequenced for *Fusarium* taxa (GenBank accessions MW319587–MW319636).

Seed viability assessment

Post-storage seed viability was assessed using two methods. Firstly, the tetrazolium chloride test was carried out following the approach of Leist and Krämer (2011). Seeds were imbibed on agar for 3 days at 20°C before a proportion of the testa was removed using a scalpel on two lateral sides to expose the endosperm. Seeds were then soaked in 1% buffered 2,3,5-triphenyl tetrazolium chloride (pH 6–8) for 2 days at 30°C in the dark. Staining patterns were recorded – embryos that completely stained dark red, or that showed dark red staining at the embryonic axis (the opposite from the haustorium) were considered viable, while light pink staining or white embryos were considered unviable. Fifty seeds per accession were tested.

The second viability test was embryo rescue. In a laminar flow, seeds were sterilised by soaking them in 96% ethanol for 3 min, followed by 20% bleach (NaOCl containing 1 drop of detergent per 100 ml) for 20 min, then seeds were rinsed three times in sterilised water. Continuing in the laminar flow with sterile forceps and scalpel, embryos were extracted from seeds. This was done using an incision in the seed coat next to the micropyle and manipulating the seed in order to split the testa; the embryo was then gently removed. Embryos were subsequently transferred onto autoclaved half MS medium (Murashige and Skoog, 1962) in tubes using long forceps with the haustorium in contact with the medium and the embryonic axis upward. Tubes containing embryos were incubated in the dark at 27°C for 14 days after which they were put in a growth chamber in the light at 27°C for an additional 14 days. Six possible observations were recorded: shoot, callus, blackened colouration, no embryo, contamination, no change. Ten seeds per accession were tested.

OTU delimitation and taxonomic identification

Sequences were manually edited with contiguous alignments using Geneious R7 v7.1.5 (Biomatters, New Zealand). Sequences were clustered into OTUs using the *de novo* method USEARCH v10.0.240 as part of the UPARSE pipeline (Edgar, 2013). As USEARCH is sensitive to fragments of different length, ITSx (Bengtsson-Palme et al., 2013) was used prior to clustering to extract the 5.8S and ITS2 regions – shown to recover more fungal OTUs when used together (Heeger et al., 2019) – while LSU fragments were manually trimmed to the same length after alignment with MUSCLE v3.8.425 (Edgar, 2004) and visualisation in AliView v1.17 (Larsson, 2014). Dereplication was performed via removal of identical sequences using the fastx_uniques functions inbuilt to USEARCH. 5.8S–partial LSU OTUs were clustered using a 99% similarity threshold, guided by the optimal threshold for species discrimination using ITS/LSU identified by Vu et al. (2019). Singletons – OTUs comprising one sequence – were not discarded, as is common practise to reduce artefacts when using next generation sequencing datasets, because each sequence originated from Sanger sequencing of an individual seed extraction, and so was assumed to be 'real'.

Preliminary identification of OTUs was made via a local BLASTn v2.6.0 search (Camacho et al., 2009) against the UNITE v8.2 database, release 04.04.2020 (Abarenkov, Zirk et al., 2020). Taxonomic identification of OTUs was inferred from the top UNITE hit, guided by Vu et al. (2019): ≥99% similarity for the same species; ≥98% similarity for the same genus; ≥96% similarity for the same family; ≥94% similarity for the same order; ≥92% similarity for the same class; and <92%

similarity for the same phylum. Similarity-based identification was corroborated with a phylogenetic approach via the Tree-Based Alignment Selector toolkit (T-BAS) v2.2 (Miller, Pfeiffer and Schwartz, 2010; Carbone, White, Miadlikowska, Arnold, Miller, Magain et al., 2019), a platform designed for preliminary placement and visualisation of unknown fungal sequences in curated multi-locus phylogenies. Representative sequences for 181 OTUs were placed in the 6-loci Pezizomycotina v2.1 and the 6-loci Fungi reference trees (James et al., 2006; Carbone, White, Miadlikowska, Arnold, Miller, Kauff et al., 2017) with default settings and using the evolutionary placement algorithm option from RAxML (Berger and Stamatakis, 2011; Stamatakis, 2014). OTU taxon assignment was altered to reflect the lowest taxonomic level in agreement between both T-BAS and UNITE, with the UNITE species level identification used if T-BAS and UNITE agreed on genus and the UNITE percentage identity was ≥99%. All filtering of classification data was done using R v3.5.3 in RStudio v1.1.463 (RStudio Team, 2015; R Core Team, 2020), the script for which is available at `https://github.com/Rowena-h/MusaEndophytes`.

Sampling effort and community analysis

For the purpose of these analyses, *Musa* subspecies and varieties were grouped under the same species. Sampling effort was assessed by producing species accumulation curves of the number of OTUs for the number of *Musa* accessions using the rarefaction method in the specaccum function from the R package vegan v2.5-6 (Oksanen et al., 2019). This was done including and excluding singleton OTUs for all *Musa* accessions (n=45) as well as distinguishing between the three best sampled species – *M. acuminata* (n=12), *M. balbisiana* (n=16) and *M. itinerans* (n=14). The impact of detection method – culturing, direct sequencing or cloning – on species recovery was quantified with analysis of similarity (ANOSIM) (Clarke, 1993) using the vegan anosim function following confirmation that data dispersion was even using the vegan betadisper function.

The RBG, Kew and oil palm plantation accessions (1 locality in Malaysia) were excluded from the following analyses due to low sample size for the habitats and the former being a geographical outlier. Endophyte community composition was explored using non-metric multidimensional scaling (NMDS) implemented in the metaMDS function in vegan. OTU counts were filtered for the eight most common OTUs (abundance greater than 20) for the 33 accessions of *M. acuminata*, *M. balbisiana* and *M. itinerans* and six dimensions were selected for the NMDS using a scree plot (Supplementary Figure S2.1). Habitat information for *Musa* accessions was interpreted from the collection notes in the MSB's metadata records (Supplementary Table S2.1). To investigate the relationships between community composition and post-storage seed viability (i.e., what proportion of seeds from the accession contained a live embryo in the tetrazolium chloride testing) and post-storage germination rate (i.e., what proportion of embryos from the individual germinated in the embryo rescue testing), test results for each *Musa* accession were fitted to the NMDS ordination using the vegan ordisurf function, which uses generalised additive models to fit a smooth response surface and is therefore appropriate for a non-linear relationship between the ordination and variable.

The impact of habitat and *Musa* species on the variation in community composition – both for the subset of common taxa visualised in the NMDS and for all OTUs including rare taxa – was tested with permutational multivariate analysis of variance (PERMANOVA) implemented in the vegan adonis and adonis2 functions using Bray-Curtis dissimilarity and 999 permutations. PERMANOVA with adonis considers variables sequentially, meaning that the test is performed on the first variable provided and the residual unexplained variance is left to be explained by the next variable, and so

on. As variables can be correlated with each other, the order in which variables are added to the adonis formula impacts the results. In order to determine the unique impact of variables irrespective of order, i.e.,, marginal effect size (marginal R^2), we used the adonis2 function with the by='margin' option, which reports the variance that is not explained by any of the other variables. The variables were then tested with adonis in order of decreasing marginal effect size to assess the total effect size (R^2). The vegan betadisper function was also used for permutational analysis of multivariate dispersions (PERMDISP) to assess whether data dispersion was uniform for each variable, as when sample sizes are unbalanced varying data dispersion can result in a significant PERMANOVA test even if group composition is not significantly different (Anderson and Walsh, 2013). The PERMDISP null hypothesis is that there is no difference in dispersion between groups, and so a significant p value indicates that dispersion is not consistent.

In order to determine which of the habitats had significantly different community composition from the others, pairwise PERMANOVA was performed on both the subset of common taxa used in the NMDS as well as all OTUs including rare taxa. This was done using the pairwise.perm.manova function from the R package RVAideMemoire v0.9-78 (Hervé, 2020) with 999 permutations and multiple testing p value correction using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Difference in diversity – according to Shannon and Simpson diversity indices, both calculated with the vegan diversity function – and abundance of fungi per *Musa* accession for each habitat was assessed using the TukeyHSD function. All results were plotted in R with the ggplot2 v3.3.0 package (Wickham, 2016). Ellipses for each habitat in the NMDS plot were generated with the stat_ellipse function in ggplot2.

Fusarium phylogenetic analysis

Given the abundance of *Fusarium* in our dataset, a genus-specific phylogeny was reconstructed to elucidate the relationships of our *Fusarium* OTUs with already known species. While the UNITE identification described above recovered many 5.8S–partial LSU OTUs to apparent species level, it has been shown that the ITS locus is not sufficiently variable for species delimitation within this particular genus (Geiser, Jiménez-Gasco et al., 2004). For this reason, OTUs based on EF1α sequences were also delimited as above for use in the phylogenetic analyses.

Representative sequences for each OTU from this study (as provided by USEARCH) were aligned with already published EF1α data and, in addition, RNA polymerase II largest subunit (RPB1) and RNA polymerase II second largest subunit (RPB2) sequences were also taken from the MycoBank website (`https://fusarium.mycobank.org/`). Taxon sampling was guided by O'Donnell, Rooney et al. (2013), with the addition of taxa from the *Fusarium oxysporum* species complex (FOSC) (Maryani et al., 2019) and *Fusarium musae* (Van Hove et al., 2011) and *Neonectria coccinea* and *Cylindrocarpon cylindroides* were selected as the outgroup (Supplementary Table S2.2). Sequences for each gene were aligned using MUSCLE v3.8.425 (Edgar, 2004) and ambiguous regions were manually delimited and removed in AliView v1.17 (Larsson, 2014). Much of the variability in EF1α that makes it a valuable marker for *Fusarium* is located across three introns (Geiser, Jiménez-Gasco et al., 2004), so introns were isolated from protein-coding regions and Gblocks v0.91b (Castresana, 2000) was used to select adequately aligned intron sites, with the 'Allow gap positions' option to prevent loss of highly variable sites. To check for topological incongruence between genes, a maximum likelihood (ML) search was performed on individual alignments – partitioned by introns and codon position for protein-coding regions – using the GTRGAMMA substitution model with bootstrapping

over 1,000 replicates in RAxML v8.2.9 (Stamatakis, 2014). Conflicts between gene trees (defined as ≥70% bootstrap support for contradictory relationships) were manually identified for each of the three pairwise comparisons with help from the compat.py script (Kauff and Lutzoni, 2002; Kauff and Lutzoni, 2003) run in Python v3.7.9 using Biopython v1.78 (Cock et al., 2009). Taxa responsible for conflicts were removed. The three loci were concatenated using SequenceMatrix (Vaidya, Lohman and Meier, 2011) and partitioned by gene, codon position and EF1α introns for the ML search, performed as above for individual gene trees – see `https://github.com/Rowena-h/MusaEndophytes` for the raw alignment and tree files. Species names were checked in Species Fungorum (`http://www.speciesfungorum.org/`) and the species tree was plotted in R using ggtree v2.3.4 (Yu et al., 2017).

2.4 Results

Most endophyte-colonised *Musa* seeds contained a single OTU

ITS–partial LSU sequences of fungal endophytes were obtained from 533 *Musa* seeds, 31% of the total 1,710 seeds used in this study (+90 control seeds). One fungal isolate per seed was most commonly found, however up to 7 unique OTUs were detected via cloning in a small number of seeds (Figure 2.1A). Of the most sampled *Musa* species, *M. acuminata* had the lowest number of fungal isolates relative to total seeds while *M. itinerans* had the highest. No fungi were detected in *M. gracilis*, however only one accession was sampled, which was also the case for *M. violascens* and *M. velutina*.

Lasiodiplodia, *Fusarium* and *Aspergillus* were the most common genera

Not including duplicate clones, 642 sequences were clustered into 181 OTUs, of which 125 (69%) were singletons. Species accumulation curves including singleton OTUs were almost linear and with a high gradient, while the curves excluding singletons approached an asymptote, indicating that many rarer OTUs remain to be discovered but a considerable proportion of the most common OTUs were captured (Figure 2.1B).

Of the 181 OTUs, UNITE and T-BAS classified the vast majority to the *Ascomycota* (162, 89.5%), with a few belonging to the *Basidiomycota* (12, 6.6%) and the remaining as unclassified *Fungi* (7, 3.9%). In almost equal proportion, most of the ascomycete OTUs fell in the classes *Dothideomycetes*, *Eurotiomycetes* and *Sordariomycetes* (in order of abundance), in the respective orders of *Botryosphaeriales*, *Eurotiales* and *Hypocreales* (Figure 2.2). The three most common genera were *Lasiodiplodia*, *Fusarium* and *Aspergillus* (with 161, 123 and 117 occurrences, respectively), which together accounted for almost two thirds of the total number of sequenced endophytes. The most abundant OTUs were recovered from all sampling approaches – culture-dependent and culture-independent (with additional cloning) – however each approach detected rare OTUs not found by the others (Figure 2.3). Data dispersion was even across methods (betadisper p=0.33) and ANOSIM indicated that, while communities were significantly different according to different detection methods (p=0.001), the strength of these differences between methods was relatively low (R=0.08). 10 OTUs from inside the seeds were also isolated pre-sterilisation on the outside of seeds (Supplementary Table S2.3), but as all the surface sterilisation imprint controls showed no fungal growth, we were confident that these OTUs were not contaminants.

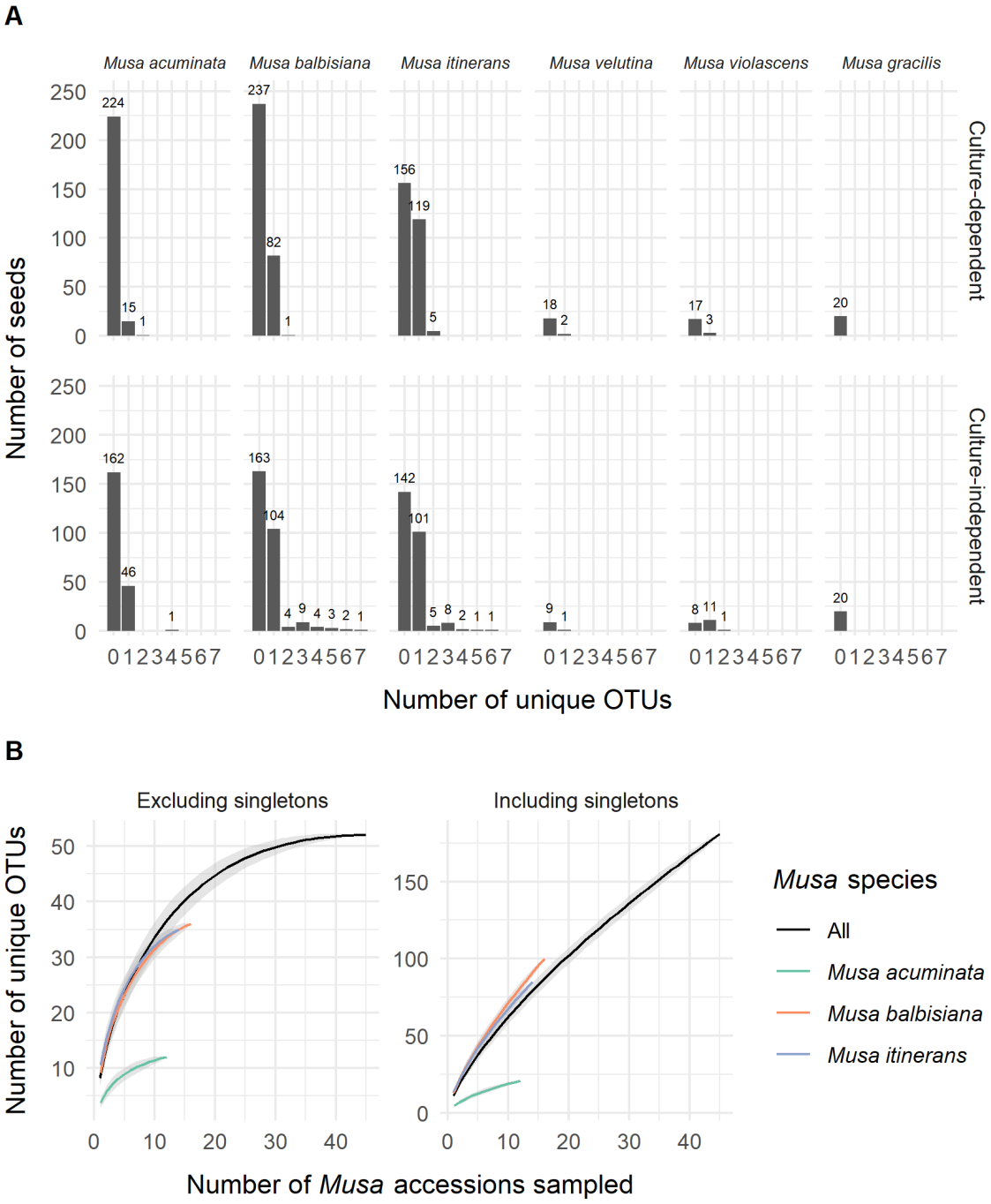


Figure 2.1: **(A)** The number of unique OTUs per seed for each species of *Musa* from both the culture-dependent and independent approaches. **(B)** Species accumulation curves of OTUs by number of *Musa* accessions sampled, both excluding and including singletons and showing distinction between the three most sampled *Musa* species. Standard error is shaded grey around the lines.

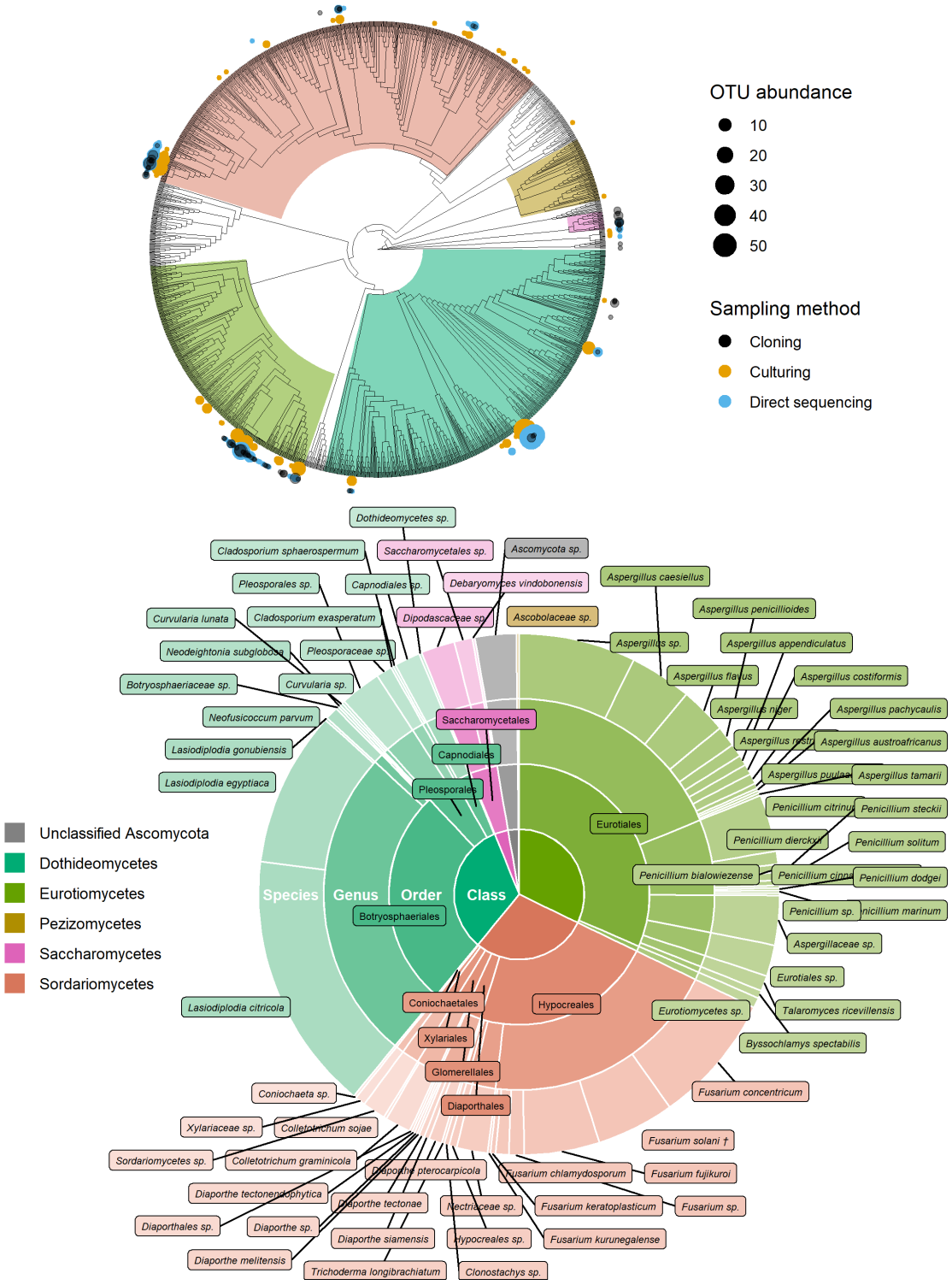


Figure 2.2: Identification of ascomycete OTUs according to UNITE and phylogenetic placement in the T-BAS *Pezizomycotina* v2.1 tree. OTUs from this study are indicated on the T-BAS tree by circles on tips (top) with size proportional to number of times the OTU was detected and colour showing sampling method. Taxon classification as agreed by UNITE and T-BAS is summarised in a pie chart (bottom). (†) *Fusarium solani* = *Neocosmospora solani* (Sandoval-Denis, Lombard and Crous, 2019).

Figure 2.3: Euler diagram showing the OTUs recovered by each sampling approach. The size of labels is proportional to the number of occurrences for that OTU. Numbers under method labels and in intersections indicate the total number of OTUs for the corresponding approach(es). The ANOSIM result in the top right indicates the statistical significance of the different approaches.

Endophyte community composition, diversity and abundance changed with habitat

There was a significant difference in endophyte communities across habitats when considering the most common OTUs (pooled from all detection methods) from *M. acuminata*, *M. balbisiana* and *M. itinerans* accessions (adonis2 marginal R^2=0.32, p=0.001; adonis R^2=0.34, p=0.001) and also when including rare taxa, although with a smaller effect size (adonis2 marginal R^2=0.18, p=0.001; adonis R^2=0.21, p=0.001). *Musa* species was not found to be a significant factor for variance of taxa (Table 2.1). PERMDISP found data dispersion of common taxa to be similar across *Musa* species but not across habitats: dispersion was greatest in the habitat with the smallest sample size (roadside), suggesting a liberal PERMANOVA bias (Supplementary Figure S2.2) (Anderson and Walsh, 2013). However, PERMANOVA, PERMDISP and NMDS together suggested that habitat was associated with both location and dispersion of the data. The NMDS visualisation showed that the ellipses for the jungle buffer, jungle edge and roadside habitats overlapped, but with data dispersion increasing with level of habitat disruption: from jungle buffer (least disrupted, most tightly clustered) to roadside (most disrupted, least tightly clustered) (Figure 2.4A). The pairwise PERMANOVA analysis confirmed that these three habitats were not significantly different to each other in community composition for the common taxa visualised in the NMDS, while they were all significantly different from the ravine habitat (Figure 2.4B), which formed a separate cluster in the NMDS (Figure 2.4A). When including rare OTUs in the pairwise PERMANOVA, however, community composition was also significantly different between jungle buffer and roadside habitats (Figure 2.4B). Both diversity and abundance of endophytes per accession showed the same trend across habitats, with greatest diversity and abundance in the ravine habitat and least in the roadside habitat, with TukeyHSD identifying three statistically distinct groups for both Shannon diversity and abundance, although Simpson diversity was not statistically significant between habitats (Figures 2.4C,D). Oil palm plantation accessions and the RBG, Kew accession were excluded from the main analyses due to low sample size (and as the latter was a geographical outlier), but endophyte abundance was comparatively low for both habitats (Supplementary Figure S2.3).

Fitting post-storage seed viability to the NMDS ordination (ordisurf adjusted R^2=0.46, p=7.46e-05) showed seed viability to have a non-linear relationship with the community structure, with accessions in the ravine habitat cluster and *Penicillium* and *Aspergillus* OTUs associated with lower viability measures and accessions in the jungle buffer habitat associated with higher viability measures (Figure 2.4A). Germination rate showed a similar relationship (ordisurf adjusted R^2=0.31, p=0.006).

Table 2.1: Results of the statistical tests on Bray-Curtis dissimilarity matrices of both the subset of common OTUs visualised in the NMDS and all taxa including rare OTUs. Significant p values are highlighted in bold.

		PERMANOVA				PERMDISP
		adonis2		adonis		Betadisper
Dataset	Variable	Marginal R^2	p	R^2	p	p
Common taxa	Habitat	0.32	**0.001**	0.34	**0.001**	0.0310
	Musa species	0.07	0.055	0.07	0.055	0.0987
All taxa	Habitat	0.18	**0.001**	0.21	**0.001**	**0.0059**
	Musa species	0.08	0.101	0.08	0.101	**1.25E-07**

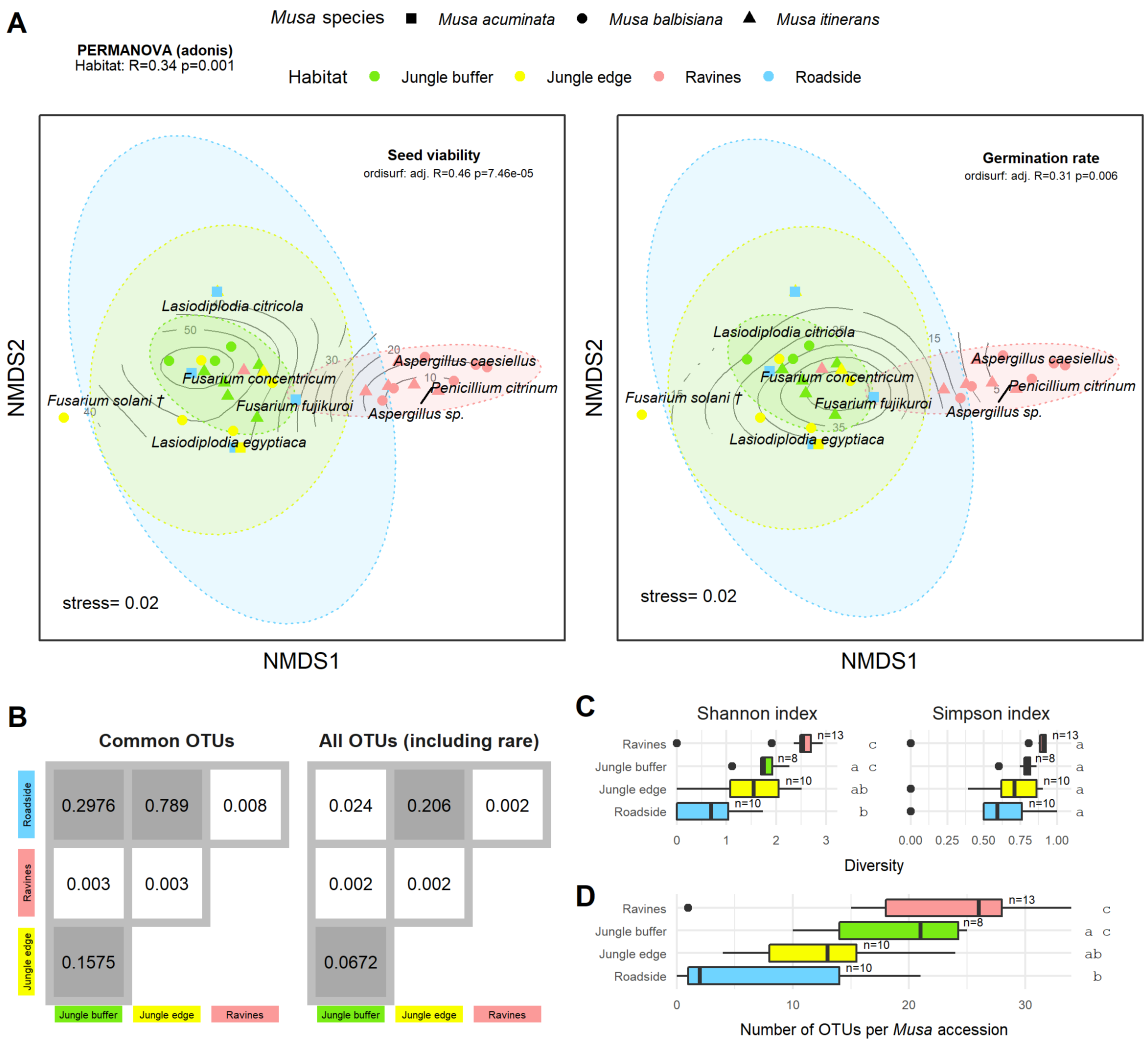


Figure 2.4: **(A)** NMDS plot of the most common OTUs, produced with metaMDS, fitted with post-storage seed viability (left) and germination rates (right). Contour lines indicate the fit of the seed viability and germination rate variables to the ordination using the ordisurf function, showing which points are associated with higher or lower seed viability. Each point represents one *Musa* accession, with shape showing host *Musa* species and colour showing habitat, while OTUs are shown in italic text. (†) *Fusarium solani* = *Neocosmospora solani* (Sandoval-Denis, Lombard and Crous, 2019). Ellipses were generated with the stat_ellipse function in ggplot2. The PERMANOVA result in the top left indicates significant difference in endophyte community composition between habitats. **(B)** Matrix of pairwise PERMANOVA p values showing whether endophyte community was significantly different between pairs of habitats, both for the subset of common OTUs visualised in the NMDS and including rare OTUs. Grey boxes indicate non-significant p values (>0.5). Diversity according to Shannon and Simpson indices **(C)** and abundance of OTUs **(D)** per *Musa* accession in each habitat. Groups with significant difference of means as calculated by TukeyHSD are shown by letters on the right of the plots. Sample size (number of accessions) is shown to the right of boxes.

Fusarium strains were phylogenetically resolved to the *Fusarium fujikuroi*, *'Fusarium' solani* and *Fusarium incarnatum-equiseti* species complexes

Additional clustering of the *Fusarium* taxa produced 10 EF1α OTUs. Phylogenetic analysis resolved these in the *incarnatum* clade of the *Fusarium incarnatum-equiseti* species complex (FIESC), in the *Fusarium solani* species complex (FSSC) – which has recently been reassigned to the genus *Neocosmospora* (Sandoval-Denis, Lombard and Crous, 2019) – and in the *Fusarium fujikuroi* species complex (FFSC), with most OTUs placed within the latter (Figure 2.5). Disregarding the naming of taxa, our phylogeny was in general agreement with the most comprehensive phylogenies of the genus (O'Donnell, Rooney et al., 2013; O'Donnell, Al-Hatmi et al., 2020), with the exception of not recovering geographically grouped clades (Asian, African and American) in the FFSC, which was also one of the only species complexes that was not significantly supported. Across the whole phylogeny, 68% of all internodes were significantly supported. Extremely short branch and internode lengths indicated rapid divergence in the FFSC and FIESC clades, as well as in other species complexes such as FOSC and *Fusarium redolens* species complex.

2.5 Discussion

In this study, we used both a culture-dependent and culture-independent approach to assess the diversity of endophytes in stored banana CWR seeds. In an example of the value of collections, we demonstrated the feasibility of endophyte discovery from seed banks, many strains of which can be isolated in culture for future study. By using cloning versus next generation sequencing methods for the culture-independent detection of seed endophytes, we were able to economically sequence individual seeds (rather than a pooled sample) to determine the endophyte capacity of the *Musa* seeds, which could then be combined with the data on number of endophytes isolated in culture per seed. Of the seeds containing endophytes, the number of unique OTUs was biased toward one for both sampling approaches (Figure 2.1A), which suggests that there is some level of competitive exclusion in the limited physical space of the seed, as posited by Raghavendra et al. (2013). This is also in agreement with recent work on seeds from various alpine plants, which showed that, while bacterial endophytes appear to interact positively, fungi are usually mutually exclusive (Wassermann et al., 2019). Similarly, during pathogenic invasion of radish seeds, it was found that a fungus altered the fungal endophyte community while a bacterium had no effect on either bacterial or fungal endophytes, although the authors noted that the different infection routes and thus microhabitats of the two pathogens could have contributed to the observed community differences (Berihuete-Azorín et al., 2018).

Our seeds were all pre-dispersal (as all MSB seeds are), so there is also the possibility that the endophyte capacity was influenced by the lack of opportunity for seeds to acquire fungi from the soil, which is known to be a source of seed endophyte diversity (e.g., U'Ren, Dalling et al., 2009; Sarmiento et al., 2017). More insight into the dynamics of endophyte seed colonisation is needed, and would benefit from experimental inoculation combined with *in situ* visualisation of the physical space endophytes inhabit within the seed (e.g., Rath et al., 2014; Vági et al., 2014). Previous work on the specific localisation of seed endophytes has established that it varies depending on the species in question: some endophytic species are known to only be found in the seed coat (Oldrup et al., 2010) while others such as grass symbionts are found in the embryo and endosperm (Philipson and Christey, 1986; Zhang, Card et al., 2017). Although in this study we did not establish the exact localisation

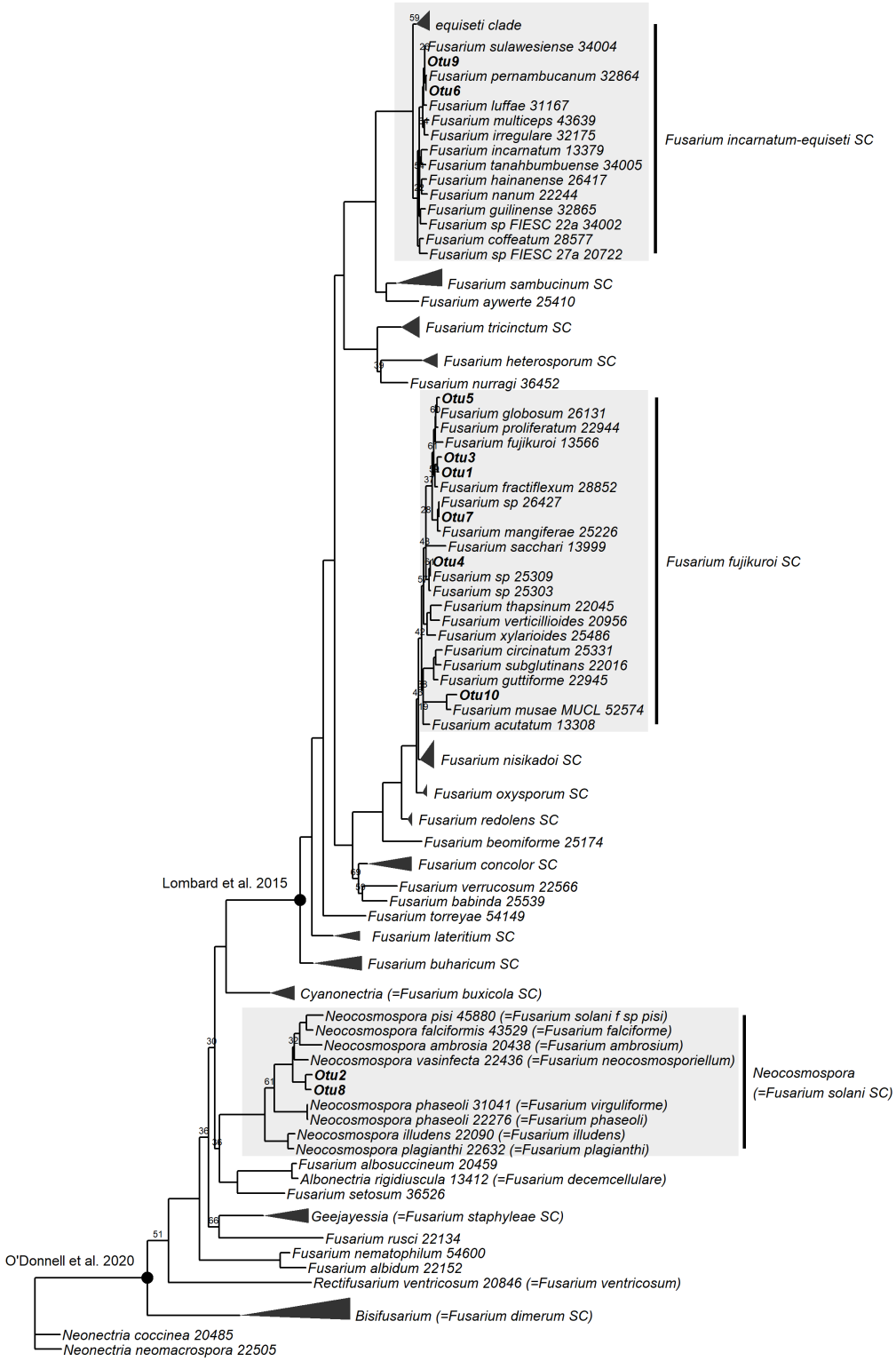


Figure 2.5: ML RAxML tree reconstructing relationships of 130 taxa of *Fusarium* and closely related genera, including the EF1α OTUs delimited in this study (indicated in bold). Bootstrap support values <70 are shown on internodes. Genera and *Fusarium* species complexes which weren't represented by any OTUs in this study are collapsed where possible – triangles are vertically scaled for ease of visualisation, with horizontal length representing the longest branch in the species complex. Circles on nodes indicate generic limits of *Fusarium* proposed by Lombard et al. (2015) and O'Donnell, Al-Hatmi et al. (2020).

of endophytes within the *Musa* seeds, the fact that we both cultured and directly sequenced many fungi from whole seeds whereas embryo rescue testing showed no or minimal 'contamination' (i.e., any fungal growth from extracted embryos; Supplementary Table S2.1) suggests that most of the OTUs may have been located outside the embryo. However, as the embryo rescue testing only applies to culturable fungi and the embryo may contain endophytes that can only be detected through direct sequencing (Figure 2.3), with this data we cannot conclusively comment on the localisation of our taxa within the seeds. We also checked for OTUs present on the seed surface (Supplementary Table S2.3) – we were confident that these OTUs were also found as endophytes inside the seeds and not contaminations as we performed culture imprint controls to confirm the efficacy of the surface sterilisation method. Being both inside and outside the seed indicates that these strains were more likely generalists, horizontally transferred, for instance, from fruit to seed, rather than vertically transmitted endophytes, which would not be expected to be found outside the seed as well.

The genera found in the *Musa* seeds were largely similar to previous studies of *Musa* endophytes from roots and leaves (Sikora et al., 2008; Wang, Min et al., 2014; Zakaria, Izham et al., 2016; Zakaria and Aziz, 2018), as well as other tropical tree endophytes, such as from cacao branches (Rubini et al., 2005), rubber leaves (Vaz et al., 2018) and tropical orchid roots (Bayman and Otero, 2006). The most commonly found genera, *Lasiodiplodia*, *Fusarium* and *Aspergillus*, are all ubiquitous in both endophytic and other contexts. The genus *Lasiodiplodia* is best known for the species *Lasiodiplodia theobromae*, a prevalent endophyte in the global tropics (Salvatore and Andolfi, 2020), but also an infamous pathogen of tropical fruit trees. For instance, *L. theobromae* has been found to cause crown rot in commercial banana (Sangeetha, Anandan and Rani, 2012) and – among other *Lasiodiplodia* strains – stem and fruit rot in papaya (Netto et al., 2014) and dieback in mango (Rodríguez-Gálvez et al., 2017). Goos, Cox and Stotzky (1961) similarly found *L. theobromae* (using the synonym *Botryodiplodia theobromae*) to be pervasive in seeds of *Musa* spp., although they did not report whether the colonised seeds or resulting plants had disease symptoms. They also found *L. theobromae* exclusively in the seed coat and micropylar plug versus the endosperm or embryo and echoed our above hypothesis that it is transferred from fruit to seed. This was also supported in *Musa ornata*, for which *Lasiodiplodia* colonisation was observed in all cases apart from those where embryos were removed from seeds under aseptic conditions (Burgos-Hernández et al., 2014). The two prevalent *Lasiodiplodia* OTUs in this study were classified as *Lasiodiplodia citricola* and *Lasiodiplodia egyptica*, both of which were first described from diseased plants: *Citrus* spp. showing 'branch dieback, cankers and fruit rot' (Abdollahzadeh et al., 2010) and mango suffering dieback (Ismail et al., 2012), respectively. *L. egyptica* has also been implicated in stem-end rot of coconut (Rosado et al., 2016) and *L. citricola* in disease of English walnut (Chen, Fichtner et al., 2013). Although we could find no reports of these species as endophytes, their relatively recent description makes it likely that their full extent of occurrence has not been revealed. Sequencing phylogenetically informative loci for these endophytic *Lasiodiplodia* strains – e.g., EF1α and beta-tubulin (TUB2) (Silva, Phillips et al., 2019) – will be desirable in the future to confirm their identity with phylogenetic analysis.

Like *Lasiodiplodia*, multiple *Fusarium* strains are phytopathogenic (Aoki, O'Donnell and Geiser, 2014), and *Fusarium oxysporum* and *Fusarium graminearum* both feature in the top 10 most economically/scientifically important fungal plant pathogens (Dean et al., 2012). This is certainly relevant to commercial banana crops, which are under threat from Foc TR4, the causal agent of Fusarium Wilt (Dita et al., 2018). Fungi in the genus *Fusarium* are also known to be common endophytes in *Musa* species, however, having been previously isolated from either wild or commercial

Musa in China, Thailand and Guatemala (zum Felde, Pocasangre and Sikora, 2003; Sikora et al., 2008; Wang, Chen et al., 2019). The species complexes represented in this study – FIESC, FFSC and 'FSSC' – are all known to comprise both phytopathogens and endophytes (Kavroulakis et al., 2007; Aoki, O'Donnell and Geiser, 2014; Niehaus et al., 2016; Bilal et al., 2018; Wang, Chen et al., 2019), and additionally both the FIESC and 'FSSC' contain species that act as opportunistic human pathogens (Zhang, O'Donnell et al., 2006; O'Donnell, Sutton et al., 2009). Even within species, *Fusarium* strains can differ greatly in their proclivity to cause disease in their plant host – *in vitro* expression of secondary metabolites (including phytohormones and mycotoxins) in an orchid endophytic *Fusarium proliferatum* strain was shown to be distinct from expression in a pathogenic *F. proliferatum* strain (Niehaus et al., 2016). It has also been demonstrated that commercial banana roots can be protected from nematodes by endophytic FOSC strains (zum Felde, Pocasangre and Sikora, 2003; Mendoza and Sikora, 2009), the same species complex to which Foc TR4 belongs. We should highlight that the taxonomy of *Fusarium* is highly contested (Summerell, 2019). Recent dismantling and splitting of certain *Fusarium* species complexes into several distinct genera (Lombard et al., 2015), including reassigning species in the 'FSSC' to the genus *Neocosmospora* (Sandoval-Denis, Lombard and Crous, 2019), has received pushback, the main opposing argument being that a broader generic concept benefits practitioners dealing with human and plant pathogens (O'Donnell, Al-Hatmi et al., 2020). Different perspectives on the limits of the generic concept of *Fusarium* (illustrated in Figure 2.5) will no doubt continue to be debated, and are discussed in more detail in Chapter 3.

Unlike the former two genera, *Aspergillus* is not known predominantly for plant associated taxa, but rather for globally distributed air and soilborne saprotrophs, with some species infamously acting as opportunistic human pathogens (Bennett, 2010; Latgé and Chamilos, 2020). Nonetheless, *Aspergillus* species are also frequently found as endophytes, and an endophytic *Aspergillus fumigatus* strain isolated from *Oxalis corniculata* roots has been shown to promote growth in rice (Bilal et al., 2018). Intriguingly, the most prevalent OTU for the genus in this study was classified as *Aspergillus caesiellus*, which has been reported as a marine endophyte of seagrasses and sponges (Liu, Li et al., 2010; Subrmaniyan, Ponnambalam and Thirunavukarassu, 2018). The second most prevalent OTU was *Aspergillus flavus*, a ubiquitous soil fungus known for contaminating stored grains with aflatoxins, and also an agent of aforementioned opportunistic diseases in animals and humans (Amaike and Keller, 2011). The range of plant–fungal interactions that are observed in these three genera emphasises the ongoing question we face for the endophytic lifestyle as a whole – how can we distinguish mutualistic or commensal endophytes from latent pathogens? Greater exploration of the genomic features and expression profiles of endophytes is required to tackle this issue, and seed banks provide an excellent resource for targeting economically, environmentally and scientifically important plant hosts from which to isolate strains for this purpose.

A relatively modest sampling effort was required to isolate the majority of common OTUs found across *Musa* species (Figure 2.1B), in agreement with other microfungi community studies (e.g., Paulus et al., 2006; Tisthammer, Cobian and Amend, 2016; Vaz et al., 2018), but the vast majority of OTUs were singleton or low-abundance, a known phenomenon in microbial diversity (Lynch and Neufeld, 2015; Jia, Dini-Andreote and Falcão Salles, 2018). A disproportionate number of rare taxa can obstruct community composition visualisation methods such as NMDS, and so low abundance taxa are often filtered out in order to visualise structural patterns (e.g., Miller, Hopkins et al., 2016; Huang, Bowman et al., 2018; U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019). This is distinct from the practise of removing rare/singleton OTUs from high throughput sequencing datasets

in case of sequencing artefacts (e.g., Brown et al., 2015). Poos and Jackson (2012) discussed two arguments for removal of rare taxa in multivariate analysis in the context of bioassessments: statistical impact ('rare species provide limited interpretative value and add noise') and biological impact ('rare taxa do not provide meaningful information beyond that captured by more common species'). For our comparison of endophyte communities between different host habitats, PERMANOVA analysis found the effect size of habitat on community variance to be greater when excluding rare taxa (Table 2.1), but when comparing the significance of individual habitats with pairwise PERMANOVA the inclusion of rare taxa revealed differences between habitats that were not found from the common taxa alone (Figure 2.4B). This challenges the 'biological impact' argument above, however removal of rare taxa remains a practical compromise to enable visualisation of at least a subset of the community structure. A valid question is whether the rare taxa that were detected are conditionally rare (i.e., their abundance is based on abiotic conditions), or permanently rare. We would need repeated samples over time to clarify this and, although outside the scope of this study, seed bank collections are excellently positioned for addressing this in the future.

While the impact of 'edge effect' – change in community structure at the boundary of habitats, whether natural or from e.g., encroaching human land use or bisecting roads – has been well documented for plant communities (Skole and Tucker, 1993; Harper et al., 2005; Kunert et al., 2015), it is far less studied in fungi (Crockatt, 2012; Ruete, Snäll and Jönsson, 2016), and, to our knowledge, the concept has not been addressed in the context of endophytes. Our results comparing the jungle buffer, jungle edge and roadside habitats suggest that there may indeed be some level of edge effect manifested in the seed mycobiome of these *Musa* accessions, both in diversity and abundance (Figures 2.4C,D). While community composition did not differ between these three habitats for the most common taxa, when including rare taxa there was a significant difference between the jungle buffer and roadside habitats (Figure 2.4B). Seeing a community difference between these habitats when including rare and not just common taxa suggests that the rare endophytes may be more sensitive to edge effects, which would be consistent with the concept of biotic homogenisation as a result of ecosystem disruption (McKinney and Lockwood, 1999; Parra-Sanchez and Banks-Leite, 2020). This is also supported by the fact that the Shannon index, which is sensitive to rare species, found a significant difference in diversity, while the Simpson index, which is sensitive to abundant species, did not (Morris, Caruso et al., 2014). These results come with the caveat that the habitats as defined in this study are interpreted from the MSB seed collection data, which were not recorded with any particular study design in mind, and as such some entries are more complete than others and there can be subjectivity in how to infer habitat from the collection notes. While the extensive metadata attached to natural history collections can be incredibly powerful for studying patterns of biodiversity (Andrew, Diez et al., 2018; Funk, 2018; Andrew, Büntgen et al., 2019; Pearce et al., 2020), the application of that data must be done with care.

Fitting post-storage seed viability and germination rate data to the NMDS visualisation showed jungle habitat accessions to be associated with highest seed viability and ravine habitat accessions to be associated with lowest seed viability. As these assessments specifically measured post-storage viability/germination, we relied on the assumption that the same collection standards and procedures were always adhered to, as other factors have been shown to impact *Musa* seed viability such as maturity of the seed at collection and the speed of drying before cold storage (Kallow et al., 2020). Nonetheless, these results highlighted *Fusarium* and *Lasiodiplodia* strains, which would be particularly interesting to trial in experimental inoculation studies, to verify whether they impact

the survivability of *Musa* seeds in storage, or indeed the germination rates of fresh seeds. Endophytic *Fusarium* strains have previously been found to promote germination and seedling growth of an Indonesian peatland grass (Tamura et al., 2008) and germination of orchid seeds (Bayman and Otero, 2006). In addition to the aforementioned roles of *L. theobromae* in tropical fruit tree diseases, it has also been implicated in seed rot, for instance of slash pine (Cilliers, Swart and Wingfield, 1993), and to cause reduced germination rates in aridan and coconut seeds (Dugan et al., 2016). The role of seedborne *L. theobromae* on germination may be more nuanced, however, as it has been found to produce fatty acid esters, which can alternately inhibit and promote tobacco seed germination and seedling growth (Uranga et al., 2016). Considering the pathogenic role of numerous *Lasiodiplodia* species discussed above, it is interesting that this study saw *Lasiodiplodia* strains to be prevalent in *Musa* accessions with comparatively high post-storage seed viability. A previous study of *in vitro* germination of both stored and fresh *M. ornata* seeds found *Lasiodiplodia* to persistently infect seeds, with the implication that these seeds then decayed (Burgos-Hernández et al., 2014). Goos, Cox and Stotzky (1961) reported a similar result for seeds of various *Musa* spp. in aseptic conditions, however they noted that germination was not significantly affected by *Lasiodiplodia* colonisation under 'greenhouse conditions'. This raises the question as to whether the pathogenic potential of *Lasiodiplodia* strains in *Musa* seeds is influenced by the abiotic conditions and/or co-occurrence of other fungi. Of course, without isolating specific strains and performing controlled pathogenicity tests, it is impossible to answer this, as different fungal strains can vary in their ability to cause disease regardless of secondary factors such as environment. It would also be interesting to look at the endosymbiotic or 'endohyphal' bacteria associated with our strains, as these have been found, in rare cases, to be capable of effecting (pre-storage) seed germination and viability in a neotropical tree (Shaffer et al., 2018).

An interesting result was that the abundance of endophytes per *Musa* accession was greatest in the ravine habitat (Figure 2.4B), the same habitat that was adversely correlated with post-storage seed viability. Returning to the ambiguity of the endophytic lifestyle, this again raises the issue that it is not the mere presence of endophytes, but the identity of specific strains that may have implications for stored seeds. The difference in abundance and community composition in the ravine habitats could partially be explained by altitude, although unfortunately there was not sufficient altitude data for all accessions in the MSB records to test this. Previous studies on the effect of altitude on endophyte communities have suggested an inconsistent relationship (Granath et al., 2007; Hashizume, Sahashi and Fukuda, 2008; Zubek et al., 2009; Bonfim et al., 2016), no doubt partially due to the large number of confounding factors associated with changing altitude, such as variation in the host plant assemblages, as host availability is believed to be a main driver of endophyte community composition (U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019). Host availability may also have been a key factor as to why accessions in oil palm plantations and a botanical garden had low endophyte abundance (Supplementary Figure S2.3). Although the sample size for these habitats was too small to include them in the main analyses, these were the only managed habitats with, presumably, the least natural co-occurring plant assemblages.

There are a number of considerations for seed banking in the context of endophytes that are important to raise for future discussion and research. Firstly, our results show that habitat of the host plants from which seeds are collected could impact the associated endophyte communities, which may potentially have downstream consequences for seed survival. Collecting seeds from individuals in a range of habitats with different co-occurring plant species may be advisable to conserve endo-

phytic diversity. As current seed bank protocol is to collect seeds pre-dispersal, before horizontal transmission of fungi from soil to seed, what, if any, impact does this have on subsequent viability of the seeds or health of the descendent plants? To our knowledge, only one study has made a direct comparison of endophytic communities in pre and post-dispersal seeds for the same plant individual, finding fewer endophytes in pre-dispersal seeds of a neotropical tree species, none of which were successfully isolated in culture (Gallery, Dalling and Arnold, 2007). Studies of buried seeds have shown that seeds acquire diverse endophytes through horizontal transmission from the soil (e.g., U'Ren, Dalling et al., 2009; Sarmiento et al., 2017), but are also vulnerable to soilborne pathogens (Gallery, Moore and Dalling, 2010). It could then be that the current protocol of storing pre-dispersal seeds is preferable, as it limits the acquisition of potential pathogens while still allowing the possibility for mutualistic endophytes to be vertically transmitted from the parent plant. The dynamics of endophyte transmission are likely to be highly variable between different plant groups, however, and more studies of seeds from different hosts, geographical areas and dispersal stages are needed to identify the optimal collection procedure for healthy microbiomes of stored seeds.

2.6 Conclusions

This study has demonstrated that seed banks provide huge potential for research into fungal endophyte communities. As well as being an untapped resource for new fungal diversity, the ability to isolate live strains from almost 40,000 global plant taxa curated by the MSB – a third of which are identified as having significant natural capital value (Liu, Breman et al., 2018) – provides far-reaching opportunities for future study of the role of endophytes in plant health. For this reason, although originally designed for conservation of plant genetic diversity, seed banks may have an equally important role in conserving the seed microbiome, and much more discussion and research is needed on how the seed collection and storage procedure can best accommodate this.

Acknowledgements

We thank the Malaysian Agricultural Research and Development Institute and the Vietnamese Plant Resources Center for providing seeds to the MSB, in particular Binti Tahir, M. Anuar Rasyidi, M. N., Ahmad Syahman, M. D., Mohd Shukri, M. A., Suryanti, B., Dang Toan Vu, Tuong Dang Vu, Le Thi Loan and Ngo Duc. We thank Toby Tydeman, Khushboo Gurung, Emily Kennedy and Tom Bance for help in processing samples. We are grateful to Jana M. U'Ren for useful discussion and Bryn T. M. Dentinger for his generosity and inspiration at the outset of the study. Thanks also to the Frontiers in Biology guest editor and the reviewers Jana M. U'Ren and Asha Janadaree Dissanayake for their valuable feedback on the published paper.

2.7 Supplementary material

For the full table of OTU classification results, see Supplementary Data Sheet 1 at `https://www.frontiersin.org/articles/10.3389/fmicb.2021.643731/full#supplementary-material`.

Supplementary Table S2.1: MSB serial numbers and metadata associated for the 45 wild *Musa* accessions. (TTC=tetrazolium chloride, ER=embryo rescue)

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
836375	*Musa balbisiana*	2014	Vietnam, Hà Tĩnh Province, Kỳ Anh District, Kỳ Hoa commune N18°1'29.4"E106°16'51.48"	56	25	0	Jungle buffer
836445	*Musa itinerans*	2014	Vietnam, Hà Tĩnh Province, Hương Sơn District, Sơn Kim commune N18°25'37.38"E105°12'53.95"	0	NA	0	Jungle buffer
836467	*Musa itinerans*	2014	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°37'14.95"E105°12'36.68"	63	63	0	Jungle buffer
836478	*Musa balbisiana*	2014	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'16.11"E105°14'15.78"	88	10	0	Jungle buffer
836489	*Musa balbisiana*	2014	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'29.8"E105°14'15.87"	86	70	0	Jungle buffer
836490	*Musa itinerans*	2014	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'14.89"E105°14'50.83"	52	70	0	Jungle buffer

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
836504	*Musa itinerans*	2014	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'30.41"E105°15'42.39"	88	0	0	Jungle buffer
836515	*Musa itinerans*	2014	Vietnam, Nghệ An Province, Con Cuông District, Châu Khê commune N19°1'48.73"E104°43'31.97"	60	80	0	Jungle buffer
880079	*Musa balbisiana* var. *bakeri*	2015	Vietnam, Lai Châu Province, Sìn Hồ District,, Phă Sô Lin commune N22°21'32.5"E103°16'37.4"	6	0	0	Ravines
880116	*Musa itinerans*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°14'41.7"E103°56'41.6"	0	10	10	Ravines
880127	*Musa balbisiana*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°14'41.6"E103°57'30.6"	0	0	10	Ravines
880138	*Musa itinerans*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°15'13.5"E103°57'1.5"	0	0	10	Ravines
880149	*Musa itinerans*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°15'10.2"E103°56'35.5"	0	0	0	Ravines

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
880161	*Musa balbisiana*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°15'2.1"E103°56'41.1"	0	10	0	Ravines
880172	*Musa balbisiana*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°14'21.3"E103°57'2.24"	9	0	0	Ravines
880264	*Musa itinerans*	2015	Vietnam, Lai Châu Province, Tam Đường District, Bản Bo N22°17'33.3"E103°40'34.4"	4	0	0	Ravines
880323	*Musa itinerans*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°21'14.3"E103°46'43"	3	0	0	Ravines
880334	*Musa itinerans*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°21'23.1"E103°47'16.1"	0	0	0	Ravines
880345	*Musa balbisiana* var. *bakeri*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°21'23.7"E103°47'11.6"	54	0	0	Ravines
880356	*Musa balbisiana*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°21'14.6"E103°46'40.5"	0	0	0	Ravines

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
880367	*Musa balbisiana*	2015	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°19'31.7"E103°46'21.5"	6	0	0	Ravines
880585	*Musa balbisiana* var. *bakeri*	2015	Vietnam Lai Châu Province Sìn Hồ District N22°6'53"E103°10'41.7"	12	75	0	Jungle edge
880600	*Musa balbisiana* var. *bakeri*	2015	Vietnam Lai Châu Province Phong Thổ District, Lả Nhì Thàng commune N22°27'41.9"E103°22'11.3"	60	80	0	Jungle edge
880622	*Musa itinerans*	2015	Vietnam Lai Châu Province Phong Thổ District, Ma Li Pho commune N22°36'8.9"E103°11'3.31"	23	30	0	Jungle edge
880633	*Musa itinerans*	2015	Vietnam Lai Châu Province Phong Thổ District, Khổng Lào commune N22°32'56.6"E103°20'34.9"	45	78	0	Jungle edge

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
880644	*Musa itinerans*	2015	Vietnam Lai Châu Province Phong Thổ District, Hoàng Thèn commune N22°34'35.6"E103°17'42.5"	51	30	0	Jungle edge
882671	*Musa acuminata*	2015	Malaysia Peninsula Malaysia Pahang N3°42'44.1"E103°2'2.04"	36	0	0	Roadside
882730	*Musa acuminata*	2015	Malaysia Peninsula Malaysia Pahang N3°52'15.66"E102°11'46.02"	46	0	0	Roadside
882741	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Pahang N3°53'47.94"E102°12'24.24"	83	70	0	Roadside
882785	*Musa acuminata*	2015	Malaysia Peninsula Malaysia Negeri Sembilan N2°29'59"E102°10'33.5"	48	56	0	Roadside
882800	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Negeri Sembilan N2°48'31.2"E102°20'38.2"	64	33	0	Roadside

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
882811	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Negeri Sembilan N2°48'31.2"E102°20'38.2"	71	0	0	Roadside
882833	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Selangor N2°56'56.7"E102°47'16.3"	73	20	0	Roadside
882877	*Musa gracilis*	2015	Malaysia Peninsula Malaysia Pahang N3°53'48.2"E102°12'24.5"	0	0	0	Roadside
882888	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Johor N2°6'53.88"E102°40'38.82"	51	50	0	Oil palm plantation
882899	*Musa acuminata* subsp. *malaccensis*	2015	Malaysia Peninsula Malaysia Johor N2°6'53.88"E102°40'38.82"	32	100	0	Oil palm plantation
928337	*Musa acuminata*	2016	Malaysia Peninsula Malaysia Pahang N3°20'51.5"E101°48'58.3"	36	40	0	Jungle edge

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
928360	*Musa violascens*	2016	Malaysia Peninsula Malaysia Pahang N3°19'17.2"E101°51'29.8"	13	0	0	Roadside
928429	*Musa acuminata*	2016	Malaysia Peninsula Malaysia Johor N2°4'44.5"E103°22'19.7"	18	40	0	Oil palm plantation
928500	*Musa acuminata* subsp. *microcarpa*	2016	Malaysia Peninsula Malyasia Pahang N4°18'23.6"E101°41'4.02"	50	0	0	Roadside
928717	*Musa balbisiana*	2015	Vietnam Hà Giang Province Hoàng Su Phì District Nậm Dịch commune N22°39'9.71"E104°41'57.87"	40	20	0	Jungle edge
928728	*Musa balbisiana* var. *balbisiana*	2016	Vietnam Hà Giang Province Hoàng Su Phì District Nậm Dịch commune N22°34'13.09"E104°47'28.63"	58	0	0	Jungle edge

Supplementary Table S2.1 continued.

Serial number	Species	Collection year	Collection location	Seed viability (TTC) (%)	Germination rate (ER) (%)	Contamination (ER) (%)	Habitat
928739	*Musa balbisiana*	2016	Vietnam Hà Giang Province Hoàng Su Phì District Nậm Dịch commune N22°34'10"E104°47'30.54"z	46	10	0	Jungle edge
928740	*Musa balbisiana* var. *balbisiana*	2016	Vietnam Hà Giang Province Hoàng Su Phì District Nậm Dịch commune N22°36'33.17"E104°45'43.36"	54	20	0	Jungle edge
944548	*Musa velutina*	2017	Royal Botanic Gardens, Kew	NA	NA	NA	Botanical garden

Supplementary Table S2.2: GenBank accession numbers for taxa used in the phylogenetic analysis. Accessions in bold were sequenced during this study.

Voucher	Species	EF1α	RPB1	RPB2
NRRL 13412	*Albonectria rigidiuscula* *(=Fusarium decemcellulare)*		JX171453	JX171567
NRRL 36160	*Bisifusarium delphinoides* *(=Fusarium delphinoides)*	HM347134	JX171535	HM347219
NRRL 20691	*Bisifusarium dimerum* *(=Fusarium dimerum)*	EU926349	JX171478	JX171592
NRRL 36168	*Bisifusarium lunatum* *(=Fusarium lunatum)*	EU926291	JX171536	JX171648
NRRL 20689	*Bisifusarium nectrioides* *(=Fusarium nectrioides)*	EU926312	JX171477	JX171591
NRRL 20711	*Bisifusarium penzigii* *(=Fusarium penzigii)*	HM347132	JX171482	HM347217
NRRL 36148	*Cyanonectria buxi* *(=Fusarium buxicola)*		JX171534	JX171647
NRRL 13308	*Fusarium acutatum*	AF160276		
NRRL 22152	*Fusarium albidum*		JX171492	JX171605
NRRL 20459	*Fusarium albosuccineum*		JX171471	JX171585
NRRL 25385	*Fusarium anguioides*		JX171511	JX171624
NRRL 32997	*Fusarium arcuatisporum* 'FIESC 7a'	GQ505624		GQ505802
NRRL 6227	*Fusarium armeniacum*	HM744692	JX171446	JX171560
NRRL 13818	*Fusarium asiaticum*	AF212451	JX171459	JX171573
NRRL 54939	*Fusarium avenaceum*		JX171551	JX171663
NRRL 25410	*Fusarium aywerte*		JX171513	JX171626
NRRL 25539	*Fusarium babinda*		JX171519	JX171632
NRRL 25174	*Fusarium beomiforme*		JX171506	JX171619
NRRL 31008	*Fusarium brachygibbosum*		JX171529	JX171642
NRRL 43638	*Fusarium brevicaudatum* 'FIESC 6a'	GQ505665		GQ505843
NRRL 13371	*Fusarium buharicum*		JX171449	JX171563
NRRL 13829	*Fusarium* cf. *compactum*		JX171460	JX171574
NRRL 25331	*Fusarium circinatum*	KM231943	JX171510	JX171623
NRRL 32871	*Fusarium clavum* 'FIESC 5a'	GQ505619		GQ505797
NRRL 28577	*Fusarium coffeatum* 'FIESC 28a'	GQ505603		GQ505781
NRRL 28387	*Fusarium commune*		JX171525	JX171638
NRRL 36323	*Fusarium compactum* 'FIESC 3a'	GQ505648		GQ505826
NRRL 13459	*Fusarium concolor*	GQ505674	JX171455	JX171569
NRRL 3020	*Fusarium croceum* 'FIESC 10a'	GQ505586		GQ505764
InaCC F983	*Fusarium cugenangense*	LS479756	LS479559	LS479307

Supplementary Table S2.2 continued.

Voucher	Species	EF1α	RPB1	RPB2
NRRL 25475	*Fusarium culmorum*	AF212463	JX171515	JX171628
NRRL 53998	*Fusarium cyanostomum*	HM626647	JX171546	JX171658
NRRL 29976	*Fusarium domesticum*	EU926286	JX171528	JX171641
NRRL 36401	*Fusarium duofalcatisporum* 'FIESC 2a'	GQ505651		GQ505829
FocMal43	*Fusarium duoseptatum* 'Race1'	LS479653		LS479207
NRRL 20697	*Fusarium equiseti*	GQ505594	JX171481	JX171595
NRRL 6548	*Fusarium flagelliforme* 'FIESC 12a'	GQ505589		GQ505767
NRRL 25473	*Fusarium flocciferum*		JX171514	JX171627
NRRL 28852	*Fusarium fractiflexum*	AF160288		
NRRL 13566	*Fusarium fujikuroi*	AF160279	JX171456	JX171570
NRRL 45417	*Fusarium gaditjirri*		JX171542	JX171654
NRRL 26131	*Fusarium globosum*	AF160285	KF466396	KF466406
NRRL 43635	*Fusarium gracilipes* 'FIESC 13a'	GQ505662		GQ505840
NRRL 31084	*Fusarium graminearum*	HM744693	JX171531	JX171644
NRRL 20692	*Fusarium graminum*		JX171479	JX171593
InaCC F820	*Fusarium grosmichelii* 'Race1'	LS479810		LS479364
NRRL 32865	*Fusarium guilinense* 'FIESC 21b'	GQ505614		GQ505792
NRRL 22945	*Fusarium guttiforme*		JX171505	JX171618
NRRL 26417	*Fusarium hainanense* 'FIESC 26a'	GQ505598		GQ505776
NRRL 20693	*Fusarium heterosporum*		JX171480	JX171594
InaCC F866	*Fusarium hexaseptatum* 'Race1'	LS479805		LS479359
NRRL 29889	*Fusarium hostae*	AY329034	JX171527	JX171640
NRRL 13379	*Fusarium incarnatum* 'FIESC 23b'	GQ505591		GQ505769
NRRL 20433	*Fusarium inflexum*	AF008479	JX171469	JX171583
NRRL 43637	*Fusarium ipomoeae* 'FIESC 1a'	GQ505664		GQ505842
NRRL 32175	*Fusarium irregulare* 'FIESC 15a'	GQ505609		GQ505787
NRRL 20423	*Fusarium lacertarum*	GQ505593	JX171467	JX171581
NRRL 54940	*Fusarium langsethiae*		JX171550	JX171662
NRRL 36372	*Fusarium longifundum* 'FIESC 11a'	GQ505649		GQ505827
NRRL 13368	*Fusarium longipes*		JX171448	JX171562
NRRL 31167	*Fusarium luffae* 'FIESC 18a'	GQ505608		GQ505786
NRRL 54252	*Fusarium lyarnte*		JX171549	JX171661
NRRL 25226	*Fusarium mangiferae*	AF160281	JX171509	JX171622
NRRL 26231	*Fusarium miscanthi*		JX171521	JX171634
NRRL 43639	*Fusarium multiceps* 'FIESC 19a'	GQ505666		GQ505844

Supplementary Table S2.2 continued.

Voucher	Species	EF1α	RPB1	RPB2
MUCL 52574	*Fusarium musae*	FN552086		
NRRL 22244	*Fusarium nanum* 'FIESC 25a'	GQ505596		GQ505774
NRRL 54600	*Fusarium nematophilum*		JX171552	JX171664
NRRL 25179	*Fusarium nisikadoi*		JX171507	JX171620
NRRL 36452	*Fusarium nurragi*		JX171538	JX171650
NRRL 54006 FocII5	*Fusarium odoratissimum* 'TR4'	LS479644	LS479459	LS479198
CAV300	*Fusarium oxysporum* f. *cubense* 'TR4'	FJ664932		
NRRL 32864	*Fusarium pernambucanum* 'FIESC 17a'	GQ505613		GQ505791
FocIndo25	*Fusarium phialophorum* 'Race1'	LS479650	LS479464	LS479204
NRRL 13714	*Fusarium poae*		JX171458	JX171572
NRRL 22944	*Fusarium proliferatum*		JX171504	JX171617
NRRL 28062	*Fusarium pseudograminearum*	AF212468	JX171524	JX171637
ATCC76244	*Fusarium purpurascens* 'Race1'	LS479645		LS479199
NRRL 22901	*Fusarium redolens*		JX171503	JX171616
NRRL 22134	*Fusarium rusci*		JX171490	JX171603
NRRL 13999	*Fusarium sacchari*		JX171466	JX171580
NRRL 22187	*Fusarium sambucinum*		JX171493	JX171606
NRRL 20472	*Fusarium sarcochroum*		JX171472	JX171586
NRRL 13402	*Fusarium scirpi*		JX171452	JX171566
NRRL 36526	*Fusarium setosum*		JX171539	JX171651
NRRL 26427	*Fusarium* sp.	AF160286		
NRRL 25309	*Fusarium* sp.	AF160284		
NRRL 25303	*Fusarium* sp.	AF160283		
NRRL 34002	*Fusarium* sp. 'FIESC 22a'	GQ505626		GQ505804
NRRL 20722	*Fusarium* sp. 'FIESC 27a'	GQ505595		GQ505773
NRRL 5537	*Fusarium* sp. 'FIESC 8a'	GQ505588		GQ505766
NRRL 29134	*Fusarium* sp. 'FIESC 9a'	GQ505605		GQ505783
836490-12	***Fusarium* sp. OTU1**	**MW319605**		
836445-03	***Fusarium* sp. OTU2**	**MW319595**		
880323-07	***Fusarium* sp. OTU3**	**MW319620**		
880334-09	***Fusarium* sp. OTU4**	**MW319629**		
880149-04	***Fusarium* sp. OTU5**	**MW319604**		
836490-20	***Fusarium* sp. OTU6**	**MW319587**		
836489-15	***Fusarium* sp. OTU7**	**MW319636**		
836445-18	***Fusarium* sp. OTU8**	**MW319589**		

Supplementary Table S2.2 continued.

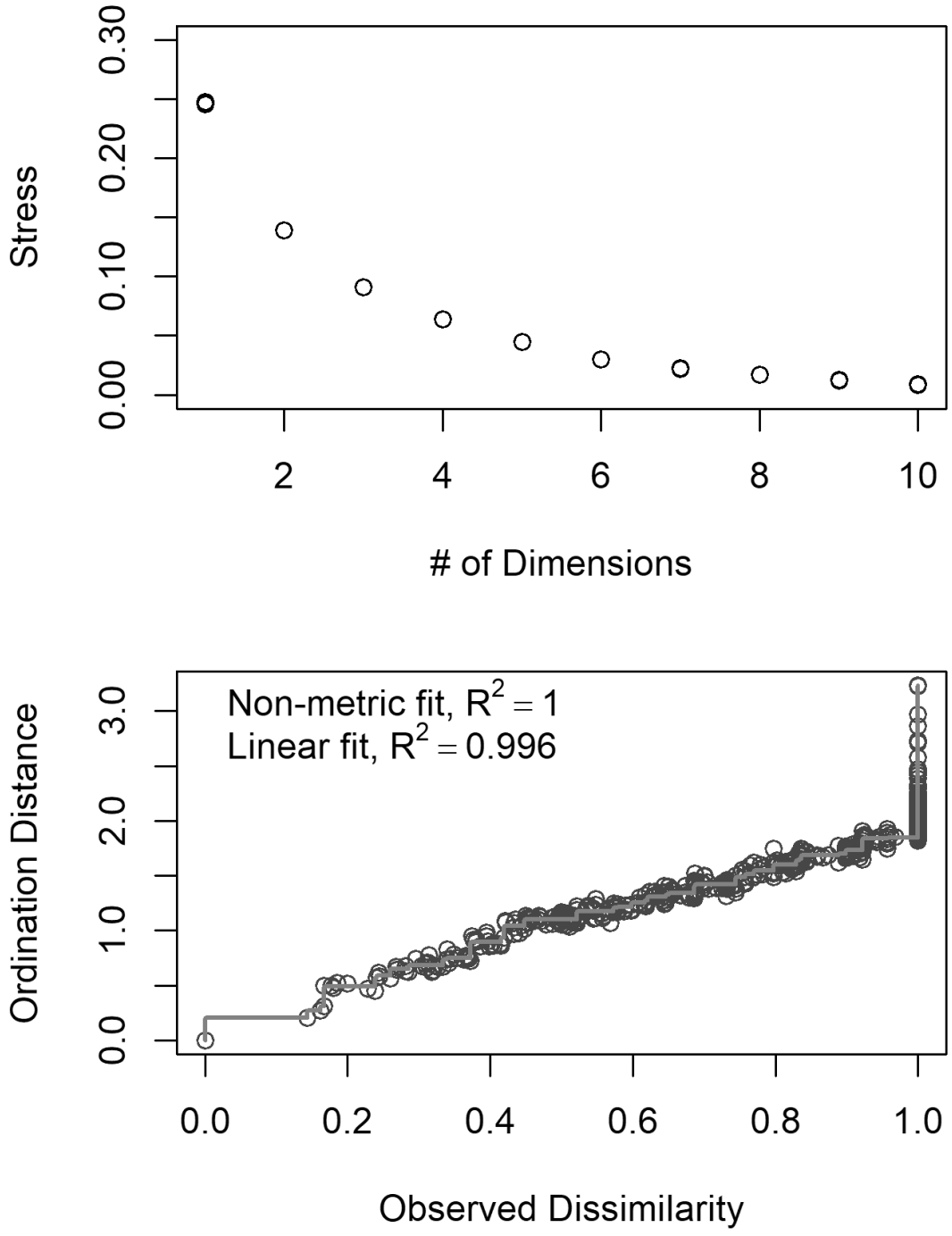
Voucher	Species	EF1α	RPB1	RPB2
880600-17	***Fusarium* sp. OTU9**	**MW319588**		
880138-05	***Fusarium* sp. OTU10**	**MW319601**		
NRRL 3229	*Fusarium sporotrichioides*	HM744665	JX171444	JX171558
NRRL 20429	*Fusarium stilboides*		JX171468	JX171582
NRRL 22016	*Fusarium subglutinans*	HM057336	JX171486	JX171599
NRRL 13384	*Fusarium sublunatum*		JX171451	JX171565
NRRL 34004	*Fusarium sulawesiense* 'FIESC 16a'	GQ505628		GQ505806
NRRL 34005	*Fusarium tanahbumbuense* 'FIESC 24a'	GQ505629		GQ505807
InaCC F956	*Fusarium tardichlamydosporum* 'Race1'	LS479727	LS479532	LS479278
NRRL 22045	*Fusarium thapsinum*		JX171487	JX171600
NRRL 54149	*Fusarium torreyae*		JX171548	JX171660
NRRL 22748	*Fusarium torulosum*		JX171502	JX171615
NRRL 25481	*Fusarium tricinctum*		JX171516	JX171629
NRRL 22196	*Fusarium venenatum*		JX171494	JX171607
NRRL 22566	*Fusarium verrucosum*		JX171500	JX171613
NRRL 20956	*Fusarium verticillioides*		JX171485	JX171598
NRRL 25486	*Fusarium xylarioides*		JX171517	JX171630
NRRL 22316	*Geejayessia atrofusca* *(=Fusarium staphyleae)*		JX171496	JX171609
NRRL 22465	*Geejayessia zealandica* *(=Fusarium zealandicum)*		JX171498	JX171611
NRRL 20438	*Neocosmospora ambrosia* *(=Fusarium ambrosium)*	AF178332	JX171470	JX171584
NRRL 43529	*Neocosmospora falciformis* *(=Fusarium falciforme)*	EF452965	JX171541	JX171653
NRRL 22090	*Neocosmospora illudens* *(=Fusarium illudens)*	AF178326	JX171488	JX171601
NRRL 45880	*Neocosmospora pisi* *(=Fusarium solani* f. sp. *pisi)*		JX171543	JX171655
NRRL 22632	*Neocosmospora plagianthi* *(=Fusarium plagianthi)*	AF178354	JX171501	JX171614
NRRL 22436	*Neocosmospora vasinfecta* *(=Fusarium neocosmosporiellum)*	AF178348	JX171497	JX171610
NRRL 22276	*Neocosmospora phaseoli* *(=Fusarium phaseoli)*	EF408415	JX171495	JX171608
NRRL 31041	*Neocosmospora phaseoli* *(=Fusarium virguliforme)*		JX171530	JX171643
NRRL 20485	*Neonectria coccinea*		JX171474	JX171588

Supplementary Table S2.2 continued.

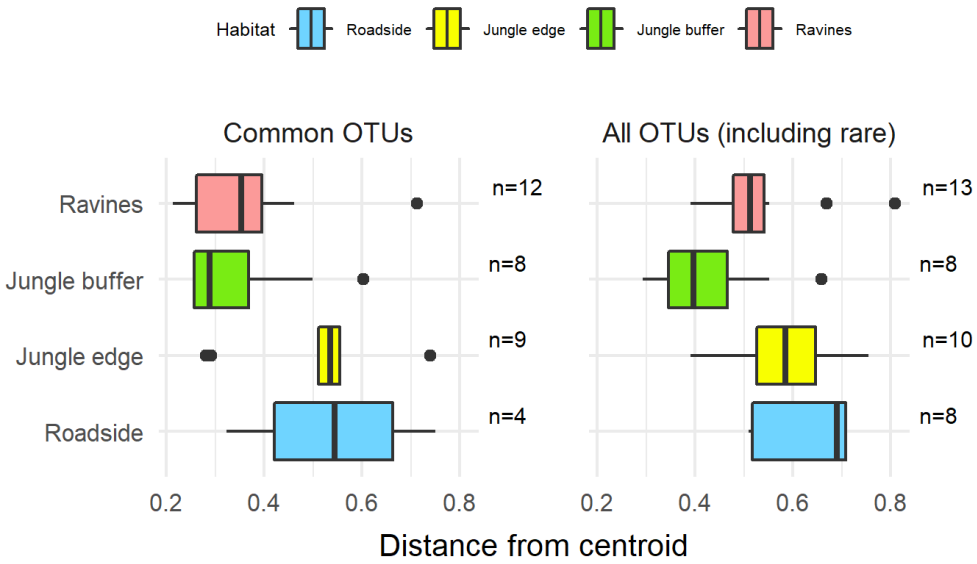
Voucher	Species	EF1α	RPB1	RPB2
NRRL 22505	*Neonectria neomacrospora (=Cylindrocarpon cylindroides)*		JX171499	JX171612
NRRL 20846	*Rectifusarium ventricosum (=Fusarium ventricosum)*		JX171484	JX171597

Supplementary Table S2.3: Summary of OTUs which were also found on the unsterilised seed surface.

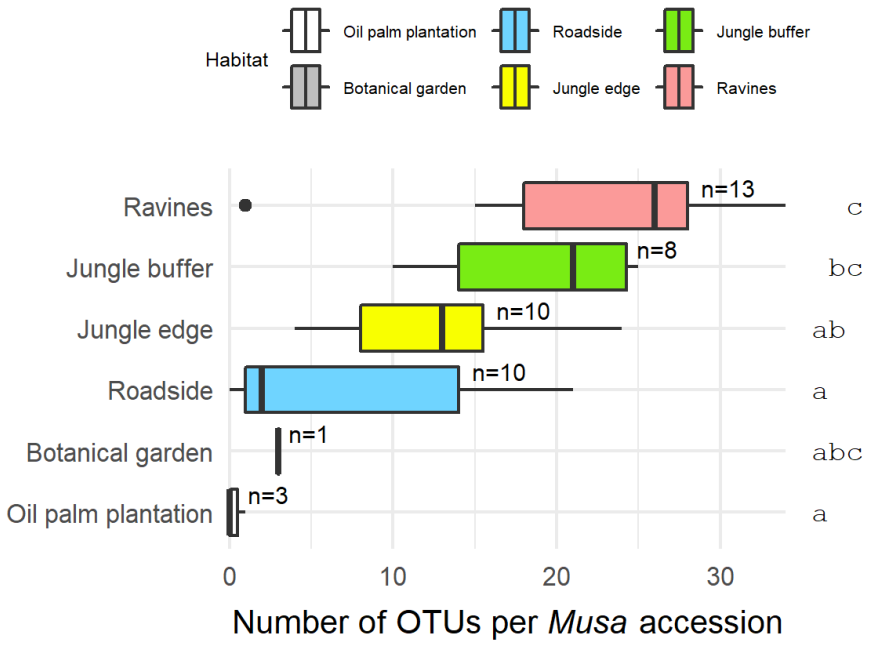
OTU	Species	Count
Otu22	*Fusarium concentricum*	5
Otu52	*Capnodiales* sp.	1
Otu6	*Lasiodiplodia citricola*	2
Otu8	*Nectriaceae* sp.	1
Otu176	*Ascomycota* sp.	1
Otu44	*Neofusicoccum parvum*	1
Otu99	*Xylariales* sp.	1
Otu13	*Aspergillus flavus*	1
Otu27	*Aspergillus niger*	7
Otu196	*Penicillium meleagrinum* var. *viridiflavum*	1
Otu26	*Capnodiales* sp.	2
Otu62	*Talaromyces ricevillensis*	1
Otu166	*Hypocreales* sp.	1
Otu96	*Ascomycota* sp.	1
Otu36	*Penicillium solitum*	2



Supplementary Figure S2.1: Scree plot up to 10 dimensions for the NMDS analysis (top) and stress plot for the chosen number of 6 dimensions (bottom).



Supplementary Figure S2.2: Data dispersion for each habitat for both the common taxa used in the NMDS and all taxa including rare OTUs, as assessed with betadisper. Sample size (number of accessions) is shown to the right of the plots. Sample size is less for the roadside habitat in the 'All OTUs' category than in the diversity and abundance analyses (Figure 2.4C,D) because accessions with no OTUs detected are removed by betadisper.



Supplementary Figure S2.3: Abundance of OTUs per *Musa* accession for each habitat including oil palm plantation and botanical garden. Groups with significant difference of means as calculated by TukeyHSD are shown by letters on the right of the plots. Sample size (number of accessions) is shown to the right of boxes.

Chapter 3

Lifestyle transitions in fusarioid fungi are frequent and lack clear genomic signatures

Publication details

This chapter has been published as the following paper:

> **Hill, R.**, Buggs, R.J.A., Vu, D.T., Gaya, E. (2022). Lifestyle transitions in fusarioid fungi are frequent and lack clear genomic signatures. *Molecular Biology and Evolution* 39(4):msac085. DOI: 10.1093/molbev/msac085.

R.H. designed the study, performed molecular lab work, implemented the analysis and wrote the paper. E.G. and R.J.A.B. supervised the study, designed the analysis and wrote the paper. D.T.V. provided the samples and read and approved the final manuscript.

3.1 Abstract

The fungal genus *Fusarium* (*Ascomycota*) includes well-known plant pathogens that are implicated in diseases worldwide, and many of which have been genome sequenced. The genus also encompasses other diverse lifestyles, including species found ubiquitously as asymptomatic-plant inhabitants (endophytes). Here, we produced structurally annotated genome assemblies for five endophytic *Fusarium* strains, including the first whole-genome data for *Fusarium chuoi*. Phylogenomic reconstruction of *Fusarium* and closely related genera revealed multiple and frequent lifestyle transitions, the major exception being a monophyletic clade of mutualist insect symbionts. Differential codon usage bias and increased codon optimisation separated *Fusarium sensu stricto* from allied genera. We performed computational prediction of candidate secreted effector proteins (CSEPs) and carbohydrate-active enzymes (CAZymes) – both likely to be involved in the host–fungal interaction – and sought evidence that their frequencies could predict lifestyle. However, phylogenetic distance described gene variance better than lifestyle did. There was no significant difference in CSEP, CAZyme, or gene repertoires between phytopathogenic and endophytic strains, although we

did find some evidence that gene copy number variation may be contributing to pathogenicity. Large numbers of accessory CSEPs (i.e., present in more than one taxon but not all) and a comparatively low number of strain-specific CSEPs suggested that there is limited specialisation among plant associated *Fusarium* species. We also found half of the core genes to be under positive selection and identified specific CSEPs and CAZymes predicted to be positively selected on certain lineages. Our results depict fusarioid fungi as prolific generalists and highlight the difficulty in predicting pathogenic potential in the group.

3.2 Introduction

Fusarium (*Hypocreales*, *Ascomycota*) is a globally distributed genus of approximately 230 species (`https://www.fusarium.org/`), many of which are implicated in devastating fungal diseases of plants. For instance, throughout the first half of the 20th century, Fusarium wilt of banana single-handedly wiped out the main globally traded banana cultivar – equivalent to losses of at least US$2.3 billion in 2000 (Ploetz, 2005). A new *Fusarium* epidemic is now affecting the current dominant banana cultivar (Ordonez et al., 2015). Moreover, on the much-cited list of the top 10 fungal plant pathogens by Dean et al. (2012), two spots belong to *Fusarium* species. Beyond plant pathogenicity, however, many species are also reported to exhibit an array of other fungal lifestyles (see Appendix A.3), and *Fusarium* strains are also frequently isolated as endophytes from inside healthy plant tissues (e.g., Parsa et al., 2016; Zakaria and Aziz, 2018; Rashmi, Kushveer and Sarma, 2019; Chapter 2). As outlined in Chapter 1.1, there is no single role that endophytes play in the plant host, as the endophytic lifestyle represents a functional range between pathogenicity and mutualism, which has been dubbed the 'endophytic continuum' (Schulz and Boyle, 2005). The need to categorise pathogenic potential of *Fusarium* taxa is obvious considering the ubiquity of *Fusarium* endophytes in our crops (e.g., Rubini et al., 2005; Leoni et al., 2013; Sandoval-Denis, Guarnaccia et al., 2018) and the ramifications of pathogenic *Fusarium* strains for food security (e.g., Menzies, Koch and Seywerd, 1990; Kokkonen et al., 2010; Okello et al., 2020).

In >200 years since *Fusarium* was first described, the generic concept has been the source of lively debate (Summerell, 2019). In recent years, many *Fusarium* species complexes have been reclassified into distinct 'fusarioid' genera based on phenotypic and phylogenetic evidence – such as *Albonectria*, *Bisifusarium*, *Cyanonectria*, *Geejayessia*, *Neocosmospora* and *Rectifusarium* (Schroers et al., 2011; Lombard et al., 2015; Sandoval-Denis, Lombard and Crous, 2019) – resulting in a narrower definition of the genus, *Fusarium sensu stricto*. This has been opposed in some quarters, with the argument that retaining a broader definition of the genus (*Fusarium sensu lato*) is desirable to facilitate communication between scientists and practitioners dealing with agriculturally and clinically important species that have historically been classified under *Fusarium* (O'Donnell, Al-Hatmi et al., 2020; Geiser, Al-Hatmi et al., 2021). Crous, Lombard et al. (2021) countered that, in light of ever-increasing species discovery and recognised chemical and morphological differences between these clades, reclassification of certain species complexes into different genera is both biologically and practically meaningful. However, both sides of the debate note that ecology is similar among many of these taxa, and so questions regarding lifestyle warrant a perspective that includes allied fusarioid genera.

An evolutionary genomics approach using genomes from diverse lifestyles of fusarioid fungi could address this issue of detecting where strains fall on the pathogenic-mutualistic spectrum. A phylo-

genomic framework could not only shed light on the timing and frequency of lifestyle transitions in the group, but also inform to what extent genetic content is shared between taxa due to ancestry versus lifestyle. In addition to comparing gene repertoires, detecting signatures of selection may also help to uncover the genetic basis of recently evolved traits. Methods based on the ratio of nonsynonymous to synonymous substitutions (dN/dS) and the phenomenon of codon usage bias – where certain codons appear more frequently than others despite encoding the same amino acid – can be used to investigate the extent of selection acting on gene content.

As outlined in Chapter 1.2, one genetic feature that can be particularly illuminating to compare between lifestyles is genes that encode effector proteins. Fungal effectors (known as CSEPs when computationally predicted) are small secreted proteins produced by fungi which mediate the plant–fungal interaction. While best-studied in the context of pathogenicity (Stergiopoulos and de Wit, 2009; de Jonge, Bolton and Thomma, 2011), we now know that effectors are also essential for mutualistic or commensal fungi to form associations with plant hosts by evading the host immune response (Rafiqi et al., 2012; Lo Presti et al., 2015; Plett and Martin, 2015). Effector repertoires have been shown to differentiate host-specific strains (*forma specialis*) in the *Fusarium oxysporum* species complex (FOSC) (van Dam, Fokkens et al., 2016), and could potentially further distinguish pathogenic and endophytic FOSC strains (Czislowski, Zeil-Rolfe and Aitken, 2021). Another frequently studied group of proteins involved in the plant–fungal interaction are CAZymes, many of which act as plant cell wall degrading enzymes (PCWDEs) (Kubicek, Starr and Glass, 2014). CAZymes are often referred to as saprotrophic features (Lebreton et al., 2021), but are also abundant in plant pathogens and endophytes (e.g., Zhao, Liu et al., 2013; Knapp et al., 2018; Mesny, Miyauchi et al., 2021), and, although present in lower numbers in mycorrhizal fungi (Kohler et al., 2015; Peter et al., 2016; Miyauchi, Kiss et al., 2020), certain CAZymes play key roles in the establishment and maintenance of the symbiosis (Veneault-Fourrey, Commun et al., 2014; Doré et al., 2017; Marqués-Gálvez et al., 2021). Comparing CSEP and CAZyme repertoires is therefore highly relevant to exploring genetic differences in plant associated lifestyles of fusarioid fungi.

Here, we performed whole genome sequencing (WGS), assembly and structural annotation of five novel endophytic *Fusarium* strains (Table 3.1) including the first WGS data and annotated assemblies for the recently described species, *Fusarium chuoi* (see Appendix A.2; Crous, Osieck et al., 2021). Using predicted genes from these and other publicly available fusarioid strains, we produced a genome-scale phylogeny of *Fusarium* and allied genera with time calibration and compared CSEP and CAZyme content to answer the following questions: 1) How are lifestyles distributed across the phylogeny? 2) Can we distinguish plant pathogens and endophytes from genome sequences alone? and 3) How is selection acting on different lifestyles?

3.3 Materials and methods

Whole genome sequencing

We selected five endophytic *Fusarium* strains for WGS which were representatives of species hypotheses that had previously been isolated and clustered into 99% similarity operational taxonomic units in Chapter 2, with taxonomic identification confirmed where possible via morphological assessment by the Westerdijk Institute (Table 3.1). For DNA extractions, a fragment of mycelium from axenic cultures was transferred to 500 ml of 2% malt extract nutrient broth using a sterile needle and grown at 25°C in ambient light conditions on an orbital shaker at 120 rpm for ∼1 week. Mycelia

were collected via vacuum filtration and frozen at -80°C before being pulverised with two sterile stainless-steel beads in a 2 ml Eppendorf using a Mixer Mill MM 400 (Retsch, Germany).

DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, CA, USA) according to the manufacturer's protocol and eluted in 70 µl of TE buffer. Sufficient DNA concentration (more than 20 ng/µl) was confirmed with a Quantus™ Fluorometer (Promega, WI, USA) and purity (260/280 absorbance ratio of approximately 1.8) confirmed with a NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, USA). DNA extractions were sent to Macrogen (Macrogen Inc., South Korea) for library preparation and sequencing: library preparation was performed using the TruSeq DNA PCR-free Sample Preparation Kit with a 550 bp insert size and 151 bp paired-end reads were sequenced using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA).

Comparison of *de novo* assembly tools

Our bioinformatics analysis pipeline is summarised in Supplementary Figure S3.1. Reads were trimmed using Trimmomatic v0.36 (Bolger, Lohse and Usadel, 2014) and quality checked using FastQC v0.11.5 (Andrews, 2018). The performance of three *de novo* assembly tools was compared – ABySS v2.0.2 (Simpson et al., 2009), MEGAHIT v1.2.9 (Li, Luo et al., 2016) and SPAdes v3.11.1 (Bankevich et al., 2012). In the case of ABySS, which requires the user to specify k-mer size, multiple assemblies were run with varying k-mer sizes to converge on an optimal k-mer size according to N50, which was calculated with the abyss-fac function. Alternatively, both MEGAHIT and SPAdes use a multiple k-mer sizes strategy. For MEGAHIT assemblies were run with k-mer sizes from 51 to 131 in steps of 8, while for SPAdes the default recommended k-mer sizes for the read length were used: 21, 33, 55, 77. Trimmed reads were mapped back onto contigs using BWA-MEM v0.7.17-r1188 (Li, 2013) and the resulting BAM files then used for polishing with Pilon v1.23 (Walker et al., 2014), which helps to corrects misassemblies and gaps. The flagstat option from SAMtools v1.9 (Li, Handsaker et al., 2009) was used to produce read mapping statistics from the BAM files in order to calculate sequencing coverage. Contigs shorter than 200 bp were removed using seqtk v1.2-r94 (`https://github.com/lh3/seqtk`) for compliance with NCBI assembly standards.

The 'best' assembly was chosen by assessing contiguity via QUAST v5.0.2 (Gurevich et al., 2013) and completeness as measured by gene sets via BUSCO v3.0.1 (Simão et al., 2015) using the hypocreales_odb10.2019-11-20 lineage dataset of 4,494 single-copy orthologues. The difference in assembly completeness between the three different assembly tools that were tested – ABySS, MEGAHIT and SPAdes – was generally minimal, with single-copy BUSCOs 10 differing by no more than 0.11% across tools for each strain (Supplementary Table S3.1), which can largely be attributed to the high sequencing coverage. QUAST Nx plots, which show the smallest contig length at which x% of the assembly is contained in contigs of at least that size, found that ABySS produced assemblies with the best contiguity in four out of five cases (Supplementary Figure S3.2). Although no single tool produced the highest completeness or best contiguity statistics for all strains, ABySS was selected as the best-performing tool on-average for consistency's sake during later biological comparison across strains. Finally, BlobTools v1.1 (Laetsch and Blaxter, 2017) was used to screen the selected ABySS assemblies and confirm the absence of contaminants using the BAM file of mapped reads and a blastn hit file of assemblies against the NCBI nucleotide database created with BLAST 2.7.1+ (Camacho et al., 2009) (Supplementary Figure S3.3). Mitochondrial contaminations flagged by NCBI during the assembly submission process were trimmed using bedtools v2.28.0 (Quinlan and Hall, 2010).

Table 3.1: Voucher and collection information for the *Fusarium* strains selected for WGS and assembly. ^{T} = ex-type material.

Name	Species hypothesis (Figure 2.5)	Voucher	Species complex	Collection location	Host
Fusarium chuoi	RH1	CBS 148465 836515-16	FFSC	Vietnam, Nghệ An Province, Con Cuông District, Châu Khê commune N19°1'48.73"E104°43'31.97"	*Musa itinerans* (seed)
F. chuoi	RH3	CBS 148464^{T} 836445-12-1	FFSC	Vietnam, Hà Tĩnh Province, Hương Sơn District, Sơn Kim commune N18°25'37.38"E105°12'53.95"	*Musa itinerans* (seed)
F. annulatum	RH5	880149-04	FFSC	Vietnam, Lào Cai Province, Sa Pa, Hoàng Liên National Park N22°15'10.2"E103°56'35.5"	*Musa itinerans* (seed)
F. sp.	RH6	836490-20	FIESC	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'14.89"E105°14'50.83"	*Musa itinerans* (seed)
F. proliferatum	RH7	836489-13	FFSC	Vietnam, Nghệ An Province, Thanh Chương District, Thanh Thủy commune N18°38'29.8"E105°14'15.87"	*Musa balbisiana* (seed)

Structural annotation

A *de novo* repeat library was generated for the ABySS assembly for each strain with RepeatModeler v2.0.1 (Smit and Hubley, 2015) and used as a custom library for softmasking with RepeatMasker v4.0.9 (Smit, Hubley and Green, 2015). Masked assemblies were annotated following the MAKER pipeline (Cantarel, Korf et al., 2008) using proteins and EST clusters downloaded from MycoCosm (`https://mycocosm.jgi.doe.gov/`; Grigoriev et al., 2014; Mesny, Miyauchi et al., 2021) to inform gene prediction: from *F. oxysporum* MPI-SDFR-AT-0094 (Fusoxy1) for *F. chuoi* RH1, *F. chuoi* RH3, *F. annulatum* RH5 and *F. proliferatum* RH7 belonging to the *Fusarium fujikuroi* species complex (FFSC); and *F. equiseti* MPI-CAGE-AA-0113 (Fuseq1) for *F.* sp. RH6 belonging to the *Fusarium incarnatum-equiseti* species complex (FIESC). The first evidence-based round of MAKER was used to train SNAP v2006-07-28 (Korf, 2004) and the resulting parameters input into a second *ab initio* MAKER round alongside the AUGUSTUS v3.2.3 (Stanke et al., 2006) pre-trained parameter set for *Fusarium*. A second iteration of SNAP training and *ab initio* prediction was then performed. Misannotations in the form of artefactually fused genes that were flagged by NCBI during the assembly submission process were checked against existing annotations in NCBI's Genome Data Viewer (Rangwala et al., 2021), and then manually edited. For compliance with NCBI standards, Genome Annotation Generator v2.0.1 (Geib et al., 2018) was used with the options -ris 10, –fix_terminal_ns and –fix_start_stop to remove introns shorter than 10 bp, remove terminal strings of Ns and ensure start and stop codons were correctly annotated. See Supplementary Table S3.1 for a summary of assembly quality statistics.

Phylogenomic analyses

Predicted genes from 57 additional publicly available strains of *Fusarium* and allied genera were downloaded from NCBI (Appendix A.3) and orthogroups (referred to here as genes) were inferred from amino acid sequences of the total 62 strains using OrthoFinder v2.4.0 (Emms and Kelly, 2019). We aligned 1,060 core (i.e., shared between all fusarioid taxa including the outgroup) single-copy genes using MAFFT v7.310 with default settings (Katoh and Standley, 2013) and removed ambiguously aligned regions using both BMGE v1.12 (Criscuolo and Gribaldo, 2010) and trimAl v1.4.rev15 with the gappyout option (Capella-Gutiérrez, Silla-Martínez and Gabaldón, 2009) to compare the impact of trimming tools on the resulting species trees.

For a concatenation-based approach, core single-copy gene alignments were concatenated with AMAS v0.98 (Borowiec, 2016). We compared two tools for maximum likelihood (ML) species tree estimation: IQ-TREE v2.1.2 (Minh et al., 2020) and RAxML-NG v1.0.1 (Kozlov et al., 2019), with the concatenated alignment partitioned by gene in both cases. For IQ-TREE, the best-fit amino acid substitution model for each partitioned gene was selected by the inbuilt tool ModelFinder (Kalyaanamoorthy et al., 2017) using Bayesian information criterion values, and branch support was computed via 1,000 ultrafast bootstrap replicates (Hoang et al., 2018). For RAxML-NG, ModelTest-NG v0.1.6 (Darriba et al., 2020) was used to select substitution models for each gene using Akaike information criterion values, and branch support was computed via 100 Felsenstein's bootstrap replicates. Bootstrap convergence was confirmed with the –bsconverge option using the default 3% cutoff for weighted Robinson-Foulds distances (Pattengale et al., 2009).

For a coalescent-based approach, ML gene trees were inferred from each core single-copy gene alignment with RAxML-NG using the best-fit model selected by ModelTest-NG during the concatenated

analysis. The resulting ML gene trees were used for coalescent-based species tree reconstruction using ASTRAL-III v5.7.3 (Zhang, Rabiee et al., 2018) with local posterior probability branch support estimation (Sayyari and Mirarab, 2016). ASTRAL-Pro v1.2 (Zhang, Scornavacca et al., 2020) was additionally run with local posterior probability support estimation on the 20,343 gene trees produced by OrthoFinder, which represented both single- and multi-copy 'total' genes. OrthoFinder itself also produces a coalescent-based species tree topology by default using STAG (Emms and Kelly, 2018), which used 3,449 core single- and multi-copy genes. All species tree topologies were compared by computing the normalised Robinson-Foulds metric using the RF.dist function from the phangorn v2.7.0 package (Schliep et al., 2017) in R v4.0.4 (R Core Team, 2020).

Molecular clock analyses

The species tree topology inferred by RAxML-NG was used to perform molecular clock analyses with MCMCTree (Yang and Rannala, 2006) in PAML v4.9 (Yang, 2007) using the top 10 'clock-like' core single-copy genes, as inferred by SortaDate based on root-to-tip variance (Smith, Brown and Walker, 2018). Divergence times were estimated using the approximate likelihood method (dos Reis and Yang, 2011) with the WAG amino acid substitution model (Whelan and Goldman, 2001). Due to the sparse fossil record for the fungi at large, a previous fossil-calibrated study of the kingdom including *Fusarium* species was used to inform secondary calibrations of the tree root at 0.9–1.35 (1 time unit being 100 My) and the node representing the split between *F. graminearum* and *'F.' solani* at 0.5–0.9 (Lutzoni et al., 2018).

We used uniform node age priors by setting both the birth rate (λ) and death rate (μ) to 1 and the sampling fraction (ρ) to 0. For the substitution rate (r) prior, the shape parameter (α) was set to 1, and the scaling parameter (β) was estimated as 4.5 using the following equation: $\beta = (\alpha \times \text{root-time}) / \text{tip-to-root}$, where the mean tip-to-root distance was calculated as 0.22 for both ML species trees using the distRoot function from the package adephylo v1.1-11 (Jombart, Balloux and Dray, 2010) in R, and the root-time being approximately 1 MY as described above. We set the rate drift (σ^2) prior parameters of α and β to 1 and 10, respectively. Two MCMC chains were run for both the independent-rates (IR) and autocorrelated-rates (AR) relaxed clock models, with 20,000 generations, posterior sampling every 10 generations and a 10% burnin per chain. Chain convergence was confirmed by plotting the posterior mean times for both chains for each clock model, and infinite-sites plots were made to confirm that sufficient molecular data was used (Supplementary Figure S3.4).

Computational prediction of CSEPs and CAZymes

CSEPs were identified from predicted genes using a framework inspired by Beckerson et al. (2019) and summarised in Figure 3.1A, including the following steps:

1. The putative secretome was identified via prediction of signal peptides with SignalP v5.0b (Almagro Armenteros, Tsirigos et al., 2019). Signal peptide prediction was additionally cross-checked against TargetP v2.0 (Almagro Armenteros, Salvatore et al., 2019) and Phobius v1.01 (Käll, Krogh and Sonnhammer, 2004).
2. Genes were removed if their predicted cellular localisation contradicted secretion, indicated by:
   (a) >1 transmembrane domain according to TMHMM v2.0c (Krogh et al., 2001) and Phobius;

(b) endoplasmic reticulum retention according to ps_scan v1.86 (de Castro et al., 2006);

(c) nuclear localisation according to NucPred v1.1 (Brameier, Krings and MacCallum, 2007);

(d) GPI-anchored according to PredGPI (Pierleoni, Martelli and Casadio, 2008), accessed using the R package ragp (Dragićević et al., 2020).

3. The remaining genes were cross checked with machine learning-based effector prediction using EffectorP 3.0 (Sperschneider and Dodds, 2021). CSEPs were predominantly less than 300 amino acids in length (Figure 3.1B), an oft-quoted cut-off for small secreted proteins in fungi (Lo Presti et al., 2015).

A custom bash script, CSEPfilter, was written to perform the filtering of gene sets at each stage. To match CSEPs to experimentally verified genes, sequences were searched against the PHI-base database (downloaded 09/02/2022; Urban et al., 2020) using a blastp search with an e-value of 1e-25 from BLAST. For CSEPs with multiple successful hits, the hit with the top bitscore was used.

CAZymes were identified from predicted genes using run_dbCAN v3.0.2 (`https://github.com/linnabrown/run_dbcan`) from the dbCAN2 CAZyme annotation server (Zhang, Yohe et al., 2018). This process involves:

1. HMMER v3.3.2 (Mistry, Finn et al., 2013) search against the dbCAN HMM (hidden Markov model) database;
2. DIAMOND v2.0.14 (Buchfink, Reuter and Drost, 2021) search against the CAZy pre-annotated CAZyme sequence database (Drula et al., 2022);
3. eCAMI (Xu et al., 2020) search against a CAZyme short peptide library for classification and motif identification.

Only genes which were predicted to be a CAZyme by all three methods were classified as such. Accepted names were retrieved via an automated search of Enzyme Commission (EC) numbers against the ExplorEnz website (McDonald, Boyce and Tipton, 2009) and webscraping of the results using the R package rvest v1.0.2 (Wickham, 2020). CAZyme families known to act on the major plant cell wall substrates of cellulose, cutin, hemicellulose, lignin and pectin were classified from the literature (Glass et al., 2013; Levasseur et al., 2013; Zhao, Liu et al., 2013; Miyauchi, Kiss et al., 2020; Hage and Rosso, 2021; Mesny, Miyauchi et al., 2021).

CSEPs and CAZymes were matched to gene orthogroups with a custom R script, orthogroup_parser.r, where a gene was defined as a CSEP/CAZyme if it was predicted to be so in at least one taxon. We checked that genome assembly quality did not significantly influence the number of predicted genes, CSEPs or CAZymes by confirming that there was no correlation between assembly N50 (extracted from NCBI metadata for assemblies produced outside this study) and number of genes/CSEPs/CAZymes using the cor.test function in R.

Comparative genomics of lifestyle

Lifestyles of all the strains used in this study were inferred from the host/substrate and other relevant data (such as pathogenicity tests) sourced from the literature, NCBI BioSample metadata and online culture collection metadata (Appendix A.3). If a strain was reported from a plant host but without sufficient clarification of whether the plant was exhibiting disease symptoms or the fungus was isolated from inside plant tissues, the strain was classified ambiguously as a ‘plant associate’.

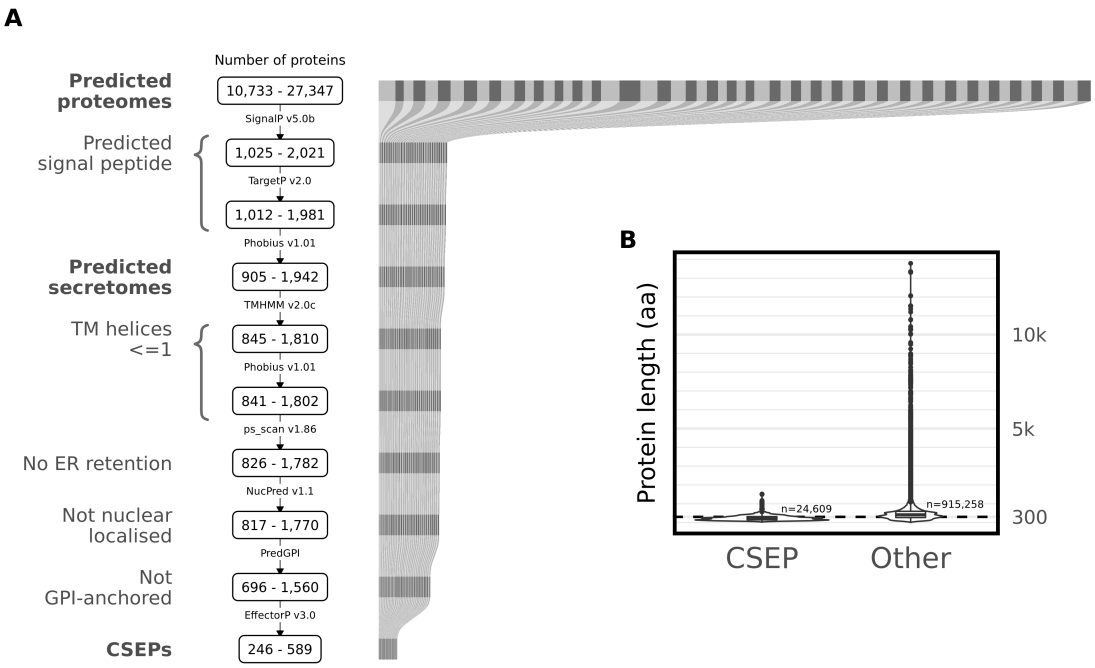


Figure 3.1: **(A)** Alluvial plot indicating the number of proteins retained at each step of the computational CSEP prediction procedure, with different taxa indicated by alternating coloured boxes. The range of number of proteins across all taxa at each step is shown to the left of boxes. **(B)** The length of CSEPs following the prediction steps in comparison with all other proteins.

In addition to the lifestyle of the specific strains used in the analyses, other lifestyle reports were collected from the literature with the help of the PlutoF platform (Abarenkov, Tedersoo et al., 2010) in order to show the range of reported lifestyles for taxa.

The impact of strain lifestyle on CSEP, CAZyme and all gene content was explored using an approach developed by Mesny and Vannier (2020) which accounts for confounding phylogenetic signal. This included principal component analysis (PCA) of phylogenetic distances from the dated species tree using the PCA function from the Python package scikit-learn v0.23.2 50 – run in Python v3.7.9 (`https://www.python.org/`) – which was used for global permutational multivariate analysis of variance (PERMANOVA) of gene/CSEP/CAZyme content with the adonis2 function from the R package vegan v2.5-7 51 using the model 'JaccardDistMatrix $\sim$ PC1 + PC2 + Lifestyle'. Two principal components were deemed sufficient to represent phylogenetic signal as together they explained over 90% of the variance (see Supplementary Figure S3.5 for PCA plots and variance explained for the first 6 principal components). Pairwise PERMANOVA was then performed using the pairwise.perm.manova function from RVAideMemoire v0.9-78 (Hervé, 2020) with Bonferroni multiple test correction.

For statistical analyses to test the difference in number of genes; number of strain-specific genes; and mean gene copy number between lifestyles, we first assessed the assumption of normality by making Q-Q plots using the ggqqplot function from ggpubr v0.4.0 (Kassambara, 2020) to ascertain approximate normality of residuals. We then assessed the assumption of homogeneity of variance using the levene_test function from the package rstatix v0.7.0 (Kassambara, 2021), where a significant p value ($p < 0.05$) means that the assumption is violated.

If we could assume homogeneity of variance (i.e., Levene's $p \geq 0.05$), we used the rstatix function an-

ova_test to compute analysis of variance (ANOVA) using the model 'value ~ PC1 + PC2 + lifestyle', as with the PERMANOVA, to once again account for phylogeny. If the ANOVA was significant ($p < 0.05$), a multiple comparison test between lifestyles was performed with the tukey_hsd rstatix function using the model 'value ~ lifestyle'. If we could not assume homogeneity of variance (i.e., Levene's $p < 0.05$), we used an aligned rank transform (ART) ANOVA with the aligned.rank.transform function from the ART v1.0 (Villacorta, 2015) R package, again using the model 'value ~ PC1 + PC2 + lifestyle'. If the ART ANOVA was significant, the games_howell_test rstatix function was used for multiple comparison testing using the formula 'value ~ lifestyle', as is recommended for multiple comparisons when classical ANOVA assumptions are violated (Sauder and DeMars, 2019).

Selection analyses

To assess whether core single-copy genes have evolved under positive selection we used HyPhy v2.5.30 (Kosakovsky Pond, Frost and Muse, 2005), which offers a suite of tools for assessing selective pressures based on dN/dS – that is, the ratio of nucleotide substitutions which alter the transcribed amino acid to those that do not. Notably, this approach assumes that synonymous substitutions are selectively neutral.

To produce the codon alignments necessary to calculate dN/dS, nucleotide sequences corresponding to the 1,060 core single-copy genes used in the phylogenomic analysis were retrieved from MAKER outputs and – for previously published taxa – GBFF files, using a custom Python script, pull_nucleotides.py. Occasionally, the corresponding nucleotide sequences were the incorrect length, and in these cases they were manually cross-checked with amino acid sequences and trimmed in AliView v1.25 (Larsson, 2014). Nucleotide sequences were then used to convert amino acid alignments into codon alignments using PAL2NAL v14.0 (Suyama, Torrents and Bork, 2006) with the -nogap option. Six genes were filtered out due to alignments not having sufficient gapless sites, leaving 1,054 core single-copy genes for selection analyses.

Codon alignments and ML gene trees were run in BUSTED v3.1 (Murrell et al., 2015) to detect gene-wide episodic positive selection ($dN/dS > 1$). To then identify specific lineages under episodic positive selection for each gene, codon alignments were run with the ML species tree in aBSREL v2.2 (Smith, Wertheim et al., 2015), which employs Holm-Bonferroni multiple testing p value correction. For both methods, all ingroup lineages were selected as foreground branches for testing. The significant difference in number of genes undergoing positive selection on external branches between different lifestyles was statistically tested as described above, the ANOVA model being 'value ~ PC1 + PC2 + lifestyle'.

To assess whether the inferred positive selection of core CSEPs on external branches could be associated with lifestyle, we used Contrast-FEL to compare differences in relative selective pressures between lifestyles (Kosakovsky Pond, Wisotsky et al., 2021). This method finds site-level differences in dN/dS between two sets of branches, and so for each gene tree we labelled branches associated with each lifestyle in turn as the 'test' set and all other lifestyles as the 'background' set. Only external branches were labelled, as we cannot definitively know the lifestyle of common ancestors associated with internodes in the tree. The labelled trees were then run with codon alignments in Contrast-FEL to calculate sites with higher or lower selective pressure in the test set relative to the background set. We used the most conservative statistic reported by Contrast-FEL to determine differences in selective pressures, the multiple testing corrected q value ($p \geq 0.05$, false discovery rate ≥ 0.2), which has the highest precision but lowest recall (Kosakovsky Pond, Wisotsky et al.,

2021). The significant difference in number of sites with a higher or lower relative selective pressure between lifestyles was statistically tested using the same process as described above, using the model 'sites ∼ lifestyle'.

Codon bias analyses

In order to explore the level of translational selection on synonymous substitutions (i.e., bias towards certain codons in more highly expressed genes), we also quantified codon optimisation of all core single-copy genes to the ribosomal protein gene pool (S) with the get.s function from the tAI v0.2 package (dos Reis, Savva and Wernisch, 2004). This first required calculation of the codon adaptation index (CAI; Sharp and Li, 1987), which compares codon usage in a given gene to a reference set of highly expressed genes. For our reference set, known ribosomal protein genes were extracted from the functionally annotated protein set of *F. graminearum* PH1 (Cuomo et al., 2007) downloaded from Mycocosm (Grigoriev et al., 2014) with the getfasta tool from bedtools v2.28.0 (Quinlan and Hall, 2010). These ribosomal protein genes were used as input for a blastp search against the predicted genes of all taxa used in this study using BLAST 2.7.1+, and then matched to core single-copy genes with a custom R script, codon_optimisation.r. A gene was defined as encoding a ribosomal protein if it had a blast hit in at least 1 taxon. The codonTable function from the coRdon v1.1.3 R package (Elek, Kuzman and Vlahovicek, 2021) was used to produce a table of codon counts for each core single-copy gene for each taxon, from which CAI was calculated in reference to the identified ribosomal protein genes using the CAI function from coRdon. The get.s function also requires the effective number of codons (Nc), which was calculated from the codon count table using the ENC function from coRdon, and GC content at the third codon position (GC3), which was calculated using the GC function from seqinr v4.2-8 (Charif and Lobry, 2007). S values were calculated for CSEP, CAZyme, non-CSEP/CAZyme and all core single-copy genes in turn. The significant difference in S values between lifestyles and between gene types (i.e., CSEP, CAZyme or other) for each lifestyle was statistically tested using the same process as described above, the ANOVA/ART ANOVA model being 'S ∼ PC1 + PC2 + lifestyle' and TukeyHSD/Games Howell test model being 'S ∼ lifestyle'.

To assess the relationship between S values and the number of reported lifestyles or 'lifestyle range' of taxa, we calculated Pearson's correlation on uncorrected data using the cor.test function in R, and used phylogenetic generalised least squares (PGLS) regression to assess correlation while correcting for phylogenetic signal in the data with the R package nlme v 3.1-152 (Pinheiro et al., 2021). For PGLS, which specifies that trait covariance between pairs of taxa decreases with time since divergence, we tested Brownian, Pagel and Blomberg phylogenetic correlation structures for the dated species tree, implemented in the R package ape v5.6-1 (Paradis and Schliep, 2019), and selected Brownian as the best model fit based on Akaike information criterion values. For number of reported lifestyles, only taxa identified to species level were included, and for species with multiple representative strains the mean S value was used. To visualise the relationship between S values and phylogeny, we used the ordisurf function from the R package vegan v2.5-7 (Oksanen et al., 2019) to fit S values to the PCA of phylogenetic distances produced in comparative analyses above (recreated in R with the vegan prcomp function). The significant difference in overall S values between *Fusarium s. str.* and allied genera was tested using the t.test function in R (having confirmed normality of residuals and homogeneity of variance with a Q-Q plot and Levene's test as above).

The uco function from seqinr v4.2-8 was used to calculate codon usage bias in terms of relative synonymous codon usage (RSCU) – the ratio of observed codon usage to expected codon usage – for

all codons across each taxon, excluding non-redundant codons encoding methionine and tryptophan and stop codons. RSCU values were then normalised using the scale function and used to produce a Euclidean distance matrix with the dist function, which was used for hierarchical clustering of taxa with the hclust function using the average agglomeration method. We compared the topology produced by hierarchical clustering with the RAxML-NG species tree topology by again computing the normalised Robinson-Foulds metric using the RF.dist function from phangorn. We calculated the p value by computing the metric for 1,000 random trees with the same number of taxa against the species tree topology to determine the number of simulations for which the metric was lower (i.e., topologically closer) than that from the hierarchical clustering.

All results were plotted in R v4.0.4 using the following packages: ape v5.6-1 (Paradis and Schliep, 2019), cowplot v1.1.1 (Wilke, 2020), deeptime v0.0.6.0 (Gearty, 2021), dendextend v1.15.2 (Galili, 2015), dplyr v1.0.6 (Wickham and Seidel, 2020), eulerr v6.1.0 (Larsson, 2020), ggplot2 v3.3.3 (Wickham, 2016), ggalluvial v0.12.3 (Brunson, 2020), ggforce v0.3.2.9000 (Pedersen, 2021), ggnewscale v0.4.6 (Campitelli, 2020), ggplotify v0.0.7 (Yu, 2021), ggpubr v0.4.0 (Kassambara, 2020), ggrepel v0.9.1 (Slowikowski, 2020), ggthemes v4.2.4 (Arnold, 2021), ggtree v2.4.2 (Yu et al., 2017), jsonlite v1.7.2 (Ooms, 2014), matrixStats v0.61.0 (Bengtsson, 2021), MCMCtreeR v1.1 (Puttick and Title, 2019), metR v0.9.2 (Campitelli, 2021), multcompView v0.1-8 (Graves et al., 2019), pBrackets v1.0.1 (Schulz, 2021), phytools v0.7-80 (Revell, 2012), plyr v1.8.6 (Wickham, 2011), reshape2 v1.4.4 (Wickham, 2007), scales v1.1.1 (Wickham and Seidel, 2020), stringi v1.6.2 (Gagolewski and Tartanus, 2021), stringr v1.4.0 (Wickham, 2019) and tidyr v1.1.3 (Wickham and Girlich, 2022). R scripts were written using RStudio v1.3.1093 (RStudio Team, 2015). This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (Butcher, King and Zalewski, 2017). Scripts of all analyses are available at `https://github.com/Rowena-h/FusariumLifestyles`.

3.4 Results

Both single- and multi-copy genes inferred the same backbone for *Fusarium s. str.*

To infer the genome-scale phylogeny of *Fusarium*, we used both concatenation and coalescent-based approaches, using single-copy genes with and without multi-copy genes also included. Including multi-copy genes had a greater impact on topology than tree building approach (i.e., concatenation versus coalescent) (Figure 3.2A). This was seen chiefly from a change in divergence order of allied genera – *Neocosmospora* (=*Fusarium solani* species complex (FSSC)), *Geejayessia* (=*Fusarium staphylae* species complex (FSTSC)) and *Albonectria* (=*Fusarium decemcellulare* species complex (FDESC)) – when including multi-copy genes (Figure 3.2B). All methods, however, produced the same divergence order for *Fusarium s. str* species concepts. Disregarding differences in the naming of species, our estimations of *Fusarium s. str* from 1,060 loci were in broad agreement with the most recent phylogenetic analyses by Crous, Lombard et al. (2021) and Geiser, Al-Hatmi et al. (2021).

We additionally compared the impact of alignment trimming tools – trimAl versus BMGE – on species tree topology. The RAxML-NG species tree was identical for both trimming tools, but trimming tool impacted topology for IQ-TREE and ASTRAL-III, with discordance in the ambrosia clade of *Neocosmospora* (Supplementary Figure S3.6). The gene trees trimmed with trimAl were selected for downstream analyses based on its reported accuracy relative to BMGE in the literature

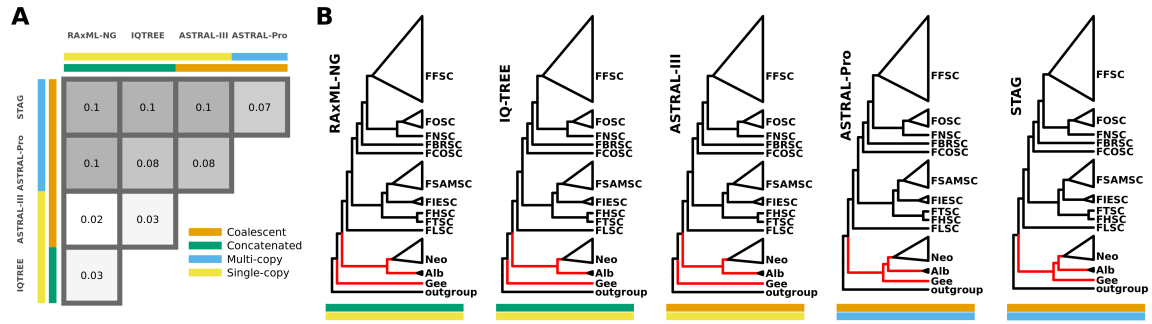


Figure 3.2: (A) Pairwise comparison of normalised Robinson-Foulds distances between topologies from all species tree estimation methods, with grid cells coloured from most similar topology (lighter) to most dissimilar (darker). **(B)** Summary of species trees with red branches indicating topological discordance between methods. Labels indicate *Fusarium* species complex (see Abbreviations) or allied genus (Neo=*Neocosmospora*, Alb=*Albonectria*, Gee=*Geejayessia*).

(Tan et al., 2015; Steenwyk, Buida et al., 2020). The RAxML-NG species tree was selected for downstream analyses as its topology was identical for both trimming tools while having branch length units as substitutions per site as opposed to coalescent units.

Dated genome-scale phylogeny of *Fusarium* and allied genera

For divergence time estimation of the RAxML-NG species tree, we used both the IR and AR relaxed clock models, implemented in MCMCTree. Testing best-fit of clock models in MCMCTree (see dos Reis, Gunnell et al., 2018) is not possible using amino acid data, and so our assessment of divergence time estimation from the two clock models was restricted to comparisons against previous studies. The IR model generally shifted nodes towards more recent divergence times in comparison to the AR model (Supplementary Figure S3.7). The crown age of *Fusarium s. lat.* was estimated to fall in the late Cretaceous by both the IR (71 Ma) and AR (84 Ma) models, although the latter was closer to the estimate by O'Donnell, Rooney et al. (2013) (83 Ma). The crown age of *Fusarium s. str.* estimated in the Eocene (49 Ma) by the same study was much closer to our result from the IR model (51 Ma) compared with the AR model (69 Ma, late Cretaceous). The middle Miocene crown age of the ambrosia clade in *Neocosmospora* from previous estimates by Kasson et al. (2013) (13 Ma) and O'Donnell, Sink et al. (2015) (9 Ma) were also in closer agreement with the IR model (7 Ma) compared with the AR model (25 Ma). The crown age of *Xyleborini* beetle hosts estimated by Jordal and Cognato (2012) (21 Ma) corresponded more closely with the IR estimate of the divergence of the ambrosia clade from non-insect mutualists (15 Ma) compared with the AR estimate (41 Ma). The dating of the diversification of various *formae speciales* in the FOSC by our IR model was also a better fit with their crop hosts having been domesticated within the last ~10,000 years (Meyer, Duval and Jensen, 2012).

Gene, CSEP and CAZyme repertoires were broadly shared across lifestyles, but plant pathogens included copy number outliers

There was no significant difference in number of genes, CSEPs or CAZymes across lifestyles (Supplementary Table S3.2). Most genes, CSEPs and CAZymes were either core (present in all fusarioid taxa) or accessory (present in more than one taxon but not all), with very few being strain-specific, indeed strain-specific CAZymes being almost non-existent (Figure 3.3A). The number of strain-specific

genes or CSEPs was not significantly different across lifestyles (Supplementary Figure S3.8A, Supplementary Table S3.2). Global PERMANOVA showed that gene, CSEP and CAZyme content were better described by phylogenetic relatedness (35–42% variance) than lifestyle (9% variance) (Figure 3.3B, Supplementary Table S3.3). Nonetheless, pairwise PERMANOVA identified the insect mutualist lifestyle as the most genetically distinct, with insect mutualist taxa having significantly different gene, CSEP and CAZyme repertoires compared with all other lifestyles other than mycoparasite. While most other lifestyles were genetically similar, endophytes and saprotrophs were also found to be significantly different in terms of CSEPs. In a similar pattern to the number of strain-specific genes, mean gene, CSEP and CAZyme copy number were not found to be significantly different between lifestyles (Supplementary Figure S3.8B, Supplementary Table S3.2), but there were extreme outliers in copy number amongst plant pathogens (Figure 3.3C). The greatest copy number outlier by a considerable margin was predicted to be both a CSEP and CAZyme belonging to *F. oxysporum* f. sp. *conglutinans*, annotated as a glycosyltransferase in the GT4 family: α,α-trehalose phosphorylase (configuration-retaining) (EC 2.4.1.231).

Almost half of core single-copy genes were under positive selection

While gene, CSEP and CAZyme repertoires may have been broadly shared, we were interested in whether genes were evolving in a lifestyle-directed manner. Of the 1,054 core single-copy genes used in the selection analyses, 469 (44%) were found to be under episodic positive selection by both BUSTED and aBSREL (Figure 3.4A). This included 11 of 31 (35%) core CSEPs and 6 of 11 (55%) core CAZymes. The branch at the root of the more conservative generic concept, *Fusarium s. str.*, was a particular 'hotspot' of positive selection, with 52 core single-copy genes positively selected according to BUSTED and aBSREL (Supplementary Figure S3.9). A few external branches also had a notably high number of positively selected core genes: insect mutualist *N. oligoseptata*; saprotrophic *F. culmorum* in the *Fusarium sambucinum* species complex (FSAMSC); and plant pathogenic *F. oxysporum* f. sp. *lycopersici* in the FOSC. There was no significant difference in the number of positively selected genes on external branches between lifestyles according to analysis of variance (ANOVA, p=0.7; Supplementary Table S3.2).

Although a minority of all CSEPs (11%) could be assigned known gene names using the PHI-base database, two core CSEPs with signatures of selection could be classified as known genes: 5680 as FGSG_00806 and 6786 as FgPR-IL-2 (Figure 3.4A). Based on PHI-base records of gene knockouts in *F. graminearum* inoculated on wheat, both FGSG_00806 and FGPR-IL-2 had the mutant phenotype of unaffected pathogenicity (Supplementary Figure S3.10). Of the six core CAZymes which had undergone positive selection, four are known to act on plant cell wall substrates (Supplementary Figure S3.11): glycoside hydrolase GH35 (β-galactosidase) on hemicellulose and pectin and GH51 (non-reducing end α-l-arabinofuranosidase) on cellulose, hemicellulose and pectin; carbohydrate esterase CE12 (rhamnogalacturonan acetylesterase) on pectin; and an enzyme of auxiliary activities AA3_2 (5'-oxoaverantin cyclase) on lignin.

Most CSEPs and CAZymes reported as positively selected by both BUSTED and aBSREL were also found to contain sites with a higher relative selective pressure in certain lifestyles by Contrast-FEL (Figure 3.4B). In most cases only one site per gene was found to have a difference in relative selective pressure. The insect mutualist lifestyle had significantly more sites per gene under higher selective pressure compared with most other lifestyles (Figure 3.4B). We should emphasise that Contrast-FEL does not inform whether positive or negative selection is occurring on a branch set, only that there

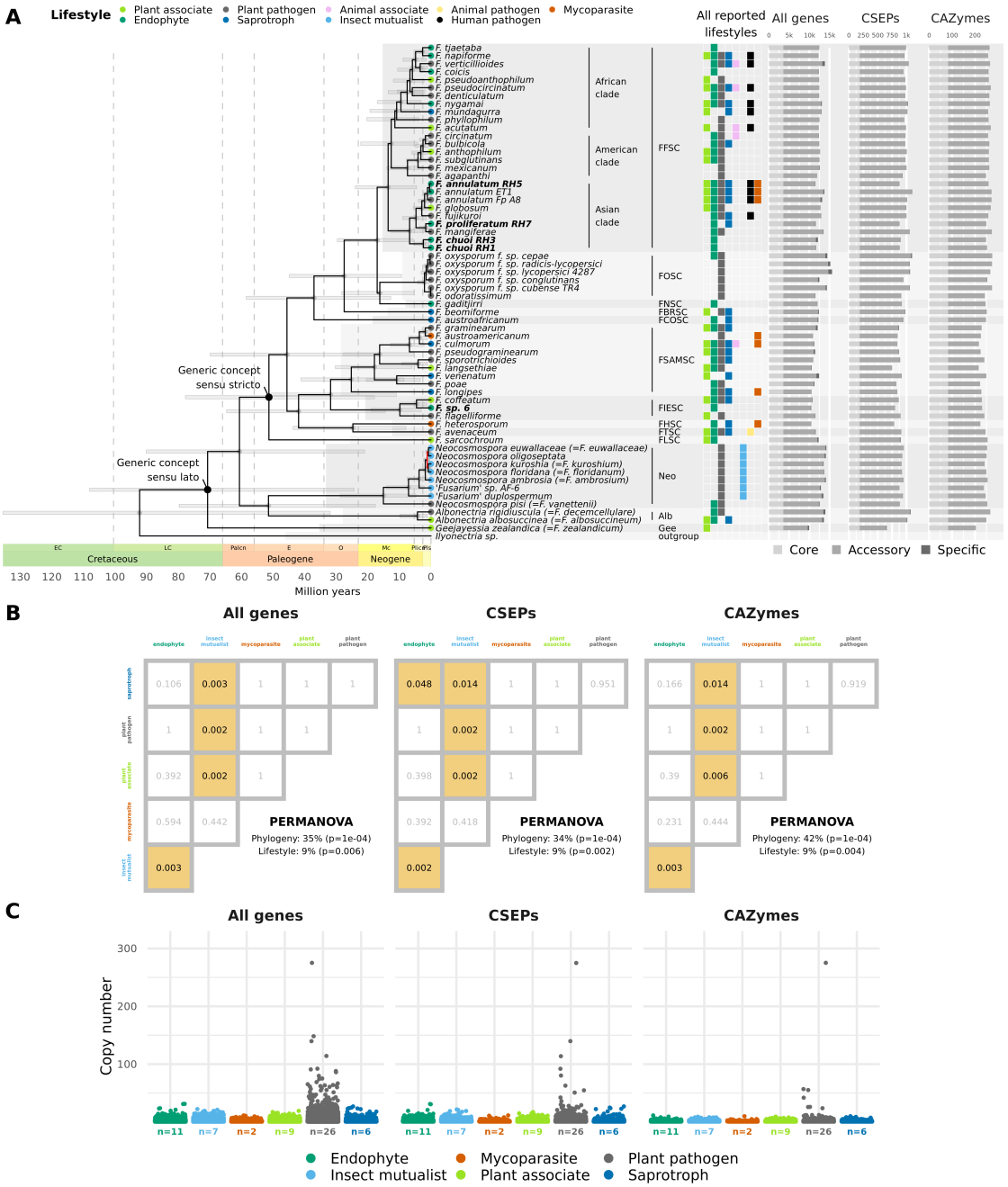


Figure 3.3: **(A)** Genome-scale phylogeny of fusarioid taxa produced by RAxML-NG from 1,060 core single-copy genes. All branches were significantly supported (≥70 Felsenstein's bootstrap replicates), except those in red. A time scale for node ages estimated by the IR relaxed clock model is shown below the phylogeny, with highest posterior density 95% confidence intervals shown as bars on nodes. For the AR model results and the exact ages and confidence intervals estimated for every node, see Supplementary Figure S3.7. Clades corresponding to species complexes (see Abbreviations) and allied genera are highlighted with alternating boxes and annotated to the right of taxon names (Neo=*Neocosmospora*, Alb=*Albonectria*, Gee=*Geejayessia*). Lifestyles of the strains used in this study are indicated by coloured circles on tips, with other lifestyles reported from the literature summarised in the central grid (see Appendix A.3 for references). Bar graphs on the right indicate the number of genes, CSEPs and CAZymes for each taxon, with lightest to darkest colour indicating whether genes are core, accessory, or strain-specific. **(B)** Matrix of p values showing whether ▼

gene, CSEP and CAZyme content were significantly different between lifestyles according to pairwise PERMANOVA. Coloured boxes indicate significant p values (<0.05). Global PERMANOVA results are reported in the bottom right of plots (see also Supplementary Table S3.3). **(C)** Scatterplot showing variation in gene copy number across all genes, CSEPs and CAZymes for different lifestyles. Points are jittered to reduce overlap. Sample size (the number of strains) is reported under x-axis labels.

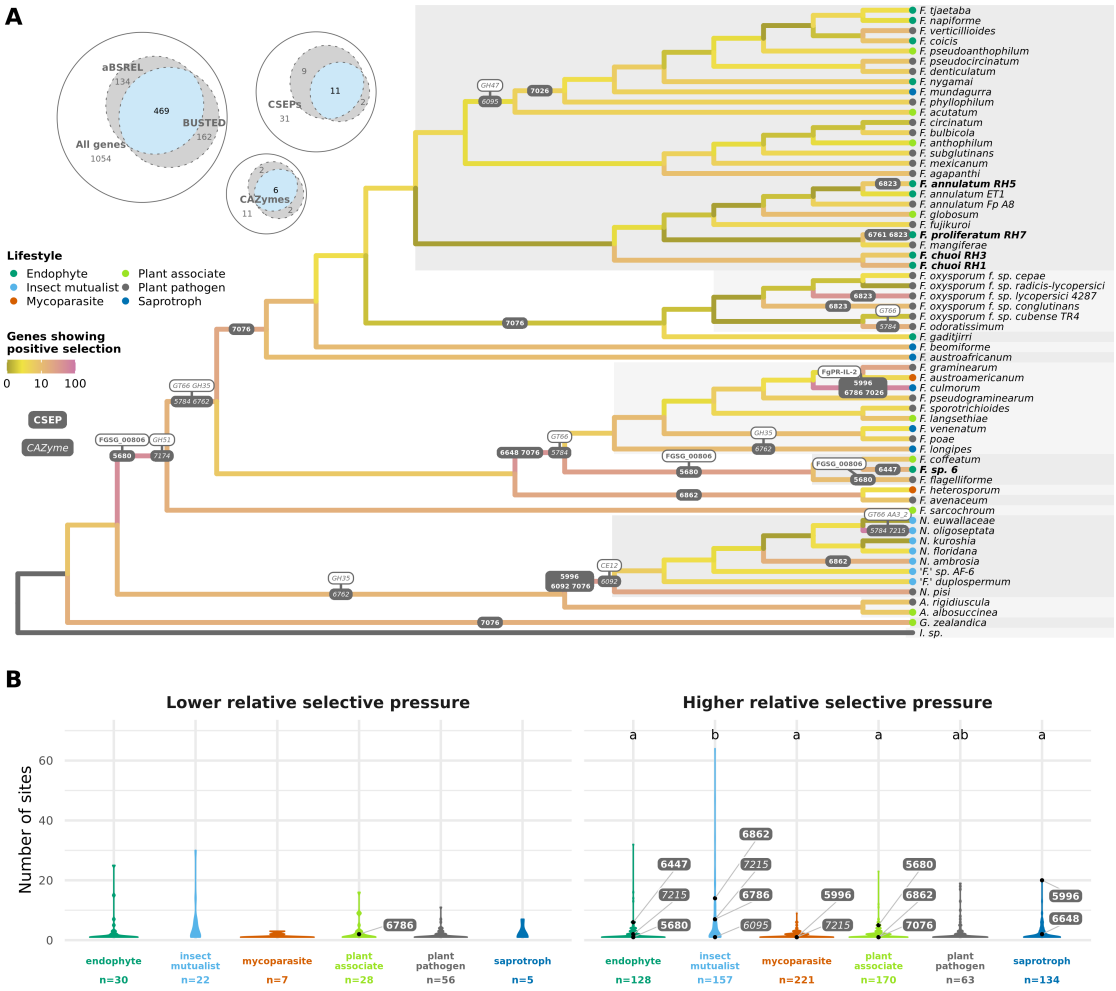


Figure 3.4: Results of dN/dS analyses on 1,054 core single-copy genes. **(A)** The Euler diagrams show the number of genes, CSEPs and CAZymes found to be under positive selection by both aBSREL and BUSTED. For the 469 cases where there was consensus between the two methods, the number of positively selected genes for each lineage according to aBSREL are shown by coloured branches on the species tree. The colour scale was pseudo log transformed for easier visualisation. For the exact number of positively selected genes on every branch, see Supplementary Figure S3.9. Branches on which CSEPs (bold) and CAZymes (italic) were positively selected are labelled with the gene ID(s) and, where possible, more detailed functional annotation is also indicated in white labels. Lifestyles of strains are indicated by coloured circles on tips. **(B)** Violin plot showing, for genes with at least 1 site with different selective pressure, the number of sites per genes for each lifestyle with lower (left) or higher (right) selective pressure relative to all other lifestyles according to Contrast-FEL. CSEPs (bold) and CAZymes (italic) that were also reported to be positively selected by BUSTED and aBSREL are indicated by points and labelled with the gene ID. Lifestyles with significant difference of means as calculated by the Games Howell test are shown by letters to the top of the plots (see Supplementary Tables S3.2 and S3.4 for full statistical test results). Sample size (the number of genes) is reported under x-axis labels.

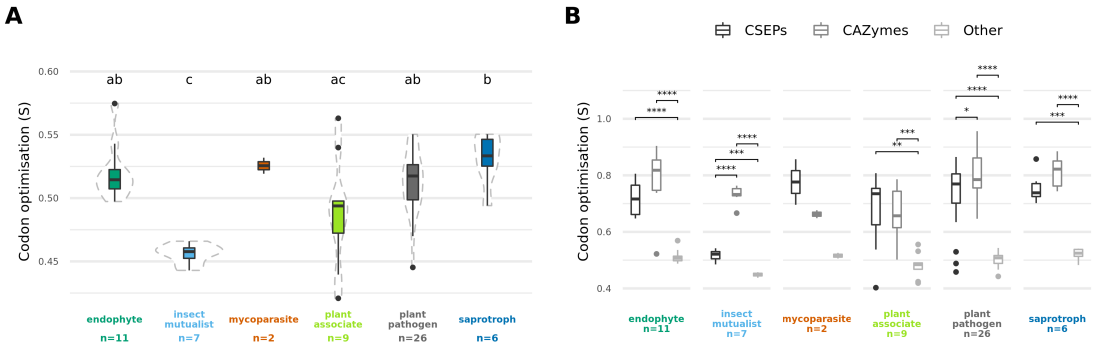


Figure 3.5: Boxplots showing codon optimisation (S) of core single-copy genes across lifestyles. Sample size (the number of strains) is reported under x-axis labels. **(A)** Difference in overall S values between lifestyles, with significant difference of means as calculated by TukeyHSD shown by letters at the top of the plot (see Supplementary Tables S3.2 and S3.4 for full statistical test results). **(B)** Difference in S values between CSEPs, CAZymes, and other genes for each lifestyle, with significant difference of means between the gene type as calculated by the Games Howell test shown by bars across significantly different categories (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001; see Supplementary Tables S3.2 and S3.4 for full statistical test results).

is a relative increase or decrease in dN/dS, and thus higher or lower selective pressure, compared with other branches. We reasoned that if a CSEP or CAZyme with higher relative selective pressure for a lifestyle was also found to be positively selected on an external lineage of that lifestyle, then it could suggest that the selective pressure is imposed by lifestyle. This was the case for 4 of the 9 core CSEPs and 1 of the 3 core CAZymes identified as positively selected on external lineages: CSEPs 6447 (*F.* sp. 6, endophyte); 5996 (*F. culmorum*, saprotroph); 6862 (*N. ambrosia*, insect mutualist); and 7076 (*Geejayessia zealandica*, plant associate); and CAZyme 7215 of lignin degrading subfamily AA3_2 (*N. oligoseptata*, insect mutualist).

Codon optimisation was higher in *Fusarium s. str.*

As dN/dS methods are biased by the erroneous assumption that all synonymous substitutions are neutral (Hershberg and Petrov, 2008; Rahman et al., 2021), we also explored whether translational selection (i.e., bias towards certain codons in more highly expressed genes) may be acting on synonymous substitutions by assessing the extent of codon optimisation (S) across fusarioid taxa (dos Reis, Savva and Wernisch, 2004). Codon optimisation of 1,054 core single-copy genes was generally high for all taxa (between 0.4 and 0.6, on a scale from -1 to 1), but it was significantly lower in insect mutualists compared with endophytes, plant pathogens and saprotrophs (Figure 3.5A, Supplementary Table S3.4). S values were found to be significantly higher in CSEPs and CAZymes than other core single-copy genes for all lifestyles (excluding mycoparasite, which could not be tested due to small sample size); furthermore, codon optimisation of CAZymes was also significantly higher than CSEPs for insect mutualists and plant pathogens (Figure 3.5B, Supplementary Table S3.4). CSEPs and CAZymes also encompassed greater extremes of codon optimisation than other core genes.

As high levels of codon optimisation has been linked to host generalism in fungi (Badet et al., 2017) and codon usage bias to wide habitat range in prokaryotes (Botzman and Margalit, 2011), we speculated that higher codon optimisation may be associated with lifestyle generalism – that is, taxa being capable of exhibiting multiple lifestyles. When no data correction was performed, there was a medium strength positive correlation between the number of reported lifestyles or 'lifestyle range'

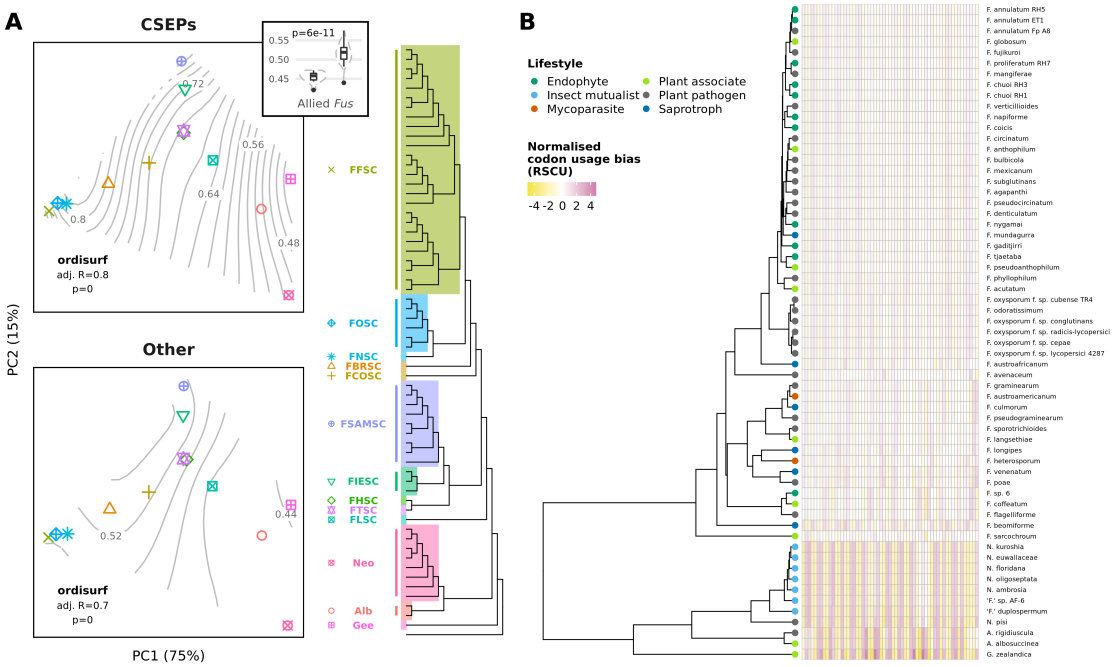


Figure 3.6: **(A)** PCA of phylogenetic distances between taxa, with points representing centroids for species complexes/allied genera, differentiated by shape and colour, as indicated by the tree legend. The percentage of variance explained by each principal component is shown on axis labels. Contours indicate the fit of codon optimisation (S values), of both core CSEPs and other core genes, to the ordination; the fit of CAZyme codon optimisation is not shown as it was not significant (p=0.2). The inset boxplot shows the significant difference (t-test, p=6e-11) in overall S values between *Fusarium* and allied genera. **(B)** Hierarchical clustering of taxa according to normalised RSCU. Heatmap columns represent codons (excluding Trp, Met, and stop codons) with cells coloured by normalised RSCU, where positive values represent higher than expected codon usage and negative values represent lower than expected codon usage.

and S values (Pearson's R=0.3, p=0.01), but the statistical significance of this correlation did not hold when accounting for phylogenetic relationships with PGLS analysis (p=0.06) (Supplementary Figure S3.12).

There was significantly higher codon optimisation in species complexes belonging to *Fusarium s. str.* compared with allied genera (t-test, p=6e-11; Figure 3.6A inset). Codon optimisation for CSEPs was shown to be strongly correlated with phylogeny as shown by the fit of S values to a PCA of phylogenetic distances (Figure 3.6A). This was not the case for CAZymes, however, for which the fit of codon optimisation to the PCA was not significant (p=0.2). Hierarchical clustering of taxa by normalised RSCU was also reasonably concordant with the species tree, with a Robinson-Foulds distance of 0.4 (p=0; Figure 3.6B), indicating that codon usage bias, for CSEPs if not CAZymes, is likely to be influenced by shared ancestry more than lifestyle.

3.5 Discussion

In this study, we inferred a phylogeny of *Fusarium* and allied genera using the greatest number of loci to date, with almost all branches significantly supported (Figure 3.3A). This adds to numerous recent efforts to produce high quality fungal phylogenies from genome-scale data (e.g., Spatafora et al., 2016; Steenwyk, Shen et al., 2019; Varga et al., 2019; Li, Steenwyk et al., 2021). Trimming method

and inclusion/exclusion of multi-copy genes had some impact on species tree topology (Figure 3.2; Supplementary Figure S3.6), but the *Fusarium s. str.* backbone was consistent across all approaches and in general agreement with the most recently published phylogeny of the group (Crous, Lombard et al., 2021). Discordance was concentrated in the ambrosia clade in *Neocosmospora*, perhaps due to the occurrence of interspecific hybridization in this lineage (Kasson et al., 2013) or horizontal gene transfer via the exchange of strains by beetles (Hulcr and Cognato, 2010). The objectives of this study were not concerned with the taxonomic debates surrounding the *Fusarium* generic concept, but our results did show that the divergence between *Fusarium s. str.* and other fusarioid taxa was associated with positive selection on a considerable number of core genes (Figure 3.4A); an upwards shift in translational selection (Figure 3.6A); and distinct patterns in codon usage bias (Figure 3.6B). While these results obviously do not directly contribute to characterisation of the taxa involved, they might be seen as a symptom of a 'larger and more abrupt' divergence than that between species within the same genus (Booth, 1978), contrary to *Fusarium s. lat.* (O'Donnell, Al-Hatmi et al., 2020; Geiser, Al-Hatmi et al., 2021).

We generally found the IR molecular clock model to produce dating estimates that were more concordant with estimates from other studies assessing divergence times of fusarioid fungi (e.g., Kasson et al., 2013; O'Donnell, Rooney et al., 2013; O'Donnell, Sink et al., 2015), which was largely to be expected considering that these studies also used IR models (but different secondary calibrations). The IR model estimated the divergence of obligate insect mutualists to correspond more closely to the crown age of their insect hosts, as estimated with insect fossil calibrations (Jordal and Cognato, 2012). By contrast, the AR model appeared to produce less congruent ages for recently diverged lineages, such as the highly specialised FOSC strains diverging before their host plants are likely to have existed. AR models have generally been thought appropriate for plants and animals considering the correlation between substitution rate and life-history traits (Lartillot, Phillips and Ronquist, 2016), and it has furthermore been suggested that AR is the norm across all kingdoms of life (Tao et al., 2019). On the other hand, Taylor and Berbee (2006) found no lineage-specific correlation of substitution rates across the kingdom Fungi. Similarly, Linder, Britton and Sennblad (2011) did not find strong evidence for rate autocorrelation across plant and simian datasets, instead finding the IR model to have more explanatory power. The AR model is not immune to bias (Lartillot and Delsuc, 2012), and has been shown to produce older estimates for simulated datasets across dating tools, including MCMCTree (Miura et al., 2020). The presence of short-term rate fluctuations in mammals suggests that mixed relaxed clock models accounting for both autocorrelation and jumps in rate variation are needed (Ho, 2009; Lartillot, Phillips and Ronquist, 2016).

Sources of error in divergence time estimation are manifold, as evidenced by the large confidence intervals in our analysis (Supplementary Figure S3.7). Beyond the difficulty surrounding choice and implementation of molecular clock models, a major source of error is the use of secondary calibrations – a necessity due to the general lack of fungal fossil data (Beimforde et al., 2014) – which can impact the precision and accuracy of divergence time estimates (Shaul and Graur, 2002; Graur and Martin, 2004; Sauquet et al., 2012; Schenk, 2016). For this reason, we incorporated the error from node ages estimated using primary fossil calibrations (Lutzoni et al., 2018) using confidence intervals to provide upper and lower bounds, as recommended when using secondary calibrations (Graur and Martin, 2004; Forest, 2009; Hipsley and Müller, 2014). An alternative approach is to expand taxon sampling until fossil data can be incorporated, although secondary calibrations have been shown to produce divergence time estimates with similar accuracy to those from distant primary calibrations,

albeit with lower precision (Powell, Waskin and Battistuzzi, 2020). Our motivation for divergence time estimation was not to test specific time-dependent hypotheses, but rather to calibrate branch lengths for more realistic measures of phylogenetic distance in subsequent comparative analyses. As with any divergence time analysis, major uncertainties are still associated with the divergence times of fusarioid fungi.

All taxa had a similar number of genes, CSEPs and CAZymes, very few of which were strain-specific (Figure 3.3A). It has previously been suggested that the number of species-specific secreted proteins (and by extension, we assume, effectors) is generally higher in fungal lifestyles which associate with plants without killing or decaying them, such as mutualistic symbionts and biotrophic pathogens, compared with saprotrophs and necrotrophic pathogens (Kim et al., 2016), the reasoning being that the former have to negotiate the plant–fungal interaction for an extended period. In the genus *Colletotrichum*, however, a reduction in the number of species-specific CSEPs was observed alongside the transition from phytopathogenicity to beneficial endophytism (Hacquard et al., 2016), showing that CSEPs and their impact on the plant–fungal interaction can be highly lineage-specific. We saw no significant difference in the number of strain-specific CSEPs (or genes) between any lifestyles (Supplementary Figure S3.8). This, combined with the fact that plant pathogens are often also reported as endophytes and vice versa (Figure 3.3A), and that plant pathogen and endophyte strains were not significantly different in terms of gene and CSEP content (Figure 3.3B), suggests that fusarioid taxa have a shared genetic capacity for phytopathogenicity and/or endophytism. Having a high proportion of species-specific CSEPs has also been associated with the connected factor of host specialisation (Spanu et al., 2010), which, considering we report very low numbers of strain-specific genes, may also explain the status of many *Fusarium* taxa as host generalists. Our results were also similar to those comparing pathogenic and non-pathogenic taxa in another genus of broad generalists, *Aspergillus* (Mead et al., 2021).

We did not identify common genetic signatures for the endophytic lifestyle in terms of gene, CSEP or CAZyme content, reinforcing the current understanding that there is no universal 'toolkit' associated with the endophytic lifestyle (Hacquard et al., 2016; Knapp et al., 2018). This contrasts with other well-defined lifestyles such as that of mycorrhizal fungi, for which specific genetic features have been associated with lifestyle in both ascomycetes and basidiomycetes (Martin, Kohler et al., 2010; Delaux et al., 2013; Kohler et al., 2015; Peter et al., 2016; Miyauchi, Kiss et al., 2020; Rich et al., 2021). One observed hallmark of the transition to mycorrhizal symbiosis is the loss of genes encoding PCWDEs (Kohler et al., 2015; Peter et al., 2016; Miyauchi, Kiss et al., 2020), but, as we found here, these are retained in various endophytic taxa (Zuccaro, Lahrmann and Langen, 2014; Lahrmann et al., 2015; Hacquard et al., 2016; Franco et al., 2021; Mesny, Miyauchi et al., 2021). As PCWDEs have often been treated predominantly as features of saprotrophy, this has fed into the hypothesis that many endophytes are latent saprotrophs, but in a broad comparison of CAZymes across the Dikarya, Zhao, Liu et al. (2013) demonstrated that plant pathogens have on average more CAZymes belonging to typical PCWDE families than saprotrophs. As there was no significant difference in total number or repertoire of CAZymes between plant pathogens, endophytes and saprotrophs, it indicates that fusarioid fungi retain the same machinery for plant cell wall degradation and/or remodelling, regardless of lifestyle. We did, however, find a significant difference in CSEP content between saprotrophs and endophytes (Figure 3.3B), which could suggest that fusarioid endophytes are more likely to be latent pathogens than saprotrophs.

The major exception to the apparent lifestyle flexibility among fusarioid fungi is the insect mutualist

lifestyle, which formed a monophyletic group (the ambrosia clade) in *Neocosmospora* (Figure 3.3A). The insect mutualist lifestyle was also the most distinct in terms of gene and CSEP content, being significantly different from all other lifestyles apart from the mycoparasitic lifestyle (Figure 3.3B), but the very small sample size for the latter will have impacted the test's power in that case (Alekseyenko, 2016). The transition to symbiotic mutualism in *Neocosmospora* was not associated with a reduction in total number of genes, CSEPs or CAZymes, in agreement with results from other ectosymbiotic insect mutualists (Biedermann and Vega, 2020). As the representative strains used in this study are all known to cause disease on the trees they colonise with their beetle partner (Freeman et al., 2013; O'Donnell, Libeskind-Hadas et al., 2016; Na et al., 2018; Aoki, Smith et al., 2019), it follows that they would have retained many of the genetic mechanisms from their (presumably) plant associated ancestors. Some strains have been found to cause disease *in vitro* in the absence of their beetle partners (e.g., Eskalen et al., 2012; Na et al., 2018), however, to our knowledge, fusarioid ambrosia fungi have never been reported as free-living in the wild.

Although we did not identify significant differences in the genetic repertoires between fusarioid endophytes and plant pathogens, we did find some evidence that copy number variation – genes or regions that are either duplicated or deleted in reference to other taxa – may be contributing to lifestyle. There was no significant difference in mean gene copy number between lifestyles, but plant pathogens included extreme outliers in gene copy number compared with other lifestyles (Figure 3.3C). Extensive gene duplication has been suggested as a key strategy for pathogenicity in basidiomycete rusts (Pendleton et al., 2014), and copy number of the pectin degrading CAZyme subfamily PL1_7 across 41 root-colonising fungi was shown to correlate with pathogenicity in *Arabidopsis* (Mesny, Miyauchi et al., 2021). Gene duplication is regarded as the primary resource for the evolution of functional novelties, and the persistence of gene duplicates is indicative of neofunctionalisation and/or subfunctionalisation, as a functionally redundant gene copy will be rapidly lost due to the absence of selective pressure to retain it (Lynch and Conery, 2000; He and Zhang, 2005). The most common functional innovations of gene copies in fungi are regulatory changes (Wapinski et al., 2007). Indeed, copy number variation is known to be correlated with differential gene expression (Stranger et al., 2007; Steenwyk and Rokas, 2018; Shao et al., 2019), and has been shown to contribute to phenotypic or pathological differences in fungi (Steenwyk, Soghigian et al., 2016; Zhao and Gibbons, 2018).

This aligns with mounting evidence that a major factor impacting lifestyle of closely related phytopathogens and endophytes is not gene repertoire itself, but expression profiles. Returning to *Colletotrichum*, Hacquard et al. (2016) found that a pathogenic taxon had a different pattern of gene expression during host colonisation, including upregulation of CSEPs, compared with a closely related and genetically similar beneficial endophyte. The authors noted that this also makes the beneficial endophyte genetically capable of reverting to pathogenicity (and, presumably, the closely related pathogens capable of inhabiting plants as endophytes). The aforementioned CAZyme subfamily PL1_7, which we found between 2 and 4 copies of in all fusarioid taxa (Supplementary Figure S3.11), was also more highly expressed in the pathogenic *Colletotrichum* taxon. The importance of expression has already been seen in *Fusarium*, where expression of secondary metabolites differed between endophytic and pathogenic strains of the same species, *F. annulatum* (as *F. proliferatum*, FFSC), despite generally sharing secondary metabolite gene clusters (Niehaus et al., 2016). Generating *in planta* expression profiles for both pathogenic and non-pathogenic strains across the group could reveal whether there is convergence in expression patterns for certain lifestyles.

Regulation of certain genes located on accessory chromosomes has also been seen to direct plant

infection phenotypes in an endophytic versus pathogenic FOSC strain (Guo et al., 2021). Accessory chromosomes – chromosomes that are not essential for survival, but potentially confer functional advantages (Bertazzoni et al., 2018) – are likely another important factor impacting lifestyle in *Fusarium*. The first acc. chromosomes in fungi were discovered in the fusarioid species *Neocosmospora haematococca* (as *Nectria haematococca*) (Coleman et al., 2009), with further reports in at least nine other fusarioid strains (Bertazzoni et al., 2018). They have mostly been studied in the FOSC, in which horizontal transfer of acc. chromosomes can confer pathogenicity (Ma, van der Does et al., 2010; Li, Fokkens et al., 2020). Not only are acc. chromosomes deemed to be a key innovation for rapid adaptation by plant pathogens (Croll and McDonald, 2012) they have also been implicated in adaptation of FOSC strains to human pathogenicity (Zhang, Yang et al., 2020). Exploring the extent of acc. chromosomes broadly across fusarioid fungi, as well as phenomena impacting genomic architecture such as transposable elements (Muszewska et al., 2019), may shed light on the mechanisms underlying lifestyle flexibility in the group (Ma, Geiser et al., 2013).

As effectors are highly diverged and often lineage-specific, if not strain-specific, only a small proportion of the CSEPs predicted here could be matched to experimentally verified genes from PHI-base. Of these, the majority were genes known to impact virulence to some degree or not at all in the hosts they have been tested on (Supplementary Figure S3.10), although the knockout mutant phenotype for a certain gene will not necessarily be the same for different fungal strains or on different hosts. PHI-base is also explicitly dedicated to pathogen-host genes, and similar high quality, curated resources are needed for genes involved in non-pathogenic fungal–host interactions. Nonetheless, our results give us a broad perspective on CSEP distributions across fusarioid fungi. Some CSEPs exhibited phylogenetic patterns (such as lower copy number in *Fusarium s. lat.* compared with *Fusarium s. str.* for MoCDIP4, which was first discovered in *Magnaporthe oryzae* (Chen, Fichtner et al., 2013) and since reported in *F. oxysporum* f. sp. *pisi* (Achari et al., 2021)), but most had scattered distributions across the group (Supplementary Figure S3.10), which may be the result of frequent horizontal gene transfer (e.g., van Dam and Rep, 2017; Peck et al., 2021).

A slightly lower proportion of core CSEPs were found to be positively selected than non-CSEPs according to dN/dS calculations (Figure 3.4A). This may be seen as surprising, as effectors that promote virulence are assumed to be under strong selective pressure during the evolutionary arms race between fungus and host (de Jonge, Bolton and Thomma, 2011; Lo Presti et al., 2015). For instance, CSEPs have been found to more frequently be under positive selection compared with non-CSEPs in phytopathogenic *Microbotryum* species (Beckerson et al., 2019). High rates of selection on CSEPs are not only a hallmark of pathogenicity, however, as these have also been observed for obligate, host-specific *Epichloë* endophytes (Schirrmann et al., 2018); the arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Schmitz, Pawlowska and Harrison, 2019); and the saprotroph *Verticillium tricorpus* (Seidl, Faino et al., 2015), emphasizing the broader roles played by effectors in host–fungal interactions. Our results could be explained by the fact that we focused on core genes, and so the CSEPs in questions are presumably contributing to integral host–fungal interactions that would be under similar selective pressure as other core functions, rather than specialised CSEPs more likely to be under strong selective pressure from the host. We should also note that detection of positive selection with dN/dS methods is biased against shorter genes (Derbyshire, Harper and Lopez-Ruiz, 2021), which CSEPs by definition are, and so this may have impacted our results.

We identified five cases where positive selection of core CSEPs and CAZymes may be connected to lifestyle by comparing aBSREL analysis of positive selection on external branches to Contrast-FEL

analysis of relative selection pressures between lifestyles. Interestingly, there were no core CSEPs with higher selective pressure in plant pathogens relative to other lifestyles, which could be interpreted as evidence that the ancestral state of the group is phytopathogenic rather than endophytic, but the unbalanced sample sizes for the different lifestyles will have influenced the Contrast-FEL results. Once again, the insect mutualist lifestyle was shown to be distinct, with a greater number of sites per gene undergoing higher selective pressure relative to other lifestyles (Figure 3.4B). This may be associated with the fact that these ambrosia taxa have evolved via insect farming, in what could be interpreted as some level of 'artificial selection' (Mueller et al., 2005). We were only able to tentatively link the positive selection of one core CAZyme to lifestyle: 5'-oxoaverantin cyclase in the AA3_2 subfamily, which was positively selected for in the insect mutualist *N. oligoseptata* (Figure 3.4A). Other members of the same subfamily are implicated in lignin degradation (Levasseur et al., 2013; Miyauchi, Navarro et al., 2017), but 5'-oxoaverantin cyclase was first identified as an intermediate in aflatoxin biosynthesis in *Aspergillus parasiticus* (Sakuno, Yabe and Nakajima, 2003). Another insect-fungus mutualism between the navel orangeworm and *A. flavus* has shown that aflatoxin tolerance is a key adaptation of the insect to its fungal diet (Niu et al., 2009; Ampt et al., 2016), and as fusarioid fungi are known to produce an array of mycotoxins (Desjardins and Proctor, 2007), it would be interesting to determine whether there is a similar dynamic in the evolution of the ambrosia mutualism.

Conventional dN/dS methods to detect selection such as aBSREL and BUSTED make the assumption that synonymous substitutions are always selectively neutral, but we now know that selection does occur on synonymous mutations (Ohta, 1996; Chen, Lee et al., 2004; Hershberg and Petrov, 2008). Subsequently dN/dS methods have been shown to overestimate the frequency of positive selection and underestimate the strength of negative selection in bacteria, even when selection on synonymous sites is weak (Rahman et al., 2021). Furthermore, using dN/dS>1 as a signifier of positive selection has been declared arbitrary (Tamuri and Dos Reis, 2021). As flexible dN/dS methods accounting for selection on synonymous substitutions have yet to be integrated into the widely used tools for detecting positive selection, this remains a caveat of our dN/dS analyses. Additionally, even a low incidence of sequence inaccuracies can results in false-positive signals of selection (Mallick et al., 2009), so ideally candidate genes should be resequenced to detect errors and confirm whether sites are truly under selection. A further limitation of the selection analyses is that they were restricted to core genes due to the requirement of a robust species tree to estimate dN/dS across lineages, which necessarily excludes a large proportion of the gene content (Derbyshire, Harper and Lopez-Ruiz, 2021). Further exploration of selection dynamics in the extensive accessory content would undoubtedly shed more light on the evolution of the group.

When exploring the issue of selection on synonymous substitutions, we showed that codon optimisation of the core single-copy genes – that is, the extent of translational selection on codon usage – was higher in CSEPs and CAZymes than other genes (Figure 3.5B), as was previously found in the *F. oxysporum* f. sp. *cepae* pangenome (Armitage et al., 2018). Insect mutualists had a much larger difference in codon optimisation between CSEPs and CAZymes (Figure 3.5B). One possible explanation for this result is that these taxa may have less translational selective pressure on CSEPs that are required for plant invasion – being farmed by insects which excavate and weaken the plant hosts – but retain higher translational selective pressure on CAZymes that are required for assimilation of nutrients, which ultimately maintains the insect-fungus mutualism. Following this broad perspective on codon optimisation, further functional annotation could allow the use of a 'reverse

ecology framework' to explore whether genes with the highest codon optimisation correspond with lifestyle (LaBella et al., 2021).

We also found that correlation between lifestyle range and codon optimisation was not significant after correcting for phylogenetic relationships (Supplementary Figure S3.12), contrary to expectation from previous studies (Botzman and Margalit, 2011; Badet et al., 2017). Our approach to assess lifestyle range was limited by the availability of published reports of fusarioid taxa, and so we will undoubtedly have underestimated the number of lifestyles exhibited by some species. Furthermore, fusarioid species are often hard to distinguish, and lifestyle reports may therefore be misattributed. To mitigate against this issue, we only included studies that used appropriate genetic markers to distinguish taxa – not, for instance, solely using internal transcribed spacer (ITS) (Geiser, Jiménez-Gasco et al., 2004) – and crosschecked phylogenetic analyses for misclassifications. Despite this, we may have inadvertently included lifestyle reports for species that were incorrectly classified in the original study. A comprehensive meta-analysis is needed to better understand the extent of lifestyle and host range for fusarioid taxa.

A major caveat of our comparative analyses is that we were forced to attribute a single lifestyle to the strains being used, despite the current understanding, which our own results support, that these lifestyles are not necessarily mutually exclusive (Selosse, Schneider-Maunoury and Martos, 2018). Furthermore, treating lifestyles as categorical traits does not accurately reflect the range of outcomes we know can exist within even one lifestyle, such as different pathogenic strains within the same species varying in 'aggressiveness' (e.g., Holtz et al., 2011; Chen, Zhou et al., 2014; Šišić et al., 2018). These both remain central issues with current approaches to fungal lifestyle comparison at large (e.g., Knapp et al., 2018; Miyauchi, Kiss et al., 2020; Franco et al., 2021; Mesny, Miyauchi et al., 2021). New methods that can effectively incorporate multiple lifestyle hypotheses, or treat lifestyles as points on a continuous spectrum, are sorely needed to encapsulate the nuance of these highly context-dependent interactions.

3.6 Conclusions

We found an apparent shared genetic capacity for phytopathogenicity and endophytism in *Fusarium*, which suggests that, while strains may be reported as plant pathogens or endophytes, their lifestyle is potentially transient. Were fusarioid taxa to make the transition to obligate, mutualistic endophytism, we might expect to see genetic hallmarks more akin to those seen in the transition to obligate symbiosis in mycorrhizal lifestyles (e.g., Delaux et al., 2013). Despite multiple reports of certain endophytic *Fusarium* strains being beneficial to certain plant hosts (e.g., Kavroulakis et al., 2007; Mendoza and Sikora, 2009; Bilal et al., 2018), large uncertainties remain as to the stability of these interactions. Our results depict fusarioid fungi as prolific generalists and highlight the difficulty in predicting pathogenic potential in the group. Considering the importance of plant immune response, biotic and abiotic conditions to the plant–fungal interaction, such endophytes may not be the 'silver bullet' for biocontrol that they are sometimes touted to be.

Acknowledgements

We thank Marcelo Sandoval-Denis and Pedro Crous for assistance with morphological identification of the strains. We also thank Mark Blaxter for helpful advice on assembly tool comparison; Mario dos

Reis for advice on using MCMCTree; and Theo Llewellyn, Laura Kelly and James Borrell for valuable discussion. We also thank the Molecular Biology and Evolution associate editor Crystal Hepp and three anonymous reviewers for their valuable feedback on the published paper. We acknowledge the assistance of the ITS Research team at Queen Mary University of London.

Data availability

WGS data and structurally annotated genome assemblies generated in this study are available on GenBank under the BioProject accession PRJNA761077. Additional data files of the raw phylogenetic trees; CSEP and CAZyme amino acid sequences; OrthoFinder output; and orthogroup metadata are deposited in Zenodo doi:10.5281/zenodo.6353640.

3.7 Supplementary material

Supplementary Table S3.1: Assembly and annotation statistics for the five *Fusarium* strains.

				QUAST						BUSCO	MAKER
			Coverage	# contigs ≥500bp	Largest contig (bp)	Total size (bp)	GC (%)	N50	L50	Completeness (single-copy BUSCOs)	# genes
F. chuoi	RH1	ABySS k124	286×	111	4,298,088	45,254,299	46.81	1,615,464	10	4,484 (99.78%)	13,380
		MEGAHIT	290×	1,874	553,685	44,681,473	47.2	115,300	117	4,481 (99.71%)	-
		SPAdes	288×	457	1,232,307	45,027,201	46.98	386,992	38	4,480 (99.69%)	-
F. chuoi	RH3	ABySS k90	304×	981	1,147,506	44,348,592	47.75	214,555	56	4,485 (99.80%)	14,313
		MEGAHIT	296×	525	1,641,087	45,474,718	46.86	308,265	43	4,484 (99.78%)	-
		SPAdes	300×	956	1,149,846	44,908,739	47.28	222,386	60	4,486 (99.82%)	-
F. annulatum	RH5	ABySS k121	317×	70	3,875,036	43,829,649	48.3	1,803,139	9	4,478 (99.64%)	12,880
		MEGAHIT	327×	64	2,611,815	43,875,029	48.29	1,638,693	11	4,480 (99.69%)	-
		SPAdes	325×	136	2,843,665	43,842,946	48.31	1,163,461	14	4,479 (99.67%)	-
F. sp.	RH6	ABySS k128	298×	110	4,722,096	39,424,294	47.66	1,717,955	7	4,476 (99.60%)	11,533
		MEGAHIT	340×	940	1,573,830	39,142,705	47.79	224,512	49	4,478 (99.64%)	-
		SPAdes	332×	176	1,999,176	39,279,569	47.7	1,117,341	14	4,478 (99.64%)	-
F. proliferatum	RH7	ABySS k127	322×	107	4,142,881	44,857,950	48.05	1,615,920	10	4,481 (99.71%)	13,009
		MEGAHIT	286×	609	1,463,918	44,731,912	48.09	324,653	43	4,481 (99.71%)	-
		SPAdes	282×	140	3,454,284	44,751,997	48.07	1,124,917	14	4,481 (99.71%)	-

Supplementary Table S3.2: Statistical test results for Levene's test for homogeneity of residual variance and ANOVA, or ART ANOVA (*) if Levene's test was significant ($p < 0.05$).

	Levene's test					ANOVA / ART ANOVA (*)				
	Formula	df1	df2	statistic	p	Formula	Effect	Df	F	p
# genes (Figure 3.3A)	All genes ~ lifestyle	5	55	0.9	0.5	All genes ~ PC1 + PC2 + lifestyle	PC1	1	1.32	0.3
							PC2	1	14.47	4E-04
							lifestyle	5	2.29	0.06
	CSEPs ~ lifestyle	5	55	1.52	0.2	CSEPs ~ PC1 + PC2 + lifestyle	PC1	1	1.5	0.2
							PC2	1	7.26	09
							lifestyle	5	2.13	0.08
	CAZymes ~ lifestyle	5	55	1.71	0.2	CAZymes ~ PC1 + PC2 + lifestyle	PC1	1	0.37	0.5
							PC2	1	6.75	0.01
							lifestyle	5	2.15	0.07
# strain specific genes (Supp. Figure 4A)	All genes ~ lifestyle	5	55	1.13	0.4	All genes ~ PC1 + PC2 + lifestyle	PC1	1	3.88	0.05
							PC2	1	0.5	0.5
							lifestyle	5	1.1	0.4
	CSEPs ~ lifestyle	5	55	0.67	0.6	CSEPs ~ PC1 + PC2 + lifestyle	PC1	1	7.73	07
							PC2	1	2.25	0.1
							lifestyle	5	1.51	0.2
Mean gene copy number (Supp. Figure 3.5B)	All genes ~ lifestyle	5	55	2.02	0.09	All genes ~ PC1 + PC2 + lifestyle	PC1	1	0	1
							PC2	1	3.14	0.08
							lifestyle	5	1.76	0.1
	CSEPs ~ lifestyle	5	55	1.64	0.2	CSEP ~ PC1 + PC2 + lifestyle	PC1	1	0	1
							PC2	1	2.53	0.1
							lifestyle	5	1.25	0.3
	CAZymes ~ lifestyle	5	55	1.75	0.1	CAZymes ~ PC1 + PC2 + lifestyle	PC1	1	0.02	0.9
							PC2	1	2.18	0.1
							lifestyle	5	1.61	0.2
# positively selected genes on external branches	num ~ lifestyle	5	52	1.5	0.206	num ~ PC1 + PC2 + lifestyle	PC1	1	1.08	0.3
							PC2	1	0.05	0.8
							lifestyle	5	0.63	0.7

Supplementary Table S3.2 continued.

	Levene's test					ANOVA / ART ANOVA (*)				
	Formula	df1	df2	statistic	p	Formula	Effect	Df	F	p
Sites with different relative evolutionary rate (Figure 3.4B)	(Higher) sites ~ lifestyle	5	867	14.94	4E-14	(Higher) sites ~ lifestyle *	lifestyle	5	9.77	4E-09
	(Lower) sites ~ lifestyle	5	142	1.64	0.2	(Lower) sites ~ lifestyle	lifestyle	5	1.61	0.2
Codon optimisation between lifestyles (Figure 3.5A)	S ~ lifestyle	5	55	1.97	0.1	S ~ PC1 + PC2 + lifestyle	PC1	1	21.84	2E-05
							PC2	1	45.37	1E-08
							lifestyle	5	2.95	0.02
Codon optimisation between gene types (Figure 3.5B)	(Endophyte)					(Endophyte)	PC1	1	0.64	0.4
	S ~ gene type	2	30	4.21	0.02	S ~ PC1 + PC2 + gene type *	PC2	1	0.02	0.9
							gene type	2	35.81	2E-08
	(Insect mutualist)					(Insect mutualist)	PC1	1	0.29	0.6
	S ~ gene type	2	18	1.9	0.2	S ~ PC1 + PC2 + gene type	PC2	1	0.29	0.6
							gene type	2	308.38	2E-13
	(Plant associate)					(Plant associate)	PC1	1	2.73	0.1
	S ~ gene type	2	24	1.83	0.2	S ~ PC1 + PC2 + gene type	PC2	1	2.03	0.2
							gene type	2	12.39	2E-04
	(Plant pathogen)					(Plant pathogen)	PC1	1	14.21	3E-04
	S ~ gene type	2	75	5.79	0.005	S ~ PC1 + PC2 + gene type *	PC2	1	24.25	5E-06
							gene type	2	80.18	4E-19
	(Saprotroph)					(Saprotroph)	PC1	1	0.26	0.6
	S ~ gene type	2	15	1.28	0.3	S ~ PC1 + PC2 + gene type	PC2	1	0.46	0.5
							gene type	2	55.57	4E-07

Supplementary Table S3.3: Statistical test results for PERMANOVA.

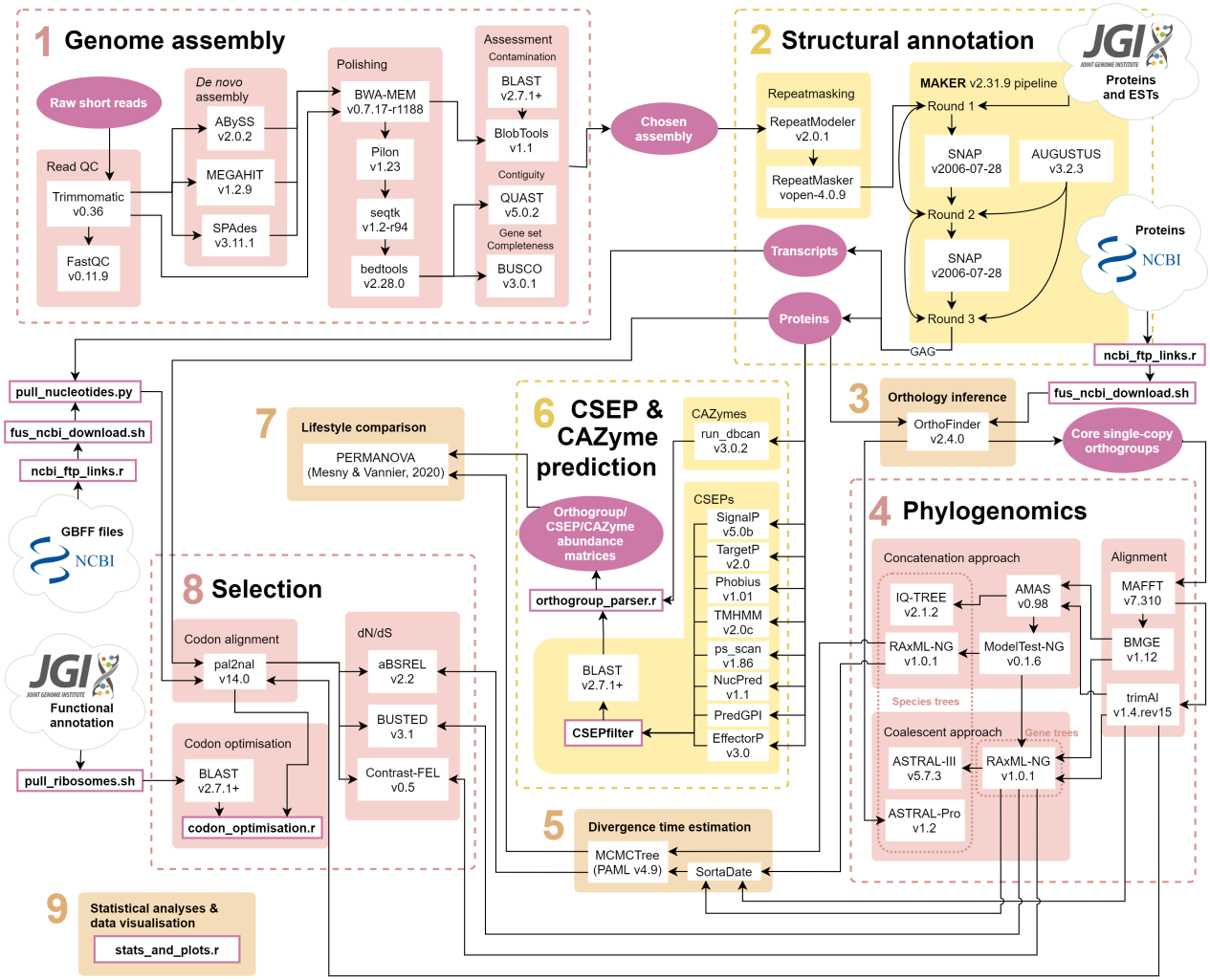
	Formula	Effect	Df	SumOfSqs	R2	F	p
All genes	JaccardDistMatrix ~ PC1 + PC2 + lifestyle	PC1	1	1.04	0.2	18.88	1.00E-04
		PC2	1	0.8	0.15	14.54	1.00E-04
		lifestyle	5	0.45	0.09	1.62	0.0063
		Residual	53	2.91	0.56		
		Total	60	5.19	1		
CSEPs	JaccardDistMatrix ~ PC1 + PC2 + lifestyle	PC1	1	2.57	0.21	19.49	1.00E-04
		PC2	1	1.63	0.13	12.39	1.00E-04
		lifestyle	5	1.13	0.09	1.71	0.002
		Residual	53	6.99	0.57		
		Total	60	12.32	1		
CAZymes	JaccardDistMatrix ~ PC1 + PC2 + lifestyle	PC1	1	0.87	0.26	27.85	1.00E-04
		PC2	1	0.52	0.16	16.59	1.00E-04
		lifestyle	5	0.3	0.09	1.91	0.0039
		Residual	53	1.66	0.5		
		Total	60	3.35	1		

Supplementary Table S3.4: Statistical test results for pairwise multiple comparisons using Tukey HSD or, if Levene's test was significant (see Supplementary Table S3.2), Games Howell test (*).

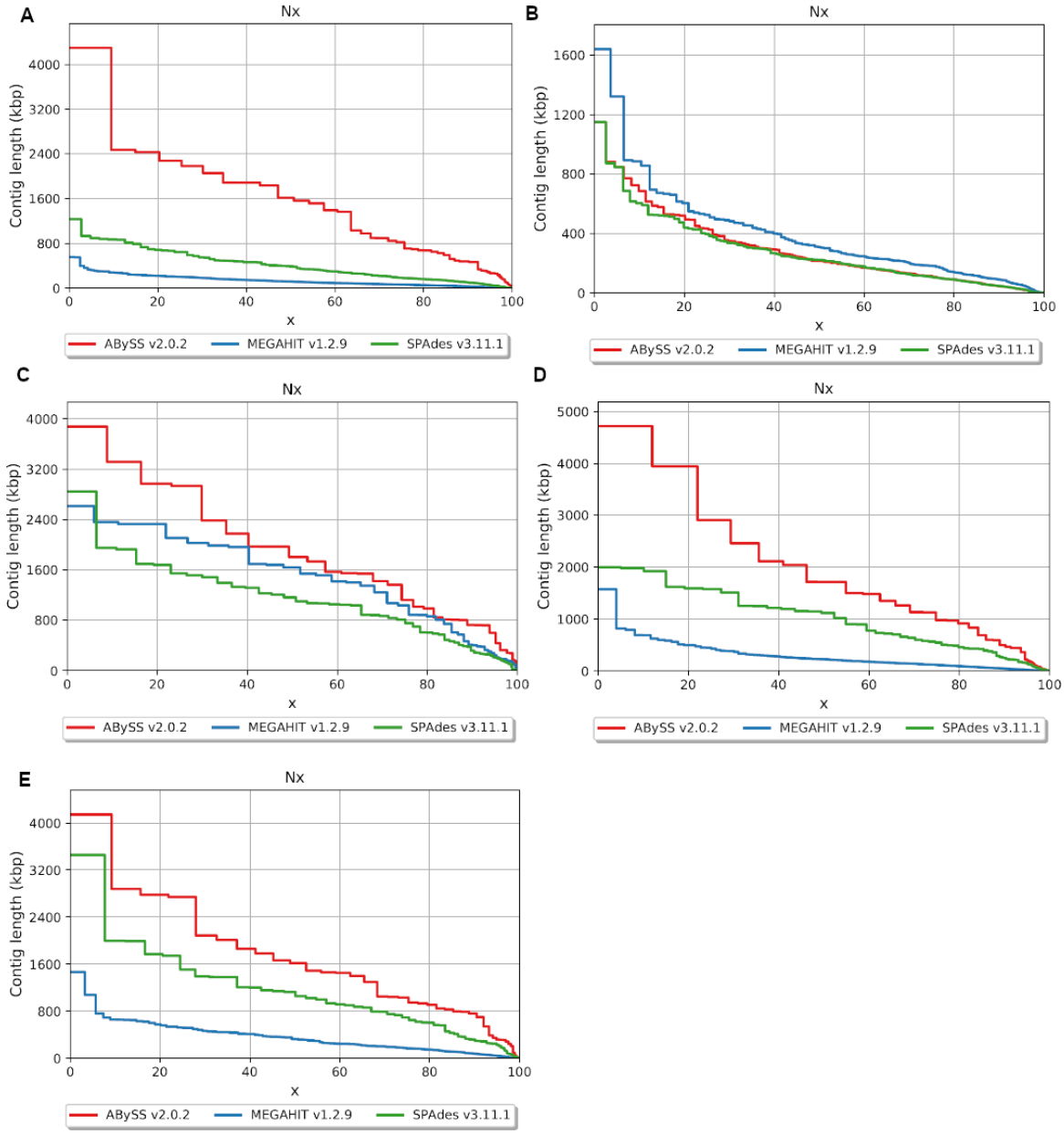
	Formula	group1	group2	estimate	conf.low	conf.high	p.adj
Sites with different relative evolutionary rate (Figure 3.4B)	(Higher) sites ∼ lifestyle *	endophyte	insect mutualist	2.86	0.81	4.91	0.001
		endophyte	mycoparasite	-0.7	-1.62	0.23	0.3
		endophyte	plant associate	-0.18	-1.21	0.86	1
		endophyte	plant pathogen	0.57	-1.19	2.32	0.9
		endophyte	saprotroph	-0.06	-1.15	1.03	1
		insect mutualist	mycoparasite	-3.55	-5.42	-1.69	2E-06
		insect mutualist	plant associate	-3.03	-4.96	-1.11	1E-04
		insect mutualist	plant pathogen	-2.29	-4.67	0.09	0.07
		insect mutualist	saprotroph	-2.92	-4.87	-0.96	4E-04
		mycoparasite	plant associate	0.52	-0.04	1.08	0.09
		mycoparasite	plant pathogen	1.26	-0.28	2.8	0.2
		mycoparasite	saprotroph	0.64	-0.03	1.31	0.07
		plant associate	plant pathogen	0.75	-0.86	2.35	0.8
		plant associate	saprotroph	0.12	-0.69	0.93	1
		plant pathogen	saprotroph	-0.63	-2.27	1.02	0.9
Codon optimisation between lifestyles (Figure 3.5A)	S ∼lifestyle	endophyte	insect mutualist	-0.06	-0.1	-0.03	8E-05
		endophyte	mycoparasite	0.01	-0.05	0.07	1
		endophyte	plant associate	-0.03	-0.07	0	0.1
		endophyte	plant pathogen	-0.01	-0.04	0.02	1
		endophyte	saprotroph	0.01	-0.03	0.05	1
		insect mutualist	mycoparasite	0.07	0.01	0.13	0.02
		insect mutualist	plant associate	0.03	-0.01	0.07	0.1
		insect mutualist	plant pathogen	0.06	0.02	0.09	7E-05
		insect mutualist	saprotroph	0.07	0.03	0.12	6E-05
		mycoparasite	plant associate	-0.04	-0.1	0.02	0.5
		mycoparasite	plant pathogen	-0.01	-0.07	0.04	1
		mycoparasite	saprotroph	0.01	-0.06	0.07	1
		plant associate	plant pathogen	0.02	-0.01	0.05	0.2
		plant associate	saprotroph	0.04	0	0.08	0.05
		plant pathogen	saprotroph	0.02	-0.02	0.05	0.7

Supplementary Table S3.4 continued.

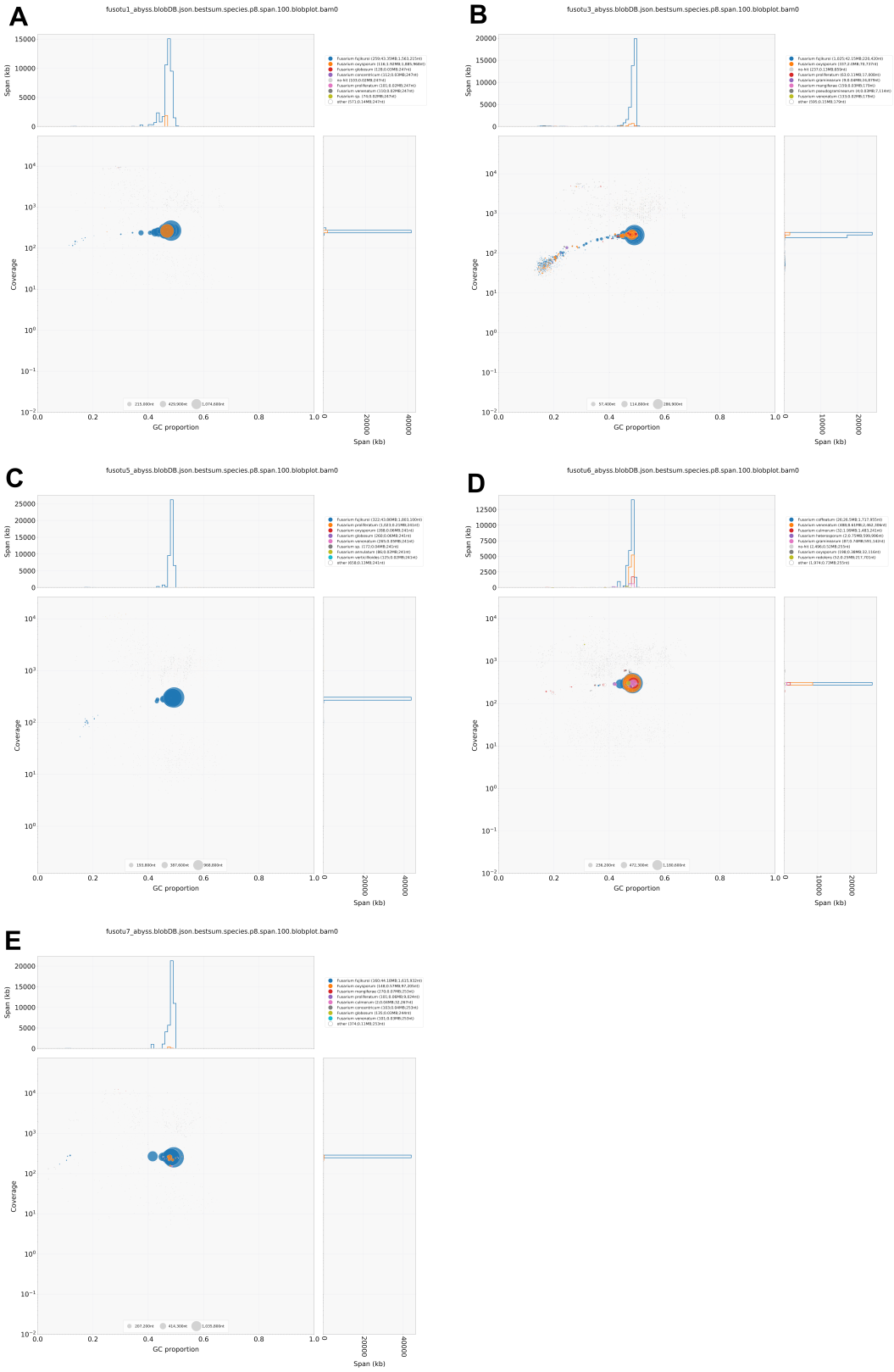
	Formula	group1	group2	estimate	conf.low	conf.high	p.adj
Codon optimisation between different gene types (Figure 3.5B)	(Endophyte)	S.CSEP	S.CAZyme	0.07	-0.03	0.16	0.2
	S ~gene type *	S.CSEP	S.other	-0.21	-0.26	-0.16	3E-07
		S.CAZyme	S.other	-0.27	-0.36	-0.19	1E-05
	(Insect mutualist)	S.CSEP	S.CAZyme	0.22	0.18	0.25	2E-12
	S ~gene type *	S.CSEP	S.other	-0.07	-0.1	-0.04	6E-05
		S.CAZyme	S.other	-0.28	-0.31	-0.25	4E-14
	(Plant associate)	S.CSEP	S.CAZyme	0	-0.12	0.11	1
	S ~gene type *	S.CSEP	S.other	-0.19	-0.31	-0.08	9E-04
		S.CAZyme	S.other	-0.19	-0.31	-0.08	0.001
	(Plant pathogen)	S.CSEP	S.CAZyme	0.07	0.01	0.13	0.03
	S ~gene type *	S.CSEP	S.other	-0.23	-0.28	-0.18	4E-11
		S.CAZyme	S.other	-0.3	-0.34	-0.26	3E-14
	(Saprotroph)	S.CSEP	S.CAZyme	0.06	-0.02	0.13	0.1
	S ~gene type *	S.CSEP	S.other	-0.24	-0.31	-0.16	1E-06
		S.CAZyme	S.other	-0.29	-0.36	-0.22	8E-08



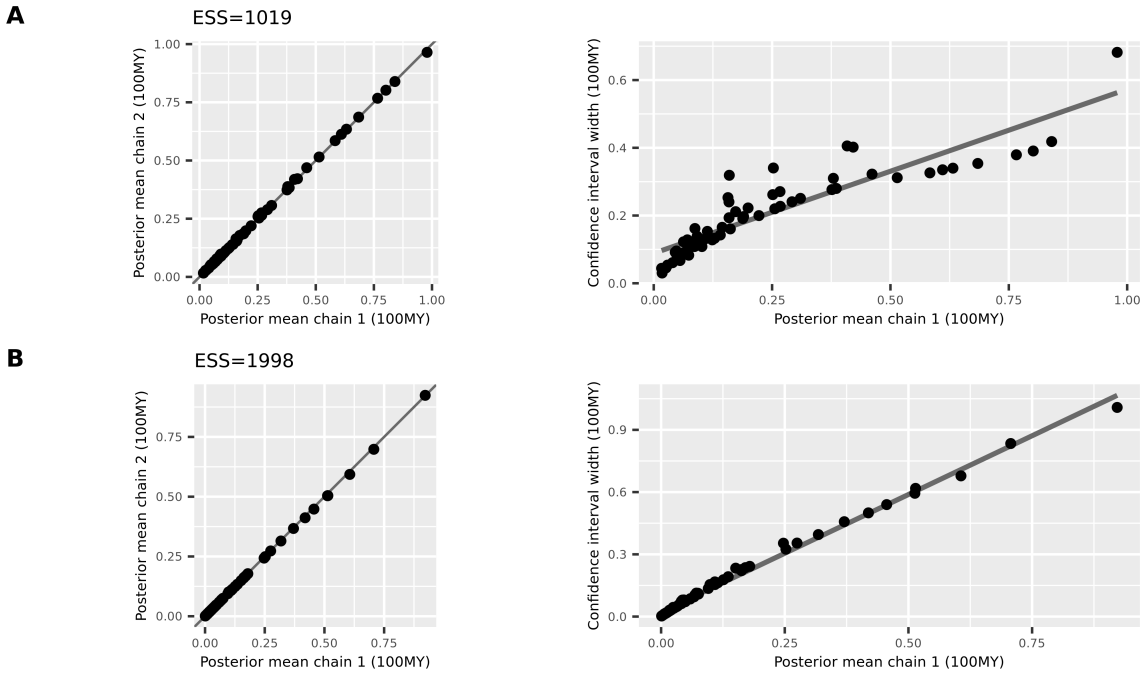
Supplementary Figure S3.1: Schematic summarising the bioinformatics analysis pipeline developed in Chapter 3, available at `https://github.com/Rowena-h/FusariumLifestyles`. Boxes outlined in pink indicate custom scripts written for this work.



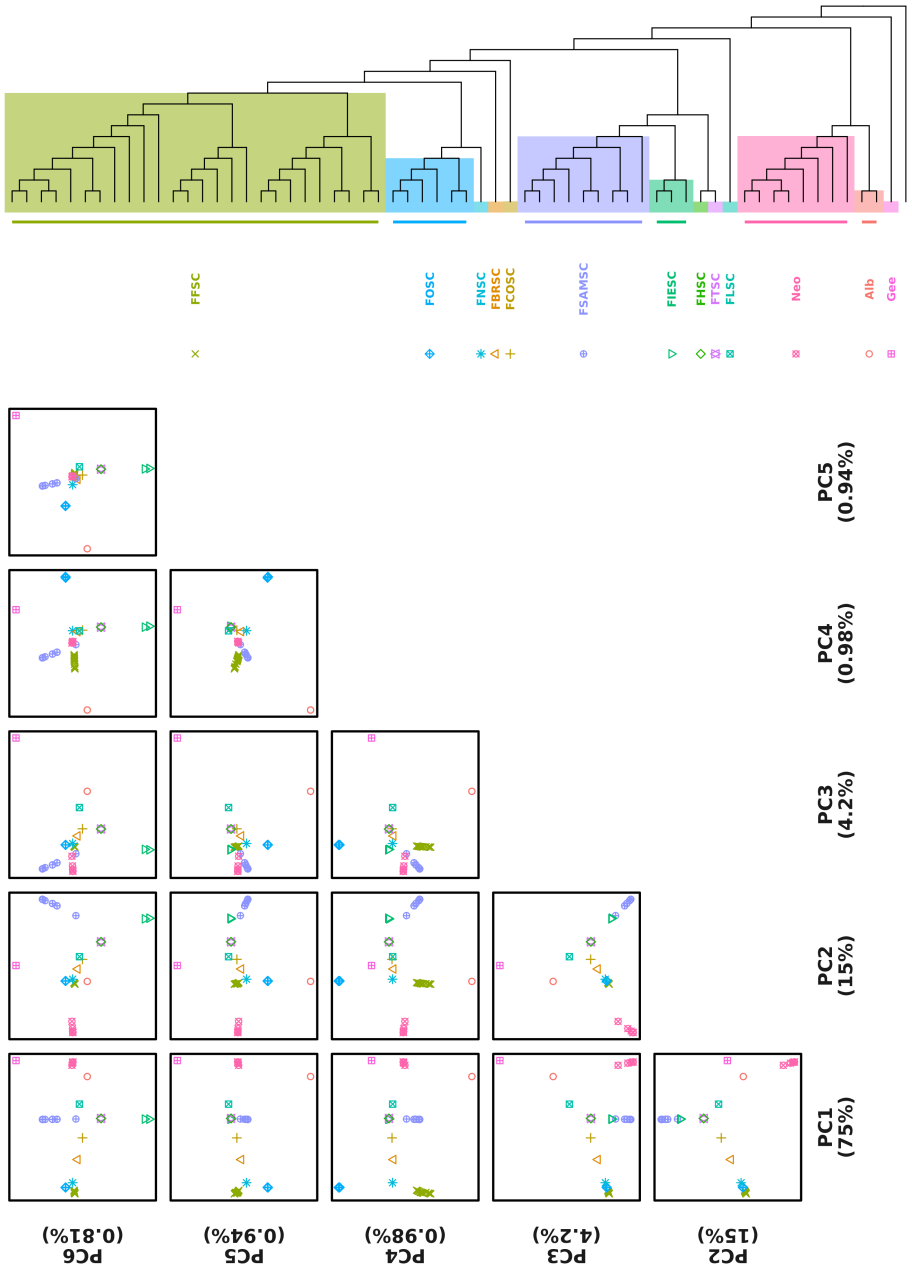
Supplementary Figure S3.2: Nx plots (the smallest contig length at which x% of the assembly is contained in contigs of at least that size) produced by QUAST for each of the strains sequenced in this chapter: **(A)** *F. chuoi* RH1 **(B)** *F. chuoi* RH3 **(C)** *F. annulatum* RH5 **(D)** *F.* sp. RH6 **(E)** *F. proliferatum* RH7.



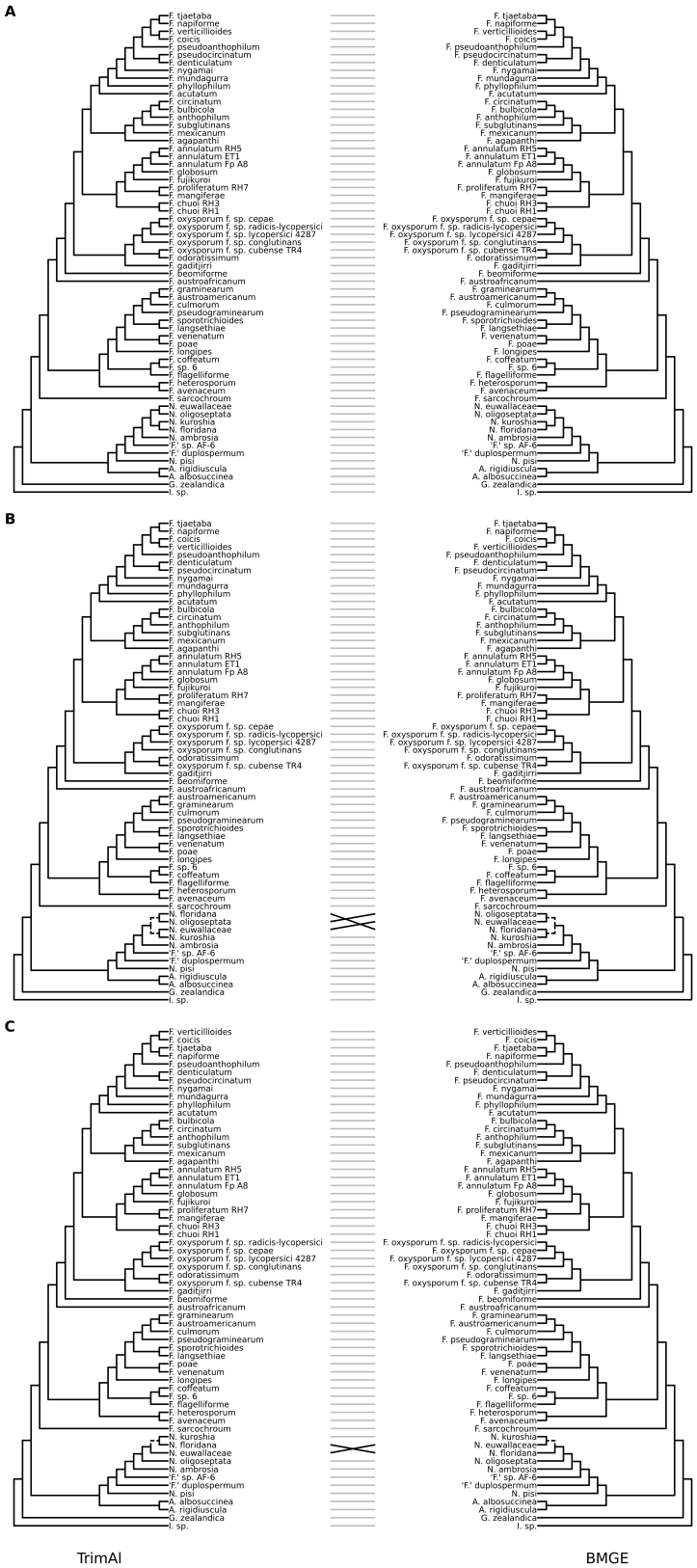
Supplementary Figure S3.3: BlobPlots showing the taxonomic classification of reads based on coverage and GC content: **(A)** *F. chuoi* RH1 **(B)** *F. chuoi* RH3 **(C)** *F. annulatum* RH5 **(D)** *F.* sp. RH6 **(E)** *F. proliferatum* RH7.



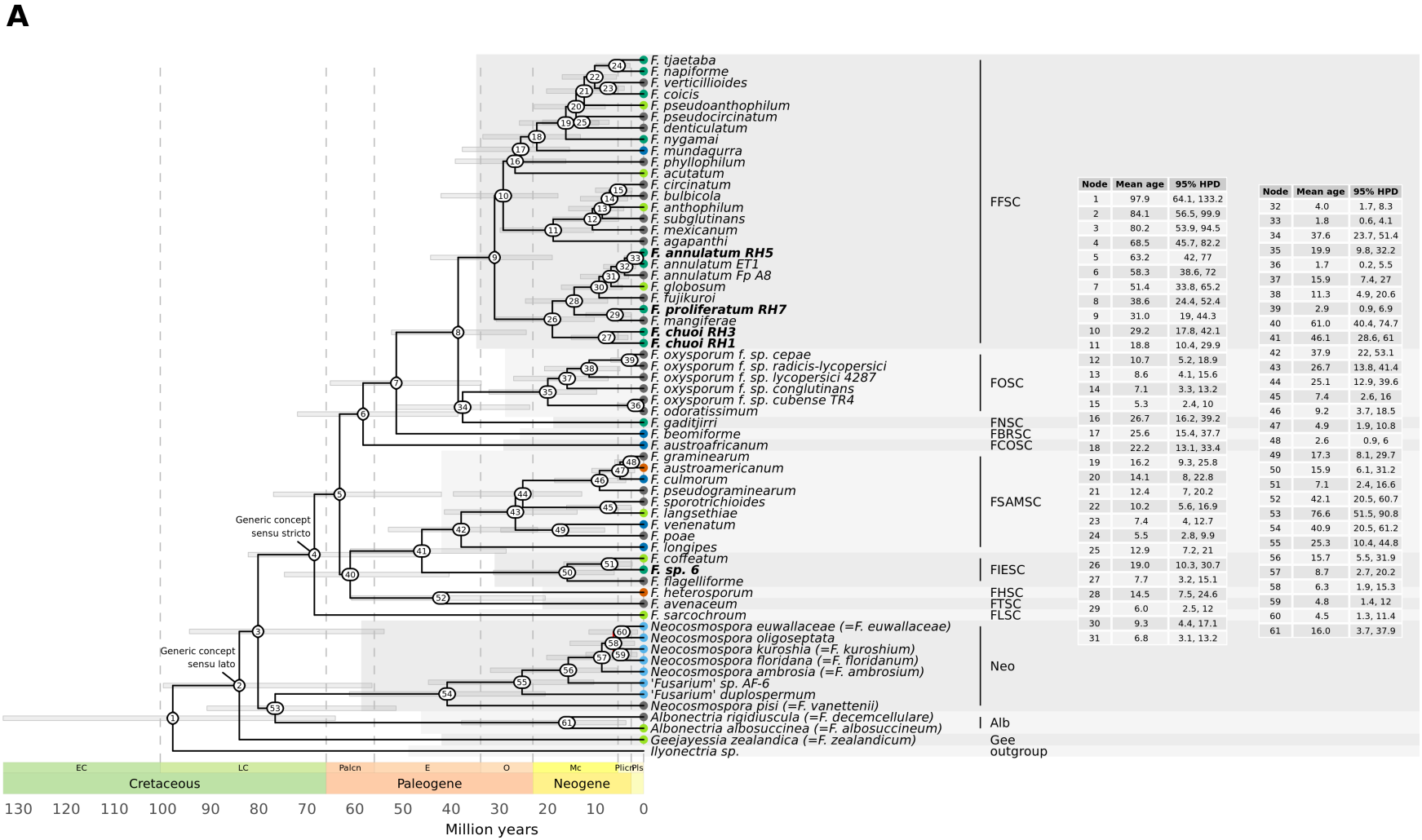
Supplementary Figure S3.4: Convergence of posterior means (left) and infinite-sites plot (right) for both MCMCTree chains for the AR clock model **(A)** and the IR clock model **(B)**.



Supplementary Figure S3.5: PCA of phylogenetic distances between taxa for the first 6 principal components, with points representing species complexes/allied genera, differentiated by shape and colour, as indicated by the tree legend. The percentage of variance explained by each principal component is shown on axis labels.



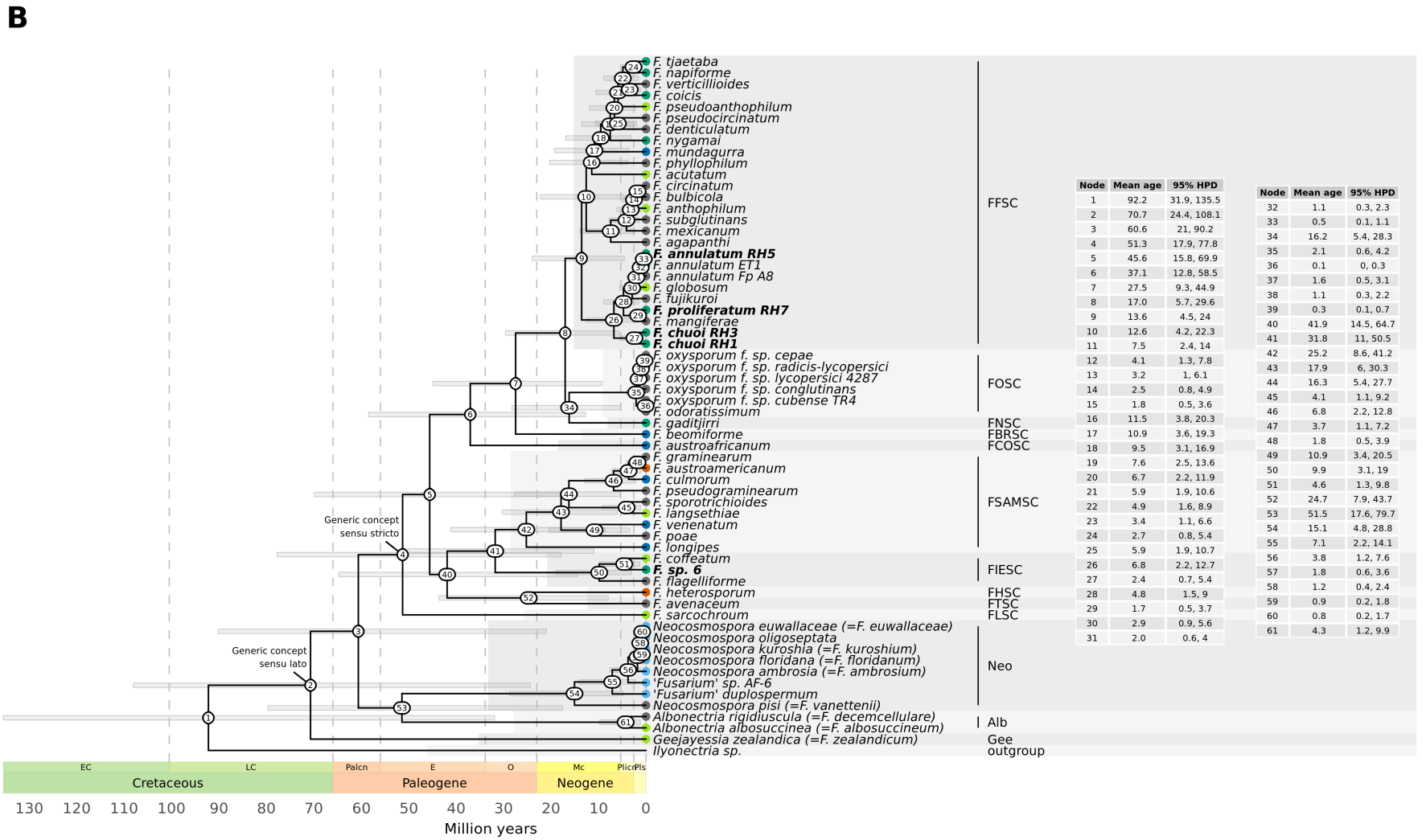
Supplementary Figure S3.6: Tanglegrams showing the difference in **(A)** RAxML-NG, **(B)** IQ-TREE and **(C)** ASTRAL-III species tree topologies when using different alignment trimming tools: TrimAl (left) and BMGE (right).



Node	Mean age	95% HPD
1	97.9	64.1, 133.2
2	84.1	56.5, 99.9
3	80.2	53.9, 94.5
4	68.5	45.7, 82.2
5	63.2	42, 77
6	58.3	38.6, 72
7	51.4	33.8, 65.2
8	38.6	24.4, 52.4
9	31.0	19, 44.3
10	29.2	17.8, 42.1
11	18.8	10.4, 29.9
12	10.7	5.2, 18.9
13	8.6	4.1, 15.6
14	7.1	3.3, 13.2
15	5.3	2.4, 10
16	26.7	16.2, 39.2
17	25.6	15.4, 37.7
18	22.2	13.1, 33.4
19	16.2	9.3, 25.8
20	14.1	8, 22.8
21	12.4	7, 20.2
22	10.2	5.6, 16.9
23	7.4	4, 12.7
24	5.5	2.8, 9.9
25	12.9	7.2, 21
26	19.0	10.3, 30.7
27	7.7	3.2, 15.1
28	14.5	7.5, 24.6
29	6.0	2.5, 12
30	9.3	4.4, 17.1
31	6.8	3.1, 13.2

Node	Mean age	95% HPD
32	4.0	1.7, 8.3
33	1.8	0.6, 4.1
34	37.6	23.7, 51.4
35	19.9	9.8, 32.2
36	1.7	0.2, 5.5
37	15.9	7.4, 27
38	11.3	4.9, 20.6
39	2.9	0.9, 6.9
40	61.0	40.4, 74.7
41	46.1	28.6, 61
42	37.9	22, 53.1
43	26.7	13.8, 41.4
44	25.1	12.9, 39.6
45	7.4	2.6, 16
46	9.2	3.7, 18.5
47	4.9	1.9, 10.8
48	2.6	0.9, 6
49	17.3	8.1, 29.7
50	15.9	6.1, 31.2
51	7.1	2.4, 16.6
52	42.1	20.5, 60.7
53	76.6	51.5, 90.8
54	40.9	20.5, 61.2
55	25.3	10.4, 44.8
56	15.7	5.5, 31.9
57	8.7	2.7, 20.2
58	6.3	1.9, 15.3
59	4.8	1.4, 12
60	4.5	1.3, 11.4
61	16.0	3.7, 37.9

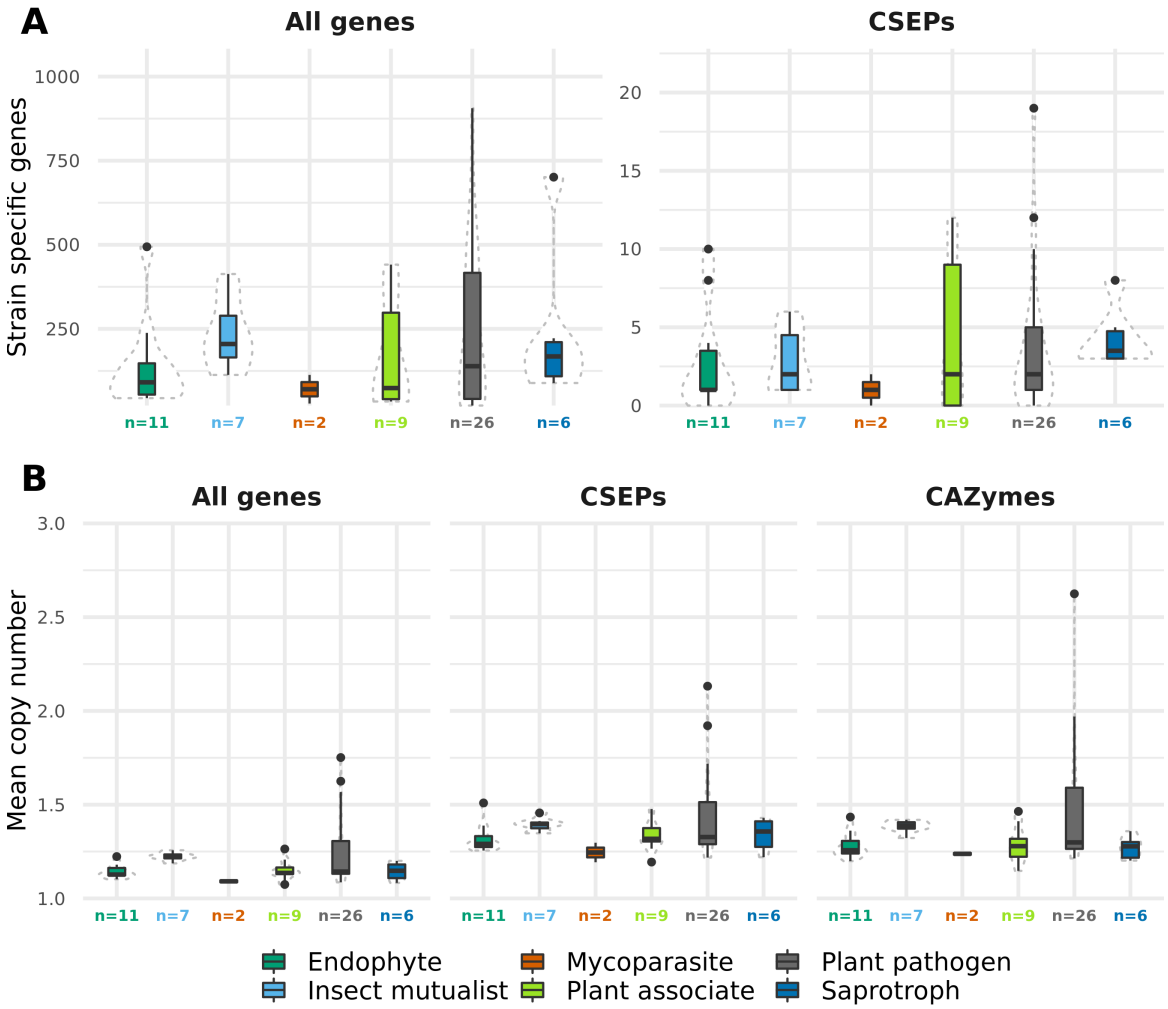
Supplementary Figure S3.7: Mean divergence times and 95% HPD confidence intervals estimated by MCMCTree for every node in the phylogeny from both the AR clock model **(A)** and the IR clock model **(B)**. ▼



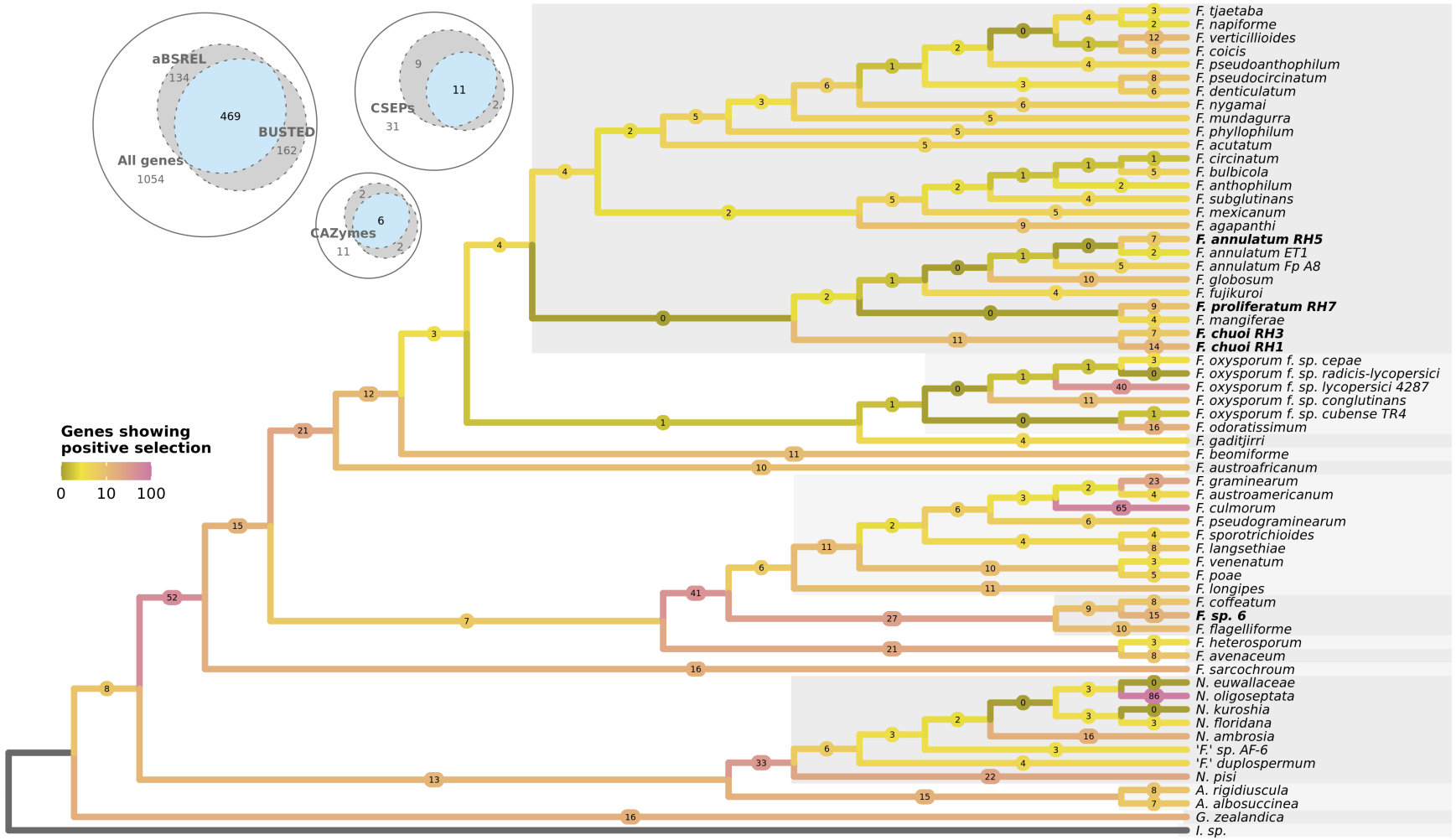
Node	Mean age	95% HPD
1	92.2	31.9, 135.5
2	70.7	24.4, 108.1
3	60.6	21, 90.2
4	51.3	17.9, 77.8
5	45.6	15.8, 69.9
6	37.1	12.8, 58.5
7	27.5	9.3, 44.9
8	17.0	5.7, 29.6
9	13.6	4.5, 24
10	12.6	4.2, 22.3
11	7.5	2.4, 14
12	4.1	1.3, 7.8
13	3.2	1, 6.1
14	2.5	0.8, 4.9
15	1.8	0.5, 3.6
16	11.5	3.8, 20.3
17	10.9	3.6, 19.3
18	9.5	3.1, 16.9
19	7.6	2.5, 13.6
20	6.7	2.2, 11.9
21	5.9	1.9, 10.6
22	4.9	1.6, 8.9
23	3.4	1.1, 6.6
24	2.7	0.8, 5.4
25	5.9	1.9, 10.7
26	6.8	2.2, 12.7
27	2.4	0.7, 5.4
28	4.8	1.5, 9
29	1.7	0.5, 3.7
30	2.9	0.9, 5.6
31	2.0	0.6, 4

Node	Mean age	95% HPD
32	1.1	0.3, 2.3
33	0.5	0.1, 1.1
34	16.2	5.4, 28.3
35	2.1	0.6, 4.2
36	0.1	0, 0.3
37	1.6	0.5, 3.1
38	1.1	0.3, 2.2
39	0.3	0.1, 0.7
40	41.9	14.5, 64.7
41	31.8	11, 50.5
42	25.2	8.6, 41.2
43	17.9	6, 30.3
44	16.3	5.4, 27.7
45	4.1	1.1, 9.2
46	6.8	2.2, 12.8
47	3.7	1.1, 7.2
48	1.8	0.5, 3.9
49	10.9	3.4, 20.5
50	9.9	3.1, 19
51	4.6	1.3, 9.8
52	24.7	7.9, 43.7
53	51.5	17.6, 79.7
54	15.1	4.8, 28.8
55	7.1	2.2, 14.1
56	3.8	1.2, 7.6
57	1.8	0.6, 3.6
58	1.2	0.4, 2.4
59	0.9	0.2, 1.8
60	0.8	0.2, 1.7
61	4.3	1.2, 9.9

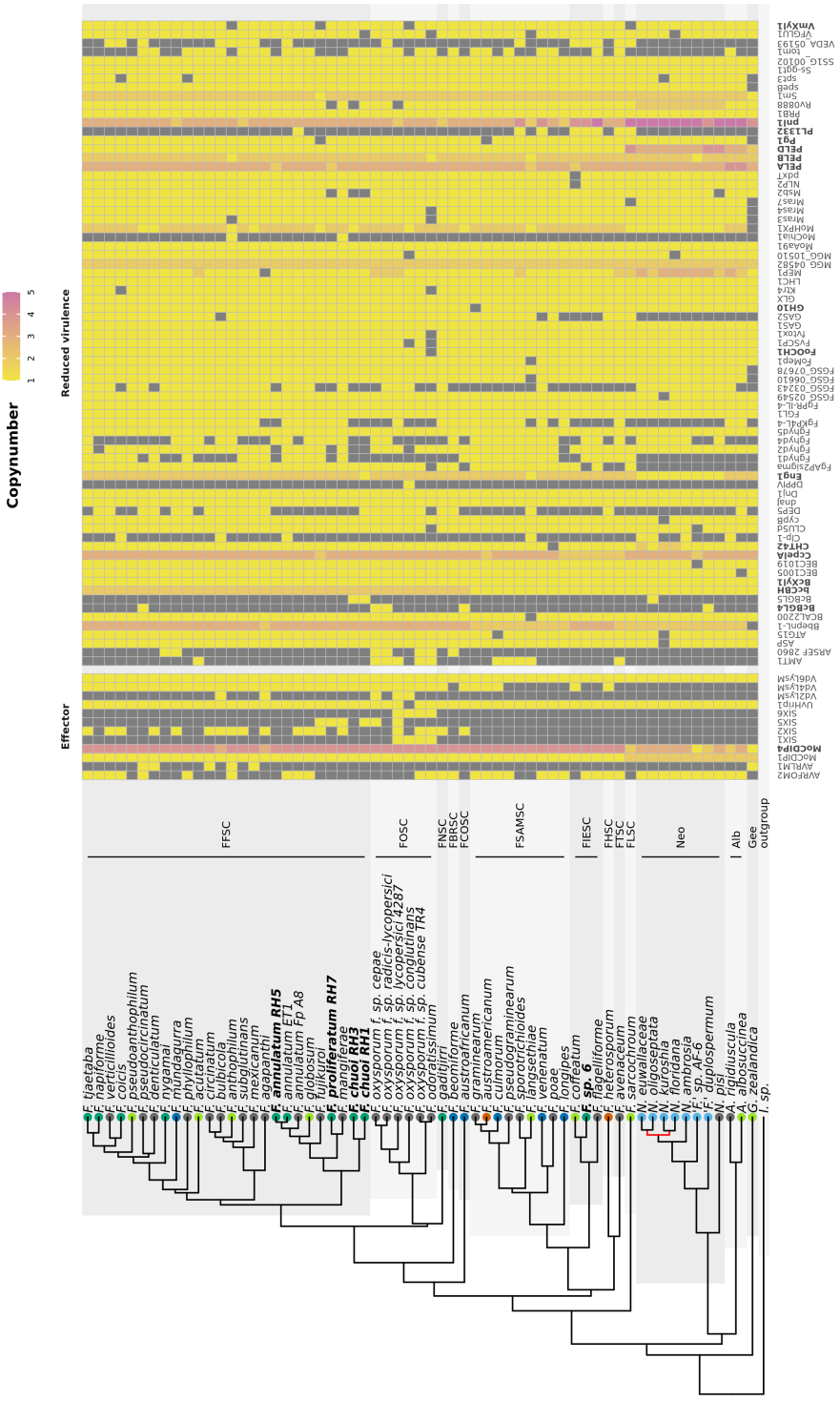
Supplementary Figure S3.7: continued.



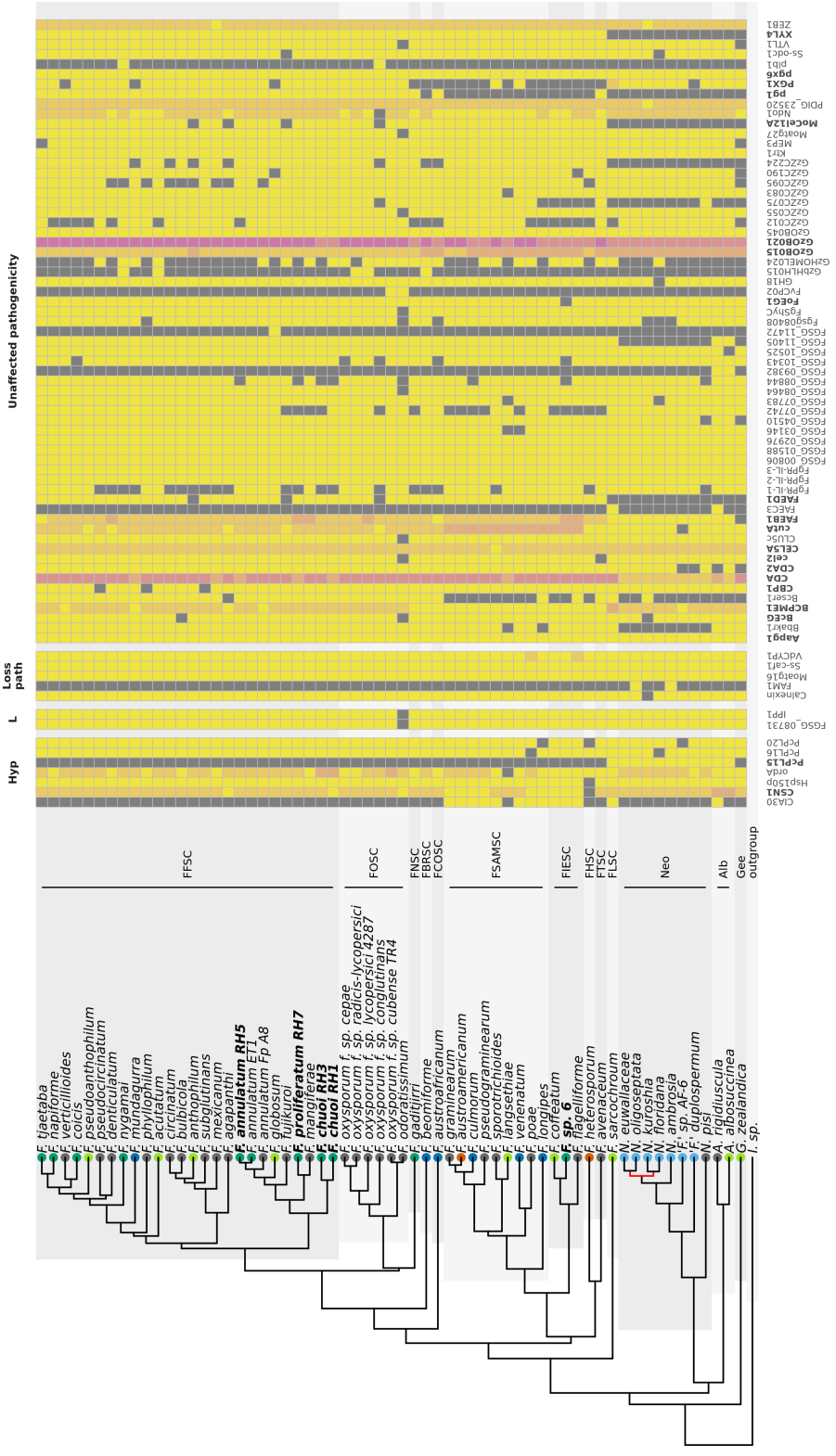
Supplementary Figure S3.8: Boxplots showing the number of strain-specific genes **(A)** and mean gene copy number **(B)** for different lifestyles. Sample size (the number of strains) is reported under x axis labels. There were no significant differences according to ANOVA (see Supplementary Table S3.2).



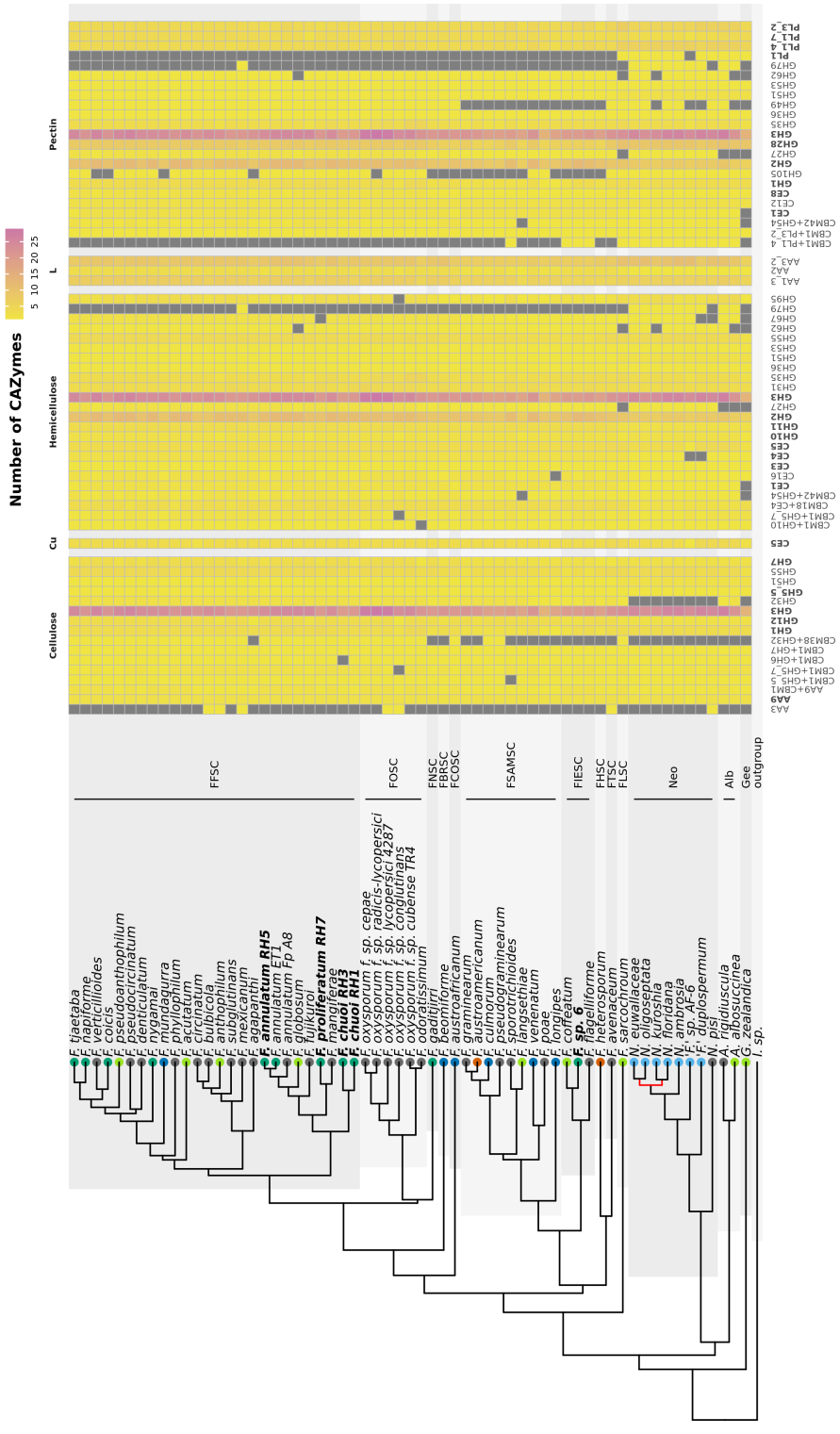
Supplementary Figure S3.9: aBSREL results showing the number of positively selected genes for every branch of the dated species tree.



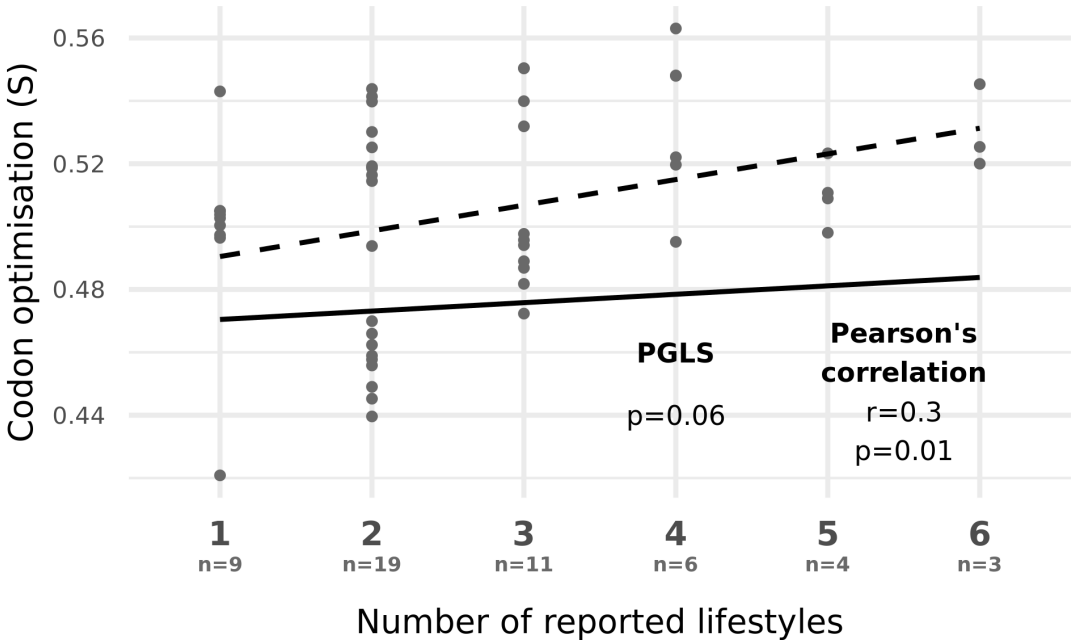
Supplementary Figure S3.10: Abundance matrix showing number of CSEPs in fusarioid taxa that could be matched to experimentally verified genes in PHI-base. Genes are grouped based on knockout mutant phenotypes curated in PHI-base (Hyp=hypervirulance, L=lethal, Loss path=loss of pathogenicity). Dark grey boxes indicate no CSEP for that taxon. Genes that are also predicted to be CAZymes. ▼



Supplementary Figure S3.10: continued.



Supplementary Figure S3.11: Abundance matrix showing number of CAZymes in fusarioid taxa belonging to families with known plant cell wall substrates (Cu=cutin, L=lignin). Dark grey boxes indicate no CAZyme family genes for that taxon. Genes that are bold were also predicted to be CSEPs.



Supplementary Figure S3.12: Scatterplot showing the relationship between codon optimisation (S) of core single-copy genes and the number of reported lifestyles for species. The dashed line indicates the best fit of uncorrected data with a linear regression model (Pearson's adj-R2=0.3, p=0.01), while the solid line indicates the phylogenetically corrected PGLS fit (p=0.06). Sample size (the number of species) is reported under x axis labels.

Chapter 4

Tapping the CABI collections for fungal endophytes: first genome assemblies for three genera and five species in the *Ascomycota*

4.1 Abstract

The *Ascomycota* form the largest phylum in the fungal kingdom and show a wide diversity of lifestyles, some involving beneficial or harmful associations with plants. Historically, whole genome sequencing (WGS) efforts have been biased towards pathogens, but improving genomic resources of commensal and mutualistic ascomycetes is fundamental if we are to fully understand plant–fungal interactions. Here, using a combination of short- and long-read technologies, we have sequenced and assembled genomes for 15 endophytic ascomycete strains from CABI's culture collections to provide valuable new resources for exploring the pathogenic–mutualistic spectrum in different lineages across the *Ascomycota*. We used phylogenetic analysis to refine the classification of taxa, which revealed that 7 of our 15 genome assemblies are the first for the genus and/or species. We also demonstrated that cytometric genome size estimates – more commonly made for plants than fungi – can act as a valuable metric for assessing assembly 'completeness', which can easily be overestimated when using BUSCOs alone. In producing these new genome resources, we emphasise the value of mining existing culture collections to produce data that can help to address major research questions relating to plant–fungal interactions.

4.2 Introduction

There is an ever mounting quantity of genomic data available for fungi and, as of October 2022, over 6,500 fungal strains had genome assemblies deposited in NCBI and MycoCosm (Chapter 1.3). Most of these genome sequencing efforts have been skewed towards pathogens and, of those, plant pathogens (Aylward et al., 2017), but recent and ongoing initiatives are rapidly increasing the num-

ber of genome assemblies available for non-pathogenic strains, such as commensal or mutualistic plant associated fungi (`https://jgi.doe.gov/our-projects/csp-plans/`; Figure 1.2). Improving genomic resources for non-pathogenic relatives of phytopathogens is key to understanding functional differences between different forms of plant associated lifestyles, and will allow us to explore how and why plant–fungal interactions evolve. This is particularly important for fungal endophytes, asymptomatic plant inhabitants which predominantly belong to the phylum *Ascomycota* (Rodriguez, White Jr et al., 2009; Hardoim et al., 2015). Factors controlling whether a fungus exhibits endophytism versus pathogenicity are not yet well defined. Case-study comparisons between closely related pathogens and endophytes – such as the one performed in Chapter 3 for the genus *Fusarium*, among others (Hacquard et al., 2016; Niehaus et al., 2016; Stauber, Prospero and Croll, 2020) – have started to reveal lineage-specific patterns or mechanisms that may contribute to lifestyle. However, we have no indication of whether they will hold true for all ascomycete endophytes, which are spread across the entire phylum (Huang, Bowman et al., 2018; U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019). If we are to better understand endophytism, and therefore improve the chance of predicting the pathogenic potential of fungal strains, comparisons across a broader taxonomic scale are needed. This is only achievable through the generation of new, high-quality genome assemblies for endophyte strains.

As described in Chapter 1.5, collections are a powerful resource for addressing all manner of research questions. It has already been demonstrated in Chapter 2 that plant collections such as Kew's Millennium Seed Bank (MSB) can act as a treasure trove for novel fungal endophyte diversity, but what of living fungal collections? The CABI collection (Egham, UK) is one of the world's largest fungal culture collections, boasting 28,000 strains spanning 100 years and 142 countries (Smith, Ryan and Caine, 2022). Access to such a wide pool of living fungal strains enables efficient data acquisition on an ambitious scale, such as helping to deliver the goal of sequencing all known species of fungi in Britain and Ireland for The Darwin Tree of Life Project (DTOL) (Smith, Kermode et al., 2020; The Darwin Tree of Life Project Consortium, 2022). Producing genomic data that links to viable fungal strains preserved in collections provides essential foundational data for future experimental and comparative research, and so increases the usefulness of accessions. Here, we capitalised on endophytic strains deposited in CABI's collection to successfully sequence, assemble and annotate genomes for 15 taxa across 8 families, 5 orders and 11 genera. For stringent quality assessment of these new genome assemblies, we additionally produced cytometric genome size estimates where possible, as recommended in Chapter 1.4.

For new genomic resources to be of use to the science community, it is of major importance to ensure accurate identification and classification of taxa. In addition to ensuring taxon names are in agreement with the current nomenclature, improving the accuracy of classifications using up-to-date molecular data is also vital. Phylogenetics has become an essential step in fungal classification, not least when dealing with cultured microfungi where morphological features are often particularly challenging to study and can be less informative, or not informative at all, for distinguishing species or even genera (Crous and Groenewald, 2005; Shivas and Cai, 2012). For the strains used here, the names from CABI's records predate the routine use of molecular data in identification, and would have been borne from morphological assessment alone (Smith, Kermode et al., 2020). Considering a third were recorded as belonging to *Phoma* – a genus which has been dismantled into numerous different genera after molecular data revealed it to be highly polyphyletic (de Gruyter et al., 2009; Aveskamp, de Gruyter, Woudenberg et al., 2010; Chen, Jiang et al., 2015; Hou et al., 2020) – incor-

porating phylogenetic analysis was essential to refine the classification of the strains sequenced here. This also revealed our assemblies to be the first for three ascomycete genera – *Collariella*, *Neodidymelliopsis* and *Neocucurbitaria* – and five species – *Ascochyta clinopodiicola*, *Didymella pomorum*, *Didymosphaeria variabile*, *Neocosmospora piperis* and *Neocucurbitaria cava*. Four more taxa – *Didymella* sp. IMI 355093, *Gnomoniopsis* sp. IMI 355080, cf. *Kalmusia* sp. IMI 367209 and *Neurospora* sp. IMI 360204 – require additional assessment to determine whether they are new or previously described species but, based on existing data, they also likely represent the first genome assemblies for their to-be-assigned species.

4.3 Materials and Methods

Extraction and sequencing of genomic DNA

The 15 endophyte strains used in this study were obtained from the CABI culture collection (Table 4.1), which uses the code 'IMI' as a prefix for its unique accessions as a relic of the now defunct Imperial Mycological Institute (`https://cabi.org/about-cabi/our-history/`. All steps involving handling of fungal material were done under sterile conditions. Strains were taken out of cryopreservation and incubated on 2% malt extract agar at 25°C for 1-2 weeks. A fragment of mycelium was transferred to flasks of 200 ml glucose yeast medium (GYM). Flasks were placed on an orbital shaker for 1 week at 25°C and shaken at 150 rpm. Mycelium was recovered via vacuum filtration, transferred to an empty petri dish and freeze dried overnight. The lyophilised material was crushed using a mortar and pestle for DNA extraction, which was done using the Qiagen DNeasy Plant Mini Kit (Qiagen, Redwood City, CA, United States) following the manufacturer's instructions. DNA concentration was quantified with a Quantus™ Fluorometer (Promega, Wisconsin, USA) and purity (260/280 absorbance ratio of approximately 1.8) was assessed with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA). To ascertain that DNA had successfully been extracted from the intended strain rather than a contaminant, 0.5 µl of DNA extraction was used for amplification and Sanger sequencing of the internal transcribed spacer (ITS) barcode, as described in Chapter 2.3. ITS sequences were searched against the UNITE database (Nilsson et al., 2019; `https://unite.ut.ee/`) and the NCBI nucleotide database (`https://ncbi.nlm.nih.gov/`) via corresponding web blastn services to identify the most similar species hypothesis (SH) for each strain. We additionally corroborated the similarity-based results by placing the ITS sequences in the 6-loci *Pezizomycotina* v2.1 reference tree (Carbone, White, Miadlikowska, Arnold, Miller, Kauff et al., 2017) of Tree-Based Alignment Selector toolkit (T-BAS) v2.3 (Carbone, White, Miadlikowska, Arnold, Miller, Magain et al., 2019) with default settings.

Table 4.1: Endophyte strains selected from CABI's collections for WGS and assembly.

IMI	CABI name	Updated name	Taxonomy	Host	Origin
355080	*Phomopsis*	*Gnomoniopsis* sp.	*Gnomoniaceae*, *Diaporthales*, *Sordariomycetes*	*Quercus ilex*	Lugano, Switzerland
355082	*Phomopsis*	*Gnomoniopsis smithogilvyi*	*Gnomoniaceae*, *Diaporthales*, *Sordariomycetes*	*Quercus ilex*	Lugano, Switzerland
355084	*Colletotrichum acutatum*	*Colletotrichum fioriniae*	*Glomerellaceae*, *Glomerellales*, *Sordariomycetes*	*Quercus ilex*	Lugano, Switzerland
355091	*Phoma sorghina*	*Didymella pomorum*	*Didymellaceae*, *Pleosporales*, *Dothideomycetes*	*Opuntia* sp.	Queensland, Australia
355093	*Phoma*	*Didymella* sp.	*Didymellaceae*, *Pleosporales*, *Dothideomycetes*	*Opuntia* sp.	Queensland, Australia
356814	*Phoma leveillei*	*Neocucurbitaria cava*	*Cucurbitariaceae*, *Pleosporales*, *Dothideomycetes*	*Quercus ilex*	Mallorca, Spain
356815	*Leptosphaeria coniothyrium*	*Didymosphaeria variabile*	*Didymosphaeriaceae*, *Pleosporales*, *Dothideomycetes*	*Quercus ilex*	Mallorca, Spain
359910	*Phoma*	*Ascochyta clinopodiicola*	*Didymellaceae*, *Pleosporales*, *Dothideomycetes*	*Dryas octopetala*	Switzerland
360193	*Microsphaeropsis*	*Didymella glomerata*	*Didymellaceae*, *Pleosporales*, *Dothideomycetes*	*Gynoxis oleifolia*	Ecuador
360204	*Gelasinospora*	*Neurospora* sp.	*Sordariaceae*, *Sordariales*, *Sordariomycetes*	*Gynoxis oleifolia*	Ecuador
364377	*Phoma nebulosa*	*Neodidymelliopsis* sp.	*Didymellaceae*, *Pleosporales*, *Dothideomycetes*	*Persea americana*[1]	Trinidad and Tobago
366226	*Colletotrichum crassipes*	*Colletotrichum tropicale*	*Glomerellaceae*, *Glomerellales*, *Sordariomycetes*	*Manilkara bidentata*	Puerto Rico
366227	*Colletotrichum crassipes*	*Collariella* sp.	*Chaetomiaceae*, *Sordariales*, *Sordariomycetes*	*Manilkara bidentata*	USA[2]
366586	*Fusarium solani*	*Neocosmospora piperis*	*Nectriaceae*, *Hypocreales*, *Sordariomycetes*	*Manilkara bidentata*	Puerto Rico
367209	*Leptosphaeria coniothyrium*	cf. *Kalmusia* sp.	*Didymosphaeriaceae*, *Pleosporales*, *Dothideomycetes*	*Manilkara bidentata*	Puerto Rico

[1]Isolated as endophyte of leaves imported by leafcutter ants into their nests.
[2]Suspected input error based on adjacent IMI records.

For short-read Illumina sequencing, DNA extractions were sent to Macrogen (Macrogen Inc., South Korea) for library preparation and sequencing: library preparation was performed using the Nextera XT DNA Library Preparation Kit and 151 bp paired-end reads were sequenced using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA). If we were able to extract ≥1 µg of DNA, strains were also processed for long-read nanopore sequencing. For each strain, the appropriate volume for 1 µg of DNA was diluted with sterile, nuclease-free water to obtain the required 47 µl of DNA for the library preparation method described here. Half of the DNA solution (23.5 µl) was then sheared to a fragment size of ∼20 Kbp by centrifuging in a g-TUBE (Covaris, Inc., Woburn, MA, USA) at 4,200 rpm for 1 minute. Sequencing libraries were prepared from the mixture of sheared and unsheared DNA using the SQK-LSK109 Ligation Sequencing Kit (Oxford Nanopore Technologies Inc., Oxford, UK) following the manufacturer's Genomic DNA by Ligation protocol (version GDE_9063_v109_revAE_14Aug2019). The Short Fragment Buffer was used during the clean-up step to purify all fragments equally. DNA repair and end-prep was performed using the NEBNext FFPE DNA Repair and Ultra II End Repair/dA-Tailing modules (New England BioLabs, Ipswich, MA, USA). The library was loaded into a FLO-MIN106 flow cell and sequenced with a MinION device (Oxford Nanopore Technologies Inc.) for ∼48 hours using the MinKNOW application (Oxford Nanopore Technologies Inc.). Fast basecalling was performed after sequencing using guppy v4.5.3 (Oxford Nanopore Technologies Inc.).

Flow cytometry

Where possible, cultures were additionally sampled for flow cytometry 10-56 days after subculturing depending on the growth rate of the sample. Two different fungal strains were used as internal calibration standards to estimate the genome sizes of the endophyte strains. The first internal fungal standard was a strain of *Coprinellus micaceus* which had been isolated and cultured from a collection made by R. Wright on 05/10/2020 at Royal Botanic Gardens Kew, UK (culture code: FTOL_0141). The genome size of *C. micaceus* was estimated directly by co-running a sample with *Arabidopsis thaliana* (L.) Heynh., 1842 (ecotype col-0 NASC) with an estimated genome size of 172.44 Mbp/1C. The sample was prepared for flow cytometry following the One-Step Protocol using LB01 buffer, as outlined by Pellicer, Powell and Leitch (2020): *C. micaceus* mycelium was co-chopped with 1 cm^2 fresh *A. thaliana* leaf tissue in a petri dish with 1 ml of LB01 buffer (Doležel, Binarová and Lucretti, 1989). A further 1 ml of LB01 was added to the sample and the contents gently mixed. The sample was then passed through a 30 µm nylon filter, stained with 100 µl propidium iodide (1 mg/ml) and incubated on ice for 10 minutes before running through a Sysmex CyFlow Space flow cytometer (Sysmex Partec GmbH, Görlitz, Germany) fitted with a 100 mW green solid state laser (532 nm, Cobolt Samba, Solna, Sweden). Each isolate was run through the flow cytometer three times to ensure reproducibility of results, with at least 1,000 nuclei analysed each time. Once the genome size of *C. micaceus* had been estimated (62.62 Mbp/1C) it was then used to calibrate a second internal standard, *Coprinopsis piacea* (52.83 Mbp/1C), which was isolated and cultured from a collection that had been made by R. Wright on 17/12/2020 at Royal Botanic Gardens Kew, UK (culture code: FTOL_0189). Preparation of each endophyte sample for flow cytometry was then completed following the same process as above, except using one of the two internal fungal standards instead of *A. thaliana*.

We used the Partec FloMax v2.4d software (Sysmex Partec GmbH) to produce histograms showing the relative fluorescence of nuclei (Supplementary Figure S4.1). FlowMax gating tools were used

to generate linear regressions to gate nuclei and quantify the number of nuclei and coefficient of variation (CV) of each peak. A polygonal region was drawn around the nuclei in the side scatter cytogram to improve the quality of the peaks by ensuring only intact nuclei were analysed. The measurement of DNA content for each isolate was considered reliable only if the CV value of the G_1 peak was below the accepted limit of 10% for fungi (Bourne et al., 2014). The holoploid 1C genome size of each strain was estimated using the following formula:

$$\frac{\text{Mean } G_1 \text{ fluorescence peak of sample} \times \text{1C nuclear DNA content of reference standard}}{\text{Mean } G_1 \text{ fluorescence peak of reference standard}}$$

Genome size in Mbp was calculated using the conversion factor 1 pg = 978 Mbp (Doležel, Bartoš et al., 2003).

De novo genome assembly

Our bioinformatics analysis pipeline is summarised in Supplementary Figure S4.2. For strains which only had short-read data, the same assembly pipeline was used as in Chapter 3.3, comparing ABySS v2.0.2 (Simpson et al., 2009), MEGAHIT v1.2.9 (Li, Luo et al., 2016) and SPAdes v3.11.1 (Bankevich et al., 2012). If we were also able to obtain long-read sequence data for strains, hybrid assembly was performed with comparison across three tools: Flye v2.6 (Kolmogorov et al., 2019), Raven v1.6.1 (Vaser and Šikić, 2021) and hybridSPAdes v3.11.1 (Antipov et al., 2016). The former two methods involved assembly using only the raw long-reads, before mapping the short-reads onto the resulting contigs using BWA-MEM v0.7.17-r1188 (Li, 2013) in order to polish with Pilon v1.2.4 (Walker et al., 2014). In contrast, hybridSPAdes used both long and short-reads to construct contigs, before similarly polishing with the short-reads using BWA-MEM and Pilon. For Flye, which requires an estimate of total genome size, cytometric genome size estimates described above were used where possible, otherwise the average genome size for the order from the analysis in Chapter 1.3 was used.

Quality assessment and contaminant removal

To select the 'best' assembly across the different assembly tools, contiguity was assessed using QUAST v5.0.2 (Gurevich et al., 2013) and completeness was assessed with BUSCO v3.0.1 (Simão et al., 2015) using the ascomycota_odb10.2020-09-10 lineage dataset of 1,706 single-copy orthologues. BlobTools v1.1 (Laetsch and Blaxter, 2017) was used to check for possible contamination in the best assemblies. To create hit files, contigs were searched against the UniRef90 database (Suzek et al. (2015); downloaded on 9$^{\text{th}}$ August 2022) using DIAMOND v2.0.15.153 (Buchfink, Reuter and Drost, 2021) and against the NCBI nucleotide database (downloaded on 17$^{\text{th}}$ August 2022) using BLAST+ v2.11.1 (Camacho et al., 2009). To create BAM files of mapped reads, long-reads were mapped back onto hybrid assemblies using minimap2 v2.5 (Li, 2018), while short-reads were mapped back onto short-read assemblies using BWA-MEM v0.7.17-r1188 (Li, 2013). Hit and BAM files were then used by BlobTools to create order-level BlobPlots. Contigs that were not assigned to orders in the correct class – as expected from the original identification by CABI and barcoding – and contigs with a coverage of less than 10× were removed from assemblies using seqtk v1.2-r94 (`https://github.com/lh3/seqtk`). Mitochondrial and adapter contamination flagged by NCBI during the assembly submission process was trimmed using bedtools v2.28.0 (Quinlan and Hall, 2010). QUAST and BUSCO were then run again on the contamination-filtered assemblies to produce final quality statistics. Mean short-read coverage was calculated by once again mapping

short-reads onto contaminant-filtered assemblies with BWA-MEM and using the stats option from SAMtools v1.9 (Li, Handsaker et al., 2009) to get the number of mapped bases, which was then divided by the total assembly length. The same approach was used for long-read coverage of hybrid assemblies, excepting the use of minimap2 in place of BWA-MEM. Assembly contiguity was visualised as snail plots using BlobToolKit v3.4.0 (Challis et al., 2020).

Assembly annotation

A *de novo* repeat library was generated for the selected assembly for each strain with RepeatModeler v2.0.1 (Smit and Hubley, 2015) and used as a custom library for softmasking with RepeatMasker v4.0.9 (Smit, Hubley and Green, 2015). Masked assemblies were structurally annotated using the Funannotate v1.8.12 pipeline (Palmer and Stajich, 2020). We used the funannotate sort command to sort and relabel contigs in preparation for annotation. Proteins and EST clusters of closely related taxa were downloaded from MycoCosm (`https://mycocosm.jgi.doe.gov/`; Grigoriev et al., 2014) to inform gene prediction: *Gnomoniopsis castanea* Behrend (Gnocas1) for IMI 355080 and IMI 355082 (unpublished); *Colletotrichum somersetensis* CBS 131599 (Colso1) for IMI 355084 and IMI 366226 (Baroncelli, Cobo-Díaz et al., 2022); *Didymella exigua* CBS 182.55 (Didex1) for IMI 355091, IMI 355093, IMI 359910, IMI 360193 and IMI 364377 (Haridas et al., 2020); *Pyrenochaeta* sp. MPI-SDFR-AT-0127 (Pyrly1) for IMI 356814 (Mesny, Miyauchi et al., 2021); *Bimuria novae-zelandiae* CBS 107.79 (Bimnz1) for IMI 356815 and IMI 367209 (Haridas et al., 2020); *Neurospora crassa* 73 trp-3 (Neucr_trp3_1) for IMI 360204 (Baker et al., 2015); *Chaetomium globosum* MPI-SDFR-AT-0079 (Chagl1) for IMI 366227 (Mesny, Miyauchi et al., 2021); and *Fusarium solani* FSSC 5 MPI-SDFR-AT-0091 (Fusso1) for IMI 366586 (Mesny, Miyauchi et al., 2021). We used the funannotate predict command to train and run three *ab initio* gene predictors – AUGUSTUS v3.3.2 (Stanke et al., 2006), GlimmerHMM (Majoros, Pertea and Salzberg, 2004) and SNAP v2006-07-28 (Korf, 2004) – and output consensus gene models according to EVidenceModeler v1.1.1 (Haas et al., 2008).

Functional prediction of the gene models was performed with InterProScan v5.57-90.0 (Jones et al., 2014) using the applications CDD v3.18 (Lu et al., 2020), Coils v2.2.1 (Lupas, 1997), Gene3D v4.3.0 (Lees et al., 2012), Hamap v2021_04 (Pedruzzi et al., 2015), MobiDBLite v2.0 (Necci et al., 2017), PANTHER v15.0 (Mi et al., 2019), Pfam v35.0 (Mistry, Chuguransky et al., 2021), Phobius v1.01 (Käll, Krogh and Sonnhammer, 2004), PIRSF v3.10 (Wu et al., 2004), PRINTS v42.0 (Attwood et al., 2012), SFLD v4 (Akiva et al., 2014), SignalP v4.1 (Nielsen, 2017), SMART v7.1 (Letunic, Doerks and Bork, 2012), SUPERFAMILY v1.75 (Gough et al., 2001), TIGRFAM v15.0 (Haft et al., 2001) and with mapping to gene ontology terms. Gene models were additionally functionally annotated using eggNOG-mapper v2.1.9-4dfcbd5 (Cantalapiedra et al., 2021) – based on the eggNOG orthology database v5.0.2 (Huerta-Cepas et al., 2019) with sequence searches using DIAMOND v2.0.15 – and using antiSMASH v6.1.1 (Blin et al., 2021). The funannotate annotate command was then used to map the InterProScan and eggNOG results onto the assembly annotations, with additional searches against UniProt v2022_02 (Bateman et al., 2021), MEROPS v12 (Rawlings, Barrett and Bateman, 2012), dbCAN v10.0 (Yin et al., 2012) and BUSCO dikarya gene models. Misannotations that were flagged by NCBI during the assembly submission process were checked and manually edited.

Phylogenetic analysis

Using our results from UNITE, NCBI and T-BAS (Supplementary Figure S4.3) to guide taxon sampling, we searched the literature for existing phylogenies and available genetic marker sequences

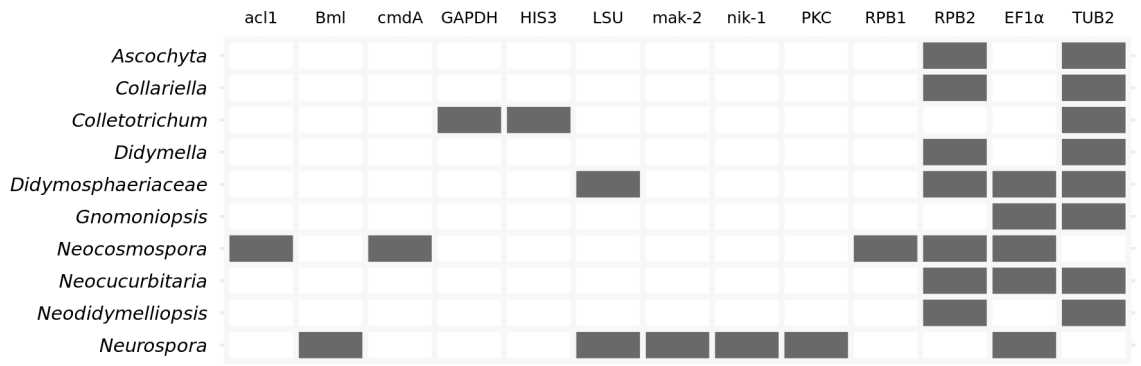


Figure 4.1: Summary of the genetic markers used for each lineage in the phylogenetic analyses.

for the different lineages to which our samples potentially belonged (Nygren et al., 2011; Wang, Houbraken et al., 2016; Wang, Han et al., 2022; Chen, Hou et al., 2017; Wanasinghe, Phookamsak et al., 2017; Crous, Schumacher et al., 2019; Crous, Lombard et al., 2021; Jaklitsch et al., 2018; Valenzuela-Lopez, Cano-Lira, Guarro et al., 2018; Hyde, Tennakoon et al., 2019; Hou et al., 2020; Scarpari et al., 2020; Vieira et al., 2020; Jiang et al., 2021; Karácsony et al., 2021; Liu, Ma et al., 2022; Wanasinghe and Mortimer, 2022). Various combinations of 13 genetic markers were selected for the different lineages (Figure 4.1), sequences for which were retrieved from GenBank – accession numbers for all taxa used in the phylogenetic analysis can be seen in Supplementary Table S4.1. A new script, GenePull (`https://github.com/Rowena-h/MiscGenomicsTools/tree/main/GenePull`), was created to extract sequences for each of the selected markers from our own genome assemblies.

We aligned each gene separately for the different lineages using MAFFT v7.480 (Katoh and Standley, 2013) and manually checked the gene alignments before trimming using trimAl v1.4.rev15 (Capella-Gutiérrez, Silla-Martínez and Gabaldón, 2009) with the -gappyout option. As multiple nuclear ribosomal large subunit (LSU) copies were extracted from the *Didymoshaeriaceae* assemblies, all of the copies were included in the *Didymoshaeriaceae* LSU alignment. A gene tree was estimated for the LSU alignment using RAxML-NG v1.0.1 (Kozlov et al., 2019) and the GTR+GAMMA model of evolution. After confirming that all copies clustered together on the LSU gene tree (Supplementary Figure S4.4), the longest sequence was selected as a representative to be included in the concatenated dataset alongside other single-copy markers. Trimmed single-copy gene alignments were concatenated using AMAS v0.98 (Borowiec, 2016) and the concatenated alignment for each lineage was run in RAxML-NG with genes partitioned and the GTR+GAMMA model of evolution.

All results were plotted in R v4.1.1 using the following packages: ape v5.5 (Paradis and Schliep, 2019), ggplot2 v3.3.5 (Wickham, 2016), ggpubr v0.4.0 (Kassambara, 2020), ggtree v3.0.4 (Yu et al., 2017) and tidyverse v1.3.2 (Wickham, Averick et al., 2019). R scripts were written using RStudio v2021.09.1+372 (RStudio Team, 2015). This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (Butcher, King and Zalewski, 2017). Scripts of all analyses are available at `https://github.com/Rowena-h/EndophyteGenomes`. New WGS data and annotated genome assemblies reported here are available on GenBank under the BioProject accession PRJNA786750.

4.4 Results

SPAdes was the optimal short-read assembly tool, but optimal hybrid assembly tool varied between Flye and Raven

For the eight short-read assemblies, SPAdes consistently produced assemblies with the best contiguity and completeness statistics compared to ABySS and MEGAHIT (Supplementary Figure S4.5, Supplementary Table S4.2). For the seven hybrid assemblies, however, hybridSPAdes resulted in markedly worse contiguity than either Flye or Raven – in the most extreme case the assembly for IMI 366227 had an N50 value ∼50 times smaller than the next best assembler (Supplementary Table S4.2). Despite comparatively poor contiguity, hybridSPAdes still produced assemblies with a similar level of completeness according to BUSCOs. There was little difference in the performance of Flye and Raven, although Raven produced the 'best' assembly for five out of seven strains (Table 4.2).

Despite originating from axenic cultures, we still detected some contaminant contigs that were removed from the assemblies. The majority of contaminants (defined here as any contigs assigned to a different taxonomic class according to BlobTools) belonged to other ascomycete fungi, although there was also some bacterial contamination found (Supplementary Figure S4.6). These contigs generally represented a small proportion of the assemblies, however, in two cases a considerable proportion of the assembly was filtered out: 19% for IMI 360204 and 12% for IMI 355082 (Table 4.2). Hybrid assemblies were less fragmented, with the largest fragments constituting between ∼6-20% of the total assembly length (versus <3% for short-read assemblies) and N50/N90 values at least one order of magnitude greater than the short-read assemblies (Figure 4.2A,B).

Flow cytometry revealed some assemblies to be less complete than BUSCOs would suggest

Genome size measurements were successfully obtained for five of the strains using flow cytometry (Table 4.3). For these strains we were able to compare total assembly length against cytometric genome size estimates, which revealed that most assemblies were notably smaller than the 'true' genome size (Figure 4.2C). This was despite assemblies having a high percentage of single-copy BUSCOs, meaning that completeness according to BUSCOS was much higher than completeness according to cytometric genome size estimates (Figure 4.2D). The exception was strain IMI 355093 (*Didymella* sp.), for which the total assembly length and the cytometric genome size measurement were very similar and thus the assembly was estimated to be highly complete according to cytometric measurements as well as BUSCOs (Figure 4.2D).

Table 4.2: Statistics for the 'best' short-read or hybrid assembly for each of the 15 endophyte strains after contaminant filtering.

						QUAST					BUSCO	Funannotate
	IMI	Tool	Coverage (SR)	Coverage (LR)	Contamination (bp removed)	# contigs ≥500bp	Largest contig (bp)	Total length (bp)	GC (%)	N50	Completeness (%)	# genes
Short-read	355080	SPAdes	112×	-	2,983,631 (7.3%)	694	423,323	38,082,340	51.69	127,272	93.14	10,907
	355091	SPAdes	252×	-	1,038,387 (2.9%)	524	908,435	34,416,163	53.52	218,427	98.94	11,427
	359910	SPAdes	139×	-	885,076 (2.6%)	1,199	259,290	33,614,440	52.55	73,892	97.48	10,203
	360193	SPAdes	253×	-	793,957 (2.3%)	724	641,373	34,727,068	53.46	179,824	98.42	10,766
	360204	SPAdes	122×	-	8,191,841 (18.5%)	3,250	166,179	36,929,578	52.64	25,999	95.25	10,020
	364377	SPAdes	186×	-	465,400 (1.5%)	1,103	382,275	30,047,231	51.51	74,885	98.01	9,755
	366226	SPAdes	86×	-	655,070 (1.2%)	1,685	305,111	54,633,813	53.59	63,560	96.42	13,995
	366586	SPAdes	116×	-	942,977 (2.2%)	1,248	470,694	41,415,286	52.32	91,570	96.31	12,790
Hybrid	355082	Flye	113×	44×	4,904,540 (12.2%)	9	7,084,357	35,292,834	50.70	6,429,383	86.64	10,375
	355084	Flye	193×	20×	32,782 (0.1%)	45	7,342,820	49,445,812	51.93	2,983,733	98.07	12,178
	355093	Raven	300×	138×	0 (0.0%)	27	1,884,042	31,528,740	52.85	1,301,886	98.65	9,918
	356814	Raven	160×	165×	0 (0.0%)	24	2,991,912	34,846,001	50.24	1,616,366	98.30	11,048
	356815	Raven	212×	216×	0 (0.0%)	11	5,345,287	39,450,705	51.25	4,705,368	98.12	12,728
	366227	Raven	316×	39×	278,263 (0.9%)	49	2,828,572	29,308,369	55.80	1,760,284	87.92	8,224
	367209	Raven	208×	27×	24,662 (0.1%)	30	4,380,344	42,784,582	49.69	2,200,773	98.18	13,561

Table 4.3: Flow cytometry genome size estimation results. * = *Coprinopsis piacea*, † = *Coprinellus micaceous*. Cytometric completeness = (genome size (Mbp/1C) / assembly length (Mbp)) × 100.

IMI	Mean G_1 peak		Mean CV		Genome size (pg/1C)	Genome size (Mbp/1C)	Cytometric completeness (%)
	Sample	Standard	Sample	Standard			
355093	150.09	242.92*	7.26	3.59*	0.033	32.03	98.44
356814	241.55	341.96†	5.03	3.42†	0.045	44.21	77.08
359910	233.33	346.14†	4.98	5.20†	0.043	42.19	77.92
360204	324.06	412.41†	4.16	5.07†	0.050	49.18	75.09
364377	175.70	412.41†	6.29	5.55†	0.040	39.55	75.95

Phylogenetic analyses classified strains as belonging to 11 genera, with 9 strains resolved to species-level

The endophyte strains were divided equally amongst the classes *Dothideomycetes* and *Sordariomycetes*. Of the former, all taxa fell in the order *Pleosporales*, with the majority belonging to so-called 'phoma-like' genera. Five strains were placed in the family *Didymellaceae*, three of these being *Didymella* spp.: IMI 355091 and IMI 360193 were resolved with significant bootstrap support as the species *D. pomorum* and *D. glomerata*, respectively (Figure 4.3A). IMI 355093 was confidently placed in a clade with *D. longicolla*, *D. dimorpha* and *D. boeremae*. The final two *Didymellaceae* taxa were IMI 359910 – resolved as *Ascochyta clinopodiicola* (Figure 4.3B) – and IMI 364377 – a *Neodidymelliopsis* species which clustered, albeit with poor support, alongside *Neod. sambuci* and an unidentifed *Neod.* species (Figure 4.3C).

The second most common pleosporalean family amongst the strains studied here was the *Didymosphaeriaceae*. IMI 356815 was resolved with significant support as *Didymosphaeria variable* (Figure 4.3D). The placement of IMI 367209 within the *Didymosphaeriaceae* was more ambiguous, as it fell within a poorly support clade alongside *Kalmusia erioi* and *Kalmusia cordylines*, but the genus *Kalmusia* was not resolved monophyletically (Figure 4.3D), and so the strain has been conservatively dubbed here as 'cf. *Kalmusia* sp.'. The placement of multi-copy LSU genes for the *Didymosphaeriaceae* corroborated the phylogenetic placement that was found by the concatenated species tree analyses (Supplementary Figure S4.4). The final pleosporalean taxon was IMI 356814, which was significantly resolved as *Neocucurbitaria cava* in the family *Cucurbitaceae* (Figure 4.3E).

Of the sordariomycete taxa, two were found to belong to the genus *Gnomoniopsis* (*Gnomoniaceae*, *Diaporthales*): IMI 355082 was confidently resolved as *G. smithogilvyi*, whilst IMI 355080 formed a distinct lineage sister to *G. paraclavulata*, which were together sister to *G. clavulata* (Figure 4.3F). Two strains were placed in the genus *Colletotrichum* (*Glomerellaceae*, *Glomerellales*): IMI 366226 was significantly resolved as *Colle. tropicale* in the *Colle. gloeosporioides* species complex, whilst IMI 355084 was significantly resolved as *Colle. fioriniae* in the *Colle. acutatum* species complex (Figure 4.3G). IMI 366227 was confidently placed in the genus *Collariella* (*Chaetomiaceae*, *Sordariales*), most closely related to *Colla. pachypodioides* and *Colla. carteri* (Figure 4.3H).

IMI 360204 was confidently placed in the genus *Neurospora* (*Sordariaceae*, *Sordariales*), although within a poorly resolved clade including *Neu. retispora*, *Neu. santi-florii* and *Neu. novoguineensis* (Figure 4.3I). Finally, IMI 366586 was resolved with significant support as *Neocosmospora piperis* (*Nectriaceae*, *Hypocreales*) (Figure 4.3J).

From all the reassessed strains, three were assigned names with accuracy to genus-level in CABI's

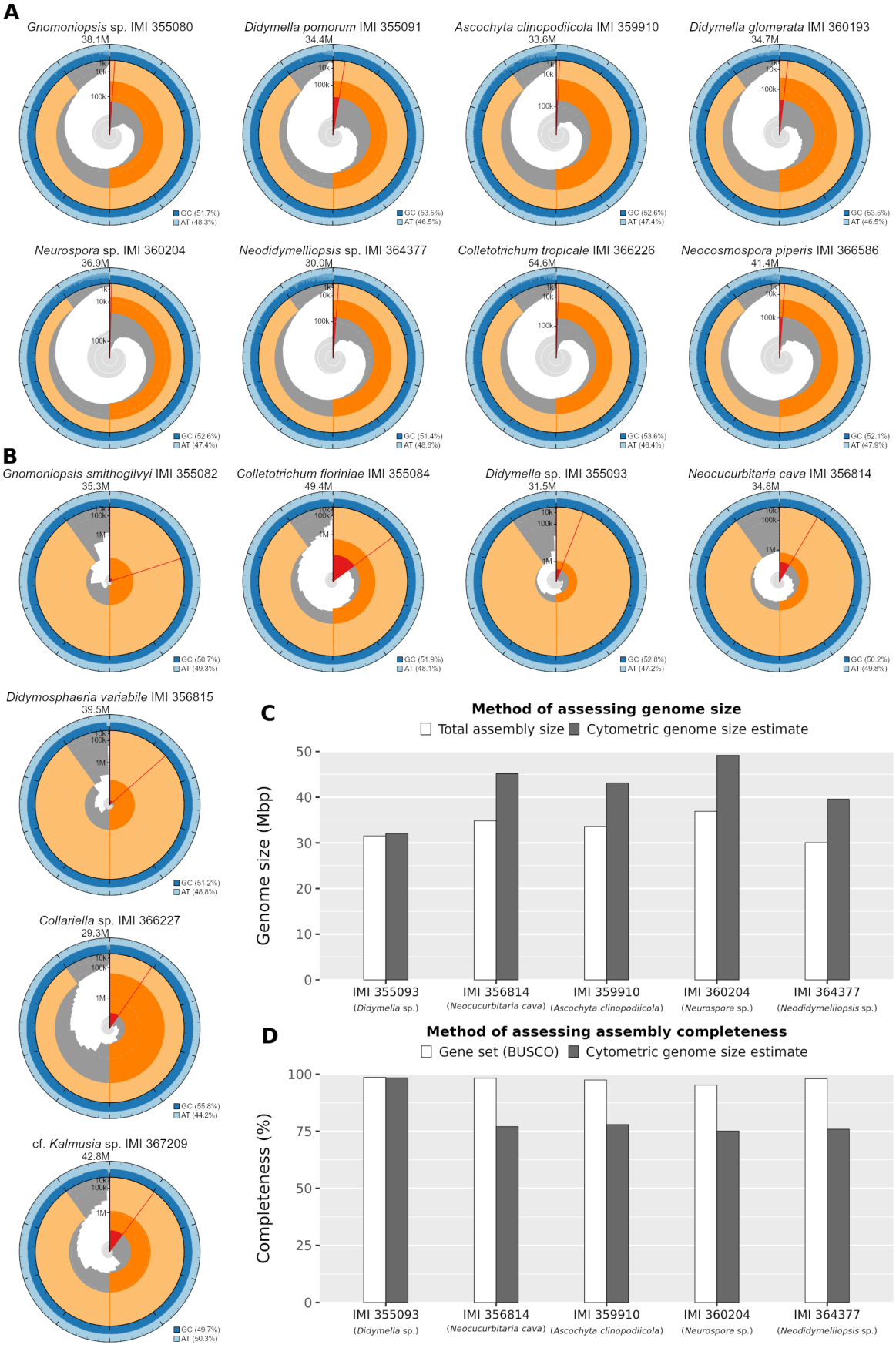

A
Gnomoniopsis sp. IMI 355080
38.1M
GC (51.7%)
AT (48.3%)
Didymella pomorum IMI 355091
34.4M
GC (53.5%)
AT (46.5%)
Ascochyta clinopodiicola IMI 359910
33.6M
GC (52.6%)
AT (47.4%)
Didymella glomerata IMI 360193
34.7M
GC (53.5%)
AT (46.5%)
Neurospora sp. IMI 360204
36.9M
GC (52.6%)
AT (47.4%)
Neodidymelliopsis sp. IMI 364377
30.0M
GC (51.4%)
AT (48.6%)
Colletotrichum tropicale IMI 366226
54.6M
GC (53.6%)
AT (46.4%)
Neocosmospora piperis IMI 366586
41.4M
GC (52.1%)
AT (47.9%)
B
Gnomoniopsis smithogilvyi IMI 355082
35.3M
GC (50.7%)
AT (49.3%)
Colletotrichum fioriniae IMI 355084
49.4M
GC (51.9%)
AT (48.1%)
Didymella sp. IMI 355093
31.5M
GC (52.8%)
AT (47.2%)
Neocucurbitaria cava IMI 356814
34.8M
GC (50.2%)
AT (49.8%)
Didymosphaeria variabile IMI 356815
39.5M
GC (51.2%)
AT (48.8%)
Collariella sp. IMI 366227
29.3M
GC (55.8%)
AT (44.2%)
cf. Kalmusia sp. IMI 367209
42.8M
GC (49.7%)
AT (50.3%)
C
Method of assessing genome size
Total assembly size
Cytometric genome size estimate
Genome size (Mbp)
0
10
20
30
40
50
IMI 355093 (Didymella sp.)
IMI 356814 (Neocucurbitaria cava)
IMI 359910 (Ascochyta clinopodiicola)
IMI 360204 (Neurospora sp.)
IMI 364377 (Neodidymelliopsis sp.)
D
Method of assessing assembly completeness
Gene set (BUSCO)
Cytometric genome size estimate
Completeness (%)
0
25
50
75
100
IMI 355093 (Didymella sp.)
IMI 356814 (Neocucurbitaria cava)
IMI 359910 (Ascochyta clinopodiicola)
IMI 360204 (Neurospora sp.)
IMI 364377 (Neodidymelliopsis sp.)

Figure 4.2: Snail plots summarising assembly contiguity for **(A)** short-read and **(B)** hybrid assemblies. The distribution of fragment lengths is shown in dark grey with the plot radius scaled to the longest fragment of the assembly, shown in red. The pale grey spiral shows the cumulative fragment count on a log scale. The orange and cream arcs show the N50 and N90 fragment lengths, respectively. The outside blue bands show the distribution of GC/AT content. **(C)** Total genome size as indicated by total assembly length versus cytometric genome size estimation. **(D)** Genome assembly completeness as measured by gene set (BUSCOs) versus cytometric genome size estimation.

records: IMI 355084 and IMI 366226 had both been identified as the correct genus, *Colletotrichum*, although not the correct species, and IMI 366586 was classified as *Fusarium solani*, a species complex which is now synonymous with *Neocosmospora* (Crous, Lombard et al., 2021). Otherwise, the names mostly corresponded to a similar – although outdated – taxonomy, with the exception of IMI 366227 being assigned in CABI's records to the *Glomerellaceae* (*Glomerellales*) instead of the *Chaetomiaceae* (*Sordariales*) (Table 4.1).

4.5 Discussion

Here, we have reported the first genome assembly for 15 fungal endophyte strains, 8 being short-read and 7 hybrid. Unsurprisingly, incorporating long-reads resulted in much less fragmented assemblies, some likely approaching chromosome-level (Figure 4.2A,B) – detection of telomere motifs and cytological karyotyping of the strains will be required to assess exactly how close. We could see no conclusive reason to explain why some strains had higher contiguity when assembled with Flye versus Raven, or vice versa, however the two strains for which Flye outperformed Raven had two of the lowest long-read sequencing coverage statistics. It is interesting that the only tool to use both long- and short-reads in the assembly process itself, hybridSPAdes, produced far more fragmented assemblies compared to both other tools that only assemble long-reads and merely use short-reads to polish. This may speak to the fact that SPAdes predates long-read assembly, and so cannot compete with tools built specifically to tackle long-reads.

In agreement with Figure 1.2 in Chapter 1.4, we found that a high-level of assembly completeness according to BUSCOs is not necessarily corroborated when calculating completeness using a cytometric genome size measurement (Figure 4.2D). We can assume that our hybrid assembly of IMI 355093 (*Didymella* sp.) is highly complete as the cytometric genome size estimate and total assembly length were very similar (Figure 4.2C). Our cytometric estimates will hopefully provide a benchmark against which future attempts to refine these assemblies can be measured. As outlined in Chapter 1.4, the genome size disparity is likely due to the difficulty of assembling non-coding repeat regions, which will have downstream consequences on functional and evolutionary inferences.

Our ability to refine classifications using phylogenetic analyses varied depending on the number of sequenced taxa and the availability of suitable marker sequences for each lineage. Better-studied genera, such as *Colletotrichum* and *Neocosmospora*, have both extensive taxon sampling and a wide pool of genetic data available, and so we were more easily able to resolve strains to species-level (Figure 4.3G,J). Others presented more of a challenge – sequencing more strains of *Neodidymelliopsis sambuci*, for instance, may help to clarify if IMI 364377 belongs to the same species (Figure 4.3B). More genes is not necessarily the key to better classification, as seen for *Neurospora*, where we used the most genes of any of the lineages (Figure 4.1), and yet failed to significantly resolve the clade in which IMI 360204 was placed (Figure 4.3I). Our results were similar to García et al. (2004), who

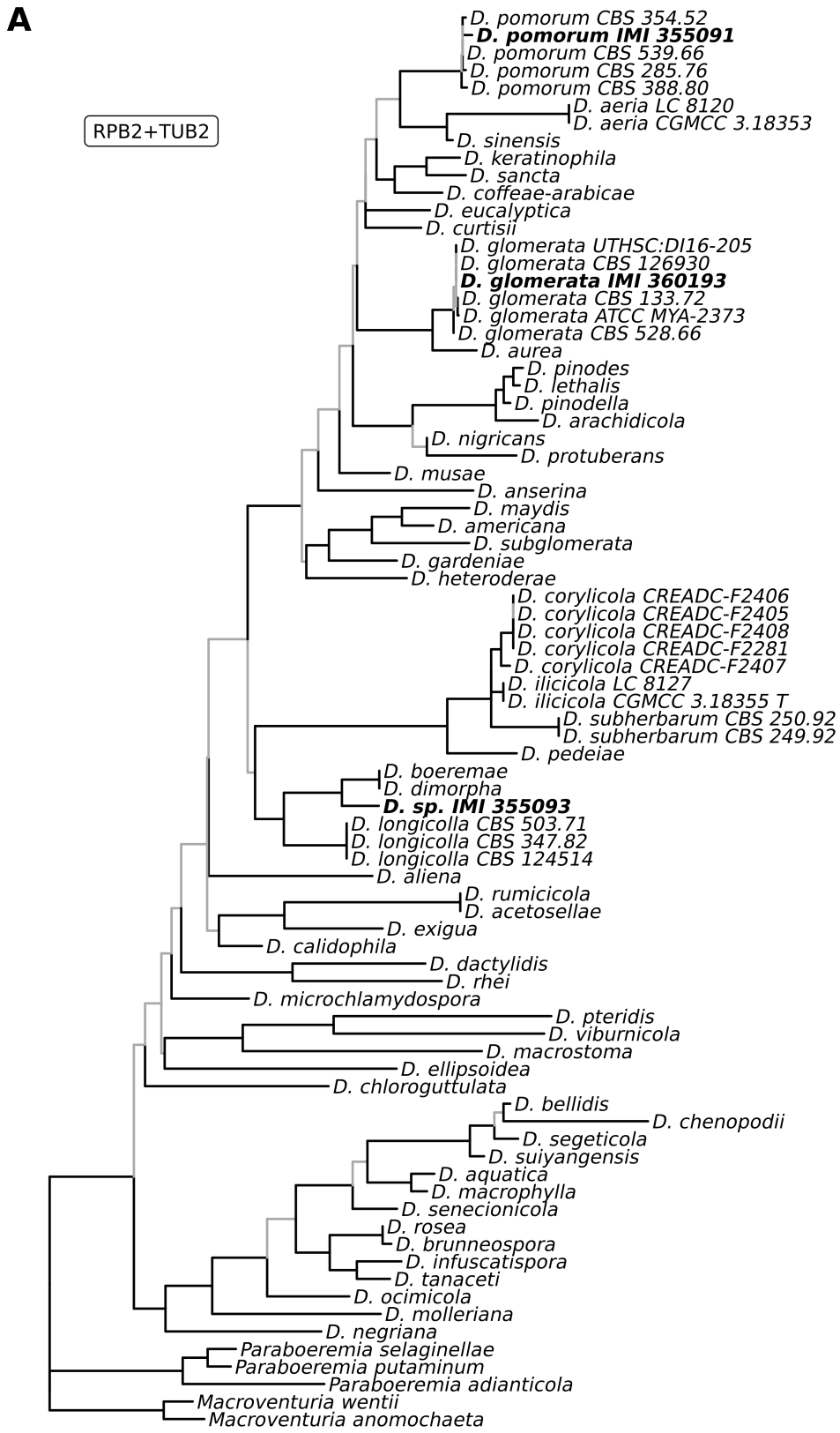


Figure 4.3: Maximum likelihood (ML) phylogenies produced using RAxML to refine classification of the 15 endophyte strains (shown in bold). Branches with significant bootstrap support (≥ 70) are in black, while others are in grey. The genetic markers used to build each tree are shown in the top left. **(A)** *D.* = *Didymella*. ▼

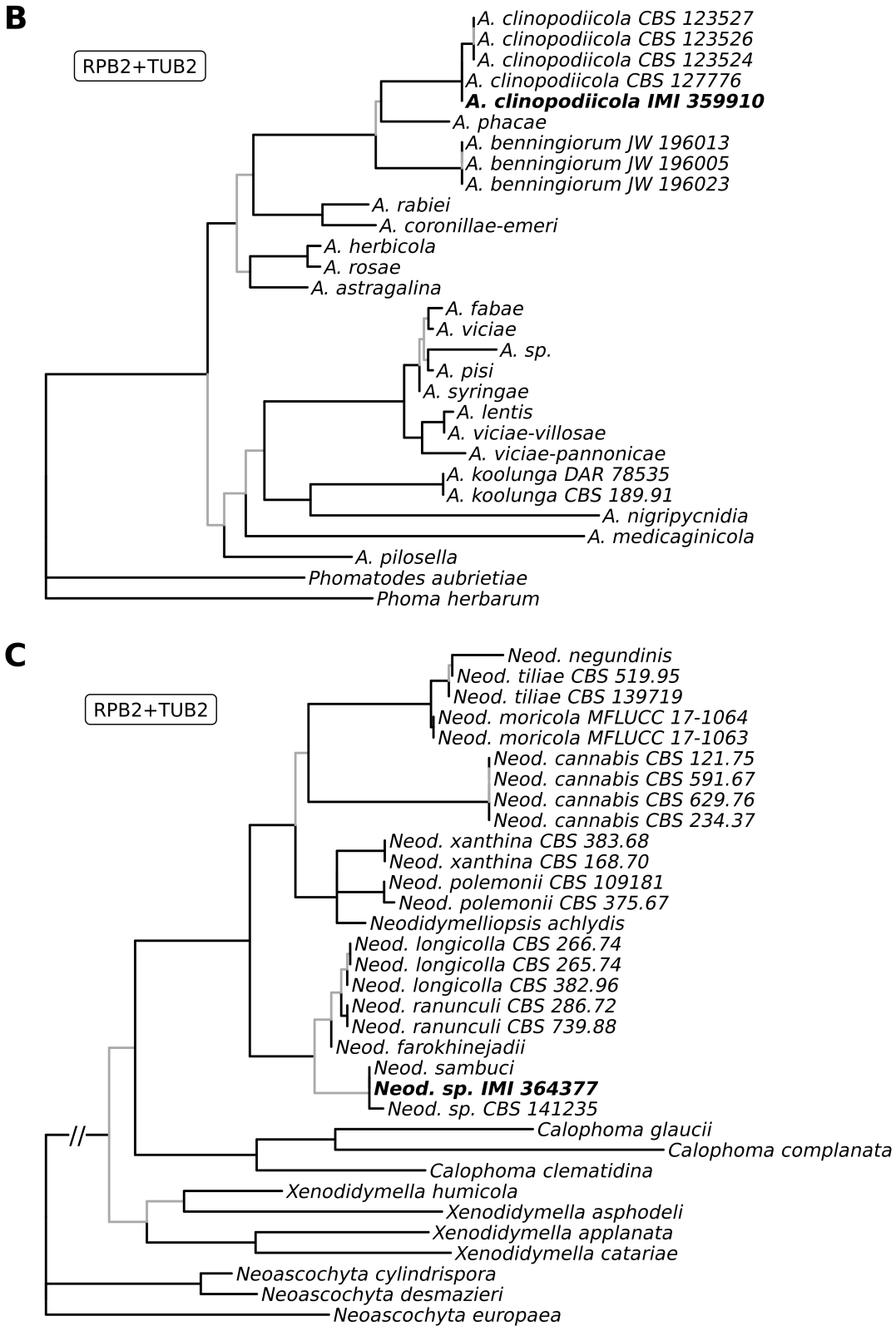


Figure 4.3: continued. **(B)** *A.* = *Ascochyta* **(C)** *Neod.* = *Neodidymelliopsis.* ▼

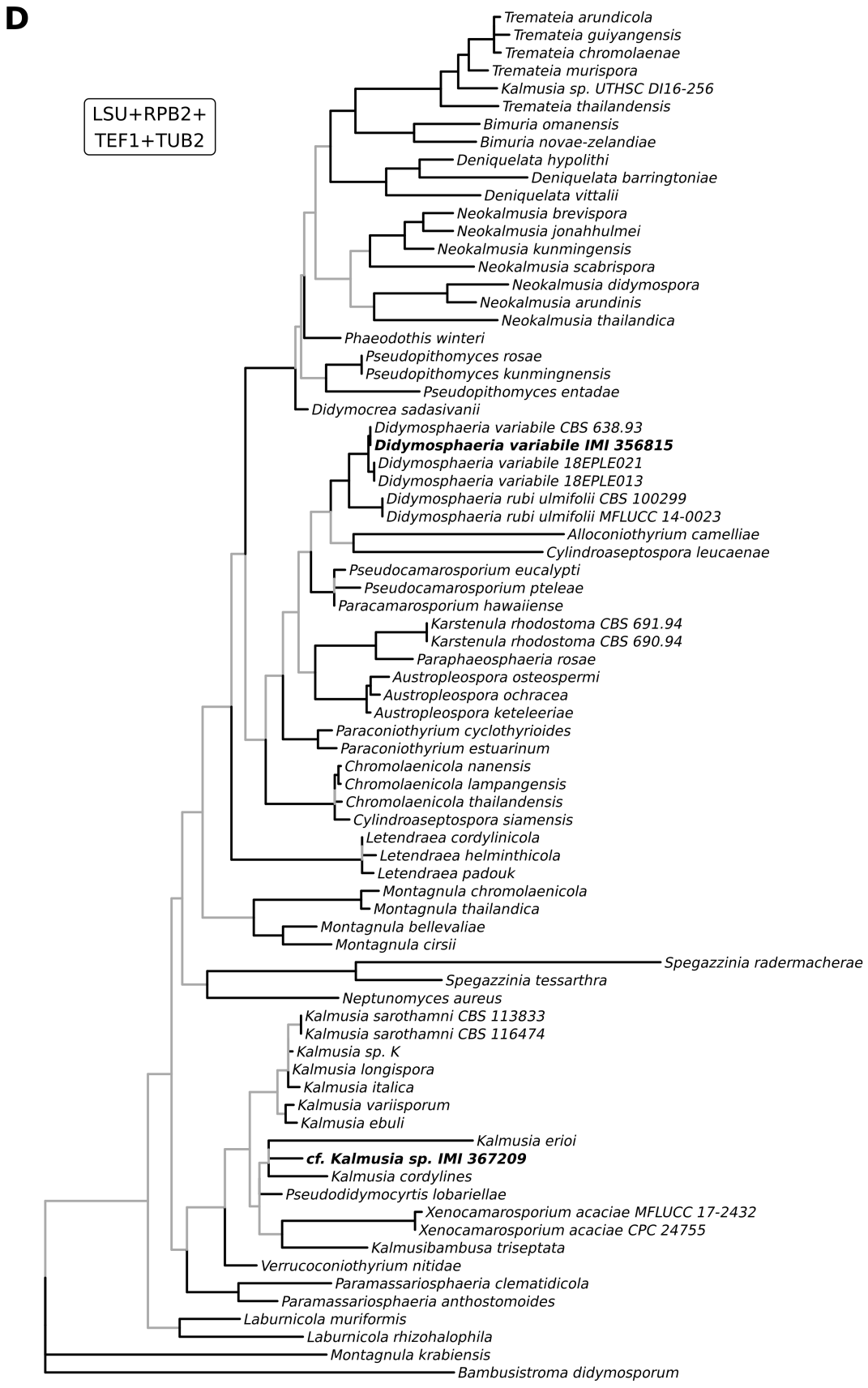


Figure 4.3: continued. **(D)** *Didymosphaeriaceae.* ▼

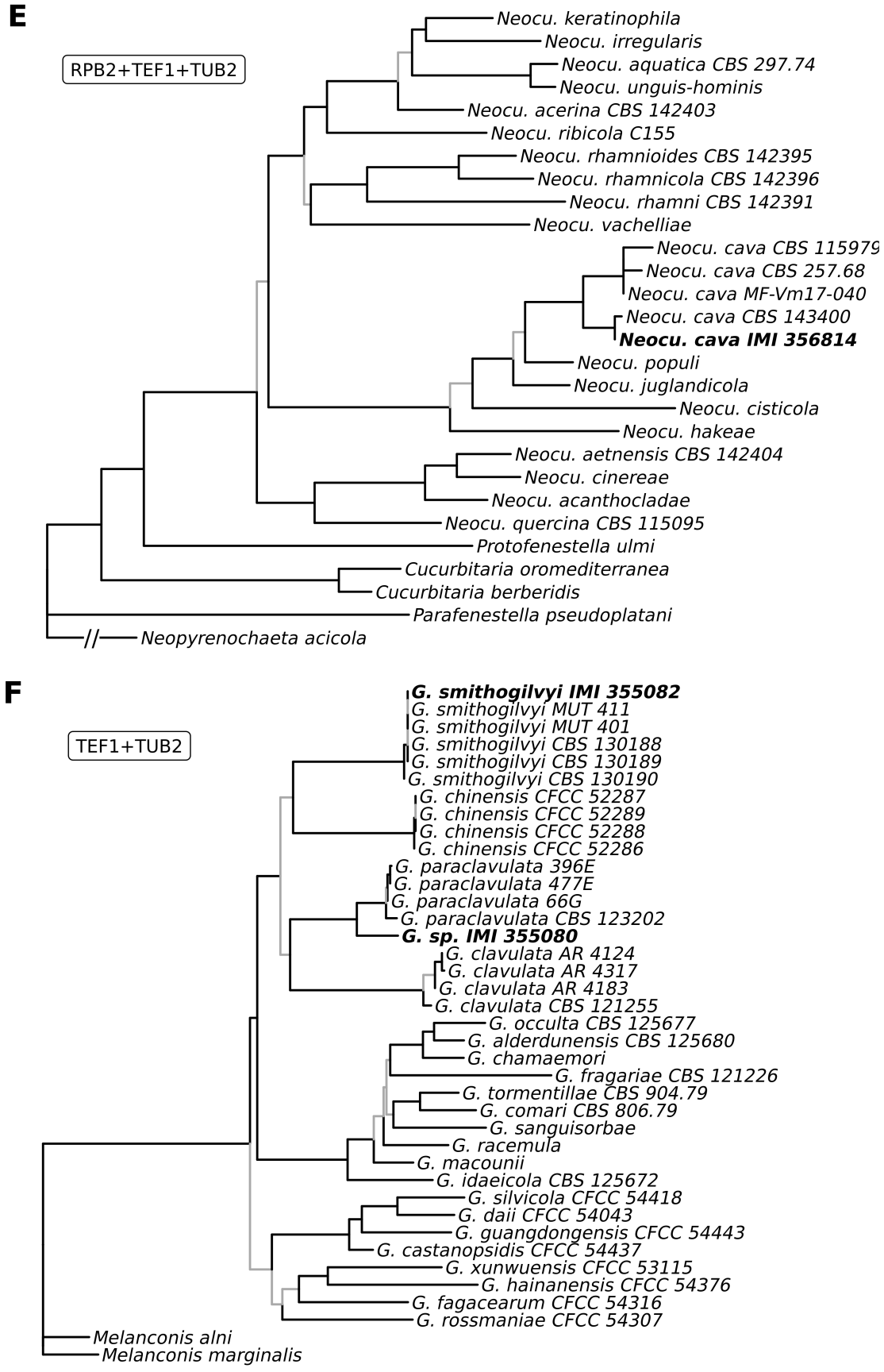


Figure 4.3: continued. **(E)** *Neocu.* = *Neocucurbitaria* **(F)** *G.* = *Gnomoniopsis.* ▼

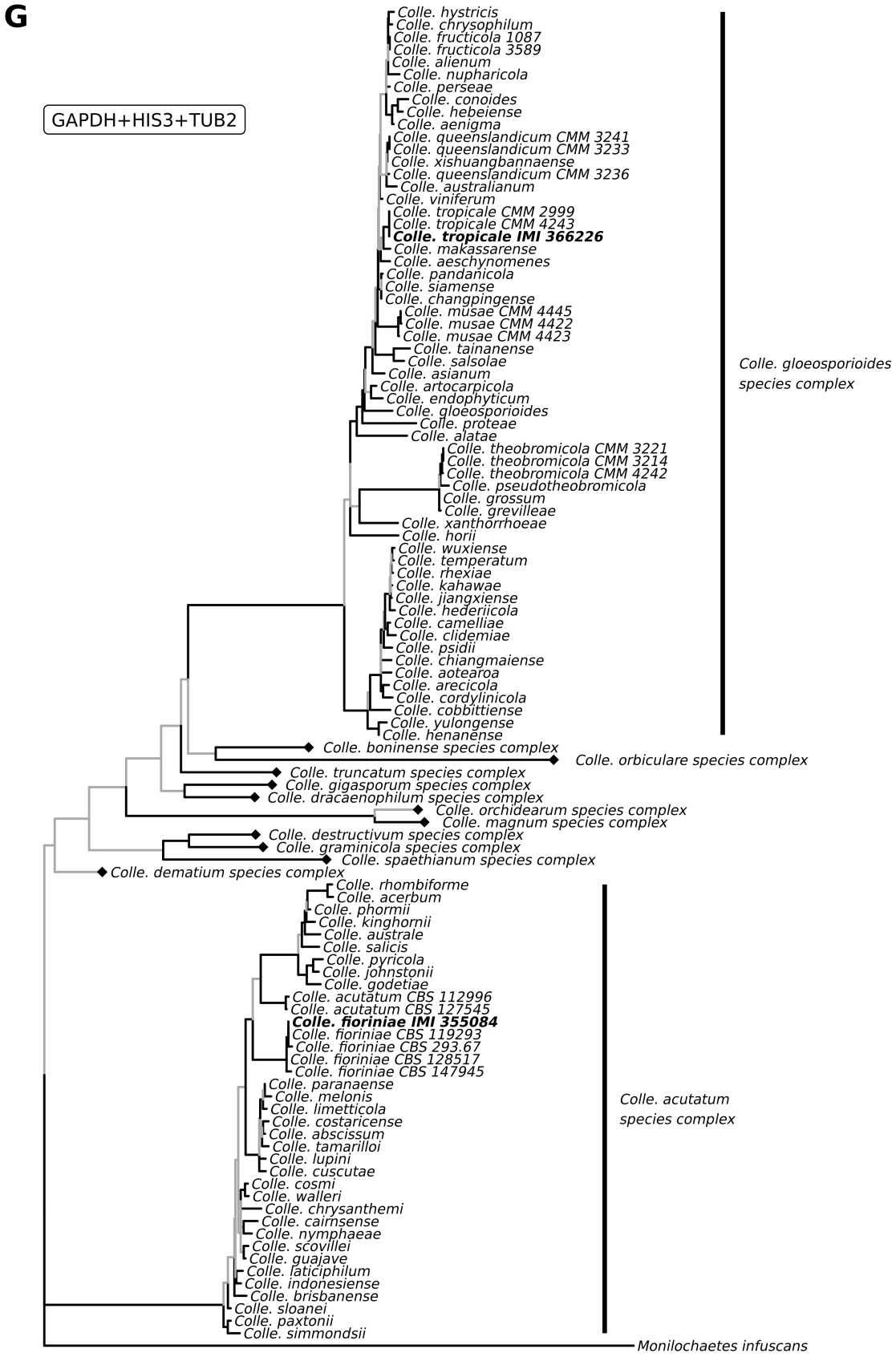


Figure 4.3: continued. **(G)** *Colle.* = *Colletotrichum*. Diamonds indicate collapsed species complexes. ▼

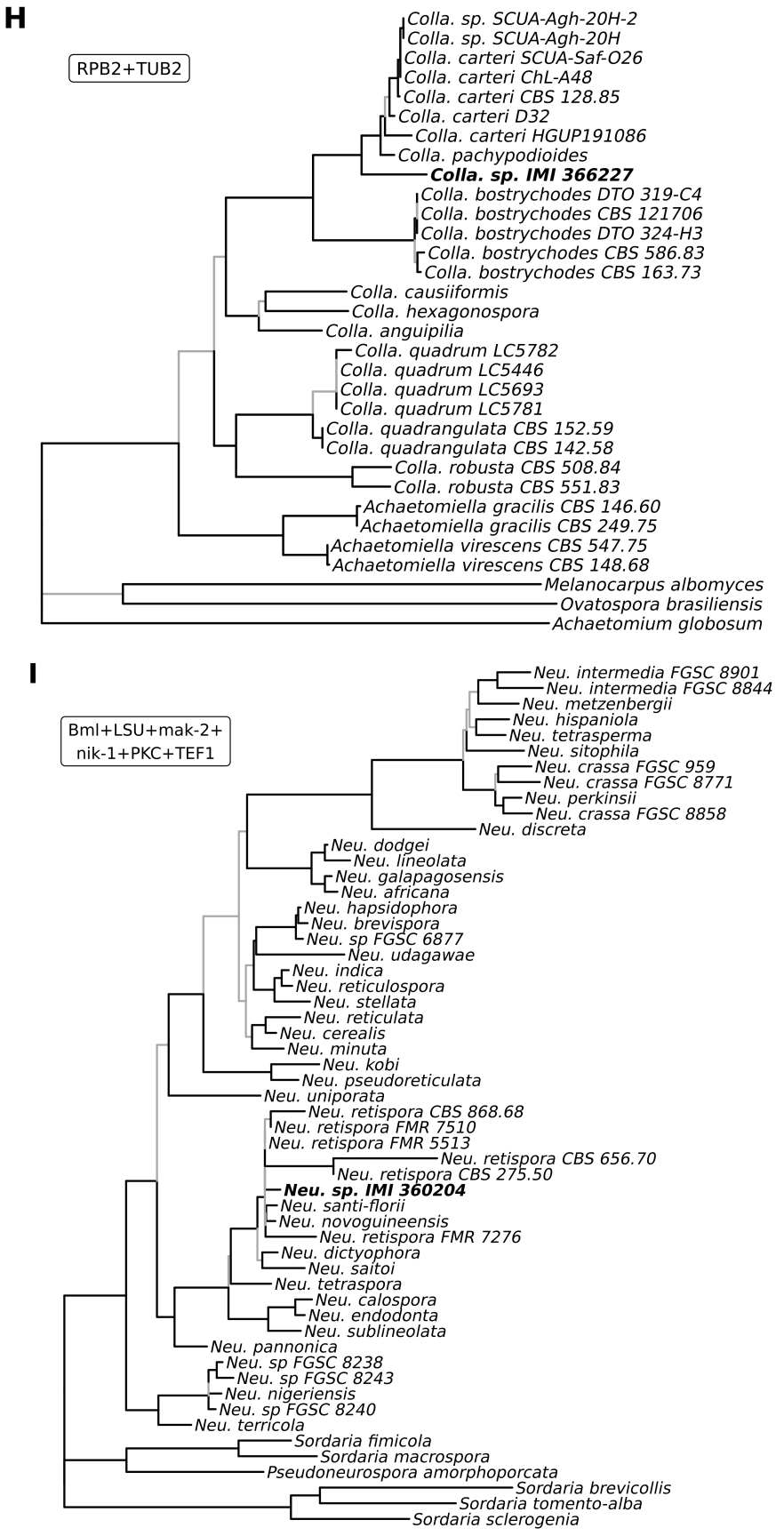


Figure 4.3: continued. **(H)** *Colla.* = *Collariella* **(I)** *Neu.* = *Neurospora.* ▼

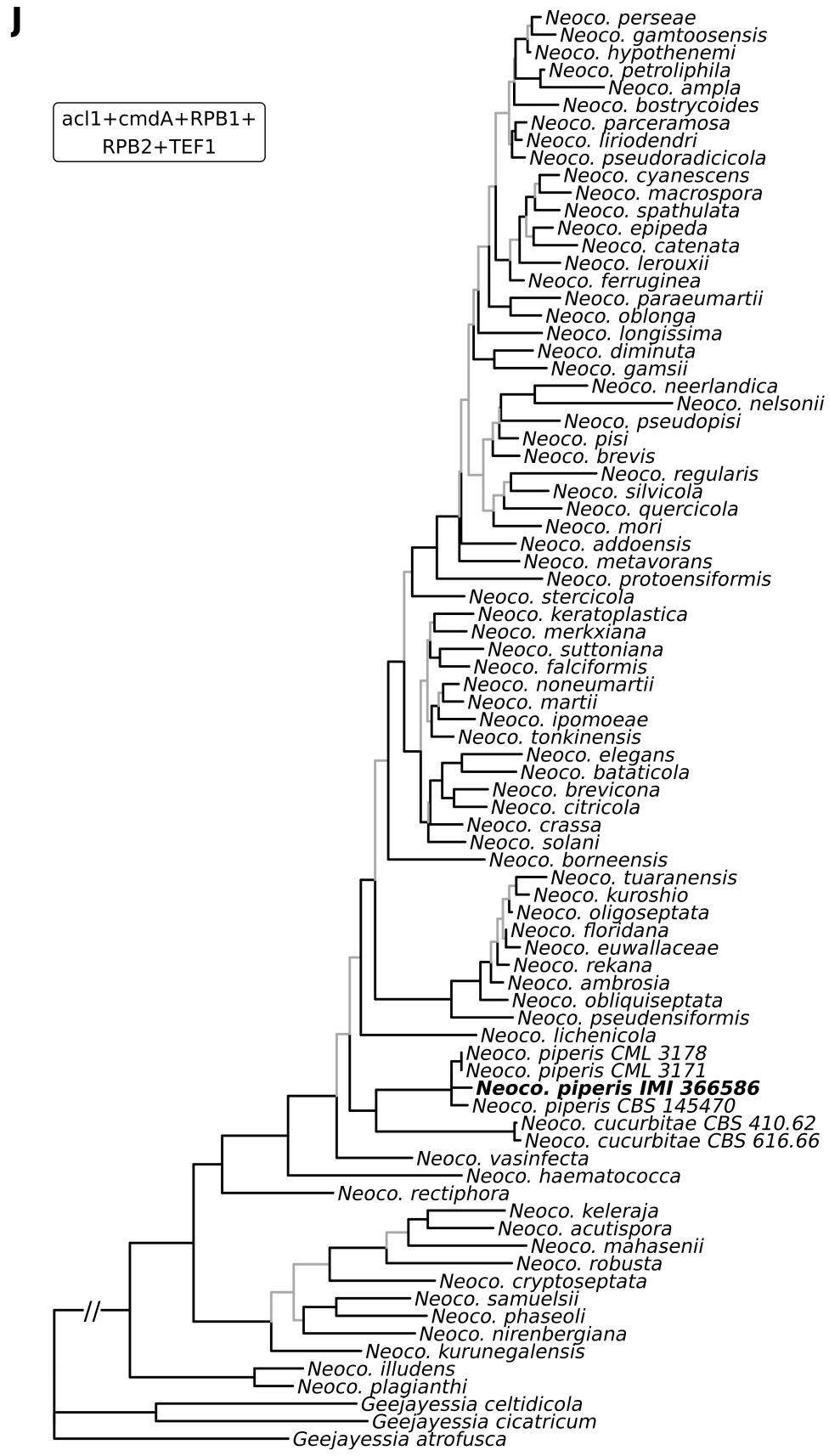


Figure 4.3: continued. **(J)** *Neoco.* = *Neocosmospora*.

also found *Neu. retispora*, *Neu. santi-florii* and *Neu. novoguineensis* to cluster together. This could partially be attributed to the need for taxonomic concepts to be revised, as is evidently the case for *Kalmusia* (Figure 4.3D). But it also emphasises that the performance of genes as phylogenetic markers depends on lineage, and that there is an unavoidable trade-off between including as many taxa as possible, and using the ideal genes for that lineage. Nonetheless, phylogenetic analyses refined the original classifications from CABI's records, with all but one assigned confidently to genus-level and nine to species-level. Of course, our results will benefit from validation through updated morphological assessment of the cultures – most importantly for the strains which we were not able to clarify to species-level – however the value of these genome assemblies has already been increased considerably with the revised names presented here.

Most of the genera or species represented here have already been reported in an endophytic context. *Colletotrichum fioriniaea* is a host-generalist phytopathogen and endophyte, globally distributed but most commonly found in temperate regions (Martin and Peter, 2021; Talhinhas and Baroncelli, 2021), which corresponds with the Swiss origin of IMI 355084 (Table 4.1). Interestingly, *Colle. fioriniaea* is also an entomopathogen of the elongate hemlock scale insect *Fiorinia externa* (Marcelino et al., 2008) – it is not uncommon for endophytic taxa to also be reported as insect pathogens, which has spurred the wider discussion on the potential use of entomopathogenic endophytes in biocontrol (Vidal and Jaber, 2015; Vega, 2018). *Colle. tropicale* was described from an endophytic strain isolated from *Theobroma cacao* in Panama (Rojas et al., 2010), again in line with the origin of IMI 366226 in Puerto Rico (Table 4.1). It has also been reported as an endophyte of tropical grass species in Thailand (Manamgoda et al., 2013) and pathogen of *Passiflora edulis* amongst other Brazilian crops (Silva, Silva et al., 2021).

Similarly to *Colletotrichum* species, *Gnomoniopsis smithogilvyi* (syn. *Gnomoniopsis castanea*) is known as both pathogen and endophyte, but with greater host specificity: it's found primarily on chestnuts (*Castanea* spp.) (Crous, Summerell et al., 2012), as well as oak (*Quercus*, as reported in this chapter), pine (*Pinus*) and ash (*Fraxinus*) across Europe, Asia and Australasia (Lione et al., 2019). Once again, *G. smithogilvyi* is also an entomopathogen of chestnut gall wasps (*Dryocosmus kuriphilus*), although any biocontrol possibility is undermined by the fact that the fungus is pathogenic on both insect and plant host (Vannini et al., 2017; Fernández, Bezos and Diez, 2018).

The genus *Kalmusia* is also known for both endophytic and phytopathogenic taxa (Gutierrez et al., 2022; Karácsony et al., 2021). Our phylogenetic analyses of the *Didymosphaeriaceae* found the genus to be polyphyletic, which echoed results from Zhang, Zhang et al. (2014). The type species of the genus, *K. ebuli*, was not in the group of *Kalmusia* taxa which IMI 367209 clustered with (Figure 4.3D), making it likely that in the future IMI 367209 and closely related '*Kalmusia*' spp. will be reclassified to another genus. As in our results, Wanasinghe and Mortimer (2022) found *Pseudodidymocyrtis lobariellae* to cluster together with '*Kalmusia*' spp., and the authors also commented on close morphological resemblance, noting that further work is needed on the delimitation between the two genera. *P. lobariellae* was described as a lichenicolous fungus isolated from *Lobariella pallida* (Flakus et al., 2019), but it has apparently also been isolated as an endophyte from *Taxus chinensis*, although the authors do not give details on how it was isolated from the host or how it was identified to be *P. lobariellae* (Cao et al., 2022). It is established that endolichenic and endophytic taxa can be closely related (Arnold, Miadlikowska et al., 2009; U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019), but further investigation into the identity and lifestyle(s) of IMI 367209 is certainly warranted.

As with many phoma-like species, *Didymella pomorum* and *D. glomerata* are primarily considered plant pathogens (e.g. Moral et al., 2018; Havenga et al., 2019; Song et al., 2021; Ilyukhin, 2022), however there are also reports of *D. glomerata* and other *Didymella* species as endophytes (e.g. Alidadi et al., 2019; Stranska et al., 2022). Both *D. pomorum* and *D. glomerata* have also been found to grow on inorganic substrates such as asbestos, cement and plaster (Aveskamp, de Gruyter and Crous, 2008). *Didymosphaeria variabile* (syn. *Paraconiothyrium variabile*) was described based on isolates from bark and 'necrotic' wood of *Prunus* spp. (Damm et al., 2008), and has been reported to cause leaf spot of *Phoenix theophrasti* (Ligoxigakis et al., 2013). However, an explicitly endophytic strain isolated from *Cephalotaxus harringtonia* has been shown to be antagonistic against the common phytopathogen *Fusarium oxysporum* (Combès et al., 2012), to the extent of reducing *F. oxysporum* lethality in *Arabidopsis* by 85% (Bärenstrauch et al., 2020). *Didymosphaeria variable* has also been found to produce the secondary metabolite taxol, which is used as an anti-cancer drug (Somjaipeng et al., 2015).

Neurospora (syn. *Gelasinospora*) species are globally distributed soilborne fungi, thought primarily to be saprotrophs (García et al., 2004; Allison et al., 2018), although endophytic strains have been isolated too (e.g. Wang, Li et al., 2017). Famously, *Neu. crassa* is a model organism with a rich history of use in scientific research (Davis and Perkins, 2002), and it has also been shown to be a naturally occurring endophyte and pathogen of *Pinus sylvestris* (Kuo et al., 2014). Perhaps surprisingly we could find little mention of the species most closely related to IMI 360204 – *Neu. retispora*, *Neu. santi-florii* and *Neu. novoguineensis* – outside of a purely taxonomic context, suggesting that lifestyles of *Neurospora* species beyond *Neu. crassa* are not well studied.

Other taxa sequenced here represent novel reports of endophytism. *Ascochyta clinopodiicola* was first isolated in Italy from a dead stem of *Clinopodium nepeta* (Hyde, Chaiwan et al., 2018), and the genus is predominantly known for pathogens of grain legumes (Tivoli and Banniza, 2007), so it is intriguing that IMI 359910 was isolated as an endophyte of a wild alpine flower, *Dryas octopetala* (Table 4.1). *Neocosmospora piperis* (syn. *F. solani* f. sp. *piperis*; *Nectria haematococca* f. sp. *piperis*) is a pathogen of *Piper nigrum*, which was described from a strain isolated in Brazil (Sandoval-Denis, Lombard and Crous, 2019). Although *Neoco. piperis* has not previously been reported as an endophyte, the genus *Neocosmospora* is known for many species capable of both pathogenicity and endophytism, as highlighted in previous chapters (Figure 2.5, Appendix A.3). The isolation of IMI 366586 in Puerto Rico is also geographically concordant with the known range of *Neoco. piperis*.

As with many other phoma-like genera, *Neodidymelliopsis* was circumscribed relatively recently, and is known for saprotrophic and potentially pathogenic species reported from Europe, Canada and Israel (Chen, Jiang et al., 2015; Hyde, Chaiwan et al., 2018; Hyde, Tennakoon et al., 2019). This makes the report of IMI 364377 as an endophyte of *Persea americana* from Trinidad and Tobago both geographically and ecologically novel (Table 4.1). *Neocucurbitaria* is a similarly recently established genus, and for which there are already diverse lifestyle reports including presumed saprotrophs (Wanasinghe, Phookamsak et al., 2017), opportunistic human pathogens (Garcia-Hermoso et al., 2019; Valenzuela-Lopez, Cano-Lira, Stchigel et al., 2019) and numerous aquatic species (Magaña-Dueñas, Stchigel and Cano-Lira, 2021). *Neocucurbitaria cava* specifically has previously been isolated from both plant material and soil in Europe (Valenzuela-Lopez, Cano-Lira, Guarro et al., 2018), as was the case for IMI 356814 isolated in Spain from *Quercus* (Table 4.1).

The genus *Collariella* is unique amongst the other taxa here in that it is the only one that is not known as plant associated, instead comprised of species isolated from substrates such as dung, soil,

dust and air (Wang, Houbraken et al., 2016). As the taxonomy of *Collariella* contradicts the name that IMI 366227 was assigned in CABI's records – *Colletotrichum crassipes* (Table 4.1) – it raises the question as to whether we sequenced an airborne contaminant. As discussed above, *Colletotrichum* species are indeed frequently reported as endophytes. However, the assembly for *Collariella* sp. IMI 366227 showed very little contamination (Supplementary Figure S4.6N), suggesting that the strain was successfully sequenced from axenic culture. We cannot rule out a contamination at the point of original isolation and deposition in CABI's collection, however under that circumstance we would presumably still expect a mixed culture when taken out of cryopreservation, unlike the axenic one found here. Based on the broadscale associations that are outlined for other fungi above, it is not implausible that *Collariella* taxa are also capable of exhibiting endophytic lifestyles.

4.6 Conclusions

Here we report the first genome assemblies, to our knowledge, for the genera *Collariella*, *Neodidymelliopsis* and *Neocucurbitaria*, and the species *Ascochyta clinopodiicola*, *Didymella pomorum*, *Didymosphaeria variabile*, *Neocosmospora piperis* and *Neocucurbitaria cava*. *Didymella* sp. IMI 355093, *Gnomoniopsis* sp. IMI 355080, cf. *Kalmusia* sp. IMI 367209 and *Neurospora* sp. IMI 360204 require morphological assessment to determine whether they are new or previously described species, but based on existing data they also likely represent the first genome assemblies for their to-be-assigned species. As well as providing the first genomic resources for taxa, these endophyte assemblies enable future work comparing endophytic and phytopathogenic strains widely across the *Ascomycota*. We also highlight that genome size statistics from assemblies can differ markedly from cytometric genome size estimates in spite of high BUSCO completeness, emphasising that using BUSCOs alone to assess assembly completeness can result in an false impression of high assembly quality. Our results demonstrate the value of mining existing culture collections to produce much-needed genomic data for neglected lineages of plant associated fungi.

Acknowledgements

We thank Rosie Woods for managing the shipment of extractions for sequencing. Many thanks to Helen Stewart for supervising lab work at CABI. Thanks to Richard Wright for help providing standards for flow cytometry, and Michael Campbell for allowing us to use unpublished EST and protein data of Gnocas1 from the JGI MycoCosm portal for structural annotation. Thank you also to the ITS Research team at Queen Mary University of London.

4.7 Supplementary material

Supplementary Table S4.1: GenBank accession numbers for taxa used in the phylogenetic analyses. ^{T} = ex-type, ^{ET} = ex-epitype. *Ascochyta* sampling informed by Hou et al. (2020).

	Name	Voucher	RPB2	TUB2
***Ascochyta* (Figure 4.3B)**	*Ascochyta astragalina*	CBS 113797 = UPSC 2222	MT018257	KT389776
	A. benningiorum	CBS 144957^{T} = JW 196005	MN824606	MN824755
	A. benningiorum	JW 196013	MN824608	MN824757
	A. benningiorum	JW 196023	MN824607	MN824756
	A. clinopodiicola	CBS 123524		MT005693
	A. clinopodiicola	CBS 123527		MT005694
	A. clinopodiicola	CBS 123526		MT005692
	A. clinopodiicola	CBS 127776		MT005695
	A. coronillae-emeri	MFLUCC 13-0820^{T}	MH069679	MH069686
	A. fabae	CBS 524.77	MT018241	GU237526
	A. herbicola	CBS 629.97 = PD 76/1017	KP330421	GU237614
	A. koolunga	CBS 189.91	MN983286	MN983711
	A. koolunga	DAR 78535^{T}	EU874849	
	A. lentis	CBS 231.79 = DAOM 170658	MT018248	MT005689
	A. medicaginicola	CBS 112.53^{T}	MT018251	GU237628
	A. nigripycnidia	CBS 116.96^{T} = PD 95/7930	MT018253	GU237637
	A. phacae	CBS 184.55^{T}	MT018255	KT389769
	A. pilosella	CBS 583.97^{T}	MT018258	MT005696
	A. pisi	CBS 122785^{T} = PD 78/517	MT018244	GU237532
	A. rabiei	CBS 237.37^{T}	MT018256	KT389773
	A. rosae	MFLUCC 15-0063^{T}	KY514409	
	A. sp.	CBS 136887	MN983295	KX033387
	A. syringae	CBS 126.82	MN983308	MN983728
	A. viciae	CBS 451.68	KT389562	KT389778
	A. viciae-pannonicae	CBS 254.92	MT018250	KT389779
	A. viciae-villosae	CBS 255.92	MT018249	MT005690
	Phoma herbarum	CBS 615.75 = IMI 199779 = PD	KP330420	FJ427133
	Phomatodes aubrietiae	CBS 627.97^{T} = PD 70/714	KT389665	GU237585

Supplementary Table S4.1 continued. *Collariella* sampling informed by Wang, Houbraken et al. (2016) and Wang, Han et al. (2022).

	Name	Voucher	RPB2	TUB2
***Collariella* (Figure 4.3H)**	*Achaetomium globosum*	CBS 332.67^{T}	KX976793	KX976911
	Melanocarpus albomyces	CBS 638.94^{T}	KX976886	KX977021
	Ovatospora brasiliensis	CBS 130174	KX976895	KX977030
	Collariella bostrychodes	CBS 163.73	KX976837	KX976983
	Colla. bostrychodes	CBS 586.83	KX976838	KX976984
	Colla. bostrychodes	DTO 319-C4		KX976985
	Colla. bostrychodes	DTO 324-H3 = DTO 324-H6	KX976839	KX976986
	Colla. bostrychodes	CBS 121706		KX976987
	Colla. causiiformis	CBS 792.83^{T}	KX976840	KX976988
	Colla. carteri	CBS 128.85^{T}	KX976841	KX976989
	Colla. gracilis	CBS 146.60^{T}	KX976842	KX976990
	Colla. gracilis	CBS 249.75	KX976843	KX976991
	Colla. quadrangulata	CBS 142.58	KX976844	KX976992
	Colla. quadrangulata	CBS 152.59	KX976845	KX976993
	Colla. robusta	CBS 551.83^{T}	KX976846	KX976994
	Colla. robusta	CBS 508.84	KX976847	KX976995
	Colla. virescens	CBS 148.68^{T}	KX976848	KX976996
	Colla. virescens	CBS 547.75	KX976849	KX976997
	Colla. anguipilia	CBS 632.83	MZ342989	MZ343028
	Colla. quadrum	CGMCC:3.17920 = LC5782	KY575873	KU746770
	Colla. quadrum	CGMCC:3.17919 = LC5781	KY575872	KU746769
	Colla. quadrum	CGMCC:3.17918 = LC5693	KY575871	KU746768
	Colla. quadrum	CGMCC:3.17917 = LC5446	KY575870	KU746767
	Colla. hexagonospora	CBS 171.84	MZ342977	MZ343016
	Colla. pachypodioides	CBS 164.52	MZ342975	MZ343014
	Colla. carteri	SCUA-Saf-O26	MW671060	MW671081
	Colla. carteri	HGUP191086		MZ724096
	Colla. carteri	D32	MG890121	
	Colla. carteri	ChL-A48		MG890023
	Colla. sp.	SCUA-Agh-20H	MN520427	MN520423
	Colla. sp.	SCUA-Agh-20H-2	MN520426	MN520422

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colletotrichum abscissum*	COAD 1877^{T}	KP843129	KP843138	KP843135
	Colle. acerbum	CBS 128530 = ICMP 12921 = PRJ 1199.3^{T}	JQ948790	JQ949450	JQ950110
	Colle. acidae	MFLUCC 17–2659^{T}	MH003691		MH003700
	Colle. acutatum	CBS 127545 = CPC 13947	JQ948714	JQ949374	JQ950034
	Colle. acutatum	CBS 112996 = ATCC 56816 = STE-U 5292^{T}	JQ948677	JQ005818	JQ005860
	Colle. aenigma	ICMP 18608^{T}	JX010044		JX010389
	Colle. aeschynomenes	ICMP 17673^{T} = ATCC 201874	JX009930		JX010392
	Colle. alatae	CBS 304.67^{T} = ICMP 17919	JX009990		JX010383
	Colle. alienum	ICMP 12071^{T}	JX010028		JX010411
	Colle. americae-borealis	CBS 136232^{T}	KM105579	KM105364	KM105504
	Colle. annellatum	CBS 129826 = CH1^{T}	JQ005309	JQ005483	JQ005656
	Colle. anthrisci	CBS 125334^{T}	GU228237	GU228041	GU228139
	Colle. antirrhinicola	CBS 102189^{T}	KM105531	KM105320	KM105460
	Colle. aotearoa	ICMP 18537^{T}	JX010005		JX010420
	Colle. arecicola	CGMCC 3.19667^{T}	MK935455		MK935498
	Colle. artocarpicola	MFLUCC 18-1167^{T}	MN435568		MN435567
	Colle. arxii	CBS 132511^{T}	KF687843	KF687858	KF687881
	Colle. asianum	ICMP 18580^{T} = CBS 130418	JX010053		JX010406
	Colle. australe	CBS 116478 = HKUCC 2616^{T}	JQ948786	JQ949446	JQ950106
	Colle. australianum	VPRI 43075^{T}	MG572127		MG572149
	Colle. beeveri	CBS 128527 = ICMP 18594^{T}	JQ005258	JQ005432	JQ005605
	Colle. bidentis	COAD 1020^{T} = CPC 21930	KF178506	KF178554	KF178602
	Colle. bletillum	CGMCC 3.15117^{T}	KC843506		JX625207
	Colle. boninense	CBS 123755 = MAFF 305972^{T}	JQ005240	JQ005414	JQ005588
	Colle. brasiliense	CBS 128501 = ICMP 18607 = PAS12^{T}	JQ005322	JQ005496	JQ005669
	Colle. brassicicola	CBS 101059 = LYN 16331^{T}	JQ005259	JQ005433	JQ005606
	Colle. brevisporum	CBS 129957	MG600822	MG600908	MG601029
	Colle. brisbanense	CBS 292.67 = DPI 11711^{T}	JQ948621	JQ949282	JQ949942
	Colle. bryoniicola	CBS 109849^{T}	KM105532	KM105321	KM105461
	Colle. cacao	CBS 119297^{T}	MG600832	MG600916	MG601039
	Colle. cairnsense	RIP 63642^{T} = CBS 140847	KU923704	KU923722	KU923688
	Colle. camelliae	CGMCC 3.14925 = LC1364^{T}	KJ954782	MZ673847	KJ955230

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colle. catinaense*	CBS 142417 = CPC 27978	KY856224	KY856307	KY856482
	Colle. catinaense	CBS 142416 = CPC 28019	KY856223	KY856306	KY856481
	Colle. cattleyicola	CBS 170.49^{T}	MG600819	MG600905	MG601025
	Colle. cereale	CBS 129663 = KS20BIG			JQ005858
	Colle. changpingense	CGMCC 3.17582^{T} = SA0016 = MFLUCC 15-0022	MZ664048		MZ673952
	Colle. chiangmaiense	MFLUCC 18-0945^{T}	MW548592		
	Colle. chrysanthemi	IMI 364540 = CPC 18930^{T}	JQ948603	JQ949264	JQ949924
	Colle. chrysophilum	CMM4268^{T}	KX094183		KX094285
	Colle. circinans	CBS 221.81^{T}	GU228247	GU228051	GU228149
	Colle. clidemiae	ICMP 18658^{T}	JX009989		JX010438
	Colle. clivicola	CBS 125375^{T}	MG600795	MG600892	MG601000
	Colle. cobbittiense	BRIP 66219^{T}	MH094133	MH094136	MH094137
	Colle. coelogynes	CBS 132504^{T}	MG600776	MG600882	MG600980
	Colle. colombiense	CBS 129818 = G2^{T}	JQ005261	JQ005435	JQ005608
	Colle. conoides	CGMCC 3.17615 = CAUG17 = LC6226^{T}	KP890162		KP890174
	Colle. constrictum	CBS 128504 = ICMP 12941^{T}	JQ005325	JQ005499	JQ005672
	Colle. corchorum-capsularis	FAFU 03	KT439361		KT439341
	Colle. cordylinicola	MFLUCC 090551^{T} = ICMP 18579	JX009975		JX010440
	Colle. cosmi	CBS 853.73 = PD 73/856^{T}	JQ948604	JQ949265	JQ949925
	Colle. costaricense	CBS 330.75^{T}	JQ948510	JQ949171	JQ949831
	Colle. curcucmae	IMI 288937^{T}	GU228285		GU228187
	Colle. cuscutae	IMI 304802 = CPC 18873^{T}	JQ948525	JQ949186	JQ949846
	Colle. cymbidicola	IMI 347923^{T}	JQ005253	JQ005427	JQ005600
	Colle. dacrycarpi	CBS 130241 = ICMP 19107^{T}	JQ005323	JQ005497	JQ005670
	Colle. dematium	CBS 125.25^{T}	GU228211	GU228015	GU228113
	Colle. destructivum	CBS 136228^{T}	KM105561	KM105347	KM105487
	Colle. destructivum	CBS 136852	KM105562	KM105348	KM105488
	Colle. dracaenophilum	CBS 118199^{T}	JX546707	JX546756	JX519247
	Colle. endophyticum	MFLUCC 13-0418 = LC0324^{T}	KC832854	MZ673839	MZ673954
	Colle. eremochloae	CBS 129661^{T} = C05			JX519245
	Colle. falcatum	CGMCC 3.14187 = CBS 147945^{T}			JQ005856
	Colle. fioriniae	CBS 293.67,DPI 13120	JQ948640	JQ949301	JQ949961
	Colle. fioriniae	CBS 128517 = ARSEF 10222 = ERL 1257 = EHS 58^{T}	JQ948622	JQ949283	JQ949943

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colle. fioriniae*	CBS 129948	JQ948674	JQ949335	JQ949995
	Colle. fioriniae	CBS 119293	JQ948644	JQ949305	JQ949965
	Colle. fructi	CBS 346.37 / CCT 4806^{T}	GU228236	GU228040	GU228138
	Colle. fructicola	1087	KX094174		KX094279
	Colle. fructicola	3589	KX094175		KX094280
	Colle. fuscum	CBS 133701^{T}	KM105524	KM105314	KM105454
	Colle. fusiforme	MFLUCC 12–0437T	KT290255		KT290256
	Colle. gigasporum	CBS 101881	KF687841	KF687861	KF687886
	Colle. gloeosporioides	IMI 356878^{T} = ICMP 17821 = CBS 112999	JX010056	JQ005413	JX010445
	Colle. godetiae	CBS 133.44^{T}	JQ948733	JQ949393	JQ950053
	Colle. graminicola	CBS 130836^{T} M1001			JQ005851
	Colle. grevilleae	CBS 132879 = CPC 15481	KC297010	KC297056	KC297102
	Colle. grossum	CGMCC3.17614 = CAUG7 = LC6227^{T}	KP890159		KP890171
	Colle. guajave	IMI 350839^{T}	JQ948600	JQ949261	JQ949921
	Colle. guizhouensis	CGMCC 3.15112^{T}	KC843507		JX625185
	Colle. hebeiense	MFLUCC13–0726^{T}	KF377495		KF288975
	Colle. hederiicola	CBS 142418 = CPC 26844^{T}	KY856270	KY856361	KY856528
	Colle. henanense	LC3030 = CGMCC 3.17354 = LF238 ^{T}	KJ954810	MZ673835	KJ955257
	Colle. higginsianum	IMI 349061 = CPC 19379^{T}	KM105535	KM105324	KM105464
	Colle. hippeastri	CBS 125376 = CSSG1^{T}	JQ005318	JQ005492	JQ005665
	Colle. horii	NBRC 7478^{T} = ICMP 10492 = MTCC 10841	GQ329681		JX010450
	Colle. hystricis	CBS 142411 = CPC 28153^{T}	KY856274	KY856365	KY856532
	Colle. incanum	ATCC 64682^{T}	KC110807		KC110816
	Colle. indonesiense	CBS 127551 = CPC 14986^{T}	JQ948618	JQ949279	JQ949939
	Colle. jiangxiense	CGMCC 3.17361^{T} = LC3266 = LF488	KJ954850		OK236389
	Colle. johnstonii	CBS 128532 = ICMP 12926 = PRJ 1139.3^{T}	JQ948775	JQ949435	JQ950095
	Colle. kahawae	IMI 319418^{T} = ICMP 17816	JX010012	MZ673838	JX010444
	Colle. karstii	CBS 111998	JQ005299	JQ005473	JQ005646
	Colle. kinghornii	CBS 198.35^{T}	JQ948785	JQ949445	JQ950105
	Colle. laticiphilum	CBS 112989 = IMI 383015 = STE-U 5303^{T}	JQ948619	JQ949280	JQ949940
	Colle. lentis	CBS 127604 = DAOM 235316 = CT21^{T}	KM105597	JQ005808	JQ005850
	Colle. lilii	CBS 109214	GU228202		GU228104
	Colle. limetticola	CBS 114.14^{T}	JQ948523	JQ949184	JQ949844

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colle. limonicola*	CBS 142410 = CPC 31141	KY856296	KY856388	KY856554
	Colle. lindemuthianum	CBS 144.31^{T}	JX546712	JQ005821	JQ005863
	Colle. lineola	CBS 125337^{T}	GU228221	GU228025	GU228123
	Colle. lini	CBS 172.51^{T}	KM105581	JQ005807	JQ005849
	Colle. liriopes	CBS 119444^{T}	GU228196		GU228098
	Colle. lobatum	IMI 79736^{T}	MG600828	MG600912	MG601035
	Colle. lupini	CBS 109225 = BBA 70884^{T}	JQ948485	JQ949146	JQ949806
	Colle. magnum	CBS 519.97^{T}	MG600829	MG600913	MG601036
	Colle. makassarense	CBS 143664^{T}	MH728820		MH846563
	Colle. malvarum	CBS 521.97^{T} = LARS 720 = Lav-4	KF178504	KF178553	KF178601
	Colle. melonis	CBS 159.84^{T}	JQ948524	JQ949185	JQ949845
	Colle. merremiae	CBS 124955^{T}	MG600825	MG600910	MG601032
	Colle. musae	CMM4422	KX094189		KX094298
	Colle. musae	CMM4423	KX094195		KX094294
	Colle. musae	CMM4445	KX094188		KX094293
	Colle. musicola	CBS 132885^{T}	MG600798	MG600895	MG601003
	Colle. navitas	CBS 125086^{T}			JQ005853
	Colle. novae-zelandiae	CBS 128505 = ICMP 12944^{T}	JQ005315	JQ005489	JQ005662
	Colle. nupharicola	CBS 470.96^{T} = ICMP 18187	JX009972		JX010398
	Colle. nymphaeae	CBS 515.78^{T}	JQ948527	JQ949188	JQ949848
	Colle. ocimi	CBS 298.94^{T}	KM105577	KM105362	KM105502
	Colle. oncidii	CBS 129828^{T}	JQ005256	JQ005430	JQ005603
	Colle. orbiculare	CBS 570.97^{T} = LARS 73	KF178490	KF178539	KF178587
	Colle. orchidearum	CBS 135131^{T}	MG600800	MG600897	MG601005
	Colle. panamense	CBS 125386^{T}	MG600826	MG600911	MG601033
	Colle. pandanicola	MFLUCC 17-0571^{T}	MG646934		MG646926
	Colle. paranaense	CBS 134729 = Col 19 = CPC 20901^{T}	KC205026	KC205004	KC205060
	Colle. parsonsiae	CBS 128525 = ICMP 18590^{T}	JQ005320	JQ005494	JQ005667
	Colle. paxtonii	IMI 165753 = CPC 18868 ^{T}	JQ948615	JQ949276	JQ949936
	Colle. perseae	CBS 141365^{T} = GA100	KX620242		KX620341
	Colle. petchii	CBS 378.94^{T}	JQ005310	JQ005484	JQ005657
	Colle. phormii	CBS 118194 = AR 3546^{T}	JQ948777	JQ949437	JQ950097
	Colle. phyllanthi	CBS 175.67 = MACS 271^{T}	JQ005308	JQ005482	JQ005655

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colle. piperis*	CPC 21195^{T}	MG600820	MG600906	MG601027
	Colle. pisicola	CBS 724.97 = LARS 60^{T}	KM105522	KM105312	KM105452
	Colle. plurivorum	CBS 125474^{T}	MG600781	MG600887	MG600985
	Colle. proteae	CBS 132882^{T} = CPC 14859	KC297009	KC297045	KC297101
	Colle. pseudomajus	CBS 571.88^{T}	KF687826	KF687864	KF687883
	Colle. pseudotheobromicola	MFLUCC 18–1602^{T}	MH853675		MH853684
	Colle. psidii	CBS 145.29^{T} = ICMP 19120	JX009967		JX010443
	Colle. pyricola	CBS 128531 = ICMP 12924 = PRJ 977.1^{T}	JQ948776	JQ949436	JQ950096
	Colle. queenslandicum	CMM3233	MF110849		MF111058
	Colle. queenslandicum	CMM3241	MF110848		MF111059
	Colle. queenslandicum	CMM3236	MF110850		MF111060
	Colle. radicis	CBS 529.93^{T}	KF687825	KF687847	KF687869
	Colle. rhexiae	Coll1026 = BPI 884112 = CBS 133134^{T}	MZ664046	MZ673834	JX145179
	Colle. rhombiforme	CBS 129953 = PT250 = RB011^{T}	JQ948788	JQ949448	JQ950108
	Colle. riograndense	ICMP 20083^{T}	KM655298		KM655300
	Colle. salicis	CBS 607.94^{T}	JQ948791	JQ949451	JQ950111
	Colle. salsolae	ICMP 19051^{T}	JX009916		JX010403
	Colle. scovillei	CBS 126529 = PD 94/921-3 = BBA 70349^{T}	JQ948597	JQ949258	JQ949918
	Colle. siamense	CBS133123	KX094186		KX094289
	Colle. sidae	CBS 504.97^{T}	KF178497	KF178545	KF178593
	Colle. simmondsii	BRIP 28519 = CBS 122122^{T}	JQ948606	JQ949267	JQ949927
	Colle. sloanei	IMI 364297 = CPC 18929^{T}	JQ948617	JQ949278	JQ949938
	Colle. sojae	ATCC 62257^{T}	MG600810	MG600899	MG601016
	Colle. spaethianum	CBS 167.49^{T}	GU228199		GU228101
	Colle. spinaceae	CBS 128.57	GU228239	GU228043	GU228141
	Colle. spinosum	CBS 515.97^{T} = LARS 465 = DAR 48942	KF178498	KF178547	KF178595
	Colle. sublineola	CBS 131301^{T} = S3.001			JQ005855
	Colle. tabacum	N150 = CPC 18945^{T}	KM105557	KM105344	KM105484
	Colle. tainanense	CBS 143666^{T}	MH728823		MH846558
	Colle. tamarilloi	CBS 129814 = T.A.6^{T}	JQ948514	JQ949175	JQ949835
	Colle. tebeestii	CBS 522.97^{T} = LARS 733 = 83-43	KF178505	KF178546	KF178594
	Colle. temperatum	CBS 133122^{T} = Coll883 = BPI 884100	MZ664045	MZ673833	JX145211
	Colle. theobromicola	CMM4242	KX094173		KX094278

Supplementary Table S4.1 continued. *Colletotrichum* sampling informed by Vieira et al. (2020) and Liu, Ma et al. (2022).

	Name	Voucher	GAPDH	HIS3	TUB2
***Colletotrichum* (Figure 4.3G)**	*Colle. theobromicola*	CMM3214	MF110847		MF111049
	Colle. theobromicola	CMM3221	MF110855		MF111048
	Colle. tofieldiae	CBS 495.85	GU228193		GU228095
	Colle. torulosum	CBS 128544 = ICMP 18586^{T}	JQ005251	JQ005425	JQ005598
	Colle. trifolii	CBS 158.83^{T}	KF178502	KF178551	KF178599
	Colle. tropicale	CMM4243	KU213601		KU213604
	Colle. tropicale	CMM2999	MF110846		MF111088
	Colle. tropicicola	CBS 127555	MG600778	MG600884	MG600982
	Colle. truncatum	CBS 151.35^{T}	GU228254		GU228156
	Colle. utrechtense	CBS 130243^{T}	KM105554	KM105341	KM105481
	Colle. verruculosum	IMI 45525^{T}	GU228198		GU228100
	Colle. vietnamense	CBS 125478^{T}	KF687832	KF687855	KF687877
	Colle. vignae	CBS 501.97 = LARS 56^{T}	KM105534	KM105323	KM105463
	Colle. viniferum	GZAAS 5.08601^{T} = yg1	JN412798		
	Colle. vittalense	CBS 181.82^{T}	MG600796	MG600893	MG601001
	Colle. walleri	CBS 125472 = BMT(HL)19^{T}	JQ948605	JQ949266	JQ949926
	Colle. wuxiense	CGMCC 3.17894^{T}	KU252045		KU252200
	Colle. xanthorrhoeae	BRIP 45094^{T} = ICMP 17903 = CBS 127831	JX009927		JX010448
	Colle. xishuangbannaense	MFLUCC 19-0107^{T}	MW537586		
	Colle. yulongense	CFCC 50818^{T}	MK108986		MK108987
	Colle. yunnanense	CBS 132135	JX546706	JX546755	JX519248
	Monilochaetes infuscans	CBS 869.96	JX546612	JQ005822	JQ005864

Supplementary Table S4.1 continued. *Didymella* sampling informed by Chen, Hou et al. (2017) and Scarpari et al. (2020).

	Name	Voucher	RPB2	TUB2
***Didymella* (Figure 4.3A)**	*Didymella acetosellae*	CBS 179.97	KP330415	GU237575
	D. aeria	LC 8120	KY742138	KY742294
	D. aeria	CGMCC 3.18353^{T}	KY742137	KY742293
	D. aliena	CBS 379.93 = PD 82/945	KP330416	GU237578
	D. americana	CBS 185.85 = PD 80/1191	KT389594	FJ427088
	D. anserina	CBS 253.80	KT389595	KT389795
	D. aquatica	CGMCC 3.18349^{T}	KY742140	KY742297
	D. arachidicola	CBS 333.75^{T} = ATCC 28333 = IMI 386092 = PREM 44889	KT389598	GU237554
	D. aurea	CBS 269.93^{T} = PD 78/1087	KT389599	GU237557
	D. bellidis	CBS 714.85 = PD 74/265	KP330417	GU237586
	D. boeremae	CBS 109942^{T} = PD 84/402	KT389600	FJ427097
	D. brunneospora	CBS 115.58 = DSM 62044	KT389625	KT389802
	D. calidophila	CBS 448.83^{T}		FJ427168
	D. chenopodii	CBS 128.93 = PD 79/140	KT389602	GU237591
	D. chloroguttulata	CGMCC 3.18351^{T}	KY742142	KY742299
	D. coffeae-arabicae	CBS 123380^{T} = PD 84/1013	KT389603	FJ427104
	D. corylicola	CREADC-F2281	MN958321	MN958331
	D. corylicola	CREADC-F2405	MN958324	MN958334
	D. corylicola	CREADC-F2406	MN958325	MN958335
	D. corylicola	CREADC-F2407	MN958326	MN958336
	D. corylicola	CREADC-F2408	MN958327	MN958337
	D. curtisii	CBS 251.92 = PD 86/1145		FJ427148
	D. dactylidis	CBS 124513^{T} = PD 73/1414		GU237599
	D. dimorpha	CBS 346.82^{T}		GU237606
	D. ellipsoidea	CGMCC 3.18350^{T}	KY742145	KY742302
	D. eucalyptica	CBS 377.91 = PD 79/210	KT389605	GU237562
	D. exigua	CBS 183.55^{T}	EU874850	GU237525
	D. gardeniae	CBS 626.68^{T} = IMI 108771	KT389606	FJ427114
	D. glomerata	CBS 133.72		FJ427115
	D. glomerata	CBS 528.66^{ET} = PD 63/590	GU371781	FJ427124
	D. glomerata	ATCC MYA-2373	MZ073895	MZ073910

Supplementary Table S4.1 continued. *Didymella* sampling informed by Chen, Hou et al. (2017) and Scarpari et al. (2020).

	Name	Voucher	RPB2	TUB2
***Didymella* (Figure 4.3A)**	*D. glomerata*	CBS 126930	MN983465	MN983856
	D. glomerata	UTHSC:DI16-205	LT593043	LT592974
	D. heteroderae	CBS 109.92^{T} = PD 73/1405	KT389601	FJ427098
	D. ilicicola	CGMCC 3.18355^{T}	KY742150	KY742307
	D. ilicicola	LC 8127	KY742151	KY742308
	D. infuscatispora	CGMCC 3.18356^{T}	KY742152	KY742309
	D. keratinophila	CBS 143032 = UTHSC:DI16–200 = FMR 13690	LT593039	LT592970
	D. lethalis	CBS 103.25	KT389607	GU237564
	D. longicolla	CBS 124514^{T} = PD 80/1189		GU237622
	D. longicolla	CBS 503.71	MN983480	MN983866
	D. longicolla	CBS 347.82	MT018160	GU237621
	D. macrophylla	CGMCC 3.18357^{T}	KY742154	KY742312
	D. macrostoma	CBS 482.95	KT389609	GU237626
	D. maydis	CBS 588.69^{T}	GU371782	FJ427190
	D. microchlamydospora	CBS 105.95^{T}	KP330424	FJ427138
	D. molleriana	CBS 229.79 = LEV 7660	KP330418	GU237605
	D. musae	CBS 463.69	LT623248	FJ427136
	D. negriana	CBS 358.71	KT389610	GU237635
	D. nigricans	CBS 444.81^{T} = PDDCC 6546		GU237558
	D. ocimicola	CGMCC 3.18358^{T}		KY742320
	D. pedeiae	CBS 124517^{T} = PD 92/612A	KT389612	GU237642
	D. pinodella	CBS 531.66	KT389613	FJ427162
	D. pinodes	CBS 525.77^{T}	KT389614	GU237572
	D. pomorum	CBS 285.76 = ATCC 26241 = IMI 176742 = VKM F-1843	KT389615	FJ427163
	D. pomorum	CBS 388.80	KT389617	FJ427165
	D. pomorum	CBS 539.66 = ATCC 16791 = IMI 122266 = PD 64/914	KT389618	FJ427166
	D. pomorum	CBS 354.52	KT389616	KT389799
	D. protuberans	CBS 381.96^{T} = PD 71/706	KT389620	GU237574
	D. pteridis	CBS 379.96^{T}	KT389624	KT389801
	D. rhei	CBS 109177 = LEV 15165 = PD 2000/9941	KP330428	GU237653
	D. rosea	BRIP 50788		KT286945

Supplementary Table S4.1 continued. *Didymella* sampling informed by Chen, Hou et al. (2017) and Scarpari et al. (2020).

	Name	Voucher	RPB2	TUB2
***Didymella* (Figure 4.3A)**	*D. rumicicola*	CBS 683.79^{T} = LEV 15094	KT389622	KT389800
	D. sancta	CBS 281.83^{T}	KT389623	FJ427170
	D. segeticola	CGMCC 3.17489^{T}	KP330414	KP330399
	D. senecionicola	CBS 160.78 = LEV 11451		GU237657
	D. sinensis	LC 8142	KY742166	KY742329
	D. subglomerata	CBS 110.92 = PD 76/1010	KT389626	FJ427186
	D. subherbarum	CBS 250.92^{T} = DAOM 171914 = PD 92/371		GU237659
	D. subherbarum	CBS 249.92 = PD 78/1088		GU237658
	D. suiyangensis	CGMCC 3.18352^{T}	KY742168	KY742331
	D. tanaceti	BRIP 50785		KT286974
	D. viburnicola	CBS 523.73 = PD 69/800	KP330430	GU237667
	Macroventuria anomochaeta	CBS 525.71	GU456346	GU237544
	Macroventuria wentii	CBS 526.71	KT389642	GU237546
	Paraboeremia adianticola	CBS 187.83 = PD 82/128	KP330401	GU237576
	Paraboeremia putaminum	CBS 130.69 = CECT 20054 = IMI 331916	LT623254	GU237652
	Paraboeremia selaginellae	CBS 122.93 = PD 77/1049	LT623255	GU237656

Supplementary Table S4.1 continued. *Didymosphaeriaceae* sampling informed by Karácsony et al. (2021) and Wanasinghe and Mortimer (2022).

	Name	Voucher	LSU	RPB2	EF1α	TUB2
***Didymosphaeriaceae* (Figure 4.3D)**	*Alloconiothyrium camelliae*	NTUCC 17-032-1^{T}	MT071270		MT232967	MT308624
	Austropleospora keteleeriae	MFLUCC 18-1551^{T}	NG_070075	MK434909	MK360045	
	Austropleospora ochracea	KUMCC 20-0020^{T}	MT799860		MT872714	
	Austropleospora osteospermi	MFLUCC 17-2429^{T}	MK347974	MK434884	MK360044	
	Bambusistroma didymosporum	MFLU 15-0057^{T}	KP761730	KP761720	KP761727	
	Bimuria novae-zelandiae	CBS 107.79^{T}	AY016356	DQ470917	DQ471087	
	Bimuria omanensis	SQUCC 15280^{T}	NG_071257		MT279046	
	Chromolaenicola lampangensis	MFLUCC 17-1462^{T}	MN325004	MN335654	MN335649	
	Chromolaenicola nanensis	MFLUCC 17-1477	MN325002	MN335653	MN335647	
	Chromolaenicola thailandensis	MFLUCC 17-1475^{T}	MN325007	MN335656	MN335652	
	Cylindroaseptospora leucaenae	MFLUCC 17-2424^{T}	NG_066310		MK360047	
	Cylindroaseptospora siamensis	MFLUCC 17-2527	NG_066311		MK360048	
	Deniquelata barringtoniae	MFLUCC 16-0271	MH260291	MH412753	MH412766	
	Deniquelata hypolithi	CBS 146988	MZ064486	MZ078201	MZ078250	
	Deniquelata vittalii	NFCCI4249^{T}	MF182395	MF168942	MF182398	
	Didymocrea sadasivanii	CBS 438.65^{T}	DQ384103			
	Didymosphaeria rubi ulmifolii	CBS 100299	JX496124			JX496350
	Didymosphaeria rubi ulmifolii	MFLUCC 14-0023^{T}	KJ436586			KJ939277
	Didymosphaeria variabile	18EPLE013			MT881834	MT881920
	Didymosphaeria variabile	CBS 638.93	JX496215			JX496441
	Didymosphaeria variabile	18EPLE021			MT881841	MT881928
	Kalmusia cordylines	ZHKU 21-0003	OL818333			
	Kalmusia ebuli	CBS 123120^{T}	JN644073			
	Kalmusia erioi	MFLU 18-0832	MN473052		MN481599	MN481603
	Kalmusia italica	MFLUCC 14-0566	KP325441			
	Kalmusia longispora	CBS 582.83^{T}	MH873371			JX496436
	Kalmusia sarothamni	CBS 116474	KF796673			
	Kalmusia sarothamni	CBS 113833	KF796671			
	Kalmusia sp.	K			MW692012	MW692021
	Kalmusia sp.	UTHSC DI16-256	LN907399	LT797014	LT797094	LT796934
	Kalmusia variisporum	CBS 121517^{T}	JX496143			JX496369
	Kalmusibambusa triseptata	MFLUCC 13-0232^{T}	KY682695			

Supplementary Table S4.1 continued. *Didymosphaeriaceae* sampling informed by Karácsony et al. (2021) and Wanasinghe and Mortimer (2022).

	Name	Voucher	LSU	RPB2	EF1α	TUB2
***Didymosphaeriaceae* (Figure 4.3D)**	*Karstenula rhodostoma*	CBS 690.94	GU301821	GU371788	GU349067	
	Karstenula rhodostoma	CBS 691.94	AB807531		AB808506	
	Laburnicola muriformis	MFLUCC 16-0290^{T}	KU743198		KU743213	KU743214
	Laburnicola rhizohalophila	CGMCC 8756	KJ125523	KJ125524	KJ125525	
	Letendraea cordylinicola	MFLUCC 11-0148^{T}	NG_059530			
	Letendraea helminthicola	CBS 884.85	AY016362	MK404164	MK404174	
	Letendraea padouk	CBS 485.70	AY849951			
	Montagnula bellevaliae	MFLUCC 14-0924^{T}	KT443902		KX949743	
	Montagnula chromolaenicola	MFLUCC 17-1469	NG_070948	MT235809	MT235773	
	Montagnula cirsii	MFLUCC 13-0680^{T}	KX274249		KX284707	
	Montagnula krabiensis	MFLUCC 16-0250^{T}	MH260303		MH412776	
	Montagnula thailandica	MFLUCC 17-1508	NG_070949	MT235810	MT235774	
	Neokalmusia arundinis	MFLU 17-0754	MT649878		MT663766	
	Neokalmusia brevispora	KT 2313	AB524601	AB539100	AB539113	
	Neokalmusia didymospora	MFLUCC 11-0613	KP091434			
	Neokalmusia jonahhulmei	KUMCC 21-0818	ON007039	ON009137	ON009133	
	Neokalmusia kunmingensis	KUMCC 18-0120	MK079889		MK070172	
	Neokalmusia scabrispora	KT 1023	AB524593	AB539093	AB539106	
	Neokalmusia thailandica	MFLUCC 16-0405	NG_059792	KY706148	KY706145	
	Neptunomyces aureus	CMG12			MK948000	MK934132
	Paracamarosporium hawaiiense	CBS 120025^{T}	JX496140			JX496366
	Paraconiothyrium cyclothyrioides	CBS 972.95^{T}	JX496232			JX496458
	Paraconiothyrium estuarinum	CBS 109850^{T}	JX496129			JX496355
	Paramassariosphaeria anthostomoides	CBS 615.86	GU205223			
	Paramassariosphaeria clematidicola	MFLU 16-0172^{T}	KU743207			
	Paraphaeosphaeria rosae	MFLUCC 17-2547^{T}	MG829044		MG829222	
	Phaeodothis winteri	CBS 182.58	GU301857			
	Pseudocamarosporium eucalypti	CBS 146084^{T} = CPC 37995	MN567657		MN556833	
	Pseudocamarosporium pteleae	MFLUCC 17-0724^{T}	MG829061		MG829233	
	Pseudodidymocyrtis lobariellae	KRAM Flakus 25130^{T}	NG_068933			
	Pseudopithomyces entadae	MFLUCC 17-0917^{T}	NG_066305	MK434899	MK360083	
	Pseudopithomyces kunmingnensis	MFLUCC 17-0314^{T}	MF173605			

Supplementary Table S4.1 continued. *Didymosphaeriaceae* sampling informed by Karácsony et al. (2021) and Wanasinghe and Mortimer (2022).

	Name	Voucher	LSU	RPB2	EF1α	TUB2
Didymosphaeriaceae	*Pseudopithomyces rosae*	MFLUCC 15-0035^{T}	MG829064			
	Spegazzinia radermacherae	MFLUCC 17-2285^{T}	MK347957	MK434893	MK360088	
	Spegazzinia tessarthra	SH 287	AB807584		AB808560	
	Tremateia arundicola	MFLU 16-1275^{T}	KX274248		KX284706	
	Tremateia chromolaenae	MFLUCC 17-1425	NG_068710	MT235816	MT235778	
	Tremateia guiyangensis	GZAAS01^{T}	KX274247		KX284705	
	Tremateia murispora	GZCC 18-2787^{T}	MK972751		MK986482	
	Tremateia thailandensis	MFLUCC 17-1430	NG_068711	MT235819	MT235781	
	Verrucoconiothyrium nitidae	CBS 119209	EU552112			
	Xenocamarosporium acaciae	CPC 24755^{T}	NG_058163			
	Xenocamarosporium acaciae	MFLUCC 17-2432	MK347983		MK360093	

Supplementary Table S4.1 continued. *Gnomoniopsis* sampling informed by Jiang et al. (2021).

	Name	Voucher	EF1α	TUB2
***Gnomoniopsis* (Figure 4.3F)**	*Gnomoniopsis alderdunensis*	CBS 125680^{T}	GU320801	GU320787
	G. castanopsidis	CFCC 54437^{T}	MZ936385	
	G. chamaemori	CBS 804.79	GU320809	GU320777
	G. chinensis	CFCC 52286^{T}	MH545370	MH545366
	G. chinensis	CFCC 52287	MH545371	MH545367
	G. chinensis	CFCC 52288	MH545372	MH545368
	G. chinensis	CFCC 52289	MH545373	MH545369
	G. clavulata	CBS 121255	GU320807	EU219211
	G. clavulata	AR 4124	EU221977	EU219167
	G. clavulata	AR 4183	EU221965	EU219190
	G. clavulata	AR 4317 = BPI 877443	EU221938	EU219214
	G. comari	CBS 806.79	GU320810	EU219156
	G. daii	CFCC 54043^{T}	MN605519	MN605517
	G. fagacearum	CFCC 54316^{T}	MZ936392	MZ936408
	G. fragariae	CBS 121226	GU320792	EU219144
	G. guangdongensis	CFCC 54443^{T}	MZ936394	MZ936410
	G. hainanensis	CFCC 54376^{T}	MZ936397	MZ936413
	G. idaeicola	CBS 125672	GU320797	GU320781
	G. macounii	CBS 121468	GU320804	EU219126
	G. occulta	CBS 125677	GU320812	GU320785
	G. paraclavulata	CBS 123202	GU320815	GU320775
	G. paraclavulata	66G	MZ078875	MZ078820
	G. paraclavulata	477E	MZ078874	MZ078819
	G. paraclavulata	396E	MZ078873	MZ078818
	G. racemula	CBS 121469^{T}	GU320803	EU219125
	G. rossmaniae	CFCC 54307^{T}	MZ936399	MZ936415
	G. sanguisorbae	CBS 858.79	GU320805	GU320790
	G. silvicola	CFCC 54418^{T}	MZ936402	MZ936418
	G. smithogilvyi	CBS 130190^{T}	KR072534	JQ910639
	G. smithogilvyi	CBS 130189	KR072535	JQ910641
	G. smithogilvyi	CBS 130188	KR072536	JQ910640
	G. smithogilvyi	MUT 401	KR072537	KR072532

Supplementary Table S4.1 continued. *Gnomoniopsis* sampling informed by Jiang et al. (2021).

Name	Voucher	EF1α	TUB2
G. smithogilvyi	MUT 411	KR072538	KR072533
G. tormentillae	CBS 904.79	GU320795	EU219165
G. xunwuensis	CFCC 53115^{T}	MK578141	MK578067
Melanconis alni	AR 3500	EU221896	EU219102
M. marginalis	AR 3442	EU221991	EU219103

Supplementary Table S4.1 continued. *Neocosmospora* sampling informed by Crous, Lombard et al. (2021).

	Name	Voucher	acl1	cmdA	RPB1	RPB2	EF1α
	Geejayessia atrofusca	CBS 125482 = DAOM 238117			MW834196	HQ897775	MW834282
	G. celtidicola	CBS 125502^{T}	HM626625		MW834197	MW834013	HM626638
	G. cicatricum	CBS 125550			MW834198	HQ897697	HM626642
	Neocosmospora acutispora	CBS 145461^{T} = NRRL 22574 = BBA 62213	MW834050	MW834122	MW834210	LR583814	LR583593
	Neoco. addoensis	CBS 146510^{T} = CPC 37128	MW218005	MW218052	MW218098	MW446575	MW248741
	Neoco. ambrosia	CBS 571.94^{ET} = NRRL 22346 = BBA 65390 = MAFF 246287			MW834211	EU329503	FJ240350
	Neoco. ampla	CBS 202.32^{T} = BBA 4170	MW834051	MW834123	MW834212	LR583815	LR583594
	Neoco. bataticola	CBS 144398^{T} = NRRL 22402 = BBA 64954 = FRC S-0567	MW218007	MW218054	MW218100	FJ240381	AF178344
	Neoco. borneensis	CBS 145462^{ET} = NRRL 22579 = BBA 65095 = GJS 85-197	MW834052	MW834124	MW834213	EU329515	AF178352
	Neoco. bostrycoides	CBS 144.25 NT	MW218008	MW218055	MW218101	LR583818	LR583597
	Neoco. brevicona	CBS 204.31^{ET} = NRRL 22659 = BBA 2123	MW218010	MW218057	MW218103	LR583821	LR583600
	Neoco. brevis	CBS 130326 = NRRL 28009 = CDC B-5543	MW834053	MW834125	MW834214	EF470136	DQ246869
	Neoco. catenata	CBS 143229^{T} = NRRL 54993 = U THSC 09-1009	MW218012	MW218059	MW218105	KC808355	KC808214
	Neoco. citricola	CBS 146513^{T} = CPC 37131	MW218015	MW218062	MW218108	MW446581	MW248747
	Neoco. crassa	CBS 144386^{T} = MUCL 11420	MW218016	MW218063	MW218109	LR583823	LR583604
	Neoco. cryptoseptata	CBS 145463^{T} = NRRL 22412 = BBA 65024	MW834054	MW834126	MW834215	EU329510	AF178351
	Neoco. cucurbitae	CBS 410.62 = NRRL 22658 = CECT 2864	MW834055	MW834127	MW834216	LR583824	DQ247640
	Neoco. cucurbitae	CBS 616.66^{T} = NRRL 22399 = BBA 64411	MW834056	MW834128	MW834217	LR583825	DQ247592
	Neoco. cyanescens	CBS 518.82^{T}	MW218017	MW218064	MW218110	LR583826	LR583605
	Neoco. diminuta	CBS 144390^{T} = MUCL 18798	MW834057	MW834129	MW834218	LR583828	LR583607
	Neoco. elegans	CBS 144396^{ET} = NRRL 22277 = MAFF 238541 = ATCC 42366	MW218020	MW218067	MW218113	FJ240380	AF178336
	Neoco. epipeda	CBS 146523^{T} = CPC 38310	MW834058	MW834130	MW834219	MW834022	MW834285
	Neoco. euwallaceae	CBS 135854^{T} = NRRL 54722			JQ038021	JQ038028	JQ038007
	Neoco. falciformis	CBS 475.67^{T} = IMI 268681	MW218021	MW218068	MW218114	LT960558	LT906669
	Neoco. ferruginea	CBS 109028^{T} = NRRL 32437	MW834060	MW834132	MW834221	EU329581	DQ246979
	Neoco. floridana	NRRL 62628^{T} = MAFF 246849			KC691593	KC691624	KC691535
	Neoco. gamsii	CBS 143207^{T} = NRRL 32323 = UTHSC 99-205	MW834062	MW834134	MW834223	EU329622	DQ247103
	Neoco. gamtoosensis	CBS 146502^{T} = VG16 = CPC 37120	MW218023	MW218070	MW218116	MW446611	MW248762

***Neocosmospora* (Figure 4.3J)**

Supplementary Table S4.1 continued. *Neocosmospora* sampling informed by Crous, Lombard et al. (2021).

	Name	Voucher	acl1	cmdA	RPB1	RPB2	EF1α
***Neocosmospora* (Figure 4.3J)**	*Neoco. haematococca*	CBS 119600^{ET} = FRC S-1832	MW834064	MW834136		LT960561	DQ247510
	Neoco. hypothenemi	CBS 145464^{T} = NRRL 52782 = ARSEF 5878	MW218024		MW218117	JF741176	JF740850
	Neoco. illudens	CBS 147303 = NRRL 22090 = BBA 67606 = GJS 82-98	MW834065	MW834137	JX171488	JX171601	AF178326
	Neoco. ipomoeae	CBS 353.87 = NRRL 22657	MW218026	MW218072	MW218119	LR583831	DQ247639
	Neoco. keleraja	CBS 125720 PT = FRC S-1837 = GJS 02-114	MW834066	MW834138	MW834225	LR583834	LR583612
	Neoco. keratoplastica	CBS 490.63^{T}	MW218028	MW218074	MW218121	LT960562	LT906670
	Neoco. kuroshio	CBS 142642^{T}	MW834068	MW834140	MW834227	LR583837	KX262216
	Neoco. kurunegalensis	CBS 119599^{T} = GJS 02-94	MW834069	MW834141	MW834228	LR583838	DQ247511
	Neoco. lerouxii	CBS 146514^{T} = CPC 37132	MW218030	MW218076	MW218123	MW446617	MW248768
	Neoco. lichenicola	CBS 623.92^{ET}	MW834071	MW834143		LR583845	LR583620
	Neoco. liriodendri	CBS 117481^{T} = NRRL 22389 = BBA 67587 = GJS 91-148	MW218031	MW218077	MW218124	EU329506	AF178340
	Neoco. longissima	CBS 126407^{T} = GJS 85-72	MW834072	MW834144	MW834230	LR583846	LR583621
	Neoco. macrospora	CBS 142424^{T} = CPC 28191	MW218032	MW218078	MW218125	LT746331	LT746218
	Neoco. mahasenii	CBS 119594^{T}	MW834073	MW834145	MW834231	LT960563	DQ247513
	Neoco. martii	CBS 115659^{ET} = FRC S-0679 = MRC 2198	MW834074	MW834146	MW834232	JX435256	JX435156
	Neoco. merkxiana	CBS 146525^{T}	MW834075	MW834147	MW834233	MW834025	MW834288
	Neoco. metavorans	CBS 135789^{T}	MW218034	MW218080	MW218127	LR583849	LR583627
	Neoco. mori	CBS 145467^{T} = NRRL 22230 = MAFF 238539	MW834077	MW834149	MW834235	EU329499	AF178358
	Neoco. neerlandica	CBS 232.34^{T}	MW834079	MW834151	MW834237	MW847903	MW847906
	Neoco. nelsonii	CBS 309.75^{T}	MW834080	MW834152	MW834238	MW847904	MW847907
	Neoco. nirenbergiana	CBS 145469^{T} = NRRL 22387 = BBA 65023 = GJS 87-127	MW834081	MW834153		EU329505	AF178339
	Neoco. noneumartii	CBS 115658^{T} = FRC S-0661	MW218036	MW218082	MW218129	MW446618	LR583630
	Neoco. obliquiseptata	NRRL 62611 = MAFF 246845			KC691606	KC691637	KC691548
	Neoco. oblonga	CBS 130325^{T} = NRRL 28008 = CDC B-4701	MW834082	MW834154	MW834239	LR583853	LR583631
	Neoco. oligoseptata	CBS 143241^{T} = NRRL 62579 = FRC S-2581 = MAFF 246283	MW834083	MW834155	KC691596	LR583854	KC691538
	Neoco. paraeumartii	CBS 487.76^{T} = NRRL 13997 = BBA 62215	MW834084	MW834156	MW834240	LR583855	DQ247549
	Neoco. parceramosa	CBS 115695^{T}	MW218037	MW218083		JX435249	JX435149
	Neoco. perseae	CBS 144142^{T} = CPC 26829	MW218038	MW218084	MW218130	LT991909	LT991902
	Neoco. petroliphila	CBS 203.32 = NRRL 13952	MW218039	MW218085	MW218131	LR583857	DQ246835

Supplementary Table S4.1 continued. *Neocosmospora* sampling informed by Crous, Lombard et al. (2021).

	Name	Voucher	acl1	cmdA	RPB1	RPB2	EF1α
***Neocosmospora* (Figure 4.3J)**	*Neoco. phaseoli*	CBS 265.50	MW834085	MW834157		KJ511278	FJ919464
	Neoco. piperis	CBS 145470^{T} = NRRL 22570 = GJS 89-14 = CML 1888	MW834086	MW834158	MW834241	EU329513	AF178360
	Neoco. piperis	CML 3171				KT943484	KT943486
	Neoco. piperis	CML 3178				KT943485	KT943487
	Neoco. pisi	CBS 123669^{ET} = NRRL 45880 = ATCC MYA-4622	MW834087	MW834159	MW834242	LR583862	LR583636
	Neoco. plagianthi	NRRL 22632 = GJS 83-146			JX171501	JX171614	AF178354
	Neoco. protoensiformis	CBS 145471^{T} = NRRL 22178 = GJS 90-168	MW834089	MW834161	MW834244	EU329498	AF178334
	Neoco. pseudensiformis	CBS 130.78 = NRRL 22575 = NRRL 22653	MW834090	MW834162	MW834245	LR583868	DQ247635
	Neoco. pseudopisi	CBS 266.50	MW834091	MW834163	MW834246	MW834027	MW834290
	Neoco. pseudoradicicola	CBS 145472^{T} = NRRL 25137 = ARSEF 2313	MW218041	MW218087	MW218133	JF741084	JF740757
	Neoco. quercicola	CBS 141.90^{T} = NRRL 22652	MW834092	MW834164	MW834247	LR583869	DQ247634
	Neoco. rectiphora	CBS 125727^{T} = GJS 02-89 = FRC S-1831	MW834094	MW834166	MW834249	LR583871	DQ247509
	Neoco. regularis	CBS 230.34^{T}	MW834096	MW834168		MW834029	LR583643
	Neoco. rekana	CMW 52862^{T}				MN249137	MN249151
	Neoco. robusta	CBS 145473^{T} = NRRL 22395 = BBA 65682		MW834169	MW834251	EU329507	AF178341
	Neoco. samuelsii	CBS 114067^{T} = GJS 89-70	MW834097	MW834170	MW834252	LR583874	LR583644
	Neoco. silvicola	CBS 123846^{T} = GJS 04-147	MW834099	MW834172	MW834254	LR583876	LR583646
	Neoco. solani	CBS 140079^{ET} = NRRL 66304 = GJS 09-1466 = FRC S-2364	MW218042	MW218088	MW218134	KT313623	KT313611
	Neoco. spathulata	CBS 145474^{T} = NRRL 28541 = UTHSC 98-1305	MW218045	MW218091	MW218137	EU329542	DQ246882
	Neoco. stercicola	CBS 142481^{T} = DSM 106211	MW834100	MW834173	MW834255	LR583887	LR583658
	Neoco. suttoniana	CBS 143214^{T} = NRRL 32858	MW218046	MW218092	MW218138	EU329630	DQ247163
	Neoco. tonkinensis	CBS 115.40^{T}	MW218048	MW218094	MW218140	LT960564	LT906672
	Neoco. tuaranensis	NRRL 22231^{T} = ATCC 16563 = MAFF 246842			KC691600	KC691631	KC691542
	Neoco. vasinfecta	CBS 533.65 = IMI 302625	MW834103	MW834176	MW834258	LR583899	LR583671

Supplementary Table S4.1 continued. *Neocucurbitaria* sampling informed by Wanasinghe, Phookamsak et al. (2017), Jaklitsch et al. (2018), Valenzuela-Lopez, Cano-Lira, Guarro et al. (2018) and Crous, Schumacher et al. (2019).

	Name	Voucher	RPB2	EF1α	TUB2
***Neocucurbitaria* (Figure 4.3E)**	*Cucurbitaria berberidis*	CBS 142401 = C241	MF795798	MF795845	MF795886
	Cucurbitaria oromediterranea	CBS 142399 = C229^{T}	MF795803	MF795849	MF795890
	Neocucurbitaria acanthocladae	CBS 142398 = C225^{T}	MF795808	MF795854	MF795894
	Neocu. acerina	CBS 142403 = C255	MF795810	MF795856	MF795896
	Neocu. aetnensis	CBS 142404 = C261^{T}	MF795811	MF795857	MF795897
	Neocu. aquatica	CBS 297.74	LT623278		LT623238
	Neocu. cava	CBS 115979	LT623273		LT623234
	Neocu. cava	CBS 257.68^{T}	LT717681		KT389844
	Neocu. cava	CBS 143400	MH108005		MH108046
	Neocu. cava	MF-Vm17-040			MZ054692
	Neocu. cinereae	CBS 142406 = KU9^{T}	MF795813	MF795859	MF795899
	Neocu. cisticola	CBS 142402 = C244^{T}	MF795814	MF795860	MF795900
	Neocu. hakeae	CBS 142109	KY173593		KY173613
	Neocu. irregularis	CBS 142791	LT593054		LT592985
	Neocu. juglandicola	CBS 142390 = BW6^{T}	MF795815	MF795861	MF795901
	Neocu. keratinophila	CBS 121759^{T}	LT623275		LT623236
	Neocu. populi	CBS 142393 = C28^{T}	MF795816	MF795862	MF795902
	Neocu. quercina	CBS 115095^{T}	LT623277		LT623237
	Neocu. rhamni	CBS 142391 = C1^{T}	MF795817	MF795863	
	Neocu. rhamnicola	CBS 142396 = C185^{T}	MF795822	MF795868	MF795906
	Neocu. rhamnioides	CBS 142395 = C118^{T}	MF795824	MF795870	MF795908
	Neocu. ribicola	CBS 142394 = C55^{T}	MF795827	MF795873	MF795911
	Neocu. unguis-hominis	CBS 111112	LT623279		LT623239
	Neocu. vachelliae	CBS 142397 = C192^{T}	MF795829	MF795875	MF795913
	Neopyrenochaeta acicola	CBS 812.95^{T}	LT623271		LT623232
	Parafenestella pseudoplatani	CBS 142392 = C26^{T}	MF795830	MF795876	MF795914
	Protofenestella ulmi	CBS 143000 = FP5^{T}	MF795833	MF795879	MF795915

Supplementary Table S4.1 continued. *Neodidymelliopsis* sampling informed by Chen, Hou et al. (2017), Hyde, Tennakoon et al. (2019) and Hou et al. (2020).

	Name	Voucher	RPB2	TUB2
***Neodidymelliopsis* (Figure 4.3C)**	*Calophoma clematidina*	CBS 108.79	KT389588	FJ427100
	Calophoma complanata	CBS 100311	KT389590	GU237594
	Calophoma glaucii	CBS 112.96	MT018230	GU237610
	Neoascochyta cylindrispora	UTHSC DI16-352	LT593101	LT593031
	Neoascochyta desmazieri	CBS 346.86	MT018304	MT005730
	Neoascochyta europaea	CBS 504.71	MT018314	MT005738
	Neodidymelliopsis achlydis	CBS 256.77^{T}		KT389829
	Neod. cannabis	CBS 121.75^{T}		GU237535
	Neod. cannabis	CBS 234.37	KP330403	GU237523
	Neod. cannabis	CBS 591.67		KT389826
	Neod. cannabis	CBS 629.76		KT389827
	Neod. farokhinejadii	CBS 142853	KY464922	KY449023
	Neod. longicolla	CBS 382.96^{T}	MT018298	KT389830
	Neod. sp.	CBS 141235		KX033382
	Neod. longicolla	CBS 265.74	MT018296	MT005725
	Neod. longicolla	CBS 266.74	MT018297	MT005726
	Neod. moricola	MFLUCC 17-1063	KY684943	KY684937
	Neod. moricola	MFLUCC 17-1064^{T}	KY684944	KY684938
	Neod. negundinis	MFLUCC 18-0083^{T}	MG564166	MG564164
	Neod. polemonii	CBS 109181^{T} = PD 83/757	KP330427	GU237648
	Neod. polemonii	CBS 375.67	MT018291	KT389828
	Neod. ranunculi	CBS 739.88	MT018295	MT005724
	Neod. ranunculi	CBS 286.72	MT018294	MT005723
	Neod. sambuci	MFLUCC 18-1565		MK049556
	Neod. tiliae	CBS 139719	MT018286	MT005720
	Neod. tiliae	CBS 519.95^{T}	MT018287	MT005721
	Neod. xanthina	CBS 383.68^{T}	KP330431	GU237688
	Neod. xanthina	CBS 168.70	MT018290	KT389831
	Xenodidymella asphodeli	CBS 499.72	MT018282	KT389853
	Xenodidymella catariae	CBS 102635	KP330404	GU237524
	Xenodidymella humicola	CBS 220.85	KP330422	GU237617

Supplementary Table S4.1 continued. *Neodidymelliopsis* sampling informed by Chen, Hou et al. (2017), Hyde, Tennakoon et al. (2019) and Hou et al. (2020).

Name	Voucher	RPB2	TUB2
Xenodidymella applanata	CBS 195.36	MT018280	KT389852

Supplementary Table S4.1 continued. *Neurospora* sampling informed by Nygren et al. (2011).

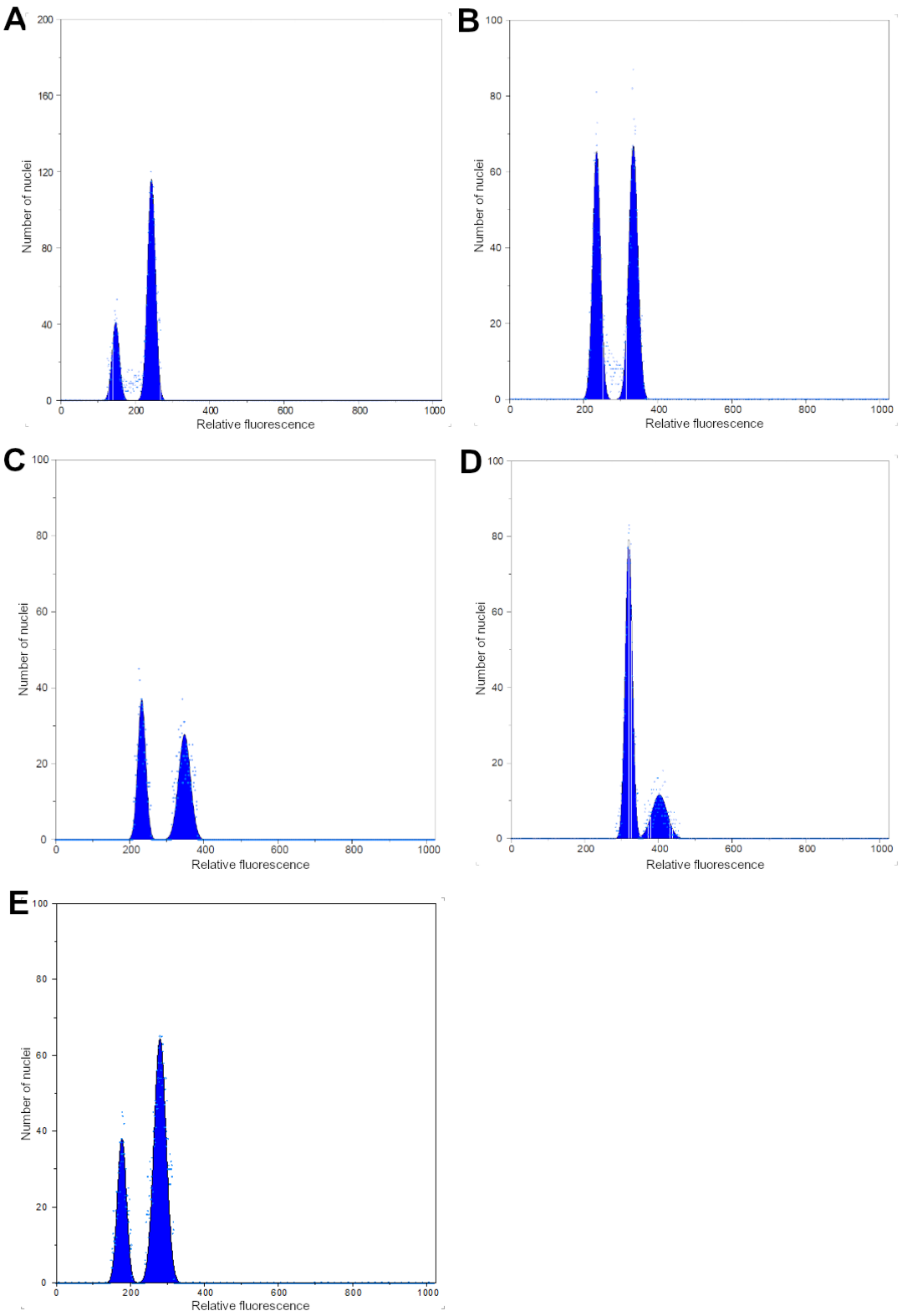
	Name	Voucher	Bml	LSU	mak-2	nik-1	PKC	EF1α
***Neurospora* (Figure 4.3I)**	*Neurospora africana*	FMR 7370	FR774319	FR774244		FR774462	FR774484	FR774369
	Neu. brevispora	FGSC 7795	FR774295	FR774245	FR774394	FR774438	FR774485	FR774345
	Neu. calospora	FGSC 958	FR774296	FR774246	FR774395	FR774439	FR774486	FR774346
	Neu. cerealis	FGSC 959	FR774297	FR774247	FR774396	FR774440	FR774487	FR774347
	Neu. crassa	FGSC 8858	FR774322	FR774250	FR774419	FR774464	FR774490	FR774371
	Neu. crassa	FGSC 8771	FR774321	FR774249	FR774418	FR774463	FR774489	FR774370
	Neu. crassa	FGSC 959	FR774320	FR774248	FR774392	FR774436	FR774488	FR774343
	Neu. dictyophora	FMR 7511	FR774298	FR774251	FR774397	FR774441	FR774491	FR774348
	Neu. discreta	FGSC 8780	FR774332	FR774252	FR774426	FR774474	FR774492	FR774381
	Neu. dodgei	FGSC 1692	FR774323	FR774253		FR774465	FR774493	FR774372
	Neu. endodonta	IMI 148369^{T}	FR774299	FR774254	FR774398	FR774442	FR774494	FR774349
	Neu. galapagosensis	FGSC 1739	FR774324	FR774255		FR774466	FR774495	FR774373
	Neu. hapsidophora	CBS 408.82^{T}	FR774300	FR774256	FR774399	FR774443	FR774496	FR774350
	Neu. hispaniola	FGSC 8817	FR774329	FR774257	FR774423	FR774471	FR774497	FR774378
	Neu. indica	FGSC 7793	FR774301	FR774258	FR774400	FR774444	FR774498	FR774351
	Neu. intermedia	FGSC 8844	FR774326	FR774260	FR774421	FR774468	FR774500	FR774375
	Neu. intermedia	FGSC 8901	FR774325	FR774259	FR774420	FR774467	FR774499	FR774374
	Neu. kobi	CBS 560.72^{T}	FR774302	FR774261	FR774401	FR774445	FR774501	FR774352
	Neu. lineolata	CBS 502.70	FR774327	FR774262		FR774469	FR774502	FR774376
	Neu. metzenbergii	FGSC 8847	FR774330	FR774263	FR774424	FR774472	FR774503	FR774379
	Neu. minuta	FMR 7512	FR774303	FR774264	FR774402	FR774446	FR774504	FR774353
	Neu. nigeriensis	FMR 5963	FR774304	FR774265	FR774403	FR774447	FR774505	FR774354
	Neu. novoguineensis	FMR 7269	FR774305	FR774266	FR774404	FR774448	FR774506	FR774355
	Neu. pannonica	FGSC 7221	FR774328	FR774267	FR774422	FR774470	FR774507	FR774377
	Neu. perkinsii	FGSC 8838	FR774331	FR774268	FR774425	FR774473	FR774508	FR774380
	Neu. pseudoreticulata	CBS 556.72	FR774306	FR774269	FR774405	FR774449	FR774509	FR774356
	Neu. reticulata	IMI 080035^{T}	FR774307	FR774270	FR774406	FR774450	FR774510	FR774357
	Neu. reticulospora	FGSC 6537	FR774308	FR774271	FR774407	FR774451	FR774511	FR774358
	Neu. retispora	FMR 7510	FR774309	FR774272	FR774408	FR774452	FR774512	FR774359
	Neu. retispora	FMR 7276		AJ579677				
	Neu. retispora	FMR 5513		AJ579544				
	Neu. retispora	CBS 868.68		MH878403				

Supplementary Table S4.1 continued. *Neurospora* sampling informed by Nygren et al. (2011).

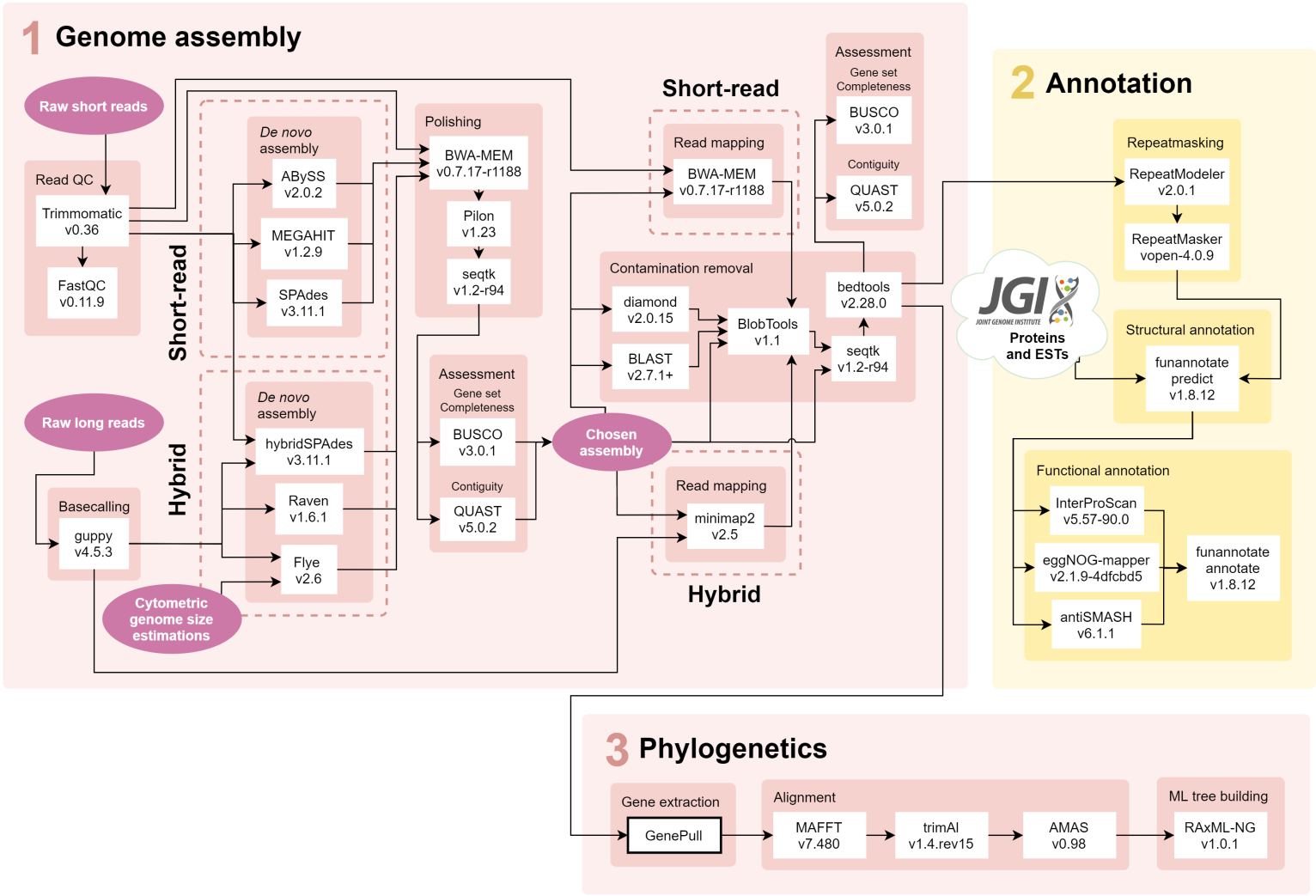
	Name	Voucher	Bml	LSU	mak-2	nik-1	PKC	EF1α
***Neurospora* (Figure 4.3I)**	*Neu. retispora*	CBS 656.70		MH871676				
	Neu. retispora	CBS 275.50^{T}		MH868127				
	Neu. saitoi	CBS 435.74^{T}	FR774311.2	FR774273	FR774410	FR774454	FR774513	FR774361
	Neu. santi-florii	FGSC 8331	FR774310	FR774274	FR774409	FR774453	FR774514	FR774360
	Neu. sitophila	FGSC 8770	FR774333	FR774275	FR774427	FR774475	FR774515	FR774382
	Neu. sp.	FGSC 8243	FR774315	FR774279	FR774414	FR774458	FR774519	FR774365
	Neu. sp.	FGSC 8240	FR774314	FR774278	FR774413	FR774457	FR774518	FR774364
	Neu. sp.	FGSC 8238	FR774313	FR774277	FR774412	FR774456	FR774517	FR774363
	Neu. sp.	FGSC 6877	FR774312	FR774276	FR774411	FR774455	FR774516	FR774362
	Neu. stellata	IFO 30242^{T}	FR774316	FR774280	FR774415	FR774459	FR774520	FR774366
	Neu. sublineolata	IMI 22388^{T}	FR774334	FR774281	FR774428	FR774476	FR774521	FR774383
	Neu. terricola	CBS 298.63^{T}	FR774335.2	FR774282	FR774429	FR774477	FR774522	FR774384
	Neu. tetrasperma	FMR 5545	FR774336	FR774283	FR774430	FR774478	FR774523	FR774385
	Neu. tetraspora	FGSC 7033	FR774317	FR774284	FR774416	FR774460	FR774524	FR774367
	Neu. udagawae	CBS 309.91^{T}	FR774318	FR774285	FR774417	FR774461	FR774525	FR774368
	Neu. uniporata	FMR 7283	FR774337	FR774286	FR774431	FR774479	FR774526	FR774386
	Pseudoneurospora amorphoporcata	CBS 626.80^{T}	FR774294	FR774287	FR774393	FR774437	FR774527	FR774344
	Sordaria brevicollis	FGSC 1904	FR774338	FR774288	FR774432	FR774480	FR774528	FR774387
	S. fimicola	FGSC 2918	FR774339	FR774289			FR774529	FR774388
	S. macrospora	FGSC 4818	FR774340	FR774290	FR774433	FR774481	FR774530	FR774389
	S. sclerogenia	FGSC 2741	FR774341	FR774291	FR774434	FR774482	FR774531	FR774390
	S. tomento-alba	CBS 260.78	FR774342.2	FR774292	FR774435	FR774483	FR774532	FR774391

Supplementary Table S4.2: Assembly statistics from all assembly tools for the 15 endophyte strains.

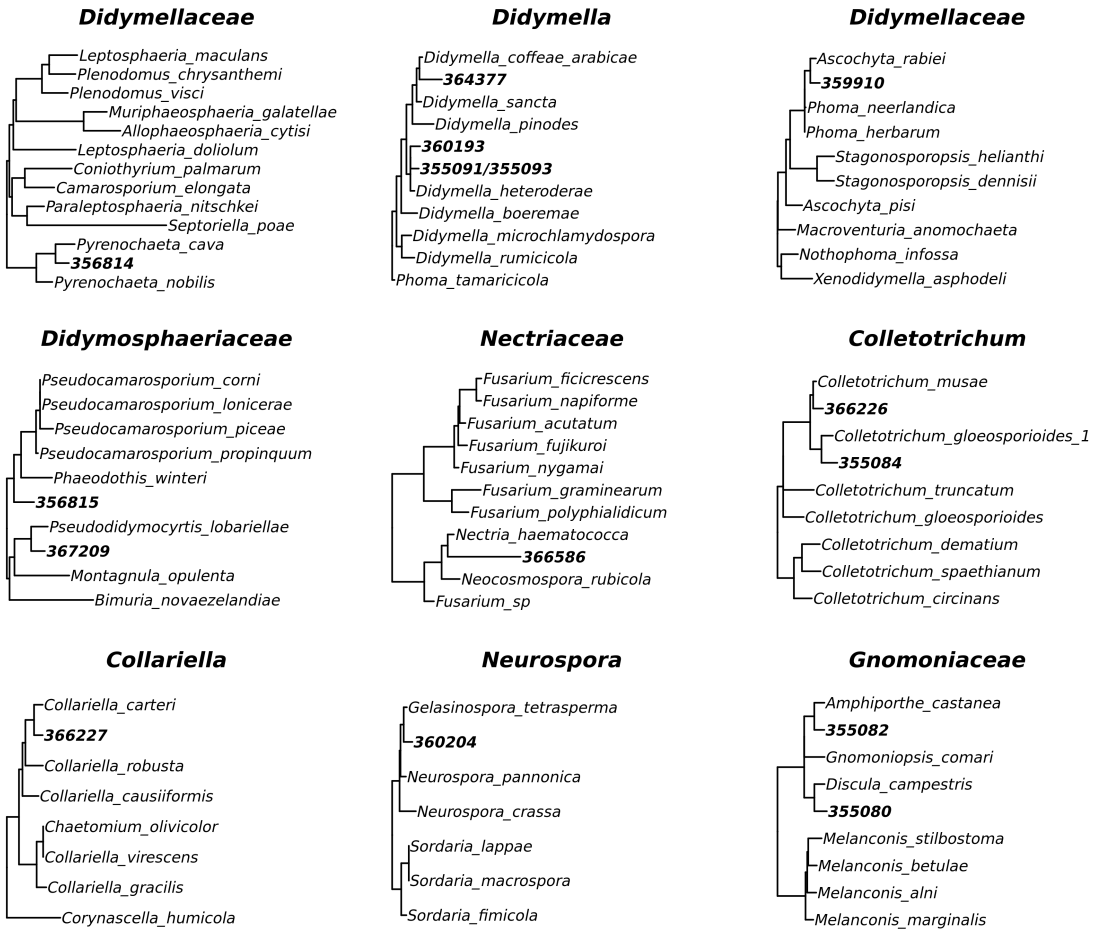
			QUAST						BUSCO
			# contigs ≥500bp	Largest contig (bp)	Total length (bp)	GC (%)	N50	L50	Single-copy BUSCOs (Completeness)
Short-read	355080	ABySS k72	1,521	331,203	40,667,570	51.60	60,408	212	1,542 (90.39%)
		MEGAHIT	1,832	202,743	40,947,610	51.59	52,207	247	1,647 (96.54%)
		SPAdes	798	564,097	41,065,971	51.57	126,266	102	1,653 (96.89%)
	355091	ABySS k80	1,184	576,573	34,975,890	53.45	132,824	81	1,682 (98.59%)
		MEGAHIT	1,526	421,407	34,840,787	53.46	103,515	101	1,683 (98.65%)
		SPAdes	610	908,435	35,454,550	53.43	214,263	44	1,688 (98.94%)
	359910	ABySS k64	1,937	202,055	33,882,229	52.73	42,393	244	1,643 (96.31%)
		MEGAHIT	2,408	173,198	34,289,510	52.52	40,722	261	1,651 (96.78%)
		SPAdes	1,357	259,290	34,499,516	52.37	73,383	152	1,667 (97.71%)
	360193	ABySS k88	1,356	476,436	35,335,285	53.37	85,694	121	1,672 (98.01%)
		MEGAHIT	1,719	325,895	35,071,700	53.46	74,746	139	1,681 (98.53%)
		SPAdes	776	641,373	35,521,025	53.41	178,807	56	1,690 (99.06%)
	360204	ABySS k72	4,475	120,442	36,976,442	52.58	16,432	622	1,567 (91.85%)
		MEGAHIT	6,029	107,474	37,710,625	52.61	13,214	809	1,604 (94.02%)
		SPAdes	4,925	166,179	45,121,419	52.54	24,778	433	1,636 (95.90%)
	364337	ABySS k72	1,734	268,349	30,277,320	51.69	50,318	178	1,643 (96.31%)
		MEGAHIT	2,116	237,638	30,535,934	51.48	45,506	201	1,663 (97.48%)
		SPAdes	1,155	382,275	30,512,631	51.44	74,080	124	1,676 (98.24%)
	366226	ABySS k64	2,708	189,189	54,213,268	53.63	39,154	421	1,627 (95.37%)
		MEGAHIT	3,295	193,772	54,922,345	53.62	34,099	481	1,628 (95.43%)
		SPAdes	1,830	305,111	55,288,883	53.54	63,290	275	1,655 (97.01%)
	366586	ABySS k72	1,656	320,053	40,977,177	52.57	50,683	239	1,657 (97.13%)
		MEGAHIT	2,221	205,757	41,829,246	52.43	42,500	299	1,663 (97.48%)
		SPAdes	1,411	470,694	42,358,263	52.22	90,196	139	1,669 (97.83%)
Hybrid	355082	Flye	12	7,084,357	40,197,374	50.70	6,429,383	3	1,668 (97.77%)
		Raven	15	7,080,637	40,228,030	50.67	4,326,196	4	1,667 (97.71%)
		hybridSPAdes	281	1,693,788	39,888,836	51.10	413,748	29	1,647 (96.54%)
	355084	Flye	58	7,342,820	49,508,467	51.90	2,983,733	6	1,643 (96.31%)
		Raven	56	3,110,953	49,524,676	51.84	1,317,902	13	1,664 (97.54%)
		hybridSPAdes	753	840,111	49,421,028	52.47	161,131	88	1,683 (98.65%)
	355093	Flye	86	2,369,202	31,358,738	52.97	1,219,652	10	1,687 (98.89%)
		Raven	27	1,884,042	31,528,740	52.85	1,301,886	10	1,684 (98.71%)
		hybridSPAdes	184	1,552,342	31,829,418	52.87	520,122	18	1,687 (98.89%)
	356814	Flye	89	3,269,191	34,410,298	50.41	1,599,529	8	1,677 (98.30%)
		Raven	24	2,991,912	34,846,001	50.24	1,616,366	9	1,678 (98.36%)
		hybridSPAdes	593	698,129	33,512,421	51.19	148,079	67	1,673 (98.07%)
	356815	Flye	54	5,272,851	38,910,400	51.57	4,473,122	4	1,680 (98.48%)
		Raven	11	5,345,287	39,450,705	51.25	4,705,368	4	1,672 (98.01%)
		hybridSPAdes	362	1,812,647	38,868,017	51.75	485,797	25	1,677 (98.30%)
	366227	Flye	162	3,665,392	30,332,852	55.81	962,134	9	1,604 (94.02%)
		Raven	52	2,828,572	29,586,632	55.79	1,760,284	7	1,499 (87.87%)
		hybridSPAdes	2,530	149,393	29,037,354	55.70	19,002	435	1,522 (89.21%)
	367209	Flye	97	4,017,923	42,713,253	49.79	1,630,038	9	1,656 (97.07%)
		Raven	31	4,380,344	42,809,244	49.69	2,200,773	7	1,675 (98.18%)
		hybridSPAdes	684	1,149,365	42,184,608	50.32	323,849	40	1,689 (99.00%)



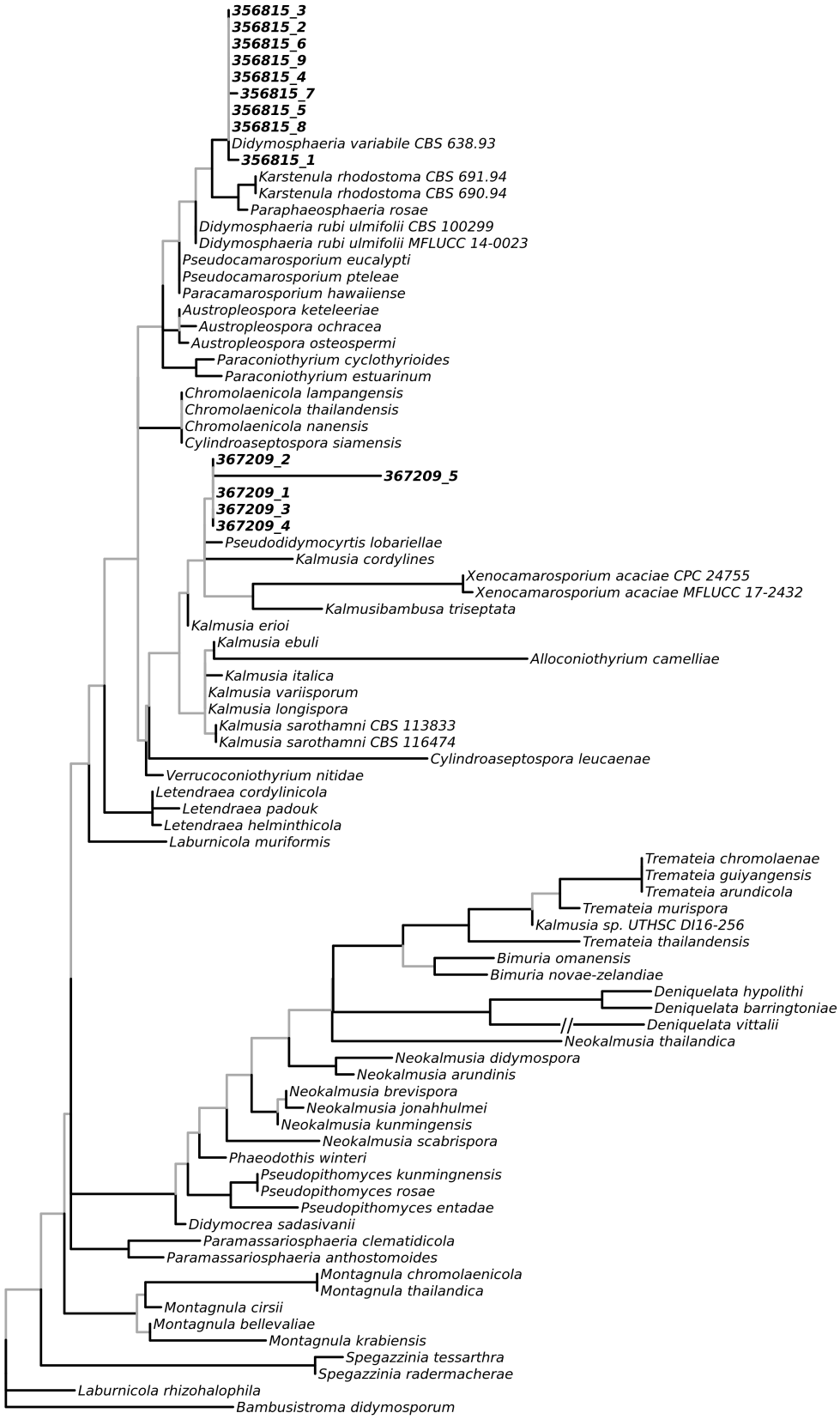
Supplementary Figure S4.1: Flow cytometry histograms showing the relative fluorescence of fungal nuclei from the sample and calibration standard. One representative histogram is shown out of the total three runs made per sample. In all cases the left-hand peak is the sample while the right-hand peak is the standard. **(A)** IMI 355093 **(B)** IMI 356814 **(C)** IMI 359910 **(D)** IMI 360204 **(E)** IMI 364377.



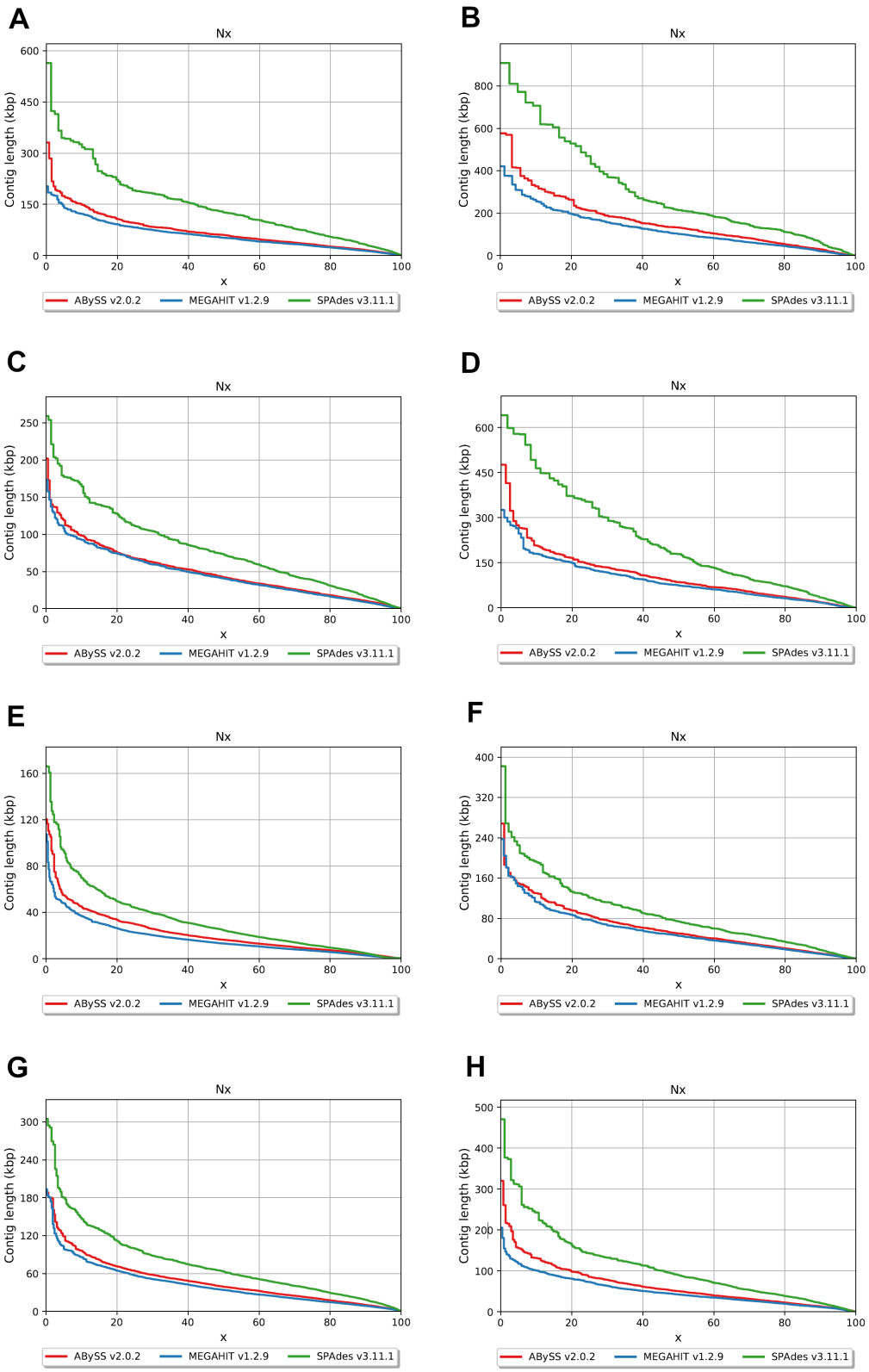
Supplementary Figure S4.2: Schematic summarising the bioinformatics analysis pipeline developed in Chapter 4, available at `https://github.com/Rowena-h/EndophyteGenomes`. Boxes outlined in black indicate custom scripts written for this work.



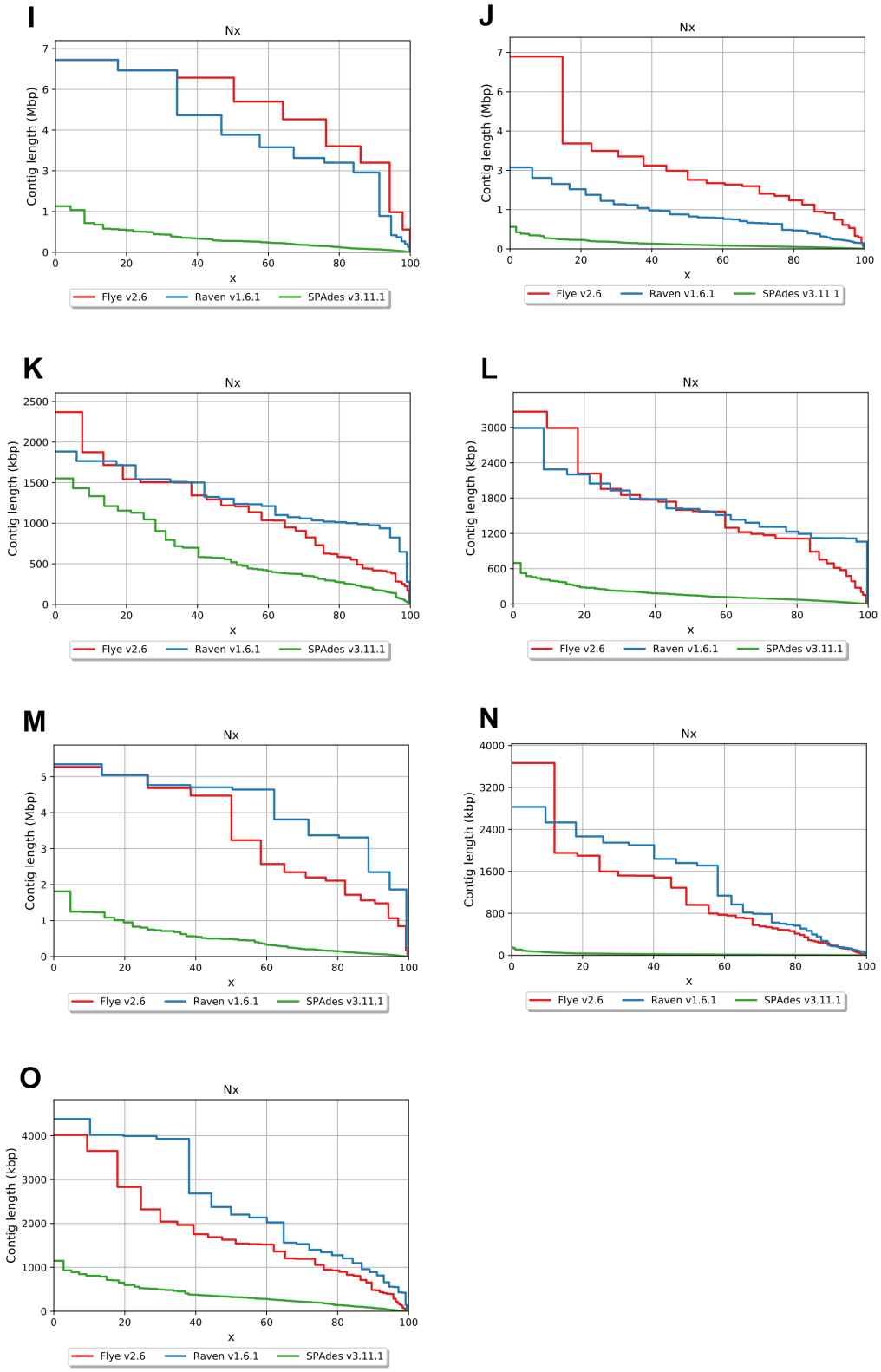
Supplementary Figure S4.3: T-BAS placements for the 15 endophyte strains. For visual clarity, clades containing our strains were extracted from the T-BAS tree and are shown separately. Due to high relatedness, IMI 355091 and IMI 355093 were grouped into a single branch in *Didymella* by T-BAS. *Pyrenochaeta cava* = *Neocucurbitaria cava*; *Gelasinospora tetrasperma* = *Neurospora tetraspora*.



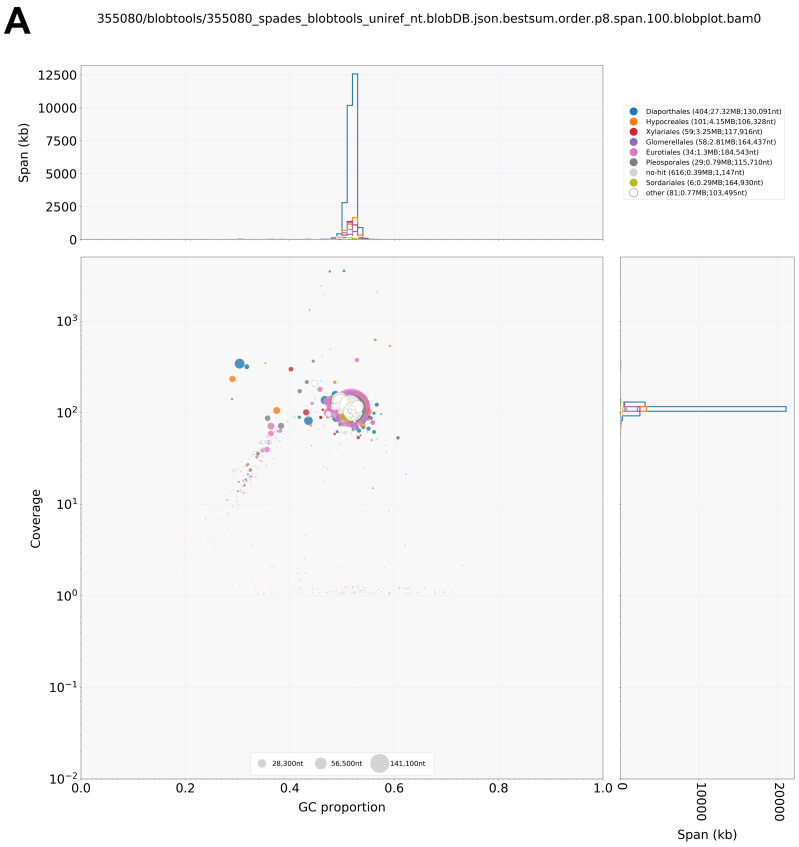
Supplementary Figure S4.4: LSU gene tree of the *Didymosphaeriaceae* produced using RAxML. Branches with significant bootstrap support ($\geq$ 70) are in black, while others are in grey. Multiple copies of LSU from strains IMI 356815 and IMI 367209 are shown in bold.



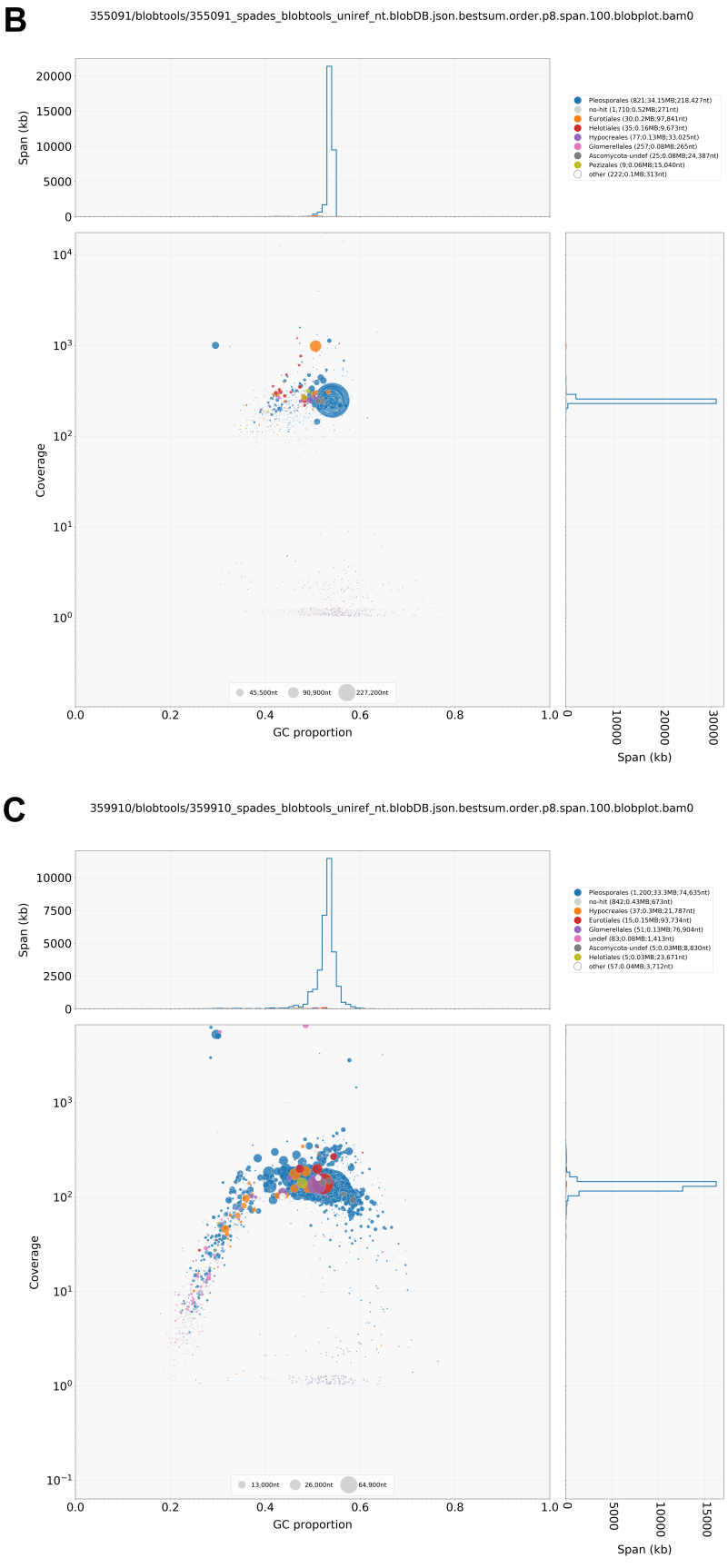
Supplementary Figure S4.5: Nx plots (the smallest contig length at which x% of the assembly is contained in contigs of at least that size) produced by QUAST for each of the strains sequenced in this chapter. Short-read assemblies: **(A)** IMI 355080 **(B)** IMI 355091 **(C)** IMI 359910 **(D)** IMI 360193 **(E)** IMI 360204 **(F)** IMI 364377 **(G)** IMI 366226 **(H)** IMI 366586. ▼



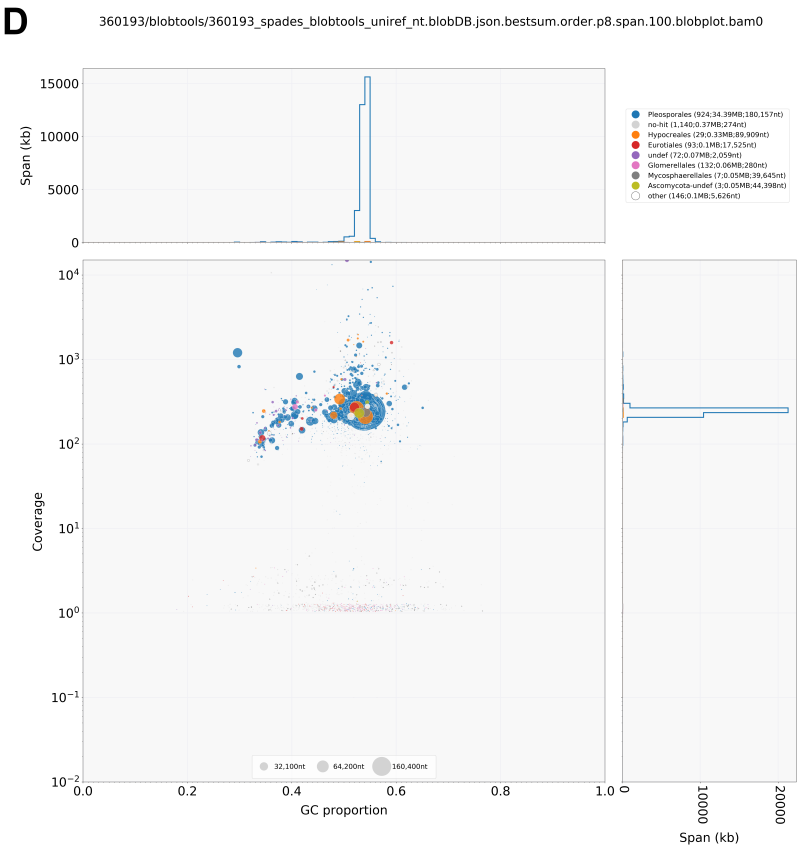
Supplementary Figure S4.5: continued. Hybrid assemblies: **(I)** IMI 355082 **(J)** IMI 355084 **(K)** IMI 355093 **(L)** IMI 356814 **(M)** IMI 356815 **(N)** IMI 366227 **(O)** IMI 367209.

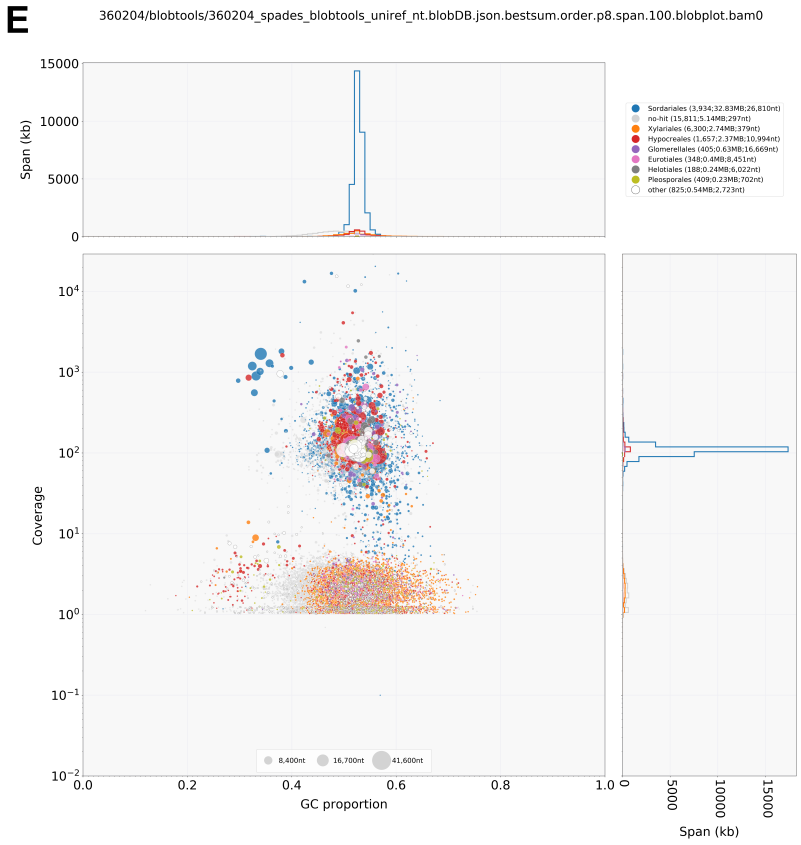


Supplementary Figure S4.6: BlobPlots for the 15 endophyte strains showing the taxonomic classification of reads based on coverage and GC content. Short-read assemblies: **(A)** IMI 355080 **(B)** IMI 355091 **(C)** IMI 359910 **(D)** IMI 360193 **(E)** IMI 360204 **(F)** IMI 364377 **(G)** IMI 366226 **(H)** IMI 366586. Hybrid assemblies: **(I)** IMI 355082 **(J)** IMI 355084 **(K)** IMI 355093 **(L)** IMI 356814 **(M)** IMI 356815 **(N)** IMI 366227 **(O)** IMI 367209. ▼

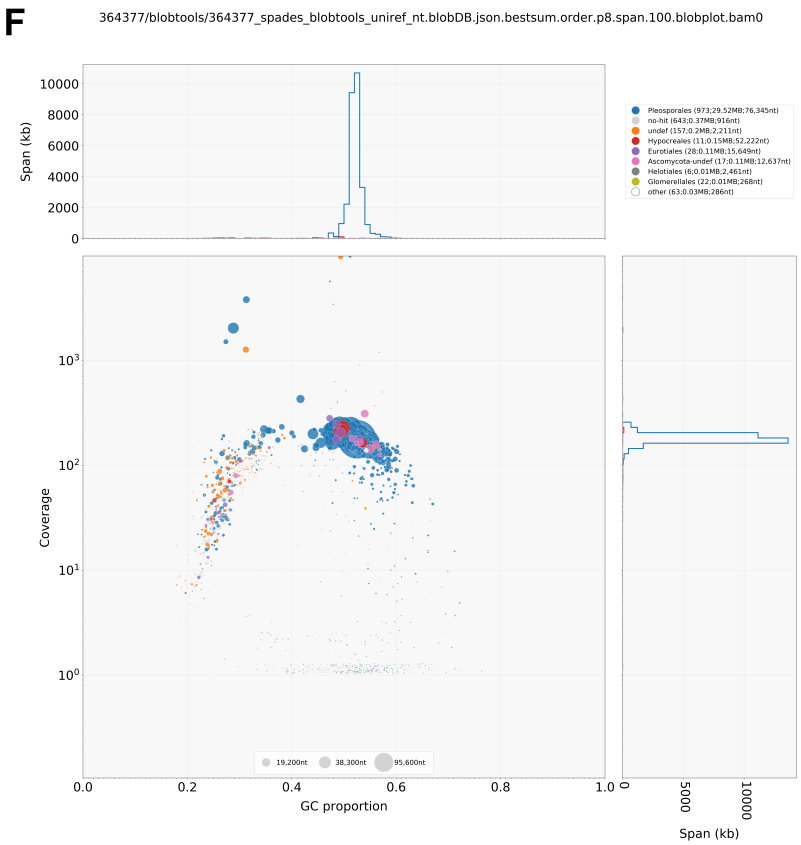


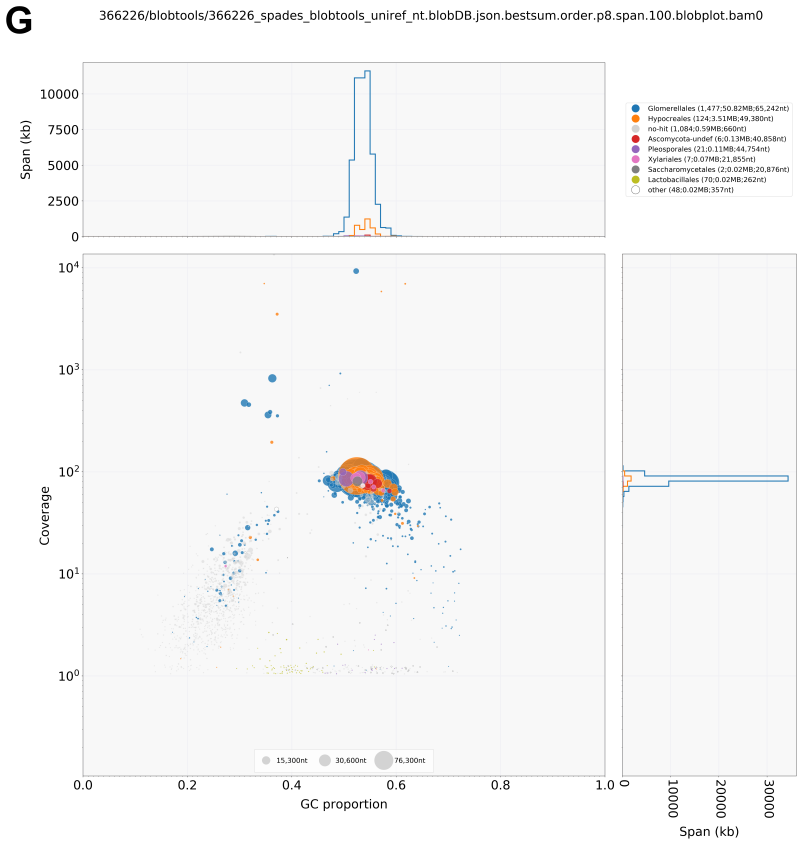
Supplementary Figure S4.6: continued. ▼



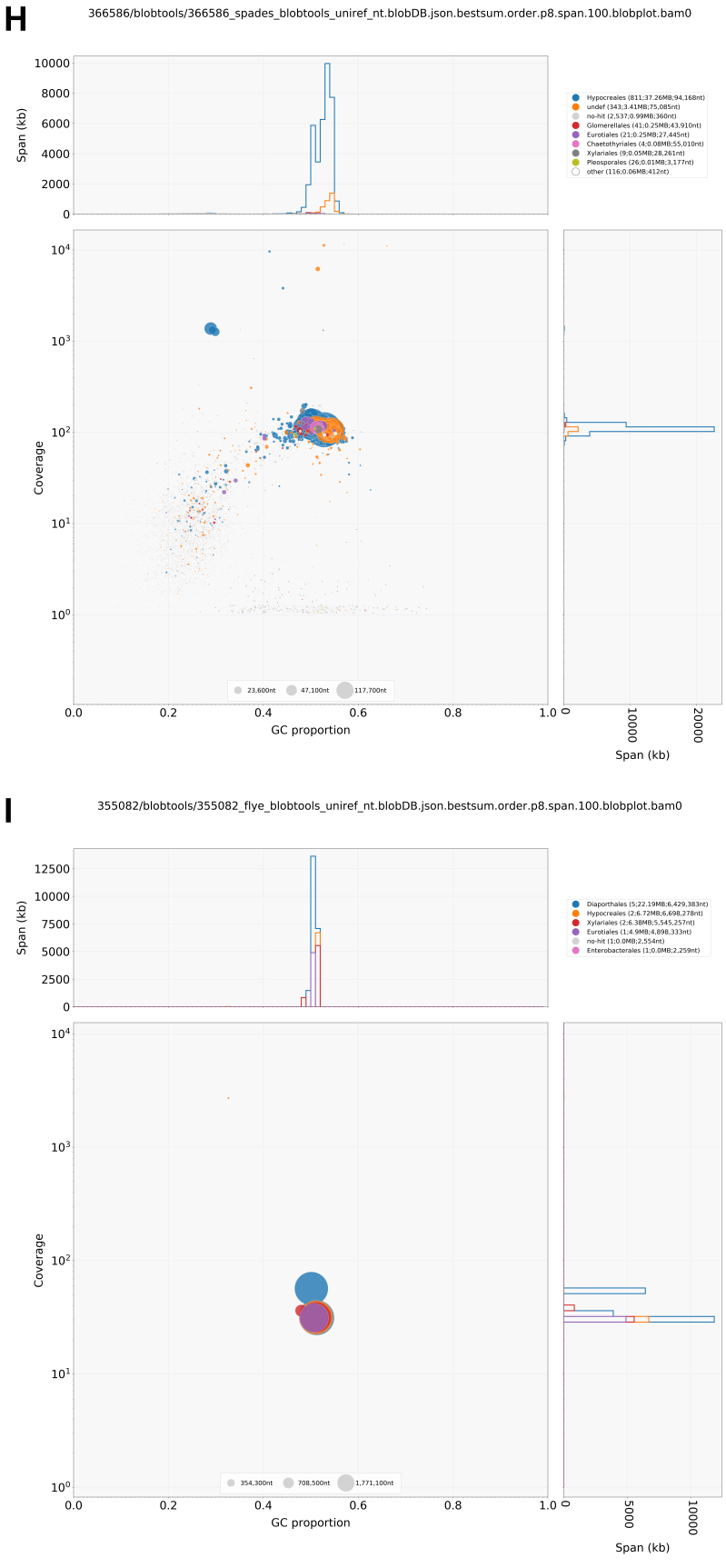


Supplementary Figure S4.6: continued. ▼

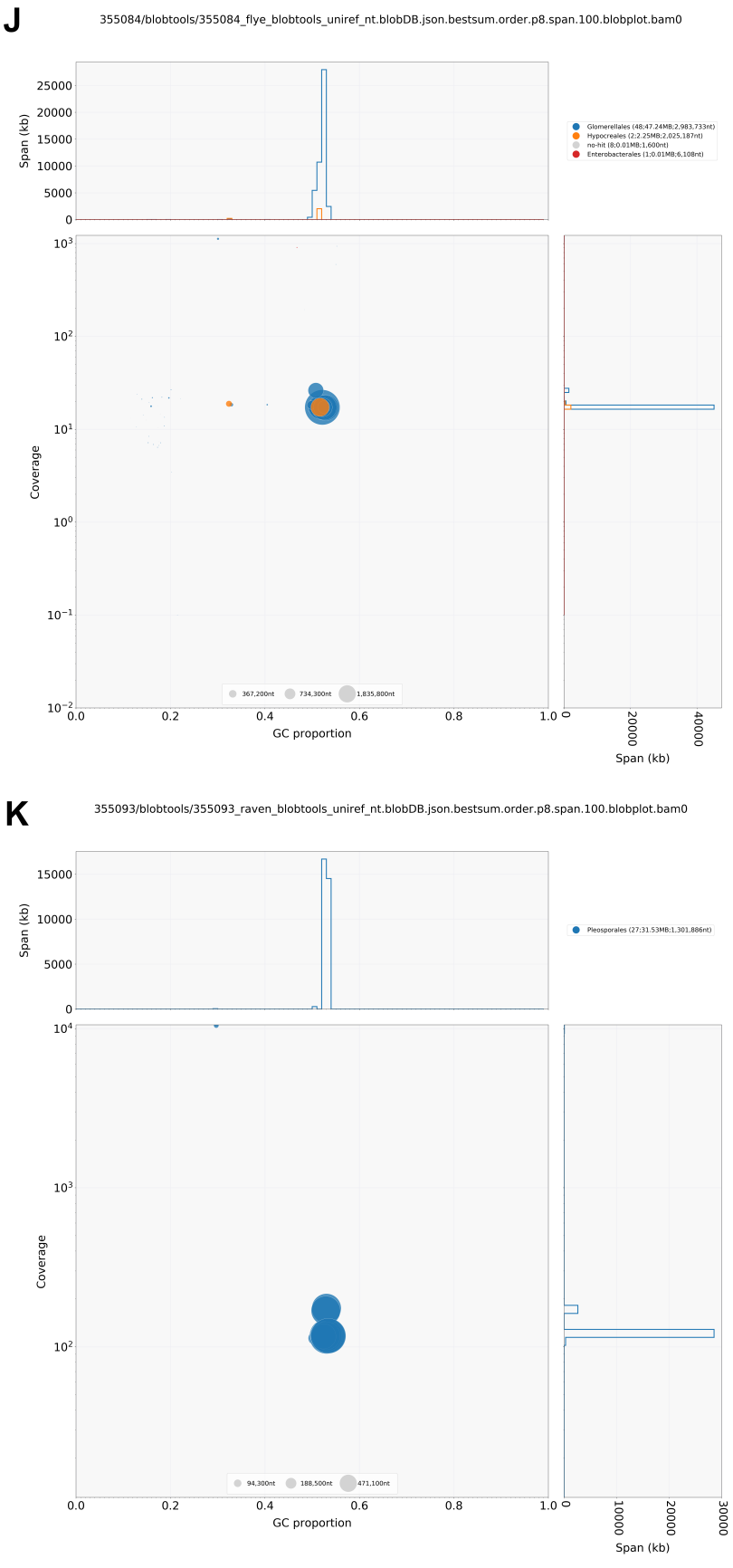




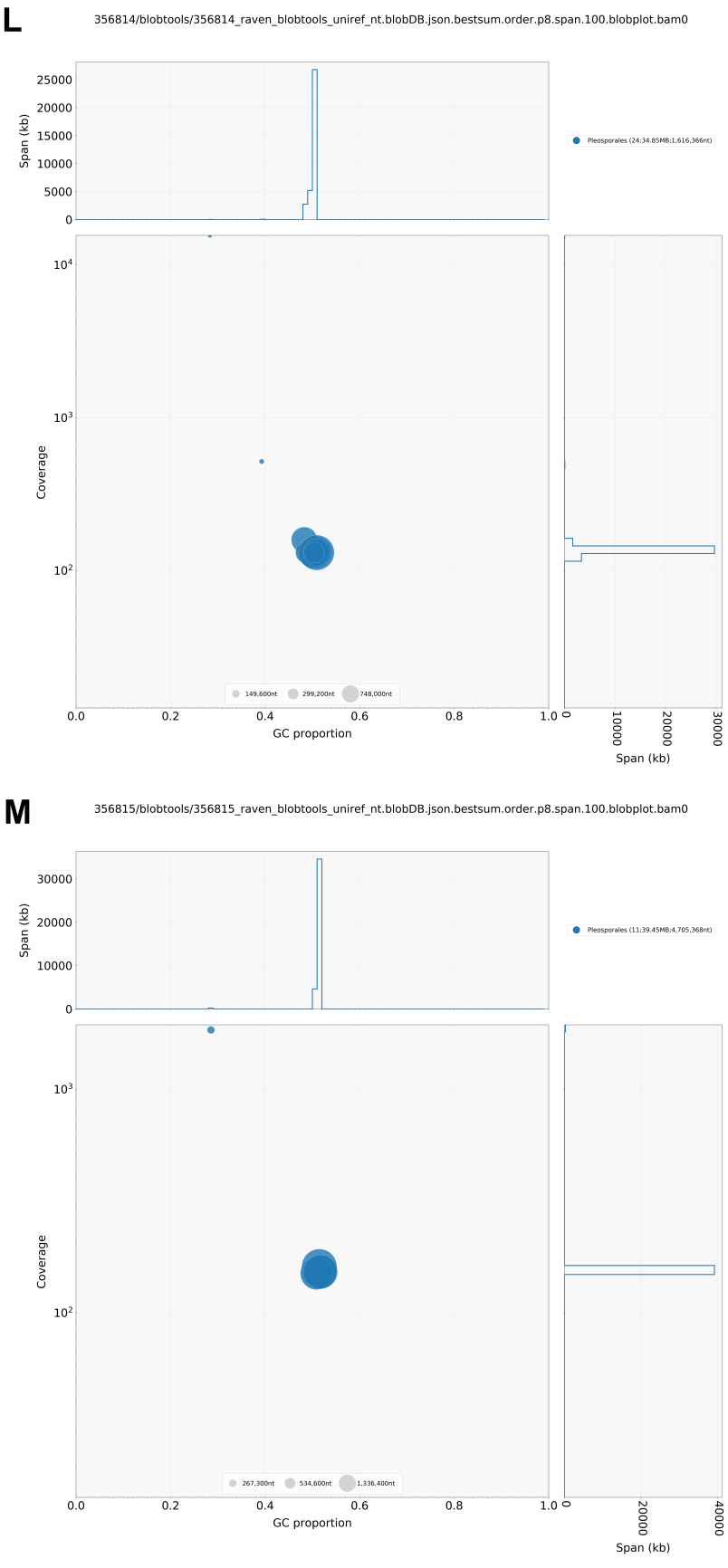
Supplementary Figure S4.6: continued. ▼



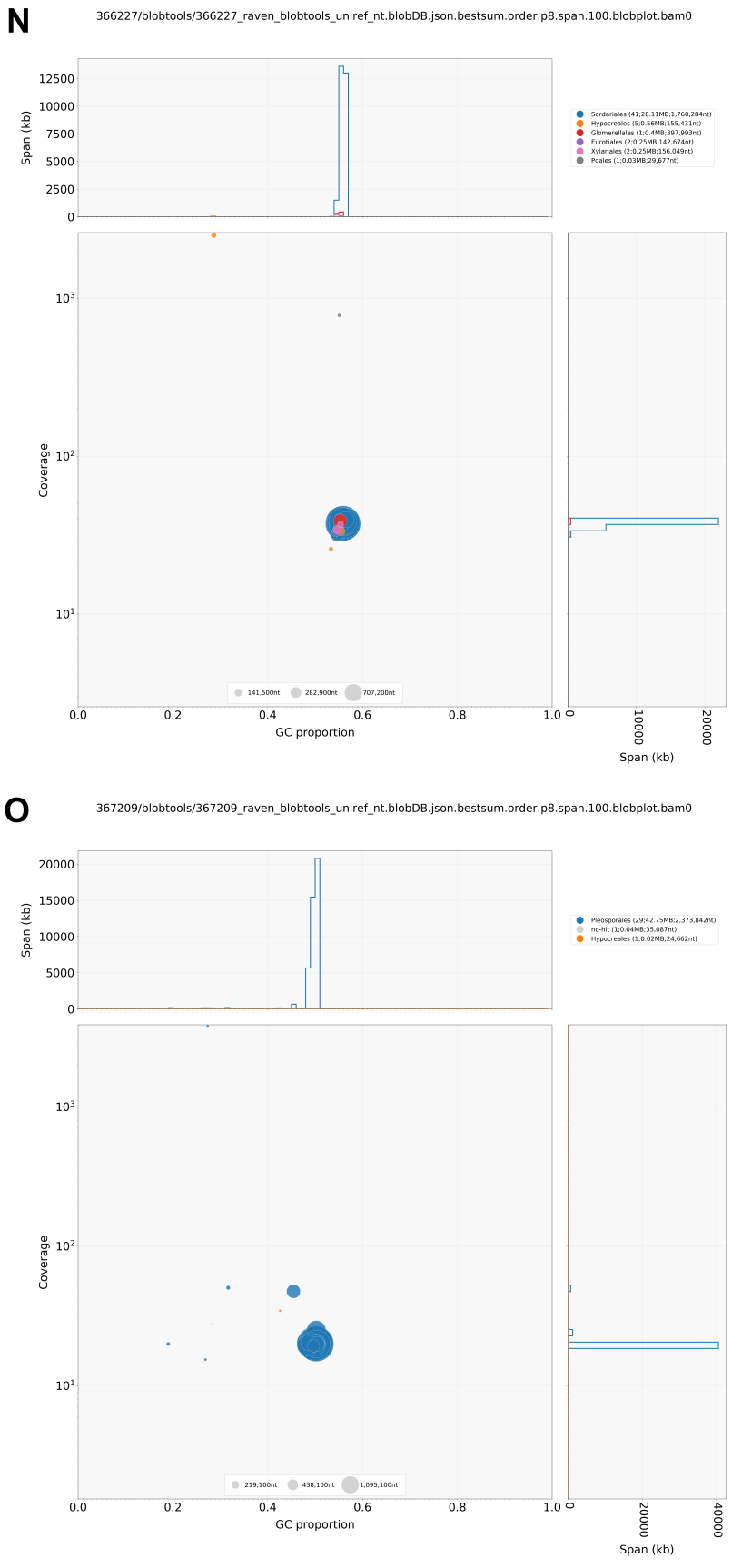
Supplementary Figure S4.6: continued. ▼



Supplementary Figure S4.6: continued. ▼



Supplementary Figure S4.6: continued. ▼



Supplementary Figure S4.6: continued.

Chapter 5

Summary and final remarks

Throughout this thesis I sought to tackle the following objectives:

Objective 1 - Explore the Millennium Seed Bank (MSB) for novel fungal endophyte diversity.

Objective 2 - Determine to what extent we can use genome data to distinguish endophytes and plant pathogens, using the genus *Fusarium* as a case study.

Objective 3 - Produce new genomic resources for a broader taxonomic range of fungal endophytes by capitalising on culture collections.

In Chapter 2 I addressed **Objective 1** using sequence data of fungal endophytes from wild banana seeds stored in the MSB. Almost 200 species hypotheses (i.e., 99% identity operational taxonomic units (OTUs)) were recovered from just one host genus, and a new species has since been formally described (see Appendix A.2). Our results provide a strong rationale for scaling up fungal research in the MSB. Firstly, there is exciting potential for fungal species discovery: even with a highly conservative assumption that there is one new fungal species to be found from each of the >5,800 plant genera in the MSB (Liu, Breman et al., 2018), that would mean thousands of new species of fungi are hidden in the collections. Secondly, as we have found the MSB to effectively be a fungal culture collection in the making, there is the opportunity to screen for strains which produce useful chemical compounds, such as new drug leads. Despite fungi famously being the source of already widely used antibiotic, antifungal and immunosuppressant drugs (Aly, Debbab and Proksch, 2011; Prescott et al., 2018), there is a strong taxonomic bias in which fungi have been investigated for bioactive compounds – only ∼800 of the more than 10,000 fungal genera have been reported in the literature as having strains known to produce certain compounds (T. Prescott, unpublished data). The MSB collections therefore represent a resource for both novel species and novel compound discovery. Finally, even though metadata was not recorded by the original collectors with the seed microbiome in mind, we revealed that habitat and seed viability/germination correlated with endophyte community structure. This has ramifications for the banking process itself, and more research is needed on whether it sufficiently accounts for associated microbes and possible implications for seed health.

There are many practical considerations that would need to be made when extending the work in Chapter 2 more widely across the MSB, and indeed to other seed collections. Large scale culturing is extremely labour intensive and requires dedicated facilities and resources. Large scale direct sequencing is arguably not as useful as isolating strains in culture, but is more logistically straightforward

and could provide data to both reveal fungal diversity patterns and inform targeted culturing efforts. Scaling up endophyte sequencing across the MSB collections would require a shift from the Sanger and cloning approach used in Chapter 2 to high-throughput metabarcoding. There is also the ability to tag and pool metabarcoding samples (Bohmann et al., 2022; Tedersoo, Bahram, Zinger et al., 2022) to optimise throughput while still linking endophytes to specific accessions or even individual seeds (e.g., Özkurt et al., 2020; Bergmann and Busby, 2021; Fort et al., 2021), although the latter may be less practical in terms of time, effort and cost. It is still difficult to determine species abundance from metabarcoding (Elbrecht and Leese, 2015; Thomas et al., 2016; Lamb et al., 2019; Matesanz et al., 2019; Piñol, Senar and Symondson, 2019; Skelton, Cauvin and Hunter, 2022), but at the very least it would provide presence-absence data. Depending on what research questions were being asked, there would also have to be careful consideration of sampling strategy and how data can be normalised to minimise bias. Seed morphology is highly variable across plants and – while it is important to strive for methodological consistency to enable comparison between host species – protocols for both culturing and directly sequencing endophytes would doubtless need adapting based on e.g., seed size or seed coat thickness. The number of seeds that are tested will also impact inferences on community composition (Oskay et al., 2022) and certain accessions may be too precious to use extensively for destructive sampling.

In terms of interpreting endophyte sequence data, in Chapter 2 we opted to cluster sequences into OTUs to use as proxies for species. OTUs are an imperfect but pragmatic solution to help enable species-level inferences from molecular barcode data, and their use has been common practise in fungal ecology (e.g., Tedersoo, Bahram, Põlme et al., 2014; Tisthammer, Cobian and Amend, 2016; U'Ren, Lutzoni, Miadlikowska, Zimmerman et al., 2019). In recent years, however, there has been the suggestion that OTUs should be superseded by amplicon sequence variants (ASVs) – also sometimes referred to as exact sequence variants or haplotypes – which are unique to the nucleotide level after accounting for noise from sequencing errors (Callahan, McMurdie and Holmes, 2017). The major benefit of ASVs is that they are not context-dependent, while OTUs will vary depending on either the reference data used for clustering or, in the case of *de novo* clustering, the other sequences included in the analysis. This means that, unlike OTUs, ASVs from different studies can generally be reused or compared with each other. The choice of ASVs versus OTUs can affect ecological inferences (e.g., Joos et al., 2020; Chiarello et al., 2022), although general patterns in community structure do not seem to be strongly impacted (Glassman and Martiny, 2018).

ASVs are also not without their own limitations. Estensmo et al. (2021) argue that ASVs require an additional clustering step to achieve species-level resolution due to intraspecific variability of the internal transcribed spacer (ITS) in fungi. Even at the strain level, if the selected genetic marker has more than one copy in the genome – which is indeed the case for ITS (Lofgren et al., 2019) – in practise one strain could be split across multiple ASVs (Schloss, 2021). In the most comprehensive study of its kind, empirical clustering of more than 24,000 ITS sequences by Vu et al. (2019) showed that 99.6% was the optimal identity threshold for fungal species, and so ASVs would not be an appropriate proxy if working at the species-level. To work towards a single, overarching ITS sequence dataset for endophytes in the MSB it would probably be sensible to process sequencing data as ASVs in the first instance to ensure consistency – the raw sequencing data produced in Chapter 2 would therefore need to be reanalysed in the same manner in order to be collated into a larger dataset. To make species-level inferences or to interrogate the data for specific research questions, ASVs could then be clustered into OTUs as appropriate. As addressed in Chapter 2, there is also the issue that

ITS is not variable enough to distinguish species in some lineages despite being the 'universal' fungal barcode, and interpretation of the data would have to take this into account.

Having successfully isolated endophyte strains from MSB seeds, in Chapter 3 I then focused on a subset of taxa belonging to the genus *Fusarium* in order to perform comparisons of gene repertoires between endophytic and plant pathogenic strains (**Objective 2**). Selecting *Fusarium* as a case-study group was pertinent because: 1) it's a comparatively well studied ascomycete genus, meaning there was genomic data for a relatively large number of strains that could be included in the analyses; 2) the genus is well known for both endophytes and plant pathogens, as well as other diverse lifestyles; and (3) *Fusarium* taxa are implicated in environmental, agricultural and clinical issues worldwide. By sequencing, assembling and annotating five new *Fusarium* endophyte strains, which also meant producing the first genomic data for the newly described species *Fusarium chuoi*, we almost doubled the number of structurally annotated endophytic *Fusarium* assemblies available on NCBI at the time (mid 2020). We then produced a time-calibrated species tree of *Fusarium sensu stricto* using the most genes to date, which resulted in a phylogeny backbone that was robust across different tools and methods and will therefore provide a valuable reference for future evolutionary studies on the genus. While Chapter 3 did not focus on the taxonomic debates surrounding the *Fusarium* generic concept, we also believe our results showing patterns of selection and codon optimisation provided evidence which supports the more conservative delimitation of the genus (*sensu* Crous, Lombard et al., 2021).

We found that gene repertoires could not distinguish endophytes and phytopathogens in *Fusarium*, suggesting that taxa generally retain the genetic machinery for both lifestyles. But what then explains the observed variation in pathogenicity amongst strains? As discussed in Chapter 3.5, the key to understanding plant-associated lifestyles in *Fusarium* may lie in gene expression profiles (e.g., Niehaus et al., 2016; Guo et al., 2021; Martínez-Soto et al., 2022), and future research examining patterns of expression in endophytic interactions across the whole genus could identify if there is a core gene set that is up/down-regulated for endophytism. There are countless factors which can contribute to differential gene expression in plant–fungal interactions and would therefore be worth investigating in *Fusarium*, such as host genotype (Sánchez-Vallet et al., 2018; Mateus et al., 2019; Porto et al., 2019); environmental stress or nutrient deprivation (Palma-Guerrero et al., 2016; Fouché et al., 2020); or gene silencing via RNA interference (Zhang, Zhao et al., 2016; Šečić et al., 2021). Copy number variation is another source of differential gene expression (Steenwyk and Rokas, 2018; Shao et al., 2019) and copy number variants have been found to be over-represented amongst genes encoding candidate secreted effector proteins (CSEPs) and carbohydrate-active enzymes (CAZymes) in the ascomycete phytopathogen *Rhynchosporium commune* (Stalder et al., 2022), indicating that this may be a key pathogenicity mechanism. In *Fusarium* we found that pathogens had numerous genes that were outliers in terms of copy number compared to other lifestyles, and so this is certainly a topic that warrants further research.

Exploring whether there are convergent patterns in genome arrangement or the presence/absence of accessory chromosomes are also needed in *Fusarium*, however these analyses require highly contiguous assemblies. Unfortunately, due to time constraints we were not able to troubleshoot culturing or DNA extraction protocols to optimise recovery of high molecular weight DNA for the *Fusarium* strains, and so were unable to extract a sufficient volume of DNA for long-read sequencing. As a result, they were not assembled to the level of completeness required to explore these chromosome-scale questions, although for short-read assemblies they were impressively contiguous, with N50 values

that were comparable to some hybrid assemblies (Supplementary Table S3.1; Table 4.2). Aside from our own assemblies, at the time of starting the analyses (mid 2020) we used all *Fusarium* assemblies with annotated gene models that were available in NCBI and only 8 of these were chromosome-level out of the total 56. Since then, many more highly contiguous assemblies have become available – as of November 2022 there are 30 chromosome- or 'complete'- level assemblies with annotated gene models (and a further 34 without annotated gene models). This shows just how much fungal genome resources are advancing in such a short time frame, and will enable us to explore whether chromosome-level phenomena impact lifestyle. The proliferation of genome assembly data provides more opportunities than ever to explore the genetics and evolution of fungal lifestyles, but it is unlikely that genomic data alone will be enough to unravel the nuances of endophytism. It is also important that we maintain transparency about the quality and limitations of genomic resources if we are to distinguish true biological phenomena from methodological artefacts.

Assessing whether our results for *Fusarium* are common to endophytes and pathogens more generally across the *Ascomycota* is currently limited by the amount of genomic data available for different endophyte/phytopathogen-rich lineages. This issue gave rise to **Objective 3** – to produce new genomic resources for a broader taxonomic range of fungal endophytes by capitalising on culture collections – which I tackled in Chapter 4 using accessions from the CABI collections. We produced 15 new endophyte genome assemblies spanning 8 families and 5 orders of the *Ascomycota*, and in this case we were able to extract high molecular weight DNA for just under half the strains, enabling additional long-read sequencing and highly contiguous hybrid assemblies. Incidentally, these are the first genome assemblies for 3 genera and 5 species – a further 4 taxa have not been confirmed to species-level, but based on existing genomic data their to-be-assigned species have also not been genome sequenced before. This emphasised how effective existing fungal collections can be for filling taxonomic as well as lifestyle sampling gaps in fungal genome sequencing and tree of life initiatives. Our phylogenetic results demonstrated just how essential such analyses are for reliable classification of endophytes, with almost all strains brought to genus- if not species-level, substantially improving the value of these genomic resources to the community. In the cases where there was insufficient existing molecular data for the genus to be able to successfully classify the endophyte strain to species-level, targeted multilocus sequencing of more species from the genus will be required to refine the classifications in the future. To our knowledge, some of the strains sequenced and assembled in Chapter 4 also represented novel reports of endophytism for the species.

As stated in Chapter 1.4, it is necessary to compare the performance of multiple assembly tools to optimise assembly quality. The results reported in this thesis reinforce the importance of doing so, as we found that ABySS outperformed SPAdes when assembling the *Fusarium* strains in Chapter 3, but the opposite was true for the various taxa that were short-read assembled in Chapter 4. I do not have a definitive explanation for this, although it is unlikely that the disparity was caused by taxonomy, as one of the strains in Chapter 4 was a closely related fusarioid taxon, *Neocosmospora piperis* IMI 366586. The same short-read sequencing platform was used in both studies, however different sequencing library preparation kits were used: the TruSeq DNA PCR-free Sample Preparation Kit (Chapter 3) fragments DNA mechanically, while the Nextera XT DNA Library Preparation Kit (Chapter 4) fragments DNA enzymatically, which can introduce sequencing bias associated with GC content (Lan et al., 2015; Tyler et al., 2016; Sato et al., 2019). Average short-read sequencing coverage was generally lower for the Chapter 4 strains and, due to the Nextera XT library preparation, sequencing coverage may also have been less uniform, which may explain the overall lower contiguity

of the short-read assemblies compared to the Chapter 3 *Fusarium* assemblies. I can only speculate as to why different assembly tools performed better across the two datasets, but perhaps ABySS and SPAdes are impacted to different degrees by uniformity of read coverage.

In another attempt to maximise genome assembly quality for the strains in Chapter 4, we ventured to produce cytometric genome size measurements, as argued for in Chapter 1.4. Much like the challenges associated with producing enough fungal biomass to extract high molecular weight DNA, we encountered some difficulty in producing adequate subcultures for flow cytometry, and so cytometric genome size measurements were only successful for 5 of the 15 strains. Using the successful measurements we showed that ∼25% of the genome was missing from the assembly in 4 out of 5 cases, meaning that the strains for which we were unable to get cytometric estimates may also be much less complete than gene sets alone would suggest. One of the ∼25% incomplete cases was a hybrid assembly, indicating that long-reads do not necessarily protect against those levels of assembly omission. These results were not altogether surprising having already visualised a similar and even greater disparity for fungal assemblies from Le Cam et al. (2019) (Figure 1.4) – indeed, I expect that the completeness of many publicly available assemblies has been overestimated in the absence of cytometric data. Aside from assembly length, there are other sequence-based genome size estimation methods which exist – such as those inferred from distributions of k-mers (short unique sequences of length 'k') – but much like assembly-based estimates they can still deviate substantially from cytometric estimates and cannot be considered a replacement for cytometry (Pflug et al., 2020). I therefore reiterate the value of cytometric genome size estimates as an additional measure of assembly quality that should be pursued more routinely for fungi, while recognising the practical challenges of obtaining such estimates.

We are evidently still in the infancy of our understanding of the endophytic continuum, and there are some basic conceptual questions concerning endophytism that remain unresolved. Most, if not all, other fungal lifestyles are defined by nutritional strategy while endophytes are defined by circumstance, and so it is perhaps not appropriate that we equate endophytes with mycorrhizas, saprotrophs, phytopathogens, etc., when categorising lifestyles. When we compare endophytes and non-endophytes, some of those endophytic taxa may actually inhabit a nutritional role that is more closely aligned with non-endophytic taxa, and yet in our ignorance of their nutritional strategy we group them together as endophytes, potentially adding noise to any emergent patterns. This is exacerbated by the reductive practise of treating taxa that are actually on a spectrum as a single discrete category – which, to varying extents, is an issue for most fungal lifestyles, not just endophytes – and current comparative approaches are inadequate for capturing the nuances of lifestyle spectrum. From the biocontrol perspective, there is also the philosophical discussion of how to even define individuality amongst hosts and their microbial associates (Skillings, 2016; O'Malley and Parke, 2020). As plant health outcomes can be so dependent on the context of the whole microbiome, it is perhaps too simplistic for us to screen individual endophyte strains in the hope of identifying reliable mutualists. Research is needed on whether it would be more productive to take a community approach (e.g., as for soil microbial communities, see Averill et al., 2022), although admittedly this would be more of a 'black box' strategy. The one certainty is that much more data is needed – genomic, transcriptomic, experimental, etc. – for us to understand the many roles of fungal endophytes. This will be important not only for their potential applications in crop health, but also the general stability of our changing ecosystems.

References

Abarenkov, K., Tedersoo, L. et al. (2010). PlutoF—a Web Based Workbench for Ecological and Taxonomic Research, with an Online Implementation for Fungal ITS Sequences. *Evolutionary Bioinformatics* 2010(6):189–196. DOI: 10.4137/EBO.S6271.

Abarenkov, K., Zirk, A. et al. (2020). *UNITE general FASTA release for Fungi. Version 04.02.2020.* DOI: 10.15156/BIO/786368.

Abbas, M. M., Malluhi, Q. M. and Balakrishnan, P. (2014). Assessment of *de novo* assemblers for draft genomes: a case study with fungal genomes. *BMC Genomics* 15(Suppl 9):S10. DOI: 10.1186/1471-2164-15-S9-S10.

Abdollahzadeh, J. et al. (2010). Phylogeny and morphology of four new species of *Lasiodiplodia* from Iran. *Persoonia* 25:1–10. DOI: 10.3767/003158510X524150.

Achari, S. R. et al. (2021). Comparative transcriptomic analysis of races 1, 2, 5 and 6 of *Fusarium oxysporum* f. sp. *pisi* in a susceptible pea host identifies differential pathogenicity profiles. *BMC Genomics* 22:734. DOI: 10.5897/ajmr12.2234.

Agnan, Y., Séjalon-Delmas, N. and Probst, A. (2013). Comparing early twentieth century and present-day atmospheric pollution in SW France: A story of lichens. *Environmental Pollution* 172:139–148. DOI: 10.1016/j.envpol.2012.09.008.

Ahrendt, S. R. et al. (2018). Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology* 3(12):1417–1428. DOI: 10.1038/s41564-018-0261-0.

Akiva, E. et al. (2014). The Structure-Function Linkage Database. *Nucleic Acids Research* 42:D521–D530. DOI: 10.1093/nar/gkt1130.

Alekseyenko, A. V. (2016). Multivariate Welch t-test on distances. *Bioinformatics* 32(23):3552–3558. DOI: 10.1093/bioinformatics/btw524.

Alidadi, A. et al. (2019). New pathogenic and endophytic fungal species associated with Persian oak in Iran. *European Journal of Plant Pathology* 155:1017–1032. DOI: 10.1007/s10658-019-01830-y.

Allison, S. D. et al. (2018). Temperature acclimation and adaptation of enzyme physiology in *Neurospora discreta*. *Fungal Ecology* 35:78–86. DOI: 10.1016/j.funeco.2018.07.005.

Almagro Armenteros, J. J., Salvatore, M. et al. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* 2(5):e201900429. DOI: 10.26508/lsa.201900429.

Almagro Armenteros, J. J., Tsirigos, K. D. et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* 37:420–423. DOI: 10.1038/s41587-019-0036-z.

Almeida, A. J. et al. (2007). Genome size and ploidy of *Paracoccidioides brasiliensis* reveals a haploid DNA content: Flow cytometry and GP43 sequence analysis. *Fungal Genetics and Biology* 44(1):25–31. DOI: 10.1016/j.fgb.2006.06.003.

Álvarez-Loayza, P. et al. (2011). Light Converts Endosymbiotic Fungus to Pathogen, Influencing Seedling Survival and Niche-Space Filling of a Common Tropical Tree, *Iriartea deltoidea*. *PLoS ONE* 6(1):e16386. DOI: 10.1371/journal.pone.0016386.

Aly, A. H., Debbab, A. and Proksch, P. (2011). Fifty years of drug discovery from fungi. *Fungal Diversity* 50:3–19. DOI: 10.1007/s13225-011-0116-y.

Amaike, S. and Keller, N. P. (2011). *Aspergillus flavus*. *Annual Review of Phytopathology* 49:107–133. DOI: 10.1146/annurev-phyto-072910-095221.

Ampt, E. A. et al. (2016). Larval Preference and Performance of *Amyelois transitella* (Navel Orangeworm, Lepidoptera: Pyralidae) in Relation to the Fungus *Aspergillus flavus*. *Environmental Entomology* 45(1):155–162. DOI: 10.1093/ee/nvv160.

Anderson, M. J. and Walsh, D. C. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs* 83(4):557–574. DOI: 10.1890/12-2010.1.

Andrew, C., Büntgen, U. et al. (2019). Open-source data reveal how collections-based fungal diversity is sensitive to global change. *Applications in Plant Sciences* 7(3):e1227. DOI: 10.1002/aps3.1227.

Andrew, C., Diez, J. et al. (2018). Fungarium specimens: a largely untapped source in global change biology and beyond. *Philosophical Transactions of the Royal Society B* 374:20170392. DOI: 10.1098/rstb.2017.0392.

Andrews, S. (2018). *FastQC: a quality control tool for high throughput sequence data*. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antipov, D. et al. (2016). HybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32(7):1009–1015. DOI: 10.1093/bioinformatics/btv688.

Aoki, T., O'Donnell, K. and Geiser, D. M. (2014). Systematics of key phytopathogenic *Fusarium* species: current status and future challenges. *Journal of General Plant Pathology* 80(3):189–201. DOI: 10.1007/s10327-014-0509-3.

Aoki, T., Smith, J. A. et al. (2019). Three novel Ambrosia *Fusarium* Clade species producing clavate macroconidia known (*F. floridanum* and *F. obliquiseptatum*) or predicted (*F. tuaranense*) to be farmed by *Euwallacea* spp. (Coleoptera: Scolytinae) on woody hosts. *Mycologia* 111(6):919–935. DOI: 10.1080/00275514.2019.1647074.

Armitage, A. D. et al. (2018). Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports* 8:13530. DOI: 10.1038/s41598-018-30335-7.

Arnold, A. E., Maynard, Z. et al. (2000). Are tropical fungal endophytes hyperdiverse? *Ecology Letters* 3(4):267–274. DOI: 10.1046/j.1461-0248.2000.00159.x.

Arnold, A. E., Miadlikowska, J. et al. (2009). A Phylogenetic Estimation of Trophic Transition Networks for Ascomycetous Fungi: Are Lichens Cradles of Symbiotrophic Fungal Diversification? *Systematic Biology* 58(3):283–297. DOI: 10.1093/sysbio/syp001.

Arnold, J. B. (2021). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. URL: https://cran.r-project.org/package=ggthemes.

Athanasopoulou, K. et al. (2022). Third-generation sequencing: The spearhead towards the radical transformation of modern genomics. *Life* 12:30. DOI: 10.3390/life12010030.

Attwood, T. K. et al. (2012). The PRINTS database: a fine-grained protein sequence annotation and analysis resource–its status in 2012. *Database* 2012:bas019. DOI: 10.1093/database/bas019.

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. URL: https://cran.r-project.org/package=gridExtra.

Averill, C. et al. (2022). Defending Earth's terrestrial microbiome. *Nature Microbiology* 7(11):1717–1725. DOI: 10.1038/s41564-022-01228-3.

Aveskamp, M. M., de Gruyter, J. and Crous, P. W. (2008). Biology and recent developments in the systematics of *Phoma*, a complex genus of major quarantine significance. *Fungal Diversity* 31:1–18. ISSN: 15602745.

Aveskamp, M. M., de Gruyter, J., Woudenberg, J. H. et al. (2010). Highlights of the *Didymellaceae*: A polyphasic approach to characterise *Phoma* and related pleosporalean genera. *Studies in Mycology* 65:1–60. DOI: 10.3114/sim.2010.65.01.

Aylward, J. et al. (2017). A plant pathology perspective of fungal genome sequencing. *IMA Fungus* 8(1):1–15. DOI: 10.5598/imafungus.2017.08.01.01.

Badet, T. et al. (2017). Codon optimization underpins generalist parasitism in fungi. *eLife* 6:e22472. DOI: 10.7554/eLife.22472.

Bainard, L. D., Klironomos, J. N. and Hart, M. M. (2010). Differential effect of sample preservation methods on plant and arbuscular mycorrhizal fungal DNA. *Journal of Microbiological Methods* 82:124–130. DOI: 10.1016/j.mimet.2010.05.001.

Baker, S. E. et al. (2015). Draft genome sequence of *Neurospora crassa* strain FGSC 73. *Genome Announcements* 3(2):e00074–15. DOI: 10.1128/genomeA.00074-15.

Baldrian, P. et al. (2022). Fungal communities in soils under global change. *Studies in Mycology* 103:1–24. DOI: 10.3114/sim.2022.103.01.

Bamisile, B. S. et al. (2018). Fungal endophytes: Beyond herbivore management. *Frontiers in Microbiology* 9:544. DOI: 10.3389/fmicb.2018.00544.

Bankevich, A. et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19(5):455–477. DOI: 10.1089/cmb.2012.0021.

Bärenstrauch, M. et al. (2020). Molecular crosstalk between the endophyte *Paraconiothyrium variabile* and the phytopathogen *Fusarium oxysporum* – Modulation of lipoxygenase activity and beauvericin production during the interaction. *Fungal Genetics and Biology* 139:103383. DOI: 10.1016/j.fgb.2020.103383.

Baroncelli, R., Cobo-Díaz, J. F. et al. (2022). Genome evolution and transcriptome plasticity associated with adaptation to monocot and eudicot plants in *Colletotrichum* fungi. *bioRxiv [Preprint]*. DOI: 10.1101/2022.09.22.508453.

Baroncelli, R., Pensec, F. et al. (2021). Complete Genome Sequence of the Plant-Pathogenic Fungus *Colletotrichum lupini*. *Molecular Plant-Microbe Interactions* 34(12):1461–1464. DOI: 10.1094/MPMI-07-21-0173-A.

Bateman, A. et al. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49:D480–D489. DOI: 10.1093/nar/gkaa1100.

Batzoglou, S. et al. (2002). ARACHNE: A Whole-Genome Shotgun Assembler. *Genome Research* 12:177–189. DOI: 10.1101/gr.208902.

Bayman, P. and Otero, J. T. (2006). 'Microbial Endophytes of Orchid Roots'. In: *Microbial Root Endophytes*. Ed. by B. Schulz, C. Boyle and T. Sieber. Berlin, Heidelberg: Springer. Chap. 9:pp. 153–177. ISBN: 3540335269. DOI: 10.1007/3-540-33526-9_9.

Beckerson, W. C. et al. (2019). Cause and Effectors: Whole-Genome Comparisons Reveal Shared but Rapidly Evolving Effector Sets among Host-Specific Plant-Castrating Fungi. *mBio* 10(6):e02391–19. DOI: 10.1128/mBio.02391-19.

Bedward, M., Eppstein, D. and Menzel, P. (2020). *packcircles: Circle Packing*. URL: https://cran.r-project.org/package=packcircles.

Behie, S. W. and Bidochka, M. J. (2014). Ubiquity of insect-derived nitrogen transfer to plants by endophytic insect-pathogenic fungi: an additional branch of the soil nitrogen cycle. *Applied and Environmental Microbiology* 80(5):1553–1560. DOI: 10.1128/AEM.03338-13.

Beimforde, C. et al. (2014). Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Molecular Phylogenetics and Evolution* 78:386–398. DOI: 10.1016/j.ympev.2014.04.024.

Bengtsson, H. (2021). *matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors)*. URL: https://cran.r-project.org/package=matrixStats.

Bengtsson-Palme, J. et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution* 4(10):914–919. DOI: 10.1111/2041-210X.12073.

Benjamini, Y. and Hochberg, Y. (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing Article. *Journal of the Royal Statistical Society* 57(1):289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Bennett, A. E. and Groten, K. (2022). The Costs and Benefits of Plant-Arbuscular Mycorrhizal Fungal Interactions. *Annual Review of Plant Biology* 73:649–672. DOI: 10.1146/annurev-arplant-102820-124504.

Bennett, J. W. (2010). 'An Overview of the Genus *Aspergillus*'. In: *Aspergillus: Molecular Biology and Genomics*. Ed. by M. Machida and K. Gomi. Caister Academic Press. Chap. 1. URL: https://www.caister.com/openaccess/pdf/aspergillus1.pdf.

Bennett, M. D. and Leitch, I. J. (2005a). 'Genome Size Evolution in Plants'. In: *The Evolution of the Genome*. Ed. by T. R. Gregory. Burlington, MA: Elsevier Academic Press. Chap. 2:pp. 89–162. ISBN: 978-0-12-301463-4. DOI: https://doi.org/10.1016/B978-0-12-301463-4.X5000-1.

– (2005b). Nuclear DNA Amounts in Angiosperms: Progress, Problems and Prospects. *Annals of Botany* 95:45–90. DOI: 10.1093/aob/mci003.

– (2011). Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany* 107(3):467–590. DOI: 10.1093/aob/mcq258.

Berger, S. A. and Stamatakis, A. (2011). Aligning short reads to reference alignments and trees. *Bioinformatics* 27(15):2068–2075. DOI: 10.1093/bioinformatics/btr320.

Bergmann, G. E. and Busby, P. E. (2021). The core seed mycobiome of *Pseudotsuga menziesii* var. *menziesii* across provenances of the Pacific Northwest, USA. *Mycologia* 113(6):1169–1180. DOI: 10.1080/00275514.2021.1952830.

Berihuete-Azorín, M. et al. (2018). Punk's not dead. Fungi for tinder at the Neolithic site of La Draga (NE Iberia). *PLoS ONE* 13(4):e0195846. DOI: 10.1371/journal.pone.0195846.

Berlin, K. et al. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature Biotechnology* 33(6):623–630. DOI: 10.1038/nbt.3238.

Bertazzoni, S. et al. (2018). Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi. *Molecular Plant-Microbe Interactions* 31(8):779–788. DOI: 10.1094/mpmi-06-17-0135-fi.

Bidartondo, M. I. et al. (2018). 'Climate change: fungal responses and effects'. In: *State of the World's Fungi 2018*. Ed. by K. J. Willis. Royal Botanic Gardens, Kew. Chap. 9:pp. 62–69. ISBN: 978-1-84246-678-0.

Biedermann, P. H. W. and Vega, F. E. (2020). Ecology and Evolution of Insect–Fungus Mutualisms. *Annual Review of Entomology* 65:431–455. DOI: 10.1146/annurev-ento-011019-024910.

Bilal, L. et al. (2018). Plant growth promoting endophytic fungi *Aspergillus fumigatus* TS1 and *Fusarium proliferatum* BRL1 produce gibberellins and regulates plant endogenous hormones. *Symbiosis* 76(2):117–127. DOI: 10.1007/s13199-018-0545-4.

Blin, K. et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49:W29–W35. DOI: 10.1093/nar/gkab335.

Boddy, L. (2016). 'Fungi, Ecosystems, and Global Change'. In: *The Fungi*. Ed. by S. C. Watkinson, L. Boddy and N. P. Money. 3rd ed. Academic Press. Chap. 11:pp. 361–400. ISBN: 9780123820341. DOI: 10.1016/B978-0-12-382034-1.00011-6.

Bohmann, K. et al. (2022). Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources* 22:1231–1246. DOI: 10.1111/1755-0998.13512.

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.

Bonfim, J. A. et al. (2016). Dark septate endophytic fungi of native plants along an altitudinal gradient in the Brazilian Atlantic forest. *Fungal Ecology* 20:202–210. DOI: 10.1016/j.funeco.2016.01.008.

Booth, C. (1978). Do you believe in genera? *Transactions of the British Mycological Society* 71(1):1–9. DOI: 10.1016/s0007-1536(78)80001-1.

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. DOI: 10.7717/peerj.1660.

Botzman, M. and Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biology* 12:R109. DOI: 10.1186/gb-2011-12-10-r109.

Bourne, E. C. et al. (2014). Large and variable genome size unrelated to serpentine adaptation but supportive of cryptic sexuality in *Cenococcum geophilum*. *Mycorrhiza* 24:13–20. DOI: 10.1007/s00572-013-0501-3.

Bradshaw, M. and Tobin, P. C. (2020). Sequencing herbarium specimens of a common detrimental plant disease (powdery mildew). *Phytopathology* 110:1248–1254. DOI: 10.1094/PHYTO-04-20-0139-PER.

Brameier, M., Krings, A. and MacCallum, R. M. (2007). NucPred—Predicting nuclear localization of proteins. *Bioinformatics* 23(9):1159–1160. DOI: 10.1093/bioinformatics/btm066.

Brown, S. P. et al. (2015). Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecology* 13:221–225. DOI: 10.1016/j.funeco.2014.08.006.

Brozynska, M., Furtado, A. and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnology Journal* 14(4):1070–1085. DOI: 10.1111/pbi.12454.

Brunson, J. (2020). ggalluvial: Layered Grammar for Alluvial Plots. *Journal of Open Source Software* 5(49):2017. DOI: 10.21105/joss.02017.

Buchfink, B., Reuter, K. and Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* 18:366–368. DOI: 10.1038/s41592-021-01101-x.

Burgess, T. I. et al. (2016). Tree invasions and biosecurity: eco-evolutionary dynamics of hitchhiking fungi. *AoB PLANTS* 8:plw076. DOI: 10.1093/aobpla/plw076.

Burgos-Hernández, M. et al. (2014). Seed germination of the wild banana *Musa ornata* (Musaceae). *Seed Science and Technology* 42:16–27. DOI: 10.15258/sst.2014.42.1.02.

Burton, J. N. et al. (2014). Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3* 4:1339–1346. DOI: 10.1534/g3.114.011825.

Butcher, S., King, T. and Zalewski, L. (2017). *Apocrita - High Performance Computing Cluster for Queen Mary University of London*. Tech. rep. Queen Mary University of London. DOI: 10.5281/zenodo.438045.

Butler, J. et al. (2008). ALLPATHS: *De novo* assembly of whole-genome shotgun microreads. *Genome Research* 18:810–820. DOI: 10.1101/gr.7337908.

Callahan, B. J., McMurdie, P. J. and Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal* 11:2639–2643. DOI: 10.1038/ismej.2017.119.

Camacho, C. et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-2105-10-421.

Campitelli, E. (2020). *ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'*. URL: https://cran.r-project.org/package=ggnewscale.

– (2021). *metR: Tools for Easier Analysis of Meteorological Fields*. DOI: 10.5281/zenodo.2593516.

Cantalapiedra, C. P. et al. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38(12):5825–5829. DOI: 10.1093/molbev/msab293.

Cantarel, B. L., Coutinho, P. M. et al. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37:D233–D238. DOI: 10.1093/nar/gkn663.

Cantarel, B. L., Korf, I. et al. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18:188–196. DOI: 10.1101/gr.6743907.

Cao, X. et al. (2022). Endophytic fungus *Pseudodidymocyrtis lobariellae* KL27 promotes taxol biosynthesis and accumulation in *Taxus chinensis*. *BMC Plant Biology* 22:12. DOI: 10.1186/s12870-021-03396-6.

Capella-Gutiérrez, S., Silla-Martínez, J. M. and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973. DOI: 10.1093/bioinformatics/btp348.

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Kauff, F. et al. (2017). T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* 33(8):1160–1168. DOI: 10.1093/bioinformatics/btw808.

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N. et al. (2019). T-BAS Version 2.1: Tree-Based Alignment Selector Toolkit for Evolutionary Placement of DNA Sequences and Viewing Alignments and Specimen Metadata on Curated and Custom Trees. *Microbiology Resource Announcements* 8(29):e00328–19. DOI: 10.1128/MRA.00328-19.

Card, S. et al. (2016). Deciphering endophyte behaviour: the link between endophyte biology and efficacious biological control agents. *FEMS Microbiology Ecology* 92(8). DOI: 10.1093/femsec/fiw114.

Carreel, F. et al. (2002). Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome* 45(4):679–692. DOI: 10.1139/g02-033.

Casarrubia, S. et al. (2018). The hydrophobin-like OmSSP1 may be an effector in the ericoid mycorrhizal symbiosis. *Frontiers in Plant Science* 9:546. DOI: 10.3389/fpls.2018.00546.

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* 17(4):540–552. DOI: 10.1093/oxfordjournals.molbev.a026334.

Center for Plant Conservation (2018). *CPC Best Plant Conservation Practices to Support Species Survival in the Wild*. Tech. rep. Center for Plant Conservation. URL: https://saveplants.org/cpc-best-plant-conservation-practices-to-support-species-survival-in-the-wild/.

Challis, R. et al. (2020). BlobToolKit - interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics* 10(4):1361–1374. DOI: 10.1534/g3.119.400908.

Chamberlain, S. A. and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research* 2:191. DOI: 10.12688/f1000research.2-191.v1.

Chang, Y. et al. (2015). Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biology and Evolution* 7(6):1590–1601. DOI: 10.1093/gbe/evv090.

Charif, D. and Lobry, J. R. (2007). 'SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis'. In: *Structural approaches to sequence evolution: Molecules, networks, populations*. Ed. by U. Bastolla et al. Berlin, Heidelberg: Springer:pp. 207–232. ISBN: 9783540353065. DOI: 10.1007/978-3-540-35306-5_10.

Chen, Q., Hou, L. W. et al. (2017). *Didymellaceae* revisited. *Studies in Mycology* 87:105–159. DOI: 10.1016/j.simyco.2017.06.002.

Chen, Q., Jiang, J. R. et al. (2015). Resolving the *Phoma* enigma. *Studies in Mycology* 82:137–217. DOI: 10.1016/j.simyco.2015.10.003.

Chen, S. F., Fichtner, E. et al. (2013). First report of *Lasiodiplodia citricola* and *Neoscytalidium dimidiatum* causing death of graft union of English walnut in California. *Plant Disease* 97(7):993. DOI: 10.1094/PDIS-10-12-1000-PDN.

Chen, S. L., Lee, W. et al. (2004). Codon usage between genomes is constrained genome-wide mutational processes. *PNAS* 101(10):3480–3485. DOI: 10.1073/pnas.0307827100.

Chen, Y., Zhou, Q. et al. (2014). Genetic diversity and aggressiveness of *Fusarium* spp. isolated from Canola in Alberta, Canada. *Plant Disease* 98:727–738. DOI: 10.1094/PDIS-01-13-0061-RE.

Cheng, S. et al. (2018). 10KP: A phylodiverse genome sequencing plan. *GigaScience* 7(3):1–9. DOI: 10.1093/gigascience/giy013.

Chiarello, M. et al. (2022). Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLoS ONE* 17(2):e0264443. DOI: 10.1371/journal.pone.0264443.

Chin, C. S., Alexander, D. H. et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* 10(6):563–569. DOI: 10.1038/nmeth.2474.

Chin, C. S., Peluso, P. et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* 13(12):1050–1054. DOI: 10.1038/nmeth.4035.

Choi, J. and Kim, S.-h. (2017). A genome Tree of Life for the Fungi kingdom. *PNAS* 114(35):9391–9396. DOI: 10.1073/pnas.1711939114.

Cilliers, A. J., Swart, W. J. and Wingfield, M. J. (1993). A review of *Lasiodiplodia theobromae* with particular reference to its occurrence on coniferous seeds. *South African Forestry Journal* 166:47–52. DOI: 10.1080/00382167.1993.9629398.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18:117–143. DOI: 10.1111/j.1442-9993.1993.tb00438.x.

Cleary, M. et al. (2019). Cryptic risks to forest biosecurity associated with the global movement of commercial seed. *Forests* 10:459. DOI: 10.3390/f10050459.

Cock, P. J. et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423. DOI: 10.1093/bioinformatics/btp163.

Coil, D., Jospin, G. and Darling, A. E. (2015). A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31(4):587–589. DOI: 10.1093/bioinformatics/btu661. arXiv: 1401.5130.

Coleman, J. J. et al. (2009). The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* 5(8):e1000618. DOI: 10.1371/journal.pgen.1000618.

Collinge, D. B., Jensen, B. and Jørgensen, H. J. (2022). Fungal endophytes in plants and their relationship to plant disease. *Current Opinion in Microbiology* 69:102177. DOI: 10.1016/j.mib.2022.102177.

Combès, A. et al. (2012). Chemical Communication between the Endophytic Fungus *Paraconiothyrium variabile* and the Phytopathogen *Fusarium oxysporum*. *PLoS ONE* 7(10):e47313. DOI: 10.1371/journal.pone.0047313.

Criscuolo, A. and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* 10:210. DOI: 10.1186/1471-2148-10-210.

Crockatt, M. E. (2012). Are there edge effects on forest fungi and if so do they matter? *Fungal Biology Reviews* 26:94–101. DOI: 10.1016/j.fbr.2012.08.002.

Croll, D. and McDonald, B. A. (2012). The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *PLoS Pathogens* 8(4):e1002608. DOI: 10.1371/journal.ppat.1002608.

Crous, P. W., Lombard, L. et al. (2021). *Fusarium*: more than a node or a foot-shaped basal cell. *Studies in Mycology* 98:100116. DOI: 10.1016/j.simyco.2021.100116.

Crous, P. W., Summerell, B. A. et al. (2012). Fungal Planet description sheets: 107–127. *Persoonia* 28:138–182. DOI: 10.3767/003158512X652633.

Crous, P. W. and Groenewald, J. Z. (2005). 'Hosts, species and genotypes: opinions versus data'. In: *15th Biennial Conference of the Australasian Plant Pathology Society*. Vol. 34. Geelong, Australia:pp. 463–470. DOI: 10.1071/ap05082.

Crous, P. W., Osieck, E. R. et al. (2021). Fungal Planet description sheets: 1284–1382. *Persoonia* 47:178–374. DOI: 10.3767/persoonia.2021.47.06.

Crous, P. W., Schumacher, R. K. et al. (2019). New and Interesting Fungi. 2. *Fungal Systematics and Evolution* 3:57–134. DOI: 10.3114/fuse.2019.03.06.

Crous, P. W., Groenewald, J. Z. et al. (2016). Global food and fibre security threatened by current inefficiencies in fungal identification. *Philosophical Transactions of the Royal Society B* 371:20160024. DOI: 10.1098/rstb.2016.0024.

Cuomo, C. A. et al. (2007). The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization. *Science* 317(5843):1400–1403. DOI: 10.1126/science.1143708.

Czislowski, E., Zeil-Rolfe, I. and Aitken, E. A. (2021). Effector profiles of endophytic *Fusarium* associated with asymptomatic banana (*Musa* sp.) hosts. *International Journal of Molecular Sciences* 22:2508. DOI: 10.3390/ijms22052508.

D'hondt, L. et al. (2011). Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status. *Molecular Plant Pathology* 12(8):815–828. DOI: 10.1111/j.1364-3703.2011.00711.x.

Dal Forno, M. et al. (2022). DNA Barcoding of Fresh and Historical Collections of Lichen-Forming Basidiomycetes in the Genera Cora and Corella (Agaricales: Hygrophoraceae): A Success Story? *Diversity* 14:284. DOI: 10.3390/d14040284.

Dalio, R. J. et al. (2018). Effector Biology in Focus: A Primer for Computational Prediction and Functional Characterization. *Molecular Plant-Microbe Interactions* 31(1):22–33. DOI: 10.1094/MPMI-07-17-0174-FI.

Damm, U. et al. (2008). Novel *Paraconiothyrium* species on stone fruit trees and other woody hosts. *Persoonia* 20:9–17. DOI: 10.3767/003158508X286842.

Dangl, J. L. and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411:826–833. DOI: 10.1038/35081161.

Darriba, D. et al. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution* 37(1):291–294. DOI: 10.1093/molbev/msz189.

Daru, B. H. et al. (2018). A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. *Philosophical Transactions of the Royal Society B* 374:20170395. DOI: 10.1098/rstb.2017.0395.

Davis, R. H. and Perkins, D. D. (2002). *Neurospora*: a model of model microbes. *Nature Reviews Genetics* 3(5):7–13. DOI: 10.1038/nrg797.

de Castro, E. et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* 34:W362–W365. DOI: 10.1093/nar/gkl124.

de Gruyter, J. et al. (2009). Molecular phylogeny of *Phoma* and allied anamorph genera: Towards a reclassification of the *Phoma* complex. *Mycological Research* 113(4):508–519. DOI: 10.1016/j.mycres.2009.01.002.

de Jonge, R., Bolton, M. D. and Thomma, B. P. (2011). How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Current Opinion in Plant Biology* 14:400–406. DOI: 10.1016/j.pbi.2011.03.005.

de Jonge, R., Esse, H. P. V. et al. (2010). Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. *Science* 329(5994):953–955. DOI: 10.1126/science.1190859.

De Silva, N. I. et al. (2019). Use of endophytes as biocontrol agents. *Fungal Biology Reviews* 33:133–148. DOI: 10.1016/j.fbr.2018.10.001.

Dean, R. et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13(4):414–430. DOI: 10.1111/j.1364-3703.2012.2011.00783.x.

Delaux, P. M. et al. (2013). Evolution of the plant-microbe symbiotic 'toolkit'. *Trends in Plant Science* 18(6):298–304. DOI: 10.1016/j.tplants.2013.01.008.

Delaye, L., García-Guzmán, G. and Heil, M. (2013). Endophytes versus biotrophic and necrotrophic pathogens-are fungal lifestyles evolutionarily stable traits? *Fungal Diversity* 60:125–135. DOI: 10.1007/s13225-013-0240-y.

Delgado-Baquerizo, M. et al. (2020). The proportion of soil-borne pathogens increases with warming at the global scale. *Nature Climate Change* 10(6):550–554. DOI: 10.1038/s41558-020-0759-3.

Dentinger, B. T. et al. (2016). Tales from the crypt: genome mining from fungarium specimens improves resolution of the mushroom tree of life. *Biological Journal of the Linnean Society* 117:11–32. DOI: 10.1111/bij.12553.

Denton, J. F. et al. (2014). Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Computational Biology* 10(12):e1003998. DOI: 10.1371/journal.pcbi.1003998.

Derbyshire, M. C., Harper, L. A. and Lopez-Ruiz, F. J. (2021). Positive Selection of Transcription Factors Is a Prominent Feature of the Evolution of a Plant Pathogenic Genus Originating in the Miocene. *Genome Biology and Evolution* 13(8):evab167. DOI: 10.1093/gbe/evab167.

Desai, A. et al. (2013). Identification of Optimum Sequencing Depth Especially for *De Novo* Genome Assembly of Small Genomes Using Next Generation Sequencing Data. *PLoS ONE* 8(4):e60204. DOI: 10.1371/journal.pone.0060204.

Desjardins, A. E. and Proctor, R. H. (2007). Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology* 119:47–50. DOI: 10.1016/j.ijfoodmicro.2007.07.024.

Desjardins, C. A., Champion, M. D. et al. (2011). Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. *PLoS Genetics* 7(10):e1002345. DOI: 10.1371/journal.pgen.1002345.

Deutekom, E. S. et al. (2019). Measuring the impact of gene prediction on gene loss estimates in Eukaryotes by quantifying falsely inferred absences. *PLoS Computational Biology* 15(8):e1007301. DOI: 10.1371/journal.pcbi.1007301.

Dita, M. et al. (2018). Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. *Frontiers in Plant Science* 9:1468. DOI: 10.3389/fpls.2018.01468.

Doležel, J., Binarová, P. and Lucretti, S. (1989). Analysis of Nuclear DNA Content in Plant Cells by Flow Cytometry. *Biologia Plantarum* 31(2):113–120. DOI: 10.1007/BF02907241.

Doležel, J., Bartoš, J. et al. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry Part A* 51A(2):127–128. DOI: 10.1002/cyto.a.10013.

Doležel, J., Greilhuber, J. and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2(9):2233–2244. DOI: 10.1038/nprot.2007.310.

Doležel, J., Sgorbati, S. and Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* 85(4):625–631. DOI: 10.1111/j.1399-3054.1992.tb04764.x.

Doré, J. et al. (2017). The ectomycorrhizal basidiomycete *Hebeloma cylindrosporum* undergoes early waves of transcriptional reprogramming prior to symbiotic structures differentiation. *Environmental Microbiology* 19(3):1338–1354. DOI: 10.1111/1462-2920.13670.

dos Reis, M., Gunnell, G. F. et al. (2018). Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Systematic Biology* 67(4):594–615. DOI: 10.1093/sysbio/syy001.

dos Reis, M., Savva, R. and Wernisch, L. (2004). Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Research* 32(17):5036–5044. DOI: 10.1093/nar/gkh834.

dos Reis, M. and Yang, Z. (2011). Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times. *Molecular Biology and Evolution* 28(7):2161–2172. DOI: 10.1093/molbev/msr045.

Dragićević, M. B. et al. (2020). ragp: Pipeline for mining of plant hydroxyproline-rich glycoproteins with implementation in R. *Glycobiology* 30(1):19–35. DOI: 10.1093/glycob/cwz072.

Drula, E. et al. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* 50:D571–D577. DOI: 10.1093/nar/gkab1045.

Dudek-Wicher, R. K., Junka, A. and Bartoszewicz, M. (2018). The influence of antibiotics and dietary components on gut microbiota. *Gastroenterology Reviews* 13(2):85–92. DOI: 10.5114/pg.2018.76005.

Dugan, F. M. et al. (2016). New Records of *Lasiodiplodia theobromae* in Seeds of *Tetrapleura tetraptera* from Nigeria and Fruit of *Cocos nucifera* from Mexico. *Journal of Phytopathology* 164:65–68. DOI: 10.1111/jph.12384.

Durán, P. et al. (2018). Microbial Interkingdom Interactions in Roots Promote *Arabidopsis* Survival. *Cell* 175:973–983. DOI: 10.1016/j.cell.2018.10.020.

Eaton, C. J. et al. (2015). A core gene set describes the molecular basis of mutualism and antagonism in *Epichloë* spp. *Molecular Plant-Microbe Interactions* 28(3):218–231. DOI: 10.1094/MPMI-09-14-0293-FI.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792–1797. DOI: 10.1093/nar/gkh340.

– (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 10(10):996–998. DOI: 10.1038/nmeth.2604.

Elbrecht, V. and Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* 10(7):e0130324. DOI: 10.1371/journal.pone.0130324.

Elek, A., Kuzman, M. and Vlahovicek, K. (2021). *coRdon: Codon Usage Analysis and Prediction of Gene Expressivity*. URL: https://github.com/BioinfoHR/coRdon.

Elliott, T. A. and Gregory, T. R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B* 370:20140331. DOI: 10.1098/rstb.2014.0331.

Emms, D. M. and Kelly, S. (2018). STAG: Species Tree Inference from All Genes. *bioRxiv [Preprint]*. DOI: doi.org/10.1101/267914.

Emms, D. M. and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238. DOI: 10.1101/466201.

English, A. C. 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.

Eskalen, A. et al. (2012). First report of a *Fusarium* sp. and its vector tea shot hole borer (*Euwallacea fornicatus*) causing Fusarium dieback on avocado in California. *Plant Disease* 96(7):1070. DOI: 10.1094/PDIS-03-12-0276-PDN.

Estensmo, E. L. F. et al. (2021). The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. *Molecular Ecology Resources* 21:1141–1148. DOI: 10.1111/1755-0998.13329.

Ewald, P. W. (1987). Transmission Modes and Evolution of the Parasitism-Mutualism Continuum. *Annals of the New York Academy of Sciences* 503:295–306. DOI: 10.1111/j.1749-6632.1987.tb40616.x.

Faino, L. et al. (2015). Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *mBio* 6(4):e00936–15. DOI: 10.1128/mBio.00936-15.

FAO (2014). *Genebank Standards for Plant Genetic Resources for Food and Agriculture*. Tech. rep. Rome: Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/3/i3704e/i3704e.pdf.

FAO (2020a). *Banana Market Review 2019*. Tech. rep. Rome: Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/publications/card/es/c/CB0168EN/.

– (2020b). *Banana Market Review Snapshot*. Tech. rep. Rome: Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/3/ca9212en/ca9212en.pdf.

Fernández, M. M., Bezos, D. and Diez, J. J. (2018). Fungi associated with necrotic galls of *Dryocosmus kuriphilus* (Hymenoptera: Cynipidae) in northern Spain. *Silva Fennica* 2(3):9905. DOI: 10.14214/sf.9905.

Field, K. J. and Pressel, S. (2018). Unity in diversity: structural and functional insights into the ancient partnerships between plants and fungi. *New Phytologist* 220:996–1011. DOI: 10.1111/nph.15158.

Fisher, M. C., Gurr, S. J. et al. (2020). Threats Posed by the Fungal Kingdom to Humans, Wildlife, and Agriculture. *mBio* 11(3):e00449–20. DOI: 10.1128/mBio.00449-20.

Fisher, M. C., Henk, D. A. et al. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484(7393):186–194. DOI: 10.1038/nature10947.

Flakus, A. et al. (2019). Biodiversity assessment of ascomycetes inhabiting *Lobariella* lichens in Andean cloud forests led to one new family, three new genera and 13 new species of lichenicolous fungi. *Plant and Fungal Systematics* 64(2):283–344. DOI: 10.2478/pfs-2019-0022.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology* 9:275–296. DOI: 10.1146/annurev.py.09.090171.001423.

Fones, H. N. et al. (2020). Threats to global food security from emerging fungal and oomycete crop pathogens. *Nature Food* 1(6):332–342. DOI: 10.1038/s43016-020-0075-0.

Forest, F. (2009). Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany* 104(5):789–794. DOI: 10.1093/aob/mcp192.

Fort, T. et al. (2021). Maternal effects shape the seed mycobiome in *Quercus petraea*. *New Phytologist* 230:1594–1608. DOI: 10.1111/nph.17153.

Fouché, S. et al. (2020). Stress-Driven Transposable Element De-repression Dynamics and Virulence Evolution in a Fungal Pathogen. *Molecular Biology and Evolution* 37(1):221–239. DOI: 10.1093/molbev/msz216.

Franceschetti, M. et al. (2017). Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. *Microbiology and Molecular Biology Reviews* 81(2):e00066–16. DOI: 10.1128/mmbr.00066-16.

Franco, M. E. et al. (2021). Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. *New Phytologist*. DOI: 10.1111/nph.17873.

Freeman, S. et al. (2013). *Fusarium euwallaceae* sp. nov.-a symbiotic fungus of *Euwallacea* sp., an invasive ambrosia beetle in Israel and California. *Mycologia* 105(6):1595–1606. DOI: 10.3852/13-066.

Funk, V. A. (2018). Collections-based science in the 21st Century. *Journal of Systematics and Evolution* 56(3):175–193. DOI: 10.1111/jse.12315.

Gagolewski, M. and Tartanus, B. (2021). *stringi: Character String Processing Facilities*. URL: https://cran.r-project.org/web/packages/stringi/index.html.

Galardini, M. et al. (2011). CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code for Biology and Medicine* 6:11. DOI: 10.1186/1751-0473-6-11.

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31(22):3718–3720. DOI: 10.1093/bioinformatics/btv428.

Gallery, R. E., Dalling, J. W. and Arnold, A. E. (2007). Diversity, host affinity, and distribution of seed-infecting fungi: A case study with *Cecropia*. *Ecology* 88(3):582–588. DOI: 10.1890/05-1207.

Gallery, R. E., Moore, D. J. and Dalling, J. W. (2010). Interspecific variation in susceptibility to fungal pathogens in seeds of 10 tree species in the neotropical genus *Cecropia*. *Journal of Ecology* 98:147–155. DOI: 10.1111/j.1365-2745.2009.01589.x.

García, D. et al. (2004). A synopsis and re-circumscription of *Neurospora* (syn. *Gelasinospora*) based on ultrastructural and 28S rDNA sequence data. *Mycological Research* 108(10):1119–1142. DOI: 10.1017/S0953756204000218.

Garcia-Hermoso, D. et al. (2019). Diversity of coelomycetous fungi in human infections: A 10-y experience of two European reference centres. *Fungal Biology* 123(4):341–349. DOI: 10.1016/j.funbio.2019.02.001.

Garnica, S. et al. (2022). Environmental stress determines the colonization and impact of an endophytic fungus on invasive knotweed. *Biological Invasions* 24:1785–1795. DOI: 10.1007/s10530-022-02749-y.

Gazis, R. et al. (2016). The genome of *Xylona heveae* provides a window into fungal endophytism. *Fungal Biology* 120:26–42. DOI: 10.1016/j.funbio.2015.10.002.

Gearty, W. (2021). *deeptime: Plotting Tools for Anyone Working in Deep Time*. URL: https://github.com/willgearty/deeptime.

Geib, S. M. et al. (2018). Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. *GigaScience* 7:1–5. DOI: 10.1093/gigascience/giy018.

Geiser, D. M., Al-Hatmi, A. et al. (2021). Phylogenomic analysis of a 55.1 kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* Species Complex. *Phytopathology* 7(111). DOI: 10.1094/phyto-08-20-0330-le.

Geiser, D. M., Jiménez-Gasco, M. D. M. et al. (2004). FUSARIUM-ID v. 1.0: A DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* 110(5-6):473–479. DOI: 10.1023/B:EJPP.0000032386.75915.a0.

Genre, A. et al. (2020). Unique and common traits in mycorrhizal symbioses. *Nature Reviews Microbiology* 18:649–660. DOI: 10.1038/s41579-020-0402-3.

Glass, N. L. et al. (2013). Plant cell wall deconstruction by ascomycete fungi. *Annual Review of Microbiology* 67:477–498. DOI: 10.1146/annurev-micro-092611-150044.

Glassman, S. I. and Martiny, J. B. H. (2018). Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. *mSphere* 3(4):e00148–18. DOI: 10.1128/mSphere.00148-18.

Gnerre, S. et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *PNAS* 108(4):1513–1518. DOI: 10.1073/pnas.1017351108.

Goos, R. D., Cox, E. A. and Stotzky, G. (1961). *Botryodiplodia theobromae* and Its Association with *Musa* Species. *Mycologia* 53(3):262–277. DOI: 10.2307/3756274.

Gough, J. et al. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology* 313:903–919. DOI: 10.1006/jmbi.2001.5080.

Govaerts, R. and Häkkinen, M. (2006). *World Checklist of Musaceae. Facilitated by the Royal Botanic Gardens, Kew*. URL: http://wcsp.science.kew.org/ (visited on 17/12/2020).

Granath, G. et al. (2007). Variation in the abundance of fungal endophytes in fescue grasses along altitudinal and grazing gradients. *Ecography* 30:422–430. DOI: 10.1111/j.2007.0906-7590.05027.x.

Graur, D. and Martin, W. (2004). Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics* 20(2):80–86. DOI: 10.1016/j.tig.2003.12.003.

Graves, S. et al. (2019). *multcompView: Visualizations of Paired Comparisons*. URL: https://cran.r-project.org/package=multcompView.

Grigoriev, I. V. et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* 42:699–704. DOI: 10.1093/nar/gkt1183.

Guo, L. et al. (2021). Metatranscriptomic comparison of endophytic and pathogenic *Fusarium–Arabidopsis* interactions reveals plant transcriptional plasticity. *Molecular Plant-Microbe Interactions* 34(9):1071–1083. DOI: 10.1094/MPMI-03-21-0063-R.

Gupta, S. et al. (2020). A critical review on exploiting the pharmaceutical potential of plant endophytic fungi. *Biotechnology Advances* 39:107462. DOI: 10.1016/j.biotechadv.2019.107462.

Gurevich, A. et al. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075. DOI: 10.1093/bioinformatics/btt086.

Gutierrez, M. et al. (2022). Severe Outbreak of Dry Core Rot in Apple Fruits cv. Fuji Caused by *Kalmusia variispora* During Preharvest in Maule Region, Chile. *Plant Disease* 106:2750. DOI: 10.1094/PDIS-02-16-0237-PDN.

Haas, B. J. et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* 9:R7. DOI: 10.1186/gb-2008-9-1-r7.

Hacquard, S. et al. (2016). Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature Communications* 7:11362. DOI: 10.1038/ncomms11362.

Haelewaters, D. et al. (2020). Draft Genome Sequence of the Globally Distributed Cockroach-Infecting Fungus *Herpomyces periplanetae* Strain D. Haelew. 1187d. *Microbiology Resource Announcements* 9(6):e01458–19. DOI: 10.1128/MRA.01458-19.

Haft, D. H. et al. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Research* 29(1):41–43. DOI: 10.1093/nar/29.1.41.

Hage, H. and Rosso, M. N. (2021). Evolution of Fungal Carbohydrate-Active Enzyme Portfolios and Adaptation to Plant Cell-Wall Polymers. *Journal of Fungi* 7(3):1. DOI: 10.3390/jof7030185.

Hajjar, R. and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* 156:1–13. DOI: 10.1007/s10681-007-9363-0.

Häkkinen, M. and Väre, H. (2008). Typification and check-list of *Musa* L. names (Musaceae) with nomenclatural notes. *Adansonia* 30(1):63–112. ISSN: 12808571.

Hardoim, P. R. et al. (2015). The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiology and Molecular Biology Reviews* 79(3):293–320. DOI: 10.1128/MMBR.00050-14.

Haridas, S. et al. (2020). 101 *Dothideomycetes* genomes: A test case for predicting lifestyles and emergence of pathogens. *Studies in Mycology* 96:141–153. DOI: 10.1016/j.simyco.2020.01.003.

Harper, K. A. et al. (2005). Edge influence on forest structure and composition in fragmented landscapes. *Conservation Biology* 19(3):768–782. DOI: 10.1111/j.1523-1739.2005.00045.x.

Harrison, J. G. and Griffin, E. A. (2020). The diversity and distribution of endophytes across biomes, plant phylogeny and host tissues: how far have we come and where do we go from here? *Environmental Microbiology* 22(6):2107–2123. DOI: 10.1111/1462-2920.14968.

Hashizume, Y., Sahashi, N. and Fukuda, K. (2008). The influence of altitude on endophytic mycobiota in *Quercus acuta* leaves collected in two areas 1000 km apart. *Forest Pathology* 38:218–226. DOI: 10.1111/j.1439-0329.2008.00547.x.

Havenga, M. et al. (2019). Canker and Wood Rot Pathogens Present in Young Apple Trees and Propagation Material in the Western Cape of South Africa. *Plant Disease* 103:3129–3141. DOI: 10.1094/PDIS-04-19-0867-RE.

Hawksworth, D. and Lücking, R. (2017). Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiology Spectrum* 5(4):FUNK–0052–2016. DOI: 10.1128/microbiolspec.FUNK-0052-2016.

He, X. and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164. DOI: 10.1534/genetics.104.037051.

Heberling, J. M. and Burke, D. J. (2019). Utilizing herbarium specimens to quantify historical mycorrhizal communities. *Applications in Plant Sciences* 7(4):e1223. DOI: 10.1002/aps3.1223.

Heckman, D. S. et al. (2001). Molecular evidence for the early colonization of land plants by fungi and plants. *Science* 293(5532):1129–1133. DOI: 10.1126/science.1061457.

Heeger, F. et al. (2019). Combining the 5.8S and ITS2 to improve classification of fungi. *Methods in Ecology and Evolution* 2019:1–10. DOI: 10.1111/2041-210x.13266.

Helmy, M., Awad, M. and Mosa, K. A. (2016). Limited resources of genome sequencing in developing countries: Challenges and solutions. *Applied and Translational Genomics* 9:15–19. DOI: 10.1016/j.atg.2016.03.003.

Hemetsberger, C. et al. (2015). The fungal core effector Pep1 is conserved across smuts of dicots and monocots. *New Phytologist* 206:1116–1126. DOI: 10.1111/nph.13304.

Henry, L., Wickham, H. and Chang, W. (2020). *ggstance: Horizontal 'ggplot2' Components*. URL: https://cran.r-project.org/package=ggstance.

Hershberg, R. and Petrov, D. A. (2008). Selection on Codon Bias. *Annual Review of Genetics* 42:287–299. DOI: 10.1146/annurev.genet.42.110807.091442.

Hervé, M. (2020). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics*. URL: ttps://cran.r-project.org/package=RVAideMemoire.

Higgins, K. L. et al. (2011). Culturing and direct PCR suggest prevalent host generalism among diverse fungal endophytes of tropical forest grasses. *Mycologia* 103(2):247–260. DOI: 10.3852/09-158.

Hill, R., Buggs, R. J. A. et al. (2022). Lifestyle Transitions in Fusarioid Fungi are Frequent and Lack Clear Genomic Signatures. *Molecular Biology and Evolution* 39(4):msac085. DOI: 10.1093/molbev/msac085.

Hill, R., Leitch, I. J. and Gaya, E. (2021). Targeting Ascomycota genomes: what and how big? *Fungal Biology Reviews* 36:52–59. DOI: 10.1016/j.fbr.2021.03.003.

Hill, R., Llewellyn, T. et al. (2021). Seed Banks as Incidental Fungi Banks: Fungal Endophyte Diversity in Stored Seeds of Banana Wild Relatives. *Frontiers in Microbiology* 12:643731. DOI: 10.3389/fmicb.2021.643731.

Hipsley, C. A. and Müller, J. (2014). Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in Genetics* 5:138. DOI: 10.3389/fgene.2014.00138.

Hiruma, K. et al. (2016). Root Endophyte *Colletotrichum tofieldiae* Confers Plant Fitness Benefits that Are Phosphate Status Dependent. *Cell* 165:464–474. DOI: 10.1016/j.cell.2016.02.028.

Ho, S. Y. (2009). An examination of phylogenetic models of substitution rate variation among lineages. *Biology Letters* 5:421–424. DOI: 10.1098/rsbl.2008.0729.

Hoang, D. T. et al. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35(2):518–522. DOI: 10.1093/molbev/msx281.

Holtz, M. D. et al. (2011). Characterization of *Fusarium avenaceum* from lupin in central Alberta: Genetic diversity, mating type and aggressiveness. *Canadian Journal of Plant Pathology* 33:61–76. DOI: 10.1080/07060661.2011.536651.

Hongsanan, S., Li, Y. M. et al. (2014). Revision of genera in *Asterinales*. *Fungal Diversity* 68:1–68. DOI: 10.1007/s13225-014-0307-4.

Hongsanan, S., Tian, Q. et al. (2015). *Meliolales*. *Fungal Diversity* 74:91–141. DOI: 10.1007/s13225-015-0344-7.

Hou, L. W. et al. (2020). The phoma-like dilemma. *Studies in Mycology* 96:309–396. DOI: 10.1016/j.simyco.2020.05.001.

Huang, X., Wang, J. et al. (2003). PCAP: A Whole-Genome Assembly Program. *Genome Research* 13:2164–2170. DOI: 10.1101/gr.1390403.1.

Huang, Y. L., Bowman, E. A. et al. (2018). Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi. *Mycologia* 110(1):47–62. DOI: 10.1080/00275514.2018.1442078.

Hubbard, M., Germida, J. J. and Vujanovic, V. (2014). Fungal endophytes enhance wheat heat and drought tolerance in terms of grain yield and second-generation seed viability. *Journal of Applied Microbiology* 116:109–122. DOI: 10.1111/jam.12311.

Huerta-Cepas, J. et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47:D309–D314. DOI: 10.1093/nar/gky1085.

Hulcr, J. and Cognato, A. I. (2010). Repeated evolution of crop theft in fungus-farming ambrosia beetles. *Evolution* 64(11):3205–3212. DOI: 10.1111/j.1558-5646.2010.01055.x.

Hyde, K. D., Chaiwan, N. et al. (2018). Mycosphere notes 169-224. *Mycosphere* 9(2):271–430. DOI: 10.5943/mycosphere/9/2/8.

Hyde, K. D., Tennakoon, D. S. et al. (2019). Fungal diversity notes 1036-1150: taxonomic and phylogenetic contributions on genera and species of fungal taxa. *Fungal Diversity* 96:1–242. DOI: 10.1007/s13225-019-00429-2.

Ilyukhin, E. (2022). First report of stem canker of *Rosa* spp. caused by *Didymella pomorum* in Canada. *Journal of Plant Pathology* 104:443. DOI: 10.1007/s42161-021-01020-y.

Irieda, H. et al. (2019). Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *PNAS* 116(2):496–505. DOI: 10.1073/pnas.1807297116.

Ismail, A. M. et al. (2012). *Lasiodiplodia* species associated with dieback disease of mango (*Mangifera indica*) in Egypt. *Australasian Plant Pathology* 41:649–660. DOI: 10.1007/s13313-012-0163-1.

Jaklitsch, W. M. et al. (2018). A preliminary account of the *Cucurbitariaceae*. *Studies in Mycology* 90:71–118. DOI: 10.1016/j.simyco.2017.11.002.

James, T. Y. et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443(7113):818–822. DOI: 10.1038/nature05110.

Jia, X., Dini-Andreote, F. and Falcão Salles, J. (2018). Community Assembly Processes of the Microbial Rare Biosphere. *Trends in Microbiology* 26(9):738–747. DOI: 10.1016/j.tim.2018.02.011.

Jiang, N. et al. (2021). Morphology and phylogeny of *Gnomoniopsis* (*Gnomoniaceae, Diaporthales*) from *Fagaceae* leaves in China. *Journal of Fungi* 7:792. DOI: 10.3390/jof7100792.

Jiao, W. B. and Schneeberger, K. (2017). The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* 36:64–70. DOI: 10.1016/j.pbi.2017.02.002.

Jombart, T., Balloux, F. and Dray, S. (2010). adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26(15):1907–1909. DOI: 10.1093/bioinformatics/btq292.

Jones, P. et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240. DOI: 10.1093/bioinformatics/btu031.

Joos, L. et al. (2020). Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genomics* 21:733. DOI: 10.1186/s12864-020-07126-4.

Jordal, B. H. and Cognato, A. I. (2012). Molecular phylogeny of bark and ambrosia beetles reveals multiple origins of fungus farming during periods of global warming. *BMC Evolutionary Biology* 12:133. DOI: 10.1186/1471-2148-12-133.

Junker, C., Draeger, S. and Schulz, B. (2012). A fine line - endophytes or pathogens in *Arabidopsis thaliana*. *Fungal Ecology* 5:657–662. DOI: 10.1016/j.funeco.2012.05.002.

Kadota, M. et al. (2020). Multifaceted Hi-C benchmarking: what makes a difference in chromosome-scale genome scaffolding? *GigaScience* 9:1–15. DOI: 10.1093/gigascience/giz158.

Kajitani, R. et al. (2019). Platanus-allee is a *de novo* haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nature Communications* 10:1702. DOI: 10.1038/s41467-019-09575-2.

Käll, L., Krogh, A. and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* 338:1027–1036. DOI: 10.1016/j.jmb.2004.03.016.

Kallow, S. et al. (2020). Challenges for *Ex Situ* Conservation of Wild Bananas: Seeds Collected in Papua New Guinea Have Variable Levels of Desiccation Tolerance. *Plants* 9:1243. DOI: 10.3390/plants9091243.

Kalyaanamoorthy, S. et al. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14(6):587–589. DOI: 10.1038/nmeth.4285.

Kandel, S. L. et al. (2017). An *In vitro* Study of Bio-Control and Plant Growth Promotion Potential of Salicaceae Endophytes. *Frontiers in Microbiology* 8(386). DOI: 10.3389/fmicb.2017.00386.

Karácsony, Z. et al. (2021). The fungus *Kalmusia longispora* is able to cause vascular necrosis on *Vitis vinifera*. *PLoS ONE* 16(10):e0258043. DOI: 10.1371/journal.pone.0258043.

Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. URL: https://cran.r-project.org/package=ggpubr.

– (2021). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. URL: https://cran.r-project.org/package=rstatix.

Kasson, M. T. et al. (2013). An inordinate fondness for *Fusarium*: Phylogenetic diversity of fusaria cultivated by ambrosia beetles in the genus *Euwallacea* on avocado and other plant hosts. *Fungal Genetics and Biology* 56:147–157. DOI: 10.1016/j.fgb.2013.04.004.

Katoh, K. and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version : Improvements in Performance and Usability. *Molecular Biology and Evolution* 30(4):772–780. DOI: 10.1093/molbev/mst010.

Kauff, F. and Lutzoni, F. (2002). Phylogeny of the Gyalectales and Ostropales (Ascomycota, Fungi): Among and within order relationships based on nuclear ribosomal RNA small and large subunits. *Molecular Phylogenetics and Evolution* 25:138–156. DOI: 10.1016/S1055-7903(02)00214-2.

– (2003). *Compat.py - a program to detect topological conflict between supported clades in phylogenetic trees*. URL: http://lutzonilab.org/downloadable-programs/.

Kauserud, H. et al. (2008). Mushroom fruiting and climate change. *PNAS* 105(10):3811–3814. DOI: 10.1073/pnas.0709037105.

Kavroulakis, N. et al. (2007). Role of ethylene in the protection of tomato plants against soil-borne fungal pathogens conferred by an endophytic *Fusarium solani* strain. *Journal of Experimental Botany* 58(14):3853–3864. DOI: 10.1093/jxb/erm230.

Khan, A. R. et al. (2018). A Comprehensive Study of *De Novo* Genome Assemblers: Current Challenges and Future Prospective. *Evolutionary Bioinformatics* 14:1–8. DOI: 10.1177/1176934318758650.

Kiers, E. T., Duhamel, M. et al. (2011). Reciprocal Rewards Stabilize Cooperation in the Mycorrhizal Symbiosis. *Science* 333(6044):880–883. DOI: 10.1126/science.1208473.

Kiers, E. T., Palmer, T. M. et al. (2010). Mutualisms in a changing world: an evolutionary perspective. *Ecology Letters* 13(12):1459–1474. DOI: 10.1111/j.1461-0248.2010.01538.x.

Kim, K.-T. et al. (2016). Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. *Frontiers in Plant Science* 7:186. DOI: 10.3389/fpls.2016.00186.

Kloppholz, S., Kuhn, H. and Requena, N. (2011). A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Current Biology* 21:1204–1209. DOI: 10.1016/j.cub.2011.06.044.

Knapp, D. G. et al. (2018). Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Scientific Reports* 8:6321. DOI: 10.1038/s41598-018-24686-4.

Kohler, A. et al. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* 47(4):410–415. DOI: 10.1038/ng.3223.

Kokkonen, M. et al. (2010). Mycotoxin production of selected *Fusarium* species at different culture conditions. *International Journal of Food Microbiology* 143:17–25. DOI: 10.1016/j.ijfoodmicro.2010.07.015.

Kolmogorov, M. et al. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 37:540–546. DOI: 10.1038/s41587-019-0072-8.

Kooij, P. W. and Pellicer, J. (2020). Genome size vs. genome assemblies: are the genomes truly expanded in polyploid fungal symbionts? *Genome Biology and Evolution* 12(12):2384–2390. DOI: 10.1093/gbe/evaa217.

Koren, S. and Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology* 23:110–120. DOI: 10.1016/j.mib.2014.11.014.

Koren, S., Schatz, M. C. et al. (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* 30(7):693–700. DOI: 10.1038/nbt.2280.

Koren, S., Walenz, B. P. et al. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27:722–736. DOI: 10.1101/gr.215087.116.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. DOI: 10.1186/1471-2105-5-59.

Kosakovsky Pond, S. L., Frost, S. D. and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. DOI: 10.1093/bioinformatics/bti079.

Kosakovsky Pond, S. L., Wisotsky, S. R. et al. (2021). Contrast-FEL—A Test for Differences in Selective Pressures at Individual Sites among Clades and Sets of Branches. *Molecular Biology and Evolution* 38(3):1184–1198. DOI: 10.1093/molbev/msaa263.

Kozlov, A. M. et al. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35(21):4453–4455. DOI: 10.1093/bioinformatics/btz305.

Krijger, J. J. et al. (2014). Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. *BMC Genomics* 15:722. DOI: 10.1186/1471-2164-15-722.

Krogh, A. et al. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* 305:567–580. DOI: 10.1006/jmbi.2000.4315.

Kubicek, C. P., Starr, T. L. and Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* 52:427–451. DOI: 10.1146/annurev-phyto-102313-045831.

Kullman, B., Tamm, H. and Kullman, K. (2005). *Fungal Genome Size Database*. URL: http://www.zbi.ee/fungal-genomesize.

Kunert, N. et al. (2015). Higher tree transpiration due to road-associated edge effects in a tropical moist lowland forest. *Agricultural and Forest Meteorology* 213:183–192. DOI: 10.1016/j.agrformet.2015.06.009.

Kuo, H. C. et al. (2014). Secret lifestyles of *Neurospora crassa*. *Scientific Reports* 4:5135. DOI: 10.1038/srep05135.

LaBella, A. L. et al. (2021). Signatures of optimal codon usage in metabolic genes inform budding yeast ecology. *PLoS Biology* 19(4):e3001185. DOI: 10.1371/journal.pbio.3001185.

Laetsch, D. R. and Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research* 6:1287. DOI: 10.12688/f1000research.12232.1.

Lahrmann, U. et al. (2015). Mutualistic root endophytism is not associated with the reduction of saprotrophic traits and requires a noncompromised plant innate immunity. *New Phytologist* 207(3):841–857. DOI: 10.1111/nph.13411.

Lamb, P. D. et al. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* 28:420–430. DOI: 10.1111/mec.14920.

Lan, J. H. et al. (2015). Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Human Immunology* 76:166–175. DOI: 10.1016/j.humimm.2014.12.016.

Lang, D. T. (2020). *RCurl: General Network (HTTP/FTP/...) Client Interface for R*. URL: https://cran.r-project.org/package=RCurl.

Langhe, E. D. et al. (2009). Why Bananas Matter: An introduction to the history of banana domestication. *Ethnobotany Research & Applications* 7:165–177. DOI: 10.17348/era.7.0.165-177.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357–359. DOI: 10.1038/nmeth.1923.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276. DOI: 10.1093/bioinformatics/btu531.

Larsson, J. (2020). *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*. URL: https://cran.r-project.org/package=eulerr.

Lartillot, N. and Delsuc, F. (2012). Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution* 66(6):1773–1787. DOI: 10.1111/j.1558-5646.2011.01558.x.

Lartillot, N., Phillips, M. J. and Ronquist, F. (2016). A mixed relaxed clock model. *Philosophical Transactions of the Royal Society B* 371:20150132. DOI: 10.1098/rstb.2015.0132.

Latgé, J.-P. and Chamilos, G. (2020). *Aspergillus fumigatus* and Aspergillosis in 2019. *Clinical Microbiology Reviews* 33(1):e00140–18. DOI: 10.1128/CMR.00140-18.

Le Cam, B. et al. (2019). Population genome sequencing of the scab fungal species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata*. *G3* 9:2405–2414. DOI: 10.1534/g3.119.400047.

Le Cocq, K. et al. (2016). Exploitation of endophytes for sustainable agricultural intensification. *Molecular Plant Pathology* 18(3):469–473. DOI: 10.1111/mpp.12483.

Lebreton, A. et al. (2021). Evolution of the Mode of Nutrition in Symbiotic and Saprotrophic Fungi in Forest Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 52:385–404. DOI: 10.1146/annurev-ecolsys-012021-114902.

Lees, J. et al. (2012). Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Research* 40:D465–D471. DOI: 10.1093/nar/gkr1181.

Leist, N. and Krämer, S. (2011). *ISTA Working Sheets on Tetrazolium Testing, Volume II, 1st Edition (2003)*.

Leoni, C. et al. (2013). *Fusarium oxysporum* f.sp. *cepae* dynamics: in-plant multiplication and crop sequence simulations. *European Journal of Plant Pathology* 137:545–561. DOI: 10.1007/s10658-013-0268-6.

Leroy, C. et al. (2019). How significant are endophytic fungi in bromeliad seeds and seedlings? Effects on germination, survival and performance of two epiphytic plant species. *Fungal Ecology* 39:296–306. DOI: 10.1016/j.funeco.2019.01.004.

Letunic, I., Doerks, T. and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 40:D302–D305. DOI: 10.1093/nar/gkr931.

Levasseur, A. et al. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels* 6:41. DOI: 10.1186/1754-6834-6-41.

Lewin, H. A. et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. *PNAS* 115(17):4325–4333. DOI: 10.1073/pnas.1720115115.

Li, D., Luo, R. 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. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997.

– (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094–3100. DOI: 10.1093/bioinformatics/bty191.

Li, H., Handsaker, B. et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, J., Fokkens, L. et al. (2020). Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environmental Microbiology* 22(12):4985–5004. DOI: 10.1111/1462-2920.15095.

Li, X. Z., Song, M. L. et al. (2017). The effect of seed-borne fungi and *Epichloë* endophyte on seed germination and biomass of *Elymus sibiricus*. *Frontiers in Microbiology* 8:2488. DOI: 10.3389/fmicb.2017.02488.

Li, Y., Steenwyk, J. L. et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Current Biology* 31(8):1653–1665. DOI: 10.1016/j.cub.2021.01.074.

Ligoxigakis, E. K. et al. (2013). First Report of Leaf Spot of *Phoenix theophrasti* Caused by *Paraconiothyrium variabile* in Greece. *Plant Disease* 97(9):1250. DOI: 10.1094/PDIS-01-13-0114-PDN.

Linder, M., Britton, T. and Sennblad, B. (2011). Evaluation of bayesian models of substitution rate evolution-parental guidance versus mutual independence. *Systematic Biology* 60(3):329–342. DOI: 10.1093/sysbio/syr009.

Lione, G. et al. (2019). The emerging pathogen of chestnut *Gnomoniopsis castaneae*: the challenge posed by a versatile fungus. *European Journal of Plant Pathology* 153:671–685. DOI: 10.1007/s10658-018-1597-2.

Liu, F., Ma, Z. et al. (2022). Updating species diversity of *Colletotrichum*, with a phylogenomic overview. *Studies in Mycology* 101:1–56. DOI: 10.3114/sim.2022.101.01.

Liu, H., Wu, S. et al. (2021). SMARTdenovo: a *de novo* assembler using long noisy reads. *Gigabyte*:1–9. DOI: 10.46471/gigabyte.15.

Liu, U., Breman, E. et al. (2018). The conservation value of germplasm stored at the Millennium Seed Bank, Royal Botanic Gardens, Kew, UK. *Biodiversity and Conservation* 27(6):1347–1386. DOI: 10.1007/s10531-018-1497-y.

Liu, U., Cossu, T. A. et al. (2020). Conserving orthodox seeds of globally threatened plants ex situ in the Millennium Seed Bank, Royal Botanic Gardens, Kew, UK: the status of seed collections. *Biodiversity and Conservation* 29:2901–2949. DOI: 10.1007/s10531-020-02005-6.

Liu, W. C., Li, C. Q. et al. (2010). Phylogenetic diversity of culturable fungi associated with two marine sponges: *Haliclona simulans* and *Gelliodes carnosa*, collected from the Hainan Island coastal waters of the South China Sea. *Fungal Diversity* 42:1–15. DOI: 10.1007/s13225-010-0022-8.

Lo Presti, L. et al. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology* 66:513–545. DOI: 10.1146/annurev-arplant-043014-114623.

Lofgren, L. A. et al. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology* 28:721–720. DOI: 10.1111/mec.14995.

Lombard, L. et al. (2015). Generic concepts in *Nectriaceae*. *Studies in Mycology* 80:189–245. DOI: 10.1016/j.simyco.2014.12.002.

Lu, S. et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* 48:D265–D268. DOI: 10.1093/nar/gkz991.

Luo, R. et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 4:18. DOI: 10.1186/s13742-015-0069-2.

Lupas, A. (1997). Predicting coiled-coil regions in proteins. *Current Opinion in Structural Biology* 7:388–393. DOI: 10.1016/S0959-440X(97)80056-5.

Lutzoni, F. et al. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* 9(5451). DOI: 10.1038/s41467-018-07849-9.

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155. DOI: 10.1126/science.290.5494.1151.

Lynch, M. D. and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology* 13(4):217–229. DOI: 10.1038/nrmicro3400.

Ma, L. J., Geiser, D. M. et al. (2013). *Fusarium* pathogenomics. *Annual Review of Microbiology* 67:399–416. DOI: 10.1146/annurev-micro-092412-155650.

Ma, L. J., van der Does, H. C. et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464(7287):367–373. DOI: 10.1038/nature08850.

Magaña-Dueñas, V., Stchigel, A. M. and Cano-Lira, J. F. (2021). New coelomycetous fungi from freshwater in Spain. *Journal of Fungi* 7:368. DOI: 10.3390/jof7050368.

Majoros, W. H., Pertea, M. and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* 20(16):2878–2879. DOI: 10.1093/bioinformatics/bth315.

Mallick, S. et al. (2009). The difficulty of avoiding false positives in genome scans for natural selection. *Genome Research* 19(5):922–933. DOI: 10.1101/gr.086512.108.

Manamgoda, D. S. et al. (2013). Endophytic *Colletotrichum* from tropical grasses with a new species *C. endophytica*. *Fungal Diversity* 61:107–115. DOI: 10.1007/s13225-013-0256-3.

Marcelino, J. et al. (2008). *Colletotrichum acutatum* var. *fioriniae* (teleomorph: *Glomerella acutata* var. *fioriniae* var. nov.) infection of a scale insect. *Mycologia* 100(3):353–374. DOI: 10.3852/07-174R.

Marqués-Gálvez, J. E. et al. (2021). Desert truffle genomes reveal their reproductive modes and new insights into plant–fungal interaction and ectendomycorrhizal lifestyle. *New Phytologist* 229:2917–2932. DOI: 10.1111/nph.17044.

Martin, F., Kohler, A. et al. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464:1033–1038. DOI: 10.1038/nature08867.

Martin, G., Cardi, C. et al. (2020). Genome ancestry mosaics reveal multiple and cryptic contributors to cultivated banana. *Plant Journal* 102:1008–1025. DOI: 10.1111/tpj.14683.

Martin, P. L. and Peter, K. A. (2021). Quantification of *Colletotrichum fioriniae* in orchards and deciduous forests indicates it is primarily a leaf endophyte. *Phytopathology* 111:333–344. DOI: 10.1094/PHYTO-05-20-0157-R.

Martínez-Soto, D. et al. (2022). Differential colonization of the plant vasculature between endophytic versus pathogenic *Fusarium oxysporum* strains. *MPMI*:1–27. DOI: 10.1094/MPMI-08-22-0166-SC.

Martino, E. et al. (2018). Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytologist* 217(3):1213–1229. DOI: 10.1111/nph.14974.

Maryani, N. et al. (2019). Phylogeny and genetic diversity of the banana Fusarium wilt pathogen *Fusarium oxysporum* f. sp. *cubense* in the Indonesian centre of origin. *Studies in Mycology* 92:155–194. DOI: 10.1016/j.simyco.2018.06.003.

Mascarin, G. M. and Jaronski, S. T. (2016). The production and uses of *Beauveria bassiana* as a microbial insecticide. *World Journal of Microbiology and Biotechnology* 32:177. DOI: 10.1007/s11274-016-2131-3.

Matesanz, S. et al. (2019). Estimating belowground plant abundance with DNA metabarcoding. *Molecular Ecology Resources* 19:1265–1277. DOI: 10.1111/1755-0998.13049.

Mateus, I. D. et al. (2019). Dual RNA-seq reveals large-scale non-conserved genotype × genotype-specific genetic reprograming and molecular crosstalk in the mycorrhizal symbiosis. *ISME Journal* 13(5):1226–1238. DOI: 10.1038/s41396-018-0342-3.

Mavromatis, K. et al. (2012). The Fast Changing Landscape of Sequencing Technologies and Their Impact on Microbial Genome Assemblies and Annotation. *PLoS ONE* 7(12):e48837. DOI: 10.1371/journal.pone.0048837.

Mawar, R., Manjunatha, B. L. and Kumar, S. (2021). Commercialization, Diffusion and Adoption of Bioformulations for Sustainable Disease Management in Indian Arid Agriculture: Prospects and Challenges. *Circular Economy and Sustainability* 1:1367–1385. DOI: 10.1007/s43615-021-00089-y.

McCluskey, K., Wiest, A. and Plamann, M. (2010). The Gungal Genetics Stock Center: a repository for 50 years of fungal genetics research. *Journal of Biosciences* 35(1):119–126. DOI: 10.1007/s12038-010-0014-6.

McDonald, A. G., Boyce, S. and Tipton, K. F. (2009). ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Research* 37:D593–D597. DOI: 10.1093/nar/gkn582.

McKinney, M. L. and Lockwood, J. L. (1999). Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology and Evolution* 14(11):450–453. DOI: 10.1016/S0169-5347(99)01679-1.

Mead, M. E. et al. (2021). An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related *Aspergillus* fungi. *Genetics* 218(2):iyab066. DOI: 10.1093/genetics/iyab066.

Mendgen, K., Hahn, M. and Deising, H. (1996). Morphogenesis and Mechanisms of Penetration By Plant Pathogenic Fungi. *Annual Review of Phytopathology* 34(1):367–386. DOI: 10.1146/annurev.phyto.34.1.367.

Mendoza, A. R. and Sikora, R. A. (2009). Biological control of *Radopholus similis* in banana by combined application of the mutualistic endophyte *Fusarium oxysporum* strain 162, the egg pathogen *Paecilomyces lilacinus* strain 251 and the antagonistic bacteria *Bacillus firmus*. *BioControl* 54(2):263–272. DOI: 10.1007/s10526-008-9181-x.

Meng, Y. et al. (2022). Genome sequence assembly algorithms and misassembly identification methods. *Molecular Biology Reports* 49:11133–11148. DOI: 10.1007/s11033-022-07919-8.

Menzies, J. G., Koch, C. and Seywerd, F. (1990). Additions to the Host Range of *Fusarium oxysporum* f. sp. *radicis-lycopersici*. *Plant Disease* 74:569–572. DOI: 10.1094/PD-74-0569.

Mesny, F., Miyauchi, S. et al. (2021). Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. *Nature Communications* 12:7227. DOI: 10.1038/s41467-021-27479-y.

Mesny, F. and Vannier, N. (2020). *Detecting the effect of biological categories on genome composition*. URL: https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition.

Meyer, R. S., Duval, A. E. and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* 196:29–48. DOI: 10.1111/j.1469-8137.2012.04253.x.

Mi, H. et al. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47:D419–D426. DOI: 10.1093/nar/gky1038.

Miller, A. N., Karakehian, J. and Raudabaugh, D. B. (2022). Next-Generation Sequencing of Ancient and Recent Fungarium Specimens. *Journal of Fungi* 8(9):932. DOI: 10.3390/jof8090932.

Miller, J. R., Koren, S. and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327. DOI: 10.1016/j.ygeno.2010.03.001.

Miller, K. E., Hopkins, K. et al. (2016). Metabarcoding of fungal communities associated with bark beetles. *Ecology and Evolution* 6(6):1590–1600. DOI: 10.1002/ece3.1925.

Miller, M. A., Pfeiffer, W. and Schwartz, T. (2010). 'Creating the CIPRES Science Gateway for inference of large phylogenetic trees'. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*. IEEE. New Orleans, LA:pp. 1–8. DOI: 10.1109/GCE.2010.5676129.

Minh, B. Q. et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37(5):1530–1534. DOI: 10.1093/molbev/msaa015.

Mishra, S., Bhattacharjee, A. and Sharma, S. (2021). An Ecological Insight into the Multifaceted World of Plant-Endophyte Association. *Critical Reviews in Plant Sciences* 40(2):127–146. DOI: 10.1080/07352689.2021.1901044.

Mistry, J., Chuguransky, S. et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research* 49:D412–D419. DOI: 10.1093/nar/gkaa913.

Mistry, J., Finn, R. D. et al. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* 41(12):e121. DOI: 10.1093/nar/gkt263.

Mita, K. et al. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Research* 11:27–35. DOI: 10.1093/dnares/11.1.27.

Miura, S. et al. (2020). A new method for inferring timetrees from temporally sampled molecular sequences. *PLoS Computational Biology* 16(1):e1007046. DOI: 10.1371/journal.pcbi.1007046.

Miyauchi, S., Kiss, E. et al. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications* 11:5125. DOI: 10.1038/s41467-020-18795-w.

Miyauchi, S., Navarro, D. et al. (2017). The integrative omics of white-rot fungus *Pycnoporus coccineus* reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. *PLoS ONE* 12(4):e0175528. DOI: 10.1371/journal.pone.0175528.

Moral, J. et al. (2018). *Didymella glomerata* causing leaf blight on pistachio. *European Journal of Plant Pathology* 151:1095–1099. DOI: 10.1007/s10658-018-1422-y.

Morris, E. K., Caruso, T. et al. (2014). Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution* 4(18):3514–3524. DOI: 10.1002/ece3.1155.

Morris, J. L., Puttick, M. N. et al. (2018). The timescale of early land plant evolution. *PNAS*:201719588. DOI: 10.1073/pnas.1719588115.

Mueller, U. G. et al. (2005). The evolution of agriculture in insects. *Annual Review of Ecology, Evolution, and Systematics* 36:563–595. DOI: 10.1146/annurev.ecolsys.36.102003.152626.

Muir, P. et al. (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biology* 17:53. DOI: 10.1186/s13059-016-0917-0.

Murashige, T. and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* 15(3):473–497. DOI: 10.1111/j.1399-3054.1962.tb08052.x.

Murphy, B. R., Doohan, F. M. and Hodkinson, T. R. (2019). 'Prospecting Crop Wild Relatives for Beneficial Endophytes'. In: *Endophytes for a Growing World*. Ed. by T. Hodkinson et al. Cambridge, UK: Cambridge University Press. Chap. 18:pp. 390–410. DOI: 10.1017/9781108607667.019.

Murphy, B. R., Hodkinson, T. R. and Doohan, F. M. (2018). Endophytic *Cladosporium* strains from a crop wild relative increase grain yield in Barley. *Biology and Environment: Proceedings of the Royal Irish Academy* 118(3):147–156. DOI: 10.3318/bioe.2018.14.

Murphy, B. R., Jadwiszczak, M. J. et al. (2018). Endophytes from the crop wild relative *Hordeum secalinum* L. improve agronomic traits in unstressed and salt-stressed barley. *Cogent Food & Agriculture* 4:1549195. DOI: 10.1080/23311932.2018.1549195.

Murrell, B. et al. (2015). Gene-wide Identification of Episodic Selection. *Molecular Biology and Evolution* 32(5):1365–1371. DOI: 10.1093/molbev/msv035.

Muszewska, A. et al. (2019). Transposable elements contribute to fungal genes and impact fungal lifestyle. *Scientific Reports* 9:4307. DOI: 10.1038/s41598-019-40965-0.

Myers, E. W. et al. (2000). A Whole-Genome Assembly of *Drosophila*. *Science* 287(5461):2196–2204. DOI: 10.1126/science.287.5461.2196.

Na, F. et al. (2018). Two novel fungal symbionts *Fusarium kuroshium* sp. nov. and *Graphium kuroshium* sp. nov. of kuroshio shot hole borer (*Euwallacea* sp. nr. *fornicatus*) cause fusarium dieback on woody host species in California. *Plant Disease* 102(6):1154–1164. DOI: 10.1094/PDIS-07-17-1042-RE.

Necci, M. et al. (2017). MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* 33(9):1402–1404. DOI: 10.1093/bioinformatics/btx015.

Nelson, A. et al. (2020). Double lives: transfer of fungal endophytes from leaves to woody substrates. *PeerJ* 8:e9341. DOI: 10.7717/peerj.9341.

Netto, M. S. et al. (2014). Species of *Lasiodiplodia* associated with papaya stem-end rot in Brazil. *Fungal Diversity* 67:127–141. DOI: 10.1007/s13225-014-0279-4.

Newman, J. A., Gillis, S. and Hager, H. A. (2022). Costs, Benefits, Parasites and Mutualists: The Use and Abuse of the Mutualism–Parasitism Continuum Concept for *Epichloë* Fungi. *Philosophy, Theory, and Practice in Biology* 14:9. DOI: 10.3998/ptpbio.2103.

Nguyen, N. H. et al. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* 20:241–248. DOI: 10.1016/j.funeco.2015.06.006.

Nic Lughadha, E. et al. (2020). Extinction risk and threats to plants and fungi. *Plants People Planet* 2:389–408. DOI: 10.1002/ppp3.10146.

Niehaus, E. M. et al. (2016). Comparative "Omics" of the *Fusarium fujikuroi* Species Complex Highlights Differences in Genetic Potential and Metabolite Synthesis. *Genome Biology and Evolution* 8(11):3574–3599. DOI: 10.1093/gbe/evw259.

Nielsen, H. (2017). 'Predicting Secretory Proteins with SignalP'. In: *Protein Function Prediction*. Ed. by D. Kihara. Vol. 1611. Humana Press, New York, NY. Chap. 6:pp. 59–73. ISBN: 978-1-4939-7015-5. DOI: 10.1007/978-1-4939-7015-5_6.

Nilsson, R. H. et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* 47:D259–D264. DOI: 10.1093/nar/gky1022.

Niu, G. et al. (2009). Comparative toxicity of mycotoxins to navel orangeworm (*Amyelois transitella*) and corn earworm (*Helicoverpa zea*). *Journal of Chemical Ecology* 35:951–957. DOI: 10.1007/s10886-009-9675-8.

Noë, R. and Kiers, E. T. (2018). Mycorrhizal Markets, Firms, and Co-ops. *Trends in Ecology and Evolution* 33(10):777–789. DOI: 10.1016/j.tree.2018.07.007.

Nygren, K. et al. (2011). A comprehensive phylogeny of *Neurospora* reveals a link between reproductive mode and molecular evolution in fungi. *Molecular Phylogenetics and Evolution* 59(3):649–663. DOI: 10.1016/j.ympev.2011.03.023.

O'Donnell, K., Al-Hatmi, A. M. S. et al. (2020). No to *Neocosmospora*: Phylogenomic and Practical Reasons for Continued Inclusion of the *Fusarium solani* Species Complex in the Genus *Fusarium*. *mSphere* 5:e00810–20. DOI: 10.1128/msphere.00810-20.

O'Donnell, K., Libeskind-Hadas, R. et al. (2016). Invasive Asian *Fusarium* – *Euwallacea* ambrosia beetle mutualists pose a serious threat to forests, urban landscapes and the avocado industry. *Phytoparasitica* 44:435–442. DOI: 10.1007/s12600-016-0543-0.

O'Donnell, K., Rooney, A. P. et al. (2013). Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genetics and Biology* 52:20–31. DOI: 10.1016/j.fgb.2012.12.004.

O'Donnell, K., Sink, S. et al. (2015). Discordant phylogenies suggest repeated host shifts in the *Fusarium-Euwallacea* ambrosia beetle mutualism. *Fungal Genetics and Biology* 82:277–290. DOI: 10.1016/j.fgb.2014.10.014.

O'Donnell, K., Sutton, D. A. et al. (2009). Novel multilocus sequence typing scheme reveals high genetic diversity of human pathogenic members of the *Fusarium incarnatum-F. equiseti* and *F. chlamydosporum* species complexes within the United States. *Journal of Clinical Microbiology* 47(12):3851–3861. DOI: 10.1128/JCM.01616-09.

O'Malley, M. A. and Parke, E. C. (2020). 'Philosophy of Microbiology'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/fall2020/entries/microbiology/.

Ohta, T. (1996). The current significance and standing of neutral and nearly neutral theories. *BioEssays* 18(8):673–677. DOI: 10.1002/bies.950180811.

Okello, P. N. et al. (2020). Characterization of species of *Fusarium* causing root rot of Soybean (*Glycine max* L.) in South Dakota, USA. *Canadian Journal of Plant Pathology* 42(4):560–571. DOI: 10.1080/07060661.2020.1746695.

Oksanen, J. et al. (2019). *vegan: Community Ecology Package*. URL: https://cran.r-project.org/package=vegan.

Oldrup, E. et al. (2010). Localization of endophytic *Undifilum* fungi in locoweed seed and influence of environmental parameters on a locoweed *in vitro* culture system. *Botany* 88(5):512–521. DOI: 10.1139/B10-026.

Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv: 1403.2805.

Ordonez, N. et al. (2015). Worse Comes to Worst: Bananas and Panama Disease—When Plant and Pathogen Clones Meet. *PLoS Pathogens* 11(11):e1005197. DOI: 10.1371/journal.ppat.1005197.

Oskay, F. et al. (2022). Seed quantity affects the fungal community composition detected using metabarcoding. *Scientific Reports* 12:3060. DOI: 10.1038/s41598-022-06997-9.

Özkurt, E. et al. (2020). Seed-Derived Microbial Colonization of Wild Emmer and Domesticated Bread Wheat (*Triticum dicoccoides* and *T. aestivum*) Seedlings Shows Pronounced Differences in Overall Diversity and Composition. *mBio* 11(6):e02637–20. DOI: 10.1128/mBio.02637-20.

Palma-Guerrero, J. et al. (2016). Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intraspecific variability in transcription profiles. *Molecular plant pathology* 17(6):845–859. DOI: 10.1111/mpp.12333.

Palmer, J. M. and Stajich, J. (2020). *Funannotate v1.8.1: Eukaryotic genome annotation*. DOI: 10.5281/zenodo.4054262.

Paradis, E. and Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3):526–528. DOI: 10.1093/bioinformatics/bty633.

Parmar, S. et al. (2018). Endophytic fungal community of *Dysphania ambrosioides* from two heavy metal-contaminated sites: evaluated by culture-dependent and culture-independent approaches. *Microbial Biotechnology* 11(6):1170–1183. DOI: 10.1111/1751-7915.13308.

Parra-Sanchez, E. and Banks-Leite, C. (2020). The magnitude and extent of edge effects on vascular epiphytes across the Brazilian Atlantic Forest. *Scientific Reports* 10:18847. DOI: 10.1038/s41598-020-75970-1.

Parsa, S. et al. (2016). Fungal endophytes in germinated seeds of the common bean, *Phaseolus vulgaris*. *Fungal Biology* 120(5):783–790. DOI: 10.1016/j.funbio.2016.01.017.

Paszkiewicz, K. and Studholme, D. J. (2010). *De novo* assembly of short sequence reads. *Briefings in Bioinformatics* 11(5):457–472. DOI: 10.1093/bib/bbq020.

Paton, A. et al. (2020). Plant and fungal collections: Current status, future perspectives. *Plants People Planet* 2:499–514. DOI: 10.1002/ppp3.10141.

Pattengale, N. D. et al. (2009). 'How Many Bootstrap Replicates Are Necessary?' In: *13th Annual International Conference on Research in Computational Molecular Biology*. Ed. by S. Batzoglou. Tucson, AZ, USA: Springer-Verlag Berlin Heidelberg:pp. 184–200. DOI: 10.1089/cmb.2009.0179.

Paulus, B. C. et al. (2006). Diversity and distribution of saprobic microfungi in leaf litter of an Australian tropical rainforest. *Mycological Research* 110(12):1441–1454. DOI: 10.1016/j.mycres.2006.09.002.

Pearce, T. R. et al. (2020). International collaboration between collections-based institutes for halting biodiversity loss and unlocking the useful properties of plants and fungi. *Plants, People, Planet* 2:515–534. DOI: 10.1002/ppp3.10149.

Peck, L. D. et al. (2021). Historical genomics reveals the evolutionary mechanisms behind multiple outbreaks of the host-specific coffee wilt pathogen *Fusarium xylarioides*. *BMC Genomics* 22:404. DOI: 10.1186/s12864-021-07831-8.

Pedersen, T. L. (2021). *ggforce: Accelerating 'ggplot2'*. URL: https://cran.r-project.org/web/packages/ggforce/index.html.

Pedruzzi, I. et al. (2015). HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Research* 43:D1064–D1070. DOI: 10.1093/nar/gku1002.

Pellicer, J., Powell, R. F. and Leitch, I. J. (2020). 'The Application of Flow Cytometry for Estimating Genome Size, Ploidy Level Endopolyploidy, and Reproductive Modes in Plants'. In: *Molecular Plant Taxonomy. Methods in Molecular Biology*. Ed. by P. Besse. Vol. 2222. New York: Humana. Chap. 17:pp. 325–361. ISBN: 978-1-0716-0997-2. DOI: 10.1007/978-1-0716-0997-2_17.

Pendleton, A. L. et al. (2014). Duplications and losses in gene families of rust pathogens highlight putative effectors. *Frontiers in Plant Science* 5:299. DOI: 10.3389/fpls.2014.00299.

Peng, Y. et al. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428. DOI: 10.1093/bioinformatics/bts174.

Peres, S. (2016). Saving the gene pool for the future: Seed banks as archives. *Studies in History and Philosophy of Biological and Biomedical Sciences* 55:96–104. DOI: 10.1016/j.shpsc.2015.09.002.

Perrier, X. et al. (2011). Multidisciplinary perspectives on (*Musa* spp.) domestication. *PNAS* 108(28):11311–11318. DOI: 10.1073/pnas.1102001108.

Peter, M. et al. (2016). Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nature Communications* 7:12662. DOI: 10.1038/ncomms12662.

Petrini, O. (1991). 'Fungal Endophytes of Tree Leaves'. In: *Microbial Ecology of Leaves*. Ed. by J. H. Andrew and S. S. Hirano. New York: Springer:pp. 179–197. ISBN: 978-1-4612-7822-1. DOI: 10.1007/978-1-4612-3168-4_9.

Pflug, J. M. et al. (2020). Measuring Genome Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in Beetles (Coleoptera). *G3* 10:3047–3060. DOI: 10.1101/761304.

Philipson, M. N. and Christey, M. C. (1986). The relationship of host and endophyte during flowering, seed formation, and germination of *Lolium perenne*. *New Zealand Journal of Botany* 24:125–134. DOI: 10.1080/0028825X.1986.10409724.

Pierleoni, A., Martelli, P. and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* 9:392. DOI: 10.1186/1471-2105-9-392.

Pinheiro, J. et al. (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. URL: https://cran.r-project.org/package=nlme.

Piñol, J., Senar, M. A. and Symondson, W. O. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology* 28:407–419. DOI: 10.1111/mec.14776.

Pirozynski, K. and Malloch, D. (1975). The origin of land plants: A matter of mycotrophism. *Biosystems* 6(3):153–164. DOI: 10.1016/0303-2647(75)90023-4.

Plett, J. M., Kemppainen, M. et al. (2011). A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Current Biology* 21:1197–1203. DOI: 10.1016/j.cub.2011.05.033.

Plett, J. M. and Martin, F. (2015). Reconsidering mutualistic plant-fungal interactions through the lens of effector biology. *Current Opinion in Plant Biology* 26:45–50. DOI: 10.1016/j.pbi.2015.06.001.

Plett, J. M., Plett, K. L. et al. (2020). Mycorrhizal effector PaMiSSP10b alters polyamine biosynthesis in *Eucalyptus* root cells and promotes root colonization. *New Phytologist* 228:712–727. DOI: 10.1111/nph.16759.

Ploetz, R. C. (2005). Panama Disease: An Old Nemesis Rears Its Ugly Head. Part 1. The Beginnings of the Banana Export Trades. *Plant Health Progress* 6(1). DOI: 10.1094/php-2005-1221-01-rv.

Põlme, S. et al. (2018). Host preference and network properties in biotrophic plant–fungal associations. *New Phytologist* 217:1230–1239. DOI: 10.1111/nph.14895.

Poos, M. S. and Jackson, D. A. (2012). Addressing the removal of rare species in multivariate bioassessments: The impact of methodological choices. *Ecological Indicators* 18:82–90. DOI: 10.1016/j.ecolind.2011.10.008.

Porto, B. N. et al. (2019). Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. *PLoS ONE* 14(4):e0215598. DOI: 10.1371/journal.pone.0215598.

Powell, C. L. E., Waskin, S. and Battistuzzi, F. U. (2020). Quantifying the Error of Secondary vs. Distant Primary Calibrations in a Simulated Environment. *Frontiers in Genetics* 11:252. DOI: 10.3389/fgene.2020.00252.

POWO (2019). *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew*. URL: http://www.plantsoftheworldonline.org/.

Prescott, T. et al. (2018). 'Useful Fungi'. In: *State of the World's Fungi 2018*. Ed. by K. Willis. Royal Botanic Gardens, Kew. Chap. 4:pp. 24–31. ISBN: 978-1-84246-678-0.

Press, M. O. et al. (2017). Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *bioRxiv [Preprint]*. DOI: 10.1101/198713.

Promputtha, I. et al. (2010). Can leaf degrading enzymes provide evidence that endophytic fungi becoming saprobes? *Fungal Diversity* 41:89–99. DOI: 10.1007/s13225-010-0024-6.

Puttick, M. and Title, P. (2019). *MCMCtreeR: Prepare MCMCtree Analyses and Plot Bayesian Divergence Time Analyses Estimates on Trees*. URL: https://cran.r-project.org/package=MCMCtreeR.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842. DOI: 10.1093/bioinformatics/btq033.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna. URL: https://www.r-project.org/.

Raffaele, S. and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 10(6):417–430. DOI: 10.1038/nrmicro2790.

Rafiqi, M. et al. (2012). Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Current Opinion in Plant Biology* 15:477–482. DOI: 10.1016/j.pbi.2012.05.003.

Raghavendra, A. K. H. et al. (2013). Exclusionary interactions among diverse fungi infecting developing seeds of *Centaurea stoebe*. *FEMS Microbiology Ecology* 84:143–153. DOI: 10.1111/1574-6941.12045.

Rahman, S. et al. (2021). Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. *PNAS* 118(20):e2023575118. DOI: 10.1073/pnas.2023575118.

Rangwala, S. H. et al. (2021). Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Research* 31:159–169. DOI: 10.1101/gr.266932.120.

Rao, S. et al. (2018). The Landscape of Repetitive Elements in the Refined Genome of Chilli Anthracnose Fungus *Colletotrichum truncatum*. *Frontiers in Microbiology* 9:2367. DOI: 10.3389/fmicb.2018.02367.

Rashmi, M., Kushveer, J. S. and Sarma, V. V. (2019). A worldwide list of endophytic fungi with notes on ecology and diversity. *Mycosphere* 10(1):798–1079. DOI: 10.5943/mycosphere/10/1/19.

Rath, M. et al. (2014). Combining microtomy and confocal laser scanning microscopy for structural analyses of plant-fungus associations. *Mycorrhiza* 24:293–300. DOI: 10.1007/s00572-013-0530-y.

Rawlings, N. D., Barrett, A. J. and Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 40:D343–D350. DOI: 10.1093/nar/gkr987.

Redkar, A. et al. (2022). Conserved secreted effectors contribute to endophytic growth and multihost plant compatibility in a vascular wilt fungus. *The Plant cell* 34(9):3214–3232. DOI: 10.1093/plcell/koac174.

Redman, R. S., Sheehan, K. B. et al. (2002). Thermotolerance Generated by Plant/Fungal Symbiosis. *Science* 298(5598):1581. DOI: 10.1126/science.1078055.

Redman, R. S., Dunigan, D. D. and Rodriguez, R. J. (2001). Fungal symbiosis from mutualism to parasitism: Who controls the outcome, host or invader? *New Phytologist* 151(3):705–716. DOI: 10.1046/j.0028-646x.2001.00210.x.

Resl, P. et al. (2022). Large differences in carbohydrate degradation and transport potential among lichen fungal symbionts. *Nature Communications* 13:2634. DOI: 10.1038/s41467-022-30218-6.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3(2):217–223. DOI: 10.1111/j.2041-210X.2011.00169.x.

Ribeiro, T. H. C. et al. (2020). Transcriptome analyses suggest that changes in fungal endophyte lifestyle could be involved in grapevine bud necrosis. *Scientific Reports* 10:9514. DOI: 10.1038/s41598-020-66500-0.

Rice, E. S. and Green, R. E. (2019). New Approaches for Genome Assembly and Scaffolding. *Annual Review of Animal Biosciences* 7:21–1–21.24. DOI: 10.1146/annurev-animal-020518-115344.

Rich, M. K. et al. (2021). Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science* 372:864–868. DOI: 10.1126/science.abg0929.

Richards, S. (2018). Full disclosure: Genome assembly is still hard. *PLoS Biology* 16(4):e2005894. DOI: 10.1371/journal.pbio.2005894.

Rillig, M. C. et al. (2019). The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* 366(6467):886–890. DOI: 10.1126/science.aay2832.

Ristaino, J. B. (2020). The Importance of Mycological and Plant Herbaria in Tracking Plant Killers. *Frontiers in Ecology and Evolution* 7:521. DOI: 10.3389/fevo.2019.00521.

Robinson, G. E. et al. (2011). Creating a Buzz About Insect Genomes. *Science* 331(6023):1386.

Rodriguez, R. J., White Jr, J. F. et al. (2009). Fungal endophytes: diversity and functional roles. *New Phytologist* 182:314–330. DOI: 10.1111/j.1469-8137.2009.02773.x.

Rodriguez, R. J., Redman, R. S. and Henson, J. M. (2004). The role of fungal symbioses in the adaptation of plants to high stress environments. *Mitigation and Adaptation Strategies for Global Change* 9:261–272. DOI: 10.1023/B:MITI.0000029922.31110.97.

Rodríguez-Gálvez, E. et al. (2017). Phylogeny and pathogenicity of *Lasiodiplodia* species associated with dieback of mango in Peru. *Fungal Biology* 121(4):452–465. DOI: 10.1016/j.funbio.2016.06.004.

Rojas, E. I. et al. (2010). *Colletotrichum gloeosporioides* s.l. associated with *Theobroma cacao* and other plants in Panamá: multilocus phylogenies distinguish host-associated pathogens from asymptomatic endophytes. *Mycologia* 102(6):1318–1338. DOI: 10.3852/09-244.

Rosado, A. W. C. et al. (2016). Phylogeny, identification, and pathogenicity of *Lasiodiplodia* associated with postharvest stem-end rot of coconut in Brazil. *Plant Disease* 100(3):561–568. DOI: 10.1094/PDIS-03-15-0242-RE.

Rouard, M. et al. (2018). Three new genome assemblies support a rapid radiation in *Musa acuminata* (Wild Banana). *Genome Biology and Evolution* 10(12):3129–3140. DOI: 10.1093/gbe/evy227.

RStudio Team (2015). *RStudio: Integrated Development for R*. Boston MA, US. URL: http://www.rstudio.com/.

Ruan, J. and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods* 17(2):155–158. DOI: 10.1038/s41592-019-0669-3.

Ruas, M. et al. (2017). MGIS: managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data. *Database* 2017:bax046. DOI: 10.1093/database/bax046.

Rubini, M. R. et al. (2005). Diversity of endophytic fungal community of cacao (*Theobroma cacao* L.) and biological control of *Crinipellis perniciosa*, causal agent of Witches' Broom Disease. *International Journal of Biological Sciences* 1:24–33. DOI: 10.7150/ijbs.1.24.

Rudgers, J. A. et al. (2012). There are many ways to be a mutualist: Endophytic fungus reduces plant survival but increases population growth. *Ecology* 93(3):565–574. DOI: 10.1890/11-0689.1.

Ruete, A., Snäll, T. and Jönsson, M. (2016). Dynamic anthropogenic edge effects on the distribution and diversity of fungi in fragmented old-growth forests. *Ecological Applications* 26(5):1475–1485. DOI: 10.1890/15-1271.

Safonova, Y., Bankevich, A. and Pevzner, P. A. (2015). dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *Journal of Computational Biology* 22(6):528–545. DOI: 10.1089/cmb.2014.0153.

Saikkonen, K., Faeth, S. H. et al. (1998). Fungal Endophytes: A Continuum of Interactions with Host Plants. *Annual Review of Ecology and Systematics* 29:319–343. DOI: 10.1146/annurev.ecolsys.29.1.319.

Saikkonen, K., Young, C. A. et al. (2016). Endophytic *Epichloë* species and their grass hosts: from evolution to applications. *Plant Molecular Biology* 90:665–675. DOI: 10.1007/s11103-015-0399-6.

Sakuno, E., Yabe, K. and Nakajima, H. (2003). Involvement of Two Cytosolic Enzymes and a Novel Intermediate, 5'-Oxoaverantin, in the Pathway from 5'-Hydroxyaverantin to Averufin in Aflatoxin Biosynthesis. *Applied and Environmental Microbiology* 69(11):6418–6426. DOI: 10.1128/AEM.69.11.6418-6426.2003.

Salvatore, M. M. and Andolfi, A. (2020). The Thin Line between Pathogenicity and Endophytism: The Case of *Lasiodiplodia theobromae*. *Agriculture* 10:488. DOI: doi:10.3390/agriculture10100488.

Sánchez-Vallet, A. et al. (2018). The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology* 56:2.1–2.20. DOI: 10.1146/annurev-phyto-080516-035303.

Sandoval-Denis, M., Guarnaccia, V. et al. (2018). Symptomatic citrus trees reveal a new pathogenic lineage in *Fusarium* and two new *Neocosmospora* species. *Persoonia* 40:1–25. DOI: 10.3767/persoonia.2018.40.01.

Sandoval-Denis, M., Lombard, L. and Crous, P. W. (2019). Back to the roots: a reappraisal of *Neocosmospora*. *Persoonia* 43:90–185. DOI: 10.3767/persoonia.2019.43.04.

Sangeetha, G., Anandan, A. and Rani, S. U. (2012). Morphological and molecular characterisation of *Lasiodiplodia theobromae* from various banana cultivars causing crown rot disease in fruits. *Archives of Phytopathology and Plant Protection* 45(4):475–486. DOI: 10.1080/03235408.2011.587986.

Sarmiento, C. et al. (2017). Soilborne fungi have host affinity and host-specific effects on seed germination and survival in a lowland tropical forest. *PNAS* 114(43):11458–11463. DOI: 10.1073/pnas.1706324114.

Sato, M. P. et al. (2019). Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Research* 26(5):391–398. DOI: 10.1093/dnares/dsz017.

Sauder, D. C. and DeMars, C. E. (2019). An Updated Recommendation for Multiple Comparisons. *Advances in Methods and Practices in Psychological Science* 2(1):26–44. DOI: 10.1177/2515245918808784.

Sauquet, H. et al. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Systematic Biology* 61(2):289–313. DOI: 10.1093/sysbio/syr116.

Sayyari, E. and Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular biology and evolution* 33(7):1654–1668. DOI: 10.1093/molbev/msw079.

Sboner, A. et al. (2011). The real cost of sequencing: higher than you think! *Genome Biology* 12:125. DOI: 10.1186/gb-2011-12-8-125.

Scarpari, M. et al. (2020). *Didymella corylicola* sp. nov., a new fungus associated with hazelnut fruit development in Italy. *Mycological Progress* 19:317–328. DOI: 10.1007/s11557-020-01562-y.

Schardl, C. L. (1996). *Epichloë* species: fungal symbionts of grasses. *Annual Review of Phytopathology* 34(67):109–130. DOI: 10.1146/annurev.phyto.34.1.109.

Schenk, J. J. (2016). Consequences of Secondary Calibrations on Divergence Time Estimates. *PLoS ONE* 11(1):e0148228. DOI: 10.1371/journal.pone.0148228.

Schirrmann, M. K. et al. (2018). Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086. DOI: 10.1111/mec.14585.

Schliep, K. et al. (2017). Intertwining phylogenetic trees and networks. *Methods in Ecology and Evolution* 8:1212–1220. DOI: 10.1111/2041-210X.12760.

Schloss, P. D. (2021). Amplicon Sequence Variants Artificially Split Bacterial Genomes into Separate Clusters. *mSphere* 6(4):e00191–21. DOI: 10.1128/msphere.00191-21.

Schmitz, A. M., Pawlowska, T. E. and Harrison, M. J. (2019). A short LysM protein with high molecular diversity from an arbuscular mycorrhizal fungus, *Rhizophagus irregularis*. *Mycoscience* 60(1):63–70. DOI: 10.1016/j.myc.2018.09.002.

Schroers, H. J. et al. (2011). A revision of *Cyanonectria* and *Geejayessia* gen. nov., and related species with *Fusarium*-like anamorphs. *Studies in Mycology* 68:115–138. DOI: 10.3114/sim.2011.68.05.

Schulz, A. (2021). *pBrackets: Plot Brackets*. URL: https://cran.r-project.org/package=pBrackets.

Schulz, B. and Boyle, C. (2005). The endophytic continuum. *Mycological Research* 109(6):661–686. DOI: 10.1017/S095375620500273X.

Schulz, B., Boyle, C. et al. (2002). Endophytic fungi: a source of novel biologically active secondary metabolites. *Mycological Research* 106(9):996–1004. DOI: 10.1186/1471-2261-9-19.

Šečić, E. et al. (2021). A novel plant-fungal association reveals fundamental sRNA and gene expression reprogramming at the onset of symbiosis. *BMC Biology* 19:171. DOI: 10.1186/s12915-021-01104-2.

Seidl, M. F., Faino, L. et al. (2015). The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Molecular Plant-Microbe Interactions* 28(3):362–373. DOI: 10.1094/MPMI-06-14-0173-R.

Seidl, M. F., Kramer, H. M. et al. (2020). Repetitive Elements Contribute to the Diversity and Evolution of Centromeres in the Fungal Genus *Verticillium*. *mBio* 11:e01714–20. DOI: 10.1128/mBio.01714-20.

Selosse, M. A., Petrolli, R. et al. (2022). The Waiting Room Hypothesis revisited by orchids: were orchid mycorrhizal fungi recruited among root endophytes? *Annals of Botany* 129(3):259–270. DOI: 10.1093/aob/mcab134.

Selosse, M. A., Schneider-Maunoury, L. and Martos, F. (2018). Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytologist* 217:968–972. DOI: 10.1111/nph.14983.

Semchenko, M. et al. (2022). Deciphering the role of specialist and generalist plant–microbial interactions as drivers of plant–soil feedback. *New Phytologist* 234(6):1929–1944. DOI: 10.1111/nph.18118.

Shaffer, J. P. et al. (2018). Context-dependent and variable effects of endohyphal bacteria on interactions between fungi and seeds. *Fungal Ecology* 36:117–127. DOI: 10.1016/j.funeco.2018.08.008.

Shao, X. et al. (2019). Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Medical Genetics* 20:175. DOI: 10.1186/s12881-019-0909-5.

Sharp, P. M. and Li, W.-H. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* 15(3):1281–1295. DOI: 10.1093/nar/15.3.1281.

Shaul, S. and Graur, D. (2002). Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene* 300:59–61. DOI: 10.1016/S0378-1119(02)00851-X.

Shearin, Z. R. et al. (2018). Fungal endophytes from seeds of invasive, non-native *Phragmites australis* and their potential role in germination and seedling growth. *Plant and Soil* 422:183–194. DOI: 10.1007/s11104-017-3241-x.

Shen, Q., Liu, Y. and Naqvi, N. I. (2018). Fungal effectors at the crossroads of phytohormone signaling. *Current Opinion in Microbiology* 46:1–6. DOI: 10.1016/j.mib.2018.01.006.

Shen, X.-X., Steenwyk, J. L. et al. (2020). Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Science Advances* 6:eabd0079. DOI: 10.1126/SCIADV.ABD0079.

Shendure, J. et al. (2017). DNA sequencing at 40: past, present and future. *Nature* 550:345. DOI: 10.1038/nature24286.

Shivas, R. G. and Cai, L. (2012). Cryptic fungal species unmasked. *Microbiology Australia* 33(1):36–37. DOI: 10.1071/ma12036.

Signorell, A. (2020). *DescTools: Tools for Descriptive Statistics*. URL: https://cran.r-project.org/package=DescTools.

Sikes, B. A. et al. (2018). Import volumes and biosecurity interventions shape the arrival rate of fungal pathogens. *PLoS Biology* 16(5):e2006025. DOI: 10.1371/journal.pbio.2006025.

Sikora, R. A. et al. (2008). Mutualistic endophytic fungi and *in-planta* suppressiveness to plant parasitic nematodes. *Biological Control* 46:15–23. DOI: 10.1016/j.biocontrol.2008.02.011.

Silva, J. L., Silva, W. F. et al. (2021). First Report of *Colletotrichum tropicale* Causing Anthracnose on *Passiflora edulis* in Brazil. *Plant Disease* 105(11):3761. DOI: 10.1094/PDIS-07-20-1440-PDN.

Silva, N. I. de, Phillips, A. J. et al. (2019). Phylogeny and morphology of *Lasiodiplodia* species associated with *Magnolia* forest plants. *Scientific Reports* 9:14355. DOI: 10.1038/s41598-019-50804-x.

Simão, F. A. et al. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212. DOI: 10.1093/bioinformatics/btv351.

Simpson, J. T. et al. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research* 19:1117–1123. DOI: 10.1101/gr.089532.108..

Singh, Y., Nair, A. M. and Verma, P. K. (2021). Surviving the odds: From perception to survival of fungal phytopathogens under host-generated oxidative burst. *Plant Communications* 2:100142. DOI: 10.1016/j.xplc.2021.100142.

Šišić, A. et al. (2018). The 'forma specialis' issue in *Fusarium*: A case study in *Fusarium solani* f. sp. *pisi*. *Scientific Reports* 8:1252. DOI: 10.1038/s41598-018-19779-z.

Skelton, J., Cauvin, A. and Hunter, M. E. (2022). Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. *Environmental DNA*. DOI: 10.1002/edn3.355.

Skillings, D. (2016). Holobionts and the ecology of organisms: Multi-species communities or integrated individuals? *Biology and Philosophy* 31(6):875–892. DOI: 10.1007/s10539-016-9544-0.

Skole, D. and Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science* 260(5116):1905–1910. DOI: 10.1126/science.260.5116.1905.

Slippers, B. and Wingfield, M. J. (2007). Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biology Reviews* 21:90–106. DOI: 10.1016/j.fbr.2007.06.002.

Slowikowski, K. (2020). *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*. URL: https://cran.r-project.org/package=ggrepel.

Smit, A. and Hubley, R. (2015). *RepeatModeler Open-1.0*. URL: http://www.repeatmasker.org.

Smit, A., Hubley, R. and Green, P. (2015). *RepeatMasker Open-4.0*. URL: http://www.repeatmasker.org.

Smith, D., Kermode, A. et al. (2020). Strengthening mycology research through coordinated access to microbial culture collection strains. *CABI Agriculture and Bioscience* 1:2. DOI: 10.1186/s43170-020-00004-9.

Smith, D., Ryan, M. J. and Caine, T. (2022). 'Contribution of CABI and culture collections to a sustainable future through the utilisation of microbial genetic resources'. In: *Importance of Microbiology Teaching and Microbial Resource Management for Sustainable Futures*. Ed. by I. Kurtböke. Elsevier Inc. Chap. 9:pp. 229–273. ISBN: 9780128182727. DOI: 10.1016/b978-0-12-818272-7.00010-9.

Smith, M. D., Wertheim, J. O. et al. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution* 32(5):1342–1353. DOI: 10.1093/molbev/msv022.

Smith, R. L., Sawbridge, T. et al. (2020). Rediscovering an old foe: Optimised molecular methods for DNA extraction and sequencing applications for fungarium specimens of powdery mildew (Erysiphales). *PLoS ONE* 15(5):e0232535. DOI: 10.1371/journal.pone.0232535.

Smith, S., Brown, J. and Walker, J. (2018). So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* 13(5):e0197433. DOI: 10.1371/journal.pone.0197433May.

Somjaipeng, S. et al. (2015). Isolation, identification, and ecology of growth and taxol production by an endophytic strain of *Paraconiothyrium variabile* from English yew trees (*Taxus baccata*). *Fungal Biology* 119:1022–1031. DOI: 10.1016/j.funbio.2015.07.007.

Sonah, H., Deshmukh, R. K. and Bélanger, R. R. (2016). Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. *Frontiers in Plant Science* 7:126. DOI: 10.3389/fpls.2016.00126.

Song, L. et al. (2021). First Report of Round Leaf Spot on *Sophora tonkinensis* Caused by *Didymella glomerata* in China. *Plant Disease* 105:498. DOI: 10.1094/PDIS-06-20-1239-PDN.

Spanu, P. D. et al. (2010). Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. *Science* 330:1543–1546. DOI: 10.1126/science.1194573.

Spatafora, J. W. et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108(5):1028–1046. DOI: 10.3852/16-042.

Sperschneider, J. and Dodds, P. N. (2021). EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *MPMI* 35(2):146–156. DOI: 10.1094/MPMI-08-21-0201-R.

Stalder, L. et al. (2022). The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology*. DOI: 10.1111/mec.16435.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312. DOI: 10.1093/bioinformatics/btu033.

Stanke, M. et al. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* 34:W435–W439. DOI: 10.1093/nar/gkl200.

Stauber, L., Prospero, S. and Croll, D. (2020). Comparative Genomics Analyses of Lifestyle Transitions at the Origin of an Invasive Fungal Pathogen in the Genus *Cryphonectria*. *mSphere* 5(5):e00737–20. DOI: 10.1128/msphere.00737-20.

Steenwyk, J. L., Buida, T. J. et al. (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology* 18(12):e3001007. DOI: 10.1371/journal.pbio.3001007.

Steenwyk, J. L. and Rokas, A. (2018). Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Frontiers in Microbiology* 9:288. DOI: 10.3389/fmicb.2018.00288.

Steenwyk, J. L., Shen, X.-X. et al. (2019). A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. *mBio* 10(4):e00925–19. DOI: 10.1128/mBio.00925-19.

Steenwyk, J. L., Soghigian, J. S. et al. (2016). Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics* 17:700. DOI: 10.1186/s12864-016-3044-0.

Stergiopoulos, I. and de Wit, P. J. (2009). Fungal Effector Proteins. *Annual Review of Phytopathology* 47:233–263. DOI: 10.1146/annurev.phyto.112408.132637.

Stergiopoulos, I., Kourmpetis, Y. A. I. et al. (2012). In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Molecular Biology and Evolution* 29(11):3371–3384. DOI: 10.1093/molbev/mss143.

Stone, J. K., Bacon, C. W. and White, J. J. F. (2000). 'An Overview of Endophytic Microbes: Endophytism Defined'. In: *Microbial Endophytes*. Ed. by C. W. Bacon and J. J. F. White. 1st ed. January. New York: Marcel Dekker. Chap. 1:pp. 3–29. DOI: 10.1163/_q3_SIM_00374.

Stranger, B. E. et al. (2007). Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science* 315(5813):848–853. DOI: 10.1126/science.1136678.

Stranska, M. et al. (2022). Fungal Endophytes of *Vitis vinifera*–Plant Growth Promoters or Potentially Toxinogenic Agents? *Toxins* 14:66. DOI: 10.3390/toxins14020066.

Strullu-Derrien, C. et al. (2018). The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytologist* 220:1012–1030. DOI: 10.1111/nph.15076.

Studholme, D. J. (2016). Genome Update. Let the consumer beware: *Streptomyces* genome sequence quality. *Microbial Biotechnology* 9:3–7. DOI: 10.1111/1751-7915.12344.

Subrmaniyan, R., Ponnambalam, S. and Thirunavukarassu, T. (2018). Inter species variations in cultivable endophytic fungal diversity among the tropical seagrasses. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences* 88(3):849–857. DOI: 10.1007/s40011-016-0817-9.

Summerell, B. A. (2019). Resolving *Fusarium*: Current Status of the Genus. *Annual Review of Phytopathology* 57:323–339. DOI: 10.1146/annurev-phyto-082718-100204.

Suyama, M., Torrents, D. and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34:W609–W612. DOI: 10.1093/nar/gkl315.

Suzek, B. E. et al. (2015). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31(6):926–932. DOI: 10.1093/bioinformatics/btu739.

Swett, C. L. and Gordon, T. R. (2015). Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecology* 13:120–129. DOI: 10.1016/j.funeco.2014.09.003.

Tadych, M., Bergen, M. S. and White Jr., J. F. (2014). *Epichloë* spp. associated with grasses: new insights on life cycles, dissemination and evolution. *Mycologia* 106(2):181–201. DOI: 10.3852/106.2.181.

Talhinhas, P. and Baroncelli, R. (2021). *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. *Fungal Diversity* 110:109–198. DOI: 10.1007/s13225-021-00491-9.

Talhinhas, P., Tavares, D. et al. (2017). Validation of standards suitable for genome size estimation of fungi. *Journal of Microbiological Methods* 142:76–78. DOI: 10.1016/j.mimet.2017.09.012.

Tamura, R. et al. (2008). Requirement for particular seed-borne fungi for seed germination and seedling growth of *Xyris complanata*, a pioneer monocot in topsoil-lost tropical peatland in Central Kalimantan, Indonesia. *Ecological Research* 23(3):573–579. DOI: 10.1007/s11284-007-0411-y.

Tamuri, A. U. and Dos Reis, M. (2021). A Mutation-Selection Model of Protein Evolution under Persistent Positive Selection. *Molecular Biology and Evolution* 39(1):msab309. DOI: 10.1093/molbev/msab309.

Tan, G. et al. (2015). Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Systematic Biology* 64(5):778–791. DOI: 10.1093/sysbio/syv033.

Tao, Q. et al. (2019). A machine learning method for detecting autocorrelation of evolutionary rates in large phylogenies. *Molecular Biology and Evolution* 36(4):811–824. DOI: 10.1093/molbev/msz014.

Tavares, S. et al. (2014). Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science* 5:422. DOI: 10.3389/fpls.2014.00422.

Taylor, J. W. and Berbee, M. L. (2006). Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98(6):838–849. DOI: 10.1080/15572536.2006.11832614.

Taylor, T. N. and Krings, M. (2005). Fossil microorganisms and land plants: Associations and interactions. *Symbiosis* 40:119–135. DOI: 10.1126/science.1175102.

Tedersoo, L., Albertsen, M. et al. (2021). Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology. *Applied Environmental Microbiology* 87(17):e00626–21. DOI: 10.1128/AEM.00626-21.

Tedersoo, L., Bahram, M., Põlme, S. et al. (2014). Global diversity and geography of soil fungi. *Science* 346(6213):1256688. DOI: 10.1126/science.1256688.

Tedersoo, L., Bahram, M., Zinger, L. et al. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology* 31:2769–2795. DOI: 10.1111/mec.16460.

The Darwin Tree of Life Project Consortium (2022). Sequence locally, think globally: The Darwin Tree of Life Project. *PNAS* 119(4):e2115642118. DOI: 10.1073/pnas.2115642118.

Thomas, A. C. et al. (2016). Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources* 16:714–726. DOI: 10.1111/1755-0998.12490.

Tintjer, T., Leuchtmann, A. and Clay, K. (2008). Variation in horizontal and vertical transmission of the endophyte *Epichloë elymi* infecting the grass *Elymus hystrix*. *New Phytologist* 179:236–246. DOI: 10.1111/j.1469-8137.2008.02441.x.

Tisthammer, K. H., Cobian, G. M. and Amend, A. S. (2016). Global biogeography of marine fungi is shaped by the environment. *Fungal Ecology* 19:39–46. DOI: 10.1016/j.funeco.2015.09.003.

Tivoli, B. and Banniza, S. (2007). Comparison of the epidemiology of ascochyta blights on grain legumes. *European Journal of Plant Pathology* 119:59–76. DOI: 10.1007/s10658-007-9117-9.

Tørresen, O. K. et al. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic acids research* 47(21):10994–11006. DOI: 10.1093/nar/gkz841.

Toruño, T. Y., Stergiopoulos, I. and Coaker, G. (2016). Plant-Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annual Review of Phytopathology* 54:419–441. DOI: 10.1146/annurev-phyto-080615-100204.Plant-Pathogen.

Tyler, A. D. et al. (2016). Comparison of sample preparation methods used for the next-generation sequencing of *Mycobacterium tuberculosis*. *PLoS ONE* 11(2):e0148676. DOI: 10.1371/journal.pone.0148676.

U'Ren, J. M., Lutzoni, F., Miadlikowska, J., Zimmerman, N. B. et al. (2019). Host availability drives distributions of fungal endophytes in the imperilled boreal realm. *Nature Ecology & Evolution* 3:1430–1437. DOI: 10.1038/s41559-019-0975-2.

U'Ren, J. M. and Arnold, A. E. (2016). Diversity, taxonomic composition, and functional aspects of fungal communities in living, senesced, and fallen leaves at five sites across North America. *PeerJ* 4:e2768. DOI: 10.7717/peerj.2768.

U'Ren, J. M., Dalling, J. W. et al. (2009). Diversity and evolutionary origins of fungi associated with seeds of a neotropical pioneer tree: a case study for analysing fungal environmental samples. *Mycological Research* 113(4):432–449. DOI: 10.1016/j.mycres.2008.11.015.

U'Ren, J. M., Lutzoni, F., Miadlikowska, J. and Arnold, A. E. (2010). Community Analysis Reveals Close Affinities Between Endophytic and Endolichenic Fungi in Mosses and Lichens. *Microbial Ecology* 60(2):340–353. DOI: 10.1007/s00248-010-9698-2.

Uranga, C. C. et al. (2016). Fatty acid esters produced by *Lasiodiplodia theobromae* function as growth regulators in tobacco seedlings. *Biochemical and Biophysical Research Communications* 472(2):339–345. DOI: 10.1016/j.bbrc.2016.02.104.

Urban, M. et al. (2020). PHI-base: the pathogen-host interactions database. *Nucleic Acids Research* 48:D613–D620. DOI: 10.1093/nar/gkz904.

Utturkar, S. M. et al. (2014). Evaluation and validation of *de novo* and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics* 30(19):2709–2716. DOI: 10.1093/bioinformatics/btu391.

Vági, P. et al. (2014). Simultaneous specific *in planta* visualization of root-colonizing fungi using fluorescence *in situ* hybridization (FISH). *Mycorrhiza* 24:259–266. DOI: 10.1007/s00572-013-0533-8.

Vaidya, G., Lohman, D. J. and Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171–180. DOI: 10.1111/j.1096-0031.2010.00329.x.

Valenzuela-Lopez, N., Cano-Lira, J. F., Guarro, J. et al. (2018). Coelomycetous *Dothideomycetes* with emphasis on the families *Cucurbitariaceae* and *Didymellaceae*. *Studies in Mycology* 90:1–69. DOI: 10.1016/j.simyco.2017.11.003.

Valenzuela-Lopez, N., Cano-Lira, J. F., Stchigel, A. M. et al. (2019). *Neocucurbitaria keratinophila*: An emerging opportunistic fungus causing superficial mycosis in Spain. *Medical Mycology* 57:733–738. DOI: 10.1093/mmy/myy132.

van Dam, P., Fokkens, L. et al. (2016). Effector profiles distinguish *formae speciales* of *Fusarium oxysporum*. *Environmental Microbiology* 18(11):4087–4102. DOI: 10.1111/1462-2920.13445.

van Dam, P. and Rep, M. (2017). The Distribution of *Miniature Impala* Elements and *SIX* Genes in the *Fusarium* Genus is Suggestive of Horizontal Gene Transfer. *Journal of Molecular Evolution* 85:14–25. DOI: 10.1007/s00239-017-9801-0.

van der Does, H. C. and Rep, M. (2017). Adaptation to the Host Environment by Plant-Pathogenic Fungi. *Annual Review of Phytopathology* 55:427–450. DOI: 10.1146/annurev-phyto-080516-035551.

van der Heijden, M. G. et al. (2015). Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist* 205:1406–1423. DOI: 10.1111/nph.13288.

Van Hove, F. et al. (2011). *Gibberella musae* (*Fusarium musae*) sp. nov., a recently discovered species from banana is sister to *F. verticillioides*. *Mycologia* 103(3):570–585. DOI: 10.3852/10-038.

Vannini, A. et al. (2017). Does *Gnomoniopsis castanea* contribute to the natural biological control of chestnut gall wasp? *Fungal Biology* 121:44–52. DOI: 10.1016/j.funbio.2016.08.013.

Varga, T. et al. (2019). Megaphylogeny resolves global patterns of mushroom evolution. *Nature Ecology & Evolution* 3:668–678. DOI: 10.1038/s41559-019-0834-1.

Vaser, R. and Šikić, M. (2021). Time- and memory-efficient genome assembly with Raven. *Nature Computational Science* 1:332–336. DOI: 10.1038/s43588-021-00073-4.

Vaz, A. B. et al. (2018). A multiscale study of fungal endophyte communities of the foliar endosphere of native rubber trees in Eastern Amazon. *Scientific Reports* 8:16151. DOI: 10.1038/s41598-018-34619-w.

Vega, F. E. (2018). The use of fungal entomopathogens as endophytes in biological control: a review. *Mycologia* 110(1):4–30. DOI: 10.1080/00275514.2017.1418578.

Veneault-Fourrey, C., Commun, C. et al. (2014). Genomic and transcriptomic analysis of *Laccaria bicolor* CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. *Fungal Genetics and Biology* 72:168–181. DOI: 10.1016/j.fgb.2014.08.007.

Veneault-Fourrey, C. and Martin, F. (2011). Mutualistic interactions on a knife-edge between saprotrophy and pathogenesis. *Current Opinion in Plant Biology* 14:444–450. DOI: 10.1016/j.pbi.2011.03.022.

Vidal, S. and Jaber, L. R. (2015). Entomopathogenic fungi as endophytes: plant-endophyte-herbivore interactions and prospects for use in biological control. *Current Science* 109(1):46–54. ISSN: 00113891. URL: http://www.jstor.org/stable/24905690.

Vieira, W. A. d. S. et al. (2020). Optimal markers for the identification of *Colletotrichum* species. *Molecular Phylogenetics and Evolution* 143:106694. DOI: 10.1016/j.ympev.2019.106694.

Vilhar, B. et al. (2001). Plant genome size measurement with DNA image cytometry. *Annals of Botany* 87(6):719–728. DOI: 10.1006/anbo.2001.1394.

Villacorta, P. J. (2015). *ART: Aligned Rank Transform for Nonparametric Factorial Analysis*. URL: https://cran.r-project.org/package=ART.

Voorhies, M. et al. (2022). Chromosome-Level Genome Assembly of a Human Fungal Pathogen Reveals Synteny among Geographically Distinct Species. *mBio* 13(1):e02574–21. DOI: 10.1128/MBIO.02574-21.

Vu, D. et al. (2019). Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Studies in Mycology* 92:135–154. DOI: 10.1016/j.simyco.2018.05.001.

Walker, B. J. 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.

Waller, F. et al. (2005). The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. *PNAS* 102(38):13386–13391. DOI: 10.1073/pnas.0504423102.

Wanasinghe, D. N., Phookamsak, R. et al. (2017). A family level rDNA based phylogeny of Cucurbitariaceae and Fenestellaceae with descriptions of new *Fenestella* species and *Neocucurbitaria* gen. nov. *Mycosphere* 8(4):397–414. DOI: 10.5943/mycosphere/8/4/2.

Wanasinghe, D. N. and Mortimer, P. E. (2022). Taxonomic and Phylogenetic Insights into Novel *Ascomycota* from Forest Woody Litter. *Biology* 11:889. DOI: 10.3390/biology11060889.

Wang, M. M., Chen, Q. et al. (2019). *Fusarium incarnatum-equiseti* complex from China. *Persoonia* 43:70–89. DOI: 10.3767/persoonia.2019.43.03.

Wang, X. W., Han, P. J. et al. (2022). Taxonomy, phylogeny and identification of *Chaetomiaceae* with emphasis on thermophilic species. *Studies in Mycology* 101:121–243. DOI: 10.3114/sim.2022.101.03.

Wang, X. W., Houbraken, J. et al. (2016). Diversity and taxonomy of *Chaetomium* and chaetomium-like fungi from indoor environments. *Studies in Mycology* 84:145–224. DOI: 10.1016/j.simyco.2016.11.005.

Wang, X. J., Min, C. L. et al. (2014). An endophytic sanguinarine-producing fungus from *Macleaya cordata*, *Fusarium proliferatum* BLH51. *Current Microbiology* 68(3):336–341. DOI: 10.1007/s00284-013-0482-7.

Wang, Y., Li, H. et al. (2017). Biodegradation of diuron by an endophytic fungus *Neurospora intermedia* DP8-1 isolated from sugarcane and its potential for remediating diuron-contaminated soils. *PLoS ONE* 12(8):e0182556. DOI: 10.1371/journal.pone.0182556.

Wapinski, I. et al. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449:54–61. DOI: 10.1038/nature06107.

Wassermann, B. et al. (2019). Seeds of native alpine plants host unique microbial communities embedded in cross-kingdom networks. *Microbiome* 7:108. DOI: 10.1186/s40168-019-0723-5.

Watson, M. and Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nature Biotechnology* 37(2):124–126. DOI: 10.1038/s41587-018-0004-z.

Weisenfeld, N. I. et al. (2014). Comprehensive variation discovery in single human genomes. *Nature Genetics* 46(12):1350–1355. DOI: 10.1038/ng.3121.

Westengen, O. T., Jeppson, S. and Guarino, L. (2013). Global *Ex-Situ* Crop Diversity Conservation and the Svalbard Global Seed Vault: Assessing the Current Status. *PLoS ONE* 8(5):e64146. DOI: 10.1371/journal.pone.0064146.

Whelan, S. and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* 18(5):691–699. DOI: 10.1093/oxfordjournals.molbev.a003851.

Wick, R. R. et al. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology* 13(6):e1005595. DOI: 10.1371/journal.pcbi.1005595.

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software* 21(12):1–20. DOI: 10.18637/jss.v021.i12.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40(1):1–29. DOI: 10.18637/jss.v040.i01.

– (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York. URL: https://ggplot2.tidyverse.org.

– (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. URL: https://cran.r-project.org/package=stringr.

– (2020). *rvest: Easily Harvest (Scrape) Web Pages*. URL: https://cran.r-project.org/package=rvest.

Wickham, H., Averick, M. et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software* 4(43):1686. DOI: 10.21105/joss.01686.

Wickham, H., François, R. et al. (2020). *dplyr: A Grammar of Data Manipulation*. URL: https://cran.r-project.org/package=dplyr.

Wickham, H. and Girlich, M. (2022). *tidyr: Tidy Messy Data*. URL: https://github.com/tidyverse/tidyr.

Wickham, H. and Pedersen, T. L. (2019). *gtable: Arrange 'Grobs' in Tables*. URL: https://cran.r-project.org/package=gtable.

Wickham, H. and Seidel, D. (2020). *scales: Scale Functions for Visualization*. URL: https://cran.r-project.org/package=scales.

Wiewióra, B., Żurek, G. and Pañka, D. (2015). Is the vertical transmission of *Neotyphodium lolii* in perennial ryegrass the only possible way to the spread of endophytes? *PLoS ONE* 10(2):e0117231. DOI: 10.1371/journal.pone.0117231.

Wijayawardene, N. N. et al. (2018). Outline of *Ascomycota*: 2017. *Fungal Diversity* 88:167–263. DOI: 10.1007/s13225-018-0394-8.

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. URL: https://cran.r-project.org/package=cowplot.

Willis, K. J., ed. (2018). *State of the World's Fungi 2018*. Royal Botanic Gardens, Kew. ISBN: 978-1-84246-678-0.

Wolinska, K. W. et al. (2021). Tryptophan metabolism and bacterial commensals prevent fungal dysbiosis in *Arabidopsis* roots. *PNAS* 118(49):e2111521118. DOI: 10.1073/pnas.2111521118.

Woo, S. L. et al. (2014). *Trichoderma*-based Products and their Widespread Use in Agriculture. *The Open Mycology Journal* 8:71–126. DOI: 10.2174/1874437001408010071.

Wu, C. H. et al. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research* 32:D112–D114. DOI: 10.1093/nar/gkh097.

Xiao, C. L. et al. (2017). MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nature Methods* 14(11):1072–1074. DOI: 10.1038/nmeth.4432.

Xu, J. et al. (2020). eCAMI: simultaneous classification and motif identification for enzyme annotation. *Bioinformatics* 36(7):2068–2075. DOI: 10.1093/bioinformatics/btz908.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24(8):1586–1591. DOI: 10.1093/molbev/msm088.

Yang, Z. and Rannala, B. (2006). Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. *Molecular Biology and Evolution* 23(1):212–226. DOI: 10.1093/molbev/msj024.

Ye, C. et al. (2016). DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Scientific Reports* 6:31900. DOI: 10.1038/srep31900. arXiv: 1410.2801.

Yin, Y. et al. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40:W445–W451. DOI: 10.1093/nar/gks479.

Yoshida, M. aki et al. (2011). Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. *Gene* 483:63–71. DOI: 10.1016/j.gene.2011.05.027.

Yu, G. (2021). *ggplotify: Convert Plot to 'grob' or 'ggplot' Object*. URL: https://cran.r-project.org/package=ggplotify.

Yu, G. et al. (2017). GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8:28–36. DOI: 10.1111/2041-210X.12628.

Zakaria, L. and Aziz, W. N. W. (2018). Molecular identification of endophytic fungi from banana leaves (*Musa* spp.) *Tropical Life Sciences Research* 29(2):201–211. DOI: 10.21315/tlsr2018.29.2.14.

Zakaria, L., Izham, M. et al. (2016). Molecular Characterisation of Endophytic Fungi from Roots of Wild Banana (*Musa acuminata*). *Tropical Life Sciences Research* 27(1):153–162. DOI: 10.21315/tlsr2018.29.2.14.

Zeng, T. et al. (2019). A lysin motif effector subverts chitin-triggered immunity to facilitate arbuscular mycorrhizal symbiosis. *New Phytologist* 225:448–460. DOI: 10.1111/nph.16245.

Zerbino, D. R. and Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18:821–829. DOI: 10.1101/gr.074492.107.

Zhang, C., Rabiee, M. et al. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19(Suppl 6):153. DOI: 10.1186/s12859-018-2129-y.

Zhang, C., Scornavacca, C. et al. (2020). ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy. *Molecular Biology and Evolution* 37(11):3292–3307. DOI: 10.1093/molbev/msaa139.

Zhang, H., Yohe, T. et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 46:W95–W101. DOI: 10.1093/nar/gky418.

Zhang, N., O'Donnell, K. et al. (2006). Members of the *Fusarium solani* species complex that cause infections in both humans and plants are common in the environment. *Journal of Clinical Microbiology* 44(6):2185–2190. DOI: 10.1128/JCM.00120-06.

Zhang, T., Zhao, Y. L. et al. (2016). Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. *Nature Plants* 2:16153. DOI: 10.1038/nplants.2016.153.

Zhang, W., Card, S. D. et al. (2017). Defining the pathways of symbiotic *Epichloë* colonization in grass embryos with confocal microscopy. *Mycologia* 109(1):153–161. DOI: 10.1080/00275514.2016.1277469.

Zhang, W., Chen, J. et al. (2011). A practical comparison of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS ONE* 6(3):e17915. DOI: 10.1371/journal.pone.0017915.

Zhang, Y., Zhang, J. et al. (2014). Neotypification and phylogeny of *Kalmusia*. *Phytotaxa* 176(1):164–173. DOI: 10.11646/phytotaxa.176.1.16.

Zhang, Y., Yang, H. et al. (2020). The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Communications Biology* 3:50. DOI: 10.1038/s42003-020-0770-2.

Zhao, S. and Gibbons, J. G. (2018). A population genomic characterization of copy number variation in the opportunistic fungal pathogen *Aspergillus fumigatus*. *PLoS ONE* 13(8):e0201611. DOI: 10.1371/journal.pone.0201611.

Zhao, Z., Liu, H. et al. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 14(274). DOI: 10.1186/1471-2164-15-6.

Zhou, Z., Wang, C. and Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nature Communications* 11:3072. DOI: 10.1038/s41467-020-16881-7.

Zimin, A. V. et al. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29(21):2669–2677. DOI: 10.1093/bioinformatics/btt476.

Zubek, S. et al. (2009). Arbuscular mycorrhizal and dark septate endophyte colonization along altitudinal gradients in the tatra mountains. *Arctic, Antarctic, and Alpine Research* 41(2):272–279. DOI: 10.1657/1938-4246-41.2.272.

Zuccaro, A., Lahrmann, U. and Langen, G. (2014). Broad compatibility in fungal root symbioses. *Current Opinion in Plant Biology* 20:135–145. DOI: 10.1016/j.pbi.2014.05.013.

zum Felde, A., Pocasangre, L. and Sikora, R. A. (2003). 'The potential use of microbial communities inside suppressive banana plants for banana root protection'. In: *Banana Root System: towards a better understanding for its productive management*. Ed. by D. W. Turner and F. E. Rosales. San José, Costa Rica:pp. 169–177. URL: https://www.cabdirect.org/cabdirect/abstract/20053211578.

Appendices

Contents

A.1 *Ascomycota* gap analysis supplementary material

A.1.1 Materials and methods

All statistical tests and visualisations were done with R v4.0.2 (R Core Team, 2020) in RStudio (RStudio Team, 2015), using the following packages: ape v5.4-1 (Paradis and Schliep, 2019), cowplot v1.1.0 (Wilke, 2020), dplyr v1.0.2 (Wickham and Seidel, 2020), DescTools v0.99.38 (Signorell, 2020), ggplot2 v3.3.2 (Wickham, 2016), ggpubr v0.4.0 (Kassambara, 2020), ggstance v0.3.4 (Henry, Wickham and Chang, 2020), ggtree v2.3.4 (Yu et al., 2017), grid v4.0.2, gridExtra v2.3 (Auguie, 2017), gtable v0.3.0 (Wickham and Pedersen, 2019), multcompView v0.1-8 (Graves et al., 2019), RCurl v1.98-1.2 (Lang, 2020), rvest v0.3.6 (Wickham, François et al., 2020), scales v1.1.1 (Wickham, 2020), stringr v1.4.0 (Wickham, 2019) and taxize v0.9.99 (Chamberlain and Szöcs, 2013). All the data and the script written for this analysis is available at `https://github.com/Rowena-h/AscomyceteGenome`.

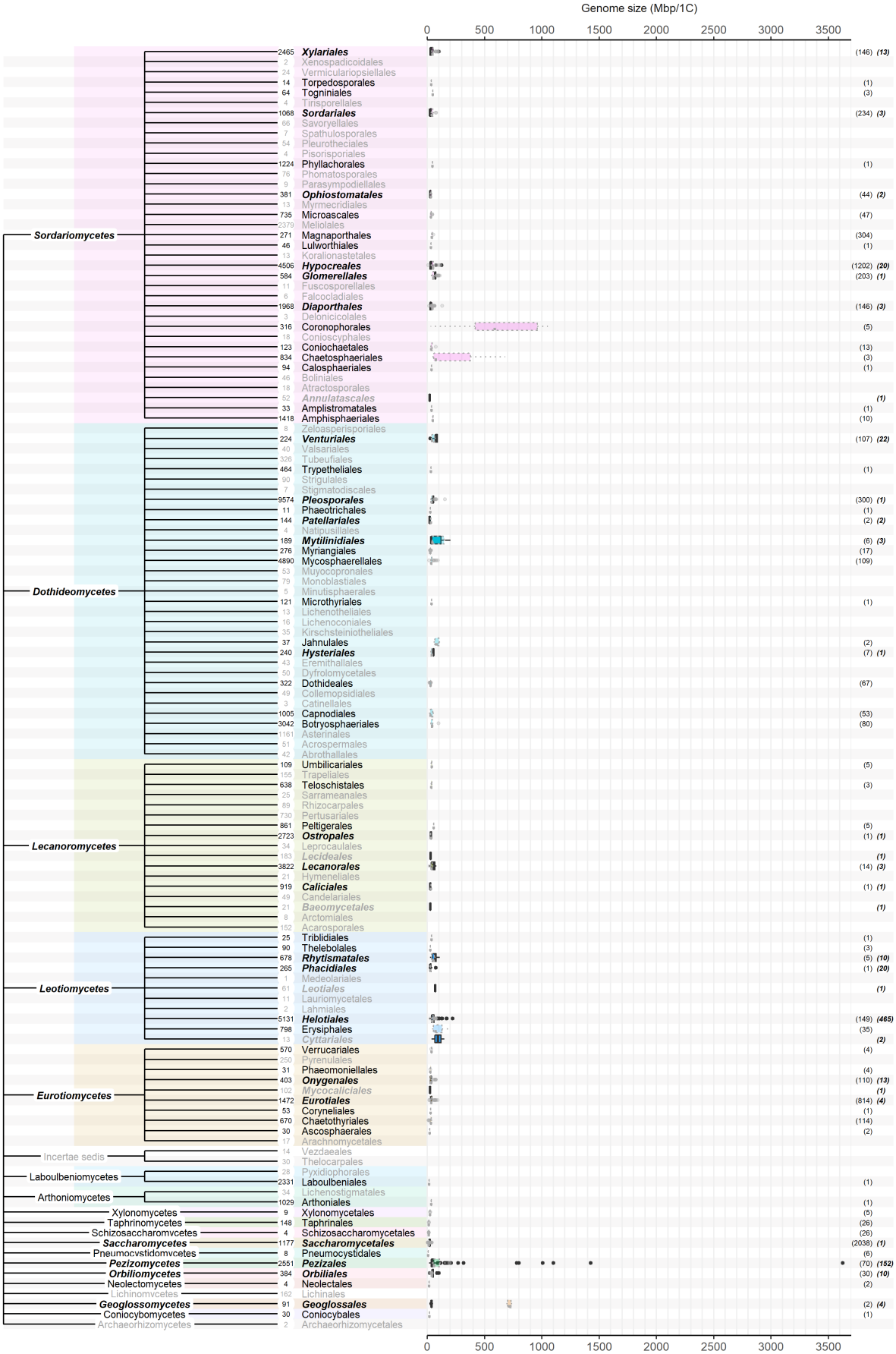
Ascomycete genome assembly data was retrieved from the NCBI genome database (`https://www.ncbi.nlm.nih.gov/genome`) and Mycocosm (Grigoriev et al., 2014; `https://mycocosm.jgi.doe.gov/mycocosm/home`) on 19/01/2021. Data was filtered to remove duplicate assemblies between the two databases and, for NCBI, the same BioSample (i.e. to remove duplicates of the same individual). Accessions were assigned to the order and class shown in Figure 1.2 by matching the genus name with the *Ascomycota* outline of Wijayawardene et al. (2018). The fg_name_search function from the taxize package was used to retrieve number of accepted species names in Species Fungorum (`http://www.speciesfungorum.org/`) for each order. The order *Superstratomycetales* was not shown in Figure 1.2 due to having no accepted species names in Species Fungorum.

Assembly-based genome size data (1C-values) was taken from the assemblies mentioned above, while measurements using cytometric methods were sourced from the Fungal Genome Size Database (`http://www.zbi.ee/fungal-genomesize/`; Kullman, Tamm and Kullman, 2005) and Le Cam et al. (2019). Measurements using the following methods were excluded from the dataset: gel electrophoresis due to the subjectivity of the method; flow cytometry using the fluorochrome DAPI due to its tendency to preferentially bind to AT bases, biasing genome size estimation (Doležel, Sgorbati and Lucretti, 1992); reassociation kinetics; and real-time PCR (Bennett and Leitch, 2011). Genome size data for fungi using all the approaches mentioned above are compiled in the Fungal Genome Size Database, but users are cautioned against using those obtained by biochemical methods and reassociation approaches as they are now considered to be unreliable, and their use has declined significantly in recent years (Bennett and Leitch, 2005a; Bennett and Leitch, 2005b). Where the genus name from the Fungal Genome Size Database did not match to the taxonomy from Wijayawardene et al. (2018), species names were checked in Species Fungorum for the accepted current name using the fg_name_search function.

For orders with at least three genome size measurements from both genome assembly and cytometric sources, data normality was tested using the Shapiro-Wilk test: if data normality could not be assumed, the Wilcoxon signed-rank test was used to compare mean genome size between the two categories, otherwise a two sample t-test was used. The number of genome assemblies and boxplots of genome size were plotted to order level.

We identified and plotted 13 case studies for species-level genome size comparisons with sufficient sample size for statistical testing: 5 in which individual species had genome size measurements from both a cytometric source and multiple genome assemblies (*Aspergillus flavus*, *A. niger*, *Paracocci-*

dioides brasiliensis, *Venturia inaequalis*, and *V. pyrina*); and 8 in which different genome assemblies gave genome size estimations which varied over 20 Mbp/1C (*Fusarium oxysporum*, *F. oxysporum* f. sp. *lini*, *F. proliferatum*, '*F.*' *solani*, *Hortaea werneckii*, *Macrophomina phaseolina*, *Saccharomyces cerevisiae*, and *S. pastorianus*). Genome assembly methods were extracted where possible from NCBI genome reports. The TukeyHSD function was used to identify statistically distinct groups.


Genome size (Mbp/1C)
0
500
1000
1500
2000
2500
3000
3500
Sordariomycetes
Dothideomycetes
Lecanoromycetes
Leotiomycetes
Eurotiomycetes
Incertae sedis
Laboulbeniomycetes
Arthoniomycetes
Xylonomycetes
Taphrinomycetes
Schizosaccharomycetes
Saccharomycetes
Pneumocystidomycetes
Pezizomycetes
Orbiliomycetes
Neolectomycetes
Lichinomycetes
Geoglossomycetes
Coniocybomycetes
Archaeorhizomycetes
2465 Xylariales (146) (13)
2 Xenospadicoidales
24 Vermiculariopsiellales
14 Torpedosporales (1)
64 Togniniales (3)
4 Tirisporellales
1068 Sordariales (234) (3)
66 Savoryellales
7 Spathulosporales
54 Pleurotheciales
4 Pisorisporiales
1224 Phyllachorales (1)
76 Phomatosporales
9 Parasympodiellales
381 Ophiostomatales (44) (2)
13 Myrmaecridiales
735 Microascales (47)
2379 Meliolales
271 Magnaporthales (304)
46 Lulworthiales (1)
13 Koralionastetales
4506 Hypocreales (1202) (20)
584 Glomerellales (203) (1)
11 Fuscosporellales
6 Falocladiales
1968 Diaporthales (146) (3)
3 Delonicicolales
316 Coronophorales (5)
18 Coniosyphales
123 Coniochaetales (13)
834 Chaetosphaeriales (3)
94 Calosphaeriales (1)
46 Boliniales
18 Atractosporales
52 Annulatascales (1)
33 Amplistromatales (1)
1418 Amphisphaeriales (10)
8 Zaldaspherisporiales
224 Venturiales (107) (22)
40 Valsariales
326 Tubeufiales
464 Trypetheliales (1)
90 Strigulales
7 Stigmatodiscales
9574 Pleosporales (300) (1)
11 Phaeotrichales (1)
144 Patellariales (2) (2)
4 Natipusillales
189 Mytilinidiales (6) (3)
276 Myriangiales (17)
4890 Mycosphaerellales (109)
53 Muyocopronales
79 Monoblastiales
5 Minutisphaerales
121 Microthyriales (1)
13 Lichenotheliales
16 Lichenoconiales
35 Kirschsteiniotheliales
37 Jahnulales (2)
240 Hysteriales (7) (1)
43 Eremithallales
50 Dyfrolomycetales
322 Dothideales (67)
49 Collemopsidiales
3 Catinellales
1005 Capnodiales (53)
3042 Botryosphaeriales (80)
1161 Asterinales
51 Acrospermales
42 Abrothallales
109 Umbilicariales (5)
155 Trapeliales
638 Teloschistales (3)
25 Sarrameanales
89 Rhizocarpales
730 Pertusariales
861 Peltigerales (5)
2723 Ostropales (1) (1)
34 Leprocaulales
183 Lecideales (1)
3822 Lecanorales (14) (3)
21 Hymeneliales
919 Caliciales (1) (1)
49 Candelariales
21 Baeomycetales (1)
8 Arctomiales
152 Acarosporales
25 Triblidiales (1)
90 Thelebolales (3)
678 Rhytismatales (5) (10)
265 Phacidiales (1) (20)
1 Medeolariales
61 Leotiales (1)
11 Lauriomycetales
2 Lahmiales
5131 Helotiales (149) (465)
798 Erysiphales (35)
13 Cyttariales (2)
570 Verrucariales (4)
250 Pyrenulales
31 Phaeomoniellales (4)
403 Onygenales (110) (13)
102 Mycocaliciales (1)
1472 Eurotiales (814) (4)
53 Coryneliales (1)
670 Chaetothyriales (114)
30 Ascosphaerales (2)
17 Arachnomycetales
14 Vezdaeales
30 Thelocarpales
28 Pyxidiophorales
2331 Laboulbeniales (1)
34 Lichenostigmatales
1029 Arthoniales (1)
9 Xylonomycetales (5)
148 Taphrinales (26)
4 Schizosaccharomycetales (26)
1177 Saccharomycetales (2038) (1)
8 Pneumocystidales (6)
2551 Pezizales (70) (152)
384 Orbiliales (30) (10)
4 Neolectales (2)
162 Lichinales
91 Geoglossales (2) (4)
30 Coniocybales (1)
2 Archaeorhizomycetales
0
500
1000
1500
2000
2500
3000
3500

A.1.2: A summary showing the taxonomy of the different classes and orders currently recognised in the *Ascomycota*, with boxplots of genome size per order, including extreme outliers. Data is from January 2021. Black taxon labels indicate taxa with representative genome assemblies versus grey for no genome assemblies and bold-italic labels indicate taxa with representative cytometric genome size estimates versus plain text for no cytometric genome size estimates. The number of species for each order is shown to the left of taxon labels. Boxplots of 762 genome size measurements (from 504 species) made using cytometric approaches are taken from Kullman, Tamm and Kullman (2005) and are shown using opaque colours while boxplots for 6,600 genome sizes (from 3,273 strains) based on genome assemblies are given in translucent colours. Sample sizes are shown on the far right, in plain text for the assembly-based estimates (left) and bold-italic for the cytometric estimates (right).

A.1.3: Genome assembly and cytometric methods listed in Figure 1.5. Asterisks (*) mark methods that are believed to be incorrect in NCBI.

	Method	Link/Description	Reference
Genome assembly tools	A5-miseq	https://sourceforge.net/p/ngopt/wiki/A5PipelineREADME/	Coil, Jospin and Darling, 2015
	ABySS	https://github.com/bcgsc/abyss	Simpson et al., 2009
	ALLPATHS	*No longer available*	Butler et al., 2008
	ALLPATHS-LG	https://software.broadinstitute.org/allpaths-lg/blog/	Gnerre et al., 2011
	Arachne	*No longer available*	Batzoglou et al., 2002
	BBMap	https://sourceforge.net/projects/bbmap/	*Unpublished*
	bowtie2	https://github.com/BenLangmead/bowtie2	Langmead and Salzberg, 2012
	Canu	https://github.com/marbl/canu	Koren, Walenz et al., 2017
	Celera	http://wgs-assembler.sourceforge.net	Myers et al., 2000
	CLC	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench/	*Unpublished*
	CONTIGuator	https://github.com/combogenomics/CONTIGuator/	Galardini et al., 2011
	DBG2OLC	https://github.com/yechengxi/DBG2OLC/	Ye et al., 2016
	dipSPAdes	http://gensoft.pasteur.fr/docs/SPAdes/3.0.0/dipspades_manual.html	Safonova, Bankevich and Pevzner, 2015
	DISCOVAR	https://software.broadinstitute.org/software/discovar/blog/	Weisenfeld et al., 2014
	FALCON	https://github.com/PacificBiosciences/FALCON/	Chin, Peluso et al., 2016
	Flye	https://github.com/fenderglass/Flye/	Kolmogorov et al., 2019
	Geneious	https://www.geneious.com/features/assembly-mapping/	*Unpublished*
	HGAP	https://github.com/ben-lerch/HGAP-3.0/	Chin, Alexander et al., 2013
	IDBA-Hybrid	https://github.com/loneknightpy/idba/	*Unpublished*
	IDBA-UD	https://github.com/loneknightpy/idba/	Peng et al., 2012
	MaSuRCA	https://github.com/alekseyzimin/masurca/	Zimin et al., 2013
	MECAT	https://github.com/xiaochuanle/MECAT/	Xiao et al., 2017
	MEGAHIT	https://github.com/voutcn/megahit/	Li, Luo et al., 2016
	MHAP	https://github.com/marbl/MHAP/	Berlin et al., 2015
	MIRA	https://github.com/bachev/mira/	*Unpublished*
	Newbler	*No longer available*	*Unpublished*
	NextDenovo	https://github.com/Nextomics/NextDenovo/	*Unpublished*
	PBcR	http://wgs-assembler.sourceforge.net/wiki/index.php/PBcR	Koren, Schatz et al., 2012
	PCAP	http://seq.cs.iastate.edu/pcap.html	Huang, Wang et al., 2003
	Phrap	http://www.phrap.org/phredphrapconsed.html	*Unpublished*

A.1.3 continued.

	Method	Link/Description	Reference
Genome assembly tools	Pilon *	Pilon is an assembly polisher, not an assembler: `https://github.com/broadinstitute/pilon/` Correct assembly method thought to be Arachne based on Desjardins, Champion et al., 2011	
	PLATANUS	`http://platanus.bio.titech.ac.jp/platanus2/`	Kajitani et al., 2019
	SMARTdenovo	`https://github.com/ruanjue/smartdenovo/`	Liu, Wu et al., 2021
	SMRT Analysis	`https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/`	*Unpublished*
	SOAPdenovo	`https://www.animalgenome.org/bioinfo/resources/manuals/SOAP.html`	Li, Fokkens et al., 2020
	SOAPdenovo2	`https://github.com/aquaskyline/SOAPdenovo2/`	Luo et al., 2012
	SPAdes	`https://github.com/ablab/spades/`	Bankevich et al., 2012
	Sprai *	Sprai is a read corrector, not an assembler: `https://anaconda.org/bioconda/sprai/` Correct assembly method unknown	
	Unicycler	`https://github.com/rrwick/Unicycler/`	Wick et al., 2017
	Velvet	`https://github.com/dzerbino/velvet/`	Zerbino and Birney, 2008
	wtdbg2	`https://github.com/ruanjue/wtdbg2/`	Ruan and Li, 2020
Cytometric methods	FC	Flow Cytometry, unspecified dye	Doležel, Greilhuber and Suda, 2007
	Fe-IC	Image Cytometry, stained with Feulgen, measuring light absorption (also called optical density, OD)	Vilhar et al., 2001
	LM	Light microscopy, stained with orcein	
	PI-FC	Flow Cytometry, stained with Propidium Iodide	Doležel, Greilhuber and Suda, 2007
	SYBR Green I FC	Flow Cytometry, stained with SYBR Green I	Almeida et al., 2007

A.2 *Fusarium chuoi* description (doi:10.3767/persoonia.2021.47.06)

Fungal Planet 1353 – 24 December 2021

Fusarium chuoi R. Hill, Gaya, D.T. Vu, Sand.-Den. & Crous, *sp. nov.*

Etymology. From *chuối*, Vietnamese vernacular name for *Musa* spp., from which the ex-type strain was isolated.

Classification — *Nectriaceae, Hypocreales, Sordariomycetes.*

On SNA and CLA, sporulation abundant from aerial conidiophores and sporodochia. *Aerial conidiophores* erect or prostrate, copiously branching laterally and sympodially, giving rise to macro-, and rarely, microconidia; *aerial conidiogenous cells* mono- and polyphialidic, subulate to subcylindrical, smooth- and thin-walled, proliferating sympodially, 6.5–40.5 × 2.5–4 µm, with apical flared collarette and periclinal thickening; aerial conidia of two types: *microconidia* often produced on prostrate conidiophores, rarely on aerial mycelium, aggregating in false heads, ellipsoidal, subcylindrical to slightly falcate, 0–1-septate, 8–15 × 2–29.5 µm; *macroconidia* fusiform to falcate, straight to apically dorsiventrally curved, apex curved to pointed, base obtuse to papillate, 1–3-septate, smooth- and thin-walled; 1-septate conidia: (14–)18–27.5(–29.5) × (2.5–)3–4 µm (av. 22.8 × 3.2 µm); 2-septate conidia: 26–28.5 × 3–4 µm (av. 27.4 × 3.6 µm); 3-septate conidia: (28–)31.5–43(–50.5) × 3–4 µm (av. 37.3 × 3.5 µm). *Sporodochia* saffron, luteous to ochreous coloured (Rayner 1970), formed abundantly on the agar surface and carnation leaves under nuv. *Conidiophores in sporodochia*, densely and irregularly branched, bearing apical whorls of 2–4 monophialides; sporodochial monophialides subcylindrical, 10–26 × 2.5–4.5 µm, smooth- and thin-walled, with a distinct apical collarette. *Sporodochial conidia* (macroconidia) falcate, almost straight to gently curved, tapering at both ends, apex curved to blunt, base poorly- to well-developed foot-shaped, 1–6-septate, hyaline, smooth- and thin-walled; 1-septate conidia: (14.5–)15–20.5(–24) × 3–4.5 µm (av. 17.9 × 3.9 µm); 2-septate conidia: 21.5–32 × 3–4.5 µm (av. 26.4 × 3.5 µm); 3-septate conidia: (33–)43–61(–71.5) × (3–)4–5 µm (av. 51.8 × 4.2 µm); 4-septate conidia: (50.5–)55–69(–74.5) × 3.5–5 µm (av. 62.3 × 4.2 µm); 5-septate conidia: 54 × 4.5 µm (rare); 6-septate conidia: (49.5–)56.5–71(–73) × (3.5–)4–4.5(–5) µm (av. 63.8 × 4.3 µm). *Chlamydospores* not observed.

Culture characteristics — Colonies on potato dextrose agar (PDA) and oatmeal agar (OA) growing in the dark at 24 °C covering and entire 9 cm Petri dish in 7 d. Colony surface peach to vinaceous, flat, velvety to felty with abundant floccose aerial mycelium forming concentric rings; colony margins undulate. Reverse flesh to salmon with diffuse coral to brick pigment throughout the medium.

Typus. Vietnam, Hà Tĩnh Province, Hương Sơn District, Sơn Kim commune, N18°25′37.38″ E105°12′53.95″, inside seed of *Musa itinerans* (*Musaceae*), 9 Nov. 2014, *D.M. Thu, L.T. Phong & T.T. Duong*, isol. *R. Hill* (holotype CBS H-24901, culture ex-type CBS 148464; ITS, LSU, *cmdA*, *rpb1*, *rpb2*, *tef1* and *tub2* sequences GenBank OK586454, OK586452, OK626304, OK626306, OK626302, OK626308 and OK626310, MycoBank MB 841865).

Colour illustrations. Flowers, fruits, leaves and seeds of *Musa itinerans* (background photo by D.T. Vu); from top to bottom and left to right: colony on PDA after 14 d at 24 °C in darkness (left = obverse,right = reverse), sporodochia formed on CLA, aerial conidiophore, aerial conidiogenous cells, aerial conidia, sporodochial conidia. Scale bars: black = 20 µm, white = 10 µm.

Additional material examined. Vietnam, Nghệ An Province, Con Cuông District, Châu Khê commune, N19°1′48.73″ E104°43′31.97″, inside seed of *M. itinerans*, 18 Nov. 2014, *L.T. Phong, V.V. Tung & T.T. Duong*, isol. *R. Hill* (culture CBS 148465; ITS, LSU, *cmdA*, *rpb1*, *rpb2*, *tef1* and *tub2* sequences GenBank OK586455, OK586453, OK626305, OK626307, OK626303, OK626309 and OK626311).

Notes — *Fusarium chuoi* resides in the Asian clade of the *Fusarium fujikuroi* species complex (FFSC: O'Donnell et al. 1998, Yilmaz et al. 2021, Crous et al. 2021b). Based on nucleotide searches using the *Fusarium* Pairwise ID engine on the Fusarioid-ID database (www.fusarium.org, Crous et al. 2021) the closest hit using the **ITS** sequence was *Fusarium siculi* (strain CBS 142422; identities = 449/450 (99 %), no gaps). The closest hit using the **LSU** sequence was *F. siculi* (strain CBS 142422; identities = 804/805 (99 %), no gaps). Closest hit using the ***cmdA*** sequence was *Fusarium fractiflexum* (strain NRRL 28852; identities = 426/434 (98 %), no gaps). Closest hit using the ***rpb1*** sequence was *F. fujikuroi* (strain NRRL 13566; identities = 687/702 (98 %), no gaps). Closest hit using the ***rpb2*** sequence was *Fusarium globosum* (strain CBS 428.97; identities = 856/867 (98 %), no gaps). Closest hit using the ***tef1*** sequence was *F. fractiflexum* (strain NRRL 28852; identities = 619/643 (96 %), 2 gaps (0.3 %)). The phylogenetic results, however, showed that *F. chuoi* is not directly related to any of the previously described species of FFSC (see Suppl. material FP1353), clustering as the second basal-most species of that clade after *F. sacchari*.

Asian *Fusarium* spp. in the FFSC are characterised by mono- and polyphialides producing oval to ellipsoid, rarely pyriform to globose (i.e., *F. annulatum*, *F. fujikuroi* and *F. globosum*) microconidia organized in chains or false heads; 3–5-septate sporodochial conidia and lacking chlamydospores. The elaborate, profusely branched aerial conidiophores of *F. chuoi* are comparable to those of *F. concentricum*, *F. lumajangense*, *F. mangiferae* and *F. sacchari*, all the latter species producing oval, ellipsoidal to allantoid microconidia on false heads. Aerial conidiophores of *F. chuoi*, however, mostly produce macroconidia, while microconidia grouped on false heads are restricted to short, mostly unbranched and prostrate conidiophores formed on the surface on the culture media.

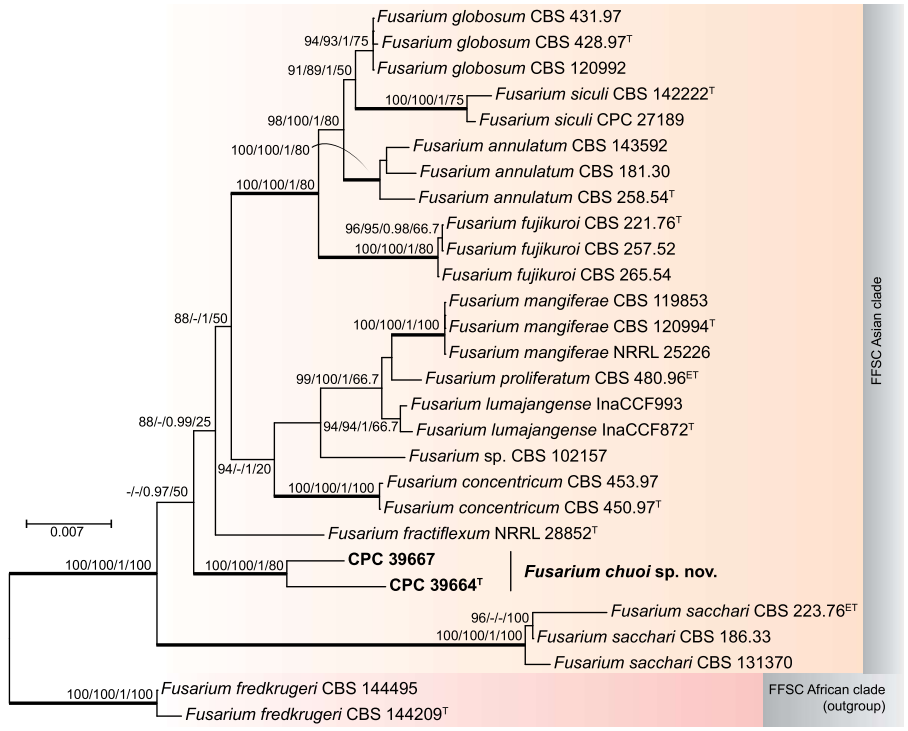
Several Asian species of the FFSC have been reported from *Musa* spp. i.e., *F. annulatum*, *F. concentricum*, *F. fujikuroi*, *F. lumajangense* and *F. sacchari* (Leslie & Summerell 2006, Maryani et al. 2019, Farr & Rossman 2021). The two strains representing *F. chuoi* were isolated as endophytes from asymptomatic seeds of wild banana (*Musa itinerans*), which had been collected predispersal and stored in the Millennium Seed Bank for ~2.5 years at -20 °C prior to isolation.

Supplementary material

FP1353 Phylogenetic tree.

Rowena Hill & Ester Gaya, Comparative Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK; e-mail: R.Hill@kew.org & EGaya@kew.org
Dang Toan Vu, Research Planning and International Cooperation Department, Plant Resources Center, An Khanh, Hoai Duc, Hanoi 152900, Vietnam; e-mail: toanvd.prc@mard.gov.vn
Marcelo Sandoval-Denis & Pedro W. Crous, Westerdijk Fungal Biodiversity Institute, Phytopathology, P.O. Box 85167, 3508 AD, Utrecht, The Netherlands; e-mail: m.sandoval@wi.knaw.nl & p.crous@wi.knaw.nl

Fungal Planet 1353 – *Fusarium chuoi*



FP1353 Maximum-likelihood phylogram inferred from the combined *cmdA*, *rpb1*, *rpb2*, *tef1* and *tub2* sequence alignment of 28 *Fusarium* strains belonging to the *Fusarium fujikuroi* species complex. The analysis included 3 983 characters including alignment gaps. The tree was obtained using RAxML v. 8.2.12 (Stamatakis 2014) on the CIPRES Science Gateway (Miller et al. 2010), and parallel analyses using MrBayes v. 3.2.7a (Ronquist & Huelsenbeck 2003) and IQ-TREE v. 2.1.2 (Nguyen et al. 2015) were run with settings as described elsewhere (Crous et al. 2021). Values at the nodes are RAxML bootstrap support (BS) (≥ 80) followed by IQ-TREE BS (≥ 95), Bayesian posterior probabilities (≥ 0.95) and IQ-TREE gene concordance factors**.** The tree is rooted with *F. fredkrugeri* (CBS 144209 and CBS 144495). The scale bar indicates expected changes per site. Ex-epitype and ex-type strains are indicated by ^{ET} and ^{T}, respectively. Fully supported branches (BS = 100 and PP = 1.0), and the new species *F. chuoi* are highlighted in **bold.**

A.3: Metadata for all the fusarioid strains used in Chapter 3. Lifestyle reports were excluded if the taxonomic classification relied solely on internal transcribed spacer (ITS) sequences, which is an unreliable barcode for *Fusarium* spp.[1] Names of strains that were sequenced and assembled in this thesis are in bold. All numbered references indicated in superscript are listed at the end of the table.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Albonectria albosuccinea =*Fusarium albosuccineum*	GCA_012931995.1	NRRL 20459	tree	**plant associate**[2]	**endophyte** (*Nectandra lineatifolia* (as *Nectriaceae*)[3]) **saprotroph** (sterile *Nectandra lineatifolia* wood (as *Nectriaceae*)[3])
Albonectria rigidiuscula =*Fusarium decemcellulare*	GCA_013266205.1	NRRL 13412	*Coffea* sp.	**plant pathogen**[4]	**endophyte** (*Theobroma gileri*[5]) **plant pathogen** (*Dimocarpus longan*, *Mangifera indica*, *Nephelium lappaceum*[6]; *Malus pumila*[7]; *Magnolia denudata*[8]; *Persea americana*[9])
Fusarium acutatum	GCA_012932015.1	NRRL 13308	?	**plant associate**[10]	**animal associate** (*Homoptera* sp.[10]) **human pathogen**[11,12,13] **plant associate** (*Cajanus* sp., *Triticum* sp.[10]) **plant pathogen** (*Cyamopsis tetragonoloba*[14])
Fusarium agapanthi	GCA_001654555.2	NRRL 31653	*Agapanthus praecox*	**plant pathogen**[15]	**plant pathogen** (*Agapanthus africanus*[16])
Fusarium annulatum ET1[a]	GCA_900067095.1	ET1	*Dendrobium moschatum*	**endophyte**[18]	**animal associate** (*Hylurgops palliatus*[19]) **endophyte** (*Austrostipa aristiglumis*[20]; grapevine[21]; *Hevea brasiliensis*[22]; *Lilium longiflorum*[23])

[a]Previously classified as *F. proliferatum*, which was epitypified[17]

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Fusarium annulatum Fp A8[b]	GCA_003615215.1	Fp_A8	*Allium cepa*	**plant pathogen**[24]	**human pathogen**[25] **mycoparasite** (Smut, *Stereum hirsutum*[27]) **plant pathogen** (*Albizia julibrissin*[28]; *Allium cepa*[29]; *Allium sativum*[30]; *Allium tuberosum*[31]; *Cannabis sativa*[32]; *Carthamus tinctorius*[33]; *Colchicum kotschyi*[34]; *Echeveria desmetiana*[35]; *Gladiolus* spp.[36]; *Gypsophila paniculata*[37]; *Ilex cornuta*[38]; *Laelia* spp.[39]; *Lilium longiflorum*[23]; *Malus sieversii*[40]; *Musa* ABB[41]; *Oryza sativa*[42]; *Polygonatum cyrtonema*[43]; *Prunus persica*[44]; *Sansevieria trifasciata*[45]; sunflower[46]; *Vaccinium corymbosum* 'O'Neal'[47]) **saprotroph** (*Arctoscopus japonicus* egg masses[48]; buried *Cunninghamia lanceolata* sticks[49]; petroleum-contaminated soil[50]; washing machines[51])
***Fusarium annulatum* RH5**	GCA_022627115.1	880149-04	*Musa itinerans*	**endophyte** (seed)[26]	
Fusarium anthophilum	GCA_013364935.1	NRRL 25214	*Hippeastrum* sp.	**plant associate**[52]	**endophyte** (*Austrostipa aristiglumis*[20]; *Vigna unguiculata*[53]) **plant pathogen** (millets[54]; sunflower[55])

[b]Previously classified as *F. proliferatum*, which was epitypified[17]

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Fusarium austroafricanum	GCA_012932025.1	NRRL 53441	plant debris in soil	**saprotroph**[56]	**endophyte** (*Pennisetum clandestinum*)[56]
Fusarium austroamericanum	GCA_013364965.1	NRRL 2903	polypore fungus	**mycoparasite**[57]	**plant pathogen** (*Hordeum vulgare*[58]; wheat[59,60])
Fusarium avenaceum	GCA_000769215.1	Fa05001	*Hordeum vulgare*	**plant pathogen**[61]	**animal associate** (*Astacus astacus*[62]; *Austrostipa aristiglumis*[20]; *Sitophilus oryzae*[63]) **endophyte** (*Abies alba*[64]; *Cucurbita maxima*[65]; *Lilium longiflorum* bulb[23]; *Salicornia europaea*[66]) **plant associate** (*Salix* spp.[67]) **plant pathogen** (*Actinidia chinensis* var. *chinensis* and var. *deliciosa*[68]; *Allium giganteum*[69]; *Cucurbita maxima*[65]; *Glycine max*[70]; *Malus sieversii*[40]; *Lepidium meyenii*[71]; *Lupinus angustifolius*[72]; *Pisum sativum*[73]; *Racomitrium japonicum*[74]; *Tanacetum cinerariifolium*[69]) **saprotroph** (*Arctoscopus japonicus* egg masses[48]; burnt *Pinus mugo* stumps[75]; saline/acidic soil[76])
Fusarium beomiforme	GCA_002980475.2	NRRL 25174	soil	**saprotroph**[77]	**plant associate** (*Sorghum bicolor* stalk[78]) **plant pathogen** (wheat cultivar Norm[79])

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Fusarium bulbicola	GCA_013758895.1	NRRL 25176	*Nerine bowdenii*	**plant pathogen**[10]	**plant pathogen** (*Glycine max* roots[80]; *Haemanthus* and *Vallota* bulbs[10]) **endophyte** (*Euterpe oleracea* (as *F. sacchari* var. *elongatum*)[81]; *Xanthorrhoea*[82]) **saprotroph** (soil[82,83])
***Fusarium chuoi* RH1**	GCA_022627125.1	836515-16	*Musa itinerans*	**endophyte** (seed)[26]	
***Fusarium chuoi* RH3**	GCA_022627105.1	836445-12-1	*Musa itinerans*	**endophyte** (seed)[26]	
Fusarium circinatum	GCA_013396185.1	NRRL 25331	*Pinus radiata*	**plant pathogen**[10]	**animal associate** (*Brachyderes incanus*, *Hylastes attenuatus*, *Hylurgops palliatus*, *Hypothenemus eruditus*, *Ips sexdentatus*, *Orthotomicus erosus*, *Pityophthorus pubescens*[19]) **endophyte** (*Zea mays*[84]) **plant pathogen** (*Solanum lycopersicum*[85])
Fusarium coffeatum	GCA_003316985.1	FIESC_28	*Sorghum bicolor*	**plant associate**[86]	**endophyte** (*Carapichea ipecacuanha* leaves and roots (as *F. chlamydosporum* var. *fuscum*)[87]) **plant associate** (*Cynodontis lemfuensis* (as *F. chlamydosporum* var. *fuscum*)[88]) **plant pathogen** (*Penniscti dandestini* (as *F. chlamydosporum* var. *fuscum*)[88]

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					saprotroph (soil (as *F. chlamydosporum* var. *fuscum*)[88])
Fusarium coicis	GCA_013781345.1	NRRL 66233	*Coix gasteenii*	**endophyte**[89]	
Fusarium culmorum	GCA_003033665.1	PV	soil	**saprotroph**[90]	**animal associate** (*Hypothenemus eruditus*, *Orthotomicus erosus*[19]; *Placospongia intermedia*[91]) **endophyte** (*Austrostipa aristiglumis*[20]; *Citrus sinensis* xylem[92]; *Leymus mollis*[93]) **mycoparasite** (*Verticillium dahlia*[94]) **plant associate** (*Ammophila arenaria*[95]; *Salix* spp.[67]) **plant pathogen** (*Brassica napus*[96]; *Cucurbita maxima*[65]; *Hordeum distichon*, *Hordeum vulgare*, *Triticum aestivum*, *Triticum turgidum* var. *durum*[97]; oat cv. Gerald and wheat cv. Claire[98]; *Solanum tuberosum*[99]) **saprotroph** (saline soil[100])
Fusarium denticulatum	GCA_013396175.1	NRRL 25311	*Ipomoea batatas*	**plant pathogen**[10]	**endophyte** (*Zea mays*)[101]
Fusarium flagelliforme	GCA_003012295.1	NRRL 13405	*Zea mays*	**plant pathogen**[102,103]	**plant associate** (*Hordeum vulgare*[104]; *Pinus nigra*, *Thuja* sp., wheat[105])
Fusarium fujikuroi	GCF_900079805.1	IMI 58289	*Saccharum officinarum*	**plant pathogen**[106,107]	**endophyte** (*Debregeasia salicifolia*[108]; *Glycine max*[109]) **human pathogen**[25]

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					plant pathogen (*Aspidosperma polyneuron*[110]; *Bletilla striata*[111]; *Canna edulis*[112]; *Lactuca serriola*[113]; *Lasia spinosa*[114]; *Oryza sativa*[42]; plum[115]; millets[54]) **saprotroph** (*Arctoscopus japonicus* egg masses[48]; *Diaphorina citri* cadavers[116]; washing machines[51])
Fusarium gaditjirri	GCA_013266175.1	NRRL 45417 FRC M-8754	*Heteropogon triticeus*	**endophyte**[117]	
Fusarium globosum	GCA_013396165.1	NRRL 26131	*Zea mays*	**plant associate** (seed)[118]	**endophyte** (*Austrostipa aristiglumis*[20]) **plant pathogen** (*Arundo donax*[119]; *Hordeum vulgare*[120])
Fusarium graminearum	GCA_000240135.3	NRRL 31084	?	**plant pathogen**[121]	**endophyte** (*Cucurbita maxima*[65]; *Solanum lycopersicum*[122]) **plant associate** (*Agarum clathratum* (marine)[123]; *Rumohra adiantiformis*[124]) **plant pathogen** (*Avena*, *Hordeum*, *Zea* spp.[125]; *Glycine max*[109]; *Ipomoea batatas*[126]; *Oryza sativa* cv. Doongara[127]; *Setaria italica*[128]; *Solanum tuberosum*[99]) **saprotroph** (*Arctoscopus japonicus* egg masses[48])
Fusarium heterosporum	GCA_013396295.1	NRRL 20693	*Claviceps purpurea*	**mycoparasite**[129]	**endophyte** (*Austrostipa aristiglumis*[20])

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					saprotroph (soil[130])
Fusarium langsethiae	GCA_001292635.1	9821-16-1 Fl201059	*Avena sativa*	**plant associate** (seed)[131]	**endophyte** (oat cv. Gerald and wheat cv. Claire[98c]) **plant pathogen** (barley, oat, wheat[61])
Fusarium longipes	GCA_003012285.1	NRRL 20695	soil	**saprotroph**[132]	**endophyte** (*Musa* sp. var. Pisang Awak pseudostem[133]) **mycoparasite** (*Sclerospora graminicola*[134]) **plant pathogen** (wheat roots and stalks[135])
Fusarium mangiferae	GCA_900044065.1	MRC7560	*Mangifera indica*	**plant pathogen**[136]	**endophyte** (*Sansevieria trifasciata*[45])
Fusarium mexicanum	GCA_013396015.1	NRRL 53147	*Mangifera indica*	**plant pathogen**[137]	**plant pathogen** (*Swietenia macrophylla*[138])
Fusarium mundagurra	GCA_013396205.1	NRRL 66235	soil	**saprotroph**[89]	**human pathogen**[139] **plant associate** (*Mangifera indica*)[89]
Fusarium napiforme	GCA_013396005.1	NRRL 25196	millet	**endophyte** (seed)[140,141]	**endophyte** (*Rhizophora mucronate*)[142] **human pathogen**[143] **plant associate** (*Sorghum caffrorum*[140]) **plant pathogen** (*Cucurbita maxima*[65]) **saprotroph** (soil[140])

[c]Only pathogenised on detached and mostly wounded leaves, latent saprotroph?

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Fusarium nygamai	GCA_002894225.1	CS10214	*Triticum* sp.	**endophyte**[144]	**endophyte** (*Austrostipa aristiglumis*[20]; *Solanum lycopersicum* roots[145]) **human pathogen**[146] **plant associate** (*Phaseolus vulgaris*[147]) **plant pathogen** (*Oryza sativa*[148]; millets[54]); Sorghum[147,149]; *Solanum tuberosum*[150]; *Striga hermonthica*[151]) **saprotroph** (soil[147]; petroleum-contaminated soil[50])
Fusarium odoratissimum =*Fusarium oxysporum* f. sp. *cubense*	GCA_000350365.1	Foc4_1.0	*Musa* spp. AAA cv. Brazilian	**plant pathogen**[152]	
Fusarium odoratissimum =*Fusarium oxysporum* f. sp. *cubense* TR4	GCA_000260195.2	NRRL 54006	*Musa* sp.	**plant pathogen**[153]	
Fusarium oxysporum f. sp. *cepae*	GCA_003615085.1	FoC_Fus2	*Allium cepa*	**plant pathogen**[154,155]	**endophyte** (>10 crop species[156,157])
Fusarium oxysporum f. sp. *conglutinans*	GCA_014154955.1	Fo5176	*Brassica oleracea*	**plant pathogen**[158]	**plant pathogen** (*Arabidopsis*[158])
Fusarium oxysporum f. sp. *lycopersici*	GCA_000149955.2	4287	*Solanum lycopersicum*	**plant pathogen**[159]	
Fusarium oxysporum f. sp. *radicis-lycopersici*	GCA_000260155.3	26381	*Solanum lycopersicum*	**plant pathogen**[153]	**plant pathogen** (>30 crop species[160])

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Fusarium phyllophilum	GCA_013396025.1	NRRL 13617	*Dracaena dermensis*	**plant pathogen**[10]	**plant pathogen** (*Aloe arborescens*[161]; *Gasteria excavata*, *Sansevieria dooneri*[10])
Fusarium poae	GCA_001675295.1	2516	*Triticum aestivum*	**plant pathogen**[162]	**endophyte** (*Austrostipa aristiglumis*[20]; oat cv. Gerald, wheat cv. Claire[98]) **plant pathogen** (alfalfa, barley, bent grasses, corn, fescue, Kentucky bluegrass, oat, rice, soybean, sunflower, timothy, tomato[163])
***Fusarium proliferatum* RH7**	GCA_022627135.1	836489-13	*Musa balbisiana*	**endophyte** (seed)[26]	**saprotroph** (soil[17])
Fusarium pseudoanthophilum	GCA_013395995.1	NRRL 25211	*Zea mays*	**plant associate**[10]	**plant pathogen** (*Capsicum annuum* var. *grossum*, *Capsicum annuum* var. *longum*, *Solanum lycopersicum*[164])
Fusarium pseudocircinatum	GCA_013396035.1	NRRL 36939	?	**plant associate**[d]	**animal associate** (*Heteropsylla incisa*[10]) **human pathogen**[165] **endophyte** (*Handroanthus chrysotrichus*[166]) **plant associate** (*Oryza sativa*[167]; *Pinus kesiya*, *Solanum* sp.[10]) **plant pathogen** (*Acacia koa*[168]; *Mangifera indica*[137]; *Sansevieria trifasciata*[45]; *Swietenia macrophylla*[138])

[d]Presumed from original description[10], in absence of associated data.

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					saprotroph (dead leaves and textile[10])
Fusarium pseudograminearum	GCA_000303195.2	CS3096	wheat	**plant pathogen**[169]	**endophyte** (*Austrostipa aristiglumis*[20]) **plant associate** (barley, oat, *Medicago truncatula*, *Phalaris paradoxa*[170]) **plant pathogen** (*Hordeum distichon*, *Hordeum vulgare*, *Triticum aestivum*, *Triticum turgidum* var. *durum*[97]) **saprotroph** (soil[170])
Fusarium sarcochroum	GCA_013266185.1	NRRL 20472	*Viscum album*	**plant associate**[171]	**endophyte** (*Citrus reticulata* and *Citrus limon* twigs and trunks[172])
***Fusarium* sp. RH6**	GCA_022627095.1	836490-20	*Musa itinerans*	**endophyte** (seed)[26]	
Fusarium sporotrichioides	GCA_003012315.1	NRRL 3299	*Zea mays*	**plant pathogen**[173]	**endophyte** (*Abies alba*[64]; Salicornia europaea[66]) **plant pathogen** (*Glycine max*[80]; *Malus sieversii*[40]; sunflower[55]; *Zea mays*[174]) **saprotroph** (*Arctoscopus japonicus* egg masses[48])
Fusarium subglutinans	GCA_013396075.1	NRRL 66333	*Zea mays*	**plant pathogen**[175]	**endophyte** (*Austrostipa aristiglumis*[20]) **plant associate** (*Oryza sativa* cv. Doongara[127])

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					plant pathogen (*Aspidosperma polyneuron*[110]; *Cymbidium hybridum*[176]; *Helianthus annuus*[55]; millets[54])
Fusarium tjaetaba	GCA_013396195.1	NRRL 66243	*Sorghum interjectum*	**endophyte**[89]	
Fusarium venenatum	GCA_900007375.1	A3/5	soil	**saprotroph**[177]	**plant associate** (*Solanum tuberosum*[99]; *Trifolium subterraneum*[178])
Fusarium verticillioides	GCA_000149555.1	NRRL 20956	*Zea mays*	**plant pathogen**[179]	**animal associate** (*Brachyderes incanus*, *Hylurgops palliatus*, *Ips sexdentatus*, *Orthotomicus erosus*[19]) **endophyte** (*Cucurbita* sp.[65]; *Oryza sativa* cv. Quest[127]; *Glycine max*[109]; *Solanum lycopersicum* roots[145]) **human pathogen**[165] **plant pathogen** (*Aspidosperma polyneuron*[110]; millets[54]; *Musa* spp.[180]; Sorghum[149]; sugarcane[181]) **saprotroph** (buried *Cunninghamia lanceolata* sticks[49]; raw milk and cheese[182])
Geejayessia zealandica =*Fusarium zealandicum*	GCA_013266195.1	NRRL 22465	?	**plant associate**[183]	**plant associate** (*Hoheria populnea*, *Plagianthus* sp. bark[184])
Ilyonectria sp.	Ilysp1	Ilysp1	*Populus deltoides*	**endophyte**[185]	

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
Neocosmospora ambrosia =*Fusarium ambrosium*	GCA_003947045.1	NRRL 20438	*Camelia sinensis*	**insect mutualist** (*Euwallacea 'fornicatus'*)[186]	**plant pathogen** (*Camellia sinensis*[187])
Neocosmospora euwallaceae =*Fusarium euwallaceae*	GCA_003957675.1	UCR1854	*Persea americana*	**insect mutualist** (*Euwallacea* sp.)[188]	**plant pathogen** (>100 tree species[189])
Neocosmospora floridana =*Fusarium floridanum*	GCA_003947005.1	NRRL 62606	*Acer negundo*	**insect mutualist** (*Euwallacea interjectus*)[190]	**plant pathogen** (*Acer negundo*[190])
Neocosmospora kuroshia =*Fusarium kuroshium*	GCA_003698175.1	UCR3666	*Persea americana*	**insect mutualist** (*Euwallacea* sp.)[191]	**plant pathogen** (*Acer negundo*, *Albizia julibrissin*, *Baccharis salicifolia*, *B. pilularis*, *Dombeya cacuminum*, *Erythrina humeana*, *Persea americana*, *Populus fremontii*, *P. nigra*, *Platanus racemosa*, *Quercus agrifolia*, *Q. suber*, *Ricinus communis*, *Robinia pseudoacacia*, *Salix gooddingii*, *S. laevigata*, *S. lasiolepis*, *Tamarix ramosissima*[192])
Neocosmospora oligoseptata	GCA_003946995.1	AF-4 NRRL 62579	*Ailanthus altissima*	**insect mutualist** (*Euwallacea validus*)[188]	**plant pathogen**[188]
Neocosmospora pisi =*Fusarium vanettenii* =*Fusarium solani* f. sp. *pisi*	GCA_000151355.1	NRRL 44580 77-13-4	*Pisum sativum*[e]	**plant pathogen**[193]	**endophyte** (*Lathyrus aphaca*, *L. ochrus*, *Lotus pedunculatus*, *Medicago arabica*, *M. polymorpha*, *Trifolium angustifolium*, *T. arvense*, *T. campestre*, *T. repens*, *T. subterraneum*, *Vicia benghalensis*, *V. hirsute*, *V. villosa*[194])

[e]'third generation cross between two field isolates: one (T2) obtained from a infected pea plant in NY and the other (T219) obtained from soil in a potato field in PA'

A.3 continued.

Species	Assembly accession	Strain/Voucher	Host/Substrate	Lifestyle	Other reported lifestyles
					plant pathogen (*Crotalaria ochroleuca*, *Galega officinalis*, *Lathyrus dymenum*, *L. gorgoni*, *L. inconspicuus*, *L. ochrus*, *L. sativus*, *L. sylvestris*, *Medicago arabica*, *M. orbicularis*, *Melilotus albus*, *Scorpiurus muricatus*, *Trifolium diffusum*, *T. palaestinum*, *T. subterraneum*, *Trigonella foenum-graecum*, *Vicia articulata*, *V. ervilia*, *V. fulgens*, *V. sativa*, *V. villosa* subsp. *varia*[194])
'Fusarium' sp. AF-6	GCA_003947015.1	NRRL 62590	*Persea americana*	**insect mutualist** (*Euwallacea* sp.)[186]	**plant pathogen**[186]
'Fusarium' duplospermum	GCA_003946985.1	NRRL 62584 AF-8	*Persea americana*	**insect mutualist** (*Euwallacea perbrevis*)[186,195]	**plant pathogen**[186]

1 Geiser, D. M. et al. (2004). FUSARIUM-ID v. 1.0: A DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* 110(5-6):473–479. DOI: 10.1023/B:EJPP.0000032386.75915.a0.

2 NCBI (2021a). *Fusarium albosuccineum NRRL 20459*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683636/ (visited on 17/05/2021).

3 Nelson, A., Vandegrift, R. et al. (2020). Double lives: transfer of fungal endophytes from leaves to woody substrates. *PeerJ* 8:e9341. DOI: 10.7717/peerj.9341.

4 ARS Culture Collection (2021). *Fusarium decemcellulare NRRL 13412*. URL: https://nrrl.ncaur.usda.gov/cgi-bin/usda/mold/report.html?nrrlcodes=13412 (visited on 25/03/2021).

5 Harry, C., Keith, A. and Sarah, E. (2003). Endophytes and mycoparasites associated with an indigenous forest tree, *Theobroma gileri*, in Ecuador and a preliminary assessment of their potential as biocontrolagents of cocoa diseases. *Mycological Progress* 2(2):149–160. DOI: 0.1007/s11557-006-0053-4.

6 Serrato-Diaz, L. M. et al. (2015). First Report of *Fusarium decemcellulare* Causing Inflorescence Wilt and Vascular and Flower Necrosis of Rambutan (*Nephelium lappaceum*), Longan (*Dimocarpus longan*), and Mango (*Mangifera indica*). *Plant Disease* 99(8). DOI: 10.1094/PDIS-09-14-0923-PDN.

7 Lee, S.-Y., Park, S.-J., Lee, J.-J., Back, C.-G. et al. (2017). First Report of Fruit Rot Caused by *Fusarium decemcellulare* in Apples in Korea. *The Korean Journal of Mycology* 45(1):54–62. DOI: 10.4489/kjm.20170006.

8 Wang, Y. X., Chen, J. Y. et al. (2015). First Report of Canker of *Magnolia denudata* Caused by *Fusarium decemcellulare* in Hubei, China. *Plant Disease* 99(7). DOI: 10.1094/PDIS-11-14-1111-PDN.

9 Darvas, J. and Kotze, J. (1987). Fungi associated with pre-and postharvest diseases of avocado fruit at Westfalia Estate, South Africa. *Phytophylactica* 19:83–85. DOI: 10520/AJA03701263_1074.

10 Nirenberg, H. I. and O'Donnell, K. (1998). New *Fusarium* species and combinations within the *Gibberella fujikuroi* species complex. *Mycologia* 90(3):434–458. DOI: 10.2307/3761403.

11 Taj-Aldeen, S. J. et al. (2006). Gangrenous necrosis of the diabetic foot caused by *Fusarium acutatum*. *Medical Mycology* 44:547–552. DOI: 10.1080/13693780500543246.

12 Gupta, C. et al. (2016). Genotyping and *In Vitro* Antifungal Susceptibility Testing of *Fusarium* Isolates from Onychomycosis in India. *Mycopathologia* 181:497–504. DOI: 10.1007/s11046-016-0014-7.

13 Muraosa, Y. et al. (2017). Epidemiological study of *Fusarium* species causing invasive and superficial fusariosis in Japan. *Medical Mycology Journal* 58E:E5–E13. DOI: 10.3314/mmj.16-00024.

14 Gautam, R., Singh, S. K. and Sharma, V. (2016). Molecular diagnosis and intraspecific genetic variability of root pathogens of arid legumes in Western Rajasthan, India. *Revista de Biología Tropical* 64(4):1505–1518. DOI: 10.15517/rbt.v64i4.23331.

15 Edwards, J. et al. (2016). *Fusarium agapanthi* sp. nov., a novel bikaverin and fusarubin-producing leaf and stem spot pathogen of *Agapanthus praecox* (African lily) from Australia and Italy. *Mycologia* 108(5):981–992. DOI: 10.3852/15-333.

16 Guarnaccia, V. et al. (2019). Soilborne diseases caused by *Fusarium* and *Neocosmospora* spp. on ornamental plants in Italy. *Phytopathologia Mediterranea* 58(1):127–137. DOI: 10.14601/Phytopathol_Mediterr-23587.

17 Yilmaz, N. et al. (2021). Redefining species limits in the *Fusarium fujikuroi* species complex. *Persoonia* 46:129–162. DOI: 10.3767/persoonia.2021.46.05.

18 Niehaus, E. M. et al. (2016). Comparative "Omics" of the *Fusarium fujikuroi* Species Complex Highlights Differences in Genetic Potential and Metabolite Synthesis. *Genome Biology and Evolution* 8(11):3574–3599. DOI: 10.1093/gbe/evw259.

19 Romón, P. et al. (2008). Fungal communities associated with pitch canker disease of *Pinus radiata* caused by *Fusarium circinatum* in northern Spain: association with insects and pathogen-saprophyte antagonistic interactions. *Canadian Journal of Plant Pathology* 30:241–253. DOI: 10.1080/07060661.2008.10540539.

20 Bentley, A. R. et al. (2007). Crop pathogens and other *Fusarium* species associated with *Austrostipa aristiglumis*. *Australasian Plant Pathology* 36:434–438. DOI: 10.1071/AP07047.

21 Akgül, D. and Ahioğlu, M. (2019). Fungal pathogens associated with young grapevine decline in the Southern Turkey vineyards. *BIO Web of Conferences* 15:01027. DOI: 10.1051/bioconf/20191501027.

22 Gazis, R. and Chaverri, P. (2010). Diversity of fungal endophytes in leaves and stems of wild rubber trees (*Hevea brasiliensis*) in Peru. *Fungal Ecology* 3(3):240–254. DOI: 10.1016/j.funeco.2009.12.001.

23 Rajmohan, N. et al. (2011). Molecular identification and mycotoxin production of *Lilium longiflorum*-associated fusaria isolated from two geographic locations in the United States. *European Journal of Plant Pathology* 131:631–642. DOI: 10.1007/s10658-011-9838-7.

24 NCBI (2021b). *Fusarium proliferatum FpA8*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN05529104/ (visited on 25/03/2021).

25 O'Donnell, K., Sarver, B. A. et al. (2007). Phylogenetic diversity and microsphere array-based genotyping of human pathogenic fusaria, including isolates from the multistate contact lens-associated U.S. keratitis outbreaks of 2005 and 2006. *Journal of Clinical Microbiology* 45(7):2235–2248. DOI: 10.1128/JCM.00533-07.

26 Hill, R. et al. (2021). Seed Banks as Incidental Fungi Banks: Fungal Endophyte Diversity in Stored Seeds of Banana Wild Relatives. *Frontiers in Microbiology* 12:643731. DOI: 10.3389/fmicb.2021.643731.

27 Torbati, M. et al. (2019). Multigene phylogeny reveals new fungicolous species in the *Fusarium tricinctum* species complex and novel hosts in the genus *Fusarium* from Iran. *Mycological Progress* 18:119–133. DOI: 10.1007/s11557-018-1422-5.

28 Wang, C.-Y., Abid, M. et al. (2019). First Report of Silk Tree (*Albizia julibrissin*) Wilt Caused by *Fusarium proliferatum* in Anhui Province of China. *Plant Disease* 103(11). DOI: 10.1094/PDIS-04-19-0876-PDN.

29 Carrieri, R. et al. (2013). *Fusarium proliferatum* and *Fusarium tricinctum* as causal agents of pink rot of onion bulbs and the effect of soil solarization combined with compost amendment in controlling their infections in field. *Crop Protection* 43:31–37. DOI: 10.1016/j.cropro.2012.09.013.

30 Quesada-Ocampo, L. M. et al. (2014). First Report of Fusarium Rot of Garlic Bulbs Caused by *Fusarium proliferatum* in North Carolina. *Plant Disease* 98(7). DOI: 10.1094/PDIS-01-14-0040-PDN.

31 Yamazaki, M. et al. (2013). *Fusarium proliferatum*, an additional bulb rot pathogen of Chinese chive. *Journal of General Plant Pathology* 79:431–434. DOI: 10.1007/s10327-013-0473-3.

32 Punja, Z. K. (2021). First report of *Fusarium proliferatum* causing crown and stem rot, and pith necrosis, in cannabis (*Cannabis sativa* L., marijuana) plants. *Canadian Journal of Plant Pathology* 43(2):236–255. DOI: 10.1080/07060661.2020.1793222.

33 Kim, S. G., Ko, H.-C. et al. (2016). First Report of Fusarium Wilt Caused by *Fusarium proliferatum* on Safflower. *Research in Plant Disease* 22(2):111–115. DOI: 10.5423/RPD.2016.22.2.111.

34 Habibi, A. and Safaiefarahani, B. (2019). First Report of Fusarium Wilt of *Colchicum kotschyi* Caused by *Fusarium proliferatum* in Iran. *Plant Disease* 103(6). DOI: 10.1094/PDIS-11-18-1924-PDN.

35 Liu, N. et al. (2020). First Report of *Fusarium proliferatum* Causing Crown and Stem Rot of *Echeveria desmetiana* in China. *Plant Disease* 104(12):3260. DOI: 10.1094/PDIS-06-20-1232-PDN.

36 Al Mahmooli, I. H. et al. (2013). First Report of Gladiolus Corm Rot Caused by *Fusarium proliferatum* in Oman. *Plant Disease* 97(2). DOI: 10.1094/PDIS-03-12-0233-PDN.

37 Lee, H. B., Kim, C. J. et al. (2011). First Report of Crown Rot on Gypsophila (*Gypsophila paniculata*) Caused by *Fusarium proliferatum* in Korea. *Plant Disease* 95(2). DOI: 10.1094/PDIS-05-10-0376.

38 Feng, F. S. and Li, H. (2019). First Report of *Fusarium proliferatum* Causing Trunk Canker on *Ilex cornuta* in China. *Plant Disease* 103(1). DOI: 10.1094/PDIS-05-18-0779-PDN.

39 Almanza-Álvarez, J. et al. (2017). Identification and control of pathogenic fungi in neotropical valued orchids (*Laelia* spp.) *Tropical Plant Pathology* 42:339–351. DOI: 10.1007/s40858-017-0171-3.

40 Cheng, Y. et al. (2019). *Fusarium* species in declining wild apple forests on the northern slope of the Tian Shan Mountains in north-western China. *Forest Pathology* 49:e12542. DOI: 10.1111/efp.12542.

41 Huang, S. P. et al. (2019). First report of sheath rot caused by *Fusarium proliferatum* on Pisang Awak Banana (*Musa* ABB) in China. *Journal of Plant Pathology* 101:1271–1272. DOI: 10.1007/s42161-019-00329-z.

42 Choi, H. W. et al. (2018). Taxonomy of *Fusarium fujikuroi* species complex associated with bakanae on rice in Korea. *Australasian Plant Pathology* 47:23–34. DOI: 10.1007/s13313-017-0536-6.

43 Zhou, X., Rao, B. et al. (2021). First report of leaf blight caused by *Fusarium proliferatum* on *Polygonatum cyrtonema* in China. *Journal of Plant Pathology* 103:369. DOI: 10.1007/s42161-020-00696-y.

44 Xie, S. H. et al. (2018). *Fusarium proliferatum*: A New Pathogen Causing Fruit Rot of Peach in Ningde, China. *Plant Disease* 102(9). DOI: 10.1094/PDIS-01-18-0103-PDN.

45 Kee, Y. J., Zakaria, L. and Mohd, M. H. (2020). Morphology, phylogeny and pathogenicity of *Fusarium* species from *Sansevieria trifasciata* in Malaysia. *Plant Pathology* 69:442–454. DOI: 10.1111/ppa.13138.

46 Ren, J. et al. (2015). First Report of Sunflower Wilt Caused by *Fusarium proliferatum* in Inner Mongolia, China. *Plant Disease* 99(9). DOI: 10.1094/PDIS-10-14-1081-PDN.

47 Pérez, B. A. et al. (2011). First Report of Root Rot Caused by *Fusarium proliferatum* on Blueberry in Argentina. *Plant Disease* 95(11). DOI: 10.1094/PDIS-04-11-0307.

48 Park, M. S. et al. (2018). Fungal diversity and enzyme activity associated with sailfin sandfish egg masses in Korea. *Fungal Ecology* 34:1–9. DOI: 10.1016/j.funeco.2018.03.004.

49 Ding, S., Hu, H. and Gu, J. D. (2015). Fungi colonizing wood sticks of Chinese fir incubated in subtropical urban soil growing with *Ficus microcarpa* trees. *International Journal of Environmental Science and Technology* 12:3781–3790. DOI: 10.1007/s13762-015-0802-5.

50 Mohammadian, E., Arzanlou, M. and Babai-Ahari, A. (2017). Diversity of culturable fungi inhabiting petroleum-contaminated soils in Southern Iran. *Antonie van Leeuwenhoek* 110:903–923. DOI: 10.1007/s10482-017-0863-1.

51 Tischner, Z. et al. (2019). Environmental characteristics and taxonomy of microscopic fungi isolated from washing machines. *Fungal Biology* 123:650–659. DOI: 10.1016/j.funbio.2019.05.010.

52 NCBI (2021c). *Fusarium anthophilum NRRL 25214*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683615/ (visited on 01/09/2021).

53 Rodrigues, A. A. C. and Menezes, M. (2005). Identification and pathogenic characterization of endophytic *Fusarium* species from cowpea seeds. *Mycopathologia* 159:79–85. DOI: 10.1007/s11046-004-7138-x.

54 Akanmu, A., Abiala, M. and Odebode, A. (2013). Pathogenic Effect of Soilborne *Fusarium* Species on the Growth of Millet Seedlings. *World Journal of Agricultural Sciences* 9(1):60–68. DOI: 10.5829/idosi.wjas.2013.9.1.1721.

55 Sharfun-Nahar and Mushtaq, M. (2006). Pathogenecity and transmission studies of seed-borne *Fusarium* species (sec. Liseola and Sporotrichiella) in sunflower. *Pakistan Journal of Botany* 38(2):487–492. ISSN: 05563321.

56 Jacobs-Venter, A. et al. (2018). Molecular systematics of two sister clades, the *Fusarium concolor* and *F. babinda* species complexes, and the discovery of a novel microcycle macroconidium–producing species from South Africa. *Mycologia* 110(6):1189–1204. DOI: 10.1080/00275514.2018.1526619.

57 NCBI (2021d). *Fusarium austroamericanum NRRL 2903*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683584 (visited on 21/05/2021).

58 Romero Cortes, T. et al. (2016). Aislamiento y caracterización morfológica de diferentes especies de hongos aislados de semillas de plantas de cebada (*Hordeum vulgare*). *Revista de Divulgación Técnica Agrícola y Agroindustrial* 65:1–10.

59 Umpiérrez-Failache, M. et al. (2013). Regional differences in species composition and toxigenic potential among Fusarium head blight isolates from Uruguay indicate a risk of nivalenol contamination in new wheat production areas. *International Journal of Food Microbiology* 166:135–140. DOI: 10.1016/j.ijfoodmicro.2013.06.029.

60 Tralamazza, S. M. et al. (2016). Fungal diversity and natural occurrence of deoxynivalenol and zearalenone in freshly harvested wheat grains from Brazil. *Food Chemistry* 196:445–450. DOI: 10.1016/j.foodchem.2015.09.063.

61 Kokkonen, M. et al. (2010). Mycotoxin production of selected *Fusarium* species at different culture conditions. *International Journal of Food Microbiology* 143:17–25. DOI: 10.1016/j.ijfoodmicro.2010.07.015.

62 Makkonen, J. et al. (2013). *Fusarium avenaceum* causes burn spot disease syndrome in noble crayfish (*Astacus astacus*). *Journal of Invertebrate Pathology* 113:184–190. DOI: 10.1016/j.jip.2013.03.008.

63 Batta, Y. A. (2012). The first report on entomopathogenic effect of *Fusarium avenaceum* (Fries) Saccardo (Hypocreales, Ascomycota) against rice weevil (*Sitophilus oryzae* L.: Curculionidae, Coleoptera). *Journal of Entomological and Acarological Research* 44:e11. DOI: 10.4081/jear.2012.e11.

64 Jankowiak, R. et al. (2016). Fungi associated with dieback of *Abies alba* seedlings in naturally regenerating forest ecosystems. *Fungal Ecology* 24:61–69. DOI: 10.1016/j.funeco.2016.08.013.

65 Rivedal, H. M. et al. (2020). Characterization of the Fungal Community Associated with Root, Crown, and Vascular Symptoms in an Undiagnosed Yield Decline of Winter Squash. *Phytobiomes Journal* 4:178–192. DOI: 10.1094/PBIOMES-11-18-0056-R.

66 Okane, I. and Nakagiri, A. (2015). Assemblages of endophytic fungi on *Salicornia europaea* disjunctively distributed in Japan: towards clarification of the ubiquity of fungal endophytes on halophytes and their ecological roles. *Current Science* 109(1):62–71. DOI: 10.18520/cs/v109/i1/62-71.

67 Corredor, A. H., Van Rees, K. and Vujanovic, V. (2012). Changes in root-associated fungal assemblages within newly established clonal biomass plantations of *Salix* spp. *Forest Ecology and Management* 282:105–114. DOI: 10.1016/j.foreco.2012.06.045.

68 Zhao, Z. et al. (2020). First Report of Fruit Blotch on Kiwifruit Caused by *Fusarium avenaceum* in China. *Plant Disease* 104(5). DOI: 10.1094/PDIS-10-19-2075-PDN.

69 Moslemi, A. et al. (2017). *Fusarium oxysporum* and *Fusarium avenaceum* associated with yield-decline of pyrethrum in Australia. *European Journal of Plant Pathology* 149:43–56. DOI: 10.1007/s10658-017-1161-5.

70 Zhou, Q., Li, N. et al. (2018). Genetic diversity and aggressiveness of *Fusarium* species isolated from soybean in Alberta, Canada. *Crop Protection* 105:49–58. DOI: 10.1016/j.cropro.2017.11.005.

71 Wei, M. et al. (2017). First Report of *Fusarium avenaceum* Causing Root Rot of Maca (*Lepidium meyenii*) in China. *Plant Disease* 101(5). DOI: 10.1094/PDIS-07-16-1061-PDN.

72 Holtz, M. D. et al. (2011). Characterization of *Fusarium avenaceum* from lupin in central Alberta: Genetic diversity, mating type and aggressiveness. *Canadian Journal of Plant Pathology* 33:61–76. DOI: 10.1080/07060661.2011.536651.

73 Feng, J., Hwang, R. et al. (2010). Genetic variation in *Fusarium avenaceum* causing root rot on field pea. *Plant Pathology* 59:845–852. DOI: 10.1111/j.1365-3059.2010.02313.x.

74 Akita, M. et al. (2011). Infection of the Sunagoke moss panels with fungal pathogens hampers sustainable greening in urban environments. *Science of the Total Environment* 409:3166–3173. DOI: 10.1016/j.scitotenv.2011.05.009.

75 Lygis, V. et al. (2014). Fungi in living and dead stems and stumps of *Pinus mugo* on coastal dunes of the Baltic Sea. *Plant Protection Science* 50(4):221–226. DOI: 10.17221/25/2014-pps.

76 Hujslová, M. et al. (2010). Diversity of fungal communities in saline and acidic soils in the Soos National Natural Reserve, Czech Republic. *Mycological Progress* 9:1–15. DOI: 10.1007/s11557-009-0611-7.

77 O'Donnell, K., Cigelnik, E. and Nirenberg, Helgard, I. (1998). Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* 90(3):465–493.

78 Mohamed Nor, N. M., Salleh, B. and Leslie, J. F. (2019). *Fusarium* species from Sorghum in Thailand. *Plant Pathology Journal* 35(4):301–312. DOI: 10.5423/PPJ.OA.03.2019.0049.

79 Laraba, I. et al. (2017). *Fusarium algeriense*, sp. nov., a novel toxigenic crown rot pathogen of durum wheat from Algeria is nested in the *Fusarium burgessii* species complex. *Mycologia* 109(6):935–950. DOI: 10.1080/00275514.2018.1425067.

80 Okello, P. N. et al. (2020). Characterization of species of *Fusarium* causing root rot of Soybean (*Glycine max* L.) in South Dakota, USA. *Canadian Journal of Plant Pathology* 42(4):560–571. DOI: 10.1080/07060661.2020.1746695.

81 Rodrigues, K. F. (1994). The foliar fungal endophytes of the Amazonian palm *Euterpe oleracea*. *Mycologia* 86(3):376–385. URL: https://www.jstor.org/stable/3760568.

82 Summerell, B. A. et al. (2011). *Fusarium* species associated with plants in Australia. *Fungal Diversity* 46:1–27. DOI: 10.1007/s13225-010-0075-8.

83 Shimada, A. et al. (2010). Nematicidal activity of beauvericin produced by the fungus *Fusarium bulbicola*. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 65(3-4):207–210. DOI: 10.1515/znc-2010-3-407.

84 Swett, C. L. and Gordon, T. R. (2015). Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecology* 13:120–129. DOI: 10.1016/j.funeco.2014.09.003.

85 Isaac, M. R. et al. (2018). Occurrence, identification, and pathogenicity of *Fusarium* spp. associated with tomato wilt in Mexico. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 46(2):484–493. DOI: 10.15835/nbha46211095.

86 NCBI (2021e). *Fusarium coffeatum FIESC_ 28*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN08667433/ (visited on 19/05/2021).

87 Ferreira, M. C., de Assis, J. C. and Rosa, L. H. (2020). Diversity of endophytic fungi associated with *Carapichea ipecacuanha* from a native fragment of the Atlantic Rain Forest. *South African Journal of Botany* 134:225–229. DOI: 10.1016/j.sajb.2019.12.031.

88 Gerlach, W. (1977). Drei neue Varietäten von *Fusarium merismoides*, *F. larvarum* und *F. chlamydosporum*. *Journal of Phytopathology* 90:31–42. DOI: 10.1111/j.1439-0434.1977.tb02883.x.

89 Laurence, M. H. et al. (2016). Six novel species of *Fusarium* from natural ecosystems in Australia. *Fungal Diversity* 77(1):349–366. DOI: 10.1007/s13225-015-0337-6.

90 de Boer, W. et al. (1997). Anti-fungal properties of chitinolytic dune soil bacteria. *Soil Biology and Biochemistry* 30(2):193–203. DOI: 10.1016/S0038-0717(97)00100-4.

91 Bolaños, J. et al. (2015). Phylogenetic Diversity of Sponge-Associated Fungi from the Caribbean and the Pacific of Panama and Their In Vitro Effect on Angiotensin and Endothelin Receptors. *Marine Biotechnology* 17:533–564. DOI: 10.1007/s10126-015-9634-z.

92 Juybari, H. Z. et al. (2019). Seasonal, tissue and age influences on frequency and biodiversity of endophytic fungi of *Citrus sinensis* in Iran. *Forest Pathology* 49:e12559. DOI: 10.1111/efp.12559.

93 Rodriguez, R. J. et al. (2008). Stress tolerance in plants via habitat-adapted symbiosis. *ISME Journal* 2(4):404–416. DOI: 10.1038/ismej.2007.106.

94 Grunden, E., Chen, W. and Crane, J. L. (2001). Fungi colonizing microsclerotia of *Verticillium dahliae* in urban environments. *Fungal Diversity* 8:129–141. ISSN: 15602745.

95 de Rooij-van der Goes, P. C., Putten, W. H. van der and Dijk, C. van (1995). Analysis of nematodes and soil-borne fungi from *Ammophila arenaria* (Marram grass) in Dutch coastal foredunes by multivariate techniques. *European Journal of Plant Pathology* 101(2):149–162. DOI: 10.1007/BF01874761.

96 Chen, Y., Zhou, Q. et al. (2014). Genetic diversity and aggressiveness of *Fusarium* spp. isolated from Canola in Alberta, Canada. *Plant Disease* 98:727–738. DOI: 10.1094/PDIS-01-13-0061-RE.

97 Backhouse, D. et al. (2004). Survey of *Fusarium* species associated with crown rot of wheat and barley in eastern Australia. *Australasian Plant Pathology* 33:255–261. DOI: 10.1071/AP04010.

98 Imathiu, S. M. et al. (2010). Evaluation of pathogenicity and aggressiveness of *F. langsethiae* on oat and wheat seedlings relative to known seedling blight pathogens. *European Journal of Plant Pathology* 126:203–216. DOI: 10.1007/s10658-009-9533-0.

99 Stefańczyk, E. et al. (2016). Diversity of *Fusarium* spp. associated with dry rot of potato tubers in Poland. *European Journal of Plant Pathology* 145:871–884. DOI: 10.1007/s10658-016-0875-0.

100 Khan, S. A. et al. (2020). Molecular diversity of halophilic fungi isolated from mangroves ecosystem of miani hor, Balochistan, Pakistan. *Pakistan Journal of Botany* 52(5):1823–1829. DOI: 10.30848/PJB2020-5(34).

101 Potshangbam, M. et al. (2017). Functional characterization of endophytic fungal community associated with *Oryza sativa* L. and *Zea mays* L. *Frontiers in Microbiology* 8:325. DOI: 10.3389/fmicb.2017.00325.

102 Proctor, R. H. et al. (2018). Evolution of structural diversity of trichothecenes, a family of toxins produced by plant pathogenic and entomopathogenic fungi. *PLoS Pathogens* 14(4):e1006946. DOI: 10.1371/journal.ppat.1006946.

103 NCBI (2021f). *Fusarium flagelliforme NRRL 13405*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN08631343/ (visited on 20/05/2021).

104 Xia, J. et al. (2019). Numbers to names – restyling the *Fusarium incarnatum-equiseti* species complex. *Persoonia* 43(1):186–221. DOI: 10.3767/persoonia.2019.43.05.

105 O'Donnell, K., Sutton, D. A. et al. (2009). Novel multilocus sequence typing scheme reveals high genetic diversity of human pathogenic members of the *Fusarium incarnatum-F. equiseti* and *F. chlamydosporum* species complexes within the United States. *Journal of Clinical Microbiology* 47(12):3851–3861. DOI: 10.1128/JCM.01616-09.

106 CABI (2021). *GRC Catalogue*. URL: https://www.cabi.org/services/microbial-services/culture-collection-microorganism-supply/grc/ (visited on 03/09/2021).

107 Wiemann, P. 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 Pathogens* 9(6):e1003475. DOI: 10.1371/journal.ppat.1003475.

108 Nisa, S. et al. (2020). Identification and Bioactivities of Two Endophytic Fungi *Fusarium fujikuroi* and *Aspergillus tubingensis* from Foliar Parts of *Debregeasia salicifolia*. *Arabian Journal for Science and Engineering* 45(6):4477–4487. DOI: 10.1007/s13369-020-04454-1.

109 Chang, X. et al. (2018). Identification of *Fusarium* species associated with soybean root rot in Sichuan Province, China. *European Journal of Plant Pathology* 151:563–577. DOI: /10.1007/s10658-017-1410-7.

110 Mazarotto, E. J. et al. (2020). Pathogenic *Fusarium* species complexes associated to seeds of indigenous Brazilian forest tree *Aspidosperma polyneuron*. *European Journal of Plant Pathology* 158:849–857. DOI: 10.1007/s10658-020-02120-8.

111 Chen, J., Zhong, J. et al. (2019). First Report of *Fusarium fujikuroi* Causing Black Rot of *Bletilla striata* (Baiji) in China. *Plant Disease* 103(2). DOI: 10.1094/PDIS-05-18-0781-PDN.

112 Jiang, S. B. et al. (2018). First Report of *Fusarium fujikuroi* Causing Stem Wilt on *Canna edulis* Ker in China. *Plant Disease* 102(6). DOI: 10.1094/PDIS-09-17-1479-PDN.

113 Kim, B. R. and Choi, Y. J. (2021). *Fusarium fujikuroi* Causing Fusarium Wilt of *Lactuca serriola* in Korea. *Plant Disease* 105(2):502. DOI: 10.1094/PDIS-06-20-1370-PDN.

114 Shen, Y. N. et al. (2020). First Report of Leaf Spot on *Lasia spinosa* Caused by *Fusarium fujikuroi* in China. *Plant Disease* 104(9). DOI: 10.1094/PDIS-01-20-0013-PDN.

115 Long, H. et al. (2021). First Report of Fruit Blotch on Plum Caused by *Fusarium fujikuroi* in China. *Plant Disease*:1–2. DOI: 10.1094/PDIS-08-20-1784-PDN.

116 Awan, U. A. et al. (2021). Isolation, fermentation, and formulation of entomopathogenic fungi virulent against adults of *Diaphorina citri*. *Pest Management Science* 77:4040–4053. DOI: 10.1002/ps.6429.

117 Phan, H. T. et al. (2004). *Gibberella gaditjirrii* (*Fusarium gaditjirrii*) sp. nov., a new species from tropical grasses in Australia. *Studies in Mycology* 50:261–272. ISSN: 01660616.

118 Rheeder, J. P., Marasas, W. F. and Nelson, P. E. (1996). *Fusarium globosum*, a new species from corn in southern Africa. *Mycologia* 88(3):509–513. DOI: 10.2307/3760891.

119 Heydari-Nezhad, A. M. et al. (2014). First report of a disease caused by *Fusarium globosum* on giant cane in Iran. *Journal of Plant Pathology* 96(4):S4.129. DOI: 10.4454/JPP.V96I4.035.

120 Gagkaeva, T. Y., Gavrilova, O. P. and Orina, A. S. (2019). First report of *Fusarium globosum* associated with barley grain in the southwestern part of Siberia. *Plant Disease* 103(3):588. DOI: 10.1094/PDIS-06-18-1108-PDN.

121 Trail, F. and Common, R. (2000). Perithecial development by *Gibberella zeae*: a light microscopy study. *Mycologia* 92(1):130–138. DOI: 10.1080/00275514.2000.12061137.

122 Akbar, A. et al. (2018). Detection, virulence and genetic diversity of *Fusarium* species infecting tomato in Northern Pakistan. *PLoS ONE* 13(9):e0203613. DOI: 10.1371/journal.pone.0203613.

123 Lee, S., Park, M. S., Lee, H., Kim, J. J. et al. (2019). Fungal Diversity and Enzyme Activity Associated with the Macroalgae, *Agarum clathratum*. *Mycobiology* 47(1):50–58. DOI: 10.1080/12298093.2019.1580464.

124 Starkey, D. E. et al. (2007). Global molecular surveillance reveals novel Fusarium head blight species and trichothecene toxin diversity. *Fungal Genetics and Biology* 44(11):1191–1204. DOI: 10.1016/j.fgb.2007.03.001.

125 Goswami, R. S. and Kistler, H. C. (2004). Heading for disaster: *Fusarium graminearum* on cereal crops. *Molecular Plant Pathology* 5(6):515–525. DOI: 10.1111/J.1364-3703.2004.00252.X.

126 Scruggs, A. C. and Quesada-Ocampo, L. M. (2016). Etiology and epidemiological conditions promoting fusarium root rot in sweetpotato. *Phytopathology* 106(8):909–919. DOI: 10.1094/PHYTO-01-16-0009-R.

127 Pak, D. et al. (2017). Reservoir of cultivated rice pathogens in wild rice in Australia. *European Journal of Plant Pathology* 147:295–311. DOI: 10.1007/s10658-016-1002-y.

128 Li, Z. Y. et al. (2015). First Report of *Fusarium graminearum* Causing Ear Rot of Foxtail Millet in China. *Plant Disease* 99(12). DOI: 10.1094/PDIS-11-14-1179-PDN.

129 CBS-KNAW (2021). *Fusarium heterosporum*. URL: https://wi.knaw.nl/page/fungal%7B%5C_%7Dtable (visited on 25/03/2021).

130 Ali, H., Backhouse, D. and Burgess, L. W. (1996). *Fusarium heterosporum* associated with paspalum ergot in eastern Australia. *Australasian Plant Pathology* 25(2):120–125. DOI: 10.1071/ap96020.

131 Torp, M. and Adler, A. (2004). The European *Sporotrichiella* project: a polyphasic approach to the biology of a new *Fusarium* species. *International Journal of Food Microbiology* 95:241–245. DOI: 10.1016/j.ijfoodmicro.2003.12.015.

132 NCBI (2021g). *Fusarium longipes NRRL 20695*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN08631279/ (visited on 21/05/2021).

133 Maryani, N. et al. (2019). New endemic *Fusarium* species hitch-hiking with pathogenic *Fusarium* strains causing Panama disease in small-holder banana plots in Indonesia. *Persoonia: Molecular Phylogeny and Evolution of Fungi* 43:48–69. DOI: 10.3767/persoonia.2019.43.02.

134 Navi, S. S. and Singh, S. D. (1993). *Fusarium longipes*: a mycoparasite of *Sclerospora graminicola* on pearl millet. *Indian Phytopathology* 46(4):365–368. ISSN: 2248-9800.

135 Besharati Fard, M., Mohammadi, A. and Darvishnia, M. (2017). *Fusarium* species associated with Wheat crown and root tissues in the Eastern Iran. *Archives of Phytopathology and Plant Protection* 50(3-4):123–133. DOI: 10.1080/03235408.2016.1275423.

136 Britz, H. et al. (2002). Two new species of *Fusarium* section *Liseola* associated with mango malformation. *Mycologia* 94(4):722–730. DOI: 10.1080/15572536.2003.11833199.

137 Otero-Colina, G. et al. (2010). Identification and characterization of a novel etiological agent of mango malformation disease in Mexico, *Fusarium mexicanum* sp. nov. *Phytopathology* 100(11):1176–1184. DOI: 10.1094/PHYTO-01-10-0029.

138 Santillán-Mendoza, R. et al. (2018). A novel disease of big-leaf mahogany caused by two *Fusarium* species in Mexico. *Plant Disease* 102(10):1965–1972. DOI: 10.1094/PDIS-01-18-0060-RE.

139 Al Yazidi, L. S. et al. (2019). Endobronchial fusariosis in a child following bilateral lung transplant. *Medical Mycology Case Reports* 23:77–80. DOI: 10.1016/j.mmcr.2019.01.002.

140 Marasas, W. F. O. et al. (1987). *Fusarium napiforme*, a New Species from Millet and Sorghum in Southern Africa. *Mycologia* 79(6):910. DOI: 10.2307/3807697.

141 Nelson, P. E., Plattner, R. D. et al. (1992). Fumonisin B1 Production by *Fusarium* Species Other Than *F. moniliforme* in Section *Liseola* and by Some Related Species. *Applied and Environmental Microbiology* 58(3):984–989. DOI: 10.1128/aem.58.3.984-989.1992.

142 Supratman, U. et al. (2019). New naphthoquinone derivatives from *Fusarium napiforme* of a mangrove plant. *Natural Product Research*. DOI: 10.1080/14786419.2019.1650358.

143 Nucci, M. and Anaissie, E. (2007). *Fusarium* Infections in Immunocompromised Patients. *Clinical Microbiology Reviews* 20(4):695–704. DOI: 10.1128/CMR.00014-07.

144 NCBI (2021h). *Fusarium nygamai*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN06240347/ (visited on 25/03/2021).

145 Bogner, C. W. et al. (2016). Fungal root endophytes of tomato from Kenya and their nematode biocontrol potential. *Mycological Progress* 15:30. DOI: 10.1007/s11557-016-1169-9.

146 Krulder, J. W. M. et al. (1996). Systemic *Fusarium nygamai* infection in a patient with lymphoblastic non-Hodgkin's lymphoma. *Mycoses* 39:121–123. DOI: 10.1111/j.1439-0507.1996.tb00113.x.

147 Burgess, L. and Trimboli, D. (1986). Characterization and Distribution of *Fusarium nygamai*, sp. nov. *Mycologia* 78(2):223–229. DOI: 10.2307/3793167.

148 Balmas, V. et al. (2000). *Fusarium nygamai* Associated with Fusarium Foot Rot of Rice in Sardinia. *Plant Disease* 84(7). DOI: 10.1094/PDIS.2000.84.7.807B.

149 Leslie, J. F. et al. (2005). Toxicity, pathogenicity, and genetic differentiation of five species of *Fusarium* from sorghum and millet. *Phytopathology* 95(3):275–283. DOI: 10.1094/PHYTO-95-0275.

150 Azil, N. et al. (2021). Identification and pathogenicity of *Fusarium* spp. associated with tuber dry rot and wilt of potato in Algeria. *European Journal of Plant Pathology* 159:495–509. DOI: 10.1007/s10658-020-02177-5.

151 Sauerborn, J. et al. (1996). '*Striga hermonthica* control with *Fusarium nygamai* in maize'. In: *Proceedings of the IX International Symposium on Biological Control of Weeds*. Ed. by V. Moran and J. Hoffmann. Stellenbosch, South Africa:pp. 461–466.

152 Guo, L. et al. (2014). Genome and transcriptome analysis of the fungal pathogen *Fusarium oxysporum* f. sp. *cubense* causing banana vascular wilt disease. *PLoS ONE* 9(4):e95543. DOI: 10.1371/journal.pone.0095543.

153 Delulio, G. A. et al. (2018). Kinome Expansion in the *Fusarium oxysporum* Species Complex Driven by Accessory Chromosomes. *mSphere* 3:e00231–18.

154 NCBI (2021i). *Fusarium oxysporum f. sp. cepae FoC_ Fus2*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN05529097/ (visited on 20/05/2021).

155 Armitage, A. D. et al. (2018). Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports* 8:13530. DOI: 10.1038/s41598-018-30335-7.

156 Abawi, G. S. and Lorbeer, J. W. (1972). Several Aspects of the Ecology and Pathology of *Fusarium oxysporum* f. sp. *cepae*. *Phytopathology* 62:870. DOI: 10.1094/phyto-62-870.

157 Leoni, C. et al. (2013). *Fusarium oxysporum* f.sp. *cepae* dynamics: in-plant multiplication and crop sequence simulations. *European Journal of Plant Pathology* 137:545–561. DOI: 10.1007/s10658-013-0268-6.

158 Thatcher, L. F. et al. (2012). A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Molecular Plant-Microbe Interactions* 25(2):180–190. DOI: 10.1094/MPMI-08-11-0212.

159 Ma, L. J. et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464(7287):367–373. DOI: 10.1038/nature08850.

160 Menzies, J. G., Koch, C. and Seywerd, F. (1990). Additions to the Host Range of *Fusarium oxysporum* f. sp. *radicis-lycopersici*. *Plant Disease* 74:569–572. DOI: 10.1094/PD-74-0569.

161 Kishi, K., Furukawa, T. and Aoki, T. (1999). Purple Spot of Aloe (*Aloe arborescens* Mill.) Caused by *Fusarium phyllophilum* Nirenberg et O'Donnell (New Disease). *Japanese Journal of Phytopathology* 65:576–587. DOI: 10.3186/jjphytopath.65.576.

162 Vanheule, A. et al. (2016). Living apart together: crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC Genomics* 17:670. DOI: 10.1186/s12864-016-2941-6.

163 Stenglein, A. S. A. (2009). *Fusarium poae*: a pathogen that needs more attention. *Journal of Plant Pathology* 91(1):25–36. URL: https://www.jstor.org/stable/41998571.

164 Amobonye, A. et al. (2021). Characterisation, pathogenicity and hydrolytic enzyme profiling of selected *Fusarium* species and their inhibition by novel coumarins. *Archives of Microbiology*. DOI: 10.1007/s00203-021-02335-1.

165 Tupaki-Sreepurna, A. et al. (2018). Phylogenetic Diversity and *In Vitro* Susceptibility Profiles of Human Pathogenic Members of the *Fusarium fujikuroi* Species Complex Isolated from South India. *Mycopathologia* 183:529–540. DOI: 10.1007/s11046-018-0248-7.

166 Cavalcanti, A. D. et al. (2020). *Fusarium massalimae* sp. nov. (*F. lateritium* species complex) occurs endophytically in leaves of *Handroanthus chrysotrichus*. *Mycological Progress* 19:1133–1142. DOI: 10.1007/s11557-020-01622-3.

167 Nicolli, C. P. et al. (2020). *Fusarium fujikuroi* species complex in Brazilian rice: Unveiling increased phylogenetic diversity and toxigenic potential. *International Journal of Food Microbiology* 330:108667. DOI: 10.1016/j.ijfoodmicro.2020.108667.

168 Shiraishi, A. et al. (2012). AFLP, Pathogenicity, and VCG Analyses of *Fusarium oxysporum* and *Fusarium pseudocircinatum* from *Acacia koa*. *Plant Disease* 96:1111–1117. DOI: 10.1094/PDIS-06-11-0491.

169 Gardiner, D. M. et al. (2012). Comparative Pathogenomics Reveals Horizontally Acquired Novel Virulence Genes in Fungi Infecting Cereal Hosts. *PLoS Pathogens* 8(9):e1002952. DOI: 10.1371/journal.ppat.1002952.

170 Aoki, T. and O'Donnell, K. (1999). Morphological and molecular characterization of *Fusarium pseudograminearum* sp. nov., formerly recognized as the Group 1 population of *F. graminearum*. *Mycologia* 91(4):597–609. DOI: 10.1080/00275514.1999.12061058.

171 NCBI (2021j). *Fusarium sarcochroum NRRL 20472*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683633/ (visited on 20/05/2021).

172 Sandoval-Denis, M. et al. (2018). Symptomatic citrus trees reveal a new pathogenic lineage in *Fusarium* and two new *Neocosmospora* species. *Persoonia* 40:1–25. DOI: 10.3767/persoonia.2018.40.01.

173 NCBI (2021k). *Fusarium sporotrichioides NRRL 3299*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN08631227/ (visited on 01/09/2021).

174 Moya-Elizondo, E. A. et al. (2013). First Report of *Fusarium sporotrichioides* Causing Foliar Spots on Forage Corn in Chile. *Plant Disease* 97(8). DOI: 10.1094/PDIS-12-12-1120-PDN.

175 NCBI (2021l). *Fusarium subglutinans NRRL 66333*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683616/ (visited on 01/09/2021).

176 Ruan, R. et al. (2021). First report of *Fusarium subglutinans* causing heart rot on *Cymbidium hybridum* in China. *Crop Protection* 144:105603. DOI: 10.1016/j.cropro.2021.105603.

177 ATCC (2021). *Fusarium venenatum Nirenberg (ATCC® 20334™)*. URL: https://www.lgcstandards-atcc.org/products/all/20334.aspx?geo%7B%5C_%7Dcountry=gb%7B%5C#%7Dhistory (visited on 21/05/2021).

178 Tan, D. C. et al. (2011). Mycotoxins produced by *Fusarium* species associated with annual legume pastures and 'sheep feed refusal disorders' in Western Australia. *Mycotoxin Research* 27:123–135. DOI: 10.1007/s12550-010-0085-0.

179 Mycocosm (2021). *Fusarium verticillioides*. URL: https://mycocosm.jgi.doe.gov/Fusve2/Fusve2.home.html (visited on 25/03/2021).

180 Salem, N. M. et al. (2020). First Report of *Fusarium verticillioides* Causing Banana Fruit Rot in Jordan. *Plant Disease* 104(12):3255. DOI: 10.1094/PDIS-05-20-1116-PDN.

181 Hilton, A. et al. (2017). Identification and characterization of pathogenic and endophytic fungal species associated with pokkah boeng disease of sugarcane. *Plant Pathology Journal* 33(3):238–248. DOI: 10.5423/PPJ.OA.02.2017.0029.

182 Marín, P., Palmero, D. and Jurado, M. (2015). Occurrence of moulds associated with ovine raw milk and cheeses of the Spanish region of Castilla La Mancha. *International Journal of Dairy Technology* 68(4):565–572. DOI: 10.1111/1471-0307.12208.

183 NCBI (2021m). *Fusarium zealandicum NRRL 22465*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN13683635/ (visited on 20/05/2021).

184 Schroers, H. J. et al. (2011). A revision of *Cyanonectria* and *Geejayessia* gen. nov., and related species with *Fusarium*-like anamorphs. *Studies in Mycology* 68:115–138. DOI: 10.3114/sim.2011.68.05.

185 Liao, H. L. et al. (2019). Fungal endophytes of *Populus trichocarpa* alter host phenotype, gene expression, and rhizobiome composition. *MPMI* 32(7):853–864. DOI: 10.1094/MPMI-05-18-0133-R.

186 Short, D. P. et al. (2017). PCR multiplexes discriminate *Fusarium* symbionts of invasive *Euwallacea* ambrosia beetles that inflict damage on numerous tree species throughout the United States. *Plant Disease* 101:233–240. DOI: 10.1094/PDIS-07-16-1046-RE.

187 O'Donnell, K., Sink, S. et al. (2015). Discordant phylogenies suggest repeated host shifts in the *Fusarium-Euwallacea* ambrosia beetle mutualism. *Fungal Genetics and Biology* 82:277–290. DOI: 10.1016/j.fgb.2014.10.014.

188 Freeman, S. et al. (2013). *Fusarium euwallaceae* sp. nov.-a symbiotic fungus of *Euwallacea* sp., an invasive ambrosia beetle in Israel and California. *Mycologia* 105(6):1595–1606. DOI: 10.3852/13-066.

189 Eskalen, A. et al. (2013). Host range of Fusarium dieback and its ambrosia beetle (Coleoptera: Scolytinae) vector in southern California. *Plant Disease* 97(7):938–951. DOI: 10.1094/PDIS-11-12-1026-RE.

190 Aoki, T., Smith, J. A. et al. (2019). Three novel Ambrosia *Fusarium* Clade species producing clavate macroconidia known (*F. floridanum* and *F. obliquiseptatum*) or predicted (*F. tuaranense*) to be farmed by *Euwallacea* spp. (Coleoptera: Scolytinae) on woody hosts. *Mycologia* 111(6):919–935. DOI: 10.1080/00275514.2019.1647074.

191 NCBI (2021n). *Fusarium kuroshium UCR3666*. URL: https://www.ncbi.nlm.nih.gov/biosample/SAMN07200645 (visited on 20/05/2021).

192 Na, F. et al. (2018). Two novel fungal symbionts *Fusarium kuroshium* sp. nov. and *Graphium kuroshium* sp. nov. of kuroshio shot hole borer (*Euwallacea* sp. nr. *fornicatus*) cause fusarium dieback on woody host species in California. *Plant Disease* 102(6):1154–1164. DOI: 10.1094/PDIS-07-17-1042-RE.

193 Coleman, J. J. et al. (2009). The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* 5(8):e1000618. DOI: 10.1371/journal.pgen.1000618.

194 Šišić, A. et al. (2018). The 'forma specialis' issue in *Fusarium*: A case study in *Fusarium solani* f. sp. *pisi*. *Scientific Reports* 8:1252. DOI: 10.1038/s41598-018-19779-z.

195 Aoki, T., Liyanage, P. N. et al. (2021). Three novel Ambrosia *Fusarium* Clade species producing multiseptate "dolphin-shaped" conidia, and an augmented description of *Fusarium kuroshium*. *Mycologia* 113(5):1089–1109. DOI: 10.1080/00275514.2021.1923300.