Arabidopsis thaliana root-associated fungi: function, evolution and genomic signatures

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Fantin Mesny
aus Pertuis - France

Köln
2022

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Pflanze-Mikroben Interaktionen (Direktor: Prof. Dr. Paul Schulze- Lefert), in der Arbeitsgruppe von Dr. Stéphane Hacquard angefertigt.

Berichterstatter*in: Prof. Dr. Bart P.H.J. Thomma
Prof. Dr. Eva H. Stukenbrock
Prüfungsvorsitzende: Prof. Dr. Alga Zuccaro
Tag der mündlichen Prüfung: 01.12.2022

Contents

Main figures

List of abbreviations

ANOVA: Analysis Of Variance
ASV: Amplicon Sequence Variant
***At*:** *Arabidopsis thaliana*
CAZyme: Carbohydrate-active enzyme
CCS: Circular Consensus Sequencing
CLR: Continuous Long Reads
COG: Clusters of Orthologous Genes (database)
CSEP: Candidate Secreted Effector Proteins
DNA: Deoxyribonucleic Acid
dbRDA: Distance-based Redundancy Analysis
ECM: Ectomycorrhiza
EF: Endophytic Fungi
ERM: Ericoid Mycorrhiza
ET: Ethylene
ETI: Effector-Triggered Immunity
FCWDE: Fungal Cell Wall-Degrading Enzyme
FDR: False Discovery Rate
GO: Gene Ontology
ITS: Internal Transcribed Spacer
JA: Jasmonic Acid
KOG: Eukaryotic Orthologous Group
log2FC: log2-tranformed Fold Change
MAMP: Microbe-Associated Molecular Pattern
MCL: Markov Clustering
MTI: MAMP-Triggered Immunity
MyM: Mycobiota Member
NCBI: National Center for Biotechnology Information
NLR: Nucleotide-binding domain Leucine-rich Repeat containing (receptor)
OMF: Orchid Mycorrhizal Fungi
PacBio: Pacific Biosciences (company)
PC: Principal Component
PCR: Polymerase Chain Reaction
PCWDE: Plant Cell Wall-Degrading Enzyme
PERMANOVA: Permutational Multivariate Analysis Of Variance

PGA: Potato Growth Agar
Pi: Orthophosphate
PPF: Plant-Pathogenic Fungi
PPI: Plant Performance Index
qPCR: quantitative Polymerase Chain Reaction
RA: Relative Abundance
rDNA: Ribosomal DNA
RNA: Ribonucleic Acid
RNA-Seq: RNA Sequencing
RPKM: Reads per kilobase of transcript per Million mapped reads
SA: Salicylic Acid
SAP: Saprotroph
SFW: Shoot Fresh Weight
SMRT: Single Molecule Real Time (sequencing technology)
SSP: Small Secreted Protein
SVM-RFE: Support Vector Machines with Recursive Feature Elimination
TSB: Tryptic Soy Broth
TukeyHSD: Tukey Honest Significant Difference (statistical test)
UPGMA: Unweighted Pair Group Method with Arithmetic mean (clustering)
WT: Wild-Type
ZMW: Zeromode Waveguide

Abstract

While mycorrhizae have been studied for decades, little is known about fungi that do not establish symbiotic structures, but have the ability to colonize roots of asymptomatic plants in nature. The non-mycorrhizal model plant *Arabidopsis thaliana* hosts in its roots diverse fungal communities, that were previously reported to negatively impact its health in absence of protection from bacterial commensals and innate immune responses. With this thesis, I aimed to better characterize this detrimental mycobiota, focusing on its function, evolution and genomic signatures. While multiple pieces of evidence pointed to the endophytism of *A. thaliana* mycobiota members, predicted evolutionary histories revealed that most of these fungi derived from pathogenic ancestors. Re-colonization experiments with individual strains highlighted diverse fungal effects on plant performance, spanning along the mutualist-pathogenic continuum. This gradient of effects was correlated to fungal root colonization efficiency, highlighting that fungi with detrimental effects dominate in natural root samples. We showed that pectin-degrading enzymes from family PL1_7 contribute in the aggressiveness of endophytic colonization. While further genomic and transcriptomic analyses corroborated the major role of carbohydrate-active enzymes in root endophytic colonization, intra-species comparative genomics of highly prevalent mycobiota member *Plectosphaerella cucumerina* revealed a genomic architecture favoring the fast evolution of effector-encoding genes. We notably identified in this species a candidate genomic region predicted to be involved in fungal adaptation to *A. thaliana*. Taken together, the results compiled in this thesis offer a better understanding of the fine line between endophytism and parasitism in the root mycobiota. They show that fungi robustly colonizing *A. thaliana* roots in nature rely on carbohydrate-active enzymes to degrade host cell walls, and likely on effectors to overcome innate immune responses.

1 Introduction

1.1 Plant-associated microorganisms

In nature, a wide diversity of micro-organisms inhabit above- and belowground plant tissues, thereby influencing the growth, health and resilience of their hosts [1]. This community - referred to as the *plant microbiota* - comprises prokaryotes (i.e. bacteria, archaea) and eukaryotes (i.e. fungi and protists), that mostly rely on plant-derived carbon compounds for nutrition. A theory is, that these microorganisms co-evolved with their host, consequently driving the system to an equilibrium (the *holobiont*) that may eventually benefit the plant [2]. While this concept remains to be conclusively demonstrated, complex plant-microbe and microbe-microbe interactions, together with environmental cues, have been shown to structure the plant microbiota and influence its composition.

1.2 Factors influencing the structure and composition of plant microbiota

1.2.1 Availability of carbon compounds

Comparison of microbiota composition across multiple hosts highlighted that different plant species host different microbial taxa [3]. Indeed, the composition of both the leaf and root microbiota of sugarcane (*Saccharum sp.*), poplar trees (genus *Populus*) and the model plant *Arabidopsis thaliana* differ substantially. One of the main factors driving microbiota differentiation is certainly host cell wall composition, which varies extensively between monocots like sugarcane (cell walls rich in xylans [4]), Brassicaceae like *A. thaliana* (rich in pectin [5]) and poplar trees (highly lignified tissues [6]). Plant microbiota members are believed to be metabolically adapted to the utilization of specific plant-derived one-carbon compounds and cell wall components [7]. Microbes encode arsenals of carbohydrate-degrading enzymes and metabolic pathways that define their ability to feed on such compounds, and likely contribute in their host preference. Genetic modification of rice plants could actually show that accumulation of cellulose in leaves reduced the abundance of specific microbial taxa, causing a shift in community composition [8]. The identity and availability of carbon compounds also modulate the community composition in different niches. For instance, the microbiota of different plants consistently differs between the rhizosphere containing soil-derived carbon compounds and secreted plant metabolites, and the root endosphere, where microbes can access plant cell walls and intracellular metabolites [2].

1.2.2 Host immunity

Plant microbiota structure and composition are also influenced by the host innate immune system. Evasion or suppression of immune outputs is essential for microorganisms to colonize plant tissues [2]. It has been proposed that plants use their immune system as a microbial management system to control microbiota assembly and host-microbe homeostasis [9].

The plant immune system is constituted by two layers [9]. In a first layer of defence, plasma membrane receptors perceive the presence of generally conserved microbial extracellular molecules such as bacterial lipopolysaccharides or fungal chitin, referred to as Microbe-Associated Molecular Patterns (MAMPs). Recognition of MAMPs leads to the activation of a broad set of immune responses (MAMP-triggered immunity; MTI) limiting microbial proliferation. However, adapted microbes have evolved strategies to evade and suppress MTI, notably through the secretion of effector molecules. To face such effector-based strategies, plants deploy a second layer of receptors, belonging to the Nucleotide-binding domain Leucine-rich Repeat containing (NLR) family. Recognition of specific effector proteins by NLR receptors activates the Effector-Triggered Immunity (ETI): a robust response that often includes localized host cell death and systemic defense signaling.

While this general organization of the plant immune system has been deciphered in the context of pathogen infection, there is now evidence for the participation of plant immunity in microbiota assembly. Some commensal and beneficial microbes were reported to evade MTI, by modifying the epitopes of MAMP-containing molecules, repressing their biosynthesis, or altering their cell wall composition [10]. However, some microbiota members still activate this first layer of the plant immune system. Different studies highlighted that MTI responses are necessary to control microbial growth, thereby preventing detrimental over-proliferation of microbiota members and maintaining plant health [11, 12]. Additionally, phytohormone-mediated immunity was shown to significantly shape plant microbiota composition. Indeed, *A. thaliana* mutants impaired in signaling pathways mediated by salicylic acid (SA) - and to a lesser extent by jasmonic acid (JA) and ethylene (ET) - showed abnormal root-associated microbial communities [13]. Unexpectedly, several bacterial taxa were actually depleted in the roots of immunocompromised plants, revealing that some microbes use SA as a carbon source or growth signal. Beneficial bacteria *Bacillus velezensis* were also shown to counteract MTI, as reactive oxygen species produced during *A. thaliana* root colonization stimulate bacterial production of auxin, promoting bacterial survival and efficient root colonization, but also plant defense against fungal pathogens [14]. Plants can therefore positively or negatively modulate the growth of specific microbiota members, using notably phytohormones and reactive oxygen species.

As pathogens, commensal microbiota members evolved strategies to suppress MTI responses, involving notably the secretion of effector molecules [9]. While such strategies were first iden-

tified by studies of individual microbes [9, 15, 16], an *A. thaliana* root-associated community of commensal bacteria was recently shown to comprise taxonomically diverse MTI suppressor strains with efficient root colonization abilities [17]. In this context, bacterial type 2 secretion systems were demonstrated to be essential for MTI suppression, pointing to extracellular mechanisms likely involving effector molecules. Recent studies linking natural genetic variation of barley and sorghum to their rhizosphere microbiota could significantly associate community structure to the presence of specific NLR-encoding genes [18, 19]. While these candidates remain to be functionally validated, these results suggest an involvement of host ETI in microbiota assembly, and a possible NLR-mediated selection of community members.

1.2.3 Environmental conditions

Soil represents the main inoculum source of root microbiota. In contrast, aboveground communities show high variability, as they also acquire microbes from aerosols, insects, pollen and/or migration via other plant tissues [2]. Soil properties (including nutrient availability), climate conditions and surrounding biodiversity therefore constitute three important environmental parameters, that influence microbiota establishment and composition.

Soil properties influence the assembly of both above- and below-ground plant microbiota. pH is considered the *master soil variable* and shapes nutrient availability by influencing soil physical properties as well as biological and chemical processes [20]. The availability of macronutrients is critical for plant health, and was shown to influence microbiota structure. Nitrogen availability in soil was linked to changes in the microbiota composition of different plant species [7]. Soil phosphorus content also influences leaf and root microbiota composition, but this impact was demonstrated to be more importantly due to plant response to nutrient stress (phosphate starvation response) than to phosphorous deficiency itself [21]. Availability of micronutrients in soil also influences plant microbiota composition. For instance, under limiting iron concentration, *A. thaliana* was shown to reshape its root microbiota by secreting coumarins that recruit beneficial microbes to alleviate iron stress [22].

While soil properties constitute the main factor driving bacterial community differentiation in *A. thaliana* roots across Europe, climate also plays a major role, influencing mainly but not exclusively the composition in filamentous eukaryotes [23]. Temperature is well-known to influence microbial growth, and diverse strains can show very different optimal growth temperatures. A recent study demonstrated that under high temperatures, *Sphagnum* peatmosses host a different set of microbes with better thermotolerance, that can provide them with thermal preconditioning, notably by activating their heat shock responses [24]. In the context of the current global change, increase of temperatures are concomitant to other environmental changes that were shown to impact microbiota composition, such as increased soil salinity [25] and reduced water availability (drought) [26,

27].

Another environmental parameter that was shown to influence the microbiota composition of a plant is its ecological context. The identity of neighbouring plants was shown to significantly shape leaf and root microbiota structures [28, 29]. Moreover, herbivores may influence community composition and play a role in plant microbiota assembly [7]. Insects also play a similar role, by importing bacteria to flowers during pollination, that are then transmitted to seeds [30].

1.2.4 Microbe-microbe interactions

Microbe-microbe interactions constitute important selective forces sculpting plant-associated microbial communities [31]. In theory, two microorganisms in a community can either interact positively, negatively or not interact. Microbes showing the highest number of interactions are referred to as *hubs* and are thought to be keystones of the general community structure. In the *A. thaliana* phyllosphere, environmental conditions and host genotypic factors were shown to directly impact the presence of two eukaryotic hub microbes (the oomycete *Albugo laibachii* and the fungus *Dioszegia sp.*), with cascading consequences on bacterial colonization capabilities, and therefore on general community composition [32].

Considering negative interactions in the plant microbiota, multiple studies highlighted mechanisms underlying microbial antagonism, relying notably on the use of antimicrobials. For instance, a yeast in the *A. thaliana* phyllosphere secretes a hydrolase with lysozyme activity to prevent one of the hub microbes, *Albugo laibachii*, to over-colonize host leaves and be detrimental for the plant [33]. In the tomato root microbiota, the fungal pathogen *Verticillium dahliae* was shown to secrete antimicrobial proteins and selectively inhibit the growth of bacteria and fungi that act as antagonists of its growth [34].

While the interaction processes favoring co-existence in plant microbiota remain poorly understood [31], microbial inter-metabolic dependencies in community have been dissected in synthetic systems [35], and general assembly principles can certainly be extended to plant-associated microbial consortia. When artificially feeding a community with a single carbohydrate polymer (a plant cell wall component, for instance), three groups of microbes can be distinguished: (1) degraders that are able to digest and acquire nutrients from the polymer; (2) exploiters that can only feed on the resulting digestion products and are therefore dependent on the presence of degraders; (3) scavengers that rely exclusively on metabolites produced by degraders and exploiters and thus depend on their presence for growth. This simplified hierarchical structure of a community illustrates that microbial inter-dependency consists essentially of cross-feeding and relies on different metabolic capabilities. It has been proposed that fungi represent hubs in plant microbiota, influencing bacterial community composition but also feeding a broad diversity of bacteria with their exsudates [31]. Since the ability to degrade plant cell wall components is a key trait of fungi [36], they may

actually represent essential degraders in plant microbiota.

1.3 Plant-associated fungi, their lifestyles and genomic signatures

Plant microbiota include a wide diversity of fungi, that importantly shape plant fitness and distribution worldwide [37]. Since the fungal kingdom diverged from other clades between 1500 and 900 million years ago [38], it can be described as phylogenetically and physiologically hyperdiverse. Fungal association with plants is described as ancestral, and may have facilitated the colonization of land by plants 500 million years ago [39]. As heterotrophs, most fungi acquire nutrients from plant material, either as saprotrophs, pathogens or mutualists [36]. Fungi have indeed evolved diverse strategies to associate with plant hosts, dead or alive, and acquire the nutrients necessary to their growth.

1.3.1 Saprotrophs

Saprotrophic fungi acquire nutrients from dead organic matter, and notably include free-living litter-decomposers and wood-decay fungi [40]. By hydrolytically degrading complex cell wall polymers from dead plant tissues, they reintroduce fixed carbon into soil and play a fundamental role in carbon cycling [39]. They act as regulators of carbon fluxes between the biosphere and the atmosphere and are estimated to contribute up to 90% of total heterotrophic respiration in woodland ecosystems [41].

To contribute in carbon cycling and acquire the nitrogen and carbon necessary to their growth, saprotrophic fungi produce and secrete highly diverse arsenals of carbohydrate-active enzymes (CAZymes), that include plant cell wall-degrading enzymes (PCWDEs). These CAZymes constitute important determinants underlying the saprotrophic lifestyle [36].

Litter-decomposing species living in forests and grasslands were reported to encode large repertoires of PCWDEs acting on the lignin and polysaccharide fractions of plant debris [42].

Wood-decay fungi are highly diverse and have evolved different strategies to degrade wood components [36]. A commonly adopted classification of these species is based on phenotype after fungal colonization: white, brown and soft rot. While white rot fungi degrade the cellulose, hemicellulose, and lignin components of plant cell walls, soft and brown rot fungi only degrade cellulosic and hemicellulosic compounds, leaving lignin oxidized and only partially affected. Genes encoding lignin degradation enzymes are overrepresented and highly diverse in the genomes of white rot fungi. They likely played an essential role in fungal adaptation to wood substrates and wood decay capabilities.

While PCWDE repertoires are key determinants of saprophytism and fungal abilities to grow on dead plant organic matter, other cellular pathways are likely as important for the utilization of

dead plant materials (e.g. pathways responsible for the detoxification and transport of carbohydrate degradation products) [36]. However, such pathways remain to be identified and characterized, to gain further insights in the functioning of saprotrophic fungi.

1.3.2 Plant-pathogenic fungi

Plant-pathogenic fungi are phylogenetically diverse and employ plethora of different strategies to colonize a host and cause disease. While necrotrophs kill their hosts to feed on dead material, biotrophs colonize living plant tissues exclusively [43]. Hemibiotrophs have both biotrophic and necrotrophic phases in their life cycle.

PCWDEs constitute essential determinants of plant pathogenicity [43]. Disruption of genes encoding cellulases, xylanases and pectinases was shown to reduce or abolish fungal virulence of phylogenetically distant species. PCWDEs are actually important for initial fungal colonization of plant tissues, prior to the activation of virulence mechanisms. Reduced expression of genes encoding cellulases was correlated to decreased tissue invasion capabilities of the rice blast fungus *Magnaporthe oryzae* [44]. Similarly, a recent study of the vascular wilt fungus *Fusarium oxysporum* identified carbohydrate-active enzymes secreted in the apoplast during early tomato root colonization that are necessary to later disease establishment [45]. However, since some plants evolved the ability to sense the presence of specific fungal PCWDEs and activate immune responses upon their perception, PCWDEs can also constitute virulence factors [43]. It is notably the case of several endopolygalacturonases from necrotrophic fungus *Botrytis cinerea*, that induce plant immune responses after being recognized as MAMPs by the *A. thaliana* receptor-like protein AtRLP42 [46].

Small secreted proteins, typically referred to as *effector proteins*, are produced by plant pathogenic fungi to support their invasion of host tissues [47]. Some effectors, such as the Tin2 protein secreted by the maize smut fungus *Ustilago maydis*, were shown to indirectly trigger changes in the structure and composition of plant cell walls, allowing consequently a better fungal colonization [48]. Another function of effectors is contributing in fungal evasion from host MTI. The tomato leaf mould fungus *Cladosporium fulvum* has been shown to secrete lysin motif (LysM) effectors, that bind chitin oligomers with high affinity and sequester them from host immune receptors [49]. Pathogenic fungi also developed effectors to subvert MTI responses that are mediated by phytohormones, such as JA, SA and ET [50]. For instance, *U. maydis* secretes high amounts of the effector Cmu1 (chorismate mutase) to reduce cytoplasmic levels of chorismate, a precursor for the synthesis of SA in plastids [51]. Finally, fungi also produce effectors targeting MTI terminal products. For instance, *C. fulvum* was shown to secrete effectors that act as inhibitors of plant antimicrobials, including chitinases [52], cysteine proteases [53] and glycoalkaloids [54].

Mycotoxins also play an important role in fungal pathogenicity, and constitute essential determinants of necrotrophy. They represent either secondary metabolites or secretory proteinaceous

toxins, with the main function of inducing plant cell death [43]. It is the case of oxalic acid, a secondary metabolite essential for the virulence of the necrotrophic fungus *Sclerotinia sclerotium*; it rapidly kills host plants, chelates calcium ions and reduces pH to activate fungal PCWDEs [55]. Wheat pathogens *Pyrenophora tritici-repentis* and *Stagonospora nodorum* secrete ToxA, a proteinaceous toxin that localizes in host chloroplasts and disrupts the photosystem activity to induce plant cell death [56]. Interestingly, while oxalic acid is a general mycotoxin affecting a broad range of plant hosts, ToxA was demonstrated to have a host-specific toxicity, genomically encoded in wheat by the sensitivity gene *Tsn1* [57].

While CAZymes, effectors and toxins often act in concert to induce disease on plant hosts [43], it is important to mention that exogenous factors are also determinant for fungal pathogenicity. Environmental variables such as temperature, humidity and atmospheric concentration in carbon dioxide are believed to modulate both pathogen and plant physiological statuses, and therefore to influence disease outcomes [58]. The extent to which variation in these environmental parameters impacts fungal pathogenicity is still poorly understood, and remains to be deciphered.

1.3.3 Mycorrhizae

Mycorrhizae represent different root-fungi mutualistic associations in which the fungus and plant exchange commodities required for their growth and survival [59]. These associations have the property to produce fungal colonization structures with specific morphologies. The two main types of mycorrhizae can be easily distinguished by visual analysis of these morphologies. While the hyphae of arbuscular mycorrhizal fungi (AMF) form arbuscule structures inside their host root cells, ectomycorrhizal fungi form an intercellular labyrinthine structure called Hartig net.

Arbuscular mycorrhiza (AM) is a widespread symbiosis between most land plants and fungi from the Glomeromycotina subphylum [16]. AMF are obligate symbionts producing extended extraradical hyphal networks that can explore soils more efficiently than plant roots. Their hyphae take up mineral phosphate, ammonium, nitrate, sulphate, potassium and water, and provide them to the plant in exchange for carbohydrates and lipids [16, 60]. The symbiosis relies therefore on plethora of plant and fungal nutrient-specific transporters [61]. For instance, plant phosphate transporters from family PHT1 contribute to an essential pathway for the establishment of the symbiosis and AMF-mediated mineral phosphate uptake [62, 63]. Importantly, AMF nutrition in carbon seems to rely essentially on plant-encoded molecular mechanisms. Comparative genomics identified that the most recent common ancestor to Glomeromycotina lost substantial amounts of genes encoding PCWDEs or involved in fatty acid metabolism [64]. In the roots, plant invertases hydrolyse sucrose into glucose and fructose, that are then transferred to the fungal symbiont via monosaccharides transporters [61]. Fatty acids are produced and metabolized in root cells hosting arbuscules, using an AM-specific pathway [65]. Glucose and acetate are metabolized into lauric

and palmitic acids in plastids, then used as substrates by cytoplasmic enzyme RAM2 to produce β-monoacyl glycerols, that can be transported through the periarbuscular membrane for fungal uptake. Plants were actually shown to employ *reward mechanisms*, modulating carbon allocation to *reward* the most collaborative symbionts [66].

Ectomycorrhiza (ECM) is a symbiosis taking place in the roots of about 6,000 tree species, that can involve different fungi from multiple lineages of Ascomycota and Basidiomycota [67]. Ectomycorrhizal fungi mediate the uptake of soil nitrogen and phosphorus for their host, in exchange of carbon resources [68]. They were shown to derive from saprotrophic ancestors. As for AM, the emergence of symbiosis was concomitant to a large-scale loss of fungal PCWDEs, especially of lignin and cellulose degrading enzymes [69]. Although mechanisms remain to be characterized, it can be hypothesized that similarly to AMF, the nutrition of ectomycorrhizal fungi relies importantly on plant carbohydrate metabolism and transport. Interestingly, fungal modulation of host immunity using effector proteins (called mycorrhiza-induced small secreted proteins; MiSSPs) was reported to be a key factor in the establishment of the symbiosis. For instance, the effector MiSSP7 secreted by *Laccaria bicolor* prevents the proteasomal degradation of the transcriptional repressor JAZ6 of *Populus spp.*, thereby maintaining the repression of JA-induced genes [15].

1.3.4 Fungal endophytes

Although *endophytism* can be referred to as a fungal lifestyle contrasting with mycorrhizae and saprotrophy, the definition behind this concept has been disputed for decades. An increasingly accepted definition suggests to describe as endophyte any microorganism spending all or part of its lifetime colonizing internal plant tissues, above- and/or below-ground [70]. This definition is descriptive of a habitat (the *endosphere*) and considers that endophytes can have different modes of nutrition and effects on plant health, spanning from pathogenic to beneficial. While this *endophytic continuum* concept was first introduced in 2005 [71], it is presently still controversial. A recent review claimed the importance of drawing a distinction between so-called *true endophytes* representing either commensals or mutualists, and pathogens causing detriments to the host plant [72]. In this thesis, the term *endophyte* will be used to describe any non-mycorrhizal fungus colonizing the endosphere of healthy plants in nature, independent of its individual effect on plant performance.

Such endophytes span a broad phylogenetic diversity, predominantly in phyla Ascomycetes and Basidiomycetes. Fungal community composition in root endosphere was demonstrated to be shaped by the local environment (climate and soil properties) [73] and host identity, but also fungal effects on plant performance [74]. The combination of these factors, together with the local competition with other microorganisms, defines whether facultative endophytic fungi inhabit the root endosphere of a healthy plant [75].

Over the last decade, functional studies of endophytism have mostly focused on specific fungal isolates and their interaction with model plant species in gnotobiotic system [76]. For instance, fungal endophytes *Colletotrichum toffeldiae* and *Serendipita indica* were well studied for promoting the growth of *A. thaliana*. Interestingly, an intact host immune system is essential for the beneficial effects of these two fungi [77, 78]. The specific effect of *C. toffeldiae* concerns plant phosphorous stress, which is alleviated by fungal solubilization and uptake of plant-inaccessible phosphate [78]. Such studies of specific fungal processes demonstrated that diverse factors modulate the effects of fungal endophytes on plants, including host identity and its physiological status, nutrient availability, environmental conditions and microbe-microbe interactions [76].

Since only a restricted number of genome sequences are available for fungal endophytes, only few genomic signatures of endophytism have yet been identified. In planta transcriptomics of diverse fungal endophytes reported PCWDE-encoding genes to be overexpressed during root colonization of different plants [77, 79, 80]. Comparison of genomes in the phylogenetic order Helotiales identified that endophytes (and ericoid mycorrhizal fungi) are enriched in CAZymes, including PCWDE families acting on hemicellulose and lignocellulosic compounds [81]. Similarly, genomes of the dark-septate endophytes *Cadophora sp.* and *Periconia macrospinosa* were shown to be enriched in genes encoding PCWDEs (but also aquaporins, secreted peptidases, and lipases) in comparison to closely related fungi with other lifestyles [82]. In the phylogenetically distant family Sebacinales, genomes of beneficial endophytes are also significantly richer in PCWDE-encoding genes than the ones of mycorrhizal fungi, with xylanases being among the most overexpressed ones in planta [77]. Additionally, in this clade, transcriptomic data suggested an importance of fungal transporters in the interaction of endophytes with *A. thaliana* roots. While PCWDE may therefore play an important role in endophytic colonization, their role in endophytism remains to be deciphered, and other genetic determinants need to be identified.

Multiple phylogenetically distant evolutionary trajectories to endophytism have been described, from pathogenic or saprotrophic ancestors [76]. In contrast with mycorrhizae, no substantial loss of PCWDE-encoding genes occurred during transitions to endophytism [77, 79, 81], reinforcing the hypothesis of their importance for the endophytic colonization of plant tissues. Genetic signatures underlying transitions from pathogenicity to asymptomatic or beneficial endophytism have been identified [72], and provide insights on molecular mechanisms substantiating the *endophytic continuum* concept. For instance, key virulence factor genes of rapeseed pathogen *Sclerotinia sclerotiorum* (encoding some PCWDE, effectors and factors involved in compound appressoria formation) were shown to be downregulated upon infection by the DNA mycovirus SsHADV-1, turning the fungus into a beneficial endophyte capable of promoting plant growth and enhancing disease resistance [83]. Another example concerns the beneficial species *C. toffeldiae* which derives from pathogenic ancestors. It encodes less effector proteins than the closely related pathogen *C. in-*

canum and exhibited a lower expression in planta of genes encoding effectors, proteases, secondary metabolism key enzymes and transporters [79]. A single *C. toffeldiae* isolate was recently shown to cause disease on *A. thaliana*, by overexpressing a fungal secondary metabolism gene cluster related to sesquiterpene abscisic acid and botrydial biosynthesis, suggested to affect host nitrogen and iron uptake [84]. Taken together, these findings revealed that subtle genomic and/or transcriptomic changes affecting virulence factors may be sufficient to operate a switch from pathogenic to non-detrimental endophytism.

1.4 The *Arabidopsis thaliana* root mycobiota

The term *mycobiota* (or *mycobiome*) refers to the fungal component of a microbiota, and generally consists of the total set of fungal taxa detected by amplicon sequencing or isolated in culture from a sample. The model plant species *A. thaliana* is as other Brassicaceae, a non-mycorrhizal host which roots are colonized by a broad diversity of fungi that do not establish symbiotic structures [85]. While the physiological relevance of these fungi remain unclear, both the ecology and functioning of the *A. thaliana* root mycobiota have been studied over the last years.

The root microbiota of natural *A. thaliana* populations was profiled in compartments rhizosphere, rhizoplane and root endosphere at 17 European sites along a latitudinal gradient [23]. Alpha diversity indices revealed a gradual decrease of fungal, bacterial, and oomycetal diversity from the soil to the root endosphere. However, sampling site showed far greater effect than root compartment on community composition in fungi and oomycetes (but not bacteria). Phylogenetic analysis of detected fungal taxa identified that the *A. thaliana* root-associated fungi mostly belong to phyla Ascomycetes and Basidiomycetes, and most represented classes were Sordariomycetes, Leotiomycetes, Dothideomycetes and Agaricomycetes. Classes Leotiomycetes and Dothideomycetes were reported to be significantly enriched in roots in comparison to surrounding soil. Comparison of *A. thaliana* root-associated communities with those of co-occurring grass species identified that only part of the mycobiota has the capacity to colonize the roots of these phylogenetically distant plants, suggesting some fungi may not have the ability to degrade and feed on monocotyledonous cell walls components. Finally, a reciprocal transplant experiment aiming at disentangling the factors behind continental-scale variation in the *A. thaliana* root microbiota (namely soil properties, climate and host genotype) highlighted that climate is the strongest parameter influencing fungal and oomycetal communities, while bacterial consortia are more impacted by soil properties. Consistent with this result, a recent study also identified a strong effect of climate conditions on root mycobiota structure in *A. thaliana* (and barley) [86]. While nutrient availability did not affect community assembly, it was suggested to influence individual fungal effects on plant growth.

The effect of the root mycobiota on plant performance was previously investigated with microbiota reconstitution experiments [87]. Culture collections of fungi, bacteria and oomycetes

recovered from healthy *A. thaliana* plants sampled in nature were assembled. Three synthetic microbial communities recapitulating the intra-kingdom phylogenetic diversity profiled in natural microbiota were designed. Plants were then grown in a soil-like substrate inoculated with these three kingdom-specific communities, individually and combined. After four weeks of culture in gnotobiotic system, plants colonized exclusively by fungi exhibited a null survival rate, revealing a strong detrimental effect of the *A. thaliana* mycobiota on plant growth and health. Interestingly, plant survival was rescued in presence of the bacterial synthetic community. Fungal alpha diversity along the soil-root continuum was shown to be reduced in presence of bacteria, suggesting that bacteria protect their host against mycobiota detrimental effects by restricting fungal growth. Importantly, only about half of the mycobiota members composing the synthetic community used in this experiment (18/34) showed significant detrimental effects on *A. thaliana* growth in mono-association, while other strains did not impact plant performance. The fungal root endophyte *Serendipita vermifera* (from phylogenetic class Sebacinales, robustly represented in the *A. thaliana* root mycobiota across Europe) was shown to protect plants synergistically with root-associated bacteria, from soil-borne fungal pathogen *Bipolaris sorokiniana* [88]. Therefore, while the *A. thaliana* mycobiota is highly detrimental as a community, it seems to comprise fungi with diverse individual effects on plants, including detrimental, neutral and protective ones. Additionally to microbial protection, plants rely on an intact immune system to survive and grow, when colonized by a detrimental mycobiota [12]. Plant genotypes impaired in different immune components were inoculated with a multi-kingdom synthetic community and showed contrasted resistance to fungal detrimental effects. While bacterial commensals are sufficient for wild-type plants to overcome these effects, plants impaired in tryptophan-derived metabolites (*cyp79b2/b3* mutants) were considerably affected, due to an increased fungal load in their roots. Both protective root microbiota members and host tryptophan metabolism are thus crucial to control fungal load and promote *A. thaliana* survival and growth.

2 Thesis aims

While mycorrhizal fungi have been studied for decades, little is known about root mycobiota members that do not establish symbiotic structures, but have the ability to colonize roots of asymptomatic plants in nature. Endophytes especially, remain to be broadly inspected from the genomic perspective, so determinants of endophytism can be identified.

Recent studies of the non-mycorrhizal plant *A. thaliana* initiated the characterization of root mycobiota assembly rules, and pointed to the influence of climate, microbe-microbe interactions and host immune system on fungal community structure [12, 23, 87]. They also highlighted clear detrimental effects of these fungi on their host, although the genetics behind this pathogenicity remain to be deciphered. During my PhD, I wished to further dissect and characterize this detrimental *A. thaliana* root mycobiota, focusing on its function, evolution and genomic signatures.

Thus, in the first chapter of this thesis, I focused on a representative set of fungi that could be isolated from the roots of healthy *A. thaliana* plants, sampled in nature. After genome sequencing of these isolates, I undertook a general characterization of the *A. thaliana* root mycobiota, questioning (1) the evolutionary history of its members; (2) their lifestyle; (3) their individual effects on plant growth; and (4) their genomic signatures. As multiple pieces of evidence pointed to the endophytism of mycobiota members, our data set offered the opportunity to characterize genomic determinants underlying root colonization by fungal endophytes. After observing a gradient of fungal effects on plant growth, ranging from beneficial to detrimental (as described by the endophytic continuum concept [71, 76]), a gene family linking fungal colonization efficiency to plant health could be identified.

Following the discovery of broadly conserved genetic determinants, I focused in a second chapter, on a single fungal species: the prevalent *A. thaliana* mycobiota member and ubiquitous soilborne fungus *Plectosphaerella cucumerina*. I aimed to identify more specific signatures behind its robust root colonization ability and detrimental effects on plant growth. After sequencing the genomes of 69 strains isolated from diverse plant hosts, intra-species evolutionary patterns could be deciphered. Genomic signatures of host adaptation have previously been identified in closely-related pathogenic fungi [89, 90]. When questioning their existence in *P. cucumerina*, I could identify a candidate genomic region predicted to be involved in adaptation to *A. thaliana*.

Taken together, the results compiled in this thesis contribute to a better understanding of the fine line separating commensal and pathogenic interactions in the *A. thaliana* root mycobiota, and offer new genomic insights into fungal endophytism.

3 Genetic determinants of endophytism in the *A. thaliana* root mycobiome

Fantin Mesny[1], Shingo Miyauchi[1,2], Thorsten Thiergart[1], Brigitte Pickel[1], Lea Atanasova[3,4], Magnus Karlsson[5], Bruno Hüttel[6], Kerrie W. Barry[7], Sajeet Haridas[7], Cindy Chen[7], Diane Bauer[7], William Andreopoulos[7], Jasmyn Pangilinan[7], Kurt LaButti[7], Robert Riley[7], Anna Lipzen[7], Alicia Clum[7], Elodie Drula[8,9], Bernard Henrissat[10], Annegret Kohler[2], Igor V. Grigoriev[7,11], Francis M. Martin[2,12] & Stéphane Hacquard[1,13].

1) Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. 2) Université de Lorraine, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, UMR Interactions Arbres/Microorganismes, Centre INRAE Grand Est-Nancy, 54280 Champenoux, France. 3) Research division of Biochemical Technology, Institute of Chemical, Environmental and Biological Engineering, Vienna University of Technology, Vienna, Austria. 4) Institute of Food Technology, University of Natural Resources and Life Sciences, Vienna, Austria. 5 Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden. 5) Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden. 6) Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany. 7) U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. 8) INRAE, USC1408 Architecture et Fonction des Macromolécules Biologiques, 13009 Marseille, France. 9) Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille Univ., 13009 Marseille, France. 10) Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. 11) Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. 12) Beijing Advanced Innovation Centre for Tree Breeding by Molecular Design (BAIC-TBMD), Institute of Microbiology, Beijing Forestry University, Tsinghua East Road Haidian District, Beijing, China. 13) Cluster of Excellence on Plant Sciences (CEPLAS), Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. *Corresponding authors: F. M. Martin (francis.martin@inrae.fr), S. Hacquard (hacquard@mpipz.mpg.de).

This chapter presents the main study of this PhD thesis and has been published in Nature Communications in December 2021 [91]. This project was initiated, coordinated and supervised by Stéphane Hacquard and Francis M. Martin. The manuscript presented here was written by myself and Stéphane Hacquard, with inputs from Francis M. Martin, Annegret Kohler and Shingo Miyauchi. The selection of 41 representative *A. thaliana* mycobiota members to be further characterized and the analysis of their ecology (see 3.2.1) were performed by Stéphane Hacquard and Thorsten Thiergart. Genome sequencing, assembly and annotation of these fungi was performed at the Joint Genome Institute (Walnut Creek, USA) by Igor V. Grigoriev, Kerrie W. Barry, Sajeet Haridas, Cindy Chen, Diane Bauer, William Andreopoulos, Jasmyn Pangilinan, Kurt LaButti, Robert Riley, Anna Lipzen and Alicia Clum. Additional functional annotations of the genomes were performed by Shingo Miyauchi (secretome, proteases, lipases), Elodie Drula and Bernard Henrissat (CAZymes). Except for several descriptive figures presented in Supplementary Information (Supplementary figures 4, 5, 6b, 7 and 14c) assembled by Shingo Miyauchi, all the genomic analyses presented in paragraphs 3.2.2, 3.2.3 and 3.2.4

were conducted by myself. Plant recolonization experiments in gnotobiotic system and fungal colonization assays discussed in paragraph 3.2.5 were performed by myself, with technical assistance from Brigitte Pickel. I also conducted the experiments of paragraph 3.2.6 and analysed the transcriptomic data sequenced at Max Planck Genome Center by Bruno Hüttel. Finally, I carried out the association analyses described in paragraph 3.2.7, and used fungal strains provided by Lea Atanasova and Magnus Karlsson to perform gnotobiotic experiments validating a candidate gene for fungal colonization ability.

3.1 Introduction

Roots of healthy plants are colonized by a rich and diverse community of microbes (i.e. bacteria and fungi) that can modulate plant physiology and development [87, 92–95]. Root colonization by arbuscular mycorrhizal, ectomycorrhizal and ericoid mycorrhizal fungi play fundamental roles in shaping plant evolution, distribution, and fitness worldwide [36, 37, 96–99]. In contrast, the physiological relevance of root mycobiota members that do not establish symbiotic structures, but have the ability to colonize roots of asymptomatic plants in nature remains unclear. These fungal endophytes are predominantly Ascomycetes [2, 100], which can either represent stochastic encounters or engage in stable associations with plant roots [23, 73, 75, 101, 102]. Multiple factors driving the assembly of endophytic fungal communities have been identified, including climatic conditions, soil properties, species identities of the host and surrounding plants and abiotic stresses [2, 23, 29, 73, 100, 101, 103–105]. Re-colonization experiments with individual fungal isolates and germ-free Brassicaceae plants—non-mycorrhizal and previously reported as hosting root endophytes colonizing a broad range of hosts [106]—revealed various effects of mycobiota members on plant performance, ranging along the parasitism-to-mutualism continuum [74, 76, 87, 107]. Importantly, the outcome of the interaction on plant health can be modulated by host genetics, host nutritional status, and local environmental conditions [77–79, 81].

While the ectomycorrhizal lifestyle was shown to have arisen independently multiple times from saprotrophic ancestors—by convergent transposon-mediated genomic expansions and simultaneous losses of plant cell wall-degrading enzymes (PCWDEs) [69, 108], some phylogenetically distant evolutionary trajectories to root endophytism have been described, from pathogenic [79, 109, 110] or saprotrophic ancestors [111]. Although genomic signatures of endophytism remain to be identified, these studies pinpointed that no contraction of PCWDE arsenals occurred during transitions to endophytism [76]. Genomes of dark-septate endophytes were shown to be enriched in genes encoding PCWDEs—but also aquaporins, secreted peptidases, and lipases—, in comparison to closely related fungi with other lifestyles [82]. Importantly, PCWDE-encoding genes were reported to be over-expressed during root colonization by diverse fungal endophytes [77, 79, 80], suggesting they might be key determinants of endophytism. Genetic factors underlying the

endophytic lifestyle could however be multiple, and also niche- and host-dependent.

Here, we aim at better characterizing the evolution and function of the root mycobiota, by studying a diverse set of 41 cultured fungi that colonize roots of the non-mycorrhizal plant *A. thaliana*. Using comparative genomics and transcriptomics in combination with plant recolonization experiments, we identified genomic determinants underlying the endophytic lifestyle. Our results suggest that repertoires of PCWDEs of the *A. thaliana* root mycobiota are key determinants of endophytism, shaping fungal endosphere assemblages and modulating host fitness.

3.2 Results

3.2.1 Cultured isolates are representative of wild *A. thaliana* root mycobiomes

Fungi isolated from roots of healthy *A. thaliana* represent either stochastic encounters or robust endosphere colonizers. From a previously established fungal culture collection obtained from surface-sterilized root fragments of *A. thaliana* and relative Brassicaceae species [87], we identified 41 isolates that could be distinguished based on their rDNA internal transcribed spacer 1 (ITS1) sequences, representing 3 phyla, 26 genera, and 38 species of the fungal root microbiota (Fig. 1a). We first tested whether these phylogenetically diverse isolates were representative of naturally occurring root-colonizing fungi. Direct comparison with rDNA ITS1 sequence tags from a continental-scale survey of the *A. thaliana* root mycobiota [23] revealed that most of the matching sequences were abundant (mean relative abundance, mean RA $> 0.1\%$, 30 out of 41 strains), prevalent (sample coverage $> 50\%$, 22 out of 41), and enriched (root vs. soil, log2FC, Mann–Whitney U test, $FDR < 0.05$, 26 out of 41) in *A. thaliana* root endosphere samples at a continental scale (Fig. 1a). Quantitatively similar results were obtained using sequence data from the independent rDNA ITS2 locus (Spearman; Sample coverage: $rho = 0.65$, $P < 0.01$; RA: $rho = 0.59$, $P < 0.01$; Fig. 1b). The cumulative RA of the sequence tags corresponding to these 41 fungi accounted for 35% of the total RA measured in root endosphere samples across European sites [23], despite the under-representation of abundant Agaricomycetes and Dothideomycetes taxa (Fig. 1c). We next assessed the worldwide distribution and prevalence of these fungal taxa across 3,582 root samples from diverse plants retrieved from the GlobalFungi database [112]. Continent-wide analysis revealed that the proportion of samples with positive hits was greater in Europe (sample coverage: up to 30%, median $= 4\%$) than in North America (sample coverage: up to 10%, median $= 0.5\%$), and largely insignificant in samples from other continents (Fig. 1a). Interestingly, only a few of these 41 isolates were detected in leaves of *A. thaliana* at two locations in Germany (data re-analyzed from [32], $n = 51$ samples), as well as in 2,647 leaf samples retrieved from the GlobalFungi database [112] (Supplementary Fig. 1). Results indicate that most of the cultured *A. thaliana* root colonizing fungi reproducibly and predominantly colonize plant roots across geographically distant sites

irrespective of differences in soil conditions and climates.

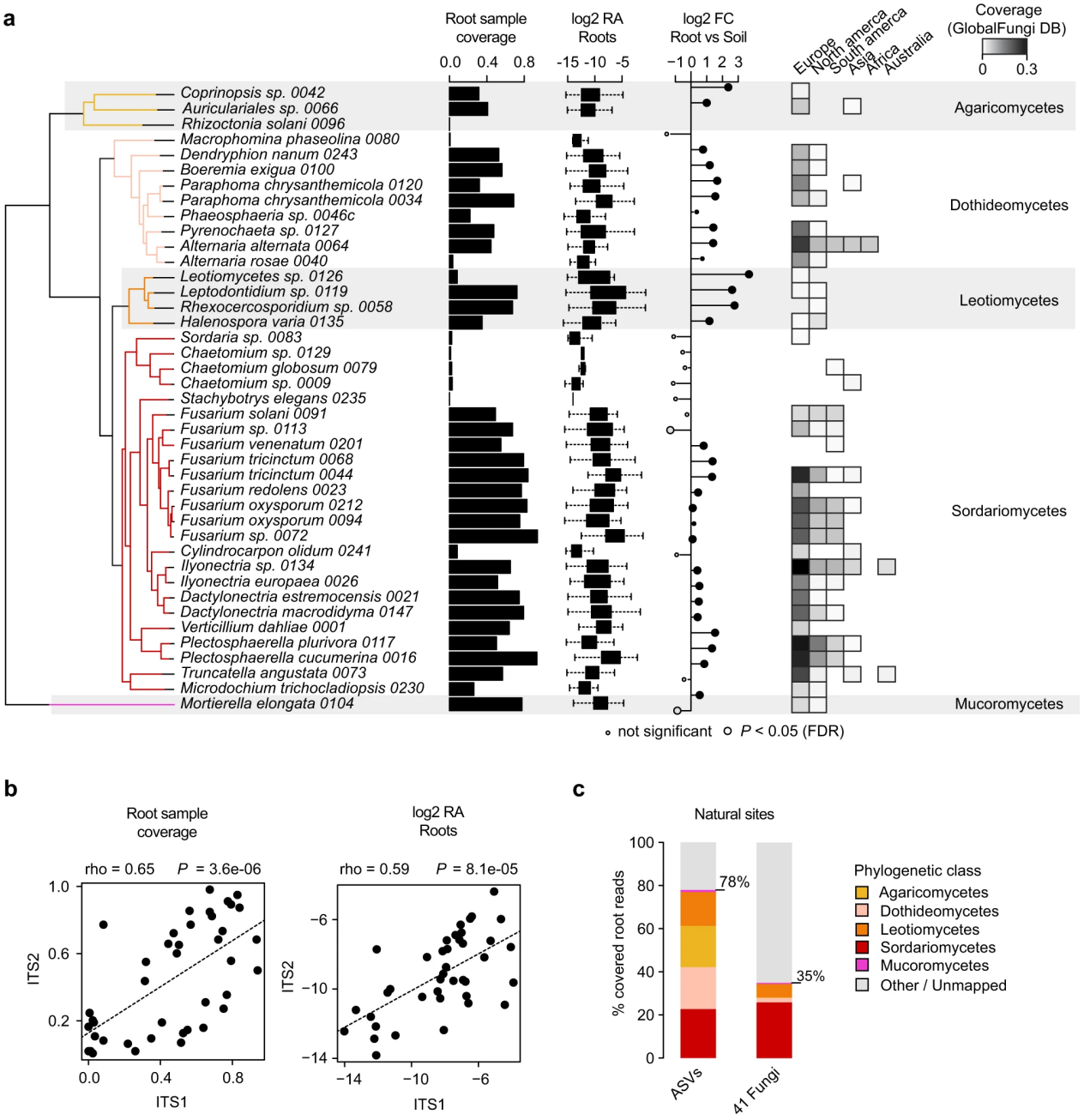


Figure 1: Prevalence and abundance profiles of 41 root-colonizing fungi across naturally occurring *A. thaliana* root mycobiomes. **a** Species names and phylogenetic relationships among the 41 selected fungi. Estimated prevalence (i.e., root sample coverage, bar-plots), relative abundance (RA, log2 transformed, box-plots), and enrichment signatures (log2FC, circles) were calculated for each fungus based on data from a previously published continental-scale survey of the *A. thaliana* root mycobiota [23]. ITS1 tags from natural site samples were directly mapped against the reference ITS1 sequences of the selected fungi. Sample coverage in roots was computed based on n=169 root samples and estimated RA were calculated for root samples having a positive hit only. On the RA boxplot, boxes are delimited by first and third quartiles

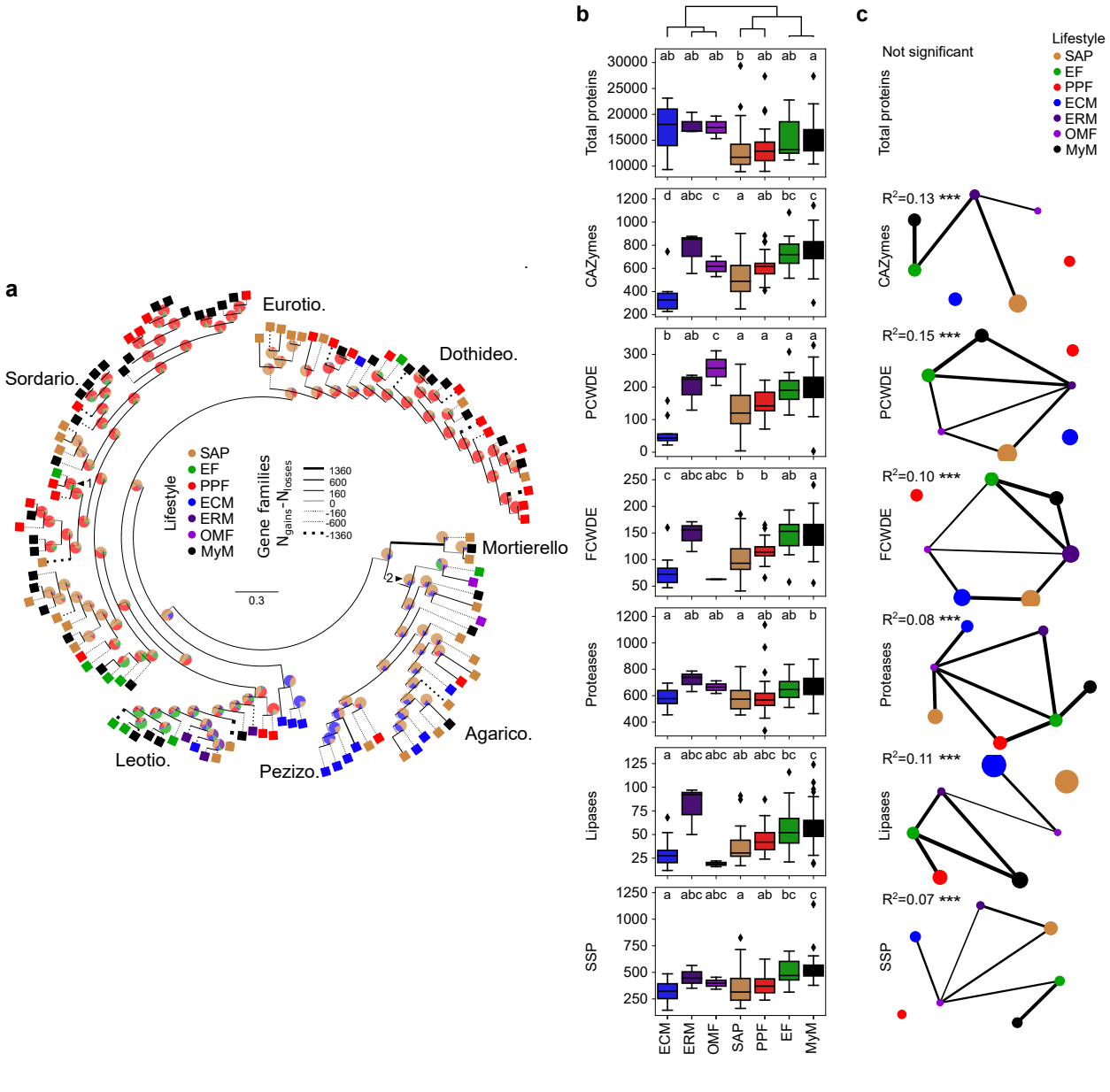
and whiskers extend to show the rest of the distribution. Log2Fold-Change (log2FC) in RA between root (n=169) and soil samples (n=223) is shown based on the mean RA measured across samples and significant differences are indicated by circle sizes (two-sided Mann–Whitney U test, $FDR < 0.05$, see detailed values in Supplementary Data 1). ITS1 sequence coverage measured across 3582 root samples retrieved from the GlobalFungi database [112]. Note that samples were analyzed separately by continent. **b** Correlation between root sample coverage (left panel) measured in ITS1 (n=169) and ITS2 (n=158) datasets for each of the 41 fungi (n=41, Spearman's rank correlation). Right panel: same correlation but based on log2 RA values (n=41, Spearman's rank correlation). **c** RA profiles of naturally occurring fungi (class level) detected in *A. thaliana* roots across 17 European sites18 ("all ASVs", left) and the corresponding distribution of the ITS1 sequences of the 41 selected fungi ("41 fungi", right). Note that the cumulative RA of the 41 fungi accounts for 35% of all sequencing reads detected in *A. thaliana* roots across European sites.

3.2.2 Root mycobiota members evolved from ancestors with diverse lifestyles

Given the broad taxonomic diversity of *A. thaliana* root mycobiota members, endosphere colonization capabilities may have evolved multiple times independently across distinct fungal lineages. We sequenced the above-mentioned 41 fungal genomes using PacBio long-read sequencing and annotated them with the support of transcriptome data (Methods), resulting in high-quality genome drafts (number of contigs: 9–919, median = 63; L50: 0.2–9.1 Mbp, median = $2.3Mbp$; Supplementary Data 1). Genome size varied between 33.3 and 121 Mb (median = $45Mbp$) and was significantly correlated with the number of predicted genes (number of genes: 10,414–25,647, median = $14,777$, Spearman $rho = 0.92$, $P = 3.82e-17$) and the number of transposable elements (Spearman $rho = 0.86$, $P = 4.13e-13$) (Supplementary Fig. 2). A comparative genome analysis was conducted with 79 additional representative plant-associated fungi with previously well-described lifestyles [113], selected in the same or closely related phylogenetic classes as the strains we sequenced. Since classifying species into unique lifestyle categories is restrictive and can introduce bias [114], both the isolation of strains and previous knowledge about their species were considered to select plant pathogens, soil/wood saprotrophs, ectomycorrhizal symbionts, ericoid mycorrhizal symbionts, orchid mycorrhizal symbionts and endophytes [79, 81, 82, 110, 115–118] (Fig. 2a, Supplementary Fig. 3 and 4, Supplementary Data 2). Arbuscular mycorrhizal fungi were excluded from the study, as they are phylogenetically distant to the strains we isolated. To decipher potential evolutionary trajectories within this large fungal set, we first defined copy numbers of gene families in the 120 fungal genomes based on orthology prediction ($n = 41,612$; OrthoFinder [119]) and subsequently predicted the ancestral genome content using the Wagner parsimony method (Count[120]). Next, we trained a Random Forest classification model linking gene family copy numbers to lifestyles, resulting in a lifestyle prediction accuracy of $R^2 = 0.70$ (Methods). Although this classifier cannot confidently assign a single lifestyle to one genome content, it can be used to estimate lifestyle probabilities, and can reveal potential evolutionary trajectories when applied to Wagner-predicted ancestral genomic compositions (see pie charts, Fig. 2a). This proba-

bilistic approach corroborated that recent ancestors of the beneficial root endophyte *Colletotrichum tofieldiae* were likely pathogenic [79], whereas those of beneficial Sebacinales—like those of ectomycorrhizal Agaricomycetes—were predicted to be saprotrophs [76, 121] (see arrows numbered 1 and 2 on Fig. 2a). According to the classifier's predictions, Agaricomycetes and Mortierellomycetes in *A. thaliana* mycobiota likely derived from soil saprotrophs, while those belonging to Dothideomycetes and Sordariomycetes were predicted to have evolved from pathogenic ancestors. The ancestral lifestyle of Leotiomycete mycobiota members remains uncertain and could be multiple (Fig. 2a). Although the composition of our data set might influence these ancestral lifestyle predictions, our results nonetheless suggest that in planta accommodation of *A. thaliana* root mycobiota members occurred multiple times independently during evolution, as these fungi evolved from ancestors with diverse lifestyles.

Figure 2: Ancestral relationships and trait convergence across root-colonizing fungal endophytes. **a** Lifestyle-annotated whole-genome phylogeny of the 41 selected mycobiota members (MyM, black) and 79 published fungal genomes (SAP saprotrophs, EF endophytic fungi, PPF plant pathogenic fungi, ECM ectomycorrhiza, ERM ericoid mycorrhiza, OMF orchid mycorrhizal fungi). Pie charts on ancestor nodes show lifestyle probabilities of each ancestor, as identified by a Random Forest model trained on 79 non-mycobiota genome compositions in gene families ($R^2 = 0.70$). Two arrows highlight ancestral lifestyle predictions which corroborate previous reports: (1) the pathogenic ancestor of the endophyte *Colletotrichum tofieldiae* (2) the saprotrophic ancestor of ectomycorrhizal fungi and Sebacinales. Branch width is proportional to the gene family gains-losses difference ($Ngains - Nlosses$). Line is dotted when this difference is negative. **b** Genomic counts ($n = 120$) of genes involved in fungal-host/environment associations (CAZymes carbohydrate-active enzymes, PCWDEs plant cell wall-degrading enzyme, FCWDEs fungal cell-wall degrading enzyme, SSPs small secreted proteins; PCWDEs and FCWDEs are CAZyme subsets). Boxes are grouped according to UPGMA hierarchical clustering on mean counts over the different categories. They are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). ANOVA-statistical testing ($Counts \sim PhylogenyPCs + Lifestyle$, Methods) identified both phylogeny and lifestyles as having an effect on genomic contents. The letters highlight the result of a two-sided post hoc TukeyHSD test that compares count differences exclusively due to the lifestyle. **c** Networks showing the results of a PERMANOVA-based comparison of gene repertoires ($JaccardDistances \sim Phylogeny + Lifestyle$, see Supplementary Data 3 for detailed R^2 and P-values). Networks for each category are labeled with Lifestyle R^2 values. ***P$<$ 0.001 (Supplementary Fig. 6). Lifestyles are connected if their gene compositions are not significantly different. Node size is proportional to the area of one lifestyle's ordination ellipse on a Jaccard-derived dbRDA plot constrained by lifestyles, and reflects the intra-lifestyle variability. Edge weights and widths are inversely proportional to the distance between ordination ellipse centroids.

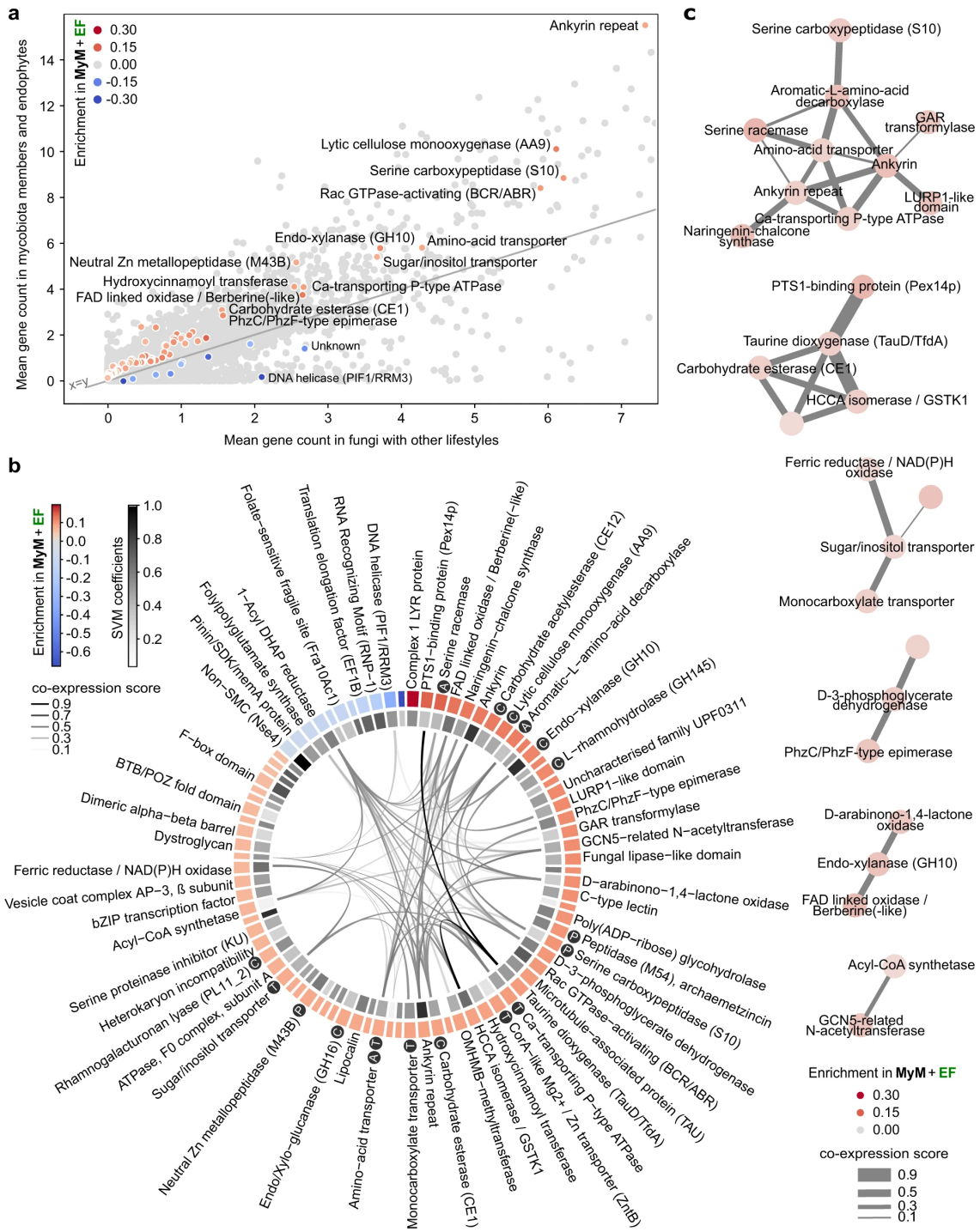


3.2.3 Functional overlap in genomes of root mycobiota members and endophytes

Isolation of mycobiota members from roots of healthy plants prompted us to test whether their gene repertoires more extensively resemble those of mycorrhizal symbionts, known endophytes, saprotrophs, or pathogens. While the genomes of ectomycorrhizal fungi were shown to be enriched in transposable elements [69, 108], the percentage of these elements remained low in genomes of root mycobiota members (0.69–28.43%, median = 5.44%, Supplementary Fig. 5). We annotated genes known to play a role in fungus-host interactions (Methods), including those encoding carbohydrate-active enzymes (CAZymes), proteases, lipases, and effector-like small secreted proteins (SSPs [122]), and then assessed differences in repertoire diversity across lifestyles (Fig. 2b).

Unlike ectomycorrhizal fungi [69, 108], but similarly to endophytes [77, 79, 81, 82, 110], the genomes of root mycobiota members retained large repertoires of genes encoding PCWDEs, SSPs, and proteases (ANOVA-TukeyHSD, $P < 0.05$, Fig. 2b). Using permutational multivariate analysis of variance (PERMANOVA) and distance-based redundancy analyses (dbRDA)—based on Jaccard dissimilarity indices between genomes calculated on the copy numbers of genes in each family—, we distinguished lifestyle from phylogenetic signals in gene repertoire composition (Fig. 2c, Supplementary Fig. 6a). This revealed that "lifestyle" significantly contributes to the variation in gene repertoire composition (phylogeny: R^2: 0.17–0.46, $P < 0.05$; lifestyle: R^2: 0.07–0.15, $P < 0.05$, Supplementary Data 3). Interestingly, the factor "lifestyle" explained the highest percentage of variance for PCWDE repertoires (phylogeny: $R^2 = 0.26$; lifestyle: $R^2 = 0.15$, Supplementary Data 3), suggesting that these CAZymes play an important role in lifestyle differentiation. Further pairwise comparisons between lifestyle groups revealed that gene repertoire composition of root mycobiota members could not be differentiated from those of endophytes (post hoc pairwise PERMANOVA, $P > 0.05$, Fig. 2c). Therefore, gene repertoires of *A. thaliana* root-colonizing fungi resemble those of endophytes more than saprotrophs, pathogens or mycorrhizal symbionts. Across the tested gene groups, the families which contribute the most in segregating genomes by lifestyles (Supplementary Fig. 6b, Methods) include two xylan esterases (CE1, CE5), two pectate lyases (PL3_2, PL1_4), one pectin methyltransferase (CE8), and one serine protease (S08A). Further analysis focusing on total predicted secretomes (Supplementary Fig. 7, Supplementary Fig. 8a) and CAZyme subfamilies (Supplementary Fig. 8b) confirmed strong genomic similarities between *A. thaliana* root mycobiota members and known endophytic fungi.

Figure 3: Minimal set of 84 gene families discriminating mycobiota members and endophytes from other lifestyles. **a** Scatterplot showing the mean per-genome copy number of each orthogroup in mycobiota members and endophytes, in comparison to other lifestyles. Light gray: all 41,612 orthogroups. The 84 discriminant orthogroups identified by SVM-RFE ($R^2 = 0.8$) are highlighted in a gradient of red or blue colors reflecting, respectively, enrichment or depletion in *A. thaliana* mycobiota members and endophytes (MyM+EF) compared to the other fungal lifestyles. **b** Functional descriptions of the 84 discriminant orthogroups. This gene set is enriched in CAZymes (Fisher, $P < 0.05$, labeled C) and also contains peptidases (labeled P), transporters (labeled T) and proteins involved in amino-acid metabolism (labeled A). The outer circle shows orthogroup enrichment/depletion as described in panel a (see Supplementary Data 4a for associated ANOVA P-values). The inner circle depicts the SVM coefficients, reflecting the contribution of each orthogroup to lifestyle differentiation. In the center, links between orthogroups indicate coexpression of associated COG families in fungi (STRING database [123]). **c** Coexpression network of gene families across published fungal transcriptomic datasets, built on discriminant orthogroups enriched in endophytes and mycobiota members and clustered with the MCL method.

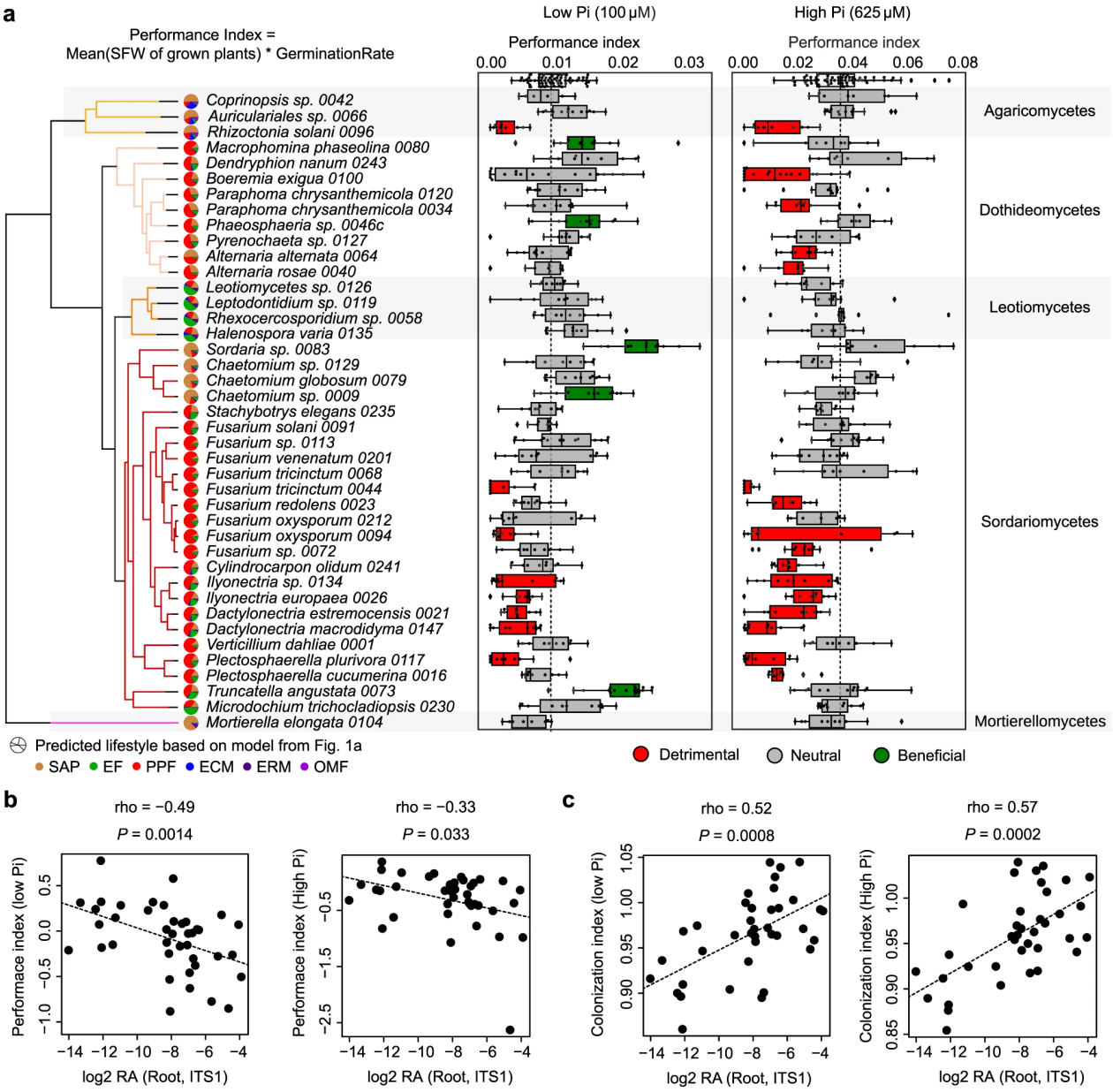


3.2.4 Genomic traits of the endophytic lifestyle

To identify unique genetic determinants characterizing both known endophytes and *A. thaliana* root mycobiota members, the 120 genomes were mined for gene families whose copy numbers allow efficient segregation of these fungi ($n = 50$) from those with other lifestyles ($n = 70$). We trained

a Support Vector Machines classifier with Recursive Feature Elimination (SVM-RFE) on the gene counts of orthogroups significantly enriched or depleted between these two groups (ANOVA, $FDR < 0.05$). A minimal set of 84 gene families that best segregated the two lifestyle groups was retained in the final SVM-RFE classifier ($R^2 = 0.80$, Fig. 3a and Supplementary Data 4a). These orthogroups can explain lifestyle differentiation independently from phylogenetic signal (PhyloGLM [124] – 83/84, $FDR < 0.05$) and were significantly enriched in enzymes (i.e., GO catalytic activity, GOATOOLS [125] $FDR = 0.002$, Supplementary Data 4b) and in CAZymes (one-sided Fisher Exact Test, $oddsratio = 7.45$, $P = 0.03$). Notably, genes encoding PCWDEs acting on pectin (CE12, GH145, PL11), cellulose (AA9), and hemicellulose (i.e., xylan: GH10, GH16, CE1) were identified, together with others encoding peptidases, transporters and proteins involved in amino acid metabolism (Fig. 3b and Supplementary Data 4a). These 84 gene families were analyzed for co-expression in published fungal transcriptomic datasets gathered in the database STRING [123]. An MCL-clustered co-expression network built on families enriched in known endophytes and *A. thaliana* mycobiota members revealed six clusters of co-expressed genes (Fig. 3c), including carbohydrate membrane transporters, and genes involved in carbohydrate metabolism (e.g., GH10) and amino acid metabolism. These functions are likely to be essential for endophytic root colonization.

Figure 4: Linking fungal outcome on host performance with root colonization patterns. **a** Performance indices (shoot fresh weights of 4-week-old plants normalized by germination rate) of *A. thaliana* plants recolonized with each of the 41 fungal strains on media containing low and high concentrations of orthophosphate (Pi). At least three independent biological replicates resulting in 2–4 values each were performed for each fungus ($n = 6$–18). Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). Differential fungal effects on plant performance were tested on both media with Kruskal–Wallis (at high and low Pi: $P < 2.2e-16$) and beneficial and pathogenic strains were identified by a two-sided Dunn test against mock-treated plants (first row in boxplots). Vertical dash lines indicate the mean performance of mock-treated plants. Left to the boxplots is displayed the strain phylogeny, together with lifestyle probabilities predicted by the Random Forest classifier trained for ancestral lifestyle prediction in Fig. 2a. **b** Spearman's rank correlation of relative fungal abundances in root samples from natural populations (log2 RA, Fig. 1a, [23]) with fungal effects on plant performance at low Pi (left) and high Pi (right) (Hedges standard effect sizes standardizing all phenotypes to the ones of mock-treated plants). **c** Spearman rank correlation of relative fungal abundances in root samples from natural populations (log2 RA, Fig. 1a, [23]) with fungal colonization indices measured by quantitative PCR in our plant recolonization experiments at low Pi (left) and high Pi (right).



3.2.5 Root colonization capabilities explain fungal outcome on plant growth

Root-colonizing fungi can span along the endophytism-parasitism continuum [71, 76]. Consistently, our previously trained Random Forest lifestyle classifier ($R^2 = 0.70$, Fig. 2a) predicted our 41 mycobiota members to be either plant pathogens, endophytes or saprotrophs (Fig. 4a). We tested the extent to which the 41 fungi can modulate host physiology by performing binary interaction experiments with germ-free *A. thaliana* plants grown in two nutrient conditions under laboratory conditions (inorganic orthophosphate, Pi: $100\mu M$ and $625\mu M\, KH_2PO_4$, Fig. 4a). We identified that seed inoculation with the independent isolates influenced both germination rate (GR, Supplementary Fig. 9) and shoot fresh weight (SFW) of four-week-old plants ($n = 7127$), and

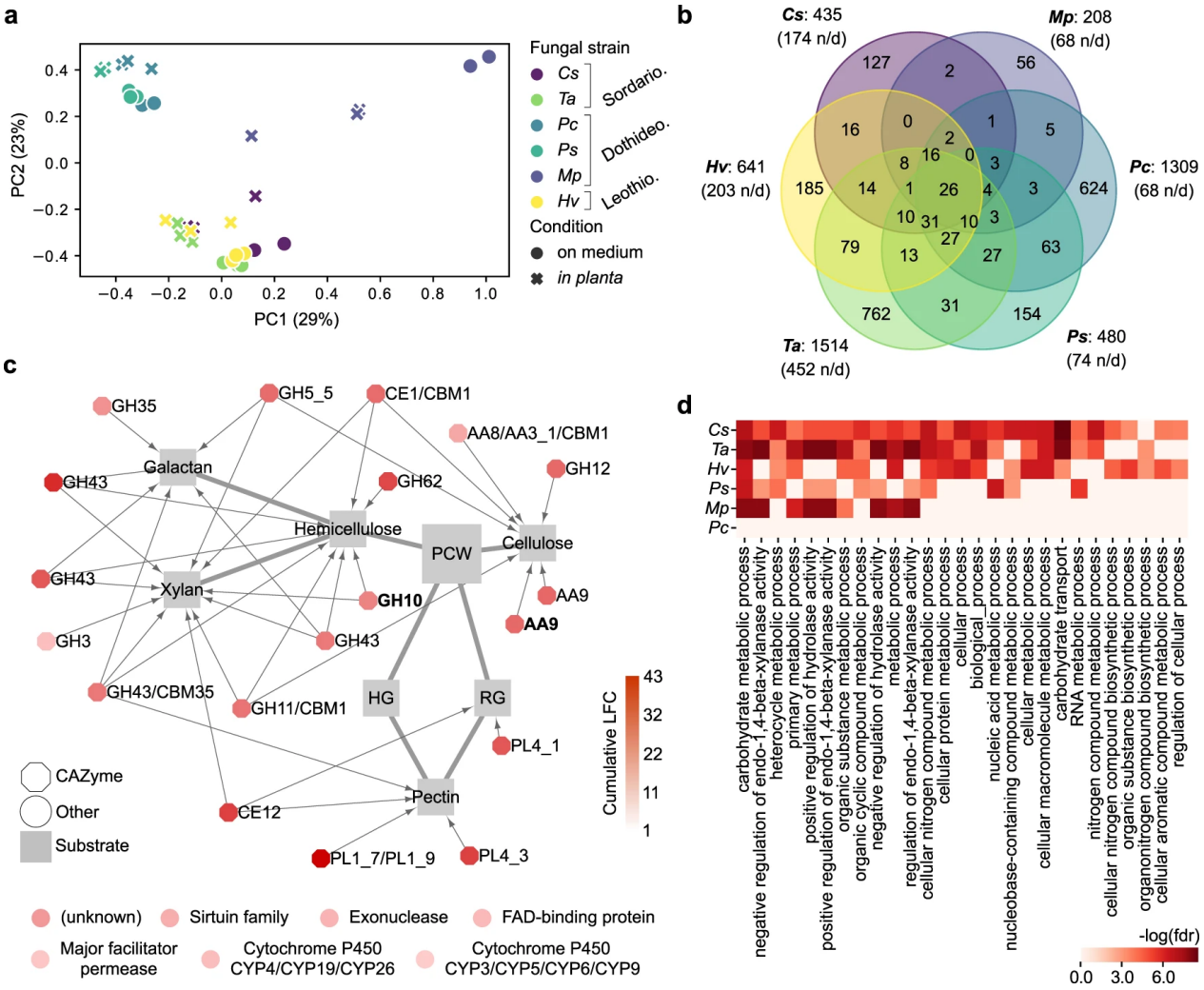
therefore calculated a plant performance index ($PPI = SFW * GR$, Methods). Under Pi-sufficient conditions, 39% of the isolates (16/41) negatively affected host performance compared to germ-free control plants, whereas 61% (25/41) had no significant effect on PPI (Kruskal–Wallis–Dunn Test, adj. $P < 0.05$, Fig. 4a). Fungal-induced change in PPI was significantly modulated by the nutritional status of the host, as depletion of bioavailable Pi in the medium was associated with a reduction in the number of fungi with pathogenic activities (20%, 8/41) and an increase of those with beneficial activities (12%, 5/41) (Kruskal–Wallis–Dunn Test, adj. $P < 0.05$, Fig. 4a). Notably, PPI measured for low and high Pi conditions was negatively correlated with strain RA in roots of European *A. thaliana* populations (Spearman, High Pi: $rho = -0.33$, $P = 0.033$; Low Pi: $rho = -0.49$, $P = 0.0014$, Fig. 4b), suggesting a potential link between the ability of a fungus to efficiently colonize roots and the observed negative effect on plant performance. Consistent with this hypothesis, fungal load measured by quantitative PCR in roots of four-week-old *A. thaliana* colonized by individual fungal isolates (Supplementary Fig. 10ab), was positively correlated with fungal RA in roots of natural populations (Spearman, High Pi: $rho = 0.57$, $P = 0.0002$; Low Pi: $rho = 0.52$, $P = 0.0008$, Fig. 4c), and was also negatively linked with PPI outcome (Spearman, High Pi: $rho = -0.44$, $P = 0.005$, Low Pi: $rho = -0.30$, $P = 0.057$) (Supplementary Fig. 10cd). Furthermore, a co-occurrence matrix based on the RA of ASVs corresponding to these isolates in naturally occurring root mycobiomes indicated that most taxa with neutral and detrimental effects often co-occurred in roots of European *A. thaliana* populations [23], whereas those with beneficial activities were rarely detected (Supplementary Fig. 11). Taken together, our results suggest that robust root colonizers have a high pathogenic potential, and that their colonization must be tightly controlled not to affect plant health.

3.2.6 A conserved set of CAZyme-encoding genes is induced in planta by diverse root mycobiota members

We tested whether putative genomic determinants of endophytism defined above by a machine learning approach were part of a core response activated in planta by root mycobiota members. Six representative fungi from three different phylogenetic classes were selected for in planta transcriptomics on low Pi sugar-free medium: *Chaetomium sp.* 0009 (*Cs*), *Macrophomina phaseolina* 0080 (*Mp*), *Paraphoma chrysantemicola* 0034 (*Pc*), *Phaeosphaeria sp.* 0046c (*Ps*), *Truncatella angustata* 0073 (*Ta*), *Halenospora varia* 0135 (*Hv*). Confocal microscopy of roots grown in mono-association with these fungi highlighted similar colonization of root surfaces and local penetrations of hyphae in epidermal cells (Supplementary Fig. 12). After mapping of RNA-seq reads on genome assemblies (Hisat2 [126]) and differential expression analysis (in planta vs. on medium, DE-Seq2 [127]), significant log2 fold-change (log2FC) values were summed by orthogroups, allowing between-strain transcriptome comparisons (Methods). Transcriptome similarity did not fully reflect

phylogenetic relationships since *Cs* and *Ta* (Sordariomycetes) clustered with *Hv* (Leotiomycete), whereas *Mp*, *Pc* and *Ps* (Dothideomycetes) showed substantial transcriptome differentiation (Fig. 5a). Although in planta transcriptional reprogramming was largely strain-specific, we identified a core set of 26 gene families that were consistently over-expressed by these distantly related fungi in *A. thaliana* roots (Fig. 5b). We observed a remarkable over-representation of genes coding for CAZymes acting on different plant cell wall components (i.e., 19/26, 73%), including cellulose, xylan and pectin (Fig. 5c). This set was also significantly enriched in families previously identified as putative determinants of endophytism by our SVM-RFE classifier (Fisher exact test, $P < 0.05$), including AA9 (lytic cellulose monooxygenase) and GH10 (xylanase) CAZyme families. Inspection of fungal genes over-expressed in planta by each strain (Supplementary Data 5), followed by independent GO enrichment analyses, corroborated that carbohydrate metabolic processes and xylanase activities were the most common fungal responses activated in planta (GOATOOLS, $FDR < 0.05$, Fig. 5d). Notably, we also observed important percentages of genes encoding effector-like SSPs induced in planta (9.8–42.4%, median $= 21.6\%$). Together, these enzymes and SSPs are likely to constitute an essential toolbox for *A. thaliana* root colonization and for fungal acquisition of carbon compounds from plant material. Analysis of corresponding *A. thaliana* root transcriptomes revealed that different responses were activated by the host as a result of its interaction with these six phylogenetically distant mycobiota members (Supplementary Fig. 13, Supplementary Data 6). Our data suggest that phylogenetically distant mycobiota members colonize *A. thaliana* roots using a conserved set of PCWDEs and have markedly different impacts on their host.

Figure 5: Comparative transcriptomics identified a core set of PCWDE-encoding genes induced in *A. thaliana* roots by diverse mycobiota members. **a** PCoA plot of Bray-Curtis distances calculated on gene family read counts from fungal transcriptome data on medium and in planta. *Cs=Chaetomium sp.* 0009, *Mp=Macrophomina phaseolina* 0080, *Pc=Paraphoma chrysantemicola* 0034, *Ps=Phaeosphaeria sp.* 0046c, *Ta=Truncatella angustata* 0073, *Hv=Halenospora varia* 0135. **b** Venn diagram showing the number of fungal gene families over-expressed in planta. It highlights 26 families commonly over-expressed by all six fungi (n/d: non-displayed interactions). **c** Commonly over-expressed gene families in planta (n=26), which include 19 plant cell-wall degrading CAZymes (octagons) linked to their substrates, as described in literature [69, 128]. The two CAZyme families highlighted in bold were identified as potential determinants of endophytism (SVM-RFE, Fig. 3a). The seven remaining (non-CAZyme) families are shown below the network. **d** Individual GO enrichment analyses performed on the genes over-expressed in planta vs. on medium by each fungal strain (GOATOOLS [125], $FDR < 0.05$).



3.2.7 Polysaccharide lyase family PL1_7 as a key component linking colonization aggressiveness to plant health

We reported above a potential link between aggressiveness in root colonization and detrimental effect of fungi on PPI. To identify underlying genomic signatures explaining this link, we employed three different methods. First, inspection of diverse gene categories across genomes of beneficial, neutral, and detrimental fungi revealed significant enrichments in CAZymes (especially polysaccharide/pectate lyases, PLs) and proteases in the genomes of detrimental fungi (Low Pi conditions, Kruskal–Wallis $P < 0.05$, and Dunn tests, Supplementary Fig. 14a, b). In these categories, three pectate lyases (PL1_4, PL1_7, PL3_2) and three peptidases (S08A, A01A, S10) contributed the most in segregating genomes by effect on plants (see the count in gene copy in Supplementary Fig. 14c). Second, multiple testing of association between secreted CAZyme counts ($n = 199$ families in total) and fungal effect on PPI identified the PL1_7 family as the only family significantly linked to detri-

mental effects (ANOVA, Bonferroni; Low Pi: $P = 0.026$; High Pi: not significant; Fig. 6a). Finally, an SVM-RFE classifier was trained on the gene counts of all orthogroups that were significantly enriched or depleted in genomes of detrimental vs. non-detrimental fungi (ANOVA, $FDR < 0.05$). While this method failed at building a classifier to predict detrimental effects at high Pi (no families significantly enriched/depleted), it successfully predicted detrimental effects at low Pi with very high accuracy ($R^2 = 0.88$). A minimal set of 11 orthogroups discriminating detrimental from non-detrimental fungi was identified (Fig. 6b, Supplementary Data 7), and includes gene families encoding membrane transporters, zinc-finger domain-containing proteins, a salicylate monooxygenase and a PL1 orthogroup containing the aforementioned PL1_7 CAZyme subfamily and related PL1_9 and PL1_10 subfamilies. Further phylogenetic instability analysis based on duplication and mutation rates (MIPhy [129]) identified PL1_9 and PL1_10 as slow-evolving clades in the gene family tree (instability = 30.94 and 18.86 respectively, Fig. 6c), contrasting with most PL1_7 genes that were located in two rapidly evolving clades (index = 85.30 and 66.12). Of note, genomic counts of PL1_7, but not PL1_9/10, remained significantly associated to detrimental host phenotypes after correction for the phylogenetic signal in our dataset (PhyloGLM [124], $FDR = 0.03$). PL1_7 was also part of the core transcriptional response activated in planta by six non-detrimental fungi (Fig. 5c) and was enriched in mycobiota members and endophytes in comparison to saprotrophs and mycorrhizal fungi (Supplementary Fig. 14d). Therefore, degradation of pectin by root mycobiota members is likely crucial for penetration of—and accommodation in—pectin-rich *A. thaliana* cell walls. However, the remarkable expansion of this gene family in detrimental compared to non-detrimental fungi predicts a possible negative link between colonization aggressiveness and plant performance. To test this hypothesis, we took advantage of the *Trichoderma reesei* QM9414 strain (WT, PL1_7-free background) and its corresponding heterologous mutant lines over-expressing *pel12*, a gene from *Clonostachys rosea* encoding a PL1_7 pectate lyase with direct enzymatic involvement in utilization of pectin [130]. By performing plant recolonization experiments at low Pi with these lines, we observed that *T. reesei pel12OE* lines negatively affected PPI with respect to their parental strain (ANOVA and TukeyHSD test, $P < 0.05$ for two out of three independent over-expressing lines, Fig. 6d), and this phenotype was associated with a significant increase in fungal load in plant roots (Kruskal–Wallis and Dunn test, $P < 0.05$, Fig. 6e). Taken together, our data indicate that pectin-degrading enzymes belonging to the PL1_7 family are key fungal determinants linking colonization aggressiveness to plant health.

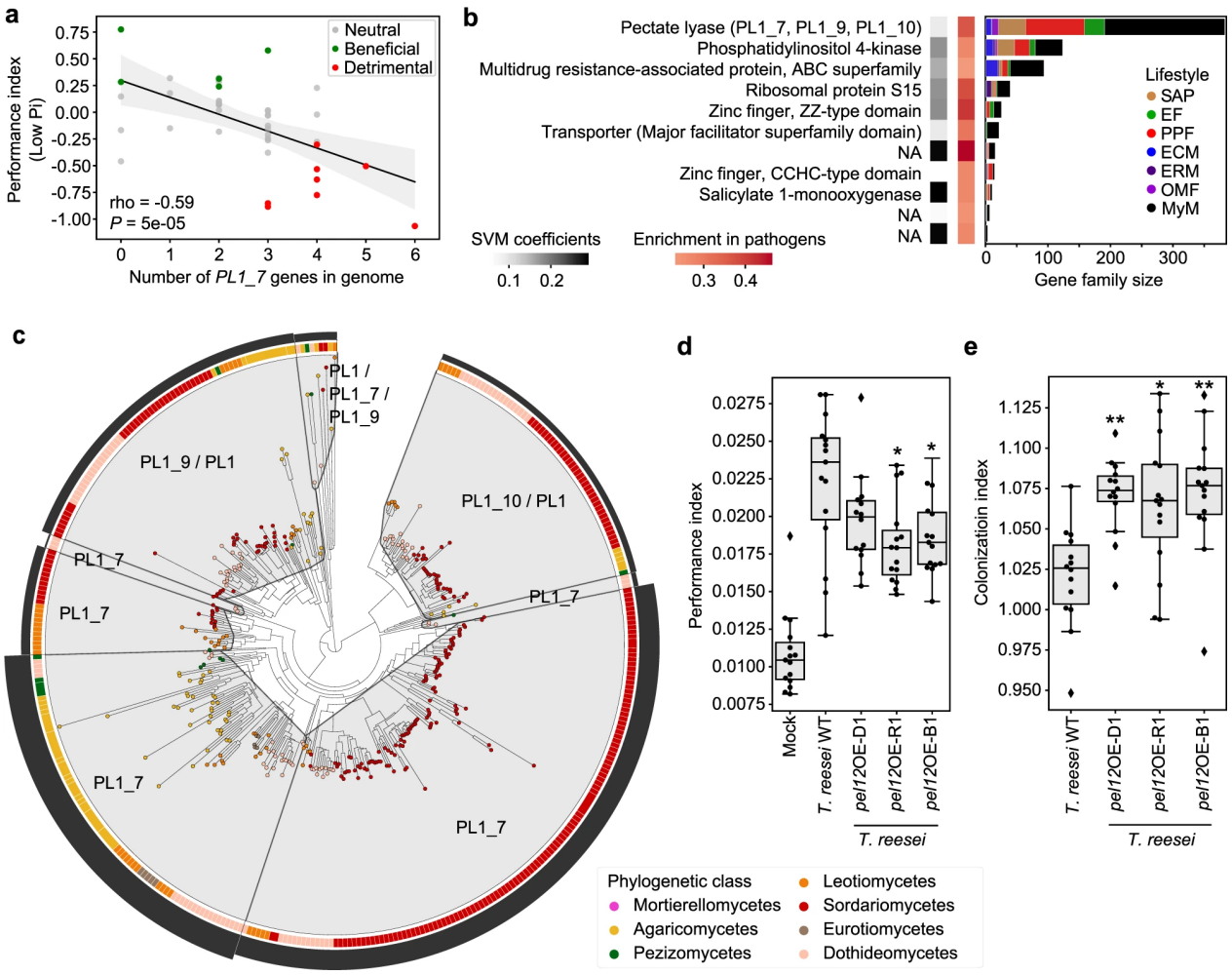


Figure 6: Genomic content in polysaccharide lyase PL1_7 links colonization aggressiveness to plant health. **a** Spearman's rank correlation between the number of genes encoding secreted PL1_7 in fungal genomes and the plant performance index at low Pi in recolonization experiments. **b** Minimal set of 11 gene families discriminating detrimental from non-detrimental fungi at low Pi (SVM-RFE $R^2 = 0.88$). The first heatmap on the left shows the SVM coefficients, reflecting the contribution of each orthogroup to the separation of the two groups, whereas the heatmap on the right shows the enrichment of these gene families in fungi identified as detrimental in recolonization experiments at low Pi. Gene family sizes and representation in the different lifestyles are shown on the barplots in the context of the whole fungal dataset ($n = 120$). NA: no functional annotation. **c** Protein family tree of the polysaccharide lyase orthogroup identified as essential for segregating detrimental from non-detrimental fungi in our SVM-RFE classification model. The tree was reconciled with fungal phylogeny and clustered into minimum instability groups by MIPhy [129]. Each group is labeled with its CAZyme annotation. The outer circle (black barplot) depicts the relative instabilities of these groups, suggesting two rapidly evolving PL1_7 groups in Sordariomycetes and Agaricomycetes. **d** Plant performance indices resulting from plant recolonization experiments at low Pi (three independent biological replicates), conducted with *Trichoderma reesei* QM9414 (WT) and three independent heterologous mutant lines (D1, R1, B1) overexpressing *pel12* from *Clonostachys rosea* (PL1_7 family, [130]). Asterisks indicate significant difference to WT, according to ANOVA ($P = 1.45e-12$) and a two-sided TukeyHSD test (WT vs. D1: adjusted $P = 0.28$; WT vs. B1: adjusted $P = 3.75e-2$; WT vs. R1:

adjusted $P = 1.19e-2$). **e** Fungal colonization measured by qPCR in colonized roots at low Pi, conducted with *T. reesei* WT and three *pel12* overexpression mutant lines. Asterisks indicate significant difference to *T. reesei* WT, according to Kruskal-Wallis ($P = 6.25e-4$) and a two-sided Dunn test (WT vs. D1: adjusted $P = 2.4e-3$; WT vs. B1: adjusted $P = 1.6e-3$; WT vs. R1: adjusted $P = 1.5e-2$). For both **d** and **e**, three independent biological replicates were performed resulting in $n = 15$ data points per condition. Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). Asterisks highlight the results of post hoc tests: **adjusted $P < 0.01$, *adjusted $P < 0.05$.

3.3 Discussion

We report here that genomes of fungi isolated from roots of healthy *A. thaliana* harbor a remarkable diversity of genes encoding secreted proteins and CAZymes. Consistent with the fact that these fungi were (1) isolated from surface-sterilized root fragments [87], (2) enriched in plant roots vs. surrounding soil samples at a continental scale [23] (Fig. 1), and (3) able to recolonize roots of germ-free plants (Supplementary Figs. 10 and 12), both the diversity and the composition of their gene repertoires resemble those of previously described endophytes [79, 81, 115] (Fig. 2). Unlike the remarkable loss in PCWDE-encoding genes in the genomes of most ectomycorrhizal fungi [69, 108], endophytism in root mycobiota members is therefore not associated with genome reduction in saprotrophic traits, as previously suggested [77]. Using a machine learning approach, together with in planta transcriptomic experiments, we identified genes encoding CAZyme families AA9 (copper-dependent lytic polysaccharide monooxygenases, acting on cellulose chains) and GH10 (xylanase) as potential determinants of endophytism (Figs. 3 and 5). Interestingly, these same families were strongly expanded in genomes of beneficial root mutualists belonging to Serendipitaceae [77, 111] compared to mycorrhizal mutualists [108] and might therefore represent key genetic components explaining adaptation to — and accommodation in — *A. thaliana* roots. It is important to note that although the 41 isolates are representative of naturally occurring *A. thaliana* root mycobiomes, a large fraction of fungi could not be included in this comparative analysis, including isolates that cannot be cultured. Therefore, it remains to test whether the genomic signatures observed here for this restricted, yet diverse set of cultured fungi, are retained across a broader range of taxonomically diverse root endophytes.

Although the 41 *A. thaliana* root mycobiota members were isolated from roots of healthy-looking plants, experiments in mono-associations with the host revealed a diversity of effects on plant performance, ranging from highly pathogenic to highly beneficial phenotypes (Fig. 4). These results are consistent with the previous reports [74, 87, 107, 131] and suggest that the pathogenic potential of detrimental fungal endophytes identified based on mono-association experiments with

the host, is largely kept at bay in a community context by the combined action of microbiota-induced host defenses and microbe-microbe competition at the soil-root interface [3, 12, 31, 87, 132]. However, we observed that robust and abundant fungal colonizers of *A. thaliana* roots defined from a continental-scale survey of the root microbiota [23] were dominated by detrimental fungi defined based on mono-association experiments with the host (Fig. 4). Based on quantitative PCR data, we also observed that fungi with beneficial activities on plant health were colonizing roots less aggressively than those with detrimental activities—as previously reported [74], suggesting a potential link between fungal colonization capabilities, abundance in natural plant populations, and plant health. A potential limitation of our qPCR-based amplification approach with the general ITS1F-ITS2 primers is linked to the fact that there is copy number variation in rDNA ITS across fungal genomes and that primer bias might distort relative fungal load measurements, thereby making direct comparisons between fungal isolates difficult [133]. Irrespective of this limitation, our results support the idea that maintenance of fungal load in plant roots is critical for plant health, and that controlled fungal accommodation in plant tissues is key for the maintenance of homeostatic plant-fungal relationships. This conclusion is indirectly supported by the fact that an intact innate immune system is needed for the beneficial activities of fungal root endophytes [12, 77, 78]. Our results, therefore, suggest that the most beneficial root mycobiota members are not necessarily the most abundant in roots of natural plant populations. In contrast, understanding how potential pathogens can dominate the endospheric microbiome of healthy plants is key for predicting disease emergence in natural plant populations [134, 135].

To identify genetic determinants explaining the link between colonization aggressiveness and detrimental effect on plant performance, we used different association methods that all converged into the identification of the CAZyme subfamily PL1_7 as one of the potential underlying determinants of this trait. Proteins from the PL1_7 family were previously characterized in different *Aspergillus* species as metabolizing pectate by eliminative cleavage of $(1 \rightarrow 4)$-α-D-galacturonan (EC 4.2.2.2) [136, 137]. Furthermore, primary cell walls of *A. thaliana* are enriched with pectin compared to those of monocotyledonous plants, which contain more hemicellulose and phenolics [4, 5]. Therefore, repertoire diversity in pectin-degradation capabilities is likely key for penetration and accommodation in pectin-rich *A. thaliana* cell walls. This is corroborated by the observation that non-detrimental fungal endophytes were also shown to consistently induce expression of this gene family in planta during colonization of *A. thaliana* roots (Fig. 5). However, re-inspection of previously published transcriptomic data indicated that genes encoding PL1_7 were induced more extensively in planta by the fungal root pathogen *Colletotrichum incanum* compared to that of its closely relative beneficial root endophyte *Colletotrichum tofieldiae* [79]. Therefore, differences in expression and diversification of this gene family are potential contributors to the differentiation between detrimental and non-detrimental fungi in the *A. thaliana* root mycobiome, especially since *A.*

thaliana cell-wall composition is a determinant factor for disease resistance [138, 139]. Notably, expansion of the PL1_7 gene family was observed in plant pathogens but also in the biocontrol fungus *C. rosea* (Sordariomycetes, Hypocreales), a fungal species with mycoparasitic and plant endophytic capacity [140, 141] that is phylogenetically closely related to multiple isolates selected in this study. Genetic manipulation of the *C. rosea pel12* gene revealed a direct involvement of the protein in pectin degradation, but not in *C. rosea* biocontrol towards the phytopathogen *Botrytis cinerea* [130]. Here, we showed that heterologous overexpression of *C. rosea pel12* in *T. reesei* does not only increase its root colonization capabilities, but also modulates fungal impact on plant performance. We, therefore, conclude that a direct link exists between expression/diversification of PL1_7-encoding genes in fungal genomes, root colonization aggressiveness, and altered plant performance. Our results suggest that the evolution of fungal CAZyme repertoires modulates root mycobiota assemblages and host health in nature.

3.4 Methods

3.4.1 Selection of 41 representative fungal strains

The 41 *A. thaliana* root mycobiota members were previously isolated from surface-sterilized root segments of *A. thaliana* and the closely related Brassicaceae species *Arabis alpina* and *Cardamine hirsuta*, as previously described [87]. Notably, this culture collection derived from fungi isolated from the roots of plants grown in the Cologne Agricultural Soil under greenhouse conditions, or from natural *A. thaliana* populations from two sites in Germany (Pulheim and Geyen) and one site in France (Saint-Dié des Vosges) [87] (Supplementary Data 1).

3.4.2 ITS sequence comparison with naturally occurring root mycobiome

Comparison of fungal ITS1 and ITS2 sequences with corresponding sequence tags from a European-scale survey of the *A. thaliana* mycobiota (17 European sites [23]) was carried out. For all 41 Fungi, sequences of the internal transcribed spacer 1 and 2 (ITS1/ITS2) were retrieved from genomes (`https://github.com/fantin-mesny/Extract-ITS-sequences-from-a-fungal-genome`) or, in the cases where no sequences could be found, via Sanger sequencing (4 of 41). All ITS sequence variants were directly aligned to the demultiplexed and quality filtered reads from previously published datasets [23] using USEARCH [142] v10.0.240 at a 97% similarity cut-off. A count table across all samples was constructed using the results from this mapping and an additional row representing all the reads that did not match any of the reference sequences was added. This additional row was based on the count data from the amplicon sequence variant (ASV) analysis from the original study, whereas the read counts from the new mapping were subtracted sample wise. To have coverage-independent information on the RA of each fungus, we calculated RA only

for the root samples where the respective fungi were found ($RA > 0.01\%$). The sample coverage was calculated across all root samples (>1000 reads, $n = 169$). Enrichment in roots was calculated for all root and soil samples (>1000 reads, $n = 169$ / $n = 223$) using the Mann–Whitney U test ($FDR < 0.05$). In the same way the RA and coverage across leaf samples from two *A. thaliana* populations [32] was calculated (two locations in Germany, samples $n = 51$). For this specific analysis of leaf samples, only ITS2 sequences were used and no fold change was calculated. In order to estimate the presence of the 41 fungi across worldwide collected samples, we used the GlobalFungi database [112] (`https://globalfungi.com`, version August 2020). The most prevalent ITS1 sequences from each genome were used to conduct a BLAST search on the website. Sample metadata for the best matching representative species hypothesis sequences were then used to determine the global sample coverage. Appearance across samples from type root/shoot was counted for each fungus and compared to the total number of root/shoot samples for each continent.

3.4.3 Whole-genome sequencing and annotation

Forty-one fungal isolates from a previously assembled culture collection2 were revived from 30% glycerol stocks stored at -80°C. Genomic DNA extractions were carried out from mycelium samples grown on Potato extract Glucose Agar (PGA) medium, with a previously described modified cetyltrimethylammonium bromide protocol [108]. Genomic DNA was sequenced using PacBio systems. Genomic DNA was sheared to 3kb, > 10kb, or 30kb using Covaris LE220 or g-Tubes or Megaruptor3 (Diagenode). The sheared DNA was treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of blunt adapters using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). The library was purified with AMPure PB beads and size selected with BluePippin (Sage Science) at > 10kb cutoff size. Sequencing was done on PacBio RSII or SEQUEL machines. For RSII sequencing, PacBio Sequencing primer was annealed to the SMRTbell template library and sequencing polymerase was bound to them. The prepared SMRTbell template libraries were sequenced on a Pacific Biosciences RSII or Sequel sequencers using Version C4 or Version 2.1 chemistry and $1*240$ or $1*600$ sequencing movie run times, respectively. The genome assembly was generated using Falcon [143] v0.7.3 with mitochondria-filtered reads. The resulting assembly was improved with finisherSC, and polished with either Quiver or Arrow. Transcriptomes were sequenced using Illumina Truseq Stranded RNA protocols with polyA selection (`http://support.illumina.com/sequencing/sequencing_kits/truseq_stranded_mrna_ht_sample_prep_kit.html`) on HiSeq2500 using HiSeq TruSeq SBS sequencing kits v4 or NovaSeq6000 using NovaSeq XP v1 reagent kits, S4 flow cell, following a $2*150$ indexed run recipe. After sequencing, the raw fastq file reads were filtered and trimmed for quality (Q6), artifacts, spike-in, and PhiX reads and assembled into consensus sequences using Trinity [144] v2.1.1.

The genomes were annotated using the JGI Annotation pipeline [145]. Species assignment was conducted by extracting ITS1 and ITS2 sequences from genome assemblies, performing a similarity search against the UNITE database [146] (`https://unite.ut.ee`, version February 2021) and a phylogenetic comparison to fungal genomes on MycoCosm [145] (`https://mycocosm.jgi.doe.gov`).

3.4.4 Comparative genomics dataset

In addition to our 41 fungal isolates from *A. thaliana* roots, we used 79 previously published fungal genomes in a comparative genomics analysis (Supplementary Data 2). While 77 genomes and annotations were downloaded from MycoCosm, the genome assemblies of fungal strains *Harpophora oryzae* R5-6-134 and *Helotiales sp.* F22930 were downloaded from NCBI (GenBank assembly accessions GCA_000733355.1 and GCA_002554605.1 respectively) and annotated with FGENESH [147] v8.8.0. Lifestyles were associated to each single strain by referring to the original publications describing their isolation, and consulting the FunGuild [113] database with the species and genus names associated to each strain. Orthology prediction was performed on this data set of 120 genomes by running OrthoFinder [119] v2.2.7 with default parameters. From this prediction, we used the generated orthogroups data, the species tree, and gene trees. OrthoFinder was also run on our 41 newly sequenced fungi to obtain a second species tree, for this subset.

3.4.5 Predicting ancestral lifestyles

To identify gene family gains and losses events, GLOOME [148] gainLoss.VR01.266 was run using the species tree and presence/absence of each orthogroup in the 120 genomes. To obtain reconstruction of ancestral genomes using the Wagner parsimony approach, Count [120] v10.04 was run using these same inputs. To associate a lifestyle to each reconstructed ancestral genome, a Random Forest classifier was trained on the copy numbers of each orthogroup in the comparative genomics data set excluding *A. thaliana* mycobiota members, and the fungal lifestyles associated to these 79 genomes. This was performed using the RandomForestClassifier() function of the Python library sklearn [149] v0.20.3. The accuracy of the model was estimated by a leave-one-out cross-validation approach, computed using the function cross_val_score(cv=KFold(n_splits=120)) in sklearn. Finally, the probabilities of ancestors to belong in each lifestyle category were retrieved using function predict_proba().

3.4.6 Genomic feature analyses

Statistics of genome assemblies (i.e., N50, number of genes and scaffolds and genome size) were obtained from JGI MycoCosm [145], and assembly-stats (`https://github.com/sanger-pathogens/`

assembly-stats). Genome completeness with single copy orthologues was calculated using BUSCO v3.0.2 with default parameters [150]. The coverage of transposable elements in genomes was calculated and visualized using a custom pipeline Transposon Identification Nominative Genome Overview (TINGO[151]). The secretome was predicted as described previously [122]. We calculated, visualized, and compared the count and ratio of total (present in the genomes) and predicted secreted CAZymes [128], proteases [152], lipases [153], and small secreted proteins [122] (SSPs) (¡300 amino acid) as a subcategory. We calculated the total count of the followings using total and predicted secreted plant cell-wall degrading enzymes (PCWDEs) and fungal cell-wall degrading enzymes (FCWDEs). Output files generated above were combined and visualized with a custom pipeline, Proteomic Information Navigated Genomic Outlook (PRINGO [69]). To compare the genomic compositions of the different lifestyle categories while taking into account phylogenetic signal, we first generated a matrix of pairwise phylogenetic distances between genomes (i.e. sum of branch lengths) using the function tree.distance() from package biopython Phylo [154], then computed a principal component analysis using the PCA(n_components=4) function of sklearn [149] v0.20.3. Components PC1, PC2, PC3 and PC4 (Supplementary Fig. 3) were then used to compare the per-genome numbers of CAZymes, proteases, lipases, SSPs, PCWDEs, and FCWDEs in the different lifestyles with an ANOVA test and a TukeyHSD post hoc test. R function aov() was used with the following formula specifying the model:

$$GeneCount \sim PC1 + PC2 + PC3 + PC4 + Lifestyle$$
$$+PC1 : Lifestyle + PC2 : Lifestyle + PC3 : Lifestyle + PC4 : Lifestyle$$

Differences in subfamily composition of the groups of genes of interest were then carried out using a PERMANOVA-based approach (https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition). This approach relies on Jaccard distances calculation (best suited for discrete variables such as copy numbers) then a PERMANOVA testing with function adonis2() from R package Vegan v2.5-7 (https://github.com/jarioksa/vegan), with the model specified by the following formula:

$$JaccardDistanceMatrix \sim PC1 + PC2 + PC3 + PC4 + Lifestyle$$
$$+PC1 : Lifestyle + PC2 : Lifestyle + PC3 : Lifestyle + PC4 : Lifestyle$$

Post hoc testing with function pairwise.perm.manova() from package RVAideMemoire v0.9-77 (https://cran.r-project.org/web/packages/RVAideMemoire) was then performed to compare pairs of lifestyles.

For each Jaccard matrix, we used the function dbRDA() from the R package Vegan, to calculate two distance-based redundancy analyses (dbRDA), respectively constrained by phylogenetic variables (formula $Jaccard \sim Condition(Lifestyle) + PCs$) and by lifestyle groups (formula $Jaccard \sim Condition(PCs) + Lifestyle$).

We determined genes discriminating groups based on the principal coordinates of a regularized discriminant analysis calculated from the count of genes coding for CAZymes, proteases, lipases, and SSPs, with R function rda(). We then used Vegan function scores() on the three first principal coordinates, and kept for each coordinate the top five high-loading gene discriminating groups.

3.4.7 Determinants of endophytism

To identify a small set of orthogroups that best segregate endophytes and mycobiota members from fungi with other lifestyles, we standardized the orthogroup gene counts with function StandardScaler() from sklearn [149] v0.20.3. Then, orthogroups that are enriched or depleted in the fungi of interest were selected with function SelectFdr(f_classif, alpha=0.05) from sklearn. On this subset of orthogroups, we trained a Support Vector Machine classifier with Recursive Feature Elimination (SVM-RFE). This was performed with functions from sklearn SVC(kernel='linear') and RFECV(step=10, cv=KFold(n_splits=120, min_features_to_select=10)), which implement a leave-one-out cross-validation allowing the estimation of the classifier accuracy at each step of the recursive orthogroup elimination. PhyloGLM models [124] were built with R package phylolm v.2.6.2 on the two groups of interest and orthogroup gene counts, with parameters $btol = 45$ and $log.alpha.bound = 7$, and the logistic_MPLE method. Further analysis of the gene families segregating fungi of interest from others ($n = 84$) was carried out by identifying a representative sequence of each orthogroup in our SVM-RFE model, and studying both its annotation and coexpression data in databases. To identify representative sequences, all protein sequences composing an orthogroup were aligned with FAMSA [155] v1.6.1. Using HMMER [156] v3.2.1, we then built a Hidden Markov Model (HMM) from this alignment with function hmmbuild, then ran function hmmsearch looking for the best hit matching this HMM within the proteins composing our orthogroup. We then considered this best hit as a representative sequence of the orthogroup and analyzed its annotation. GO enrichment analysis was performed by running GOATOOLS [125] v1.0.3 using the GO annotations associated to the representative sequences. To obtain coexpression data linking the orthogroups retained in our SVM-RFE model, we searched the String-db [123] website (`https://string-db.org`, version August 2020) for COG protein families matching our set of representative protein sequences in fungi. Each protein was associated to one COG (Supplementary Data 4a), and coexpression data were downloaded. A coexpression network was then built on the families enriched in endophytes and mycobiota members ($n = 73$) and clustered with algorithm MCL (granularity = 5) using Cytoscape [157] v3.7.2 and clusterMaker2 [158] v1.3.1.

3.4.8 Plant recolonization experiments assessing the effect of each fungal strain on plant growth

A. thaliana seeds were sterilized 15min in 70% ethanol, then 5min in 8% sodium hypochlorite. After six washes in sterile double-distilled water and one wash in 10mM $MgCl_2$, they were stratified 5–7 days at 4°C in the dark. Seed inoculation with fungal strains was carried out by crushing 50mg of mycelium grown for 10 days on Potato extract Glucose Agar medium (PGA) in 1ml of 10mM $MgCl_2$ with two metal beads in a tissue lyser, then adding $10\mu M$ of this inoculum in $250\mu l$ of seed solution for 5min. Seeds were then washed twice with $MgCl_2$ before seven were deposited on each medium-filled square Petri plate. Mock-inoculated seeds were also prepared by simple washes in $MgCl_2$. The two media used in this study — 625 and 100μM Pi — were previously described [159]. They were prepared by mixing 750μM $MgSO_4$, 625μM/100μM KH_2PO_4, 10,300μM NH_4NO_3, 9400μM KNO_3, 1500μM $CaCl_2$, 0.055μM $CoCl_2$, 0.053μM $CuCl_2$, 50μM H_3BO_3, 2.5μM KI, 50μM $MnCl_2$, 0.52μM Na_2MoO_4, 15μM $ZnCl_2$, 75μM Na-Fe-EDTA, and 1000μM MES pH5.5, 0μM/525μM KCl, then adding Difco Agar (ref. 214530, 1% final concentration), and finally adapting the pH to 5.5 prior to autoclaving. Plants were grown for 28 days at 21°C, for 10h with light (intensity 4) at 19°C and 14h in the dark in growth chambers. While roots were harvested and flash-frozen, SFW was measured for each plant. To distinguish seeds that did not germinate from plants that could not develop because of a fungal effect, we introduced a per-plate PPI corresponding to the average SFW of grown plants multiplied by the proportion of grown plants. In further correlation analyses, we used plant-performance indexes normalized to mock controls (standard effect sizes) using the Hedges' g method [160].

3.4.9 Fungal colonization of roots assay

Frozen root samples (one per plate) were crushed and total DNA was extracted from them using a QIAGEN Plant DNEasy Kit. Fungal colonization of these root samples was then measured by quantitative PCR. For each sample, two reactions were conducted with primers ITS1F (5-CTTGGTCATTTAGAGGAAGTAA-3) and ITS2 (5-GCTGCGTTCTTCATCGATGC-3) which target the fungal ITS1 sequence, and two with primers UBQ10F (5-TGTTTCCGTTCCTGTTATCT-3) and UBQ10R (5-ATGTTCAAGCCATCCTTAGA-3) that target the *Ubiquitin10 A. thaliana* gene. Each reaction was performed by mixing 5μl of iQ™ SYBR® Green Supermix with 2μl of 10μM forward primer, 2μl of 10μM reverse primer and 1μl of water containing 1ng template DNA. A BioRad CFX Connect Real-Time system was used with the following programme: 3min of denaturation at 95°C, followed by 39 cycles of 15sec at 95°C, 30s at 60°C and 30s at 72°C. We then calculated a single colonization index for each sample using the following formula: $Index = 2^{-(Cq(ITS1)/Cq(UBQ10))}$.

3.4.10 Confocal microscopy of root colonization by fungi

Roots of plants grown for 28 days in mono-association with fungi were harvested and conserved in 70% ethanol. They were then rinsed in ddH2O, and stained with propidium iodide (PI) and wheat germ agglutinin conjugated to fluorophore Biotium CF®488 (WGA-CF488). This was carried out by dipping the root samples for 15min in a solution of 20μg/ml PI and 10μg/ml WGA-CF488 buffered at pH 7.4 in phosphate-buffered saline (PBS). Samples were then washed in PBS and imaged with a Zeiss LSM700 microscope and the associated software ZEN v2.3 SP1.

3.4.11 Plant-fungi interaction transcriptomics

Dual RNAseq of six different plant-fungi interactions was carried out by performing three independent plant recolonization experiments on our low Pi medium, as described above. Total roots per plates were harvested after 28 days in culture, flash frozen, and crushed in a tissue lyser, and then total RNA was extracted with a QIAGEN RNeasy Plant Mini kit. As a control condition, sterile Nucleopore Track-Etched polyester membranes were deposited on low Pi medium, then 10μl drops of fungal inoculum (50mg/ml of mycelium in 10mM MgCl2) were placed on each one. The membranes were collected and processed as the root samples of our test condition. PolyA-enrichment was carried out on the RNA extracts, then an RNAseq library was prepared with the NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was then performed in single read mode on a HiSeq 3000 system. RNAseq reads were trimmed using Trimmomatic [161] v0.38 and parameters TRAILING:20 AVGQUAL:20 HEADCROP:10 MINLEN:100. We then used HiSat2 [126] v2.2.0 to map the trimmed reads onto reference genomes. Six independent HiSat2 indexes were prepared, each based on the TAIR10 *A. thaliana* genome and one of the six fungal genome assemblies of interest. We then performed six mappings, and counted the mapped reads using featureCounts [162] v2.0.0. RPKM (Reads Per Kilobase Million) values were computed from the featureCounts output. Differential gene expression analyses were then carried out on these counts using DESeq2 [127] v1.24.0. log2FC values were corrected by shrinkage with the algorithm apeglm [163] v1.6.0. To compare the transcriptomes of the six different fungi, significant log2FC values were summed per orthogroup. For each orthogroup, we used annotation of the most representative sequence, as previously described. GO enrichment analyses were carried out with GOATOOLS [125] v1.0.3, using the MycoCosm [145] GO annotation for fungi, and the TAIR annotation for *A. thaliana*.

3.4.12 Determinants of detrimental effects on plants and analysis of pectate lyases

Determinants of detrimental effects at low Pi were identified with the same method as previously described for determinants of endophytism/mycobiota: standard scaling of the orthogroup gene

counts, then training of an SVM classifier with RFE and leave-one-out cross validation. Instability analysis was carried out by submitting the species tree generated by OrthoFinder [119] to MIPhy [129] (`http://miphy.wasmuthlab.org`, version October 2020), together with the gene tree of our orthogroup of interest, with default parameters. PhyloGLM [124] models were built with R library phylolm v.2.6.2 on the two groups detrimental/non-detrimental and CAZyme gene counts, using our 41-genome species tree with default parameters and the logistic_MPLE method. *T. reesei* strain QM9414 and three heterologous overexpression lines of *pel12* generated previously [130], were revived on PGA medium and then inoculated into seeds for plant recolonization experiments on low Pi medium as previously described.

3.4.13 Statistics

Except for statistical methods described in the previous paragraphs, statistical testing was performed in R v3.5.1. Function aov() was used for ANOVA tests. Two-sided TukeyHSD post hoc testing was performed using function TukeyHSD(), which compares values associated to the different categories of one factor, respective of the variance that was attributed to this factor by the previous ANOVA test. When data were abnormally distributed, the non-parametric Kruskal–Wallis test was used by running function kruskal.test(), and the two-sided Dunn post hoc test was performed with function DunnTest() from package DescTools v0.99.28 (`https://github.com/AndriSignorell/DescTools/`).

3.5 Data availability

The genomic data generated in this study have been deposited in the GenBank database under the following BioProject accession codes: PRJNA371205 (assembly JAHBNJ000000000), PRJNA347188 (assembly JAHBNI000000000), PRJNA441695 (assembly JAHBNH000000000), PRJNA370201 (assembly JAGJXA000000000), PRJNA571620 (assembly JAGIZQ000000000), PRJNA370120 (assembly JAHBOE000000000), PRJNA347200 (assembly JAHBOF000000000), PRJNA371203 (assembly JAGPYM000000000), PRJNA370196 (assembly JAGMUU000000000), PRJNA500112 (assembly JAGMUV000000000), PRJNA370194 (assembly JAGMWT000000000), PRJNA455444 (assembly JAHBOG000000000), PRJNA370199 (assembly JAHBOO000000000), PRJNA347190 (assembly JAHEWL000000000), PRJNA455442 (assembly JAHEVI000000000), PRJNA347185 (assembly JAGMUX000000000), PRJNA370198 (assembly JAGTJS000000000), PRJNA347189 (assembly JAGPXF000000000), PRJNA455443 (assembly JAGMVH000000000), PRJNA500113 (assembly JAGMVI000000000), PRJNA347186 (assembly JAGPNQ000000000), PRJNA347191 (assembly JAHLEZ000000000), PRJNA370195 (assembly JAGTJR000000000), PRJNA370119 (assembly JAGTJQ000000000), PRJNA347187 (assembly JAGSXK000000000),

PRJNA500111 (assembly JAHEWK000000000), PRJNA347192 (assembly JAGTJP000000000), PRJNA347193 (assembly JAGMWG000000000), PRJNA538399 (assembly JAGMVJ000000000), PRJNA459235 (assembly JAGTJN000000000), PRJNA347194 (assembly JAGMVK000000000), PRJNA371204 (assembly JAGPXD000000000), PRJNA570880 (assembly JAGSXJ000000000), PRJNA347196 (assembly JAGTJM000000000), PRJNA347195 (assembly JAGTJL000000000), PRJNA371202 (assembly JAGTJO000000000), PRJNA370118 (assembly JAGPNK000000000), PRJNA370200 (assembly JAGPNJ000000000), PRJNA347197 (assembly JAGPXC000000000), PRJNA519173 (assembly JAHEWJ000000000), and PRJNA370197 (assembly JAHEWH000000000). The transcriptomic data generated in this study have been deposited in the Gene Expression Omnibus database under accession code GSE169629. The processed transcriptomic data are also available in this GEO entry. We referred to three online databases for analysis: UNITE (`https://unite.ut.ee`, version February 2021), GlobalFungi (`https://globalfungi.com`, version August 2020) and String-db (`https://string-db.org`, version August 2020). The plant phenotypic data and fungal colonization values are provided in the Source Data file. Source data are provided with this paper.

3.6 Code availability

All the scripts used for data processing and analysis were written in Python v3.7.3 and R v3.5.1 (except for transcriptomic analyses in which R v3.6.1 was used). Scripts are available at GitHub (`https://github.com/fantin-mesny/Scripts-from-Mesny-et-al.-2021`).

3.7 Acknowledgements

The sequencing project was funded by the U.S. Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231 within the framework of CSP 1974 "1KFG: Deep-sequencing of ecologically relevant Dikaria". This work was supported by funds to S.Hac from a European Research Council starting grant (MICRORULES 758003), the 'Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DECRyPT 2125)' and the Cluster of Excellence on Plant Sciences (CEPLAS), both funded by the Deutsche Forschungsgemeinschaft. F.M. salary was covered by the DECRyPT 2125 programme. This research was also supported by the Laboratory of Excellence ARBRE (ANR- 11-LABX-0002-01), the Region Lorraine, the European Regional Development Fund, and the Plant–Microbe Interfaces Scientific Focus Area in the Genomic Science Program, the Office of Biological and Environmental Research in the US DOE Office of Science (to F.M.M.). M.K. acknowledges funding from the SLU Centre for Biological Control (CBC). The Austrian Science Fund FWF project P30460-B32 is acknowledged for funding

L.A. We would like to thank Nathan Vannier for regular discussions and ideas about data analysis and method development. Finally, we thank Paul Schulze-Lefert, Ruben Garrido-Oter, Ryohei Thomas Nakano, Gregor Langen and Rozina Kardakaris for providing helpful comments regarding the manuscript or during departmental seminars and thesis advisory committee meetings.

4 Genomic signatures of host-adaptation in a ubiquitous fungal endophyte

Fantin Mesny[1], Thorsten Thiergart[1], Bruno Hüttel[2], Pedro W. Crous[3,4,5,•], Jose G. Maciá-Vicente[6,•], CABI[7,•], Hannah Rivedal[8,9,•], Ahmad M. Fakhoury[10,•], Soledad Sacristán[11,12,•], Isabelle Batisson[13,•], Stefano Dumontet[14,•], Wade H. Elmer[15,•], Jana Henzelyová[16,•], Joanna S. Kruszewska[17,•], Jessica M. Nelson[18,•], Cara M. Santelli[19,•], Stéphane Hacquard[1,20,*].

1) Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany. 2) Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany. 3) Westerdijk Fungal Biodiversity Institute, P.O. Box 85167, 3508 AD Utrecht, The Netherlands. 4) Microbiology, Department of Biology, Faculty of Science, Utrecht University,Padualaan 8, 3584 CT Utrecht, The Netherlands. 5) Wageningen University and Research Centre (WUR), Laboratory of Phytopathology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. 6) Plant Ecology and Nature Conservation, Wageningen University & Research, PO Box 47, 6700 AA Wageningen, The Netherlands. 7) CABI, Bakeham Lane, Englefield Green, Egham TW20 9TY, United Kingdom. 8) Forage Seed and Cereal Research Unit, U.S. Department of Agriculture Agricultural Research Service, Corvallis, OR 97331, USA. 9) Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA. 10) School of Agricultural Sciences, Southern Illinois University, Carbondale, IL 62901, USA. 11) Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA/CSIC), Universidad Politécnica de Madrid (UPM), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA/CSIC), Madrid, Spain. 12) Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid (UPM), Madrid, Spain. 13) Université Clermont-Auvergne, CNRS, Laboratoire Microorganismes: Génome et Environnement, F-63000 Clermont-Ferrand, France. 14) Department of Sciences and Technologies, University of Napoli 'Parthenope', Centro Direzionale, Isola C4, 80143 Napoli, Italy. 15) Department of Plant Pathology and Ecology, The Connecticut Agricultural Experiment Station, New Haven, Connecticut 06504, USA. 16) Department of Genetics, Institute of Biology and Ecology, Faculty of Science, Pavol Jozef Šafárik University in Košice, Mánesova 23, 04154, Košice, Slovakia. 17) Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland. 18) Maastricht Science Programme, Maastricht University, Maastricht, The Netherlands. 19) Department of Earth and Environmental Sciences, University of Minnesota Twin Cities, Minneapolis, MN, USA. 20) Cluster of Excellence on Plant Sciences (CEPLAS), Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany.

*Corresponding author: S. Hacquard (hacquard@mpipz.mpg.de).

This chapter presents a study initiated and coordinated by myself, with inputs, support and supervision from Stéphane Hacquard. I wrote the manuscript presented in this section, with inputs from Stéphane Hacquard. The 72 *Plectosphaerella* isolates described in this study were obtained from collaborators marked with "•" in the authors list above (see also Supplementary Table), and sequenced at Max Planck Genome Center by Bruno Hüttel. Except for the re-analysis of *A. thaliana* microbiota profiling data presented on Figure 7 (panels a and b) that was performed by Thorsten Thiergart, all the analyses and experiments described in the following manuscript were performed by myself.

4.1 Introduction

Both the above- and belowground tissues of plants are inhabited by a broad diversity of fungi, collectively forming their mycobiota [3, 74, 86, 91]. Recent characterization of the *Arabidopsis thaliana* root mycobiota suggested that Brassicaceae (non-mycorrhizal plants) are essentially colonized by endophytes, spanning along the mutualist-pathogen continuum [85, 86, 91]. As a community, this endophytic mycobiota is highly detrimental for its host, which growth and survival depend on a fully functional immune system and protective bacterial root commensals [12, 87]. A link between fungal negative effects on host performance and root colonization efficiency was identified, and could be associated to a conserved plant cell wall-degrading enzyme (PCWDE) family: the pectate lyase PL1_7 [91]. However, we can expect that besides the use of conserved PCWDEs, phylogenetically distant mycobiota members have evolved different strategies to accommodate in *A. thaliana* tissues.

Genome architecture was demonstrated to play an important role in the adaptation of fungal pathogens to novel environments, hosts and niches [164]. A well-known example concerns accessory chromosomes, first shown to confer virulence towards pea to specific strains of *Nectria haematococca* [165], then identified to be lineage-specific and determining host-specificity of infection in *Fusarium oxysporum* [166]. Besides from these chromosomes, specific genomic compartments identified in fungal pathogens showed important plasticity, with remarkably high mutation and/or recombination rates; sometimes under the pressure of transposable elements [164, 167–171]. Both these genomic regions and accessory chromosomes are enriched in effector-encoding genes playing a key role in fungal fitness, through the evasion of host immune responses, virulence and/or antimicrobial activity against bacterial and fungal competitors [47, 164, 172]. The rapid evolution of effector sequences and repertoires constitutes a major force driving fungal adaptation to hosts and environment, notably keeping pathogens undetectable by constantly co-evolving plant surveillance systems [47, 164, 170, 173].

While the genomes of fungal endophytes remain poorly studied, their compartmentalization and the existence of accessory regions enriched in effector-encoding genes were demonstrated [174]. As pathogens, endophytes are exposed to plant immune receptors in planta, and may depend on rapidly evolving effectors to grow and survive in plant endospheres. Interestingly, the widespread endophytic species *Plectosphaerella cucumerina* (also described as an emergent pathogen) was shown to express numerous effector-encoding genes after inoculation on *A. thaliana* leaves, contrasting with its epiphytic relative *Plectosphaerella plurivora* [175]. Effectors may therefore constitute major determinants of endophytism, supporting fungal colonization of plant tissues with various strategies. In the endophytic species *Epichloë typhinia*, five effectors were identified as putative determinants of host adaptation [176], after demonstrating subspecies host preference with

reciprocal infection experiments [177]. While the importance of effectors for host-specific fungal pathogenicity is well documented [89, 90], it remains to be investigated how widespread host specialization is among endophytes, and what are the mechanisms supporting it. As host preference was previously demonstrated for *A. thaliana* root-associated bacteria [178], we wondered if genomic signatures of host adaptation could be found in the root mycobiota of the model Brassicaceae plant.

Here, we studied the genomics of fungal endophyte *P. cucumerina*, that represents a highly prevalent member of the *A. thaliana* root mycobiota in Europe [23, 91]. First, we characterized its effect on plant growth, and assessed the intra-species conservation of this effect. After sequencing the genomes of 69 *P. cucumerina* strains isolated from different hosts around the world, we profiled their architecture and observed highly plastic genomic compartments enriched in effector-encoding genes. Then, we aimed at deciphering potential genomic signatures of adaptation to *A. thaliana*. Using statistical testing and with support from transcriptomic data, we could identify a single genomic region significantly associated to *A. thaliana*-isolation, that is expressed in planta.

4.2 Results

4.2.1 Ubiquitous soil-borne species *P. cucumerina* is highly prevalent in the *A. thaliana* mycobiota

Given the reported high variability of the *A. thaliana* root mycobiota composition [23, 86], we re-analysed previously published data [23, 87] to identify fungal taxa that robustly colonize the roots of the model plant across Europe. Relative abundance in surface-sterilized roots and prevalence across 18 sampling sites were calculated for each rDNA ITS1 sequence variant, highlighting five core taxa (Fig. 7ab), belonging to three different genera in class Sordariomycetes: *Fusarium* (n=3), *Plectosphaerella* (n=1) and *Ilyonectria* (n=1). The second most prevalent taxa could be confidently assigned to species *Plectosphaerella cucumerina* [179, 180]. We inspected the Global Soil Mycobiome data set [181] to assess the worldwide prevalence of *P. cucumerina* in soils. Curated rDNA ITS1 variants of the species [180] could be detected in 780/3194 soil samples (sample coverage= 24.4%), originating from all continents except Antarctica (Fig. 7c). While *P. cucumerina* is significantly depleted in soils sampled in tropical environments and coniferous forests (Fig. 7d), it is particularly enriched in anthropogenic biomes as well as in temperate grasslands and broadleaf forests (Fisher's exact test, $FDR < 0.05$ - see Methods). Taken together, our results revealed that *P. cucumerina* is a widespread soil-borne fungus that has the ability to robustly colonize the roots of *A. thaliana*, and shows preference for specific biomes.

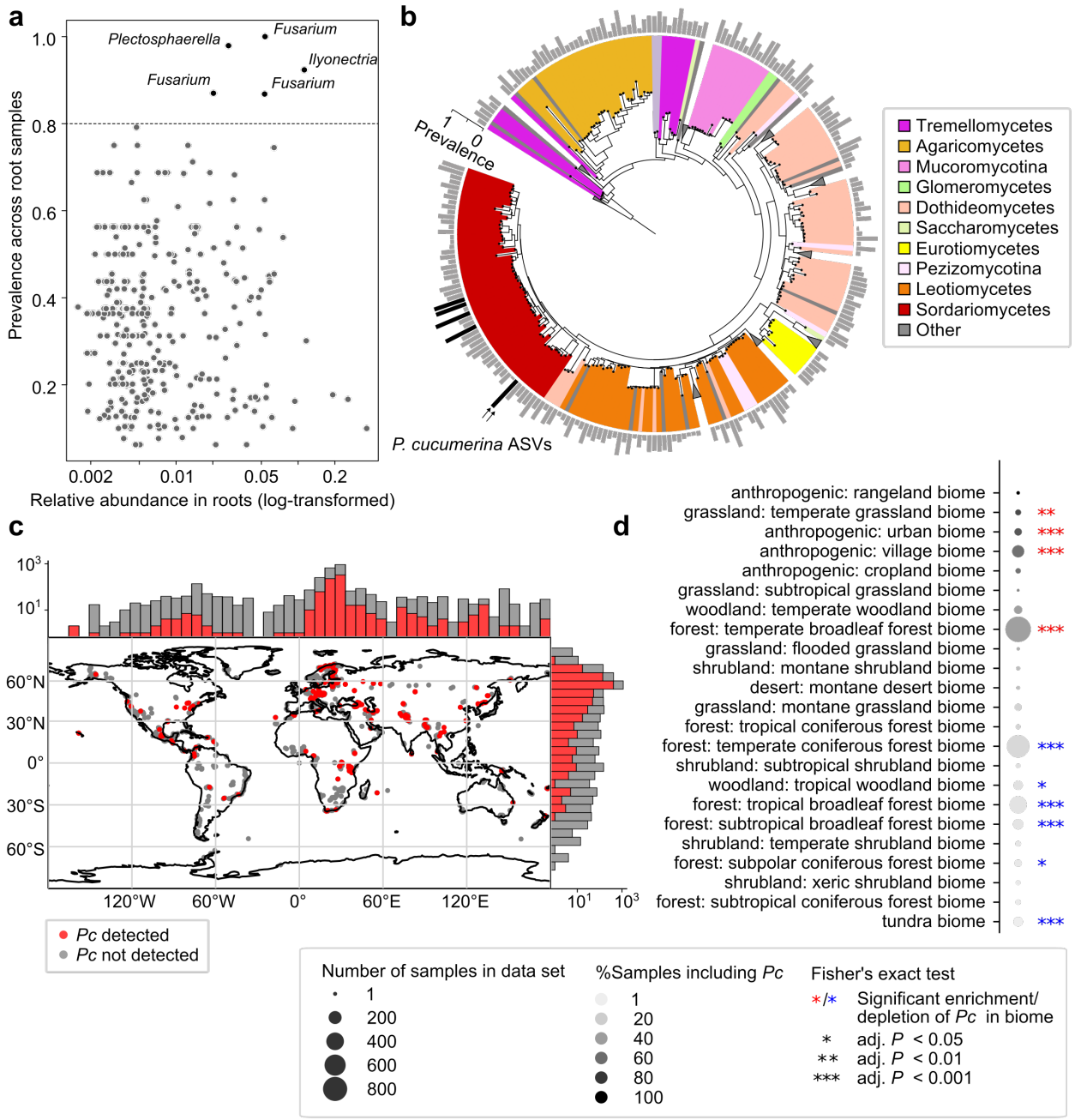


Figure 7: Core *A. thaliana* root mycobiota members and ecology of *Plectosphaerella cucumerina*. **a** Amplicon sequence variants (ASVs) sequenced from European root samples of *A. thaliana* (data: [23, 87]) are plotted according to their log-transformed relative abundance in root samples, relative to their prevalence across all 18 European sites. Relative abundance values are only calculated across samples where ASVs were detected (minimum 0.1% per sample). Only root samples sequenced with more than 1000 reads were considered for analysis. The five most prevalent ASVs are highlighted in black, and labelled with their taxonomic affiliation identified by Warcup [179]. **b** Maximum-likelihood phylogenetic tree reconstructed from the ITS1 sequences of all 338 ASVs detected in roots of *A. thaliana* sampled across Europe (data: [23, 87]). Clades are colored accoring to taxonomic affiliaton. Clades including more than one unclassified ASVs

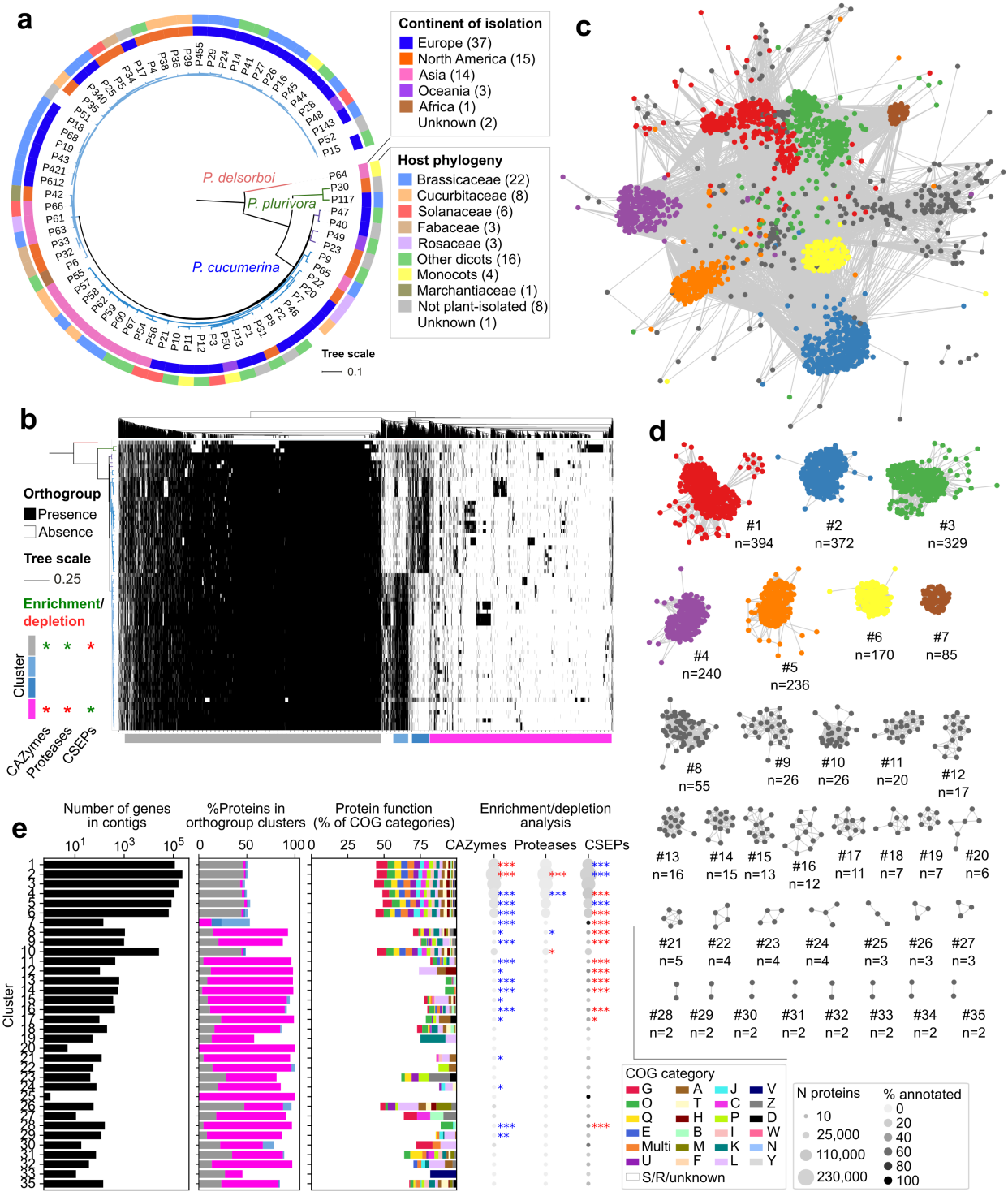
were collapsed. The barplot depicts average site prevalence. The five most prevalent ASVs are highlighted in black, as on panel **a**. **c** Detection of species *Plectospharella cucumerina* (presence of curated ITS1 sequences [180]) in soil samples originating from all continents (n=3194) (Global Soil Mycobiome data set [181]). **d** Soil biome enrichment/depletion in samples containing *Plectosphaerella cucumerina*, as tested by Fisher's exact test (see Methods) using the Global Soil Mycobiome data set.

4.2.2 Detrimental effect of *P. cucumerina* on *A. thaliana* growth

We aimed to better understand the effect of prevalent *A. thaliana* root mycobiota member *P. cucumerina* on host growth. We previously reported the detrimental effect on plant performance (i.e. growth and germination rate) of a *P. cucumerina* strain (P16) isolated from *A. thaliana* roots [91]. We could confirm this result in an agar-based system, by inoculating the roots of 10-day-old sterile seedlings with 500 and 1000 spores of this same isolate. Shoot fresh weight (SFW) values measured 28 days after inoculation revealed a clear detrimental effect of *P. cucumerina* on host growth (ANOVA $P = 4.54e-12$, Tukey Honest Significant Difference $P = 1e-7$, Supplementary Fig. 15a). Although no significant difference between plants inoculated with 500 and 1000 spores was detected (Tukey Honest Significant Difference $P = 0.55$), we observed a lower mean SFW for plants inoculated with 1000 spores. This suggests a potential dose-effect of *P. cucumerina*, in line with the previously reported correlation between fungal load in roots and detrimental effects on *A. thaliana* growth [91]. While plant health was significantly affected by *P. cucumerina* in our agar-based system (Supplementary Fig. 15b), the strain we used was isolated from the roots of healthy-looking plants. We suspect that in nature, bacterial root commensals protect plants from the detrimental effects of *P. cucumerina*, as previously shown at the microbiota scale [87]. A broad variety of *A. thaliana* root-isolated bacteria (including Pseudomonadaceae, Comamonadaceae and Rhizobiaceae) were actually shown to antagonize the growth of *P. cucumerina* strains, in binary competition assays [87]. We then aimed to identify whether the observed detrimental effects are conserved in the *P. cucumerina* species. We assembled a collection of 72 *Plectosphaerella* isolates, including 69 from species *P. cucumerina*. These strains were isolated from diverse hosts and environments, on five different continents (Supplementary Table, Fig. 8a). Eleven strains were isolated from the tissues of healthy-looking *A. thaliana* plants: 8 from roots (harvested at 5 distinct European sites or grown in Cologne Agricultural Soil [87]) and 3 from leaves. We sequenced the genomes of *Plectosphaerella* isolates using PacBio long-read sequencing, resulting in high-quality assemblies (sizes: $35.5-40.5$ Mbp, median=36.4 Mbp; number of contigs: 9–188, median$=27.5$; N50: 0.77–7.66 Mbp, median$=2.69$ Mbp; Supplementary Fig. 16). After gene prediction in these assemblies, we used four different methods to reconstruct the phylogeny of our fungal collection (see Methods) and identified that *P. cucumerina* is essentially divided into two subspecies (Fig. 8a, Supplementary Fig. 17). We inoculated independently the 72 isolates in the roots of sterile *A. thaliana* seedlings. SFW values 28 days after inoculation revealed detrimental effects of most *P.*

cucumerina strains (ANOVA $P = 2e-16$, Dunnett's post-hoc test vs. mock $P < 0.05$ for $44/69$ *P. cucumerina* strains), contrasting with neighbouring species *P. plurivora* and *P. delsorboi* (Supplementary Fig. 18a). While intra-species variation of detrimental effects was identified, it did not exhibit any clear phylogenetic patterns, and seemed independent from isolation source. We specifically tested whether strains retrieved from *A. thaliana* tissues are more pathogenic on this plant species than strains isolated from other hosts (Supplementary Fig. 18b). Our results showed that effects on *A. thaliana* performance are not linked to hosts of isolation. However, we cannot exclude the possibility that strains isolated from *A. thaliana* colonize the tissues of this plant species more efficiently than others. Such observation would suggest host adaptation in *P. cucumerina*, and could explain the remarkable prevalence and abundance of the species in the roots of European *A. thaliana* populations (Fig. 7ab).

Figure 8: Comparative genomics of *P. cucumerina* isolates identifies core and variable genomic compartments. **a** Phylogeny of the 73 *Plectosphaerella* isolates, annotated with their continent of isolation (inner circle of color strips) and the phylogenetic group of their host (outer circle). The phylogenetic tree was reconstructed from the concatenated alignments of 5,466 single copy orthologues (see Methods). In the *P. cucumerina* species, two subspecies can be distinguished and are depicted in two shades of blue. **b** Hierarchical clustering (UPGMA method) of gene families by their presence/absence, presented across the data set phylogeny. For simplification, single copy orthologues present in all strains were excluded from this analysis. Four major clusters were defined (flat clustering of the UPGMA hierarchy with distance threshold= 5) and are depicted in different colors under the clustermap. Results of Fisher's exact tests analyzing the enrichment/depletion in CAZymes, proteases and CSEPs in each of these clusters are summarized on the left (*: FDR<0.05, red: Odd's ratio<1, green: Odd's ratio>1). **c** Similarity network of contigs in our genome collection, built from pairwise genome mapping data. Each node consists of a contig, and one contig is connected with another if at least 10% of its sequence could be mapped on it. Edges are weighted by mapped sequence lengths, and node positions were defined by the edge-weighted spring-embedded layout of Cytoscape [157]. Different colors highlight major clusters of contigs identified by MCL clustering (granularity=5) with clusterMaker2 [158]. All the defined clusters (n=35) and their sizes are depicted on panel **d**. **e** Functional analysis of the 35 contig clusters, showing from left to right: the number of genes annotated in the total contig cluster, clusters of orthogroups these genes belong to (same colors as panel **b**), gene annotation in COG categories, and the results of an enrichment/depletion analysis in CAZymes, proteases and CSEPs with Fisher's exact test (red: Odd's ratio<1, green: Odd's ratio>1; *: $FDR < 0.05$, **: $FDR < 0.01$, ***: $FDR < 0.001$).


a
Continent of isolation
Europe (37)
North America (15)
Asia (14)
Oceania (3)
Africa (1)
Unknown (2)
Host phylogeny
Brassicaceae (22)
Cucurbitaceae (8)
Solanaceae (6)
Fabaceae (3)
Rosaceae (3)
Other dicots (16)
Monocots (4)
Marchantiaceae (1)
Not plant-isolated (8)
Unknown (1)
P. delsorboi
P. plurivora
P. cucumerina
Tree scale
0.1
b
Orthogroup
Presence
Absence
Tree scale
0.25
Enrichment/
depletion
Cluster
CAZymes
Proteases
CSEPs
c
d
#1
n=394
#2
n=372
#3
n=329
#4
n=240
#5
n=236
#6
n=170
#7
n=85
#8
n=55
#9
n=26
#10
n=26
#11
n=20
#12
n=17
#13
n=16
#14
n=15
#15
n=13
#16
n=12
#17
n=11
#18
n=7
#19
n=7
#20
n=6
#21
n=5
#22
n=4
#23
n=4
#24
n=4
#25
n=3
#26
n=3
#27
n=3
#28
n=2
#29
n=2
#30
n=2
#31
n=2
#32
n=2
#33
n=2
#34
n=2
#35
n=2
e
Number of genes
in contigs
%Proteins in
orthogroup clusters
Protein function
(% of COG categories)
Enrichment/depletion
analysis
CAZymes
Proteases
CSEPs
Cluster
COG category
G
O
Q
E
Multi
U
A
T
H
B
M
F
J
C
P
I
K
L
V
Z
D
W
N
Y
S/R/unknown
N proteins
10
25,000
110,000
230,000
% annotated
0
20
40
60
80
100

4.2.3 Genomes of *P. cucumerina* isolates are composed of core and variable regions

We aimed to better understand intra-species genome evolution of prevalent *A. thaliana* root mycobiota member *P. cucumerina*. After performing an orthology prediction in our data set of 72 *Plectosphaerella* genomes defining gene families (OrthoFinder [119], n=15,269 orthogroups), hierarchical clustering was conducted on their presence/absence in each strain, revealing two major clusters and two smaller subspecies-specific ones (Fig. 8b). A major cluster was composed of gene families conserved across the *P. cucumerina* species, and was significantly enriched in genes encoding CAZymes (Fisher's exact test: Odds Ratio= 1.85, $FDR = 3e-13$) and proteases (Odds Ratio= 1.77, $FDR = 1e-4$). In contrast, another major cluster composed of highly-variable gene families occurring generally in a low number of strains was found to be enriched in genes encoding candidate secreted effector proteins (CSEPs; Odds Ratio= 1.51, $FDR = 1e-14$). Consistently with these results, we identified a lower intra-species variation in numbers of genes encoding CAZymes and proteases than of those encoding CSEPs (Supplementary Fig. 19). We hypothesized that the genomic architecture of *P. cucumerina* drives the conservation of specific genes and the contrasting high polymorphy of CSEP repertoires. By performing pairwise genome alignments then building a weighted similarity network of contigs, we could observe different genomic compartments (Fig. 8c), that we defined precisely using Markov clustering (MCL; Fig. 8d). While such clusters of contigs could correspond to chromosomes, in absence of experimental evidence, we will refer to them as *genomic compartments*. Due to its low percentage in G/C, we can notably expect cluster #7 to correspond to the mitochondrial chromosome (Supplementary Fig. 20). While seven of the 35 genome compartments are conserved across the *P. cucumerina* species, the others show a sparse representation in our genome collection (Supplementary Fig. 20). An analysis of genes contained in the different compartments revealed that the majority of core gene families (observed on Fig. 8b) are located on the conserved genomic compartments #1 to #6 (Fig. 8e). On the contrary, most of the highly variable gene families are found in the other compartments. While most of the CAZyme-encoding genes are located on genomic compartments #1 and #2 (significant enrichment - Fisher's exact test: Odds Ratio= 1.21 and 1.10, $FDR = 7e-40$ and $6e-22$ respectively), 10 less-conserved compartments show an enrichment in CSEP-encoding genes (together with conserved compartments #4 and #6). Thus, both the analyses of orthology data and contig similarity highlighted core and an variable genomic regions. Our results are in line with the *two-speed genome* evolution model [170, 171], which describes the localization of effector-encoding genes in highly polymorphic and unstable genomic compartments, permitting the rapid adaptation of fungi to new hosts and environments.

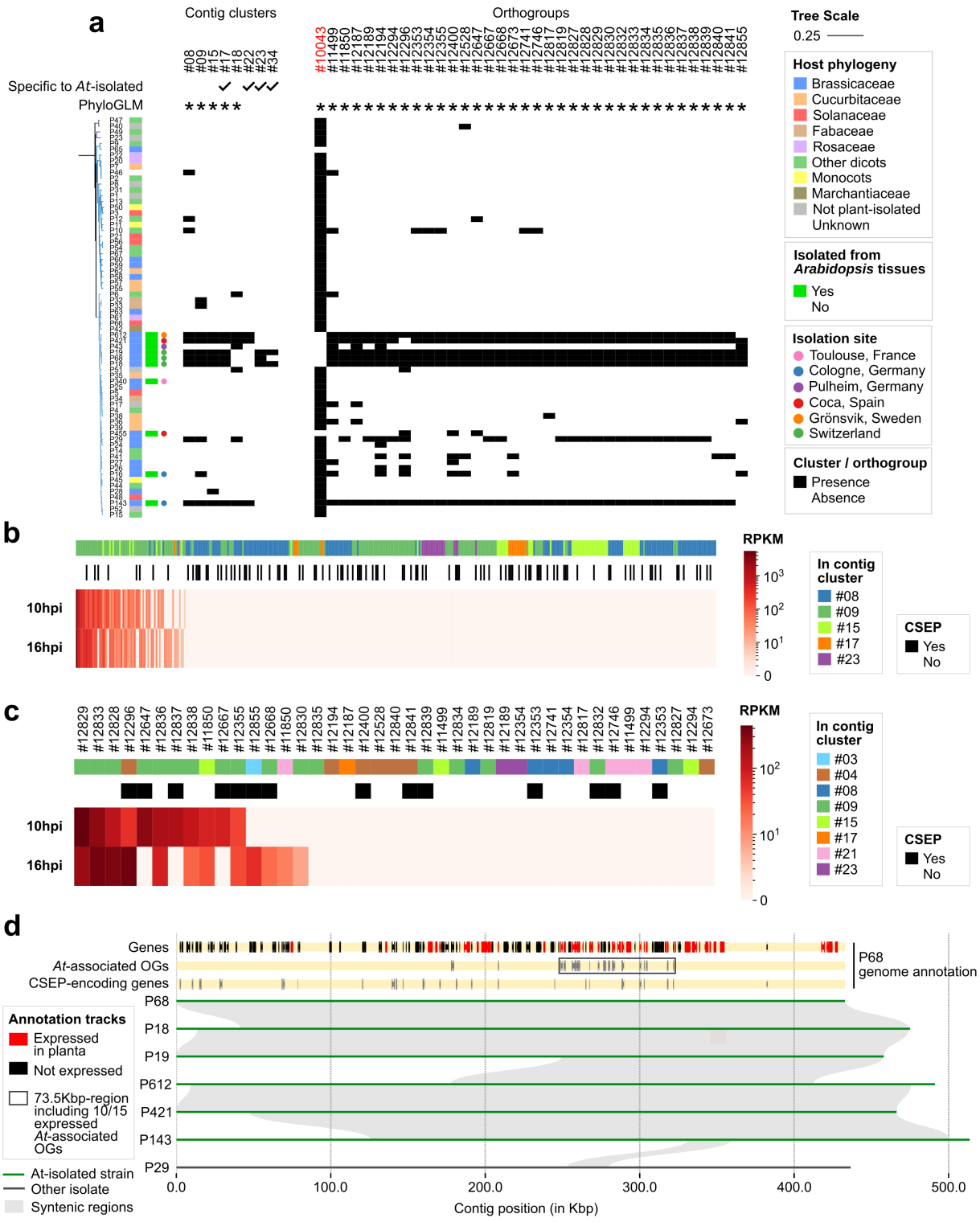


Figure 9: A candidate genomic region of *P. cucumerina* significantly associates to *A. thaliana*-isolation. **a** Genomic compartments (contig clusters, left) and orthogroups (right) which presence/absence in the different *P. cucumerina* strains significantly associates to the isolation from *A. thaliana*, independently

from strain phylogeny (according to PhyloGLM [124] and Benjamini-Hochberg's correction). Contig clusters that are exclusively present in the genomes of *A. thaliana*-isolated strains without any significant association to this host (#22,#23,#34), are also shown. A single orthogroup (highlighted in red) was found to negatively associate to the host. **b** Previously-published transcriptomic data from strain P68 [175] reveal which genes from genomic compartments associated to fungal isolation from *A. thaliana* (see panel **a**) are expressed in planta. RPKM=Reads Per Kilobase of transcript per Million mapped reads. hpi=hours post-inoculation. **c** In planta expression of P68 genes which orthogroups are significantly associated to strain isolation from *A. thaliana* (see panel **a**). Most of these candidate genes for adaptation to *A. thaliana* are located in genomic compartments #9. **d** Genomic conservation within contig cluster #9, focusing on contig 29 from P68 that includes majority of genes expressed in planta and of genes from orthogroups associated to *A. thaliana*-isolation (*At-associated OGs*). A syntenic block of candidate genes for adaptation to *A. thaliana* can be distinguished in 6 strains isolated from this host. Strains P16, P32 and P33 that have contigs in genomic compartment #9 were excluded from the analysis, as they revealed very low (or no) similarity with the contig 29 of P68. Genome annotation at the top of the panel is from strain P68.

4.2.4 Genomic signatures of adaptation to *A. thaliana* in *P. cucumerina*

In our collection, ten *P. cucumerina* strains were isolated from *A. thaliana* plants. We tested whether these isolates comprise specific genomic features, underlying potential host adaptation. By performing statistical testing with PhyloGLM [124], we identified 5 genomic compartments and 36 gene families that are significantly associated to *A. thaliana*-isolation, independently from phylogenetic signal (Fig. 9a). To estimate whether these genomic features include genes important for *A. thaliana* tissue colonization, we re-analyzed a previously published RNA-Seq data set [175], profiling the transcriptome of strain PcBMM (P68) in the leaves of *A. thaliana* at 10 and 16 hours post-inoculation. First, we focused on genes located in genomic compartments that significantly associate with *A. thaliana*-isolation (Fig. 9b), and identified that most genes detected as expressed in planta are located in contig cluster #9 (52/66, 79%). Then, we inspected the gene families significantly associated to *A. thaliana*-isolation (Fig. 9c), and found that they include 15 genes expressed by PcBMM during plant colonization. While similarity-based annotation of these genes (EggNog [182]) was poorly informative (Table), it revealed that one gene encodes a putative transcription factor with sequence similarity to CON7, previously described as a master regulator of virulence-related morphogenesis in rice-blast fungus *Magnaporthe grisea* [183] and in tomato-infecting strains of *F. oxysporum* [184]. Of note, seven of the 15 genes encode CSEPs (Fig. 9c), with no sequence similarity to previously characterized effectors. Consistently with the results of our previous approach, we found that 11 of these 15 genes are located in genomic compartment #9. We further analyzed the genomic position of these genes, and identified a genomic region of about 73.5kbp in compartment #9 comprising 10 of these 15 genes, that includes in total 14 genes from orthogroups significantly associated to *A. thaliana*-isolation (Fig. 9d). This region constitutes a syntenic block conserved in 6 of our 10 *A. thaliana*-isolated strains, and may represent a genomic

feature underlying host adaptation. Molecular biology techniques will be employed to functionally validate this genomic region and assess its role in the interaction of *P. cucumerina* with *A. thaliana*.

Gene ID	Contig cluster	RPKM 10hpi	RPKM 16hpi	CSEP	Functional description (curated from EggNog annotations)
contig_29_gene_66	9	442.24	235.57	No	CON7-like transcription factor (K21455)
contig_29_gene_73	9	273.45	372.55	No	Methyltransferase domain (PF13489)
contig_29_gene_65	9	131.64	337.98	No	-
contig_32_gene_105	4	56.12	395.14	Yes	Tetratricopeptide repeat
contig_29_gene_83	9	274.84	0.00	Yes	-
contig_29_gene_77	9	186.50	82.95	No	-
contig_29_gene_82	9	172.68	0.00	Yes	Ankyrin repeat (PF00023), NACHT domain (PF05729), phosphorylase superfamily (PF01048)
contig_29_gene_84	9	113.50	22.54	No	-
contig_39_gene_15	15	77.77	30.62	No	-
contig_29_gene_88	9	81.94	0.00	Yes	Putative DNA-binding domain in centromere protein B (PF04218), Tc5 transposase DNA-binding domain (PF03221)
contig_29_gene_52	9	38.62	34.29	Yes	-
contig_2_gene_19	3	0.00	58.81	Yes	-
contig_29_gene_79	9	0.00	19.44	Yes	Basic region leucin zipper (PF07716)
contig_13_gene_13	21	0.00	13.20	No	-
contig_29_gene_68	9	0.00	5.88	No	-

Table: In planta expression and functional annotation of PcBMM genes in orthogroups significantly linked to *A. thaliana*-isolation

4.3 Discussion

We report here the assembly of a large culture collection of isolates from soil-borne species *P. cucumerina* and its sequencing. We showed that this species represents a core member of the endophytic *A. thaliana* root mycobiota across Europe, while being highly prevalent in soils worldwide (Fig. 7). Although frequently colonizing healthy *A. thaliana* plants in nature, *P. cucumerina* shows detrimental effects on plant growth in binary interaction, after both root (Supplementary Fig. 15 and 18a, [87, 91]) and shoot inoculation [175]. Also described as an emergent pathogen [175], diverse crops were reported to be affected by *P. cucumerina* infections in the last decade, including notably Brassicaceae [185–187], Cucurbitaceae [188, 189], Solanaceae [190–192] and Fabaceae [193, 194]. While strains isolated from hosts of these families are overrepresented in the culture collection we assembled, we also imported four strains isolated from monocotyledonous plants, and one from a *Marchantia* livewort. With this broad diversity of hosts, *P. cucumerina* constitutes an adequate species to study host adaptation in endophytes. Besides from one clade constituted exclusively of Brassicaceae-isolated strains, the phylogeny of our strain collection did not reveal any clear clustering by host phylogeny (Fig. 8a), contrary to other species like *F. oxysporum* in which lineages have defined host-specificity [164, 195]. After performing recolonization experiments of sterile *A. thaliana* plants with individual *P. cucumerina* strains from our collection, we could not identify any significant difference between detrimental effects of *A. thaliana*-isolated strains and

the ones of other isolates (Supplementary Fig. 18). However, an important variation between biological replicates was observed, possibly inherent to the agar-based gnotobiotic system we used. While differences in the phenotypes of inoculated plants may be observed in a different experimental setup, we will also consider to measure in planta fungal burden, to test if *A. thaliana*-isolated strains are better than others at colonizing their host.

We used the PacBio long-read technology to sequence the genomes of our *P. cucumerina* isolates and obtained overall high assembly quality (Supplementary Fig. 16). With a median contig number per assembly of 27.5, we however did not reach chromosomal resolution. To study genome architecture in absence of chromosome-resolved genomes, we developed an approach consisting in building a similarity network of total contigs in our dataset, then clustering it. Resulting clusters may represent chromosomes, as corroborated by the clustering of highly-similar contigs with low percentages in guanine and cytosine, likely corresponding to mitochondrial chromosomes (cluster #7, Supplementary Fig. 20). Clustering-defined genomic compartments were found to be either well conserved, or sparsely distributed across the phylogeny of our strain collection. These results are to a certain extent consistent with the *two-speed* genome model, previously described in other fungal species (including *Verticillium dahliae*, a close relative of *P. cucumerina*) [171]. This concept can be criticized for providing a simplified and biased vision of genome evolution, and generalizing specific observations [170]. In our specific case, we could clearly identify an enrichment of effector-encoding genes in the most variable genomic compartments of *P. cucumerina* (Fig. 8e). The high inter-connectivity of these compartments on our similarity network (Fig. 8c) suggests that recombination events may occur frequently in these genomic regions, as stated by the model. During a synteny analysis focusing on a region of interest (Fig. 9d), we could identify a clear recombination event that occurred in a poorly conserved genomic compartment (#9). However, contrary to previously reported in other species [168, 169], contigs in variable genomic compartments of *P. cucumerina* do not show any clear enrichment in transposable elements, in comparison to conserved ones (Supplementary Fig. 20). In the future, we aim to better characterize the physical position of variable genomic compartments in the *P. cucumerina* genome. Taking into consideration that our approach encompasses clustering biases, we wish to understand if the variable genomic compartments we defined represent highly plastic regions of conserved chromosomes, or correspond to small supernumerary chromosomes. Additionally, to further investigate genome evolution and signatures of *P. cucumerina* adaptation, we will calculate across the genomic compartments we identified, nucleotide diversity indices (π), allele frequencies (Tajima's *D*) and ratios of non-synonymous to synonymous substitutions (dN/dS). Such approach will allow us to identify genomic regions under positive selection, and previously permitted to distinguish candidate genes for host adaptation in host-specialized subspecies of fungal endophyte *Epichloë typhinia* [176].

Host preference patterns have previously been identified in the bacterial root microbiota of *A. thaliana* [178]. We questioned the existence of such properties in its root mycobiota, knowing that it includes taxa in which adaptation to other hosts was demonstrated (e.g. the *F. oxysporum* species [89, 90]) and considering previous demonstration of host specialization in non-pathogenic endophytes [177]. We used our collection of *Plectosphaerella* genomes and identified candidate regions underlying putative host adaptation. In nature, *P. cucumerina* colonizes a broad diversity of plant hosts, with different plant cell wall compositions. We therefore expected that it relies on a diverse set of CAZymes to degrade these cell walls and colonize hosts. While the species encodes a large number of PCWDEs in comparison to some close relatives [91], we identified low intra-species variability of CAZyme repertoires (Supplementary Fig. 19). We observed that CAZymes are essentially encoded by well-conserved gene families, contrasting with CSEPs which numbers and repertoires vary extensively intra-species (Fig. 8; Supplementary Fig. 19). From the genomic perspective, we could therefore expect effectors rather than CAZymes, to mediate potential host specialization of *P. cucumerina* strains. However, the colonization of different hosts was shown to trigger differential expression of both CAZymes and CSEPs in various fungi [196]. It can therefore not be excluded that *P. cucumerina* strains evolved different host-specialized transcriptomes, in which CAZymes play an important role. Nevertheless, multiple effectors were previously demonstrated to contribute in host adaptation of diverse fungi [89, 90]. In *P. cucumerina*, our analyses pointed to a set of closely located genes (syntenic block) that are significantly associated to *A. thaliana*-isolation and expressed in plant leaves (Fig. 9). This region notably encodes seven CSEPs, and was found in the genomes of six *A. thaliana*-isolated strains from our collection: five closely related ones in the Brassicaceae-specific clade, and one phylogenetically distant isolate. Of note, these six strains were collected from four different European sites, located in Sweden, Germany, Switzerland and Spain (Fig. 9, Supplementary Table). Although recombination events were identified, the contigs carrying this region showed good conservation in length and sequences across multiple strains (Fig. 9d), suggesting they may correspond to a supernumerary chromosome. In the future, we will karyotype multiple *P. cucumerina* strains to test this hypothesis. We also aim to investigate next whether the genes encoded in this 73.5Kbp-region confer a fitness advantage to *P. cucumerina* for endophytic root and/or shoot colonization of *A. thaliana*. Since depleting such a large genomic region by molecular methods is to our knowledge very challenging, we will knock-out individual genes in this region, starting with the transcription factor from the CON7 family and the four effectors expressed in planta (Table, Fig. 9d). Alternatively, we may use a *P. cucumerina* strain that does not carry the region of interest in its genome to produce heterologous expression lines. With this technically less challenging approach, we could test whether fungal colonization of *A. thaliana* tissues is improved upon addition of individual genes (i.e. the eleven expressed genes from the genomic region of interest) in the genome of the chosen strain.

4.4 Methods

4.4.1 Re-analysis of *A. thaliana* mycobiota profiling data

Data from a European transect study profiling the root mycobiota of *A. thaliana* at 17 sites [23] were re-analysed, together with profiling data of *A. thaliana* roots grown in Cologne agricultural soil [87]. Paired sequencing reads were joined and demultiplexed using QIIME and its scripts *join_paired_reads* and *split_libraries_fastq* [197]. Quality filtering step was performed with phred score of 30 or higher. For fungal reads that could not be merged, the respective forward reads were kept for further processing. Filtered and demultiplexed reads were trimmed to an equal length of 220bp using USEARCH with option *-fastx_truncate* [142]. Those reads were dereplicated and forwarded to the UNOISE3 pipeline [198]. Amplicon sequence variants (ASVs) were determined using the USEARCH *–unoise3* command, where reads with sequencing errors are corrected and possible chimeras are removed, leaving a set of correct amplicon sequences. ASVs reads were checked against an ITS sequences database (full-length ITS sequences from the National Center for Biotechnology Information (NCBI)) to remove non-fungal reads. Fungal ASVs were classified using the Warcup database [179]. To receive count-tables, ASV sequences were then mapped back against the whole read dataset at a 97% similarity cut-off (using USEARCH *-usearch_global*).

4.4.2 Analysis of *P. cucumerina* prevalence in soils worldwide

Previously curated ITS1-5.8S-ITS2 rDNA sequences of species *P. cucumerina* [180] were downloaded and ITS1 sequences were extracted from them using ITSx [199]. We then linked these ITS1 sequences to ASVs from the Global Mycobiome Global Mycobiome data set using blastn with a percentage of identity threshold of 100% [181]. These ASVs were detected in 780/3194 soil samples (relative abundance > 0). Metadata of sites at which ASVs were detected were then analysed, focusing on latitude, longitude and biome types. We tested the enrichment of specific biomes in soils containing *P. cucumerina* using a random sampling approach. We randomly sampled 780 soil samples 99,999 times, and each time counted the number of soil samples associated to each biome. We used a Fisher's exact test to compare whether the set of soils in which *P. cucumerina* was detected contains significantly more/less samples associated to each biome than the randomly sampled sets of soils. Statistical testing was performed with Scipy 1.5.3 and the function *stats.fisher_exact(alternative='two-sided')* [200].

4.4.3 Plant recolonization experiments with fungal spores

We performed plant recolonization experiments with individual *Plectosphaerella* strains in an agar-based system. First, to prepare fungal spore stock solutions, fungi were grown on oatmeal medium.

This medium was prepared by crushing 5g of oatmeal in liquid nitrogen with a mortar and pestle until obtaining a powder. This power was boiled for 30min in 200mL of water on a heating plate at 180°C. Then, 4g of Difco Agar (ref. 214530) was added to the mixture prior to autoclaving. Fungi were revived from 30% glycerol stocks and grown on this medium for 10-14 days. Sterile water was poured on the grown mycelium, pipeted then filtered on Miracloth, then centrifugated at 4000xg for 10min to concentrate spores. Spores were counted on a Malassez cell, and their concentration was adjusted to 10^7 spores/ml in 25% glycerol. These stock solutions were flash-frozen in liquid nitrogen, then kept at -80°C.

A. thaliana Col-0 seeds were sterilized 15 min in 70% ethanol, then 5 min in 8% sodium hypochlorite. They were then washed six times in sterile double-distilled water and one time in 10 mM MgCl2. They were kept 5 days at 4°C in the dark for stratification. Culture medium was prepared by mixing 2.2g Murashige Skooge (Duchefa Biochemie, ref. M0222) and 0.5g MES buffer (Roth, ref. 4256.5) powders in 1L of double-distilled water. Then, pH was adjusted to 5.7, and 10g of Difco Agar (ref. 214530, 1% final concentration) was added. After autoclaving, 120x120mm petri dishes were filled each with 65mL of medium. The top 2cm of agar were cut out, and 7 seeds were deposited where the cut was made. Plates were sealed with micropore tape and their bottoms were wrapped (up to the seeds level) in paper to hide roots from the light. Plants were grown for 10 days at 21°C (10h with light intensity 4 at 19°C, and 14 h in the dark) in growth chambers.

After 10 days, culture plates were re-opened and only the four biggest plants in each plate were kept. To prepare a spore inoculum, 5 μl of defrosted spore stock solution was pipeted and mixed with 995 μl of sterile water. This mixture was centrifugated at $9,000xg$ for 10min. The top $950\mu l$ in the tube were replaced with fresh sterile water. With this procedure, glycerol was removed and the concentration of spores in the tube was set to 500 spores$/10\mu l$. Ten-days old seedlings were inoculated by pipeting down 10 μl of inoculum (500 spores) on their primary roots. Plates were then re-sealed and put back in growth chambers, with settings mentioned above. After 28 more days in culture, plants were phenotyped by measuring their shoot fresh weight.

4.4.4 Assembly of a culture collection of *Plectosphaerella* isolates

We previously isolated 7 strains of *P. cucumerina* and 1 strain of *P. plurivora* from surface-sterilized *A. thaliana* roots sampled across Europe [87]. We imported from 17 collaborators or culture collections 64 additional strains, sampled from different hosts (or the environment) on different continents. See Supplementary Table for details.

4.4.5 Genome sequencing

Two strains in our collection (P16 and P117) were previously sequenced with the PacBio technology and assembled [91]. Their genomes made available on database GenBank: bioprojects PRJNA371204 and PRJNA570880. While we used previously published genome assembly JAGPXD000000000 for P16, we used an improved version of JAGSXJ000000000 for P117, which was downloaded from the Mycocosm portal with the identifier *Plecucu2* [145]. For P30, we used a genome assembled from Illumina short reads made available on the Mycocosm portal with identifier *Plecu1*.

Other genomes used in this study were newly sequenced with PacBio sequencing. Individual strains were revived on Potato Growth Agar medium (Roth, ref. CP74.1), and cultured for 7 days.

Six strains (P43, P143, P340, P421, P455, P612) were sequenced with PacBio CLR technology. Their gDNA was extracted according to a previously introduced cetyltrimethylammonium bromide protocol [108]. RNA of these strains was sequenced with Illumina short reads to support gene prediction in the assemblies.

All the other strains (n=63) were sequenced with PacBio CCS technology (HiFi reads). Their gDNA was extracted according to a previously introduced cetyltrimethylammonium bromide protocol [108], without the described purification step of the DNA extracts. Native fungal DNA was found in many cases to be recalcitrant to sequencing. Therefore, PacBio barcoded ultra-low libraries were prepared according to the protocol "Procedure Checklist - Preparing HiFi SMRTbell® Libraries from Ultra-Low DNA Input" starting with DNA fragmentation using g-Tubes (Covaris). Typically, five libraries were pooled and then sequenced on a Sequel IIe SMRT cell with 8 mio. ZMWs with Binding kit 2.0 or Binding kit 2.2 and the Sequel II Sequencing Kit 2.0 for 30 h. If necessary gDNA or final libraries were additionally size-selected on a Blue Pippin (Sage Sciences) device to remove smaller DNA fragments. After sequencing, the HiFi data was demultiplexed with SMRTlink10.0, and adapter sequences were trimmed away with cutadapt v3.5 [201].

4.4.6 Genome assembly and gene prediction

Genomes were assembled and polished from PacBio reads using Flye v2.9-b1768 with parameter *–genome-size* 40m, decided based on the size of previously sequenced *P. cucumerina* genomes [91, 175]. Parameter *–pacbio-hifi* was used for genomes sequenced with PacBio CCS and *–pacbio-raw* was used for genomes sequenced with PacBio CLR.

Gene prediction was carried out in all the genome assemblies except for P16 and P117, for which we used the gene predictions from the JGI annotation pipeline [145]. We used FGENESH v8.0.0 with similarity matrix *Torrubiella hemipterigena* and parameters -skip_bad_prom -skip_bad_term.

4.4.7 Gene functional annotations

Gene functions were predicted with emapper v2.1.5 [182] based on eggNOG orthology data [202], using DIAMOND as a similarity search algorithm [203]. CAZymes were annotated using dbCAN v3.0.6 with default parameters [204]. Proteins annotated as CAZymes by any of the three databases implemented in dbCAN were considered as such. Effectors (CSEPs) were predicted using EffectorP v3.0 in fungal mode (-f) [205]. We annotated as proteases all the proteins linked by emapper to a PFAM identifier associated to a MEROPS family [152]. Transposons were annotated in genome assemblies with reasonaTE v1.0 with parameters *-mode annotate -tool all* [206].

Principal components analyses (PCA) revealing similarities of CAZyme, protease and CSEP catalogs were performed by counting the number of annotated genes per family (CAZy and MEROPS families for CAZymes and proteases, orthogroups for CSEPs), then calculating pairwise Jaccard distances between strains. Jaccard dissimilarity matrices were calculated with R package Vegan v2.5-7 (`https://github.com/jarioksa/vegan`), using function *vegdist(method="jaccard")*. PCA plots were then reconstructed from these matrices using function *prcomp* in R.

4.4.8 Orthology prediction

Orthology prediction was performed on the total set of proteins predicted in our data set using OrthoFinder v2.5.4 [119] with parameter *-S blast*. When analysing the results of this orthology prediction (Fig. 8b), we performed hierarchical clustering of the gene families that are present in at least two genomes by their presence/absence. To do so, we used the function *hierarchy.linkage(method='average')* implemented in Scipy [200], which consists in an unweighted pair group method with arithmetic mean (UPGMA). We then performed flat clustering of the UPGMA hierarchy using function *hierarchy.fcluster(t=5,criterion='distance')*, defining four major group of orthogroups.

4.4.9 Phylogeny reconstruction

We used four different methods to reconstruct the phylogeny of our *Plectosphaerella* collection, based on genome sequences. First two trees (Supplementary Fig. 17ab) were built from 5,466 single copy orthologues, identified in our data set by OrthoFinder [119]. Sequences of these 5,466 gene families were aligned independently using MAFFT v7.407 with parameter *–auto* [207], then trimmed with Trimal v1.4.rev22 using default settings [208]. These trimmed alignments were used as input for IQ-TREE v2.1.1 [209]. The algorithm ModelFinder [210] implemented in IQ-TREE identified maximum-likelihood model *JTT+F+I+G4* to be the most adapted one considering the input data. We therefore reconstructed phylogenies with IQ-TREE and RAxML-NG v1.1.0 [211] using this model. While prior alignment concatenation using AMAS [212] was necessary

in the case of RAxML-NG, the IQ-TREE programme carries out concatenation before phylogeny reconstruction.

Additionally, we reconstructed phylogenies with two different coalescent approaches (Supplementary Fig. 17cd). First, we used method STAG [213], as implemented in OrthoFinder [119]. STAG built individual gene family trees (n=13,694), then reconstructed a coalescent species tree from them. Finally, we used Astral v5.7.1 [214] with default parameters, to reconstruct a coalescent species tree from all the single copy orthologue gene trees generated by OrthoFinder (n=5,466).

The phylogenetic tree used as a reference in this study and presented on multiple figures is the one generated by IQ-TREE.

4.4.10 Contig similarity network and genomic compartment analysis

To profile the different genomic compartments represented in our genomic data set, we built a contig similarity network (Fig. 8c). First, we computed all pairwise alignments of the 69 *P. cucumerina* genome assemblies, using minimap2 v2.24-r1122 with parameters *-ax asm5 –eqx* [215]. We then parsed the resulting alignment files and measured the length of mapping between each pair of contigs. We then used Cytoscape v3.7.2 [157] to build a network in which each contig is a node. A query contig was linked with an edge to a reference one, if at least 10% of its length was mapped. The *Edge-weighted spring embedded layout* implemented in Cytoscape was used to group contigs by length of mapping.

To define precise genomic compartments (Fig. 8d), we performed network clustering with ClusterMaker2 [158]. We used Markov clustering (MCL) with a granularity parameter (inflation value) of 5, using the length of mapping as an *array source*.

To analyse the gene functions represented in each cluster, we considered the annotations of genes located in the clustered contigs (see Fig. 8e). We focused on COG categories annotated by emapper [182], and performed enrichment/depletion analyses in CAZymes, Proteases and CSEPs. To do so, we used a random sampling approach. We randomly sampled in our data set as many genes as included in a cluster 99,999 times, then compared proportions of annotated CAZymes, proteases and CSEPs in the random sets of genes with the actual gene set contained in a cluster using a Fisher's exact test. Statistical testing was performed with the function *stats.fisher_exact(alternative='two-sided')* of Scipy 1.5.3 [200].

4.4.11 Association of genomic features to strain hosts

To identify whether presence/absence of specific genomic compartments and gene families could be significantly linked to strain isolation from *A. thaliana*, we used PhyloGLM [124] then Benjamini-Hochberg p-value correction (with R function *p.adjust(method="fdr")*). PhyloGLM models were

built between the two binary variables with formula:

$$featurePresenceAbsence \sim isolationFromArabidopsis.$$

Method *logistic_MPLE* was used, according to default parameters and our IQ-TREE phylogeny.

4.4.12 Transcriptomic analysis

Previously published RNA-Seq data of P68 (PcBMM) in *A. thaliana* Col-0 plant leaves [175] were downloaded from the National Center for Biotechnology Information Sequence Read Archive (SRA) (bioProject PRJNA614936). Reads from the two biological replicates of conditions 10 and 16 hours post-inoculation were used for analysis. We built an index comprising both *A. thaliana* TAIR10 genome and our newly sequenced assembly for P68, then mapped these sets of reads independently on the index using Hisat2 v2.2.0 [126]. Then we used featureCounts v2.0.0 [162] to link mapped reads to genes, according to genome annotations. Finally, we calculated Reads Per Kilobase of transcript, per Million mapped reads (RPKM) values which we used for graphical represenation and to tell which genes were expressed in planta.

4.4.13 Conservation and synteny analysis of a genomic region linked *A. thaliana*-isolation

To analyse the conservation of our region of interest, located in *contig_29* of the genome of P68, we mapped all the contigs clustered in genomic compartment #9 on this specific contig using minimap2 v2.24-r1122 with parameters *-ax asm5 –eqx* [215]. We then used SyRI v1.5.4 to characterize the synteny between pairs of contigs that produced significant alignments [216], and used plotsr [217] to represent the results graphically.

4.5 Data availability

Genomic data generated for this study will be deposited on public databases upon publication of this chapter in a scientific journal.

4.6 Code availability

All the scripts used for data processing and analysis were written in Python v3.8.10 and R v3.5.1. Scripts are available at GitHub (`https://github.com/fantin-mesny/Genomic-signatures-of-host-adaptation-in-a-ubiquitous-fungal-endophyte`).

4.7 Acknowledgements

We wish to thank all the scientific collaborators, culture collections and companies that provided *Plectosphaerella* strains for our study, especially Bevan Weir (ICMP culture collection, Manaaki Whenua), Helen Stewart (CABI culture collection) and Shoshi Kikuchi (NARO Genebank). The sequencing project was funded by a European Research Council starting grant (MICRORULES 758003) allocated to Stéphane Hacquard, and by the Plant-Microbe Interaction department of Max Planck Institute for Plant Breeding Research. We wish to thank Paul Schulze-Lefert for the financial support. Fantin Mesny's salary was covered by the DECRyPT 2125 programme. Thanks should also go to Bart Thomma, Ruben Garrido-Oter and Gregor Langen, for their suggestions and comments during thesis advisory committee meetings.

5 General discussion

During my PhD, I aimed to better characterize the fungi that colonize the roots of the non-mycorrhizal model plant *A. thaliana* in nature. I employed a variety of approaches - including plant recolonization experiments in gnotobiotic system, in planta transcriptomics and genomic data mining - to decipher the function, evolutionary properties and genomic signatures of *A. thaliana* mycobiota members.

Both comparative genomics and confocal microscopy provided evidence that these fungi are endophytes, colonizing the root epidermis of *A. thaliana*. While the definition of endophytism is still disputed [70, 72], the first chapter of my PhD thesis is among the first studies to characterize such a broad set of fungal endophytes, offering functional and genomic insights into endophytism. Although isolated from healthy-looking plants in nature, *A. thaliana* root mycobiota members showed diverse effects on host performance when cultivated in mono-association with their host, in line with the endophytic continuum concept [71, 76]. Since the most robust colonizers in nature have a detrimental effect on their host in mono-association, their qualification as endophytes is conflicting with another accepted definition of endophytism, stating commensalism (or mutualism) as a ground rule [72]. However, our study provides evidence that changes in nutrient concentration can modulate fungal detrimental effects on host. More importantly, such detrimental effects did not affect plants in nature, prior to fungal isolation, and we know that the presence of specific bacterial root commensals can inhibit the detrimental effects of a fungus [87]. The *A. thaliana* immune system was also shown to play a major role in preventing such fungal dysbiosis [12]. Taken together, these results suggest that slight changes in environmental conditions (nutrient concentration), local biodiversity (presence of competing microbes) and host status (capacity to activate immune responses) influence the effect of an endophyte on its host. Considering this condition-dependent pathogenicity of fungi isolated from *A. thaliana* root endosphere, it seems inappropriate to draw a clear distinction between fungal endophytes and pathogens. I thus stand for a definition of endophytism that is descriptive of a fungal habitat (the plant endospheres) and independent from fungal effects on plants.

Using phylogenetic reconstruction and machine learning predictions, we identified that *A. thaliana* mycobiota members took different evolutionary trajectories to endophytism, originating essentially from pathogenic ancestors and in some cases, from saprotrophic ones. These predictions are in line with previously identified transitions from pathogen to commensal/beneficial endophyte [79, 83], revealing that such transitions might occur frequently in nature. Introduced in 2009, the *waiting room* hypothesis suggests that root endophytes are *in the wait* for co-evolution with their host and development into a mycorrhizal symbiosis [218, 219]. In some cases, endophytism may therefore constitute an intermediate step in the evolution of a pathogen into a mutualist symbiont. We in-

vestigated whether endophytes carry signatures of host adaptation in their genomes, as previously identified in pathogenic and mycorrhizal fungi. We could identify in the widespread endophyte *P. cucumerina*, one candidate genomic region for host adaptation, significantly associated to fungal isolation from *A. thaliana*. Additionally, we noticed that the architecture of the *P. cucumerina* genome might favor fungal adaptation to hosts and environment. With genomic compartments exhibiting high plasticity and enrichment in effector-encoding genes, *P. cucumerina* likely has the capacity to evolve quickly its effector catalog, and remain undetected by plant surveillance systems, as pathogens do [47, 164]. Since the universality of this genome architecture among fungal pathogens is questioned [170], its representation among phylogenetically distant endophytes also needs to be verified. Because of the broad phylogenetic diversity spanned by endophytes, we can expect distinct evolutionary properties to shape fungal fitness in plant endospheres.

Contrary to the emergence of some mycorrhizae [69, 108], transitions to endophytic lifestyles did not involve losses of plant cell wall-degrading enzymes [76]. When studying a broad set of *A. thaliana* root mycobiota isolates, we actually identified genes encoding plant cell wall-degrading enzymes to be enriched in endophytes and we predicted them to constitute genomic determinants of endophytism. We also found them to be commonly over-expressed in planta by diverse mycobiota members. In *P. cucumerina*, we could identify conservation of these genes, and localization in slow-evolving core regions of the genome, contrasting notably with effector-encoding genes. Taken together, these results highlight the importance of plant cell wall-degrading enzymes for endophytes. Although their major role in endophytism was presumed [76], we could demonstrate the involvement of a specific enzyme - the pectate lyase PL1_7 - in *A. thaliana* root colonization. Additionally, we showed that PL1_7-mediated fungal colonization affects plant performance, revealing a potential mechanism behind the endophytic continuum: plant cell wall-degradading enzymes are necessary for fungi to colonize their host (and acquire the carbon they need to grow) but their action can be detrimental for the plant if uncontrolled. PCWDEs have a broad diversity of targets, and we can imagine that distinct subsets of enzymes are necessary to colonize plants with different cell wall compositions. Therefore, pectate lyase PL1_7 may be especially important for fungi to colonize pectin-rich dicotyledonous plants like *A. thaliana*. While the PL1_7 family is well represented across the fungal kingdom, we can expect phylogenetically distant endophytes to have evolved additional specific mechanisms to efficiently colonize a host. Genomic analyses of *P. cucumerina* revealing high plasticity of effector sequences and catalogs support this idea. However, individual effectors remain to be functionally validated in the context of endophytism.

Taken together, our results offer a better understanding of the fine line between endophytism and parasitism in the root mycobiota. They show that fungi robustly colonizing *A. thaliana* roots in nature are very well equipped to degrade its cell walls and to overcome innate immune responses. The growth of these endophytic colonizers is tightly controlled in planta by host-encoded

and bacteria-mediated processes, which likely minimize their negative effects and promote their beneficial activities. This work therefore predicts a key role of CAZymes and effector proteins as determinants driving mycobiota assembly and host specificity.

Acknowledgements

First of all, I would like to thank Dr. Stéphane Hacquard, for giving me the chance to carry out my PhD research in his group. I am very grateful for his thoughtful supervision, and the trust I received when exploring my own ideas. Many thanks to Prof. Dr. Francis M. Martin and Dr. Annegret Kohler who co-supervised the first research project presented in this thesis, and made it possible.

My PhD would not have been possible without the *Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota* (SPP DECRyPT 2125) from the German Research Foundation (DFG), that provided the funds for my salary and experimental work. I wish to thank all the scientists involved in this programme - especially its coordinator Prof. Dr. Alga Zuccaro - for the interesting scientific discussions that took place during symposiums and workshops.

I would also like to thank my Thesis Advisory Committee members, Dr. Ruben Garrido-Oter and Dr. Gregor Langen. Their tips and suggestions were always very helpful and truly appreciated.

Thanks should also go to all the scientists of the Plant-Microbe Interaction department at Max Planck Institute for Plant Breeding Research, especially to its director Prof. Dr. Paul Schulze-Lefert. Scientific discussions that took place during departmental seminars and root group meetings were truly helpful and inspiring. Special thanks also go to the (actual and ex-) members of the Hacquard group, especially to Dr. Nathan Vannier who helped me getting started with experimental work, and with whom I had great discussions regarding data analysis methods.

I would like to express my deepest gratitude to Dr. Stephan Wagner, PhD coordinator at the Max Planck Institute, for his invaluable help with administrative procedures and always offering precious advice when needed.

Finally, I am extremely grateful to Prof. Dr. Bart Thomma and Prof. Dr. Eva Stukenbrock who accepted to review and evaluate this thesis, and to anyone who will read it.

References

1. Gross, M. How plants grow their microbiome. *Current Biology* **32,** R97–R100. doi:10.1016/J.CUB.2022.01.044 (2022).

2. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology* **18,** 607–621. doi:10.1038/s41579-020-0412-1 (2020).

3. Getzke, F., Thiergart, T. & Hacquard, S. Contribution of bacterial-fungal balance to plant and animal health. *Current Opinion in Microbiology* **49,** 66–72. doi:10.1016/J.MIB.2019.10.009 (2019).

4. Vogel, J. Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* **11,** 301–307. doi:10.1016/j.pbi.2008.03.002 (2008).

5. Bacic, A. Breaking an impasse in pectin biosynthesis. *PNAS* **103,** 5639–5640. doi:10.1073/pnas.0601297103 (2006).

6. Studer, M. H. *et al.* Lignin content in natural *Populus* variants affects sugar release. *PNAS* **108,** 6300–6305. doi:10.1073/pnas.1009252108 (2011).

7. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The plant microbiota: systems-level insights and perspectives. *Annual Review of Genetics* **50,** 211–234. doi:10.1146/ANNUREV-GENET-120215-034952 (2016).

8. Roman-Reyna, V. *et al.* The rice leaf microbiome has a conserved community structure controlled by complex host-microbe interactions. *bioRxiv,* 615278. doi:10.1101/615278 (2019).

9. Teixeira, P. J. P., Colaianni, N. R., Fitzpatrick, C. R. & Dangl, J. L. Beyond pathogens: microbiota interactions with the plant immune system. *Current Opinion in Microbiology* **49,** 7–17. doi:10.1016/J.MIB.2019.08.003 (2019).

10. Hacquard, S., Spaepen, S., Garrido-Oter, R. & Schulze-Lefert, P. Interplay between innate immunity and the plant microbiota. *Annual Review of Phytopathology* **55,** 565–589. doi:10.1146/ANNUREV-PHYTO-080516-035623 (2017).

11. Xin, X. F. *et al.* Bacteria establish an aqueous living space in plants crucial for virulence. *Nature* **539,** 524–529. doi:10.1038/nature20166 (2016).

12. Wolinska, K. W. *et al.* Tryptophan metabolism and bacterial commensals prevent fungal dysbiosis in *Arabidopsis* roots. *PNAS* **118,** 21115–21118. doi:10.1073/pnas.2111521118 (2021).

13. Lebeis, S. L. *et al.* Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349,** 860–864. doi:10.1126/science.aaa8764 (2015).

14. Tzipilevich, E., Russ, D., Dangl, J. L. & Benfey, P. N. Plant immune system activation is necessary for efficient root colonization by auxin-secreting beneficial bacteria. *Cell Host & Microbe* **29,** 1507–1520.e4. doi:10.1016/J.CHOM.2021.09.005 (2021).

15. Plett, J. M. *et al.* Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *PNAS* **111,** 8299–8304. doi:10.1073/pnas.1322671111 (2014).

16. Berger, F. & Gutjahr, C. Factors affecting plant responsiveness to arbuscular mycorrhiza. *Current Opinion in Plant Biology* **59,** 101994. doi:10.1016/J.PBI.2020.101994 (2021).

17. Teixeira, P. J. *et al.* Specific modulation of the root immune system by a community of commensal bacteria. *PNAS* **118,** e2100678118. doi:10.1073/PNAS.2100678118/SUPPL_FILE/PNAS.2100678118.SD07.XLSX (2021).

18. Deng, S. *et al.* Genome-wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *The ISME Journal* **15,** 3181–3194. doi:10.1038/s41396-021-00993-z (2021).

19. Escudero-Martinez, C. *et al.* Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nature Communications* **13,** 1–14. doi:10.1038/s41467-022-31022-y (2022).

20. Neina, D. The role of soil pH in plant nutrition and soil remediation. *Applied and Environmental Soil Science.* doi:10.1155/2019/5794869 (2019).

21. Finkel, O. M. *et al.* The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. *PLOS Biology* **17,** e3000534. doi:10.1371/JOURNAL.PBIO.3000534 (2019).

22. Harbort, C. J. *et al.* Root-secreted coumarins and the microbiota interact to improve iron nutrition in *Arabidopsis*. *Cell Host & Microbe* **28,** 825–837.e6. doi:10.1016/J.CHOM.2020.09.006 (2020).

23. Thiergart, T. *et al.* Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nature Ecology and Evolution* **4,** 122–131. doi:10.1038/s41559-019-1063-3 (2020).

24. Carrell, A. A. *et al.* Habitat-adapted microbial communities mediate *Sphagnum* peatmoss resilience to warming. *New Phytologist* **234,** 2111–2125. doi:10.1111/NPH.18072 (2022).

25. King, W. L. *et al.* Soil salinization accelerates microbiome stabilization in iterative selections for plant performance. *New Phytologist* **234,** 2101–2110. doi:10.1111/NPH.17774 (2022).

26. Naylor, D., Degraaf, S., Purdom, E. & Coleman-Derr, D. Drought and host selection influence bacterial community dynamics in the grass root microbiome. *The ISME Journal* **11,** 2691–2704. doi:10.1038/ismej.2017.118 (2017).

27. Xie, J. *et al.* Drought stress triggers shifts in the root microbial community and alters functional categories in the microbial gene pool. *Frontiers in Microbiology* **12,** 3066. doi:10.3389/FMICB.2021.744897/BIBTEX (2021).

28. Meyer, K. M. *et al.* Plant neighborhood shapes diversity and reduces interspecific variation of the phyllosphere microbiome. *The ISME Journal* **16,** 1376–1387. doi:10.1038/s41396-021-01184-6 (2022).

29. Vannier, N., Bittebiere, A. K., Mony, C. & Vandenkoornhuyse, P. Root endophytic fungi impact host plant biomass and respond to plant composition at varying spatio-temporal scales. *Fungal Ecology* **44,** 100907. doi:10.1016/j.funeco.2019.100907 (2020).

30. Prado, A., Marolleau, B., Vaissière, B. E., Barret, M. & Torres-Cortes, G. Insect pollination: an ecological process involved in the assembly of the seed microbiota. *Scientific Reports* **10,** 1–11. doi:10.1038/s41598-020-60591-5 (2020).

31. Hassani, M. A., Durán, P. & Hacquard, S. Microbial interactions within the plant holobiont. *Microbiome* **6,** 1–17. doi:10.1186/S40168-018-0445-0 (2018).

32. Agler, M. T. *et al.* Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLOS Biology* **14,** e1002352. doi:10.1371/JOURNAL.PBIO.1002352 (2016).

33. Eitzen, K., Sengupta, P., Kroll, S., Kemen, E. & Doehlemann, G. A fungal member of the *Arabidopsis thaliana* phyllosphere antagonizes *Albugo laibachii* via a GH25 lysozyme. *eLife* **10,** 1. doi:10.7554/ELIFE.65306 (2021).

34. Snelders, N. C. *et al.* Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nature Plants* **6,** 1365–1374. doi:10.1038/s41477-020-00799-5 (2020).

35. Pontrelli, S. *et al.* Metabolic cross-feeding structures the assembly of polysaccharide degrading communities. *Science Advances* **8,** 3076. doi:10.1126/sciadv.abk3076 (2022).

36. Nagy, L. G. *et al.* in *The Fungal Kingdom* 35–56 (ASM Press, Washington, DC, USA, 2017). doi:10.1128/9781555819583.ch2.

37. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356,** eaad4501. doi:10.1126/science.aad4501 (2017).

38. Gan, T. *et al.* Cryptic terrestrial fungus-like fossils of the early Ediacaran period. *Nature Communications 2021 12:1* **12,** 1–12. doi:10.1038/s41467-021-20975-1 (2021).

39. Peay, K. G., Kennedy, P. G. & Talbot, J. M. Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology 2016 14:7* **14,** 434–447. doi:10.1038/nrmicro.2016.59 (2016).

40. Lindahl, B. D. & Tunlid, A. Ectomycorrhizal fungi – potential organic matter decomposers, yet not saprotrophs. *New Phytologist* **205,** 1443–1447. doi:10.1111/NPH.13201 (2015).

41. Crowther, T. W., Boddy, L. & Hefin Jones, T. Functional and ecological consequences of saprotrophic fungus–grazer interactions. *The ISME Journal* **6,** 1992. doi:10.1038/ISMEJ.2012.53 (2012).

42. Hage, H. & Rosso, M. N. Evolution of fungal carbohydrate-active enzyme portfolios and adaptation to plant cell-wall polymers. *Journal of Fungi* **7,** 1–16. doi:10.3390/JOF7030185 (2021).

43. Doehlemann, G., Ökmen, B., Zhu, W. & Sharon, A. Plant pathogenic fungi. *Microbiology Spectrum* **5,** 5.1.14. doi:10.1128/MICROBIOLSPEC.FUNK-0023-2016 (2017).

44. Van Vu, B., Itoh, K., Nguyen, Q. B., Tosa, Y. & Nakayashiki, H. Cellulases belonging to glycoside hydrolase families 6 and 7 contribute to the virulence of *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions* **25,** 1135–1141. doi:10.1094/MPMI-02-12-0043-R (2012).

45. Redkar, A. *et al.* Conserved secreted effectors contribute to endophytic growth and multihost plant compatibility in a vascular wilt fungus. *The Plant Cell* **34,** 3214–3232. doi:10.1093/PLCELL/KOAC174 (2022).

46. Zhang, L. *et al.* Fungal endopolygalacturonases are recognized as microbe-associated molecular patterns by the *Arabidopsis* receptor-like protein RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiology* **164,** 352–364. doi:10.1104/PP.113.230698 (2014).

47. Lo Presti, L. *et al.* Fungal effectors and plant susceptibility. *Annual Reviews in Plant Pathology* **66,** 513–545. doi:10.1146/ANNUREV-ARPLANT-043014-114623 (2015).

48. Tanaka, S. *et al.* A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife* **3.** doi:10.7554/ELIFE.01355.001 (2014).

49. De Jonge, R. *et al.* Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science* **329,** 953–955. doi:10.1126/SCIENCE.1190859 (2010).

50. Shen, Q., Liu, Y. & Naqvi, N. I. Fungal effectors at the crossroads of phytohormone signaling. *Current Opinion in Microbiology* **46,** 1–6. doi:10.1016/J.MIB.2018.01.006 (2018).

51. Djamei, A. *et al.* Metabolic priming by a secreted fungal effector. *Nature* **478,** 395–398. doi:10.1038/nature10454 (2011).

52. Van Den Burg, H. A., Harrison, S. J., Joosten, M. H., Vervoort, J. & De Wit, P. J. Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Molecular Plant-Microbe Interactions* **19,** 1420–1430. doi:10.1094/MPMI-19-1420 (2007).

53. Van Esse, H. P. *et al.* The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *The Plant Cell* **20,** 1948–1963. doi:10.1105/TPC.108.059394 (2008).

54. Okmen, B. *et al.* Detoxification of α-tomatine by *Cladosporium fulvum* is required for full virulence on tomato. *New Phytologist* **198,** 1203–1214. doi:10.1111/NPH.12208 (2013).

55. Dutton, M. V. & Evans, C. S. Oxalate production by fungi: its role in pathogenicity and ecology in the soil environment. *Canadian Journal of Microbiology* **42,** 881–895. doi:10.1139/M96-114 (2011).

56. Manning, V. A., Hardison, L. K. & Ciuffetti, L. M. Ptr ToxA interacts with a chloroplast-localized protein. *Molecular Plant-Microbe Interactions* **20,** 168–177. doi:10.1094/MPMI-20-2-0168 (2007).

57. Faris, J. D. *et al.* A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. *PNAS* **107,** 13544–13549. doi:10.1073/pnas.1004090107 (2010).

58. Velásquez, A. C., Castroverde, C. D. M. & He, S. Y. Plant–pathogen warfare under changing climate conditions. *Current Biology* **28,** R619–R634. doi:10.1016/J.CUB.2018.03.054 (2018).

59. Brundrett, M. Diversity and classification of mycorrhizal associations. *Biological Reviews* **79,** 473–495. doi:10.1017/S1464793103006316 (2004).

60. Kakouridis, A. *et al.* Routes to roots: direct evidence of water transport by arbuscular mycorrhizal fungi to host plants. *New Phytologist* **236,** 210–221. doi:10.1111/NPH.18281 (2022).

61. Salvioli di Fossalunga, A. & Novero, M. To trade in the field: the molecular determinants of arbuscular mycorrhiza nutrient exchange. *Chemical and Biological Technologies in Agriculture* **6,** 1–12. doi:10.1186/S40538-019-0150-7 (2019).

62. Karandashov, V. & Bucher, M. Symbiotic phosphate transport in arbuscular mycorrhizas. *Trends in Plant Science* **10,** 22–29. doi:10.1016/j.tplants.2004.12.003 (2005).

63. Bucher, M. Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. *New Phytologist* **173,** 11–26. doi:10.1111/J.1469-8137.2006.01935.X (2007).

64. Malar C, M. *et al.* The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology* **31,** 1570–1577.e4. doi:10.1016/J.CUB.2021.01.058 (2021).

65. Keymer, A. & Gutjahr, C. Cross-kingdom lipid transfer in arbuscular mycorrhiza symbiosis and beyond. *Current Opinion in Plant Biology* **44,** 137–144. doi:10.1016/J.PBI.2018.04.005 (2018).

66. Kiers, E. T. *et al.* Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333,** 880–882. doi:10.1126/science.1208473 (2011).

67. Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C. & Hibbett, D. S. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14,** 760–773. doi:10.1038/nrmicro.2016.149 (2016).

68. Liu, Y., Li, X. & Kou, Y. Ectomycorrhizal fungi: Participation in nutrient turnover and community assembly pattern in forest ecosystems. *Forests* **11,** 453. doi:10.3390/F11040453 (2020).

69. Miyauchi, S. *et al.* Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nature Communications* **11,** 1–17. doi:10.1038/s41467-020-18795-w (2020).

70. Hardoim, P. R. *et al.* The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews* **79,** 293–320. doi:10.1128/MMBR.00050-14 (2015).

71. Schulz, B. & Boyle, C. The endophytic continuum. *Mycological research* **109,** 661–686. doi:10.1017/S095375620500273X (2005).

72. Collinge, D. B., Jensen, B. & Jørgensen, H. J. Fungal endophytes in plants and their relationship to Plant Disease. *Current Opinion in Microbiology* **69,** 102177. doi:10.1016/J.MIB.2022.102177 (2022).

73. Glynou, K. *et al.* The local environment determines the assembly of root endophytic fungi at a continental scale. *Environmental Microbiology* **18,** 2418–2434. doi:10.1111/1462-2920.13112 (2016).

74. Kia, S. H. *et al.* Influence of phylogenetic conservatism and trait convergence on the interactions between fungal root endophytes and plants. *ISME Journal* **11,** 777–790. doi:10.1038/ismej.2016.140 (2017).

75. Glynou, K., Nam, B., Thines, M. & Maciá-Vicente, J. G. Facultative root-colonizing fungi dominate endophytic assemblages in roots of nonmycorrhizal *Microthlaspi* species. *New Phytologist* **217,** 1190–1202. doi:10.1111/nph.14873 (2018).

76. Fesel, P. H. & Zuccaro, A. Dissecting endophytic lifestyle along the parasitism/mutualism continuum in *Arabidopsis*. *Current Opinion in Microbiology* **32,** 103–112. doi:10.1016/j.mib.2016.05.008 (2016).

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

78. Hiruma, K. *et al.* 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 (2016).

79. Hacquard, S. *et al.* Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nature Communications* **7,** 1–13. doi:10.1038/ncomms11362 (2016).

80. Hettiarachchige, I. K. *et al.* Global changes in asexual *Epichloë* transcriptomes during the early stages, from seed to seedling, of symbiotum establishment. *Microorganisms* **9,** 991. doi:10.3390/MICROORGANISMS9050991 (2021).

81. Almario, J. *et al.* Root-associated fungal microbiota of nonmycorrhizal *Arabis alpina* and its contribution to plant phosphorus nutrition. *PNAS* **114,** E9403–E9412. doi:10.1073/pnas.1710455114 (2017).

82. Knapp, D. G. *et al.* 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 (2018).

83. Zhang, H. *et al.* A 2-kb mycovirus converts a pathogenic fungus into a beneficial endophyte for *Brassica* protection and yield enhancement. *Molecular Plant* **13,** 1420–1433. doi:10.1016/J.MOLP.2020.08.016 (2020).

84. Hiruma, K. *et al.* A fungal secondary metabolism gene cluster enables mutualist-pathogen transition in root endophyte *Colletotrichum tofieldiae*. *bioRxiv*. doi:10.1101/2022.07.07.499222 (2022).

85. Poveda, J., Díaz-González, S., Díaz-Urbano, M., Velasco, P. & Sacristán, S. Fungal endophytes of Brassicaceae: Molecular interactions and crop benefits. *Frontiers in Plant Science,* 2605. doi:10.3389/FPLS.2022.932288 (2022).

86. Maciá-Vicente, J. G. *et al.* Nutrient availability does not affect community assembly in root-associated fungi but determines fungal effects on plant growth. *mSystems* **7** (ed Schadt, C. W.) e00304–22. doi:10.1128/MSYSTEMS.00304-22 (2022).

87. Durán, P. *et al.* Microbial interkingdom interactions in roots promote Arabidopsis survival. *Cell* **175,** 973–983.e14. doi:10.1016/j.cell.2018.10.020 (2018).

88. Mahdi, L. K. *et al.* The fungal root endophyte *Serendipita vermifera* displays inter-kingdom synergistic beneficial effects with the microbiota in *Arabidopsis thaliana* and barley. *The ISME Journal* **16,** 876–889. doi:10.1038/s41396-021-01138-y (2021).

89. Li, J., Cornelissen, B. & Rep, M. Host-specificity factors in plant pathogenic fungi. *Fungal Genetics and Biology* **144,** 103447. doi:10.1016/J.FGB.2020.103447 (2020).

90. Ayukawa, Y. *et al.* A pair of effectors encoded on a conditionally dispensable chromosome of Fusarium oxysporum suppress host-specific immunity. *Communications Biology* **4,** 1–12. doi:10.1038/s42003-021-02245-4 (2021).

91. Mesny, F. *et al.* Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. *Nature Communications* **12,** 1–15. doi:10.1038/s41467-021-27479-y (2021).

92. Hou, S. *et al.* A microbiota–root–shoot circuit favours *Arabidopsis* growth over defence under suboptimal light. *Nature Plants* **7,** 1078–1092. doi:10.1038/s41477-021-00956-4 (2021).

93. Van der Heijden, M. G., de Bruin, S., Luckerhoff, L., van Logtestijn, R. S. & Schlaeppi, K. A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. *The ISME Journal* **10,** 389–399. doi:10.1038/ismej.2015.120 (2015).

94. Carrión, V. J. *et al.* Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366,** 606–612. doi:10.1126/SCIENCE.AAW9285 (2019).

95. Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E. E. & van der Heijden, M. G. A. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications* **10,** 1–10. doi:10.1038/s41467-019-12798-y (2019).

96. Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist* **220,** 1108–1115. doi:10.1111/nph.14976 (2018).

97. Delavaux, C. S. *et al.* Mycorrhizal fungi influence global plant biogeography. *Nature Ecology and Evolution* **3,** 424–429. doi:10.1038/s41559-019-0823-4 (2019).

98. Soudzilovskaia, N. A. *et al.* Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nature Communications* **10,** 1–10. doi:10.1038/s41467-019-13019-2 (2019).

99. Steidinger, B. S. *et al.* Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569,** 404–408. doi:10.1038/s41586-019-1128-0 (2019).

100. Lugtenberg, B. J., Caradus, J. R. & Johnson, L. J. Fungal endophytes for sustainable crop production. *FEMS Microbiology Ecology* **92,** fiw194. doi:10.1093/FEMSEC/FIW194 (2016).

101. U'Ren, J. M. *et al.* Host availability drives distributions of fungal endophytes in the imperilled boreal realm. *Nature Ecology and Evolution* **3,** 1430–1437. doi:10.1038/s41559-019-0975-2 (2019).

102. Maciá-Vicente, J. G., Piepenbring, M. & Koukol, O. Brassicaceous roots as an unexpected diversity hot-spot of helotialean endophytes. *IMA Fungus* **11,** 1–23. doi:10.1186/s43008-020-00036-w (2020).

103. Oita, S. *et al.* Climate and seasonality drive the richness and composition of tropical fungal endophytes at a landscape scale. *Communications Biology* **4,** 1–11. doi:10.1038/s42003-021-01826-7 (2021).

104. Jumpponen, A., Herrera, J., Porras-Alfaro, A. & Rudgers, J. in *Biogeography of Mycorrhizal Symbiosis* 195–222 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-56363-3_10.

105. Bokati, D., Herrera, J. & Poudel, R. Soil influences colonization of root-associated fungal endophyte communities of maize, wheat, and their progenitors. *Journal of Mycology* **2016,** 1–9. doi:10.1155/2016/8062073 (2016).

106. Card, S. D. *et al.* Beneficial endophytic microorganisms of *Brassica* – A review. *Biological Control* **90,** 102–112. doi:10.1016/J.BIOCONTROL.2015.06.001 (2015).

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

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

109. Spatafora, J. W., Sung, G.-H., J.-M., S., Hywel-Jones, N. L. & White, J. F. Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. *Molecular Ecology* **16,** 1701–1711. doi:10.1111/J.1365-294X.2007.03225.X (2007).

110. Xu, X. H. *et al.* The rice endophyte *Harpophora oryzae* genome reveals evolution from a pathogen to a mutualistic endophyte. *Scientific Reports* **4,** 1–9. doi:10.1038/srep05783 (2014).

111. Weiß, M., Waller, F., Zuccaro, A. & Selosse, M.-A. Sebacinales – one thousand and one interactions with land plants. *New Phytologist* **211,** 20–40. doi:10.1111/nph.13977 (2016).

112. Větrovský, T. *et al.* GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Scientific Data* **7,** 1–14. doi:10.1038/s41597-020-0567-7 (2020).

113. Nguyen, N. H. *et al.* 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 (2016).

114. Selosse, M.-A., Schneider-Maunoury, L. & Martos, F. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytologist* **217,** 968–972. doi:10.1111/NPH.14983 (2018).

115. Zuccaro, A. *et al.* Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathogens* **7,** e1002290. doi:10.1371/journal.ppat.1002290 (2011).

116. David, A. S. *et al.* Draft genome sequence of *Microdochium bolleyi*, a dark septate fungal endophyte of beach grass. *Genome Announcements* **4.** doi:10.1128/GENOMEA.00270-16 (2016).

117. Walker, A. K. *et al.* Full genome of *Phialocephala scopiformis* DAOMC 229536, a fungal endophyte of spruce producing the potent anti-insectan compound rugulosin. *Genome Announcements* **4.** doi:10.1128/GENOMEA.01768-15 (2016).

118. Wu, W. *et al.* Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Applied Microbiology and Biotechnology,* 2603–2618. doi:10.1007/s00253-017-8091-1 (2017).

119. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20,** 1–14. doi:10.1186/s13059-019-1832-y (2019).

120. Csűös, M. Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26,** 1910–1912. doi:10.1093/BIOINFORMATICS/BTQ315 (2010).

121. Shah, F. *et al.* Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytologist* **209,** 1705–1719. doi:10.1111/nph.13722 (2016).

122. Pellegrin, C., Morin, E., Martin, F. M. & Veneault-Fourrey, C. Comparative analysis of secretomes from ectomycorrhizal fungi with an emphasis on small-secreted proteins. *Frontiers in Microbiology* **6,** 1278. doi:10.3389/fmicb.2015.01278 (2015).

123. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47,** D607–D613. doi:10.1093/nar/gky1131 (2019).

124. Si Tung Ho, L. & Ané, C. A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Systematic Biology* **63,** 397–408. doi:10.1093/sysbio/syu005 (2014).

125. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports* **8,** 1–17. doi:10.1038/s41598-018-28948-z (2018).

126. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37,** 907–915. doi:10.1038/s41587-019-0201-4 (2019).

127. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550. doi:10.1186/s13059-014-0550-8 (2014).

128. Cantarel, B. I. *et al.* The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research* **37,** 233–238. doi:10.1093/nar/gkn663 (2009).

129. Curran, D. M., Gilleard, J. S. & Wasmuth, J. D. MIPhy: Identify and quantify rapidly evolving members of large gene fam. *PeerJ* **2018,** e4873. doi:10.7717/peerj.4873 (2018).

130. Atanasova, L. *et al.* Evolution and functional characterization of pectate lyase PEL12, a member of a highly expanded *Clonostachys rosea* polysaccharide lyase 1 family. *BMC Microbiology* **18,** 1–19. doi:10.1186/s12866-018-1310-9 (2018).

131. Keim, J., Mishra, B., Sharma, R., Ploch, S. & Thines, M. Root-associated fungi of *Arabidopsis thaliana* and *Microthlaspi perfoliatum*. *Fungal Diversity* **66,** 99–111. doi:10.1007/s13225-014-0289-2 (2014).

132. Vannier, N., Agler, M. & Hacquard, S. Microbiota-mediated disease resistance in plants. *PLOS Pathogens* **15,** e1007740. doi:10.1371/journal.ppat.1007740 (2019).

133. Lofgren, L. A. *et al.* Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology* **28,** 721–730. doi:10.1111/MEC.14995 (2019).

134. Karasov, T. L. *et al. Arabidopsis thaliana* and *Pseudomonas* pathogens exhibit stable associations over evolutionary timescales. *Cell Host and Microbe* **24,** 168–179.e4. doi:10.1016/j.chom.2018.06.011 (2018).

135. Karasov, T. L. *et al.* The relationship between microbial population size and disease in the *Arabidopsis thaliana* phyllosphere. *bioRxiv.* doi:10.1101/828814 (2020).

136. Benen, J. A., Kester, H. C., Pařenicová, L. & Visser, J. Characterization of *Aspergillus niger* pectate lyase A. *Biochemistry* **39,** 15563–15569. doi:10.1021/bi000693w (2000).

137. Bauer, S., Vasu, P., Persson, S., Mort, A. J. & Somerville, C. R. Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *PNAS* **103,** 11417–11422. doi:10.1073/pnas.0604632103 (2006).

138. Bacete, L. *et al. Arabidopsis* response reGUlator 6 (ARR6) modulates plant cell-wall composition and disease resistance. *Molecular Plant-Microbe Interactions* **33,** 767–780. doi:10.1094/MPMI-12-19-0341-R (2020).

139. Molina, A. *et al. Arabidopsis* cell wall composition determines disease resistance specificity and fitness. *PNAS* **118,** e2010243118. doi:10.1073/pnas.2010243118 (2021).

140. Sun, Z.-B. *et al.* Biology and applications of *Clonostachys rosea*. *Journal of Applied Microbiology* **129,** 486–495. doi:10.1111/JAM.14625 (2020).

141. Broberg, M. *et al.* Comparative genomics highlights the importance of drug efflux transporters during evolution of mycoparasitism in *Clonostachys* subgenus *Bionectria* (Fungi, Ascomycota, Hypocreales). *Evolutionary Applications* **14,** 476–497. doi:10.1111/EVA.13134 (2021).

142. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461. doi:10.1093/bioinformatics/btq461 (2010).

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

144. Grabherr, M. G. *et al.* Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology* **29,** 644. doi:10.1038/NBT.1883 (2011).

145. Grigoriev, I. V. *et al.* MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42,** D699–D704. doi:10.1093/nar/gkt1183 (2014).

146. Nilsson, R. H. *et al.* 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 (2019).

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

148. Cohen, O., Ashkenazy, H., Belinky, F., Huchon, D. & Pupko, T. GLOOME: gain loss mapping engine. *Bioinformatics* **26,** 2914–2915. doi:10.1093/bioinformatics/btq549 (2010).

149. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *The Journal of Machine Learning Research* **12,** 2825–2830 (2011).

150. Seppey, M., Manni, M. & Zdobnov, E. M. in *Methods in Molecular Biology* 227–245 (Humana Press Inc., 2019). doi:10.1007/978-1-4939-9173-0_14.

151. Morin, E. *et al.* Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytologist* **222,** 1584–1598. doi:10.1111/nph.15687 (2019).

152. Rawlings, N. D., Barrett, A. J. & Finn, R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **44,** D343–D350. doi:10.1093/nar/gkv1118 (2016).

153. Fischer, M. & Pleiss, J. The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research* **31,** 319–321. doi:doi.org/10.1093/nar/gkg015 (2003).

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

155. Deorowicz, S., Debudaj-Grabysz, A. & Gudys, A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Scientific Reports* **6,** 1–13. doi:10.1038/srep33964 (2016).

156. Eddy, S. R. Accelerated profile HMM searches. *PLoS Computational Biology* **7.** doi:10.1371/journal.pcbi.1002195 (2011).

157. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research* **13,** 2498–2504. doi:10.1101/gr.1239303 (2003).

158. Morris, J. H. *et al.* ClusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **12,** 436. doi:10.1186/1471-2105-12-436 (2011).

159. Gruber, B. D., Giehl, R. F., Friedel, S. & von Wirén, N. Plasticity of the *Arabidopsis* root system under nutrient deficiencies. *Plant Physiology* **163,** 161–179. doi:10.1104/pp.113.218453 (2013).

160. Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* **6,** 107–128. doi:10.3102/10769986006002107 (1981).

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

162. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930. doi:10.1093/bioinformatics/btt656 (2014).

163. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35,** 2084–2092. doi:10.1093/bioinformatics/bty895 (2019).

164. Möller, M. & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* **15,** 756–771. doi:10.1038/nrmicro.2017.76 (2017).

165. Temporini, E. D. & VanEtten, H. D. Distribution of the pea pathogenicity (PEP) genes in the fungus *Nectria haematococca* mating population VI. *Current Genetics* **41,** 107–114. doi:10.1007/S00294-002-0279-X (2002).

166. Ma, L. J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464,** 367–373. doi:10.1038/nature08850 (2010).

167. Feurtey, A. *et al.* Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria. *BMC Genomics* **21,** 1–15. doi:10.1186/S12864-020-06871-W/FIGURES/5 (2020).

168. Seidl, M. F. & Thomma, B. P. Transposable elements direct the coevolution between plants and microbes. *Trends in Genetics* **33,** 842–851. doi:10.1016/j.tig.2017.07.003 (2017).

169. Faino, L. *et al.* Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research* **26,** 1091–1100. doi:10.1101/GR.204974.116 (2016).

170. Torres, D. E., Oggenfuss, U., Croll, D. & Seidl, M. F. Genome evolution in fungal plant pathogens: looking beyond the two-speed genome model. *Fungal Biology Reviews* **34,** 136–143. doi:10.1016/J.FBR.2020.07.001 (2020).

171. Dong, S., Raffaele, S. & Kamoun, S. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35,** 57–65. doi:10.1016/J.GDE.2015.09.001 (2015).

172. Snelders, N. C., Rovenich, H. & Thomma, B. P. H. J. Microbiota manipulation through the secretion of effector proteins is fundamental to the wealth of lifestyles in the fungal kingdom. *FEMS Microbiology Reviews* **2022,** 1–16. doi:10.1093/FEMSRE/FUAC022 (2022).

173. Stukenbrock, E. H. & McDonald, B. A. Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions* **22,** 371–380. doi:10.1094/MPMI-22-4-0371 (2009).

174. Redkar, A., Sabale, M., Zuccaro, A. & Di Pietro, A. Determinants of endophytic and pathogenic lifestyle in root colonizing fungi. *Current Opinion in Plant Biology* **67,** 102226. doi:10.1016/J.PBI.2022.102226 (2022).

175. Muñoz-Barrios, A. *et al.* Differential expression of fungal genes determines the lifestyle of *Plectosphaerella* strains during *Arabidopsis thaliana* colonization. *Molecular Plant-Microbe Interactions* **33,** 1299–1314. doi:10.1094/MPMI-03-20-0057-R (2020).

176. Schirrmann, M. K. *et al.* Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* **27,** 3070–3086. doi:10.1111/MEC.14585 (2018).

177. Schirrmann, M. K. & Leuchtmann, A. The role of host-specificity in the reproductive isolation of *Epichloë* endophytes revealed by reciprocal infections. *Fungal Ecology* **15,** 29–38. doi:10.1016/J.FUNECO.2015.02.004 (2015).

178. Wippel, K. *et al.* Host preference and invasiveness of commensal bacteria in the *Lotus* and *Arabidopsis* root microbiota. *Nature Microbiology 2021 6:9* **6,** 1150–1162. doi:10.1038/s41564-021-00941-9 (2021).

179. Deshpande, V. *et al.* Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* **108,** 1–5. doi:10.3852/14-293 (2017).

180. Giraldo, A. & Crous, P. Inside Plectosphaerellaceae. *Studies in mycology* **92,** 227–286 (2019).

181. Tedersoo, L. *et al.* The Global Soil Mycobiome consortium dataset for boosting fungal diversity research. *Fungal Diversity* **111,** 573–588. doi:10.1007/s13225-021-00493-7 (2021).

182. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* **38,** 5825–5829. doi:10.1093/MOLBEV/MSAB293 (2021).

183. Odenbach, D. *et al.* The transcription factor Con7p is a central regulator of infection-related morphogenesis in the rice blast fungus *Magnaporthe grisea*. *Molecular Microbiology* **64,** 293–307. doi:10.1111/J.1365-2958.2007.05643.X (2007).

184. Ruiz-Roldán, C., Pareja-Jaime, Y., González-Reyes, J. A. & Roncero, M. I. G. The transcription factor Con7-1 is a master regulator of morphogenesis and virulence in *Fusarium oxysporum*. *Molecular Plant-Microbe Interactions* **28,** 55–68. doi:10.1094/MPMI-07-14-0205-R (2014).

185. Gao, D. J. *et al.* First report of Chinese cabbage wilt caused by *Plectosphaerella cucumerina* in inner Mongolia of China. *Plant Disease*. doi:10.1094/PDIS-10-21-2210-PDN (2022).

186. Li, P. L., Chai, A. L., Shi, Y. X., Xie, X. W. & Li, B. J. First report of root rot caused by *Plectosphaerella cucumerina* on cabbage in China. *Mycobiology* **45,** 110–113. doi:10.5941/MYCO.2017.45.2.110 (2018).

187. Garibaldi, A., Gilardi, G., Ortu, G. & Gullino, M. L. First report of *Plectosphaerella cucumerina* on greenhouse cultured wild rocket (*Diplotaxis tenuifolia*) in Italy. *Plant Disease* **96,** 1825. doi:10.1094/PDIS-06-12-0583-PDN (2012).

188. Rivedal, H. M., Tabima, J. F., Stone, A. G. & Johnson, K. B. Identity and pathogenicity of fungi associated with root, crown, and vascular symptoms related to winter squash yield decline. *Plant Disease* **106,** 1660–1668. doi:10.1094/PDIS-09-20-2090-RE (2022).

189. Rivedal, H. M., Stone, A. G., Severns, P. M. & Johnson, K. B. 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 (2020).

190. Alam, M. W., Malik, A., Rehman, A., Sarwar, M. & Mehboob, S. First report of potato wilt caused by *Plectosphaerella cucumerina* in Pakistan. *Journal of Plant Pathology* **103,** 687–687. doi:10.1007/S42161-021-00771-Y (2021).

191. Xu, J., Xu, X. D., Cao, Y. Y. & Zhang, W. M. First report of greenhouse tomato wilt caused by *Plectosphaerella cucumerina* in China. *Plant Disease* **98,** 158. doi:10.1094/PDIS-05-13-0566-PDN (2013).

192. Carlucci, A., Raimondo, M. L., Santos, J. & Phillips, A. J. *Plectosphaerella species* associated with root and collar rots of horticultural crops in southern Italy. *Persoonia: Molecular Phylogeny and Evolution of Fungi* **28,** 34–48. doi:10.3767/003158512X638251 (2012).

193. Zhao, Y. Q., Shi, K., Yu, X. Y. & Zhang, L. J. First Report of *Alfalfa* root rot caused by *Plectosphaerella cucumerina* in inner Mongolia autonomous region of China. *Plant Disease* **105,** 87. doi:10.1094/PDIS-03-21-0515-PDN (2021).

194. Yang, L., Lu, X. H., Li, S. D. & Wu, B. M. First report of common bean (*Phaseolus vulgaris*) root rot caused by *Plectosphaerella cucumerina* in China. *Plant Disease* **102,** 1849. doi:10.1094/PDIS-10-17-1659-PDN (2018).

195. Zhang, Y. *et al.* The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. *Communications Biology* **3,** 1–12. doi:10.1038/s42003-020-0770-2 (2020).

196. Petre, B., Lorrain, C., Stukenbrock, E. H. & Duplessis, S. Host-specialized transcriptome of plant-associated organisms. *Current Opinion in Plant Biology* **56,** 81–88. doi:10.1016/J.PBI.2020.04.007 (2020).

197. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336. doi:10.1038/nmeth.f.303 (2010).

198. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv,* 081257. doi:10.1101/081257 (2016).

199. Bengtsson-Palme, J. *et al.* 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,** 914–919. doi:10.1111/2041-210X.12073 (2013).

200. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17,** 261–272. doi:10.1038/s41592-019-0686-2 (2020).

201. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12. doi:10.14806/EJ.17.1.200 (2011).

202. Huerta-Cepas, J. *et al.* 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 (2019).

203. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12,** 59–60. doi:10.1038/nmeth.3176 (2014).

204. Zhang, H. *et al.* dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46,** W95–W101. doi:10.1093/NAR/GKY418 (2018).

205. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Molecular Plant-Microbe Interactions* **35,** 146–156. doi:10.1094/MPMI-08-21-0201-R (2022).

206. Riehl, K., Riccio, C., Miska, E. A. & Hemberg, M. TransposonUltimate: Software for transposon classification, annotation and detection. *Nucleic Acids Research* **50,** e64–e64. doi:10.1093/NAR/GKAC136 (2022).

207. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30,** 772–780. doi:10.1093/MOLBEV/MST010 (2013).

208. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25,** 1972–1973. doi:10.1093/BIOINFORMATICS/BTP348 (2009).

209. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37,** 1530–1534. doi:10.1093/MOLBEV/MSAA015 (2020).

210. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6* **14,** 587–589. doi:10.1038/nmeth.4285 (2017).

211. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35,** 4453–4455. doi:10.1093/BIOINFORMATICS/BTZ305 (2019).

212. Borowiec, M. L. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4,** e1660. doi:10.7717/PEERJ.1660/SUPP-2 (2016).

213. Emms, D. M., Kelly & Affiliations. STAG: Species tree inference from all genes. *bioRxiv,* 267914. doi:10.1101/267914 (2018).

214. Zhang, C., Sayyari, E. & Mirarab, S. in *RECOMB international workshop on comparative genomics* 53–75 (2017). doi:10.1007/978-3-319-67979-2_4.

215. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094–3100. doi:10.1093/BIOINFORMATICS/BTY191 (2018).

216. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* **20,** 1–13. doi:10.1186/S13059-019-1911-0 (2019).

217. Goel, M. & Schneeberger, K. plotsr: Visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38,** 2922–2926. doi:10.1093/BIOINFORMATICS/BTAC196 (2022).

218. Selosse, M. A., Dubois, M. P. & Alvarez, N. Do Sebacinales commonly associate with plant roots as endophytes? *Mycological research* **113,** 1062–1069. doi:10.1016/J.MYCRES.2009.07.004 (2009).

219. Selosse, M. A. *et al.* The waiting room hypothesis revisited by orchids: Were orchid mycorrhizal fungi recruited among root endophytes? *Annals of Botany* **129,** 259–270. doi:10.1093/AOB/MCAB134 (2022).

Supplementary data

Direct links to the supplementary data files are provided here with their descriptions. These files can also be accessed via the Supplementary Material of the publication [91].

Supplementary Data 1

Description of the 41 newly-sequenced strains of *Arabidopsis thaliana* root mycobiota members. This table provides information regarding the phylogeny of the 41 strains (phylum, class, order, species), as well as information regarding the isolation of these strains (host, location) and statistics regarding the genome assemblies (Assembly size, number of contigs, L50, N50), and number of predicted genes. Genbank bioproject, biosample and accession IDs are provided in the last three columns of the table.
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM4_ESM.xlsx`

Supplementary Data 2

Description of the comparative genomics dataset, comprising our 41 newly-sequenced strains and 79 published fungal genomes. This table provides information regarding the phylogeny of the 120 strains (phylum, class, order, strain), as well as information regarding the isolation of these strains (host, location, niche), statistics about the genome assemblies (Assembly size, number of contigs, L50, N50), together with publication IDs (PMID) and URL from which the genomes were downloaded. It also shows the lifestyle attributed to each fungal strain (Assigned lifestyle) and the FunGuild [113] description of the associated strain or genus (Guild FG, Trophic mode FG).
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM5_ESM.xlsx`

Supplementary Data 3

Results of PERMANOVA analyses testing the effect of phylogeny and lifestyle on genomic compositions in gene repertoires. This table provides the detailed results of independent PERMANOVA analyses testing the effect of phylogeny and lifestyle on Jaccard distance matrices reflecting the genomic compositions in each gene category of interest. Tested factors are the first four Principal Components (PC) of a phylogenetic PCA, fungal lifestyle, and the interaction of each phylogenetic PC with fungal lifestyle.
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM6_ESM.xlsx`

Supplementary Data 4

Description of the 84 orthogroups segregating endophytes and mycobiota members from other fungi.
a) This table provides information regarding the 84 gene families that best segregate endophytes and mycobiota members from others, according to our SVM-RFE classifier ($R^2 = 0.8$), including enrichment/depletion scores in the fungi of interest (Enrichment in EF+MyM) and associated ANOVA P-values (FDR), support vector coefficients in the classifier (SVM coefficients), representative sequence of the family, information about functional annotation (curated description, curated group) and associated COG family used for co-expression analysis.
b) This table provides the results of a GO enrichment analysis performed with GOATOOLS [125] on the 84 orthogroups determinant for endophytism. GOATOOLS performs a two-sided Fisher's exact test.Correction of p-values into FDR was performed using the Benjamini-Hochberg method.
c) Coexpression scores of the 84 gene families in fungal transcriptomic data sets, according to STRING-db [123].
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM7_ESM.xlsx`

Supplementary Data 5

Differential fungal gene expression in planta vs. on medium. These tables provide read mapping statistics (RPKM values in control samples and in planta test samples) differential expression statistics (baseMean, log2FoldChange, lfcSE, pvalue, padj) from DESeq2 [127], together with functional annotation information (SSP, CAZyme IDs and descriptions, MEROPS Protease IDs and descriptions, Lipase IDs, KOG IDs and descriptions, EC IDs and descriptions, InterPro IDs and descriptions, GO IDs and descriptions. Statistical testing for differential expression was performed by a two-sided Wald test as implemented in DESeq2. Correction of p-values into adjusted p-values (padj) was performed with the DESeq2 built-in method.
5a) *Chaetomium sp.* MPI-CAGE-AT-0009 (*Cs*)
5b) *Macrophomina phaseolina* MPI-SDFR-AT-0080 (*Mp*)
5c) *Paraphoma chrysanthemicola* MPI-GEGE-AT-0034 (*Pc*)
5d) *Phaeosphaeria sp.* MPI-PUGE-AT-0046c (*Ps*)
5e) *Truncatella angustata* MPI-SDFR-AT-0073 (*Ta*)
5f) *Halenospora varia* MPI-CAGE-AT-0135 (*Hv*).
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM8_ESM.xlsx`

Supplementary Data 6

Differential *Arabidopsis thaliana* gene expression inoculated with fungi vs. mocktreated. These tables provide read mapping statistics (RPKM values in control mock-treated samples and inoculated test samples) differential expression statistics (baseMean, log2FoldChange, lfcSE, pvalue, padj) from DESeq2 [127], and short gene descriptions from TAIR10. Statistical testing for differential expression was performed by a two-sided Wald test as implemented in DESeq2. Correction of p-values into adjusted p-values (padj) was performed with the DESeq2 built-in method.

6a) *A. thaliana* inoculated with *Chaetomium sp.* MPI-CAGE-AT-0009 (*Cs*)

6b) *A. thaliana* inoculated with *Macrophomina phaseolina* MPI-SDFR-AT-0080 (*Mp*)

6c) *A. thaliana* inoculated with *Paraphoma chrysanthemicola* MPI-GEGE-AT-0034 (*Pc*)

6d) *A. thaliana* inoculated with *Phaeosphaeria sp.* MPI-PUGE-AT-0046c (*Ps*)

6e) *A. thaliana* inoculated with *Truncatella angustata* MPI-SDFR-AT-0073 (*Ta*)

6f) *A. thaliana* inoculated with *Halenospora varia* MPI-CAGE-AT-0135 (*Hv*).

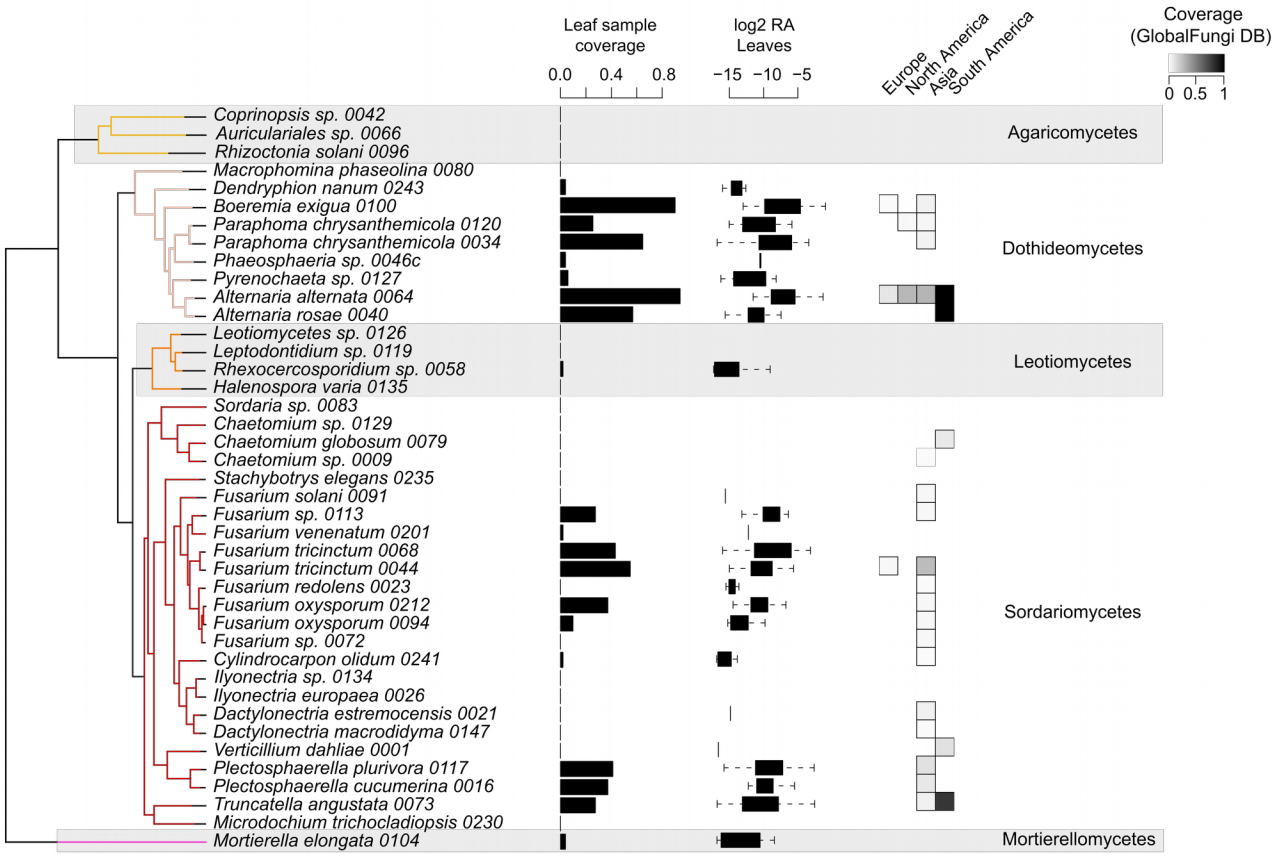
`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM9_ESM.xlsx`

Supplementary Data 7

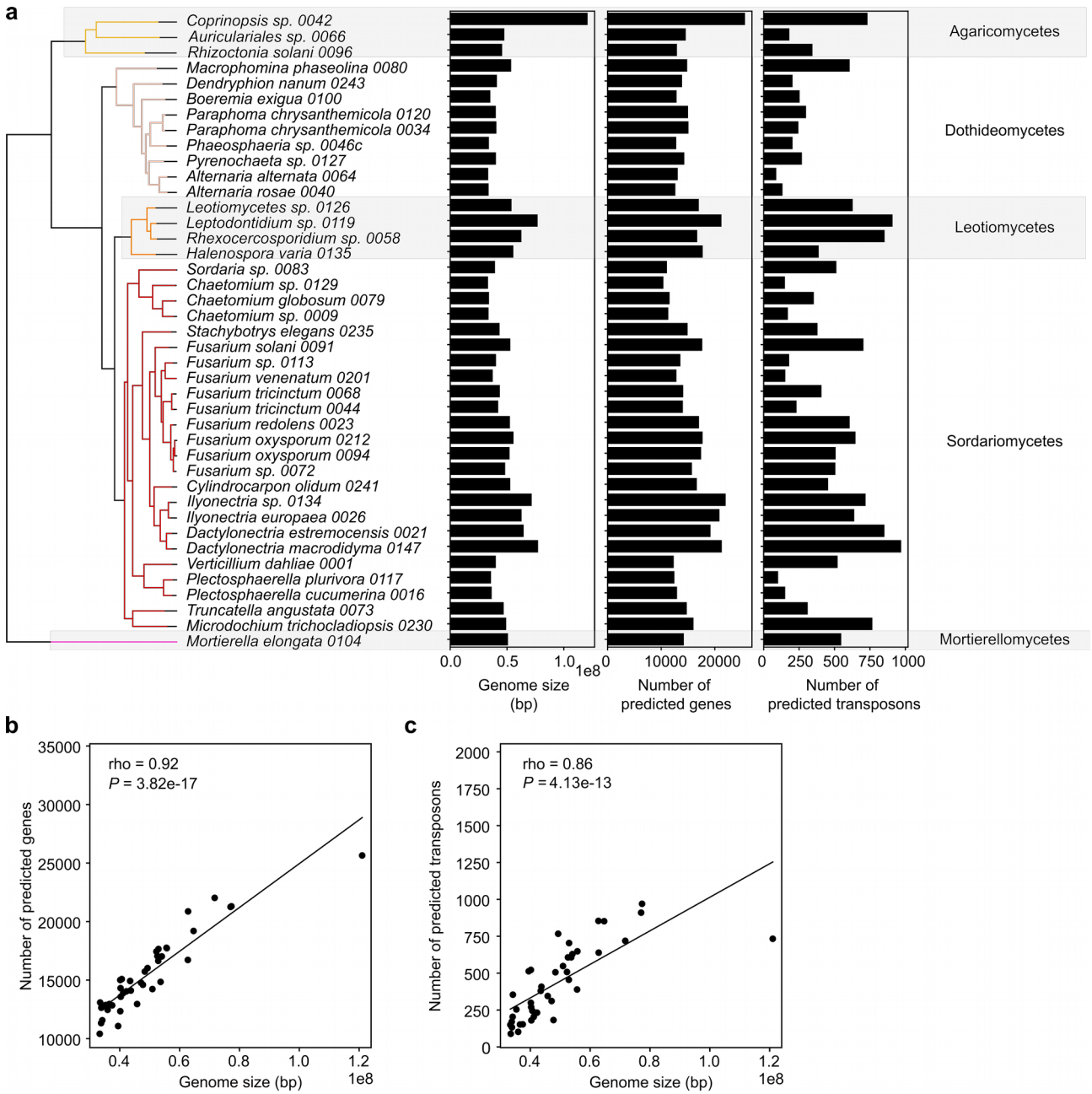
Description of the 11 orthogroups segregating detrimental mycobiota members from others. This table provides information regarding the 11 gene families that best segregate detrimental mycobiota members from neutral and beneficial ones, according to our SVM-RFE classifier ($R^2 = 0.88$), including enrichment/depletion scores in the fungi of interest (Enrichment in detrimental fungi) and associated ANOVA P-values (FDR), support vector coefficients in the classifier (SVM coefficients), representative sequence of the family and functional annotation (curated description).

`https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-021-27479-y/MediaObjects/41467_2021_27479_MOESM10_ESM.xlsx`

Supplementary figures and tables

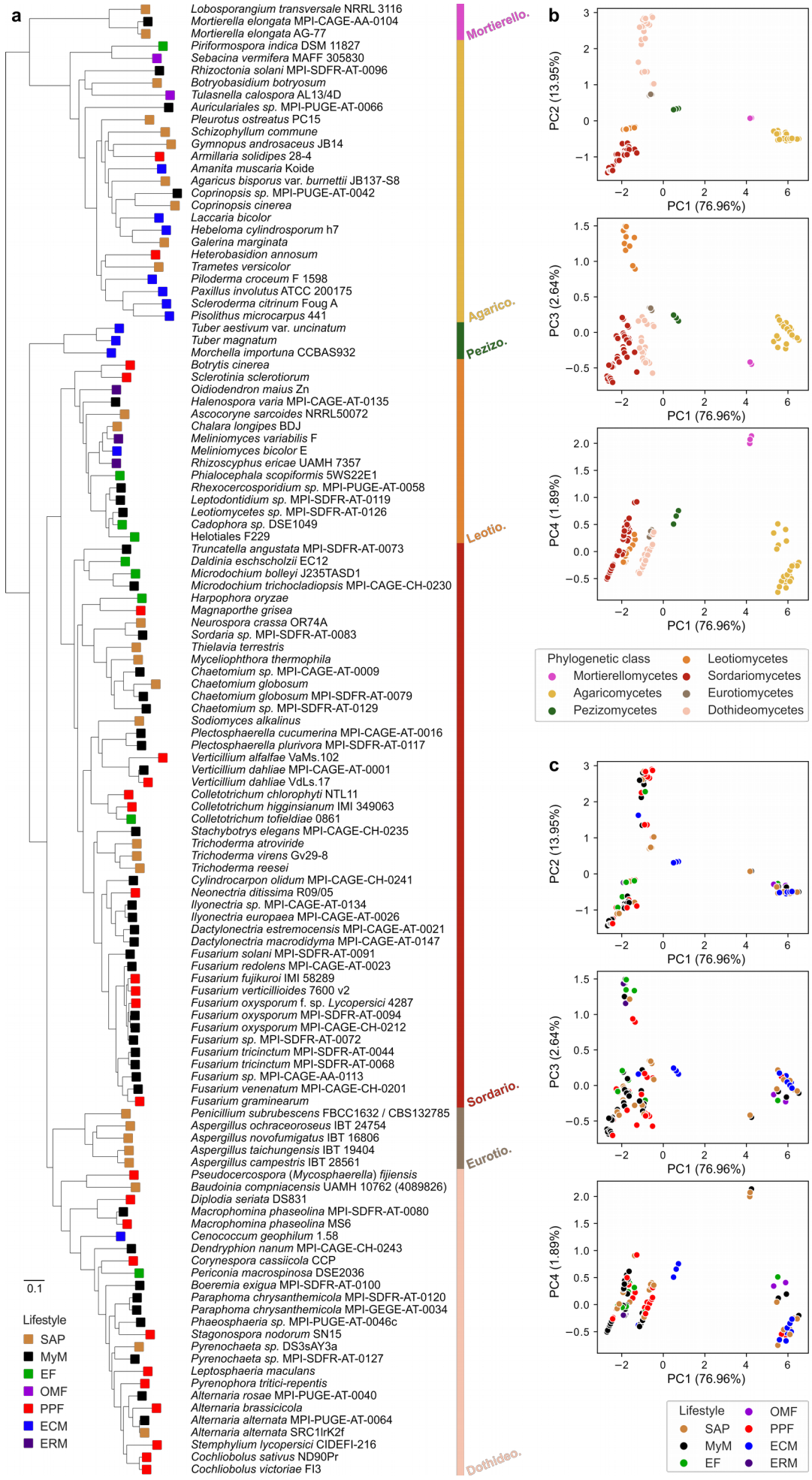


Supplementary figure 1: Prevalence and abundance profiles of 41 root-colonizing fungi across naturally occurring *A. thaliana* shoot mycobiomes.
Relative abundance and sample coverage across leaf samples taken from wild *A. thaliana* plants harvested at two locations in Germany (Cologne and Tübingen [32]). Fungal rDNA ITS2 sequences were directly mapped to raw sequencing reads with a 97% similarity threshold. Leaf sample coverage refers to the percentage of leaf samples with relative abundances superior to 0.01% (leaf samples n=51). On relative abundance boxplots, boxes are delimited by first and third quartiles and whiskers extend to minimum and maximum values. Unmapped reads were used to estimate relative abundance of other fungal species. The coverage of global shoot samples was estimated by checking the occurrence of rDNA ITS1 sequences in 2,647 shoot samples (Asia n=250, Europe n= 1759, North America n= 602, South America n=36) retrieved from the GlobalFungi database [112].

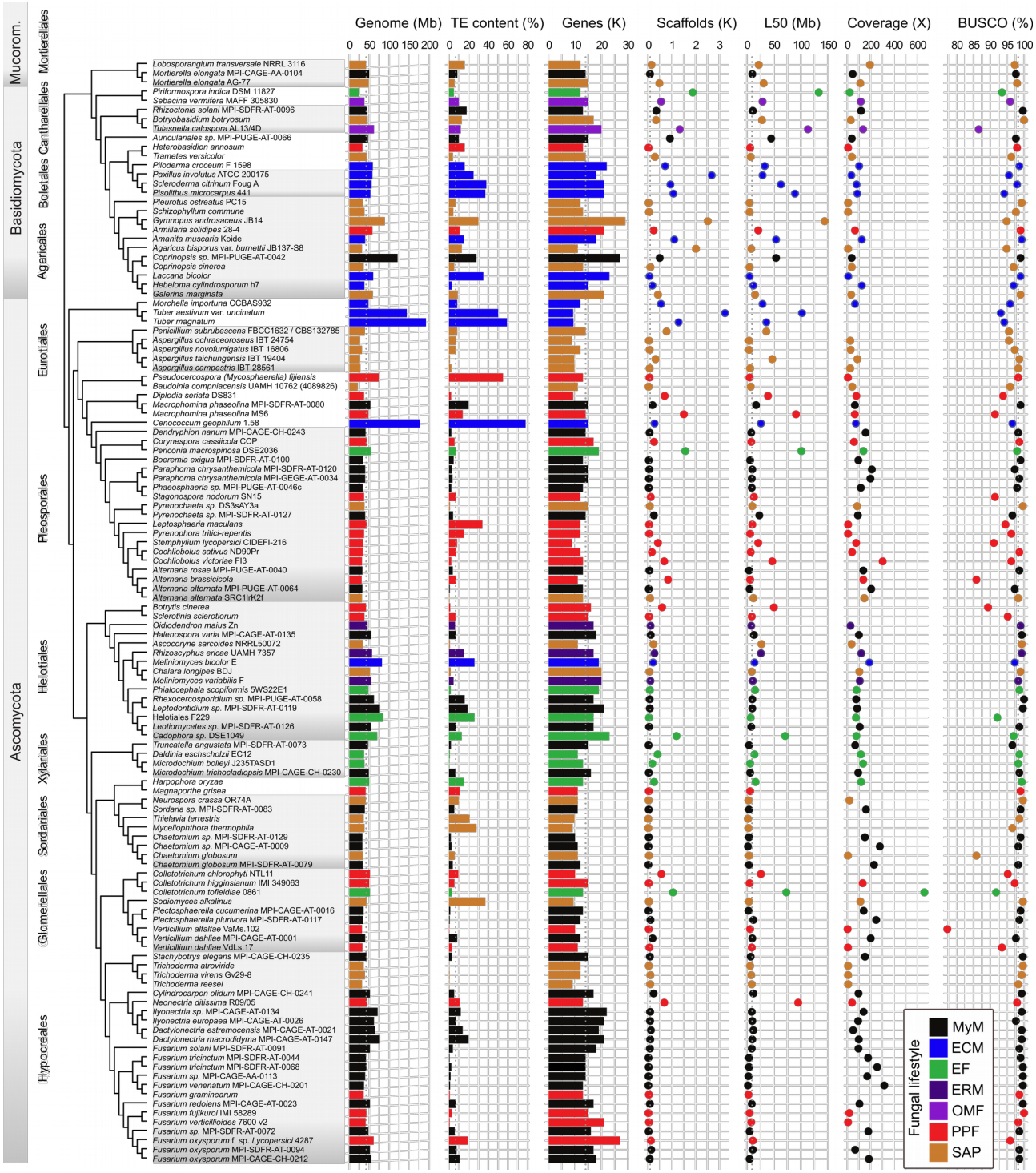


Supplementary figure 2: Link between genome size, number of genes and number of transposons across the 41 newly-sequenced fungal strains.

a, Genome assembly size, number of predicted genes and number of identified transposons in the genomes of the 41 *A. thaliana* mycobiota members. **b**, Spearman's rank correlation (rho, $P < 0.05$) between genome size and number of predicted genes. **c**, Spearman's rank correlation (rho, $P < 0.05$) between genome size and number of predicted transposable elements.

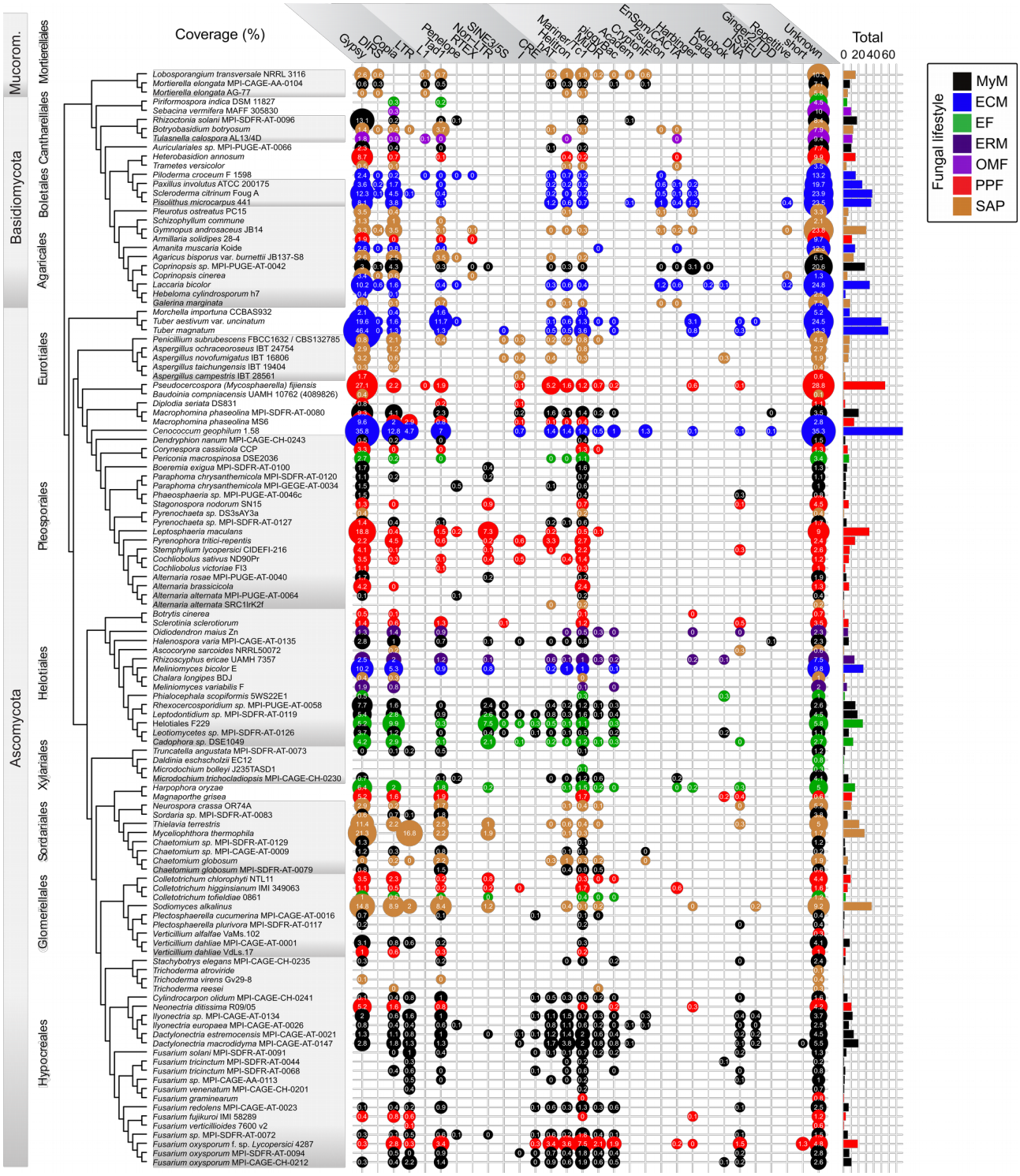

a
Lobosporangium transversale NRRL 3116
Mortierella elongata MPI-CAGE-AA-0104
Mortierella elongata AG-77
Piriformospora indica DSM 11827
Sebacina vermifera MAFF 305830
Rhizoctonia solani MPI-SDFR-AT-0096
Botryobasidium botryosum
Tulasnella calospora AL13/4D
Auriculariales sp. MPI-PUGE-AT-0066
Pleurotus ostreatus PC15
Schizophyllum commune
Gymnopus androsaceus JB14
Armillaria solidipes 28-4
Amanita muscaria Koide
Agaricus bisporus var. burnettii JB137-S8
Coprinopsis sp. MPI-PUGE-AT-0042
Coprinopsis cinerea
Laccaria bicolor
Hebeloma cylindrosporum h7
Galerina marginata
Heterobasidion annosum
Trametes versicolor
Piloderma croceum F 1598
Paxillus involutus ATCC 200175
Scleroderma citrinum Foug A
Pisolithus microcarpus 441
Tuber aestivum var. uncinatum
Tuber magnatum
Morchella importuna CCBAS932
Botrytis cinerea
Sclerotinia sclerotiorum
Oidiodendron maius Zn
Halenospora varia MPI-CAGE-AT-0135
Ascocoryne sarcoides NRRL50072
Chalara longipes BDJ
Meliniomyces variabilis F
Meliniomyces bicolor E
Rhizoscyphus ericae UAMH 7357
Phialocephala scopiformis 5WS22E1
Rhexocercosporidium sp. MPI-PUGE-AT-0058
Leptodontidium sp. MPI-SDFR-AT-0119
Leotiomycetes sp. MPI-SDFR-AT-0126
Cadophora sp. DSE1049
Helotiales F229
Truncatella angustata MPI-SDFR-AT-0073
Daldinia eschscholzii EC12
Microdochium bolleyi J235TASD1
Microdochium trichocladiopsis MPI-CAGE-CH-0230
Harpophora oryzae
Magnaporthe grisea
Neurospora crassa OR74A
Sordaria sp. MPI-SDFR-AT-0083
Thielavia terrestris
Myceliophthora thermophila
Chaetomium sp. MPI-CAGE-AT-0009
Chaetomium globosum
Chaetomium globosum MPI-SDFR-AT-0079
Chaetomium sp. MPI-SDFR-AT-0129
Sodiomyces alkalinus
Plectosphaerella cucumerina MPI-CAGE-AT-0016
Plectosphaerella plurivora MPI-SDFR-AT-0117
Verticillium alfalfae VaMs.102
Verticillium dahliae MPI-CAGE-AT-0001
Verticillium dahliae VdLs.17
Colletotrichum chlorophyti NTL11
Colletotrichum higginsianum IMI 349063
Colletotrichum tofieldiae 0861
Stachybotrys elegans MPI-CAGE-CH-0235
Trichoderma atroviride
Trichoderma virens Gv29-8
Trichoderma reesei
Cylindrocarpon olidum MPI-CAGE-CH-0241
Neonectria ditissima R09/05
Ilyonectria sp. MPI-CAGE-AT-0134
Ilyonectria europaea MPI-CAGE-AT-0026
Dactylonectria estremocensis MPI-CAGE-AT-0021
Dactylonectria macrodidyma MPI-CAGE-AT-0147
Fusarium solani MPI-SDFR-AT-0091
Fusarium redolens MPI-CAGE-AT-0023
Fusarium fujikuroi IMI 58289
Fusarium verticillioides 7600 v2
Fusarium oxysporum f. sp. Lycopersici 4287
Fusarium oxysporum MPI-SDFR-AT-0094
Fusarium oxysporum MPI-CAGE-CH-0212
Fusarium sp. MPI-SDFR-AT-0072
Fusarium tricinctum MPI-SDFR-AT-0044
Fusarium tricinctum MPI-SDFR-AT-0068
Fusarium sp. MPI-CAGE-AA-0113
Fusarium venenatum MPI-CAGE-CH-0201
Fusarium graminearum
Penicillium subrubescens FBCC1632 / CBS132785
Aspergillus ochraceoroseus IBT 24754
Aspergillus novofumigatus IBT 16806
Aspergillus taichungensis IBT 19404
Aspergillus campestris IBT 28561
Pseudocercospora (Mycosphaerella) fijiensis
Baudoinia compniacensis UAMH 10762 (4089826)
Diplodia seriata DS831
Macrophomina phaseolina MPI-SDFR-AT-0080
Macrophomina phaseolina MS6
Cenococcum geophilum 1.58
Dendryphion nanum MPI-CAGE-CH-0243
Corynespora cassiicola CCP
Periconia macrospinosa DSE2036
Boeremia exigua MPI-SDFR-AT-0100
Paraphoma chrysanthemicola MPI-SDFR-AT-0120
Paraphoma chrysanthemicola MPI-GEGE-AT-0034
Phaeosphaeria sp. MPI-PUGE-AT-0046c
Stagonospora nodorum SN15
Pyrenochaeta sp. DS3sAY3a
Pyrenochaeta sp. MPI-SDFR-AT-0127
Leptosphaeria maculans
Pyrenophora tritici-repentis
Alternaria rosae MPI-PUGE-AT-0040
Alternaria brassicicola
Alternaria alternata MPI-PUGE-AT-0064
Alternaria alternata SRC1lrK2f
Stemphylium lycopersici CIDEFI-216
Cochliobolus sativus ND90Pr
Cochliobolus victoriae FI3
Mortierello.
Agarico.
Pezizo.
Leotio.
Sordario.
Eurotio.
Dothideo.
0.1
Lifestyle
SAP
MyM
EF
OMF
PPF
ECM
ERM
b
PC2 (13.95%)
PC1 (76.96%)
PC3 (2.64%)
PC4 (1.89%)
Phylogenetic class
Mortierellomycetes
Agaricomycetes
Pezizomycetes
Leotiomycetes
Sordariomycetes
Eurotiomycetes
Dothideomycetes
c
Lifestyle
SAP
MyM
EF
OMF
PPF
ECM
ERM

Supplementary figure 3: Phylogeny of the 120-genome data set used for comparative genomics.
a, Species tree describing the phylogeny of the 120 fungal genomes used for comparative genomics. This tree was inferred from full sets of proteins by OrthoFinder [119] after orthology prediction. Leaf tip colors represent fungal lifestyles, and color strips on the right highlight the different phylogenetic classes. **bc**, Principal component analysis (PCA) calculated on phylogenetic pairwise distances extracted from the aforementioned species tree. Panels **b** and **c** represent the same PCA plot with different colors representing either the phylogenetic class (**b**) or the fungal lifestyle (**c**).
SAP: Saprotrophs, MyM: *A. thaliana* mycobiota members, EF: Endophytic Fungi, OMF: Orchid Mycorrhizal Fungi, PPF: Plant Pathogenic Fungi, ECM: Ectomycorrhiza, ERM: Ericoid Mycorrhiza.



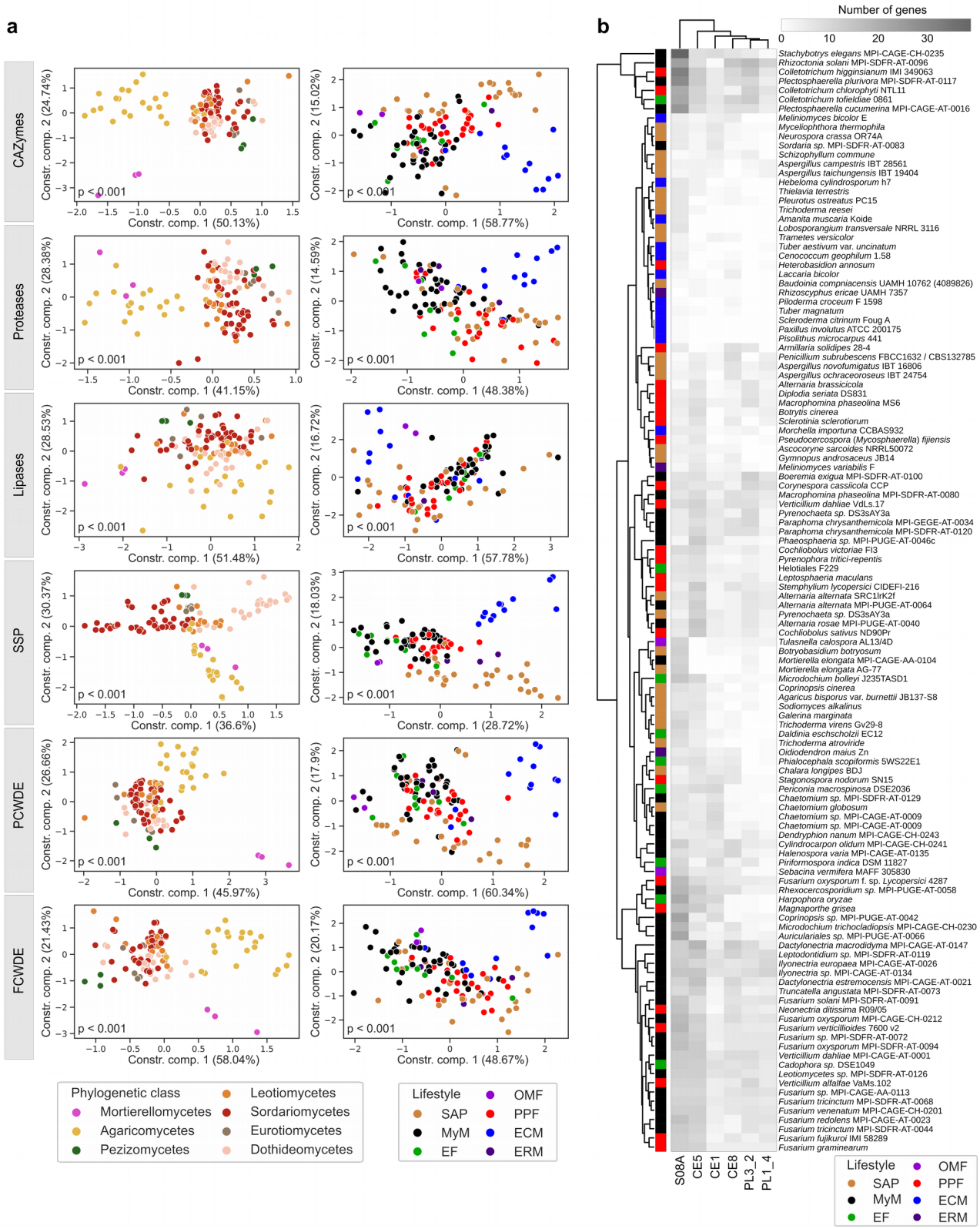
Supplementary figure 4: Genome sizes and properties of the 41 root mycobiota members, along with 79 previously published genomes used for comparative genomics.

Fungal lifestyle is depicted in color. Median values are marked with dotted line. Genome: genome size. TE content: the coverage of transposable elements in the genomes. Genes: the number of genes. Secreted: the number of predicted secreted proteins (**Methods**). Scaffolds: the number of scaffolds. L50: N50 length. Coverage: sequencing depth in fold. BUSCO: genome completeness. MyM: *A. thaliana* mycobiota members, ECM: Ectomycorrhiza, EF: Endophytic Fungi, ERM: Ericoid Mycorrhiza, OMF: Orchid Mycorrhizal Fungi, PPF: Plant Pathogenic Fungi, SAP: Saprotrophs.

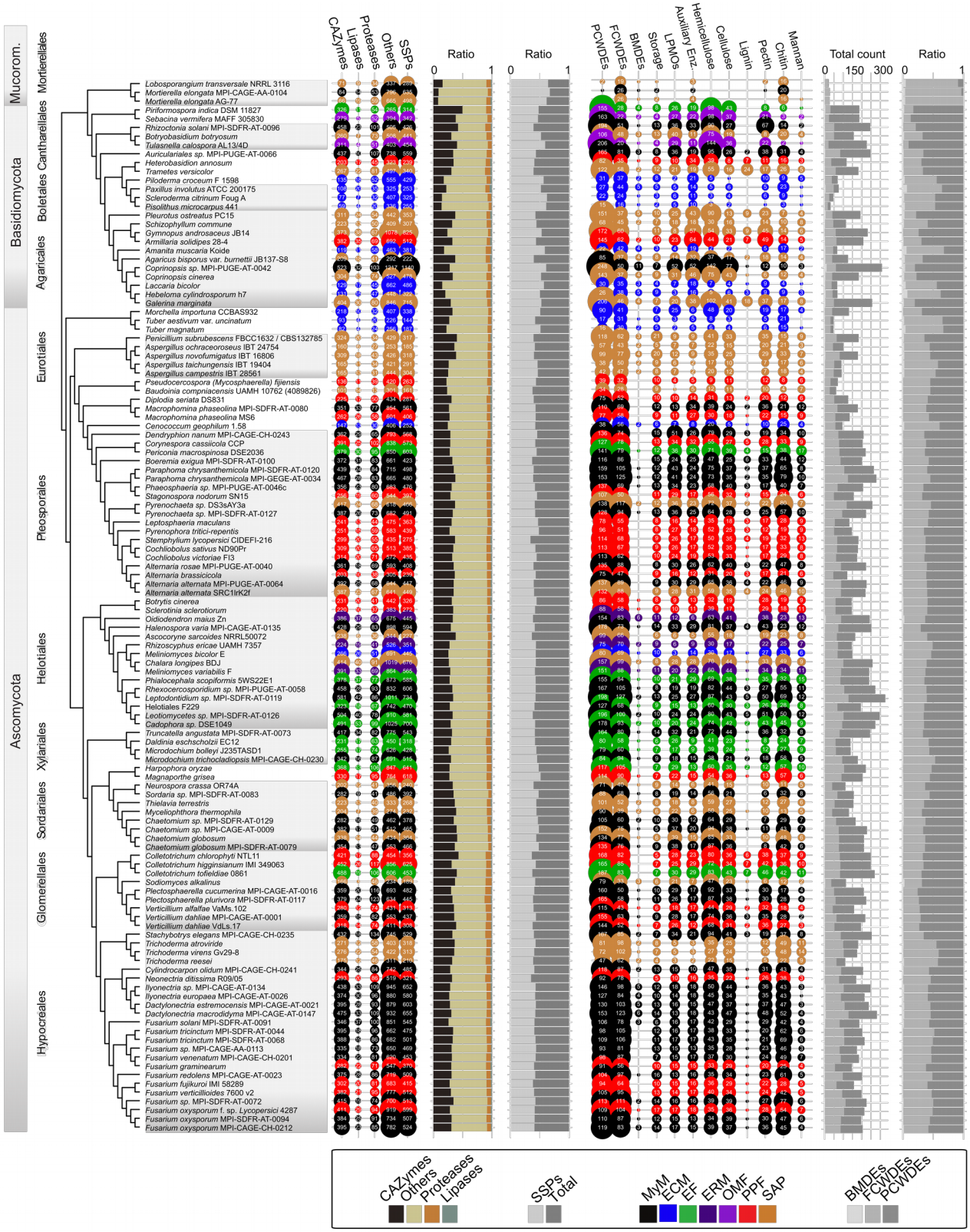


Supplementary figure 5: Compositions in transposable elements of the 120 fungal genomes used for comparative genomics.

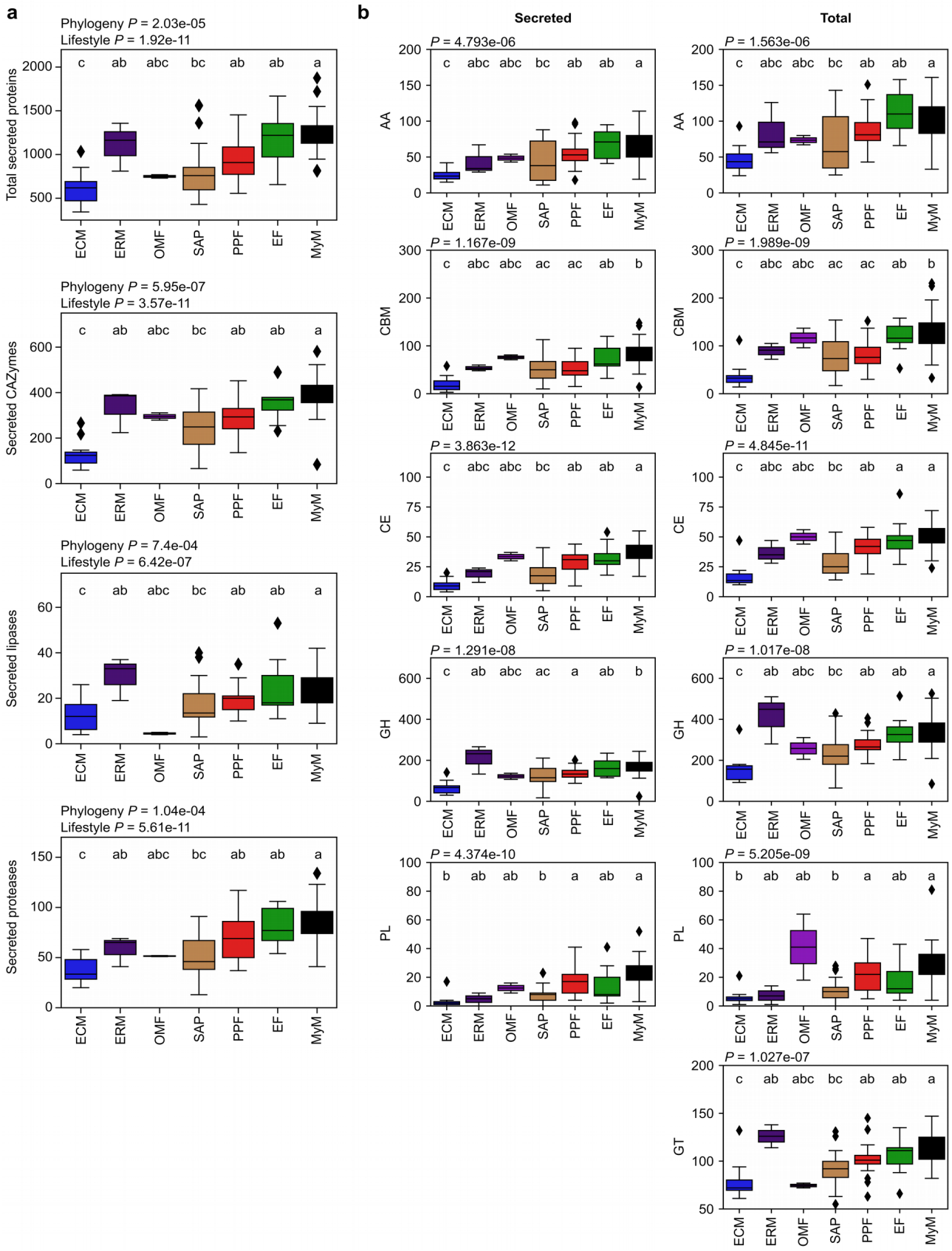
Coverage of transposable elements in the 120 fungal genome dataset. LTR: long-terminal repeat retrotransposons. Non-LTR: non-long-terminal repeat retrotransposons. DNA: DNA transposons. Repetitive/short: simple repeats. Unknown: unclassified repeated sequences. The bubble size is proportional to the coverage of each of the transposable elements (shown inside the bubbles). The barplot on the right shows the total transposon coverage per genome. MyM: *A. thaliana* mycobiota members, ECM: Ectomycorrhiza, EF: Endophytic Fungi, ERM: Ericoid Mycorrhiza, OMF: Orchid Mycorrhizal Fungi, PPF: Plant Pathogenic Fungi, SAP: Saprotrophs.


a
b
CAZymes
Proteases
Lipases
SSP
PCWDE
FCWDE
Constr. comp. 1 (50.13%)
Constr. comp. 2 (24.74%)
Constr. comp. 1 (58.77%)
Constr. comp. 2 (15.02%)
Constr. comp. 1 (41.15%)
Constr. comp. 2 (28.38%)
Constr. comp. 1 (48.38%)
Constr. comp. 2 (14.59%)
Constr. comp. 1 (51.48%)
Constr. comp. 2 (28.53%)
Constr. comp. 1 (57.78%)
Constr. comp. 2 (16.72%)
Constr. comp. 1 (36.6%)
Constr. comp. 2 (30.37%)
Constr. comp. 1 (28.72%)
Constr. comp. 2 (18.03%)
Constr. comp. 1 (45.97%)
Constr. comp. 2 (26.66%)
Constr. comp. 1 (60.34%)
Constr. comp. 2 (17.9%)
Constr. comp. 1 (58.04%)
Constr. comp. 2 (21.43%)
Constr. comp. 1 (48.67%)
Constr. comp. 2 (20.17%)
p < 0.001
Phylogenetic class
Mortierellomycetes
Agaricomycetes
Pezizomycetes
Leotiomycetes
Sordariomycetes
Eurotiomycetes
Dothideomycetes
Lifestyle
SAP
MyM
EF
OMF
PPF
ECM
ERM
Number of genes
0
10
20
30
Stachybotrys elegans MPI-CAGE-CH-0235
Rhizoctonia solani MPI-SDFR-AT-0096
Colletotrichum higginsianum IMI 349063
Plectosphaerella plurivora MPI-SDFR-AT-0117
Colletotrichum chlorophyti NTL11
Colletotrichum tofieldiae 0861
Plectosphaerella cucumerina MPI-CAGE-AT-0016
Meliniomyces bicolor E
Myceliophthora thermophila
Neurospora crassa OR74A
Sordaria sp. MPI-SDFR-AT-0083
Schizophyllum commune
Aspergillus campestris IBT 28561
Aspergillus taichungensis IBT 19404
Hebeloma cylindrosporum h7
Thielavia terrestris
Pleurotus ostreatus PC15
Trichoderma reesei
Amanita muscaria Koide
Lobosporangium transversale NRRL 3116
Trametes versicolor
Tuber aestivum var. uncinatum
Cenococcum geophilum 1.58
Heterobasidion annosum
Laccaria bicolor
Baudoinia compniacensis UAMH 10762 (4089826)
Rhizoscyphus ericae UAMH 7357
Piloderma croceum F 1598
Tuber magnatum
Scleroderma citrinum Foug A
Paxillus involutus ATCC 200175
Pisolithus microcarpus 441
Armillaria solidipes 28-4
Penicillium subrubescens FBCC1632 / CBS132785
Aspergillus novofumigatus IBT 16806
Aspergillus ochraceoroseus IBT 24754
Alternaria brassicicola
Diplodia seriata DS831
Macrophomina phaseolina MS6
Botrytis cinerea
Sclerotinia sclerotiorum
Morchella importuna CCBAS932
Pseudocercospora (Mycosphaerella) fijiensis
Ascocoryne sarcoides NRRL50072
Gymnopus androsaceus JB14
Meliniomyces variabilis F
Boeremia exigua MPI-SDFR-AT-0100
Corynespora cassiicola CCP
Macrophomina phaseolina MPI-SDFR-AT-0080
Verticillium dahliae VdLs.17
Pyrenochaeta sp. DS3sAY3a
Paraphoma chrysanthemicola MPI-GEGE-AT-0034
Paraphoma chrysanthemicola MPI-SDFR-AT-0120
Phaeosphaeria sp. MPI-PUGE-AT-0046c
Cochliobolus victoriae FI3
Pyrenophora tritici-repentis
Helotiales F229
Leptosphaeria maculans
Stemphylium lycopersici CIDEFI-216
Alternaria alternata SRC1lrK2f
Alternaria alternata MPI-PUGE-AT-0064
Pyrenochaeta sp. DS3sAY3a
Alternaria rosae MPI-PUGE-AT-0040
Cochliobolus sativus ND90Pr
Tulasnella calospora AL13/4D
Botryobasidium botryosum
Mortierella elongata MPI-CAGE-AA-0104
Mortierella elongata AG-77
Microdochium bolleyi J235TASD1
Coprinopsis cinerea
Agaricus bisporus var. burnettii JB137-S8
Sodiomyces alkalinus
Galerina marginata
Trichoderma virens Gv29-8
Daldinia eschscholzii EC12
Trichoderma atroviride
Oidiodendron maius Zn
Phialocephala scopiformis 5WS22E1
Chalara longipes BDJ
Stagonospora nodorum SN15
Periconia macrospinosa DSE2036
Chaetomium sp. MPI-SDFR-AT-0129
Chaetomium globosum
Chaetomium sp. MPI-CAGE-AT-0009
Chaetomium sp. MPI-CAGE-AT-0009
Dendryphion nanum MPI-CAGE-CH-0243
Cylindrocarpon olidum MPI-CAGE-CH-0241
Halenospora varia MPI-CAGE-AT-0135
Piriformospora indica DSM 11827
Sebacina vermifera MAFF 305830
Fusarium oxysporum f. sp. Lycopersici 4287
Rhexocercosporidium sp. MPI-PUGE-AT-0058
Harpophora oryzae
Magnaporthe grisea
Coprinopsis sp. MPI-PUGE-AT-0042
Microdochium trichocladiopsis MPI-CAGE-CH-0230
Auriculariales sp. MPI-PUGE-AT-0066
Dactylonectria macrodidyma MPI-CAGE-AT-0147
Leptodontidium sp. MPI-SDFR-AT-0119
Ilyonectria europaea MPI-CAGE-AT-0026
Ilyonectria sp. MPI-CAGE-AT-0134
Dactylonectria estremocensis MPI-CAGE-AT-0021
Truncatella angustata MPI-SDFR-AT-0073
Fusarium solani MPI-SDFR-AT-0091
Neonectria ditissima R09/05
Fusarium oxysporum MPI-CAGE-CH-0212
Fusarium verticillioides 7600 v2
Fusarium sp. MPI-SDFR-AT-0072
Fusarium oxysporum MPI-SDFR-AT-0094
Verticillium dahliae MPI-CAGE-AT-0001
Cadophora sp. DSE1049
Leotiomycetes sp. MPI-SDFR-AT-0126
Verticillium alfalfae VaMs.102
Fusarium sp. MPI-CAGE-AA-0113
Fusarium tricinctum MPI-SDFR-AT-0068
Fusarium venenatum MPI-CAGE-CH-0201
Fusarium redolens MPI-CAGE-AT-0023
Fusarium tricinctum MPI-SDFR-AT-0044
Fusarium fujikuroi IMI 58289
Fusarium graminearum
S08A
CE5
CE1
CE8
PL3_2
PL1_4
Lifestyle
SAP
MyM
EF
OMF
PPF
ECM
ERM

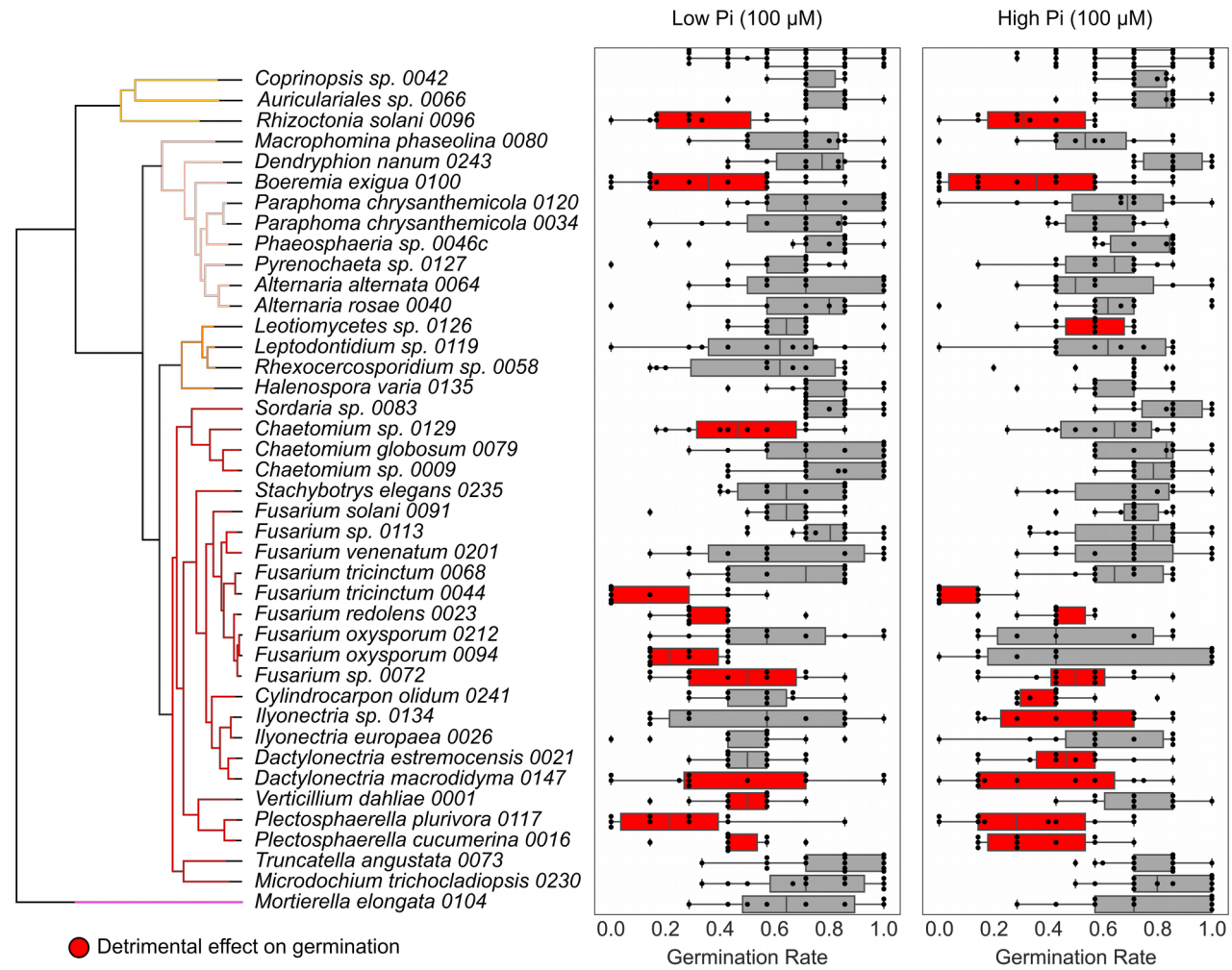
Supplementary figure 6: Differential composition in CAZyme, protease, lipase and SSP repertoires according to fungal phylogeny and lifestyle.
a, Distance-based redundancy analyses (dbRDA) of Jaccard distances calculated on the genomic compositions in subfamilies of CAZymes, proteases, lipases, small secreted proteins (SSPs), plant cell wall degrading enzymes (PCWDEs) and fungal cell-wall degrading enzymes (FCWDEs). The left column shows the result of a dbRDA constrained by phylogeny, while the right one shows the results of dbRDA constrained by lifestyle. Both of these factors significantly explain genomic compositions (dbRDA $P < 0.05$; PERMANOVA *JaccardMatrix~Phylogeny+Lifestyle*, $P < 0.05$ - see **Supplementary Data 3** for details). **b,** Double-clustering heatmap of high-loading genes annotated as CAZymes, proteases, lipases or SSPs, which gene counts best segregate lifestyles. S08A: a subfamily S8A secreted serine proteases from proteinase K subfamily. CE: Carbohydrate esterases. PL: Polysaccharide lyases. Colors indicate the fungal lifestyle. Principal components were calculated on total gene counts. High-loading genes were determined based on the first three principal components.
SAP: Saprotrophs, MyM: *A. thaliana* mycobiota members, EF: Endophytic Fungi, OMF: Orchid Mycorrhizal Fungi, PPF: Plant Pathogenic Fungi, ECM: Ectomycorrhiza, ERM: Ericoid Mycorrhiza.


Mucorom.
Basidiomycota
Ascomycota
Mortierellales
Cantharellales
Boletales
Agaricales
Eurotiales
Pleosporales
Helotiales
Xylariales
Sordariales
Glomerellales
Hypocreales
Lobosporangium transversale NRRL 3116
Mortierella elongata MPI-CAGE-AA-0104
Mortierella elongata AG-77
Piriformospora indica DSM 11827
Sebacina vermifera MAFF 305830
Rhizoctonia solani MPI-SDFR-AT-0096
Botryobasidium botryosum
Tulasnella calospora AL13/4D
Auriculariales sp. MPI-PUGE-AT-0066
Heterobasidion annosum
Trametes versicolor
Piloderma croceum F 1598
Paxillus involutus ATCC 200175
Scleroderma citrinum Foug A
Pisolithus microcarpus 441
Pleurotus ostreatus PC15
Schizophyllum commune
Gymnopus androsaceus JB14
Armillaria solidipes 28-4
Amanita muscaria Koide
Agaricus bisporus var. burnettii JB137-S8
Coprinopsis sp. MPI-PUGE-AT-0042
Coprinopsis cinerea
Laccaria bicolor
Hebeloma cylindrosporum h7
Galerina marginata
Morchella importuna CCBAS932
Tuber aestivum var. uncinatum
Tuber magnatum
Penicillium subrubescens FBCC1632 / CBS132785
Aspergillus ochraceoroseus IBT 24754
Aspergillus novofumigatus IBT 16806
Aspergillus taichungensis IBT 19404
Aspergillus campestris IBT 28561
Pseudocercospora (Mycosphaerella) fijiensis
Baudoinia compniacensis UAMH 10762 (4089826)
Diplodia seriata DS831
Macrophomina phaseolina MPI-SDFR-AT-0080
Macrophomina phaseolina MS6
Cenococcum geophilum 1.58
Dendryphion nanum MPI-CAGE-CH-0243
Corynespora cassiicola CCP
Periconia macrospinosa DSE2036
Boeremia exigua MPI-SDFR-AT-0100
Paraphoma chrysanthemicola MPI-SDFR-AT-0120
Paraphoma chrysanthemicola MPI-GEGE-AT-0034
Phaeosphaeria sp. MPI-PUGE-AT-0046c
Stagonospora nodorum SN15
Pyrenochaeta sp. DS3sAY3a
Pyrenochaeta sp. MPI-SDFR-AT-0127
Leptosphaeria maculans
Pyrenophora tritici-repentis
Stemphylium lycopersici CIDEFI-216
Cochliobolus sativus ND90Pr
Cochliobolus victoriae FI3
Alternaria rosae MPI-PUGE-AT-0040
Alternaria brassicicola
Alternaria alternata MPI-PUGE-AT-0064
Alternaria alternata SRC1lrK2f
Botrytis cinerea
Sclerotinia sclerotiorum
Oidiodendron maius Zn
Halenospora varia MPI-CAGE-AT-0135
Ascocoryne sarcoides NRRL50072
Rhizoscyphus ericae UAMH 7357
Meliniomyces bicolor E
Chalara longipes BDJ
Meliniomyces variabilis F
Phialocephala scopiformis 5WS22E1
Rhexocercosporidium sp. MPI-PUGE-AT-0058
Leptodontidium sp. MPI-SDFR-AT-0119
Helotiales F229
Leotiomycetes sp. MPI-SDFR-AT-0126
Cadophora sp. DSE1049
Truncatella angustata MPI-SDFR-AT-0073
Daldinia eschscholzii EC12
Microdochium bolleyi J235TASD1
Microdochium trichocladiopsis MPI-CAGE-CH-0230
Harpophora oryzae
Magnaporthe grisea
Neurospora crassa OR74A
Sordaria sp. MPI-SDFR-AT-0083
Thielavia terrestris
Myceliophthora thermophila
Chaetomium sp. MPI-SDFR-AT-0129
Chaetomium sp. MPI-CAGE-AT-0009
Chaetomium globosum
Chaetomium globosum MPI-SDFR-AT-0079
Colletotrichum chlorophyti NTL11
Colletotrichum higginsianum IMI 349063
Colletotrichum tofieldiae 0861
Sodiomyces alkalinus
Plectosphaerella cucumerina MPI-CAGE-AT-0016
Plectosphaerella plurivora MPI-SDFR-AT-0117
Verticillium alfalfae VaMs.102
Verticillium dahliae MPI-CAGE-AT-0001
Verticillium dahliae VdLs.17
Stachybotrys elegans MPI-CAGE-CH-0235
Trichoderma atroviride
Trichoderma virens Gv29-8
Trichoderma reesei
Cylindrocarpon olidum MPI-CAGE-CH-0241
Neonectria ditissima R09/05
Ilyonectria sp. MPI-CAGE-AT-0134
Ilyonectria europaea MPI-CAGE-AT-0026
Dactylonectria estremocensis MPI-CAGE-AT-0021
Dactylonectria macrodidyma MPI-CAGE-AT-0147
Fusarium solani MPI-SDFR-AT-0091
Fusarium tricinctum MPI-SDFR-AT-0044
Fusarium tricinctum MPI-SDFR-AT-0068
Fusarium sp. MPI-CAGE-AA-0113
Fusarium venenatum MPI-CAGE-CH-0201
Fusarium graminearum
Fusarium redolens MPI-CAGE-AT-0023
Fusarium fujikuroi IMI 58289
Fusarium verticillioides 7600 v2
Fusarium sp. MPI-SDFR-AT-0072
Fusarium oxysporum f. sp. Lycopersici 4287
Fusarium oxysporum MPI-SDFR-AT-0094
Fusarium oxysporum MPI-CAGE-CH-0212
CAZymes
Lipases
Proteases
Others
SSPs
Ratio
Ratio
PCWDEs
FCWDEs
BMDEs
Storage
LPMOs
Auxiliary Enz.
Hemicellulose
Cellulose
Lignin
Pectin
Chitin
Mannan
Total count
0
150
300
Ratio
CAZymes
Others
Proteases
Lipases
SSPs
Total
MyM
ECM
EF
ERM
OMF
PPF
SAP
BMDEs
FCWDEs
PCWDEs

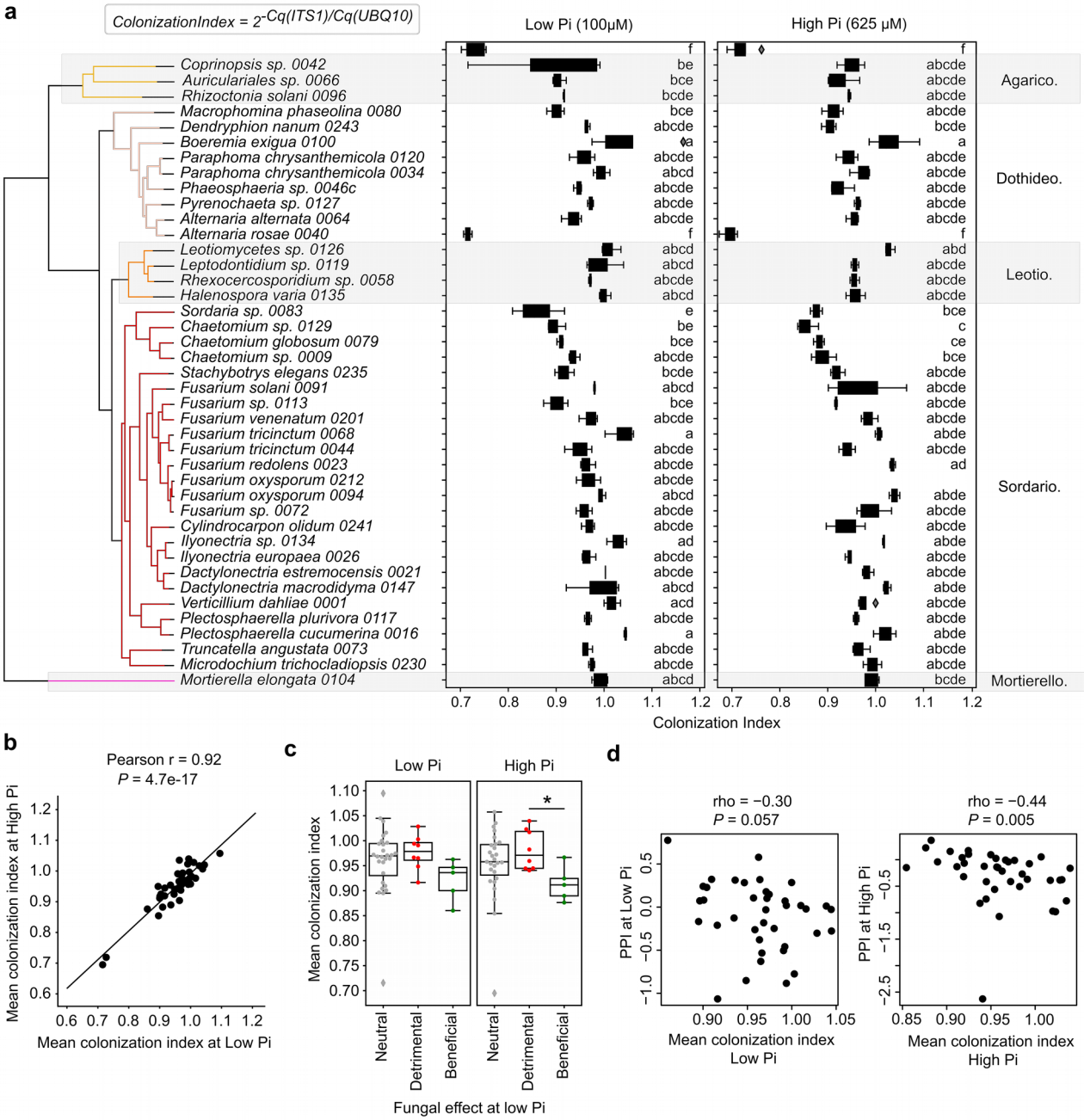
Supplementary figure 7: Descriptions and compositions of predicted fungal secretomes.
The first bubble plot (on the left) shows the number of secreted genes for CAZymes, lipases, proteases, and others (*i.e.*, all secreted proteins not in these first three groups). The group SSPs is a subcategory showing the number of secreted proteins < 300 aa. The size of bubbles corresponds to the number of genes. The fungi are colored according to their ecology. The first bar plots (in the middle) represent the ratio of CAZymes, lipases and proteases, to all secreted proteins (left); and the ratio of SSPs among the entire secretome (right). The second bubble plot (on the right) shows CAZymes grouped according to their functions including plant cell-wall degrading enzymes (PCWDEs) and fungal cell wall degrading enzymes (FCWDEs), peptidoglycans (*i.e.*, bacterial membrane) degrading enzymes (BMDEs), trehalose, starch, glycogen degrading enzymes (Storage), lytic polysaccharide monooxygenase (LPMOs), substrate-specific enzymes for cellulose, hemicellulose, lignin, and pectin (plant cell walls); chitin, glucan, mannan (fungal cell walls). The second bar plots (far right) show the total count of genes including PCWDEs, FCWDEs, and BMDEs (left); and the proportion of PCWDEs, FCWDEs, and BMDEs (right).
MyM: *A. thaliana* mycobiota members, ECM: Ectomycorrhiza, EF: Endophytic Fungi, ERM: Ericoid Mycorrhiza, OMF: Orchid Mycorrhizal Fungi, PPF: Plant Pathogenic Fungi, SAP: Saprotrophs.


a
Phylogeny P = 2.03e-05
Lifestyle P = 1.92e-11
Total secreted proteins
Phylogeny P = 5.95e-07
Lifestyle P = 3.57e-11
Secreted CAZymes
Phylogeny P = 7.4e-04
Lifestyle P = 6.42e-07
Secreted lipases
Phylogeny P = 1.04e-04
Lifestyle P = 5.61e-11
Secreted proteases
ECM ERM OMF SAP PPF EF MyM
b
Secreted
Total
AA
CBM
CE
GH
PL
GT
P = 4.793e-06
P = 1.563e-06
P = 1.167e-09
P = 1.989e-09
P = 3.863e-12
P = 4.845e-11
P = 1.291e-08
P = 1.017e-08
P = 4.374e-10
P = 5.205e-09
P = 1.027e-07

Supplementary figure 8: Genomic counts of secreted CAZymes (and subfamilies), proteases and lipases across fungal lifestyles.
a, Genomic counts across our comparative genomics data set (n = 120) of total secreted proteins, and secreted CAZymes, lipases and proteases. ANOVA-statistical testing (*Counts~PhylogenyPCs+Lifestyle*, **Methods**) identified both phylogeny and lifestyles as having an effect on genomic contents ($P < 0.05$ - see values on figure); letters result from two-sided post-hoc TukeyHSD testing. **b**, Gene counts across our comparative genomics data set (n = 120) of CAZyme families (AA: Auxiliary Activities, CBM: Carbohydrate-Binding Modules, CE: Carbohydrate Esterases, GH: Glycoside Hydrolases, PL: Polysaccharide Lyases), predicted as secreted (extracellular, left) and total (intra and extracellular, right). Statistical testing with a Kruskal-Wallis test (*Counts~Lifestyle*, $P < 0.05$ - see values on figure) identified lifestyle as having an effect on genome contents. Letters result from post-hoc testing with a two-sided Dunn test.
ECM: Ectomycorrhiza, ERM: Ericoid Mycorrhiza, OMF: Orchid Mycorrhizal Fungi, SAP: Saprotrophs, PPF: Plant Pathogenic Fungi, EF: Endophytic Fungi, MyM: *A. thaliana* mycobiota members. Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges).



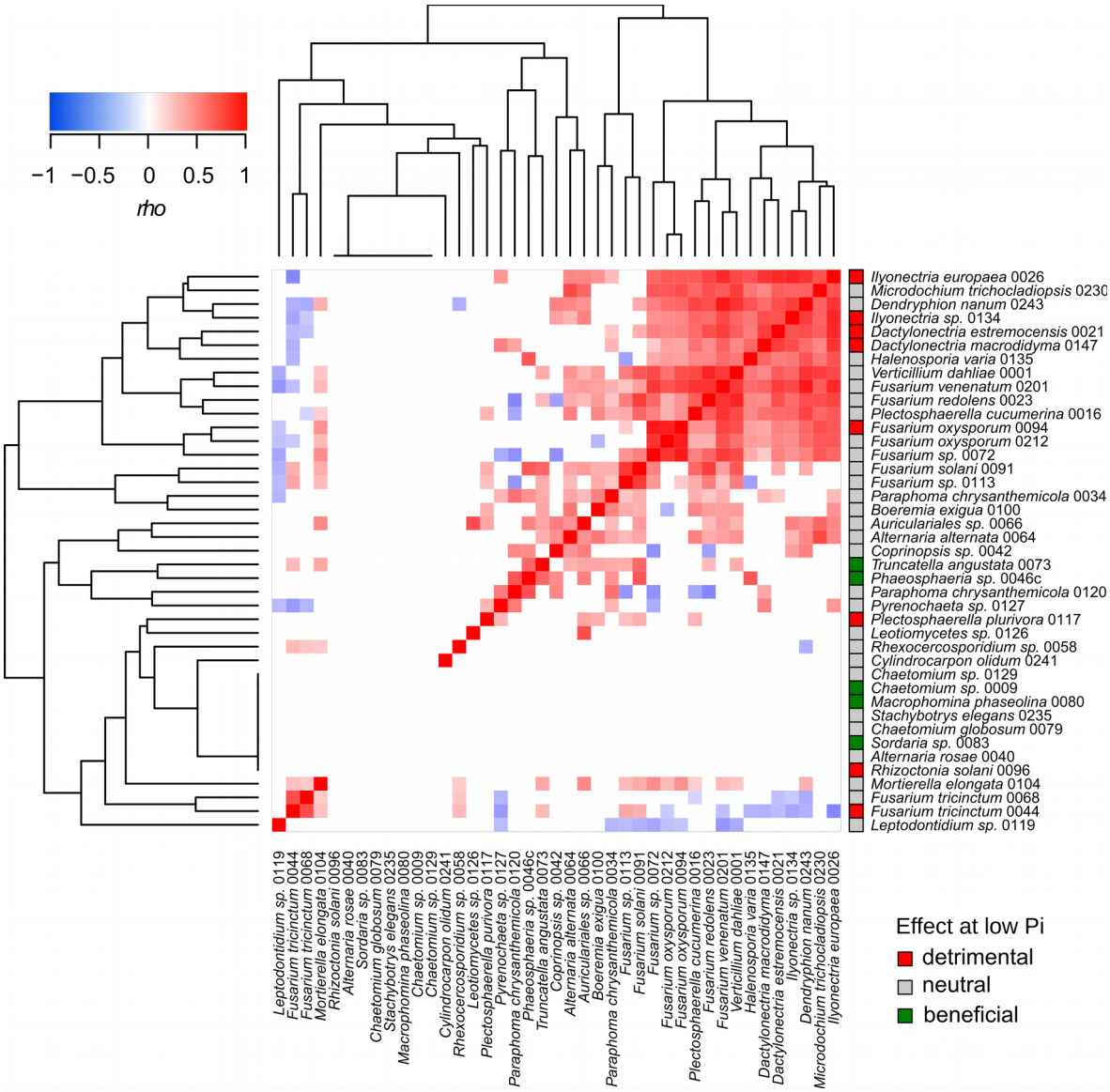
Supplementary figure 9: Fungal effects on *A. thaliana* germination.
Germination rate — proportion of plants that developed over the total number of seeds sowed on each culture plate (n = 6-18) — of *A. thaliana* plants which seeds were inoculated with each of the 41 fungal strains on media containing low and high concentrations of orthophosphate (Pi). Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles). Differential fungal effects on germination rates were tested on both media with Kruskal-Wallis ($P < 10^{-15}$) and beneficial and pathogenic strains were identified by a two-sided Dunn test against mock-treated plants (first row in boxplots).



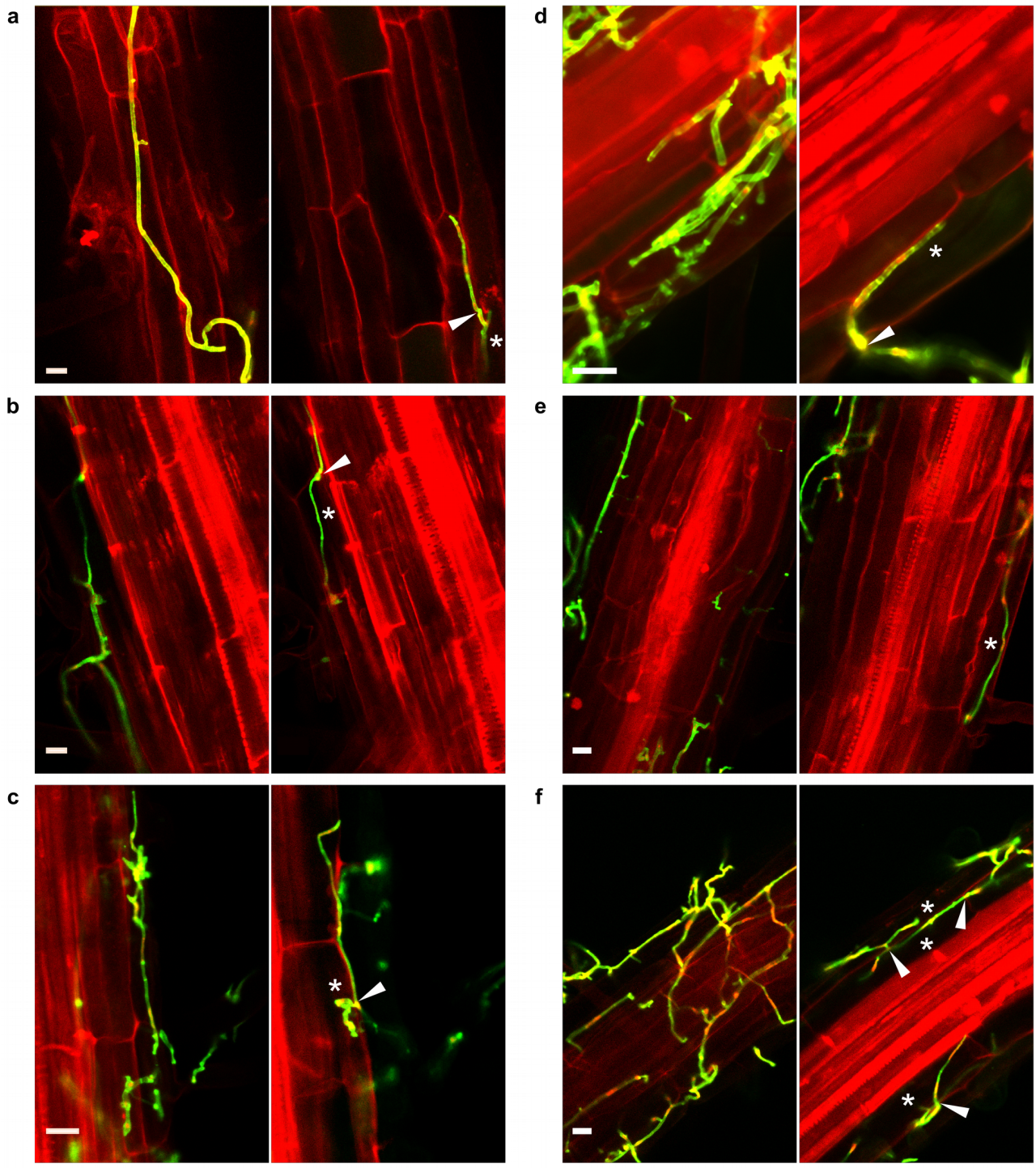
Supplementary figure 10: Fungal colonization of *A. thaliana* roots after 28 days of culture in mono-association.

a, Fungal colonization of plant roots mono-inoculated with different mycobiota members, estimated by quantitative PCR (**Methods**). Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). Statistical difference across treatments was identified by ANOVA (*ColonizationIndex~Treatment, P* < 1e-17), and two-sided post-hoc testing was performed with TukeyHSD. Colonization was measured in n = 3 root samples per condition. **b**, Pearson correlation (r, *P* < 0.05, n = 41) between mean colonization indexes at low and high Pi concentrations. **c**, Differences in colonization indices between neutral (n = 28), beneficial (n = 5), and detrimental (n = 8) fungi at low Pi. Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). Significant differences across fungal groups were identified at high Pi, by ANOVA (*P* = 0.0343) and two-sided TukeyHSD tests. (*: adjusted *P* = 0.0297). No significant difference was identified at low Pi

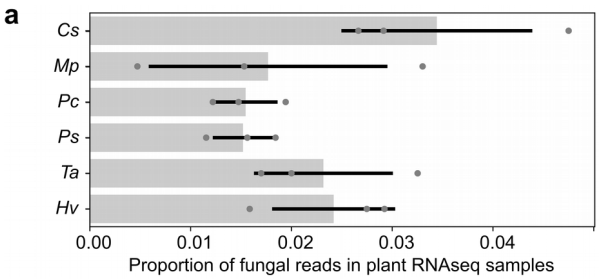
(ANOVA $P = 0.2$) **d**, Correlation between plant performance index and mean colonization index at low Pi (left) and high Pi (right) (n = 41; Spearman's rank correlation rho, $P < 0.05$).

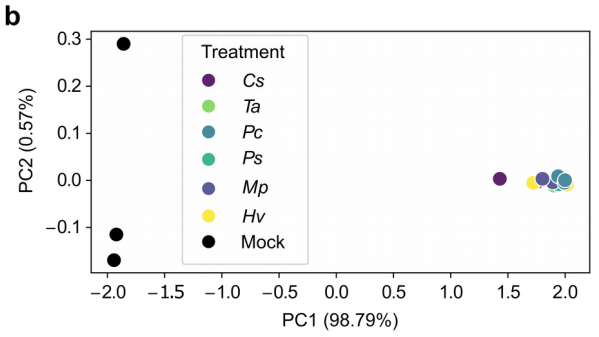


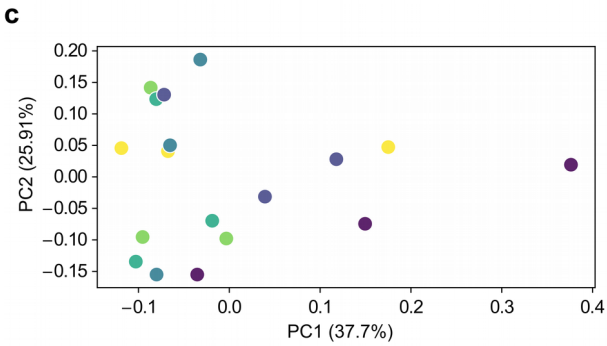
Supplementary figure 11: Correlation matrix comparing the relative abundance profiles of the 41 root mycobiota members in naturally occurring root mycobiomes. Correlation of the 41 fungal taxa relative abundance values across root samples from the European transect data [23]. Spearman's rank correlation (*rho*) was calculated for each fungal pair if these are co-occurring in at least 10 root samples. Only the *rho* values of significant correlations ($P < 0.05$) are plotted. Right to the heatmap, a color-strip indicates if one fungal isolate was identified as having a beneficial, neutral or detrimental effect on plant growth, in mono-association on low Pi agar medium (see Figure 4).

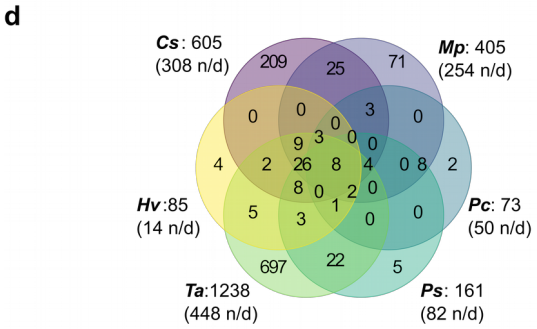


Supplementary figure 12: Confocal imaging of *A. thaliana* root surface and epidermis colonized by six different root mycobiota members.
Roots grown for 4 weeks in mono-association with six diverse fungi, double-stained with Propidium Iodide and Wheat Germ Agglutinin coupled to fluorophore CF®488A (WGA-CF488; Biotium), imaged by confocal microscopy. Left and right picture belong to a single z-stack, respectively focusing on the root surface and the root endosphere where colonization of epidermal cells can be observed. 10µm-scale bars are shown on the left of each panel. Arrows indicate penetration sites and asterisks infected root cells. Similar colonization patterns were observed on 7 different plants in 3 biological replicates. **a,** *Cs = Chaetomium sp. 0009.* **b**, *Mp = Macrophomina phaseolina 0080.* ***c****, Pc= Paraphoma chrysanthemicola 0120.* ***d***, *Ps = Phaeosphaeria sp. 0046c.* **e**, *Ta = Truncatella angustata 0073.* **f**, *Hv = Halenospora varia 0135*.

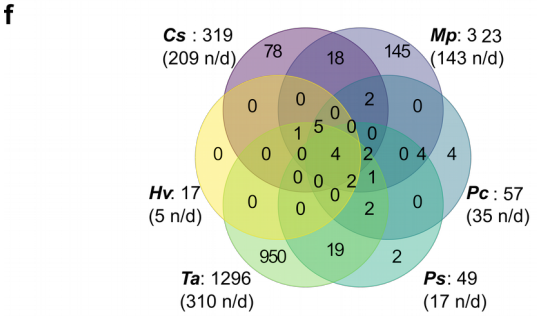




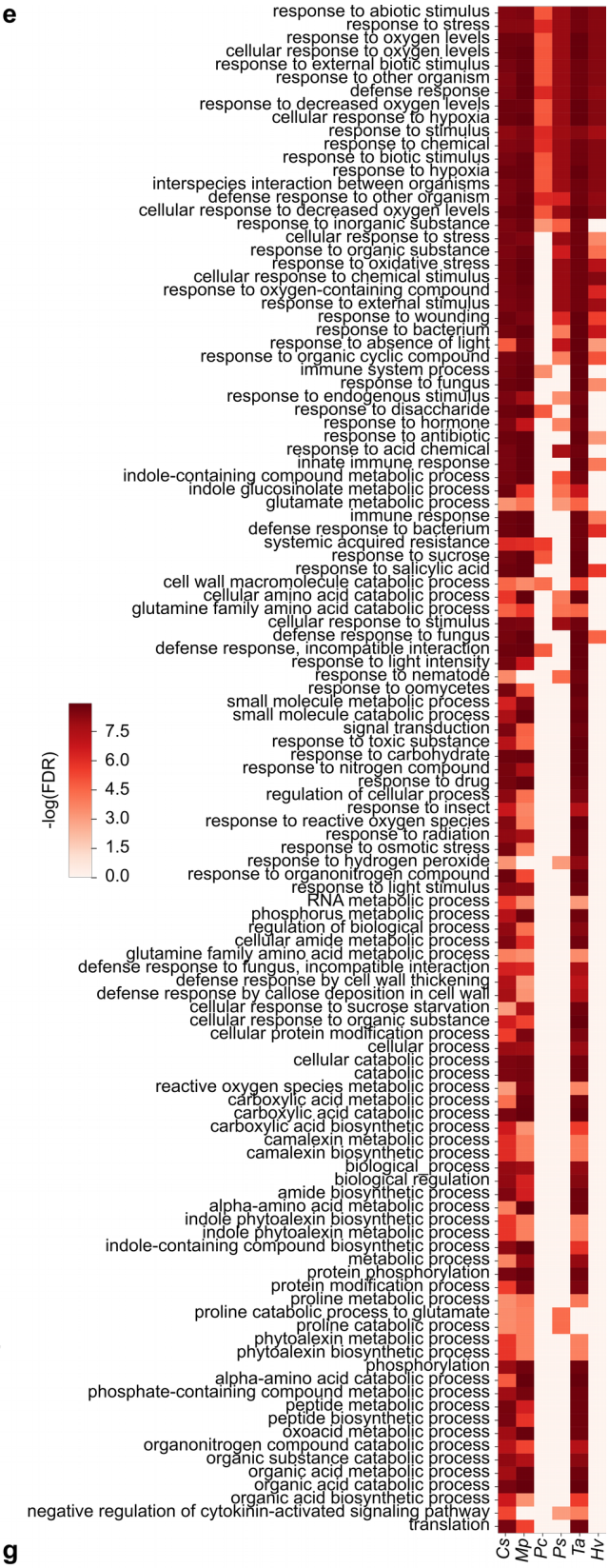


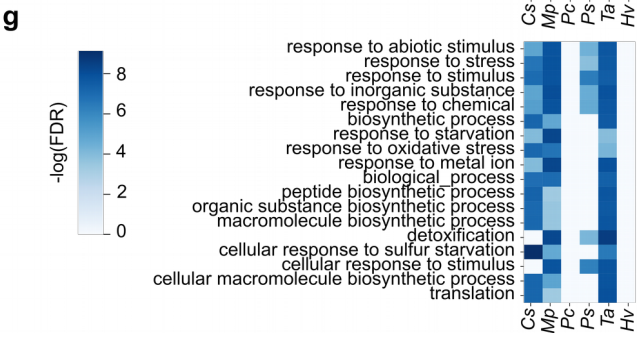


AT2G43610: Chitinase family, **AT1G22890:** STMP2 (Secreted peptide), **AT1G26410:** FAD-binding Berberine family, **AT5G20250:** Raffinose synthase 6, **AT4G14630:** Germin-like protein 9, **AT1G26240:** Extensin 19, **AT1G78290:** SNF1-related protein kinase 2.8, **AT5G63160:** BTB and TAZ domains

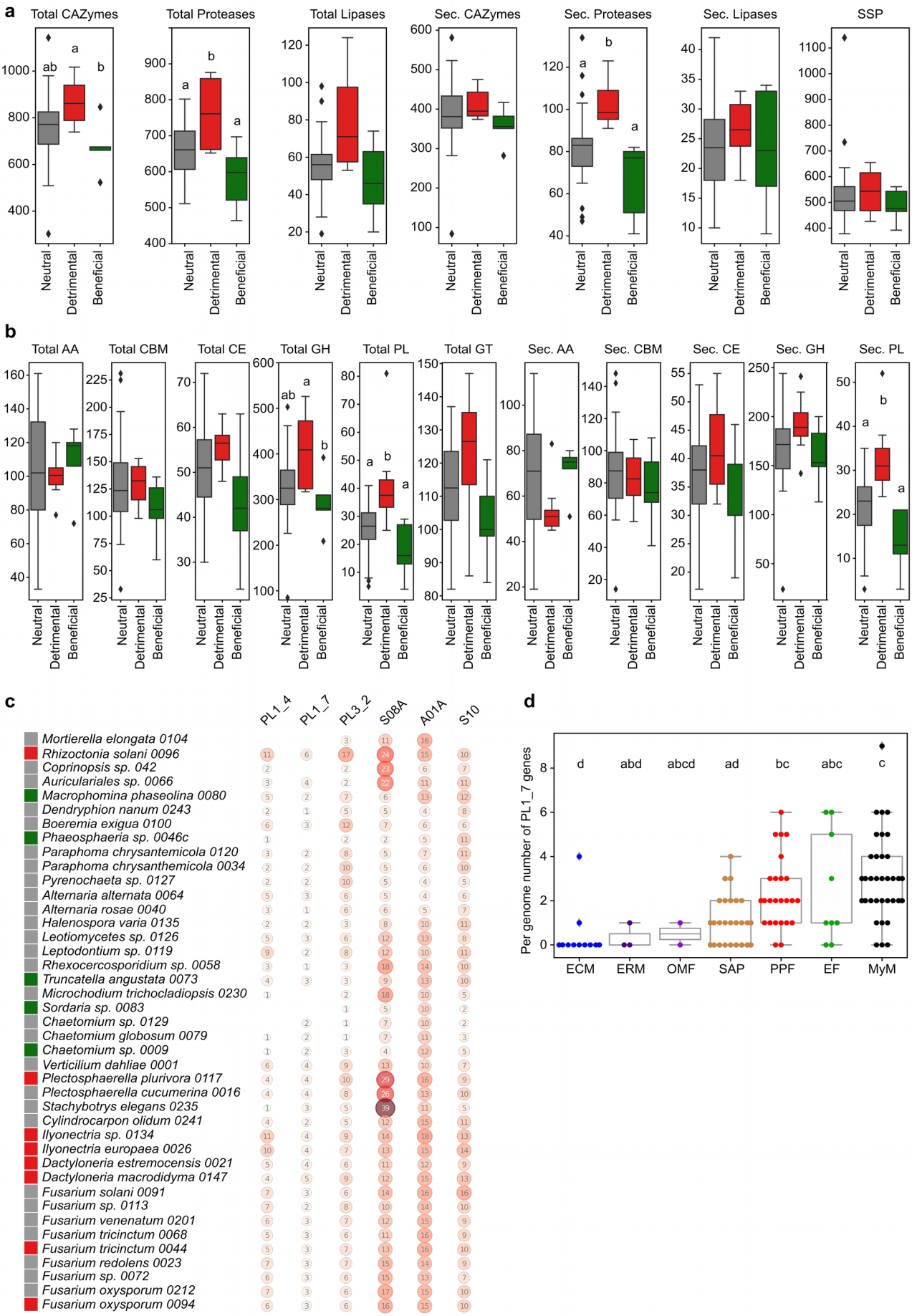


AT1G22150: sulfate transporter Sultr1;3, **AT4G31330:** transmembrane protein DUF599, **AT5G14565:** MIR398C, microRNA targetting CSD and Cytc oxidase family, **AT5G14570:** Nitrate transporter ATNRT2.7

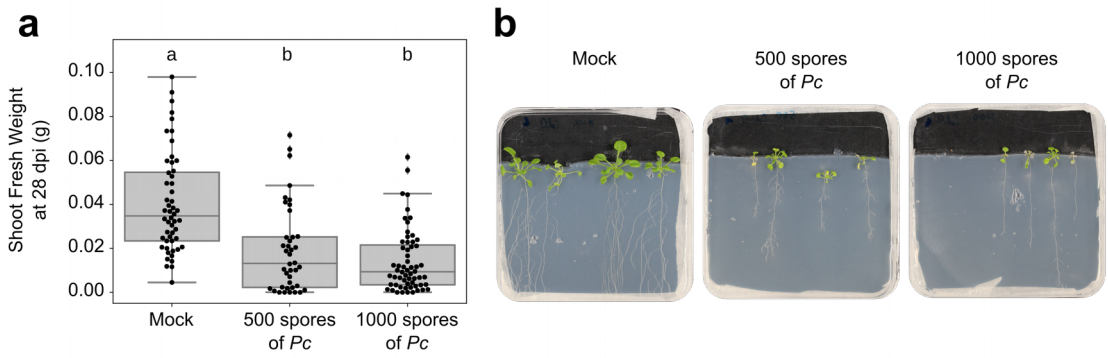




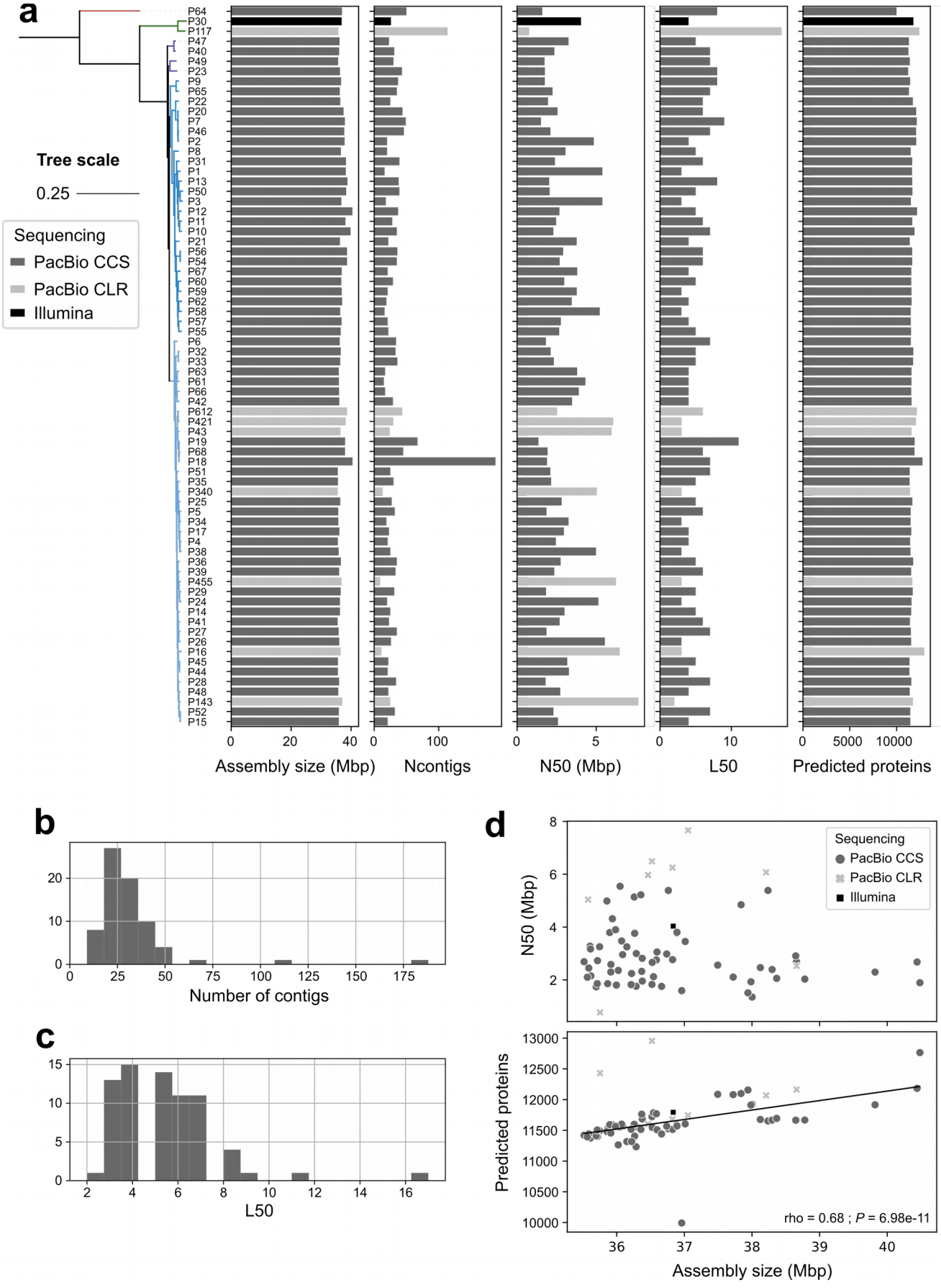
Supplementary figure 13: *A. thaliana* transcriptional reprogramming upon colonization by six different root mycobiota members.
a, Proportion of reads in RNA-Seq samples mapped on fungal genomes. Mean values are shown by the bar plot. Error bars show standard deviation values. n=3 samples per condition, from three independent biological replicates **b**, Principal Component Analysis of Bray-Curtis distances calculated over *A. thaliana* gene read counts. **c**, Principal Component Analysis of Bray-Curtis distances calculated over *A. thaliana* gene read counts, excluding mock-treated samples to reveal sample differences due to the different fungi. *Cs* = *Chaetomium sp. 0009*, *Mp* = *Macrophomina phaseolina* 0080, *Pc* = *Paraphoma chrysantemicola* 0034, *Ps* = *Phaeosphaeria sp.* 0046c, *Ta* = *Truncatella angustata 0073*, *Hv* = *Halenospora varia* 0135. **d**, Venn diagram showing *A. thaliana* commonly over-expressed genes in response to fungal inoculations. Below is the list of genes over-expressed in response to all six fungi. **e**, Independent GO enrichment analyses performed on the *A. thaliana* genes over-expressed in response to each fungus (GOATOOLS [125], FDR < 0.05). **f**, Venn diagram showing *A. thaliana* commonly under-expressed genes in response to fungal inoculations. Below is the list of genes under-expressed in response to all six fungi. **g**, Independent GO enrichment analyses performed on the *A. thaliana* genes under-expressed in response to each fungus (GOATOOLS, FDR < 0.05).


a
Total CAZymes
Total Proteases
Total Lipases
Sec. CAZymes
Sec. Proteases
Sec. Lipases
SSP
Neutral
Detrimental
Beneficial
b
Total AA
Total CBM
Total CE
Total GH
Total PL
Total GT
Sec. AA
Sec. CBM
Sec. CE
Sec. GH
Sec. PL
c
PL1_4
PL1_7
PL3_2
S08A
A01A
S10
Mortierella elongata 0104
Rhizoctonia solani 0096
Coprinopsis sp. 042
Auriculariales sp. 0066
Macrophomina phaseolina 0080
Dendryphion nanum 0243
Boeremia exigua 0100
Phaeosphaeria sp. 0046c
Paraphoma chrysantemicola 0120
Paraphoma chrysanthemicola 0034
Pyrenochaeta sp. 0127
Alternaria alternata 0064
Alternaria rosae 0040
Halenospora varia 0135
Leotiomycetes sp. 0126
Leptodontium sp. 0119
Rhexocercosporidium sp. 0058
Truncatella angustata 0073
Microchodium trichocladiopsis 0230
Sordaria sp. 0083
Chaetomium sp. 0129
Chaetomium globosum 0079
Chaetomium sp. 0009
Verticilium dahliae 0001
Plectosphaerella plurivora 0117
Plectosphaerella cucumerina 0016
Stachybotrys elegans 0235
Cylindrocarpon olidum 0241
Ilyonectria sp. 0134
Ilyonectria europaea 0026
Dactyloneria estremocensis 0021
Dactyloneria macrodidyma 0147
Fusarium solani 0091
Fusarium sp. 0113
Fusarium venenatum 0201
Fusarium tricinctum 0068
Fusarium tricinctum 0044
Fusarium redolens 0023
Fusarium sp. 0072
Fusarium oxysporum 0212
Fusarium oxysporum 0094
d
Per genome number of PL1_7 genes
ECM
ERM
OMF
SAP
PPF
EF
MyM

Supplementary figure 14: Genomic signatures in polysaccharide lyase repertoires explain lifestyle differentiation among root mycobiota members.
a, Distribution of genes encoding secreted and total CAZymes, lipases, proteases and SSPs, in the genomes of the 41 mycobiota members. **b,** Distribution of genes inside each CAZyme family. In **a** and **b**, boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles, and marked by lozenges). The different letters indicate significant difference (FDR < 0.05; Kruskal-Wallis and two-sided Dunn test). Beneficial (n = 5), neutral (n = 26), pathogenic (n = 10). **c,** Key secreted protein coding genes discriminating fungal lifestyles of 41 endophytic fungi. Three fungal effects on plant growth (in mono-association on low Pi medium, see Figure 4) are depicted in different colors (green: beneficial, red: detrimental, grey: neutral). The selected genes coding for secreted polysaccharide lyases (PLs) and proteases discriminate between pathogenic, neutral, and beneficial fungi. Fungal taxa are displayed according to the phylogenetic order. Bubbles with numbers contain the number of genes. **d,** Comparative genomics of the PL1_7 CAZyme subfamily, showing the number of PL1_7 genes in the genomes (n=120) associated to different lifestyles. Boxes are delimited by first and third quartiles, central bars show median values, whiskers extend to show the rest of the distribution, but without covering outlier data points (further than 1.5 interquartile range from the quartiles). Statistics were performed using an ANOVA test (*Counts~Lifestyle*, $P = 1.07e\text{-}06$) and a two-sided TukeyHSD post-hoc test (adjusted $P < 0.05$).



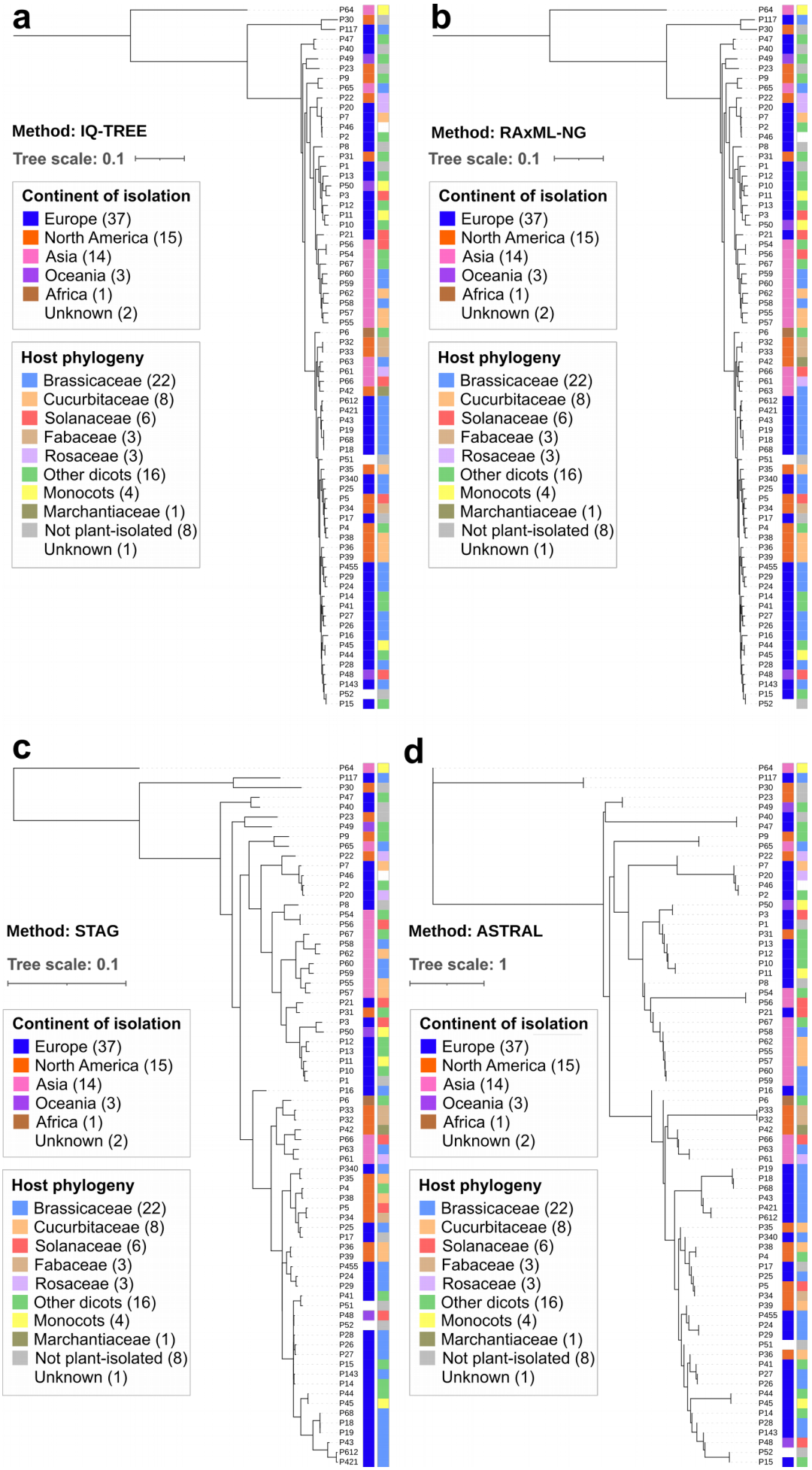
Supplementary figure 15: Effect of a P. cucumerina isolate on plant growth.
a, Shoot fresh weights of *A. thaliana* plants 28 days after root inoculation with *P. cucumerina* (n=48 plants per condition, 2 independent biological replicates). ANOVA test revealed a significant effect of fungal inoculation on plant growth (ANOVA *P*=4.54e-12). Letters on the boxplot highlight significant pairwise differences identified by post-hoc test TukeyHSD. Representative plant phenotypes resulting from this experiment can be observed on panel **b**. Fungal isolate used in this experiment (here labeled *Pc*) corresponds to previously described strain *P. cucumerina* MPI-CAGE-AT-0016 [91], also referred to as P16.



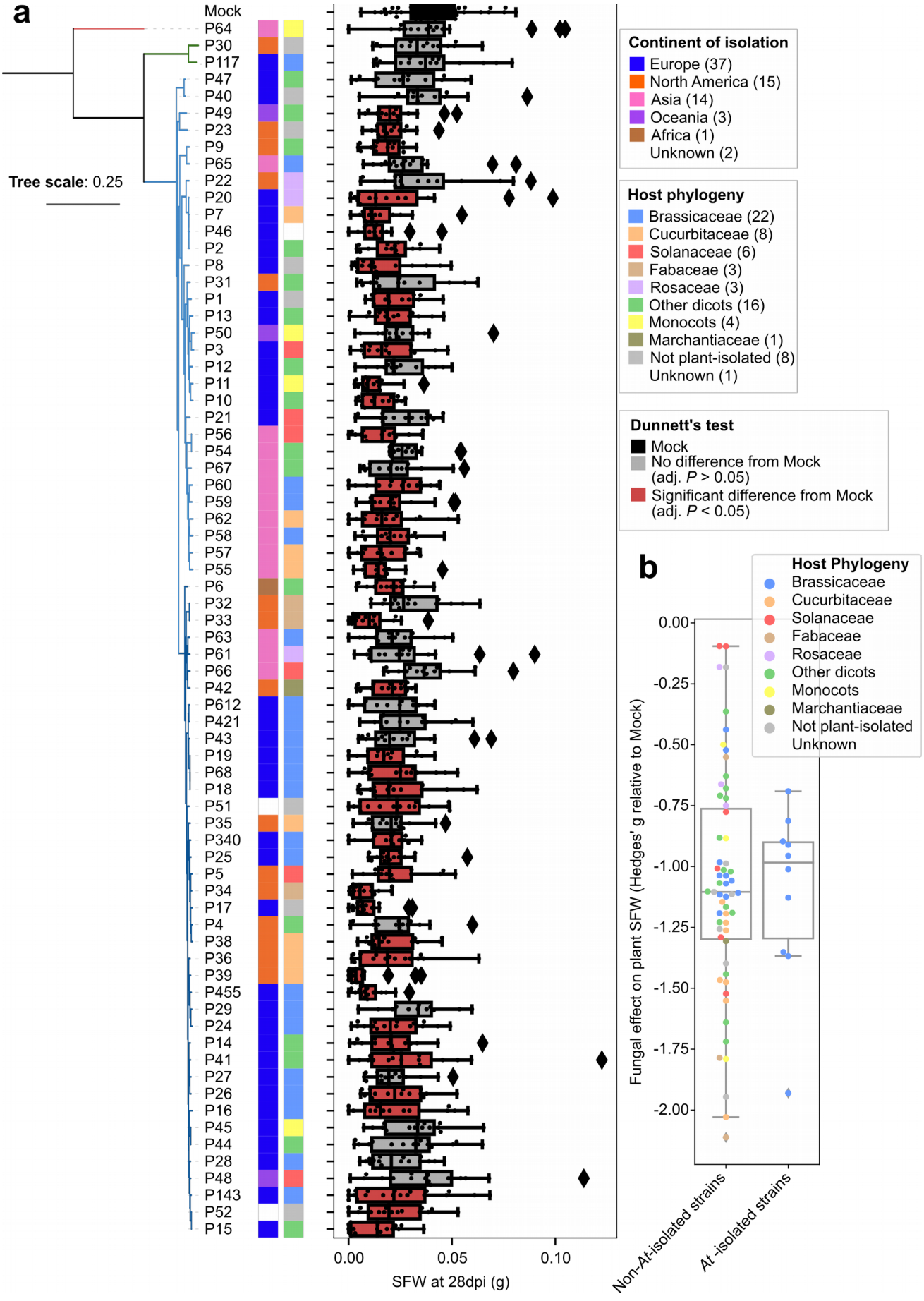
Supplementary figure 16: Statistics reflecting assembly sizes and quality for the 72 *Plectosphaerella* genomes used in this study.

a, For each genome of our genomic data set, the following values are represented: assembly size, number of contigs in the assembly, N50 value (sequence length of the

shortest contig at 50% of the total genome length), L50 (number of longest contigs needed to reach 50% of the genome), number of predicted genes/proteins in the assembly. These values are represented along the collection phylogeny calculated with IQ-TREE [209]. **b,** Distribution of the number of contigs in our genomic data set (n=72, one value per genome assembly). **c**, Distribution of the L50 values in our genomic data set (n=72)**. d**, N50 values (top) and numbers of predicted genes/proteins (bottom), plotted related to assembly sizes. The number of predicted proteins is significantly correlated to the assembly size according to Spearman's rank correlation (see rho and *P* values on graph).

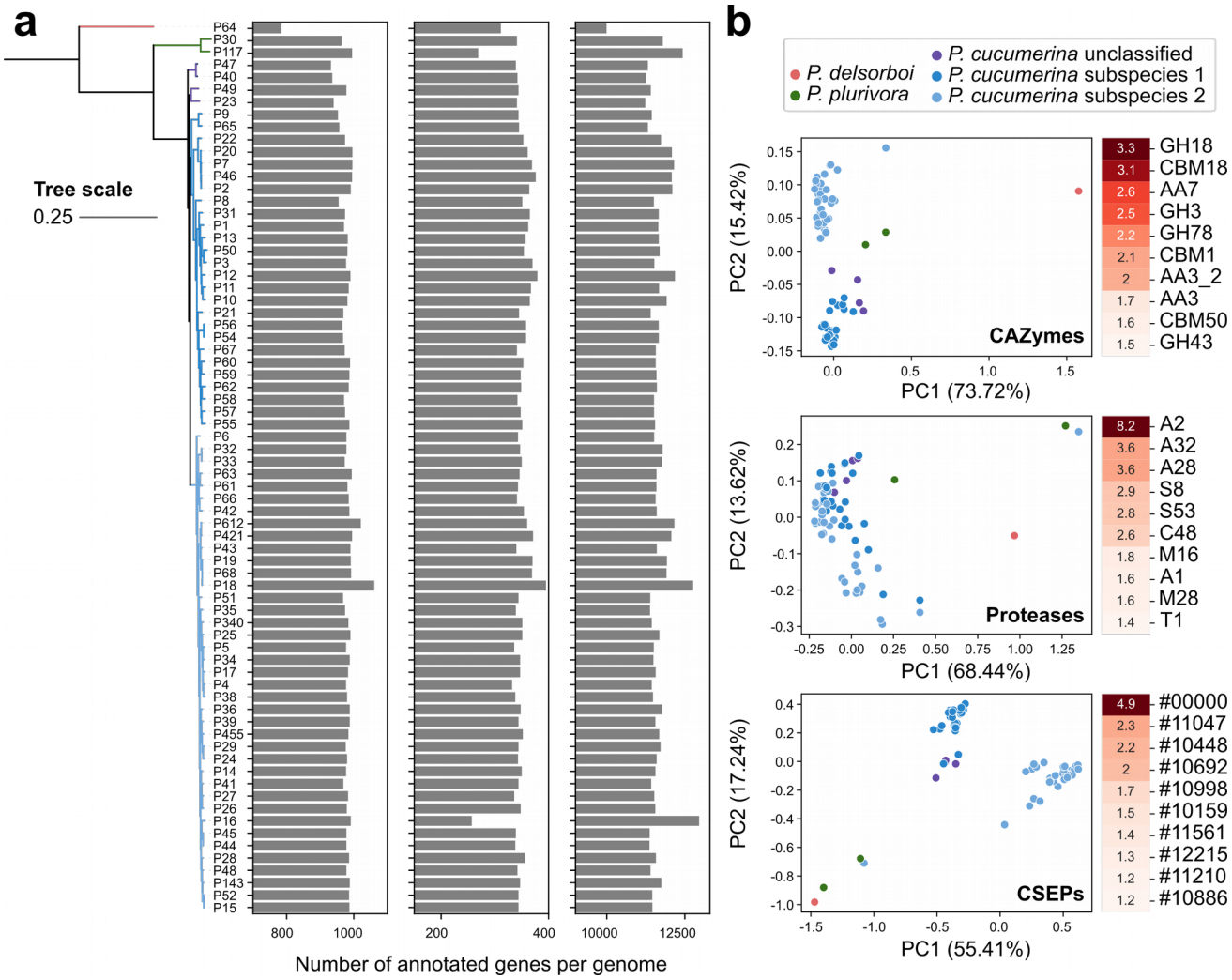

a
Method: IQ-TREE
Tree scale: 0.1
Continent of isolation
Europe (37)
North America (15)
Asia (14)
Oceania (3)
Africa (1)
Unknown (2)
Host phylogeny
Brassicaceae (22)
Cucurbitaceae (8)
Solanaceae (6)
Fabaceae (3)
Rosaceae (3)
Other dicots (16)
Monocots (4)
Marchantiaceae (1)
Not plant-isolated (8)
Unknown (1)
b
Method: RAxML-NG
Tree scale: 0.1
Continent of isolation
Europe (37)
North America (15)
Asia (14)
Oceania (3)
Africa (1)
Unknown (2)
Host phylogeny
Brassicaceae (22)
Cucurbitaceae (8)
Solanaceae (6)
Fabaceae (3)
Rosaceae (3)
Other dicots (16)
Monocots (4)
Marchantiaceae (1)
Not plant-isolated (8)
Unknown (1)
c
Method: STAG
Tree scale: 0.1
Continent of isolation
Europe (37)
North America (15)
Asia (14)
Oceania (3)
Africa (1)
Unknown (2)
Host phylogeny
Brassicaceae (22)
Cucurbitaceae (8)
Solanaceae (6)
Fabaceae (3)
Rosaceae (3)
Other dicots (16)
Monocots (4)
Marchantiaceae (1)
Not plant-isolated (8)
Unknown (1)
d
Method: ASTRAL
Tree scale: 1
Continent of isolation
Europe (37)
North America (15)
Asia (14)
Oceania (3)
Africa (1)
Unknown (2)
Host phylogeny
Brassicaceae (22)
Cucurbitaceae (8)
Solanaceae (6)
Fabaceae (3)
Rosaceae (3)
Other dicots (16)
Monocots (4)
Marchantiaceae (1)
Not plant-isolated (8)
Unknown (1)

Supplementary figure 17: Phylogeny of the 72 *Plectosphaerella* strains, reconstructed with four different methods.
a, Phylogenetic tree reconstructed with IQ-TREE [209] (model *JTT+F+I+G4*, as defined by ModelFinder [210]) from the trimmed concatenated alignments of 5,466 single copy orthologues identified by orthology prediction with OrthoFinder [119]. **b**, Phylogenetic tree reconstructed with RAxML-NG [211] (model *JTT+F+I+G4*, as defined by ModelFinder) from the trimmed concatenated alignments of 5,466 single copy orthologues identified by orthology prediction with OrthoFinder. **c**, Coalescent phylogenetic tree reconstructed from total gene family trees generated by OrthoFinder, with implemented method STAG [213]. **d**, Coalescent phylogenetic tree reconstructed with ASTRAL [214] from the OrthoFinder-reconstructed gene family trees of 5,466 single copy orthologues. See Methods for further details.

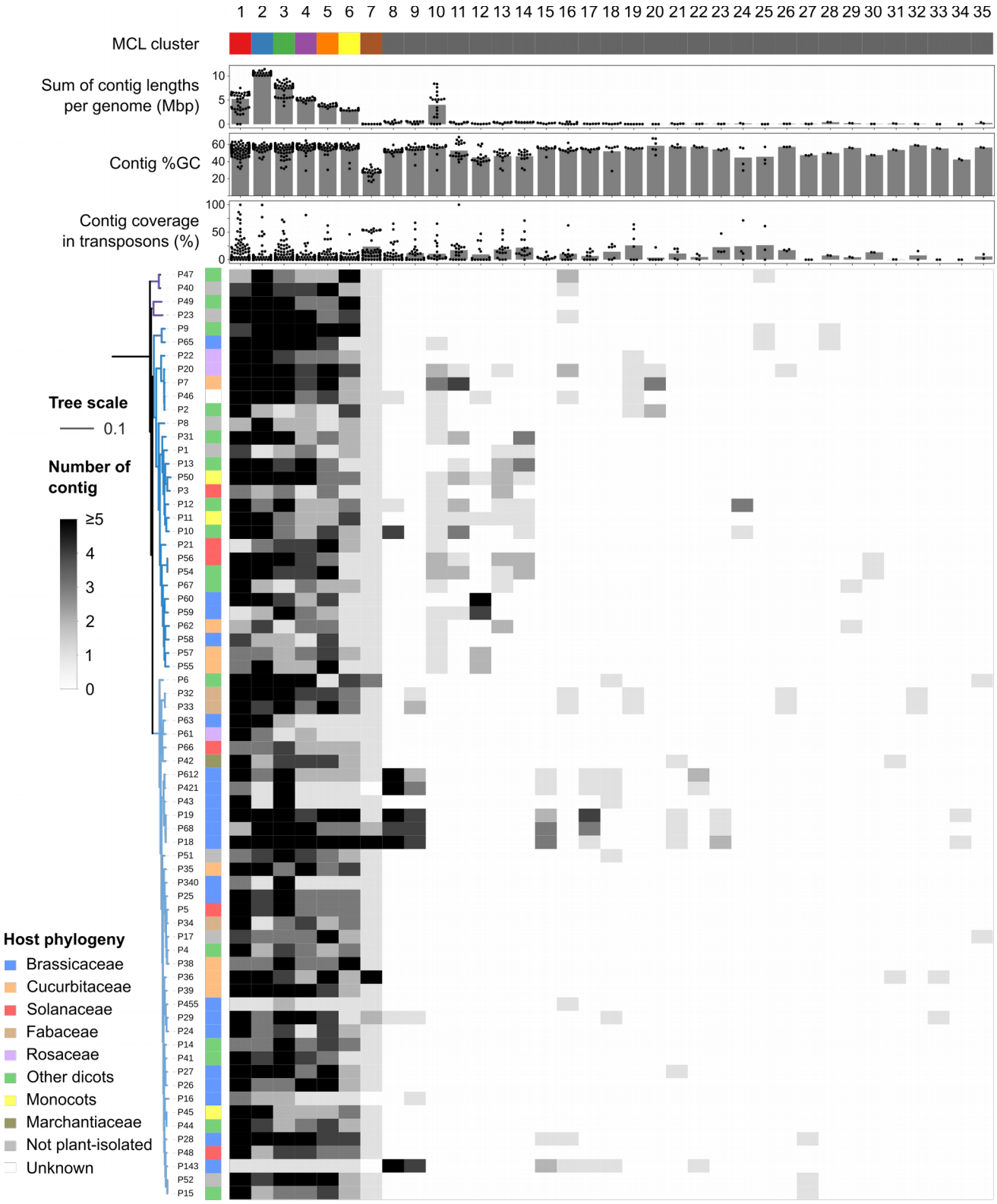


Supplementary figure 18: Fungal effects of each *Plectosphaerella* strain on *A. thaliana* performance in binary interactions.

a, Shoot fresh weight (SFW) of plants which roots were inoculated with 500 spores of each strain, after 28 days in culture in an agar-based gnotobiotic system (for each fungus: 4 independent biological replicates, 4 plants per replicate). These weight values reflect individual fungal effects on plants. Boxes colored in red reveal a significantly different between the fungal inoculum and mock-treated plants, according to ANOVA and Dunnett's test ($P<0.05$). Boxes are represented along the strain phylogeny, reconstructed with IQ-TREE [209]. **b**, Standard effect sizes (Hedges' g) calculated against mock reference and representing fungal effect on *A. thaliana* growth are plotted and show no difference between *A. thaliana* isolated strains and other *P. cucumerina* isolates (Student t-test: $P>0.05$).



Supplementary figure 19: Genome compositions in CAZymes, proteases and CSEPs. **a,** Number of carbohydrate-active enzyme- (CAZyme), protease- and candidate secreted effector protein (CSEP)-encoding genes annotated in each genome. These values are represented along the phylogeny of the data set, reconstructed with IQ-TREE [209]. **b**, Principal component analyses (PCA) calculated on pairwise Jaccard distances between genomes reflecting differences of CAZyme, protease and CSEP repertoires. On the right of each PCA plot are shown the ten most variable families in our data set, with their variance values. While we used annotated CAZy and MEROPS [152] families for CAZymes and proteases, the identifiers of CSEPs reflect orthogroups defined by OrthoFinder [119].



Supplementary figure 20: Descriptive statistics of each genomic compartment and representation across the *Plectosphaerella* phylogeny.
Each column in the heatmap corresponds to a genomic compartment, as defined by MCL clustering. For each one are presented the cumulated size of contigs per genome (top bar-chart), the percentage in guanine and cytosine (%GC) of each contig (second bar-chart), the sequence coverage in transposons as annotated by reasonaTE [206] (third bar-chart), and the number of contigs in the cluster for each *Plectosphaerella* strain, organized along strain phylogeny reconstructed with IQ-TREE [209].

Identifier	Source institution/company/culture collection	Collaborator/contact person	Strain name	Host	Isolation site	Related publication (if any)	Other name
P1	Parthenope University of Naples	Stefano Dumontet	8-10	damp tip of a calthemite straw stalactite	Naples, Italy	https://doi.org/10.1016/j.ibiod.2019.03.005	
P2	Universite Clermont Auvergne	Isabelle Batisson	AR1	submerged Alnus leaf	France	https://doi.org/10.3389/fmicb.2018.03167	
P3	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 286.64	Hair root of Nicotiana tabacum	Heverlee, Belgium		
P4	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 101958	Endophyte in leaves and stems of Galium	Alberta, Canada		
P5	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 400.58	Lycopersicon esculentum	Canada		
P6	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 367.73	Viola odorata	Egypt		
P7	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 131739	Collar of Cucumis melo	Foggia, Italy 41.450000 N 15.533330 E		
P8	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 137.37	Paper	Pisa, Italy		
P9	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146211	Carica papaya	Mexico		CPC 15979
P10	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146213	Begonia sp.	Netherlands		CPC 34755
P11	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146215	Carex	Netherlands		CPC 34096
P12	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146214	Hebe sp.	Netherlands		CPC 35056
P13	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146216	Hebe sp.	Netherlands		CPC 35054
P14	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146210	Petroselinum crispum	Netherlands		CPC 35048
P15	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 355.36	Root of Viola tricolor	Netherlands		
P16	MPIPZ	Stephane Hacquard	MPI-CAGE-AT-0016	Arabidopsis thaliana (roots)	Cologne, Germany	https://doi.org/10.1038/s41467-021-27479-y	
P17	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 567.78	Unknown fungus	USSR		
P18	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 632.94	Arabidopsis sp. (leaves)	Switzerland		
P19	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 101014	Arabidopsis thaliana (leaves)	Switzerland		
P20	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 619.74	Leaf of Pyrus malus	Basel, Switzerland		
P21	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 137.33	Nicotiana tabacum	Bristol, England		
P22	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 146212	Rubus phoenicolasius	USA		CPC 25081
P23	Westerdijk Fungal Biodiversity Institute	Pedro Crous	CBS 139.60	Unknown	USA		
P24	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P1868	Microthlaspi perfoliatum	Bulgaria 42.57 N 22.69 E	https://doi.org/10.1111/1462-2920.13112	DSM 110859
P25	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P6208	Brassica napus	France 48.80 N 7.18 E	https://doi.org/10.1080/17550874.2018.1504332	DSM 110857
P26	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P2062	Microthlaspi perfoliatum	France 47.20 N 5.43 E	https://doi.org/10.1111/1462-2920.13112	DSM 110858
P27	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P6215	Brassica napus	Germany 49.52 N 9.40 E	https://doi.org/10.1080/17550874.2018.1504332	DSM 110848
P28	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P1943	Microthlaspi perfoliatum	Germany 50.37 N 7.22 E	https://doi.org/10.1111/1462-2920.13112	DSM 110850
P29	Goethe University Frankfurt / Wageningen University and Research	Jose Macia-Vicente	P2540	Microthlaspi perfoliatum	Greece 38.91 N 21.83 E	https://doi.org/10.1111/1462-2920.13112	DSM 110833
P30	University of Minnesota	Cara Santelli	DS2psM2a2	AMD passive treatment system	USA		
P31	The Connecticut Agricultural Experiment Station	Wade Elmer	PcCAES1	Heuchera sanguine	New England, USA	https://doi.org/10.1111/jph.12867	
P32	Southern Illinois University Carbondale	Ahmad Fakhoury	MNSO1_2_1	Glycine max	Minnesota, USA		
P33	Southern Illinois University Carbondale	Ahmad Fakhoury	W_ILSO2_5_4	Glycine max	Illinois, USA		
P34	Southern Illinois University Carbondale	Ahmad Fakhoury	W_MNSO2_5_25	Glycine max	Minnesota, USA		
P35	Hermiston Agricultural Research & Extension Center, Oregon State University	Hannah Rivedal	J55	Cucurbita moshata	Oregon, USA 44.529483 N -123.370432 E	https://doi.org/10.1094/PBIOMES-11-18-0056-R	
P36	Hermiston Agricultural Research & Extension Center, Oregon State University	Hannah Rivedal	I652	Cucurbita maxima	Oregon, USA 45.050674 N -122.803997 E	https://doi.org/10.1094/PBIOMES-11-18-0056-R	
P38	Hermiston Agricultural Research & Extension Center, Oregon State University	Hannah Rivedal	K107	Cucurbita pepo	Oregon, USA 44.477608 N -122.781642 E	https://doi.org/10.1094/PBIOMES-11-18-0056-R	
P39	Hermiston Agricultural Research & Extension Center, Oregon State University	Hannah Rivedal	M118	Cucurbita maxima	Oregon, USA 45.23516 N -122.916037 E	https://doi.org/10.1094/PBIOMES-11-18-0056-R	
P40	Institute of Biochemistry and Biophysisc Polish Academy of Sciences	Joanna Kruszewska	IBBPKBobr2	soil	Kampinos Forest, Poland	https://doi.org/10.2166/wh.2020.096	
P41	Pavol Jozef Šafárik University in Košice	Jana Henzelyova	2SF1-3	Hypericum maculatum	Botanical Garden UPJŠ, Slovakia	https://doi.org/10.1055/a-1130-4703	
P42	Maastricht University	Jessica Nelson	JN0335	Marchantia sp.	Tenessee University Bryophytes garden 35.95713, -83.92493	https://doi.org/10.1111/nph.15012	
P43	MPIPZ	Stephane Hacquard	MPI-PUGE-AT-0043	Arabidopsis thaliana (roots)	Pulheim, Germany		
P44	CABI	Helen Stewart	IMI312228	Dancus carota	Denmark		
P45	CABI	Helen Stewart	IMI312232	Hordeum	Denmark		
P46	CABI	Helen Stewart	IMI358508		United Kingdom		
P47	CABI	Helen Stewart	IMI506094b	Senecio vulgaris	United Kingdom		
P48	ICMP – Landcare Research	Bevan Weir	ICMP16464	Solanum tuberosum	Mid Canterbury, New Zealand		
P49	ICMP – Landcare Research	Bevan Weir	ICMP15226	Papaver nudicaule	Auckland, New Zealand		
P50	ICMP – Landcare Research	Bevan Weir	ICMP10923	Anthurium	Fiji Islands		
P51	DSMZ		DSM3048	soil	unknown		
P52	Universidad Politécnica de Madrid	Soledad Sacristan	DSM2127	wheat bulb fly	unknown	https://doi.org/10.1094/MPMI-03-20-0057-R	
P54	NARO Genbank	Shoshi Kikuchi	MAFF238632	Persian buttercup	Chiba, Japan		
P55	NARO Genbank	Shoshi Kikuchi	MAFF306720	Cucumber	Chiba, Japan		
P56	NARO Genbank	Shoshi Kikuchi	MAFF238634	Bell pepper	Ehime, Japan		
P57	NARO Genbank	Shoshi Kikuchi	MAFF238969	Oriental pickling melon	Ibaraki, Japan		
P58	NARO Genbank	Shoshi Kikuchi	MAFF241235	Turnip rape	Ibaraki, Japan		
P59	NARO Genbank	Shoshi Kikuchi	MAFF238964	Radish	Kanagawa, Japan		
P60	NARO Genbank	Shoshi Kikuchi	MAFF240260	Radish	Kanagawa, Japan		
P61	NARO Genbank	Shoshi Kikuchi	MAFF241233	Common plum	Kanagawa, Japan		
P62	NARO Genbank	Shoshi Kikuchi	MAFF238968	Cucumber	Mie, Japan		
P63	NARO Genbank	Shoshi Kikuchi	MAFF238966	Radish	Miyazaki, Japan		
P64	NARO Genbank	Shoshi Kikuchi	MAFF238802	Curcuma	Okinawa, Japan		
P65	NARO Genbank	Shoshi Kikuchi	MAFF240789	Turnip	Okinawa, Japan		
P66	NARO Genbank	Shoshi Kikuchi	MAFF243382	Potato	Okinawa, Japan		
P67	NARO Genbank	Shoshi Kikuchi	MAFF244724	Bergamot mint, orange mint	Tokyo, Japan		
P68	Universidad Politécnica de Madrid	Soledad Sacristan	PcBMM	Arabidopsis thaliana (leaves)	Switzerland	https://doi.org/10.1094/MPMI-03-20-0057-R	
P117	MPIPZ	Stephane Hacquard	MPI-SDFR-AT-0117	Arabidopsis thaliana (roots)	Saint-Die, France	https://doi.org/10.1038/s41467-021-27479-y	
P143	MPIPZ	Stephane Hacquard	MPI-CAGE-AT-0143	Arabidopsis thaliana (roots)	Cologne, Germany		
P340	MPIPZ	Stephane Hacquard	MPI-FR1-AT-0340	Arabidopsis thaliana (roots)	Toulouse, France		
P421	MPIPZ	Stephane Hacquard	MPI-SP2-AT-0421	Arabidopsis thaliana (roots)	Spain		
P455	MPIPZ	Stephane Hacquard	MPI-SP2-AT-0455	Arabidopsis thaliana (roots)	Spain		
P612	MPIPZ	Stephane Hacquard	MPI-SW4-AT-0612	Arabidopsis thaliana (roots)	Sweden		

Supplementary table: Description of the *Plectosphaerella* strains (n=72) used in in our comparative study.

Erklärung zur Dissertation
gemäß der Promotionsordnung vom 12. März 2020

Diese Erklärung muss in der Dissertation enthalten sein.
(*This version must be included in the doctoral thesis*)

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht."

Teilpublikationen:

Mesny, F., Miyauchi, S., Thiergart, T., Pickel, B., Atanasova, L., Karlsson, M., ... & Hacquard, S. (2021). Genetic determinants of endophytism in the Arabidopsis root mycobiome. Nature communications, 12(1), 1-15.

Datum, Name und Unterschrift

24.08.2022, Fantin Mesny

Erklärung zum Gesuch um Zulassung zur Promotion

gemäß der Promotionsordnung vom 12. März 2020

1. Zugänglichkeit von Daten und Materialien

Die Dissertation beinhaltet die Gewinnung von Primärdaten oder die Analyse solcher Daten oder die Reproduzierbarkeit der in der Dissertation dargestellten Ergebnisse setzt die Verfügbarkeit von Datenanalysen, Versuchsprotokollen oder Probenmaterial voraus.

☐ Trifft nicht zu

☒ Trifft zu.
In der Dissertation ist dargelegt wie diese Daten und Materialien gesichert und zugänglich sind (entsprechend den Vorgaben des Fachgebiets beziehungsweise der Betreuerin oder des Betreuers).

2. Frühere Promotionsverfahren

Ich habe bereits einen Doktortitel erworben oder ehrenhalber verliehen bekommen.

Oder: Für mich ist an einer anderen Fakultät oder Universität ein Promotionsverfahren eröffnet worden, aber noch nicht abgeschlossen.

Oder: Ich bin in einem Promotionsverfahren gescheitert.

☒ Trifft nicht zu

☐ Zutreffend
Erläuterung:

3. Straftat

Ich bin nicht zu einer vorsätzlichen Straftat verurteilt worden, bei deren Vorbereitung oder Begehung der Status einer Doktorandin oder eines Doktoranden missbraucht wurde.

Ich versichere, alle Angaben wahrheitsgemäß gemacht zu haben.

Datum	Name	Unterschrift
24.08.2022	Fantin Mesny