UNIVERSITÀ DEGLI STUDI DI TORINO

UNIVERSITÉ DE LYON

Lyon 1

Università degli Studi di Torino
Scuola di Dottorato in Scienze della Natura
e Tecnologie Innovative

Université Claude Bernard Lyon1
Ecole Doctorale (N°341): Evolution,
Ecosystèmes, Microbiologie et Modélisation

N°d'ordre NNT : 2018LYSE1088

DOCTORAT EN BIOLOGIE SPÉCIALITÉ MICROORGANISMES, INTERACTIONS, INFECTIONS
Discipline: "Ecologie Microbienne"

TESI DI DOTTORATO DI RICERCA IN SCIENZE BIOLOGICHE E BIOTECNOLOGIE APPLICATE
XXX CICLO

31st May 2018

Martino Adamo

Diversity of fungal DyP-type peroxidases and their potential contribution to organic matter degradation

Members of the Thesis Board:

Franzetti, Andrea	Professeur Associé	Università di Milano-Bicocca, DISAT	Rapporteur
Morel-Rouhier,Mélanie	Professeure	Université del Lorranne, INRA, Nancy	Rapporteure
Cocolin, Luca Simone	Professeur	Università degli Studi di Torino, DISAFA	Examinateur
Rosso, Marie-Noëlle	Directrice de Recherche	Université Aix Marseille, INRA, Marseille	Examinateurice
Moënne-Loccoz, Yvan	Professeur	Univesité de Lyon, CNRS, Villeurbanne	Examinateur
Corno, Gianluca	Chercheur	Institute of Ecosystem Study, Verbania	Examinateur
Marmeisse Roland	Directeur de Recherche	Univesité de Lyon, CNRS, Villeurbanne	Directeur de thèse
Girlanda Mariangela	Professeure associée	Università degli Studi di Torino, DBIOS	Directeurice de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	M. le Professeur Frédéric FLEURY
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directrice Générale des Services	M.me Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G.RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : M. la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : M. la Professeure VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : M. la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. Professeur MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Research partners

UNIVERSITÀ
DEGLI STUDI
DI TORINO

Financial partners

MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Tutto questo lavoro è dedicato a Kiki.

«Cuando despertó, el dinosaurio todavía estaba allí»
Augusto Monterroso

Index

Chapter 4 - Metabarcoding on both environmental DNA and RNA better resolves differences between fungal communities.

This page was intentionally left blank

Abstract

Lignin is one of the most abundant natural polymers on earth whose resistance to degradation contributes to the mechanical strength of plant cell walls and protects plant cells from pathogen attack. Few organisms are can degrade lignin efficiently; most of them belong to the fungal Kingdom. In most cases, plant cell wall degradation results from enzymatic attack. Recent studies also underline the contribution of secreted fungal enzymes in many areas:

- many of these enzymes are already in use in numerous industrial processes,

- thanks to the comparison between fungal genomes, through the study of genes implicated in organic matter degradation, it becomes obvious that fungal evolution and their ecology are two tightly linked phenomena,

- saprotrophic fungi degrading organic matter are the main decomposers in terrestrial ecosystems. Among other, they contribute in a decisive way to the carbon biogeochemical cycle in temperate and boreal ecosystems. The study and a better comprehension of carbon fluxes are of prime importance in the evaluation of climate changes.

The fungal enzymatic machinery involved in organic matter decomposition is compozed of dozens of enzymes whose functions are diversely understood. The roles of several of them need to be clarified. This is the case for the DyP peroxidases, an enzyme family recently described but already well known in the field of biotechnology for their capacity of

catalyzing many reactions. Their natural role in natural ecosystems are however matter of discussion. It has been suggested that they could participate in lignin degradation although a role in detoxification during biomass degradation cannot be excluded.

In the course of this thesis, we highlighted the potential roles of these enzymes.

The DyP gene family had been divided in different sub-families but no study specifically dealt with the phylogeny of fungal DyPs. Such an analysis revealed the unsuspected existence of both intracellular and extracellular DyPs in fungi

To better understand the potential roles of this fungal gene family we developed ecological analyses that first required the development of specific tools such as a protocol to extract RNA directly from decomposing wood.

Following a molecular ecology approach, we evaluated the source and diversity of DyP-producing fungi in three distinct habitats; grassland soils, forest soils and decomposing wood. A metabarcoding analysis of the fungal communities present in these different environments has first been conducted and has revealed the beneficial impact of performing metabarcoding on both environmental DNA and RNA to accurately describe fungal communities.

The study of DyPs expressed within fungal communities colonizing these different habitats has been conducted by sequence capture on environmental RNA.

Preliminary results demonstrate the validity of this approach to isolate the corresponding full-length genes from all studies environmental samples. Several of these environmental DyP genes were transformed in the fungus Podospora anserina and the expression of one of them in this heterologous host was demonstrated.

In conclusion, DyP peroxidases still represent a family of fungal enzymes of unclear role. We suggest that extracellular and intracellular DyPs may play complementary roles in both lignin degradation and detoxification of toxic environmental compounds, respectively.

This enzyme family is more specifically present in the genomes of basidiomycete fungi capable of enzymatic deconstrustion of lignin. A restricted number of DyP genes has been isolated from each of the different studied environmental samples, thus suggesting that the corresponding enzymes are not abundantly produced although present in all environments.

Résumé

La lignine est un des polymères naturels les plus abondants sur la terre particulièrement résistant à la dégradation du fait de sa structure particulière qui lui permet de participer au renforcement des parois végétales ainsi qu'à la protection de celles-ci contre l'attaque de pathogènes. Peu d'organismes peuvent dégrader efficacement la lignine et la plupart de ceux-ci appartiennent au Règne Fongique. Dans la plupart des cas la dégradation des composants de la paroi végétale est réalisée par des enzymes sécrétées. Des

études récentes soulignent combien l'arsenal enzymatique des champignons est intéressant dans plusieurs domaines :

- nombreuses de ces enzymes sont déjà utilisées en biotechnologies dans de nombreux et divers processus industriels,

- grace à la possibilité de comparer les génomes de nombreuses espèces fongiques, au travers de l'étude des gènes impliqués dans la matière organique il apparait que l'évolution fongique et l'écologie de ces microorganismes sont étroitement liés,

- les champignons dégradateurs de la matière organique végétale sont les principaux décomposeurs en milieu terrestre. Ils contribuent notamment de manière décisive dans le cycle biogéochimique du carbone en milieu forestier tempéré et boréal. L'étude et une meilleure compréhension des flux de carbone est importante pour mieux appréhender les phénomènes de changements climatiques.

L'appareil enzymatique fongique permettant la dégradation de la matière organique est composé de dizaines, sinon de centaines d'enzymes aux rôles bien décrits pour certaines. Mais pour d'autres, leurs rôles restent à définir. C'est le cas des peroxydases de type DyP, une famille enzymatique découverte récemment et déjà bien connue dans le domaine des biotechnologies, étant capables de catalyser de nombreuses réactions. Leurs rôles dans la nature sont toutefois incertains. Il a été suggéré qu'elles puissent jouer un rôle dans la dégradation de la lignine, mais certaines études suggèrent plutôt un rôle de détoxification durant la dégradation de la biomasse.

Au cours de cette thèse, nous nous sommes attachés à mettre en lumière les rôles potentiels joués par ces enzymes

La famille des DyP a récemment été divisée en sous-groupes, mais pour ce qui concerne les DyP fongiques, il n'existait pas d'analyse approfondie de leur évolution phylogénique. Un tel travail nous a conduit à mettre en lumière l'existence ignorée jusqu'à présent de DyP aussi bien intracellulaires qu'extracellulaires.

Afin d'améliorer nos connaissances sur cette famille enzymatique nous nous sommes orientés sur des analyses écologiques pour lesquelles nous avons dû au préalable développer des outils appropriés tel qu'un protocole d'extraction d'ARN à partir de bois en décomposition.

Par une approche d'écologie moléculaire, nous avons évalué quelle est la source et la diversité des champignons producteurs de DyP dans 3 habitats distincts : des sols forestiers, des sols de prairies et du bois en décomposition. Une analyse par métabarcoding des communautés fongiques de ces environnements a au préalable été réalisée et a démontré le bénéfice que l'on peut tirer en métabarcoding de l'analyse comparative de l'ADN et de l'ARN environnemental pour une description optimale de ces communautés.

L'étude de la diversité des DyP exprimées au sein des communautés fongiques présentes dans ces différents habitats a été réalisée par capture de séquences à partir d'ARN environnementaux.

Les résultats préliminaires obtenus démontrent la capacité de l'approche expérimentale à isoler les gènes pleine longueur correspondants de tous les échantillons étudiés. Certaines de ces DyP environnementales ont été

transformées dans le champignon Podospora anserina, et l'expression de l'une d'entre elle dans ce système hétérologue a pu être démontrée.

En conclusion, les peroxydases de type DyP restent encore une famille aux rôles incertains. Nous suggérons toutefois que des rôles complémentaires dans la dégradation de la lignine et la détoxification de molécules pourraient être joués par les DyPs extracellulaires d'une part et intracellulaires d'autre part.

Cette famille enzymatique est plus particulièrement présente au sein des espèces Basidiomycètes impliquées dans la dégradation enzymatique de la lignine. L'aspect fonctionnel, abordé au travers de la capture de séquence a conduit à isoler un nombre restreint de DyP de chacun des échantillons étudiés suggérant qu'il d'agit d'enzymes peu abondantes mais présentes dans tous les environnements.

Riassunto

La lignina è un polimero naturale, oltre ad essere uno dei polimeri naturali più diffusi sulla Terra è anche una della molecole organiche naturali più difficili da decomporre, in quanto possiede una struttura peculiare che le permette di avere un ruolo sia strutturale, sia difensivo, di grande importanza nella parete cellulare vegetale.

Pochi organismi sono in grado di decomporre efficacemente la lignina e molti di essi appartengono al Regno dei Funghi. Nella maggior parte dei casi la degradazione delle moltissime componenti della parete cellulare vegetale,

lignina inclusa, avviene tramite la secrezione enzimatica. Studi recenti hanno evidenziato come l'apparato enzimatico utilizzato dai funghi nella degradazione della parete cellulare sia di forte interesse per la ricerca su più fronti:

- molte di queste molecole sono potenzialmente o già utilizzate nelle più svariate applicazioni biotecnologiche e industriali,

- grazie alla possibilità di comparare interi genomi di svariate specie fungine, attraverso lo studio dei geni legati alla degradazione, si è capito come evoluzione ed ecologia fungina siano aspetti legati molto più intimamente di quanto si sospettasse,

- i funghi degradatori della materia organica di origine vegetale sono tra i principali decompositori negli ambienti terrestri. In particolare è noto il loro forte apporto nel ciclo biogeochimico del carbonio nelle foreste a clima boreale e temperato. Lo studio e la migliore comprensione del flusso del carbonio è importante per comprendere al meglio il fenomeno del cambiamento climatico.

L'apparato enzimatico fungino per la degradazione della materia organica è composto da decine, se non da centinaia di enzimi, il cui ruolo in molti casi è noto, ma per molti altri resta ancora nascosto. Questo è il caso delle DyP-type Perossidasi (DyP), una famiglia enzimatica scoperta recentemente e già ben nota a chi si occupa di applicazioni biotecnologiche, infatti essa è in grado di catalizzare moltissime reazioni diverse, ma quale sia il suo ruolo in natura non è ben chiaro. Per alcuni ricercatori sembra che le DyP siano coinvolte nella degradazione della lignina, mentre altri sostengono che

siano parte del sistema di detossificazione chi il fungo usa durante l'aggressione del substrato vegetale.

In questa tesi di dottorato si è provato a far luce su questa problematica, sotto diversi punti di vista.

La famiglia enzimatica delle DyP è stata recentemente organizzata in sottogruppi di enzimi, ma per quanto riguarda le DyP fungine non esisteva un'analisi approfondita dell'assetto filogenetico. Nel cercare di ricomporlo si è visto come un problema all'apparenza semplice, nascondesse peculiarità insospettate, come il possibile dualismo tra enzimi extracellulari ed enzimi intracellulari.

Per migliorare la conoscenza di questa famiglia enzimatica si è deciso di focalizzarsi verso un'analisi ecologica. Perciò, è stato necessario il settaggio degli strumenti appropriati sviluppando un protocollo di estrazione dell'RNA specifico per legno in decomposizione.

Tramite un approccio di ecologia molecolare, abbiamo valutato quale fosse il potenziale bacino di "funghi produttori di DyP" in tre habitat distinti suoli forestali, suoli di ambiente aperto e legno in decomposizione. Attraverso un approccio di metabarcoding si è visto come la comunità microbica attiva (effettiva) sia meglio spiegata e rappresentata da campioni basati sull'RNA ambientale (eRNA), ma che essa non può discernere dal dato sul DNA (eDNA) per la costruzione di un dataset ottimale.

Lo studio della diversità funzionale e delle DyP nello specifico, presente nei diversi habitat é stato condotto con la "sequence capture" a partire dall'RNA ambientale.

I risultati preliminari sembrano indicare un'effettiva capacità del metodo di isolare i geni corrispondenti da tutti i campioni analizzati. Alcune delle DyP isolate sono state trasformate con successo in un sistema eterologo basato sul fungo ascomicete *Podospra anserina*, la cui effettiva espressione è stata dimostrata per almeno uno dei ceppi ottenuti.

Le DyP Perossidassi restano ancora una famiglia enzimatica dal ruolo poco chiaro. Vogliamo pero suggerire la presenza di ruoli complementari per le DyP. La degradazione della lignina e la detossicazione potrebbero essere ruoli complementari attuati dalle DyP extracellulari e DyP intracellulari.

Questa famiglia enzimatica è particolarmente presente tra i funghi basidiomiceti attivi nella degradazione enzimatica della lignina. L'aspetto funzionale, affrontato con la Sequence Capture, ha evidenziato come un numero relativamente piccolo di DyP sia presente in un singolo campione, in quanto forse sono enzimi relativamente rari in natura, ma al contempo ubiquitari, in quanto rinvenuti in tutte le matrici ambientali analizzate.

Abbreviations list

AA	Auxiliary Activities
AM	Arbuscular Mycorrhiza
bp	Base Pair
BRF	Brown-rot Fungi
C	Carbon
ca.	circa (about)
CAZymes	Carbohydrate-Active Enzymes
cDNA	Complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
DyP	Dye-decolorizing Peroxidase (or DyP-type Peroxidase)
e.g.	exempli gratia (for instance)
EC	Enzyme Commission number
ECM	Ectomycorrhiza
EF1α	Elongation Factor 1-alpha
GH	Glycoside Hydrolase
i.e.	id est (namely)
ITS	Internal Transcribed Spacer
LiP	Lignine Peroxidases
Mm	Millimolar
MnP	Manganèse Peroxidase
mRNA	Messenger Ribonucleic acid
N	Nitrogen
P	Phosphorous
PCR	Polymerase Chain Reaction
PCW	Plant Cell Wall
POM	Plant Organic Matter
rDNA	Ribosomal Deoxyribonucleic Acid
RNA	Ribonucleic acid

rRNA	Ribosomal Ribonucleic acid
SOM	Soil Organic Matter
SRF	Soft-rot Fungi
VP	Versatile Peroxidase
WRF	White-rot Fungi

Chapter 01

General introduction

1.0 Foreword

Fungi are certainly less studied compared to members of other taxonomic groups such as animals or plants. Nevertheless, fungi are essential for the functioning of almost all terrestrial ecosystems as well as aquatic ones. They are also found in "extreme" environments, often considered as populated exclusively by bacteria and archaea. Estimated biodiversity of Fungi is about 3 millions of species (Treseder & Lennon, 2015), but only about 100,000 have been described by now (Liu et al., 2015). A large fraction of the unknown species is likely to occur in poorly explored and described habitats (Hyde 2001), such as the tropical forests but also in or on the surface of plants, insects, lichens, ... (Liu et al., 2015).

Thanks to their unique morphology and diversified metabolic potential, fungi colonize and explore numerous, sometime hostile environments. Many fungi secrete large quantities of different enzymes, and metabolites that enable them to attack the most diverse and recalcitrant substrates and to compete and interact with other micro or macroorganisms. With respect to mankind, humans have long ago realized the importance of fungi that are part of the human's diet and pharmacopoeia as illustrated by the case of the "Ötzi man" (Pleszczyńska et al., 2017). Indirectly and probably unconsciously, man has also domesticated/recruited many fungal species, mostly unicellular yeasts, to produce novel aliments from raw animal or plant products through fermentation (e.g. wine, beer, bread). From these historical applications, to pioneer work in the field of microbiology (by e.g. van Leeuwenhoek or Pasteur), fungi can be considered as an integral part of our daily life.

Biotechnology is a fast evolving scientific domain and the exploration and use of natural resources represents a modus operandi of this discipline. As an illustration, it has been shown that several fungi produce enzymes with high redox potential that can degrade lignin, an abundant and certainly the most recalcitrant component of the plant cell wall and therefore of plant organic matter. Several of these enzymes are of interest for the removal of

lignin in the paper industry and for the production of second-generation bioethanol. Moreover, these enzymes, that can catalyze many different chemical reactions, are also active on other natural or synthetic substrates, different from lignin. From these biochemical studies essentially motivated by biotechnological applications we now know many details on the biochemistry of lignolytic enzymes and thanks to the increasing availability of fungal genome data, thousands of sequences potentially corresponding to these enzymes have been identified. This apparent enormous wealth of knowledge masks however an important knowledge gap which regards the actual role of these enzymes in fungal biology. How do they act in nature? Under which conditions? What are their roles in saprotrophic fungi biology? Are they replaceable? These are not ordinary questions, but essential to understand essential ecological processes.

1.1 What are Fungi?

From a phylogenetic perspective, the kingdom "Fungi" together with the Metazoa (animals) and other minor taxonomic groups form the super group of the Opistokonts characterized by flagellated cells with a single posterior flagella. Flagellated cells have been lost in most extant fungi but are still present in zoospores produced by Chytridiomycota (Spatafora et al., 2016). The origin of fungi has been dated back to 460 and 455 million years ago (Deacon 2013), in a period where terrestrial environments were already colonized by plants similar to the Bryophytes (Kendrick & Crane 1997).

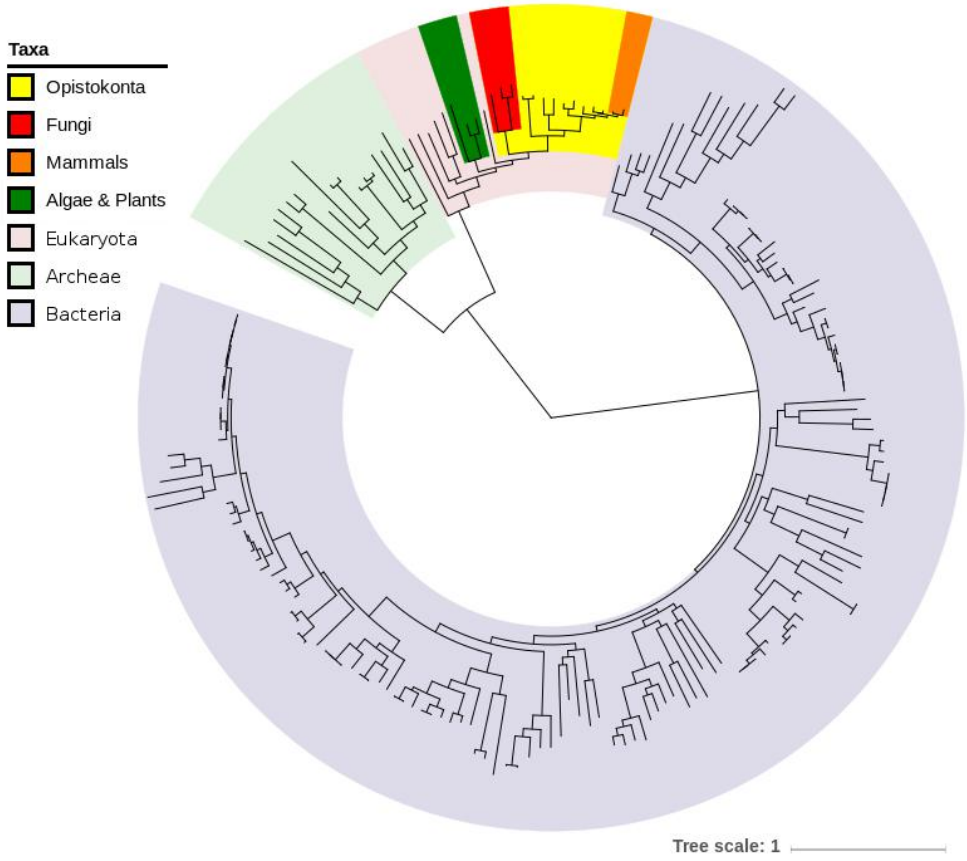


Figure 1 - Tree of Life: In a tree representative of all the life kingdoms, the Opistokonts are highlighted in yellow and, inside them, the fungal clade appears in red. Orange and green colors indicate the position of respectively: mammals and plants. Animals are phylogenetically closer to fungi than plants (Cavalier-Smith & Chao 1995). [Image from: https://itol.embl.de/itol.cgi].

Most Fungi typically grow as filaments, called hyphae (singular: hypha), which extend only at their tips. The hyphae together form a network called mycelium. Some are truly unicellular organisms called yeasts; some others switch between a yeast form and a filamentous form (dimorphic species). Most Fungi, unlike most other eukaryotes, have haploid nuclei in their vegetative cells. It is however common to find several, genetically different nuclei inside a single fungal cell (*heterokaryosis*). Fungi reproduce by both sexual and asexual ways, typically through spores. Spore features and the structures that give rise to them vary considerably from one taxonomic group / species to another and their description has been originally at the base of fungal classification. Fungi are heterotrophic organisms; they absorb soluble nutrients by diffusion through their cell walls and plasma membrane in a process called osmotrophy. Their cell walls are typically composed of chitin and glucans (glucose polymers with β-1,3 and β-1,6 as dominant linkages). Mannitol, trehalose and glycogen are the most common storage compounds in fungi.

1.2 What Fungi do in nature?

A majority of the Fungi are true saprotrophic decomposers that degrade dead organic matter. Many species are pathogenic, living at the expense of a living host (mostly plants, but also animals or other taxonomic groups, including fungi themselves); other species are mutualistic symbionts, essentially of plants, as illustrated by the plant root mycorrhizal fungal species.

Fungi can live in a wide variety of habitats. In forest ecosystems, for example, living fungi occur in most compartments of the ecosystem: tree foliage as pathogens or endophytes, bark surface, decomposing wood, roots as pathogens or mutualistic symbionts, forest litter as saprotrophs, bulk soil and rhizosphere, associated to soil invertebrates as symbionts or pathogens (Baldrian 2017). In that respect fungi play essential roles in the global

ecosystem functioning, but are mainly implicated as saprotrophs in nutrient cycling (Baldrian & Lindahl 2011; Purahong et al., 2014; Hiscox et al., 2015).

Referring to saprotrophic species, in term of ecosystem functioning, fungi are commonly placed at the base of the food chains. This view is strongly reductive but it is true that in terrestrial ecosystems, fungi are the primary ad main decomposers of dead organic matter, essentially of plant origin (Boberg 2009; Lindahl & Tunlid 2015). As opposed to other unicellular microorganisms, filamentous fungi are indeed well adapted to move and efficiently explore the three-dimensional structure of the solid (by opposition of liquid) terrestrial substrates (e.g. soils) (Boddy et al., 2009). Exploration by connected hyphae allow mycelia to find nutrient-rich patches of substrate, that are quickly colonized, the content of which transferred over long distances to connected hyphae growing in nutrient-poor environments in search of other nutrient-rich patches of substrate (Boddy et al., 2009). This ability to move as a connected structure inside solid heterogeneous environmental matrices gives to fungal mycelia a competitive advantage over unicellular microorganisms with a limited capacity of self-movement (Barnerd et al., 2013).

"Nutrient cycling" and fungal degradation of organic matter essentially takes place in the pedosphere (the most superficial part of the lithosphere that includes soils) and more specifically in the most superficial soil horizons where the dead plant material accumulates (Bardgett 2010). Plant organic matter (POM) degradation is equally crucial for both pedogenesis and sustaining life cycles. For the above-described reasons, fungi are well-adapted to live in soil, degrading soil organic matter (SOM), thus contributing to the decomposition and transformation of recalcitrant organic carbon-rich material that contribute to the recycling of organic nitrogen and phosphorus (Baldrian & Lindahl 2011; Purahong et al., 2014; Hiscox et al., 2015). Few lifestyles (free-living mycelium, mycorrhiza and yeast-like forms) reflect the main “functional traits” of fungi, with essential consequences in their role in ecosystems (Treseder & Lennon 2015).

1.3 What is Plant Organic Matter?

Plant Organic Matter (POM) includes both the dead and living parts of a plant. In mass, excluding water, the main part of this material is essentially composed of the Plant Cell Wall (PCW) components (Bardgett 2010). The exact PCW composition and structure is variable across the plant kingdom, and each part of a plant and each plant cell type has a specific PCW composition. However, in terms of "bricks" or single molecule families the PCW is relatively homogeneous between land plants (Keegstra 2010).

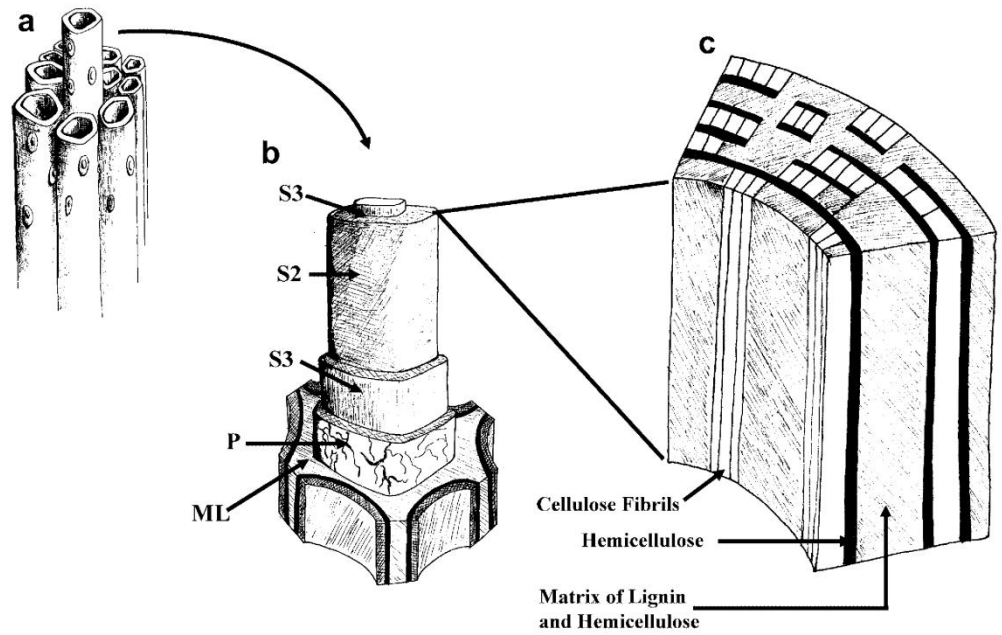


Figure 03 - Plant Cell Wall structure: Configuration of wood tissue: (a) Adjacent cells positioning, (b) cell wall layers: S1, S2, S3 are the secondary cell wall layers, P is the primary wall and ML is the middle lamella. (c) Schematic distribution of lignin, hemicellulose and cellulose in the secondary wall. [Image adapted from Perez et al., 2002].

Basically, the PCW is a highly organized and dynamic network of both "simple" and complex polymers. The PCW plays many different roles. As for other cell walls (in fungi for example), it maintains the structural integrity of the plant cell and of the plant as a whole. It confers resistance, rigidity and protection to the cell against biotic and abiotic stresses, it senses extra-cellular

information, and mediates signaling processes. (Ochoa-Villarreal et al., 2012). The main components of the plant cell wall are essentially polymers; including polysaccharides, lignin and proteins (Popper & Tuohy 2010). It also contains low molecular weight organic molecules (secondary metabolites, water and ions) (Valette et al., 2017). The different biomechanical properties of the plant cell wall are mainly defined by the content of the polymers cellulose, hemicelluloses and pectins (Pettersen 1984). The structural support is the leading role of the lignin (Perez et al., 2002). Wood represent the most highly lignified plant tissues.

1.3.1 Cellulose

Cellulose is the most abundant natural polymer on Earth. It represents about the 40% of plant cell wall material (Eriksson et al., 1990) and therefore it forms a significant part of the plant litter (between 10% and 50%) (Berg & McClaugherty 2014). Cellulose, a structural component of the plant cell walls (Keegstra 2010), it is a relatively simple polymer; it consists in linear polymers of D-glucose linked by β(1→4)-glycosidic bonds (Pettersen 1984). Different linear cellulose chains assemble in PCW to form cellulose microfibrils (Pereira et al.,2011).

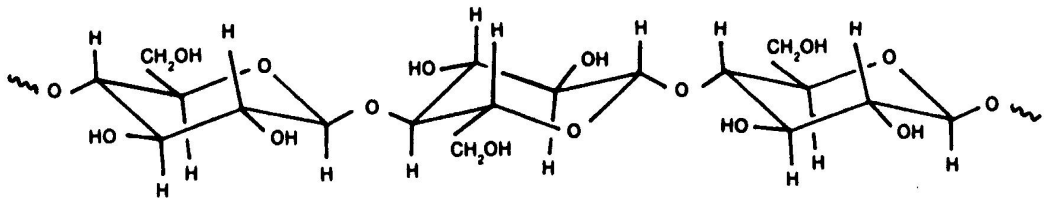


Figure 04 - Cellulose structure: [Image adapted from Pettersen 1984].

1.3.2 Hemicellulose

Cellulose microfibers are connected to each others by short chains of hemicelluloses. They are a complex set of heteropolymers composed of different monosaccharides (the principal ones being glucose, mannose xylose, arabinose, fucose and galactose) (Sarkar et al., 2009), that form branched or

linear chains. Hemicelluloses, with other macromolecules, exert a protective role on cellulose, which alone is easy to degrade (Willats et al., 2011). Hemicelluloses represent from 20 to 40% of the plant litter (Berg & McClaugherty, 2014). Composition and relative abundance of the different monomers depend on plant species and tissues, for example approximately 15% of Scots pine (*Pinus sylvestris* L.) needles, are linear or branched polymers containing several different sugar units (Berg & McClaugherty, 2003).

In addition, angiosperm hemicelluloses are rich in xylose and poor in mannose while gymnosperm ones are rich in mannose and poor in xylose. Due to their complex structure, full degradation of hemicelluloses necessitates a large set of different enzymes (Eriksson et al., 1990).

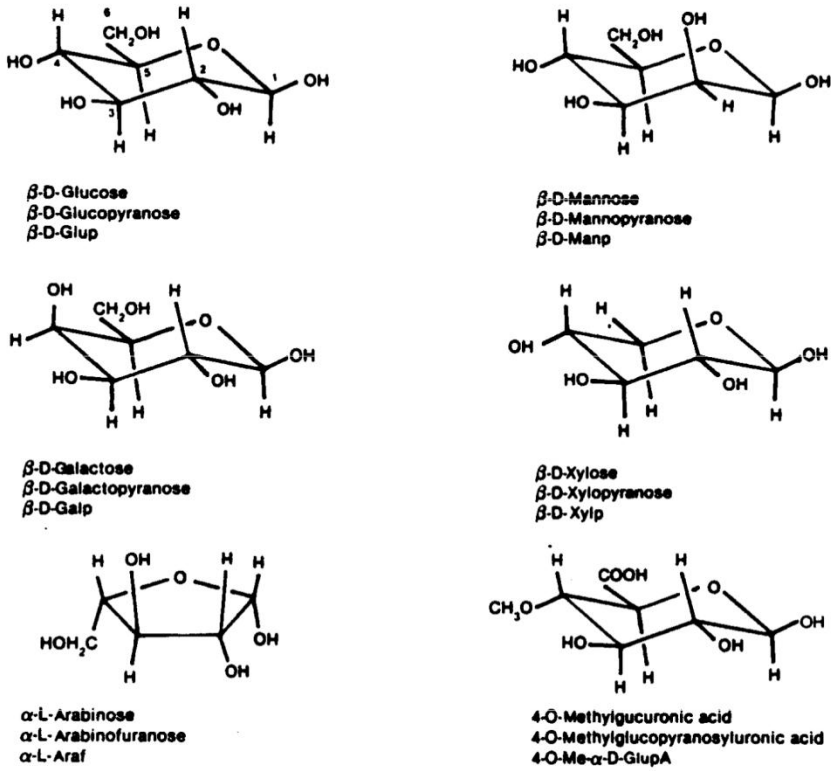


Figure 05 - Hemicellulose monomers: [Image adapted from Pettersen 1984].

1.3.3 Lignin

Lignin is the second most abundant polymer on Earth. It represents

about 30% of the organic carbon available in the biosphere (Boerjan et al., 2003). It is a complex aromatic heteropolymer, whose biological role in plants is to increase cell wall integrity and resistance to biotic and abiotic stresses (Brown & Chang, 2014). The lignin heterogeneity resides in both its structure and in its composition. Its structure is not well defined because of the lack of repeated units (Eriksson et al., 1990).

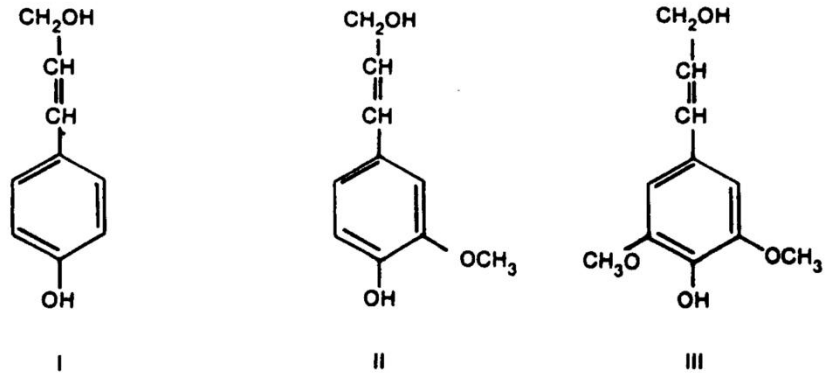


Figure 06 - Lignin precursors: The precursors of lignin biosynthesis are p-coumaryl alcohol (I), coniferyl alcohol (II), and sinapyl alcohol (III). [Image adapted from Pettersen et al., 1984]

Moreover, the relative abundance of each single monolignol monomers is highly variable. There are three basic structure methoxylated in various degrees: p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. These monomers, when incorporated in the lignin structure are in the form of the phenylpropanoids p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S). (Boerjan et al., 2003; Ralph et al., 2004; Davin & Lewis 2005; Weng & Chapple 2010). Relative abundance of the different monomers is species and tissue-specific. Lignin biological oxidation is a very complex enzymatic and/or non-enzymatic process performed by a limited group of (micro)organisms (Baldrian & Valášková 2008; Arantes et al., 2012; Brown & Chang 2014).

1.3.4 Pectin

Pectins are the most abundant heteropolysaccharide forming the

matrix in which lignin, cellulose and hemicelluloses are embedded (Willats et al., 2001). It is the principal component of the primary cell wall of terrestrial plants. Pectins are a heterologous group of compounds, the main ones being homogalacturonans, rhamnogalacturonan I and rhamnogalacturonan II. They are linear and branched polymers composed of galacturonic acid and rhamnose (Sarkar et al., 2009). Pectin functions are multiple and they are the main actors in developmental tissue modifications (Willats et al., 2001), the best-known example of this phenomena being fruit ripening. As for the hemicelluloses, the diversified structure of the pectin requires the action of multiple enzymes for its degradation (Manavalan et al., 2015).

1.3.5 Wood extractives

Aside from the lignocellulose, wood tissues contain a variety of low molecular weight organic compounds, called extractives. They do not have structural roles and are present in variable amounts (Valette et al., 2017). There are thousands of different extractives, but these compounds can be classified into 2 main groups: (1) aliphatic and alicyclic compounds and (2) phenolic compounds (Stenius 2000). They are the source of diverse bio-active molecules which principal role is the protection of the plant from attack by insects and microorganisms. Extractives represent with lignin the two main weapons developed by plants against fungal pathogen attack on wood (Valette et al., 2017).

1.4 Which Fungi Degrade Plant Organic Matter?

The kingdom Fungi (Mycota or Eumycota intended as the assemblage of the true fungi) is traditionally subdivided into five phyla: Chytridiomycota, Zygomycota, Glomeromycota, Ascomycota and Basidiomycota (Hibbett et al., 2007). The fungal taxonomy is a complex and fast evolving discipline (Schüßler et al., 2001; James et al., 2006; Hibbett et al., 2007; Spatafora et al., 2016), this paragraph is not intended as a taxonomic reference, but a way to

briefly describe the main groups of fungi in relation to POM degradation.

1.4.1 Chytridiomycota and Glomeromycota

Both these two taxa do not play a significant role in POM degradation. Chytridiomycota that still produce flagellated spores are preferentially aquatic and many of them are pathogens of animals or of other microorganisms. Fungi that belong to the Glomeromycota are ancient fungi forming the arbuscular mycorrhiza (AM)(James et al., 2006; Smith & Read 2008), a plant-fungus symbiosis that probably was fundamental in the colonization of the terrestrial environment by plants (Brundett 2004). About 80% of the vascular plants form this kind of symbiosis (Smith et al., 2003). Glomeromycota species are obligate symbionts that cannot be grown in vitro and are not involved in POM degradation but instead involved in POM production, as reflected by an almost absence of genes encoding PCW-degrading enzymes in their genomes (Kohler et al., 2015).

1.4.2 Zygomycota

Zygomycota is a complex and probably polyphyletic group of fungi (Spatafora et al., 2016) with cell wall composed of chitin, chitosan and polyglucuronic acid and coenocytic mycelia. From the ecological point of view, they are known as “sugar fungi” because they grow very fast absorbing simple and soluble carbon sources but seem to have little capacity at degrading complex PCW polymers (Fukusawa et al., 2011; Battaglia et al., 2011; Treseder & Lennon 2015). They are very common in the first phases of colonization of POM (Schneider et al., 2012). They are abundant in soil and dung or other peculiar environments (Vanderwolf et al., 2012).

1.4.3 Dikarya

The Dikarya subkingdom includes Fungi that belong to the divisions Ascomycota and Basidiomycota. Both of them produce dikaryons at some

stages in their life cycle. Dikaryotic cells contain two genetically distinct nuclei that will ultimately fuse before meiosis. The term Dikarya is frequently used because it includes the so-called "higher fungi", and many anamorphic species often called molds and classified as "imperfect fungi" in historical literature.

The phylum Ascomycota phylum contains the 75% of the described fungal species (Deacon 2013). It is thus the largest and most diverse group of fungi, that are primarily defined by the formation of sexual spores (ascospores) within a specialized cell called the ascus Ascomycota species encompass all trophic status found in the fungi. Most of them are either saprotrophic species living on plant detritus, as exemplified by the fast growing molds that profusely produce asexual spores, or certainly the most common plant pathogens. Besides saprotrophs and plant pathogens Ascomycota also include animal pathogens and symbionts forming either ectomycorrhizas with plant roots (e.g. the truffles) or lichens with algae.

The phylum Basidiomycota is certainly the most familiar form of fungi outside of the scientific domain. It indeed includes most of the fungal mushroom-forming species. With more than 30,000 known species (Kirk et al., 2001) they are characterized by the “basidium”, a specialized cell where meiosis takes place and which differentiates sexual spores (basidiospores) on the outside. In fungal ecology, and especially in forest ecosystems, Basidiomycota species, and more specifically Agaromycetes species, are often considered as the main wood and litter decomposers (Boddy 1995; Baldrian 2008; Lonsdale et al., 2008; Johnston et al., 2016). Indeed, Agaricomycetes saprotrophic species are the only known fungal species that have been shown to degrade efficiently all wood components, including lignin (Nagy et al., 2017).

1.5 The Dikarya and the plant organic matter decay

1.5.2 Free-living filamentous fungi

Concentrating on the Basidiomycota, according to Kohler et al., 2015

and Martin et al., 2016, their common ancestor could have been a free-living saprotrophic filamentous species. Taking the definition of Deacon (2006) “Saprotrophs (saprophytes) are organisms that feed on dead organic matter (sapros = death; trophy = feeding)”, this excludes the feeding on living organisms as performed by pathogens. In the fungi, this distinction between "true saprotrophs" and pathogen is often not clear-cut as many pathogens behave as true saprotrophs at some stages of their life cycles and many species involved, especially in the early stages of dead POM degradation live in healthy plant tissues before plant death.

According to the mode of attack of the plant material, the final outcome of the attack and the nature of the plant material, saprotrophic basidiomycetes have historically been classified as "white rot", "brown rot" or "litter" species (Eichlerová et al., 2015). The first two attach wood, while the third group degrade plant litter.

Figure 07 - Wood rots : In the two photos WRF (on the left) and BRF (on the right) effect on wood pieces are shown. White-rot fungi degrade lignin leaving cellulose fibers partially intact, for this reason rotten wood has a pale yellow characteristic color. On the contrary BRF attack wood leaving lignin mostly intact, in this case rot wood has a typical dark color. [Images by David Mallock, Mycology Web Page]

White-Rot Fungi

White-Rot Fungi (WRF) are basidiomycetes able to enzymatically break down both lignin (first) and (then) cellulose leaving behind a moist, spongy and white/yellow fibrous wood-derived material (Bugg et al., 2011). WRF specificity is to produce efficient and specific enzyme cocktails containing unique enzymes (such as so-called Class II, lignin, manganese or versatile Peroxidases) capable to perform the complete degradation of lignin,

one of the most recalcitrant polymer in nature (Riley et al., 2014). This ability confers to the WRF the status of unique organisms with a demonstrated capacity to directly degrade lignin in nature. It has been suggested that several bacteria could also achieve lignin degradation, but less efficiently (Bugg et al., 2011). During the course of basidiomycete evolution, the WR phenotype originated soon after the appearance of the Agaricomyctes (the main and best known clade of the Basidiomycota) (Floudas et al., 2012).

Brown-rot Fungi

Brown-rot Fungi (BR) principally break-down enzymatically or via non enzymatic oxidative mechanisms hemicellulose and cellulose, but not lignin, thus living behind a brownish brittle material enriched in modified lignin (Otjien et al., 1987). Certain BR species, such as *Serpula lacrymans* or *Coniophora puteana*, have also the characteristic to leave a dry surface, and sometimes literature refers to these species as dry-rots (Shmidt 2006). BRF are also basidiomycetes that achieve via a non-enzymatic strategy (Fenton cycle) the modification, by demethylation, of lignin thus allowing the diffusion through the lignin barrier of enzymes and the degradation products of polysaccharides. BR species attack on wood is therefore a two-step process (Arantes et al., 2012).

Recent phylogenomic studies comparing the genomes of Basidiomycete species and focusing more specifically on the distribution and diversification of the functional genes involved in plant material degradation, suggest that the WRF/BRF paradigm is inadequate to describe the wood-degrading Basidiomycota (Otjen et al., 1987; Worrall et al., 1997). It would more correct to classify the species along a continuum of differentiated modes to degrade wood (Riley et al., 2014). Nevertheless the BRF and WRF classification remains valid and convenient if intended to describe simple phenotypes. This argument will be treated further later on in this chapter.

Litter Decomposing Fungi

Litter Decomposing Fungi or Soil Saprotrophs (SS) are Basidiomycetes fungi specialized in plant litter decomposition. Plant litter is the principal

source of carbon in forest soil (Voříšková & Baldrian 2013). Its composition is variable depending on the plant organs (leaves, stem, wood, etc), the plant species and the degradation stage. Consequently, different litters will have different chemical compositions and differ with respect to the proportions of the main constitutive polymers and secondary metabolites (hemicellulose, cellulose, lignin, low molecular weight phenolics (tannins), wax…) and consequently recruit different saprotrophic fungal communities (Valette et al., 2017). With respect to the WR/BR paradigm defined for wood rot fungi, many (most) SS fungi could be WR species specializing on litter; this is supported by the genome content of several fully sequenced species (Eastwood et al., 2011; Floudas et al., 2012; Kohler et al., 2015).

Ascomycetes relevant in POM decay

To the difference of basidiomycetes, there is no clear demonstration of the involvement of ascomycetes in the degradation of lignin (Rajala et al., 2011; Manavalan et al., 2015). Therefore, most of the numerous saprotrophic ascomycetes seem to specialize on poorly lignified POM such as grassland litter, agricultural residues, fruits and other soft plant tissues. In that respect ascomycete saprotrophic taxa (sometime referred as "saprotrophic microfungi") are also abundant in soil, although their contribution to plant litter degradation may be secondary in specific habitats such as forests characterized by high lignin content POM (Thevenot et al., 2010). Paradoxically several saprotrophic ascomycete species are the main sources of commercialized enzymes related to cellulose, hemicellulose, pectin degradation that are abundantly used in many different industrial field (Marmeisse et al., 2017). This is the case of e.g. *Penicillium* or *Trichoderma* species.

Although a direct involvement of Ascomycete taxa in lignin degradation has not been established, many Ascomycete taxa are associated with wood. These Wood Associated (WA) taxa are sometime referred to as Soft-rot Fungi (SRF) (Worrall et al., 1997; Rajala et al., 2011; Manavalan et al., 2015), but their action on wood structure is highly variable and it is difficult to assign their activity on wood to a clear category. Typical examples of WA taxa are species

of the genus *Chetomium* and of the order Xylariales (Worrall et al., 1997; Osono & Takeda 2002; Eichlerová et al., 2015; Purahong et al., 2017). These species do not attack lignin, but secrete enzymes that degrade cellulose, resulting in the formation of micro cavities making the wood spongy (Shmidt 2006). Sometimes SRF live in extreme conditions, too harsh be colonized by BRF and WRF, in a sort of residual niche, other times they act as pioneer species (Osono & Takeda 2002; Baldrian & Lindahl 2011).

1.5.2 Symbiotic mycorrhizal fungi and POM decay

Numerous taxa of filamentous Dikarya, mostly in the Basidiomycota and to a lesser extent in the Ascomycota are mutualistic symbiotic species associated to the roots of living plants (Smith & Read 2008). These mycorrhizal associations are of several kinds and classified according to the mode of "penetration" of the fungal hyphae within the roots and the host plants. While a few Ascomycota associate to Ericales plants (ericoid mycorrhizas) and a few Basidiomycota to orchids (Orchid mycorrhizae), several thousands of both Ascomycota and Basidiomycota associate to the roots of trees and shrubs to form the so-called ectomycorrhizal (EM) symbiosis (Selosse & Le Tacon 1998; Brundrett 2004). A classical and certainly simplistic functional view of the mycorrhizal plant/fungus association is the beneficial bi-directional exchange of essential nutrients (Smith & Read 2008; Martin et al., 2016). On the one side, the fungal hyphae explore a large volume of soil where they assimilate poorly mobile and complex forms of nutrients such as phosphorus and organic forms of nitrogen (Agerer 2001). These nutrients are then transferred to the plant partner in exchange, on the other side, of simple carbon sources (monosaccharides) directly derived from the photosynthesis (Smith & Read 2008). The fungus thus relies on photosynthesis for its carbon nutrition and not on the degradation of dead POM as for saprotrophic species (Högberg et al., 2001).

From a phylogenetic perspective, the EM phenotype is neither ancestral nor monophyletic, EM fungal lineages appeared several time

independently during the course of fungal evolution from white rot saprotrophic lineages (Floudas et al., 2012; Kohler et al., 2015).

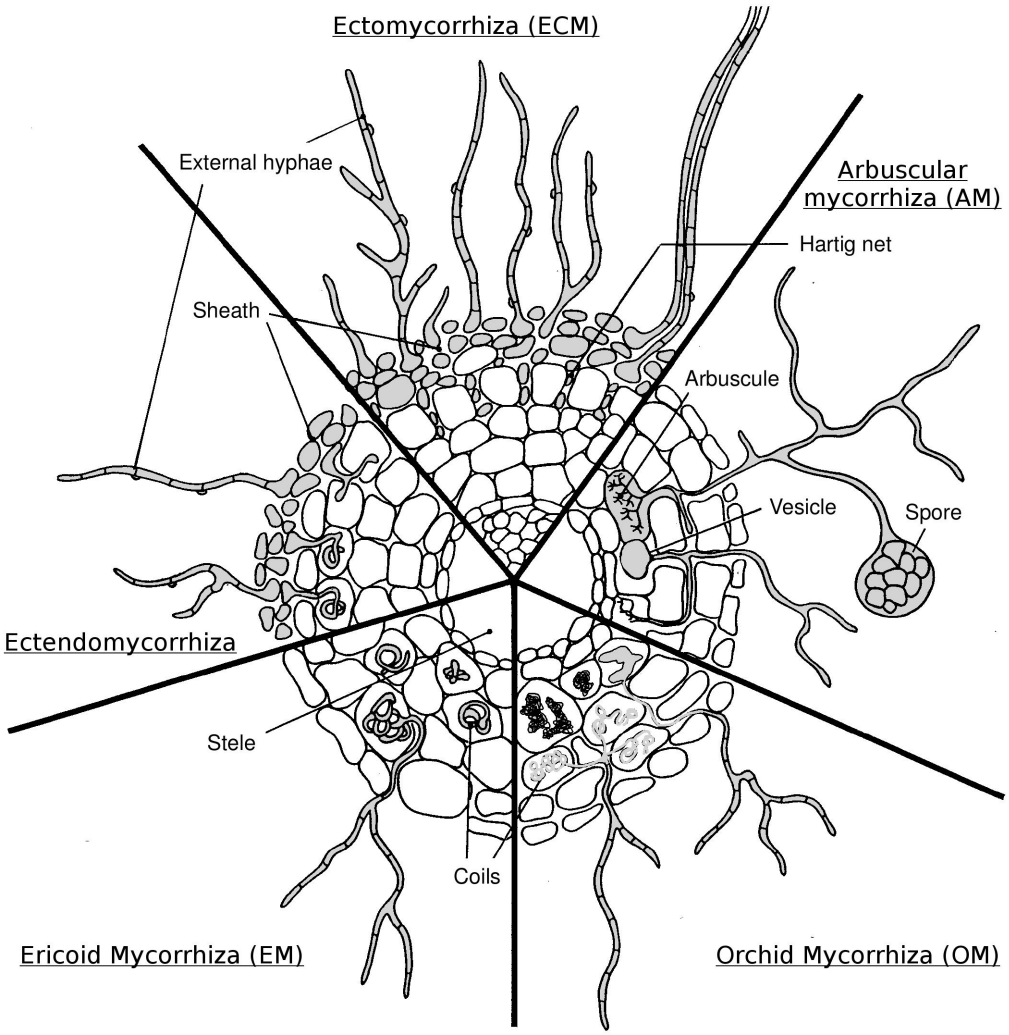


Figure 08 - Mycorrhizas : Different types of mycorrhizas are schematically represented in the figure. Plant structures are white, while fungal structures are pale grey colored. [Image modified from Selosse & Le Tacon (1998)]

The genome sequencing of several EM fungal species belonging to distinct EM lineages demonstrated that the transition from a saprotrophic to a symbiotic trophic mode was accompanied by the systematic reduction (if not complete lost) of many of the gene families encoding enzymes participating to POM degradation (Kohler et al., 2015; Nagy et al., 2016). Despite this reduction of the "saprotrophic toolkit", several EM species may have retained

a certain capacity of degrading/modifying lignin in order to access the organic nitrogen trapped in dead POM. POM degradation by EM fungi could proceed through the Fenton chemistry as in the case of brown rot wood fungi (Rineau et al., 2012; Tunlid et al., 2013) or through the secretion of class II peroxidases as suggested for *Cortinarius* species (Lindahl & Tunlid 2015; Shah et al., 2016).

From an ecological perspective, the EM symbiosis is the main form of mycorrhizal association in temperate and boreal forests dominated by EM tree species (Smith & Read 2008; Lilleskov et al., 2010). In the corresponding forest soils, extramatricial mycelia from EM fungi could contribute up to one-third of the total microbial biomass (Högberg & Högberg 2002). Quantification of the relative importance of EM and saprotrophic fungal species in POM alteration in this environment represents therefore a priority to understand nutrient cycling in high latitude forest ecosystems (Boberg et al., 2009; Martin et al., 2016).

1.6 Soil fungal ecology

1.6.1 A rapid definition of soil

Soil is a thin (from a few cm to a few meters in depth) layer of "non-solid material" that covers emerged lands. It results from the degradation of the underlying parental rock material and the accumulation of organic material through the combined and synergistic action of physico-chemical and biological factors. It represents the natural substratum for land plans and soil structure and chemistry is largely determined by the nature of the parental material, topography, climate and vegetation. The soil biota, especially soil microorganisms, contribute to pedogenesis and soil fertility. Soils are not homogeneous but are usually composed of distinct and discrete layers (horizons) whose physico-chemical characteristics differ from one another (Tripathi et al., 1992).

A typical soil profile in temperate climates consists of four main layers; from

top to bottom; the uppermost *O horizon* essentially composed of organic matter; the *A horizon*, organic rich, depleted in soluble ions; the *B horizon*, poor in organic matter but enriched in soluble ions and the *C horizon* (physically and chemically weathered parent material) that makes the transition and penetrates into the unaltered parent material (Tripathi et al., 1992). Above the O horizon may exist a layer of poorly degraded dead plant material called *litter* (Bardgett et al., 2010). In terrestrial ecosystems, the plant-litter input to soil represents the main source of energy and matter for an extraordinarily diverse community of soil organisms connected through complex networks of interactions (Hättenschwiler et al., 2005)

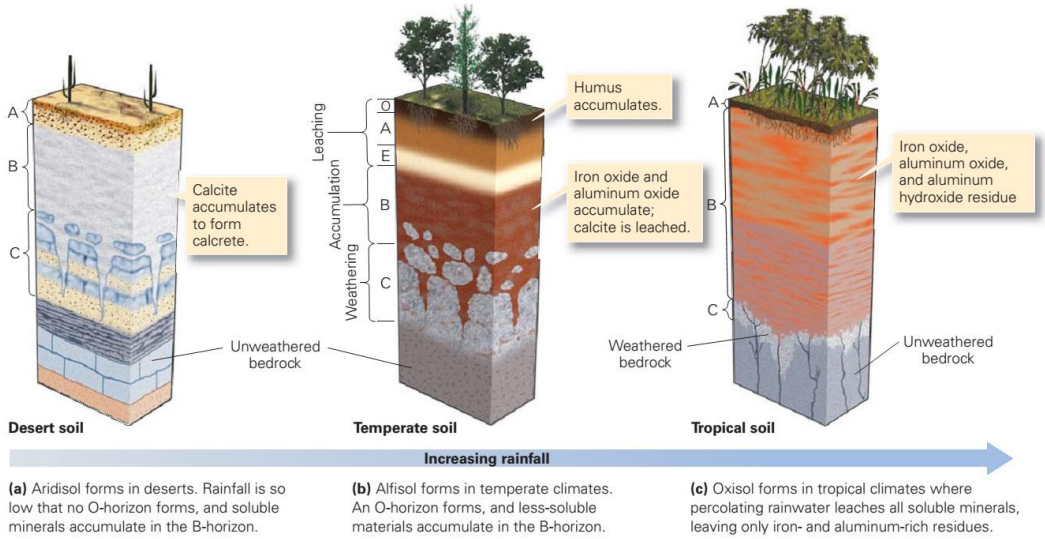


Figure 02 - Soils horizons: soil profile is influenced by many environmental variables (climate, relief, organisms activity, parent material,...) that contribute to shape pedosphere along the geological scale of time. For these reasons there is not a single type of soil. Three strongly different soils are illustrated in the figure following a rainfall gradient: from few mm of rain per year of the desert soil up to more than 2000 mm of rain per year of the pluvial forests, soil characteristics are extremely different. [Image from: Learning Geology Blogspot].

As intuited from the soil layer composition and the driving forces, the two main resources of material for the soil formation are: the parent material (bedrocks or sediments) for what concern the mineral component and the plant organic matter (POM) for what concern the organic component (Anderson 1988; Kögel-Knabner 2002).

1.6.2 The soil biota

Decomposition of dead plant material is a complex process that involves an extended network of interactions between a huge diversity of organisms which in some way cooperate to contribute to the formation of the soil itself (Bardgett et al., 2010). By means of several different actions, soil inhabitants contribute to one or more processes, the main ones being the physical destruction of soil particles, particles mixing, chemical transformation of the organic material, solubilisation and assimilation of the inorganic nutrients, ... This variety of actions is performed by organisms which are classically (and arbitrarily) divided according to their body size (**Figure 09**).

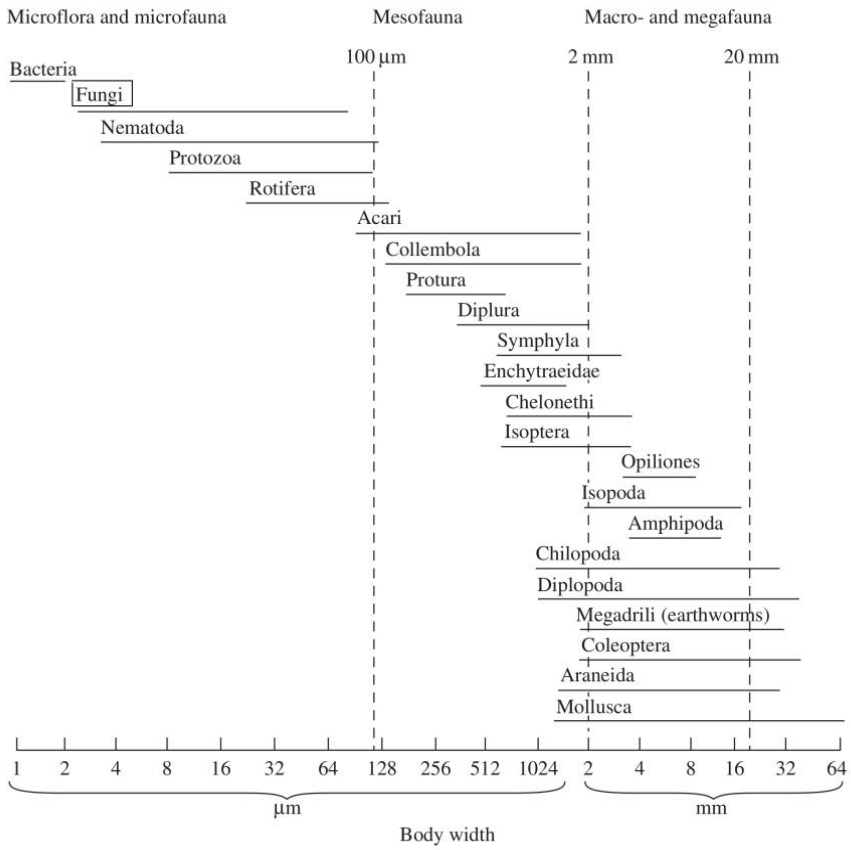


Figure 09 - Body size classification of the soil biota: Image extracted from Bardgett 2010.

The macrofauna groups all organisms larger than 2mm, mesofauna

includes organisms whose sizes range from 0.1 to 2 mm, and the microfauna consists in organisms smaller than 0.1mm (Jeffery et al., 2010). The macrofauna and mesofauna that include numerous animals participate to the first steps of the degradation of the dead plant organic matter by performing the mechanical fragmentation and burial of the organic matter in soil. The microfauna includes eukaryotic unicellular protists but also rotifers and nematodes, which mainly live as predators or parasites of other species (Bardgett et al., 2010). The most relevant and certainly most diverse part of the microfauna is however represented by heterotrophic but non-phagotrophic microorganisms, essentially bacteria and fungi that are the main agents of the Soil Organic Matter (SOM) transformation (Hättenschwiler et al., 2005). Microorganisms are the most numerous inhabitant of soil in terms of biomass and species numbers. With 10^7-10^{12} microbial cells in one gram of soil (Watt et al., 2006) they represent a hotspot of biodiversity (Kuzyakov & Blagodatskaya, 2015). Microbes are involved in organic matter decomposition, thus participating to the recycling of organic carbon (C), an essential step in the carbon cycle (Hättenschwiler et al., 2005; Boberg et al., 2009; Hibbett et al., 2016). Through the mineralization of dead plant organic matter, decomposer organisms contribute to the recycling of essential inorganic elements (e.g. N, P, S) essential to sustain plant growth. They thus control soil fertility (Maeder et al., 2002; Brundrett 2004; Acacio et al., 2012) and consequently, plant growth and below- and above-ground community structure (van der Heijden et al., 1998).

As discussed in the previous sections, fungi play major roles in the cycling of soil organic matter and nutrients (Schneider et al., 2012; Kellner et al., 2014). All the different fungal species that live in a specific environment form the local soil fungal community, which is shaped by several biotic and abiotic features. Fungal community composition is highly variable and is controlled by multiple ecological factors; from the mineral soil composition, to the composition of the soil organic matter, to the local climate and so on (Bardgett 2010; Wu et al., 2013, Burns et al., 2015; Baldrian et al., 2016; Žifcáková et al., 2016).

1.6.3 Forest soil fungal communities

Soil fungal communities present in boreal and temperate forests are among the most extensively studied. In these forests dominated by ectomycorrhizal tree species the main feature of the corresponding soil communities is the presence of large quantities of ECM mycelia (Högberg & Högberg 2002). Symbiotic fungi extract, to the benefit of the associated plants, N and P from the SOM, with therefore an important role in soil fertility (Brundrett 2004; Martin et al., 2016). Mycorrhizal fungi colonize essentially the organic layer of soil, while saprotrophic fungi lives preferentially in the litter, the most organic and superficial layer of soil, composed of mostly unmodified dead plant matter (Baldrian et al., 2012). Fungal activity is variable in time; during the vegetative season, mycorrhizal fungi are particularly active, presumably because their activity is strictly related to the below-ground allocation of soluble C by the photosynthetically-active trees. Saprotrophic fungi that are not directly dependent from actual photosynthetic activity seem to be less affected by the seasons (Žifčáková et al., 2016). They however vary in feeding strategy. During winter (especially when snow covers the ground) they are most active on the most recalcitrant substrates, while in other seasons, they preferentially degrade more labile substrates in an efficient energy consumption strategy (Baldrian et al., 2016).

1.6.4 Grassland soil fungal communities

Today natural grasslands are very rare. The great majority of them has been converted in agricultural lands. Residues of this natural environment remain in the North American prairie, the South American pampas, the African savanna and the Eurasian steppe. Other natural grassland residues occur in specific environments, for instance in the alpine one (alpine grasslands). Because of the vastness of the cited areas and their climatic

diversity, in the next paragraphs I will concentrate on the boreal/temperate environments as previously done for the forest fungal communities.

Grasslands are important biomes. They comprise approximately 32 % of the natural vegetation of earth's surface (Adams et al., 1990) and store 28–37 % of the terrestrial organic soil carbon (Lal 2004). Function and composition of the plant communities in grassland ecosystems have been proposed to be sensitive to variability in climatic conditions (Knapp & Smith 2001; Gao & Reynolds 2003). Water availability and temperature are the major abiotic parameters that influence grassland carbon exchange on an ecosystem level both above- and below-ground (Xu et al., 2004; Davidson & Janssens, 2006).

Grassland soils are different from the forest ones with respect to several features. Quantity of "input" dead plant material is smaller in grassland soils compared to forest ones and the quality is also different, with for example a far lower concentration of lignin and an absence of woody debris (Conant et al., 2001; Zak et al., 2003; Six et al., 2008). Grassland soils also characterized by a lower spatial heterogeneity if compared with the forest environment, that is characterized by a multi-layered vegetation. Consequently grassland soils hosts fungal communities different from the forest soils ones (Spurgeon et al., 2013). Due to the dominance of herbaceous plants there is, for example, an abundance of AM (Arbuscular Mycorrhiza) species belonging to the phylum Glomeromycota (Öpik et al., 2003; Liu et al., 2015). Another well represented phylum is the phylum Ascomycota, which includes in grassland ecosystems a great variety of fungi active as pathogens of herbaceous plants, as well as opportunistic saprophytes (Deacon et al., 2006; Porras-Alfaro et al., 2011; Liu et al., 2015). Even if more prevalent in the forest soils, basidiomycetes species are ecologically important in the grasslands. Some basidiomycetes (*Agaricus arvensis*, *Marasmius oreades*, ...) characteristics of the grasslands are well known for the "*fairy rings*" formation (Gramss et al., 2005; Liu et al., 2015). They are involved in the degradation of a litter with low contents in lignin or in other cases in the degradation of

material of animal origin like dung (Spatafora et al., 2016). Although of little importance in term of biomass they represent, terrestrial orchids are good indicators of the environmental status of many grasslands (Sindaco et al., 2009), orchid mycorrhiza are mainly characterized by the association with basidiomycetes of the Cantharellales order (*Tulasnella sp.* and *Ceratobasidium sp.*), and of the Sebacinales order (*Serendipita sp.*) (Ercole et al., 2015).

1.7 Wood fungal ecology

Wood is a specific habitat for many organisms, in a single forest it represents a very large environment if we consider the wood of the living trees, the bark and the decaying wood. Many fungal species live in wood, but only a restricted number of specialized Basidiomycetes are seemingly able to truly perform its degradation (Worrall et al., 1997; Hammel & Cullen 2008).

1.7.1 Bark

Tree bark is not properly considered as wood, but is connected to it. Bark is the first defense line of the tree, for this reason it represents a material rich of compounds recalcitrant to microbial decomposition, such as suberin and extractives (Baldrian 2017). A rich fungal community live on the bark, including lichens and their fungal components (Beck et al., 2014), yeasts and yeasts-like fungi (Bhadra et al., 2008) and in general numerous Ascomycetes (Morrison-Whittle & Goddard 2015).

1.7.2 Living trees

Living trees represent with soil, litter and decaying wood, one of the largest habitat in the forest ecosystem. Two main fungal trophic guilds dominate the fungal diversity inside the stems of the living trees; endophytes and parasites (or pathogens) (Giordano et al., 2009). Endophytes may turn out

as saprophytes upon tree death and potential saprophytes also grow inside the stems, highlighting the thin line that divide living and decaying wood (Song et al., 2017).

1.7.3 Decaying wood

Decaying wood is very peculiar habitat at the center of many studies (Lonsdale et al., 2008; Baldrian et al., 2008; Seibold et al., 2015; Hiscox et al., 2015; Purahong et al., 2016; Baldrian 2017). It is degraded by peculiar organisms that are confronted to a habitat depleted in several essential nutrients such as N and extremely rich in recalcitrant (e.g. lignin) or toxic (wood "extractives") compounds (Boddy et al., 1995; Cornwell et al., 2009; Valette et al., 2017).

Deadwood abundance is strictly connected to age and management of the forest (Blaser et al., 2013; Purahong et al., 2016). In a temperate forest, its volume can range from 1200 m^3/ha in the case of "natural forests" down to 2 - 65 m^3/ha for managed forests from which wood remnants are removed (Stockland et al., 2012). Fresh wood is impermeable, rich in lignin and poor in available N (Shmidt 2006). Due to this features it is difficult to be attacked by bacteria and only specific fungi (BRF or WRF) are able to start the long degradation process (which may last more than ten years for a single tree log). Wood-decay fungi are key determinants of nutrient cycling and carbon sequestration rates in forest ecosystems (Hibbett et al., 2016; Hu et al., 2017).

Early colonizers of wood are often pioneers species arriving as spores (Osono & Takeda 2002; Baldrian & Lindahl 2011), or species previously present in the living tree (Song et al., 2015; Giordano et al., 2009). Later, specific fungi arrive as spores or via the soil as mycelium (in the case of dead wood pieces in direct contact with soil) and contribute in the formation of the specific fungal community (Boddy et al., 2009), which is generally dominated by basidiomycetes and xylariaceous ascomycetes (Baldrian et al., 2016). The initial dominance of decomposers is followed by the increase in the proportion of ECM fungi during late degradation stages (Rajala et al., 2011; Baldrian et al., 2016) as decomposed wood gradually mixes with soil.

Wood fungal community seems to be strongly influenced by a location (Hiscox et al., 2016) and a priority effect (Yamashita et al., 2015; Hiscox et al., 2015), influencing the wood degradation process. Different fungi and different fungal communities, such as BRF of WRF can create highly different wood environments from a chemical and physical point of view (Baldrian 2008; Ottosson & de Boer 2015), leading to strict selection of the following fungal species in the ecological succession. Although deadwood is initially poor in N, early during decomposition it becomes a specific habitat for bacteria, and an association between colonization by N-fixing bacteria and fungal growth has been suggested (Hoppe et al., 2014). The succession of different communities is slow and extended in time. During decomposition many ecological drivers can operate, but it seems that nitrogen content is one of the most important driver of fungal biomass content, community composition and enzyme activity (Baldrian et al., 2016). Recent studies highlighted also the importance of C quality, while during early decaying stages fungi are dominant, the accumulation of residual (low-quality) C may benefit to bacteria (Hu et al., 2017).

1.8 Fungal hydrolytic enzymes

1.8.1 What make fungi efficient in dead plant material degradation?

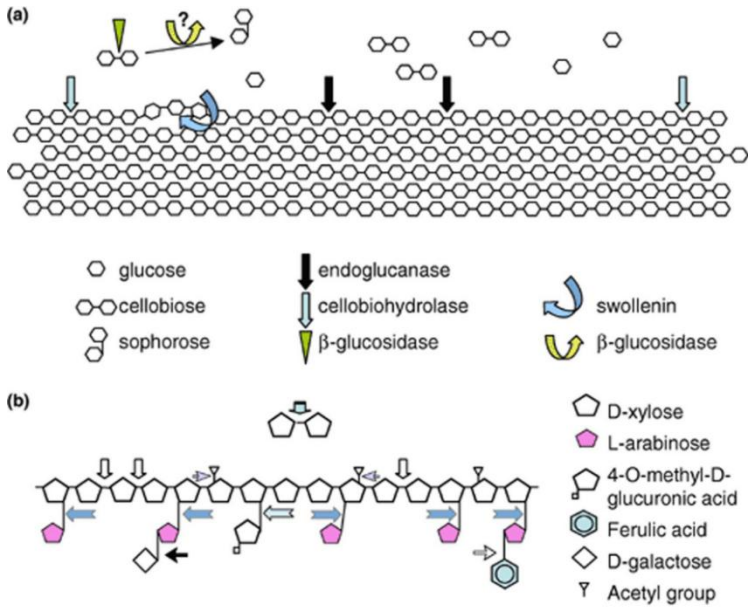
As previously underlined, the decomposition of soil C (POM) is tightly related to the activity of fungi (Bruce et al., 2013). They are able to degrade all of the major forms of “raw” C supplied to the soil, namely the different polymers of plant origin. Fungal degradation is mainly based on the production of extracellular hydrolytic enzymes that convert complex polymers into soluble, low molecular weight compounds, that enter the fungal cells to be further degraded intracellularly (Deacon 2013).

Plant organic matter is essentially composed of plant cell walls composed mainly of polysaccharides and lignin. The different enzymes active on these compounds are classified in the reference CAZy (for

carbohydrate-active enzymes) database (http://www.cazy.org/, Lombard et al., 2014). CAZy includes all enzymes active on carbohydrates whatever the nature of the carbohydrates (e.g. structural and storage carbohydrates, "protein decoration") and their activity on these carbohydrates (synthesis, degradation, modification). This enzyme database therefore extends far beyond the field of plant cell wall degradation. Because lignin (a non carbohydrate) is invariably associated to polysaccharides in plant cell walls, enzyme implicated in its degradation are also listed in this database as "Auxiliary enzymes" (AA categories). In CAZy, enzymes are primarily classified according to their mode of action on carbohydrates. Accordingly they can, for example, be "Glycoside Hydrolases" (GH) performing the hydrolysis and/or rearrangement of glycosidic bonds or, "Polysaccharide Lyases" (PL) performing non-hydrolytic cleavage of glycosidic bonds. Within each of these different broad categories, enzymes are further grouped in "families" that include all phylogenetically-related proteins that share a common ancestor. This classification therefore does not group enzymes according to their substrates and their "exact" catalytic activities, nor according to their taxonomic origin. As an illustration if we refer to the cleavage of internal β,1-4 bonds within cellulose (endo β,1-4 glucanase or endocellulase activity), enzymes presenting this activity are found in different GH families (e.g. the GH5, GH7 or GH16 families) and within several of these families, all enzymes do not necessarily present an endocellulase activity (e.g. GH5 enzymes can also correspond to endomannanases or endoxylanases active on hemicelluloses).

If we take cellulose as an example of structural component of the plant cell wall, its full degradation to glucose by fungi requires the synergistic action of different secreted and/or intracellular enzymes all active on β-1,4-glycosidic bonds (Perez et al., 2002). Endoglucanases (endo-1,4-β-glucanases, from e.g. families GH5, GH7 or GH16) hydrolyze internal bonds in the cellulose amorphous regions, releasing new terminal ends. Cellobiohydrolases (exo-1,4-β-glucanases, that can belong to families

GH6 or GH7) cleave at the extremities of cellulose chains to produce essentially cellobiose, which is finally cleaved to glucose by either extracellular or intracellular β-glucosidases (Aro et al., 2005). Regarding hemicelluloses, the higher chemical complexity and diversity of this polysaccharide results in a higher diversity of carbohydrate active enzymes necessary for its full decomposition (Perez et al., 2002; Aro et al., 2005).



SYMBOL	ENZYME	EC-NUMBER	LINKAGE HYDROLYSED
⇩	Endo-1,4-β-xylanase	3.2.1.8	β-1,4
	β-D-xylosidase	3.2.1.37	β-1,4
	α-L-arabinofuranosidase Arabinoxylan arabino-furanohydrolases	3.2.1.55	α-L-1,2; α-L-1,3 and α-L-1,5
	α-D-glucuronidase	3.2.1	α-1,2
	feruloyl esterase	3.2.1	ester bond
→	α-D-galactosidase	3.2.1.22	α-1,6
	asetylxylan esterase	3.1.1.6	ester bond

Figure 10 - Cellulose (A) and hemicellulose (B) depolymerization process: Image extracted from Aro et al., 2005.

1.8.2 Lignin degradation

Lignin degradation represents a "biochemical challenge" due to its high molecular weight, insolubility and its random polymeric structure (Pettersen 1984). White-rot fungi are the microorganisms that better degrade lignin from

wood; among them *Phanerochaete chrysosporium* (but also *Pleurotus spp.* and other Polyporales) has been the most extensively studied (Martinez et al., 2004). Peroxidases and possibly laccases are two main families of enzymes involved in ligninolysis (Aro et al., 2005; Sinsabaugh 2010). A particular group of fungi, the Brown rot fungi (BRF), use diffusible low molecular weight compounds to carry out a partial lignin degradation without a direct enzymatic action (Arantes et al., 2012). In the CAZy database, several of the enzymes putatively involved in lignin hydrolysis are classifies in "Auxiliary Activities" (AA) families (Lombard et al., 2014).

Laccases

Laccases (AA1.1, EC 1.10.3.2) are multicopper oxidases that use diphenols and related substances as donors with oxygen as final acceptor of electrons (Hofrichter et al., 2010). The AA1 family includes 3 subfamilies: laccases, ferroxidases and laccase-like multicopper oxidases (Lombard et al., 2014). Laccases are present in different groups of organisms (Bacteria, Fungi, Oomycetes) where they fulfill different but often ambiguous and controversial roles (Cragg et al., 2015). In fungi their implication in processes as diverse a pigment (melanin-like) production, aggressiveness of plant pathogens, detoxification of extracellular toxic molecules and lignin degradation has been suggested (Lundell et al., 2010). This latter role remains however debated as many of the fungi that produce large amounts of laccase do not degrade lignin although, in vitro, laccase can effectively act on this polymer (Xie et al., 2014).

Peroxidases

Peroxidases are oxidoreductases that act on a peroxide as acceptor of electrons as described in the following reaction (Hofrichter et al., 2010):

$$ROOR' + \text{electron donor } (2e^-) + 2H^+ \rightleftharpoons ROH + R'OH$$

In the EC classification, peroxidases (EC 1.11.1) are the sole sub-class of the oxidoreductases (EC 1.11.1). This sub-class includes enzymes that use peroxides as electron acceptors. For many of these enzymes, the optimal

substrate is hydrogen peroxide (H_2O_2), while for others they are organic hydro-peroxides. Many peroxidases contain a heme group within their active sites (Perez et al., 2002). These enzymes are present across the entire tree of life covering a great variety of roles (Zamocky et al., 2014). The biochemical/physiological role of a peroxidase can either be the removal of toxic peroxides or on the contrary the oxidation of one or several molecules that participate as electron donors in the reaction (Lundell et al., 2010; Valette et al., 2017). In term of diversity, several phylogenetically unrelated protein families present a peroxidase activity and heme-containing peroxidases can themselves be affiliated to distinct unrelated families (Hofrichter et al., 2010; Zamocky et al., 2014).

Regarding the fungi and their role in POM degradation/modification, at least three families of secreted heme peroxidases have been suggested to participate to this process (Lundell et al., 2010).

1.8.3 Class II Peroxidases (AA2)

Class II peroxidases (CIIP or CAZy AA2) belong to the peroxidase-catalase super-family that is currently the most abundant peroxidase superfamily represented in various protein databases (Fawal et al., 2013; Lombard et al., 2014). Representatives of this superfamily are present in Bacteria, Archaea, Fungi and Plants as well as in other Eukaryotic taxa, such as the Metazoa (Zamocky et al., 2014). Within the superfamily, Class II peroxidases represent a small group, specific to the fungi and particularly widespread and abundant in the Basidiomycota where several of them are instrumental in lignolysis (Hammel & Cullen 2008; Manavalan et al., 2015). Indeed CIIP correspond to enzymes with different ranges of substrates and different, sometime poorly known physiological roles. Several of them, known as Lignin peroxidases (LiP), were the first "truly" lignolytic enzymes isolated from a fungus (*Phanerochaete chrysosporium*) (Lundell et al., 2010). The CIIP family encompasses also the so-called Manganese Peroxidases (MnP) and the Versatile Peroxidases (VP), which are probably also involved in lignin

cleavage (Lundell et al., 2010). Recent studies identified ancestral Class II Peroxidases, such as the Generic Peroxidases (GP) and "*Coprinopsis cinerea* Peroxidases" (CiP) (Zamocky et al., 2015; Liers et al., 2013). These ancestral enzymes, exhibit a low redox potential, and therefore may not be active in lignin degradation (Riley et al., 2014).

1. POX + H_2O_2 -> POX-Comp I + H_2O

2. POX-Comp I + A -> POX-Comp II + $A^{\bullet}$

3. POX-Comp II + A -> POX + $A^{\bullet}$ + H_2O

Formula 01 : General heme peroxidase (POX) catalytic cycle (Lundell et al., 2010)

Lignin Peroxidases (LiP)

LiPs (EC 1.11.1.14) are exclusive of white rot fungi (WRF) that belongs to the Polyporales order (Ruiz-Dueñas et al., 2013). Many species possess several LiP genes in their genomes (Floudas et al., 2012; Nagy et al., 2017). LiPs are glycoproteins whose molecular mass ranges from 38 to 43 kDa and are preferentially active at acidic pH (Lundell et al., 2010). Thanks to their high redox potential LiPs are able to oxidize phenolic and non-phenolic compounds (including the phenolic and non-phenolic moieties of lignin), amines, aromatic ethers, and polycyclic aromatics (Martinez et al., 2009).

Manganese Peroxidases (MnP)

MnPs (EC 1.11.1.13) are enzymes similar to LiPs but with higher molecular masses (45 - 60 kDa) (Perez et al., 2002). MnPs have a typical peroxidase catalytic cycle, but with Mn^{2+} as substrate, oxidizing Mn^{2+} to Mn^{3+} (Lundell et al., 2010). MnPs oxidizes only phenolic compounds and, in the case of lignin, its phenolic moiety, as opposed to LiPs, that are able to degrade the whole structure of lignin. MnP generates phenoxy-radicals, starting lignin depolymerization (Ruiz-Dueñas & Martinez 2009). *In vitro* studies highlighted the possibility to degrade non-phenolic lignin model compounds in the presence of Mn^{2+} via peroxidation of unsaturated lipids (Perez et al., 2012).

Versatile Peroxidases (VP)

Versatile peroxidases (VP, EC 1.11.1.16) are key enzymes in lignin degradation, they display both manganese and lignin peroxidase activities (Scheibner et al., 2008). VP can oxidize hydroquinone in the absence of exogenous H_2O_2 when Mn^{2+} is present in the medium (Martinez et al., 2009). It has been suggested that the chemical oxidation of hydroquinones promoted by Mn^{2+} could be important during the initial steps of wood biodegradation because ligninolytic enzymes are too large to penetrate into non-modified wood cell walls (Perez et al., 2002).

1.8.4 DyP-type Peroxidases

DyP-type Peroxidases (DyP) (EC 1.11.1.19) have been discovered relatively recently. In 1999 the first enzyme was described from the white-rot basidiomycete *Bjerkandera adusta*. Today, 20 years later, less than ten fungal DyP enzymes have been characterized to some extent (Liers et al., 2015; Lauber et al., 2014). Conversely, thanks to the development of fungal genomic programs, a far larger number of sequences homologous to characterized DyPs are available in the public databases (Grigoriev et al., 2014).

DyPs are versatile enzymes active on several artificial and natural substrates (Singh & Eltis 2015). They are have been named (DyP = "Dye decolorizing peroxidases") because of their capacity to de-colorize various dyes (including the recalcitrant anthraquinone-based industrial dyes) (Kim et al., 1995); but they can also perform the oxidation of phenolic compounds as described for class II peroxidases (Kim & Shoda 1999). They can also catalyze peculiar reactions as β-carotene cleavage that has led to their utilization in the dairy industry for the removal of carotene dyes (Scheibner et al., 2008). Some DyPs have also been shown to oxidase non-phenolic methoxylated aromatics and Mn^{2+} (Singh & Eltis 2015). In addition to the wide spectrum of reactions that they can catalyze, DyPs appear to be very stable under high pressure and temperatures (Puhse et al., 2009) and they have a high redox potential

(1.1–1.2V) (Liers et al., 2013). For these reasons DyP-type peroxidases are interesting for several biotechnological applications (Husain 2006; Sugano et al., 2007; Scheibner et al., 2008; Szweda et al., 2013; Colpa et al., 2014). Finally, it has been suggested, especially in the case of bacteria, that DyPs could also participate to lignin degradation (Liers 2010; Ruiz-Duenas & Martinez 2010).

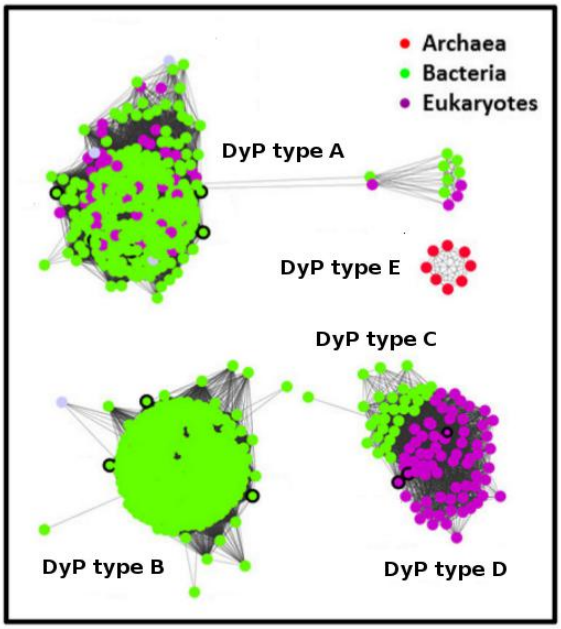


Figure 11 - Network analysis of the DyP family showing its division into subfamilies: Groups of protein sequences with ≥70% pairwise identity are represented by dots colored according to the domain of life of their host organism. The DyP family forms 4 subgroups, one of them hosts the two enzymatic subfamilies DyP type C and DyP type D.

DyPs are present in several domains and kingdoms of life: Bacteria, Archaea, Fungi and lower Eukaryotes (Amoebozoa, Excavata,...) (Zamocky et al., 2015). They may originate from bacterial from which they spread to other taxonomic groups via several Horizontal Gene Transfer (HGT) events. (Zamocky et al., 2014; Zamocky et al., 2015). Phylogenetic analysis of the global DyP family identified four distinct phylogenetic groups (Sugano et al., 2009; Zamocky et al., 2015). The ancestral clade is probably the DyP-type A subfamily, from which evolved the DyP-type C subfamily and the DyP-type D subfamily clades. The DyP-type B subfamily clade seems to be the result of a distinct evolutionary process, this latter clade is now the most diversified

DyP subfamily (Sugano 2009). A fifth DyP subfamily is the DyP-type E subfamily. DyP-type E were recently discovered in Archaea and several pathogenic Bacteria. It has been suggested that they could be involved in stress response (Singh & Eltis, 2015). Recent analyses on the protein structure suggested to define only 3 families within the DyP family by merging the DyP-type C with the DyP-type D subfamilies (Yoshida & Sugano 2015).

Most probably, three different HGT events permitted the passage of DyPs from bacterial to eukaryotic genomes. Two events are recognizable within the DyP-type B subfamily clade, where are present sequences from Ascomycota, but also from protist species (Zamocky et al., 2015). The third HGT events probably generated the DyP-type D subfamily clade that is specific to filamentous fungi from the two phyla Ascomycota and Basidiomycota.

1.8.5 Unspecific Peroxygenases

Unspecific Peroxygenases (or UPOs) are also often cited as putative fungal ligninolytic enzymes (Hofrichter et al., 2010; Lier et al., 2013; Hofrichter & Ullrich 2014; Kellner et al., 2014; Zamocky et al., 2015). This recently discovered group of heme-containing enzymes is interesting because of its high redox potential (Hofrichter et al., 2010). As in the case of the DyP-type Peroxidases, they are principally studied from biotechnological and biochemical point of views, and almost nothing is known about their role in fungal biology and in natural ecosystems.

Unspecific Peroxygenases (EC 1.11.2.1) were described for the first time in 2004 from *Agrocybe aegerita* (Basidiomycota) (Ullrich et al., 2004), but only later they were contextualized as independent enzymatic family (Hofrichter & Ullrich 2011), being subjected to frequent change of nomenclature (APO, HTP, CPO) (Zamocky et al., 2015).

$$\text{A-H} + \text{ROOH} \rightarrow \text{A-OH} + \text{R-OH}$$

Formula 02 - UPO typical reaction: A-H is the substrate, ROOH is a hydroperoxide, R represent an organic substituent or a hydrogen atom, A-OH is the hydroxylated product and R-OH is the

reduced hydroperoxide (or H_2O).

Putative UPO sequences are widespread in fungal genomes where more than 2000 sequences have been identified (H. Kellner, personal communication). The largest part of these sequences comes from the Dikarya, but they are present also among Zygomycota, Glomeromycota and the Stramenopiles (an horizontal gene transfer from the fungi is probably at the origin of their presence in this latter taxon) (Zamocky et al., 2015). Recent analysis revealed that peroxygenases can be classified as “short peroxygenases” or “long peroxygenases”, based on protein length (Hofrichter et al., 2015). Short UPO have an average molecular mass of 29 kDa. Short-UPO (or group I) are present in all fungal phyla, while the long-UPO (or group II) with an average molecular mass of 44 kDa and a di-sulfide bridge is present only in the Dikarya (Kellner et al., 2014).

UPOs present peroxidase and/or peroxygenase activities and are active on numerous substrates. They can oxidize halides and aryl alcohols, aromatic, heterocyclic and aliphatic substrates with an elevated unique oxygen atom transfer potential (Hofrichter et al., 2015). This characteristic is at the origin of the research interest expressed on this family of enzymes.

UPO enzymes show a high catalytic versatility. The wide range of substrates they can modify suggest an involvement in detoxification processes (i.e. detoxification of plant phytoalexins, microbial toxins, xenobiotic compounds). Their involvement in lignin and humus degradation/modification is a realistic hypothesis as well as the involvement in the biosynthetic pathways of secondary metabolites (Hofrichter et al., 2015). At the moment, the role of UPO enzymes in nature still remains to be defined.

1.9 Gene repertoires and the evolution of trophic modes in the Fungi

We have seen in the previous paragraphs that fungi "interact" in

different ways with plant organic matter and that these different patterns of interaction led to their classification in different, sometime overlapping broad "trophic categories", such as plant pathogens, white rot saprotrophic species, brown and soft rot ones, soil saprotrophs, symbiotic mycorrhizal fungi and so on. We have also seen that these traditional classification schemes are not clear cut, as for example many plant pathogens often behave as endophytic or soil saprotrophs at some stage of their life cycles and that several ectomycorrhizal species can modify organic matter in a way similar to modification by brown rot species. Finally, we have also seen that degradation of organic matter is largely the result of enzymatic attack by sometime seemingly redundant enzyme families and also that the role and genuine function of several enzyme families (e.g. the peroxidases) need to be clarified.

In the last years these questions have been tentatively addressed through phylogenomics, i.e. the comparisons of the gene repertoires of the genomes of different fungal species affiliated to the different trophic categories (Eastwood et al., 2011; Floudas et al., 2012; Kohler et al., 2015; Shah et al., 2016). In the following paragraphs I will concentrates mostly on Agaromycetes genome analyses since this taxonomic group certainly encompasses the largest number of trophic modes, including the white rot species, the only one known to be able to completely degrade lignin enzymatically.

First, the separate analysis of the genomes of saprotrophic species have all shown that they always contain complex repertoires of genes encoding hydrolytic enzymes participating to POM degradation. For example, in the case of the white rot *Phanerochaete chrysosporium* (Polyporales), 31 lignocellulose-degrading oxidoreductases and 75 CAZymes (Martinez et al., 2004). This high number is partly explained by a high degree of redundancy that results from multiple copies in a single genome of genes encoding homologous enzymes (so-called gene families) and also by the presence of genes encoding non homologous proteins with identical or very similar enzyme activities (e.g. both GH5_5 and GH7 endocellulases or both GH10 and GH11 endoxylanases).

The comparative analysis of the genome content of several species classified in different "trophic groups" also highlighted clear-cut differences that could explain these different trophic modes (Nagy et al., 2017). For example, compared to the WR species *P. chrysosporium* whose genome contains several typical Lignin (LiP) and Mn (MnP) class II peroxidases involved in lignin hydrolysis, the genome of the Brown rot (BR) *Postia placenta* (that does not hydrolyze lignin) lacks members of this gene family (Eastwood et al., 2011). More globally, most fungal species characterized by "well defined" trophic modes possess characteristic gene repertoires that distinguish them from species belonging to other trophic groups (Koheler et al., 2015).

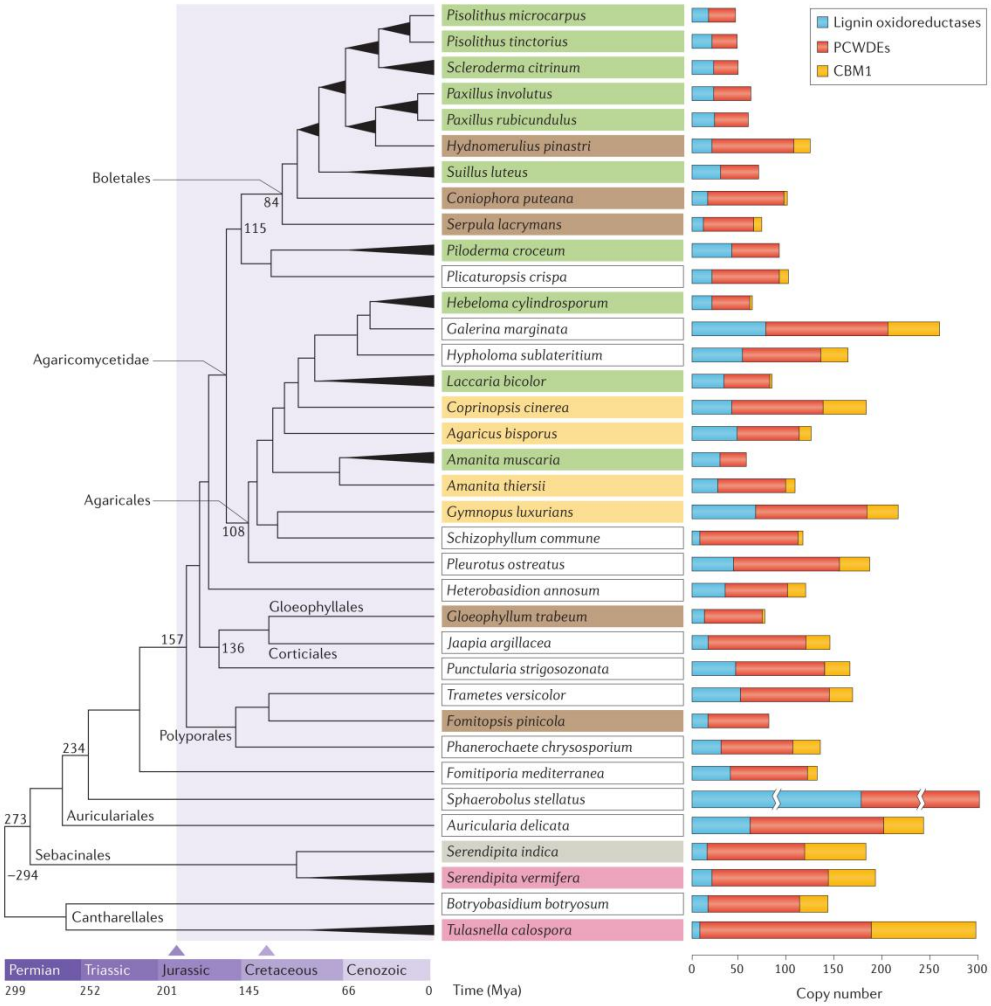


Figure 12 - Evolution of trophic modes in the Basidiomycota: Phylogenomic analysis of the genes

that encode enzymes that function in lignocellulose oxidation (blue bars) the degradation of cellulose and hemicellulose (red bars), and other aspects of plant cell wall degradation (yellow bars). Affiliation to different trophic guilds is polyphyletic, as taxa of ectomycorrhizal fungi (green), orchid mycorrhizal fungi (pink), endophyte (grey), brown-rot fungi (brown), white-rot fungi (white) and soil or litter saprotrophs (yellow) do not always cluster. [Figure adapted Martine et al., 2016].

For example confrontation of the proteomes of several symbiotic ectomycorrhizal (EM) species to the proteomes of saprotrophic WR species showed that the evolutionary transition of the EM phenotype from WR ancestors was always accompanied with a strong reduction, if not complete loss, in both gene numbers and gene families encoding plant cell wall degrading enzymes (Eastwood et al., 2011; Floudas et al., 2012; Nagy et al., 2017). These gene loss events are interpreted in terms of "inutility" of these enzymes for a symbiotic fungus that derives most of its carbon from simple plant sugars and also in term of "dangerousness" of these enzymes that could alter the plant cell wall integrity of the symbiotic plant partner.

Several species and phenotypes however do not fit these models of gene gain/gene loss. For example, typical mycorrhizal symbionts of Ericaceae or of orchids possess in their genomes extended repertoires of POM degrading enzymes (Kohler et al., 2015), that a "blind test" would classify them as typical saprotrophic species (Martin et al., 2016). The sequencing of large numbers of basidiomycete genomes also suggest that the simple white rot/brown rot dichotomy is not as clear-cut as previously thought (Riley et al., 2015). Several species present intermediate gene repertoires that could be interpreted in terms of progressive evolutionary transition to a BR phenotype from a WR ancestor (Floudas et al., 2015).

Recently comparative phylogenomic approaches were also implemented to identify all gene families whose pattern of diversification correlate with the evolution of the WR phenotype in the Basidiomycetes (Nagy et al., 2017). Such an approach recovered, as expected, gene families essential for lignin degradation such as the one encoding class II peroxidases. Other gene families identified through this approach encoded thus far uncharacterized proteins or enzymes whose precise roles in lignin

degradation are still debated. This is the case of the DyP-type peroxidases.

1.10 Objectives of the Thesis

Diversity of fungal DyP-type peroxidases and their potential contribution to organic matter degradation

To degrade plant biomass Fungi secrete numerous extracellular enzymes. Many of them, like the enzymes active on polysaccharides, are well characterized. For others, in particular those active on lignin, our knowledge is much more limited. For several of the enzymes suspected to be active on lignin nothing is known of their role in fungal biology and natural ecosystems and many of these enzymes have essentially been characterized in a context of biotechnology. This is the case of the DyP-type peroxidases. In this context, the objective of my PhD thesis is to address the potential role of fungal DyP peroxidases in organic matter degradation.

Several approaches are in possible to address the roles of any protein. The most obvious and straightforward approaches are the "direct ones". Biochemical and genetic analyses are indeed widely (and preferentially) used to address the functions of specific enzymes. These approaches have however thus far failed to assign a clear biological role to DyPs. Genetic approaches, i.e. the functional analysis of fungal mutant strains deleted of the corresponding genes, are difficult to develop because most fungal species possess several DyP-coding genes and always possess other peroxidase-coding genes thus making difficult the functional analysis of the corresponding mutants.

Therefore, for his thesis we opted for correlative approaches trying to deduce the role of fungal DyP using phylogenetic and ecological information. The phylogenetic approach tries to answer to the question:

- does DyP diversification follows diversification of trophic modes in the fungi? To answer to this specific question we performed a global phylogenetic analysis of the fungal DyP family and in a phylogenenomic

analysis tried to correlate DyP gene diversity to known trophic modes in the Agaromycetes.

The purpose of the ecological approach was to address the question:

- does DyP phylogenetic diversity correlates with habitat diversity? We tested the hypothesis that, if fungal DyPs are essential in lignin degradation, we may observe a higher abundance and diversity of DyP genes in the (meta)transcriptomes of fungal communities present in habitats characterized by high lignin contents and degradation rates. As prerequisites to the implementation this approach we first developed specific protocols to isolate nucleic acids (RNA) from specific environments (wood) and then developed protocols to isolate full-length DyP cDNAs from fungal metatranscriptomes.

1.11 Bibliography

Adams JM, Faure H, Faure-Denard L, McGlade JM, Woodward FI. 1990. Increases in terrestrial carbon storage from the Last Glacial Maximum to the present. *Nature* **348**: 711–714.

Agerer R. 2001. Exploration types of ectomycorrhizae: A proposal to classify ectomycorrhizal mycelial systems according to their patterns of differentiation and putative ecological importance. *Mycorrhiza* **11**: 107–114.

Anderson DW. 1988. The effect of parent material and soil development on nutrient cycling in temperate ecosystems. *Biogeochemistry* **5**: 71–97.

Arantes V, Jellison J, Goodell B. 2012. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied Microbiology and Biotechnology* **94**: 323–338.

Aro N, Pakula T, Penttilä M. 2005. Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiology Reviews* **29**: 719–739.

Atta-Krah K (Kwesi), Sanginga N, Tripathi BR, Psychas PJ. 1992. *The AFNETA alley farming training manual*. Alley Farming Network for Tropical Africa.

Baldrian P. 2008. Wood-inhabiting ligninolytic basidiomycetes in soils: Ecology and constraints for applicability in bioremediation. *Fungal Ecology* **1**: 4–12.

Baldrian P. 2017. Forest microbiome: Diversity, complexity and dynamics. *FEMS Microbiology Reviews* **41**: 109–130.**Baldrian P, Lindahl B. 2011**. Decomposition in forest ecosystems: After decades of research still novel findings. *Fungal Ecology* **4**: 359–361.

Baldrian P, Valášková V. 2008. Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiology Reviews* **32**: 501–521.

Baldrian P, Voříšková J, Dobiášová P, Merhautová V, Lisá L, Valášková V. 2011. Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi

from the upper layers of forest soil. *Plant and Soil* **338**: 111–125.

Baldrian P, Zrůstová P, Tláskal V, Davidová A, Merhautová V, Vrška T. 2016. Fungi associated with decomposing deadwood in a natural beech-dominated forest. *Fungal Ecology* **23**: 109–122.

Bardgett R. 2010. *The Biology of Soil: A community and ecosystem approach*. Oxford University Press.

Bardgett RD, Van Der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. *Nature* **515**: 505–511.

Barnard RL, Osborne CA, Firestone MK. 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *ISME J* **7**: 2229–2241.

Battaglia E, Benoit I, van den Brink J, Wiebenga A, Coutinho PM, Henrissat B, de Vries RP. 2011. Carbohydrate-active enzymes from the zygomycete fungus Rhizopus oryzae: A highly specialized approach to carbohydrate degradation depicted at genome level. *BMC Genomics* **12**: 38.

Beck A, Peršoh D, Rambold G. 2014. First evidence for seasonal fluctuations in lichen- and bark-colonising fungal communities. *Folia Microbiologica* **59**: 155–157.

Berg B, McClaugherty C. 2003. Decomposition of fine root and woody litter. In: Plant Litter. Berlin, Heidelberg: Springer Berlin Heidelberg, 173–189.

Bhadra B, Rao RS, Singh PK, Sarkar PK, Shivaji S. 2008. Yeasts and yeast-like fungi associated with tree bark: Diversity and identification of yeasts producing extracellular endoxylanases. *Current Microbiology* **56**: 489–494.

Blaser S, Prati D, Senn-Irlet B, Fischer M. 2013. Effects of forest management on the diversity of deadwood-inhabiting fungi in Central European forests. *Forest Ecology and Management* **304**: 42–48.

Boberg J. 2009. *Litter decomposing fungi in boreal forests: Their function in carbon and nitrogen circulation*.

Boddy L, Hynes J, Bebber DP, Fricker MD. 2009. Saprotrophic cord systems: Dispersal mechanisms in space and time. *Mycoscience* **50**: 9–19.

Boddy L, Watkinson SC. 1995. Wood decomposition, higher fungi, and their role in nutrient redistribution. *Canadian Journal of Botany* **73**: 1377–1383.

Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* **54**: 519–546.

Brown ME, Chang MCY. 2014. Exploring bacterial lignin degradation. *Current Opinion in Chemical Biology* **19**: 1–7.

Brundrett M. 2004. Diversity and classification of mycorrhizal associations. *Biological Reviews* **79**: 473–495.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* **28**: 1883.

Burns RG, DeForest JL, Marxsen J, Sinsabaugh RL, Stromberger ME, Wallenstein MD, Weintraub MN, Zoppini A. 2013. Soil enzymes in a changing environment: Current knowledge and future directions. *Soil Biology and Biochemistry* **58**: 216–234.

Cavalier-Smith T, Chao EE-Y. 1995. The Opalozoan Apusomonas is Related to the Common Ancestor of Animals, Fungi, and Choanoflagellates. *Proc R Soc Lond B Biol Sci* **261**: 1–6.

Colpa DI, Fraaije MW, Van Bloois E. 2014. DyP-type peroxidases: A promising and versatile class of enzymes. *Journal of Industrial Microbiology and Biotechnology* **41**: 1–7.

Conant RT, Paustian K, Elliott ET. 2009. Grassland Management and Conversion into Grassland : Effects on Soil Carbon. *Ecological Society of America* **11**: 343–355.

Cornwell WK, Cornelissen JHC, Allison SD, Bauhus J, Eggleton P, Preston CM, Scarff F, Weedon JT, Wirth C, Zanne AE. 2009. Plant traits and wood fates across the globe: Rotted, burned, or consumed? *Global Change Biology* **15**: 2431–2449.

Cragg SM, Beckham GT, Bruce NC, Bugg TDH, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison J, McGeehan JE, *et al.* 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology* **29**: 108–119.

Davidson EA, Janssens IA. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**: 165–173.

Davin LB, Lewis NG. 2005. Lignin primary structures and dirigent sites. *Current Opinion in Biotechnology* **16**: 407–415.Atta-Krah K (Kwesi), Sanginga N, Tripathi BR, Psychas PJ. 1992. The AFNETA alley farming

Deacon J. 2013. *Fungal Biology: 4th Edition*. Malden, MA USA: Blackwell Publishing Ltd.

Deacon LJ, Janie Pryce-Miller E, Frankland JC, Bainbridge BW, Moore PD, Robinson CH. 2006. Diversity and function of decomposer fungi from a grassland soil. *Soil Biology and Biochemistry* **38**: 7–20.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, *et al.* 2011. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science* **333**: 762–765.

Eichlerová I, Homolka L, Žifčáková L, Lisá L, Dobiášová P, Baldrian P. 2015. Enzymatic systems involved in decomposition reflects the ecology and taxonomy of saprotrophic fungi. *Fungal Ecology* **13**: 10–22.

Ercole E, Adamo M, Rodda M, Gebauer G, Girlanda M, Perotto S. 2015. Temporal variation in mycorrhizal diversity and carbon and nitrogen stable isotope abundance in the wintergreen meadow orchid Anacamptis morio. *New Phytologist* **205**: 1308–1319.

Eriksson K-EL, Blanchette RA, Ander P. 1990. Biodegradation of Cellulose. In: Springer, Berlin, Heidelberg, 89–180.

Fawal N, Li Q, Savelli B, Brette M, Passaia G, Fabre M, Mathé C, Dunand C. 2013. PeroxiBase: A database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Research* **41**: D441–D444.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science* **336**: 1715–1719.

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

Fukasawa Y, Osono T, Takeda H. 2011. Wood decomposing abilities of diverse lignicolous fungi on nondecayed and decayed beech wood. *Mycologia* **103**: 474–482.

Gao Q, Reynolds JF. 2003. Historical shrub-grass transitions in the northern Chihuahuan Desert: Modeling the effects of shifting rainfall seasonality and event size over a landscape gradient. *Global Change Biology* **9**: 1475–1493.

Giordano L, Gonthier P, Varese GC, Miserere L, Nicolotti G. 2009. Mycobiota inhabiting sapwood of healthy and declining Scots pine (Pinus sylvestris L.) trees in the Alps. *Fungal Diversity*: 69–83.

Gramss G, Voigt KD, Bergmann H. 2005. Factors influencing water solubility and plant availability of mineral compounds in the tripartite fairy rings of Marasmius oreades (BOLT.: FR.) FR. *Journal of Basic Microbiology* **45**: 41–54.

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–D704.

Hammel KE, Cullen D. 2008. Role of fungal peroxidases in biological ligninolysis. *Current Opinion in Plant Biology* **11**: 349–355.

Hättenschwiler S, Tiunov A V., Scheu S. 2005. Biodiversity and Litter Decomposition in Terrestrial Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **36**: 191–218.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, *et al.* 2007. A higher-level phylogenetic classification of the Fungi. *Mycological Research* **111**: 509–547.

Hibbett D, Blanchette R, Kenrick P, Mills B. 2016. Climate, decay, and the death of the coal forests. *Current Biology* **26**: R563–R567.

Hiscox J, Savoury M, Johnston SR, Parfitt D, Müller CT, Rogers HJ, Boddy L. 2016. Location, location, location: priority effects in wood decay communities may vary between sites. *Environmental Microbiology* **18**: 1954–1969.

Hiscox J, Savoury M, Müller CT, Lindahl BD, Rogers HJ, Boddy L. 2015. Priority effects during fungal community establishment in beech wood. *The ISME Journal* **9**: 2246–2260.

Hofrichter M, Kellner H, Pecyna MJ, Ullrich R. 2015. Fungal unspecific peroxygenases: Heme-thiolate proteins that combine peroxidase and cytochrome P450 properties. *Advances in Experimental Medicine and Biology* **851**: 341–368.

Hofrichter M, Ullrich R. 2011. New Trends in Fungal Biooxidation. *The Mycota X, Industrial Applications*: 425–444.**Hofrichter M, Ullrich R. 2014**. Oxidations catalyzed by fungal peroxygenases. *Current Opinion in Chemical Biology* **19**: 116–125.

Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T. 2010. New and classic families of secreted fungal heme peroxidases. *Applied Microbiology and Biotechnology* **87**: 871–897.

Högberg MN, Högberg P. 2002. Extramatrical ectomycorrhizal mycelium contributes one-third of microbial biomass and produces, together with associated roots, half the dissolved organic carbon in a forest soil. *New Phytologist* **154**: 791–795.

Högberg P, Nordgren A, Buchmann N, Taylor AFS, Ekblad A, Högberg MN, Nyberg G, Ottosson-Löfvenius M, Read DJ. 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature* **411**: 789–792.

Hoppe B, Kahl T, Karasch P, Wubet T, Bauhus J, Buscot F, Krüger D. 2014. Network analysis reveals ecological links between N-fixing bacteria and wood-decaying fungi. *PLoS ONE* **9**.

Hu Z, Xu C, McDowell NG, Johnson DJ, Wang M, Luo Y, Zhou X, Huang Z. 2017. Linking microbial community composition to C loss rates during wood decomposition. *Soil Biology and Biochemistry* **104**: 108–116.Atta-Krah K (Kwesi), Sanginga N, Tripathi BR, Psychas PJ. 1992. The AFNETA alley farming

Husain Q. 2006. Potential applications of the oxidoreductive enzymes in the

decolorization and detoxification of textile and other synthetic dyes from polluted water: A review. *Critical Reviews in Biotechnology* **26**: 201–221.

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, *et al.* 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**: 818–822.

Jeffery S, Gardi C, Jones A, Montanarella L, Marmo L, Miko L, Ritz K, Peres G, Römbke J, van der Putten WH. 2010. *European Atlas of soil biodiversity*. European Commission.

Johnston SR, Boddy L, Weightman AJ. 2016. Bacteria in decomposing wood and their interactions with wood-decay fungi. *FEMS microbiology ecology* **92**: 1–12.

Keegstra K. 2010. Plant Cell Walls. *PLANT PHYSIOLOGY* **154**: 483–486.

Kellner H, Luis P, Pecyna MJ, Barbi F, Kapturska D, Kruger D, Zak DR, Marmeisse R, Vandenbol M, Hofrichter M. 2014. Widespread occurrence of expressed fungal secretory peroxidases in forest soils. *PLoS ONE* **9**.

Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* **389**: 33–39.

Kim SJ, Ishikawa K, Hirai M, Shoda M. 1995. Characteristics of a newly isolated fungus, Geotrichum candidum Dec 1, which decolorizes various dyes. *Journal of Fermentation and Bioengineering* **79**: 601–607.

Kirk PM, Cannon PF, Minter DW, Stalpers JA. 2009. *Ainsworth and Bisby's Dictionary of the Fungi (10th edition)*. Emerald Group Publishing Limited.

Knapp AK, Smith MD. 2001. Variation among biomes in temporal dynamics of aboveground primary production. *Science* **291**: 481–484.

Kogel-Knabner I. 2002. ChemInform Abstract: The Macromolecular Organic Composition of Plant and Microbial Residues as Inputs to Soil Organic Matter. *Soil Biology & Biochemistry* **34**: 139–162.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Kowalchuk G, Buma D, de Boer W, Klinkhamer P, van Veen J. 2002. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms RID C-4298-2011 RID C-2737-2011 RID C-3697-2011. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **81**: 509–520.

Kuzyakov Y, Blagodatskaya E. 2015. Microbial hotspots and hot moments in soil: Concept & review. *Soil Biology and Biochemistry* **83**: 184–199.**Lal R. 2004**. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *American Association for the Advancement of Science* **304**: 1623–7.

Lauber C, Schwarz T, Nguyen QK, Lorenz P, Lochnit G, Zorn H. 2017. Identification, heterologous expression and characterization of a dye-decolorizing peroxidase of Pleurotus sapidus. *AMB Express* **7**: 164.

Liers C, Aranda E, Strittmatter E, Piontek K, Plattner DA, Zorn H, Ullrich R, Hofrichter M. 2014. Phenol oxidation by DyP-type peroxidases in comparison to fungal and plant peroxidases. *Journal of Molecular Catalysis B: Enzymatic* **103**: 41–46.

Liers C, Bobeth C, Pecyna M, Ullrich R, Hofrichter M. 2010. DyP-like peroxidases of the jelly fungus Auricularia auricula-judae oxidize nonphenolic lignin model compounds and high-redox potential dyes. *Applied Microbiology and Biotechnology* **85**: 1869–1879.

Liers C, Pecyna MJ, Kellner H, Worrich A, Zorn H, Steffen KT, Hofrichter M, Ullrich R. 2013. Substrate oxidation by dye-decolorizing peroxidases (DyPs) from wood- and litter-degrading agaricomycetes compared to other fungal and plant heme-peroxidases. *Applied Microbiology and Biotechnology* **97**: 5839–5849.

Lilleskov EA, Fahey TJ, Horton TR, Lovett GM. 2011. Belowground Ectomycorrhizal Fungal Community Change Over a Nitrogen Deposition Gradient in Alaska. *Ecology* **83**: 104–115.

Lindahl BD, Tunlid A. 2015. Ectomycorrhizal fungi - potential organic matter decomposers, yet not saprotrophs. *New Phytologist* **205**: 1443–1447.

Liu JK, Hyde KD, Jones EBG, Ariyawansa HA, Bhat DJ, Boonmee S, Maharachchikumbura SSN, McKenzie EHC, Phookamsak R, Phukhamsakda C, *et al.* 2015. Fungal diversity notes 1–110: taxonomic and phylogenetic contributions to fungal species. *Fungal Diversity* **72**.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-5.

Lonsdale D, Pautasso M, Holdenrieder O. 2008. Wood-decaying fungi in the forest: Conservation needs and management options. *European Journal of Forest Research* **127**: 1–22.

Lundell TK, Mäkelä MR, Hildén K. 2010. Lignin-modifying enzymes in filamentous basidiomycetes - Ecological, functional and phylogenetic review. *Journal of Basic Microbiology* **50**: 5–20.

Mäder P, Fließbach A, Dubois D, Gunst L, Fried P, Niggli U. 2002. Soil fertility and biodiversity in organic farming. *Science* **296**: 1694–1697.

Manavalan T, Manavalan A, Heese K. 2015. Characterization of Lignocellulolytic Enzymes from White-Rot Fungi. *Current Microbiology* **70**: 485–498.

Marmeisse R, Kellner H, Fraissinet-Tachet L, Luis P. 2017. Discovering Protein-Coding Genes from the Environment: Time for the Eukaryotes? *Trends in Biotechnology* **xx**: 1–12.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14**: 760–773.

Martinez D, Larrondo LF, Putnam N, Sollewijn Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F, *et al.* 2004. Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. *Nature Biotechnology* **22**: 695–700.

Martínez T, Ruiz-Dueñ As FJ, Jesú S Martínez M, Del Río JC, Gutié Rrez A. 2009. Enzymatic delignification of plant cell wall: from nature to mill. *Current Opinion in Biotechnology* **20**: 348–357.

Morrison-Whittle P, Goddard MR. 2015. Quantifying the relative roles of selective and neutral processes in defining eukaryotic microbial communities. *ISME Journal* **9**: 2003–2011.

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R. 2011. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils (J Gilbert, Ed.). *PLoS ONE* **6**: e17000.

Nagy LG, Riley R, Bergmann PJ, Krizsan K, Martin FM, Grigoriev I V., Cullen D, Hibbett DS. 2017. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Molecular Biology and Evolution* **34**: 35–44.

Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson KH, *et al.* 2016. Comparative genomics of early-diverging mushroom-forming fungi

provides insights into the origins of lignocellulose decay capabilities. *Molecular Biology and Evolution* **33**: 959–970.

Navarrete AA, Tsai SM, Mendes LW, Faust K, De Hollander M, Cassman NA, Raes J, Van Veen JA, Kuramae EE. 2015. Soil microbiome responses to the short-term effects of Amazonian deforestation. *Molecular Ecology* **24**: 2433–2448.

Ochoa-Villarreal M, Aispuro-Hernndez E, Vargas-Arispuro I, Ngel M. 2012. Plant Cell Wall Polymers: Function, Structure and Biological Activity of Their Derivatives. In: Polymerization. InTech.

Opik M, Moora M, Liira J, Koljalg U, Zobel M, Sen R. 2003. Divergent arbuscular mycorrhizal fungal communities colonize roots of Pulsatilla spp. in boreal Scots pine forest and grassland soils. *New Phytologist* **160**: 581–593.

Osono T, Takeda H. 2002. Comparison of litter decomposing ability among diverse fungi in a cool temperate deciduous forest in Japan. *Mycologia* **94**: 421–427.

Otjen L, Blanchette R, Effland M, Leatham G. 1987. Assessment of 30 White Rot Basidiomycetes for Selective Lignin Degradation. *Holzforschung* **41**: 343–349.

Pereira H, Graca J, Rodrigues JC. 2004. Wood Chemistry in Relation to Quality. *ChemInform* **35**: 80–81.

Pérez J, Muñoz-Dorado J, De La Rubia T, Martínez J. 2002. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *International Microbiology* **5**: 53–63.

Pettersen R. 1984. The chemical composition of wood. *The chemistry of solid wood*: 1–9.

Pleszczyńska M, Lemieszek MK, Siwulski M, Wiater A, Rzeski W, Szczodrak J. 2017. Fomitopsis betulina (formerly Piptoporus betulinus): the Iceman's polypore fungus with modern biotechnological potential. *World Journal of Microbiology and Biotechnology* **33**: 83.

Popper ZA, Tuohy MG. 2010. Beyond the Green: Understanding the Evolutionary Puzzle of Plant and Algal Cell Walls. *PLANT PHYSIOLOGY* **153**: 373–383.

Porras-Alfaro A, Herrera J, Natvig DO, Lipinski K, Sinsabaugh RL. 2011. Diversity and distribution of soil fungal communities in a semiarid grassland. *Mycologia* **103**: 10–21.

Purahong W, Kapturska D, Pecyna MJ, Schulz E, Schloter M, Buscot F, Hofrichter M, Kruger D. 2014. Influence of different forest system management practices on leaf litter decomposition rates, nutrient dynamics and the activity of ligninolytic enzymes: A case study from Central European forests. *PLoS ONE* **9**: 1–11.

Purahong W, Pietsch KA, Lentendu G, Schöps R, Bruelheide H, Wirth C, Buscot F, Wubet T. 2017. Characterization of unexplored deadwood mycobiome in highly diverse subtropical forests using culture-independent molecular technique. *Frontiers in Microbiology* **8**.

Rajala T, Peltoniemi M, Hantula J, Mäkipää R, Pennanen T. 2011. RNA reveals a succession of active fungi during the decay of Norway spruce logs. *Fungal Ecology* **4**: 437–448.

Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, *et al.* 2004. Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochemistry Reviews* **3**: 29–60.

Remy W, Hass H, Kerp H. 1994. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 11841–11843.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard

V, Morin E, Otillar R, *et al.* 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences* **111**: 9923–9928.

Rineau F, Roth D, Shah F, Smits M, Johansson T, Canbäck B, Olsen PB, Persson P, Grell MN, Lindquist E, *et al.* 2012. The ectomycorrhizal fungus Paxillus involutus converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. *Environmental Microbiology* **14**: 1477–1487.

Ruiz-Dueñas FJ, Lundell T, Floudas D, Nagy LG, Barrasa JM, Hibbett DS, Martínez AT. 2013. Lignin-degrading peroxidases in Polyporales: an evolutionary survey based on 10 sequenced genomes. *Mycologia* **105**: 1428–1444.

Ruiz-Dueñas FJ, Martínez AT. 2010. Structural and functional features of peroxidases with a potential as industrial biocatalysts. In: Biocatalysis Based on Heme Peroxidases: Peroxidases as Potential Industrial Biocatalysts. 37–59.

Ruiz-Dueñas FJ, Martínez ÁT. 2009. Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microbial Biotechnology* **2**: 164–177.

Sarkar P, Bosneaga E, Auer M. 2009. Plant cell walls throughout evolution: Towards a molecular understanding of their design principles. *Journal of Experimental Botany* **60**: 3615–3635.

Scheibner M, Hülsdau B, Zelena K, Nimtz M, De Boer L, Berger RG, Zorn H. 2008. Novel peroxidases of Marasmius scorodonius degrade β-carotene. *Applied Microbiology and Biotechnology* **77**: 1241–1250.

Schmidt O. 2006. *Wood and Tree Fungi*. Springer Berlin Heidelberg.

Schüßler A, Schwarzott D, Walker C. 2001. A new fungal phylum, the Glomeromycota: Phylogeny and evolution. *Mycological Research* **105**: 1413–1421.

Seibold S, Bässler C, Brandl R, Gossner MM, Thorn S, Ulyshen MD, Müller J. 2015. Experimental studies of dead-wood biodiversity - A review identifying global gaps in knowledge. *Biological Conservation* **191**: 139–149.

Selosse MA, Le Tacon F. 1998. The land flora: A phototroph-fungus partnership? *Trends in Ecology and Evolution* **13**: 15–25.

Shah F, Nicolás C, Bentzer J, Ellström M, Smits M, Rineau F, Canbäck B, Floudas D, Carleer R, Lackner G, *et al.* 2016. Ectomycorrhizal fungi decompose soil organic matter using oxidative mechanisms adapted from saprotrophic ancestors. *New Phytologist* **209**: 1705–1719.

Schneider T, Keiblinger KM, Schmid E, Sterflinger-Gleixner K, Ellersdorfer G, Roschitzki B, Richter A, Eberl L, Zechmeister-Boltenstern S, Riedel K. 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *The ISME Journal* **6**: 1749–1762.

Sindaco R, Savoldelli P, Selvaggi A. 2009. *La Rete Natura 2000 in Piemonte - I Siti di Importanza Comunitaria*. Torino: Regione Piemonte.

Singh R, Eltis LD. 2015. The multihued palette of dye-decolorizing peroxidases. *Archives of Biochemistry and Biophysics* **574**: 56–65.

Sinsabaugh RL. 2010. Phenol oxidase, peroxidase and organic matter dynamics of soil. *Soil Biology and Biochemistry* **42**: 391–404.

Six J, Elliott ET, Paustian K, Doran JW. 1998. Aggregation and Soil Organic Matter

Accumulation in Cultivated and Native Grassland Soils. *Soil Science Society of America Journal* **62**: 1367.

Smith SE. **2003**. Mycorrhizal Fungi Can Dominate Phosphate Supply to Plants Irrespective of Growth Responses. *Plant Physiology* **133**: 16–20.

Smith SE, Read DJ. **1997**. *Mycorrhizal symbiosis*. Academic Press.

Song Z, Kennedy PG, Liew FJ, Schilling JS. **2017**. Fungal endophytes as priority colonizers initiating wood decomposition (A Bennett, Ed.). *Functional Ecology* **31**: 407–418.

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, *et al.* 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028–1046.

Spurgeon DJ, Keith AM, Schmidt O, Lammertsma DR, Faber JH. **2013**. Land-use and land-management change: relationships with earthworm and fungi communities and soil structural properties. *BMC ecology* **13**: 46.

Stenius P. **2000**. Papermaking Science and Technology. Book 3, Forest Products Chemistry. *Helsinki, Fapet oy*.

Stokland JN, Siitonen J, Jonsson BG. **2012**. *Biodiversity in Dead Wood*. Cambridge University Press.

Sugano Y. **2009**. DyP-type peroxidases comprise a novel heme peroxidase family. *Cellular and Molecular Life Sciences* **66**: 1387–1403.

Sugano Y, Sasaki K, Shoda M. **1999**. cDNA cloning and genetic analysis of a novel decolorizing enzyme, peroxidase gene dyp from Geotrichum candidum Dec 1. *Journal of Bioscience and Bioengineering* **87**: 411–417.

Szweda RT, Schmidt K, Zorn H. **2013**. Bleaching of colored whey and milk by a multiple-enzyme system. *European Food Research and Technology* **237**: 377–384.

Thevenot M, Dignac MF, Rumpel C. **2010**. Fate of lignins in soils: A review. *Soil Biology and Biochemistry* **42**: 1200–1211.

Treseder KK, Lennon JT. **2015**. Fungal Traits That Drive Ecosystem Dynamics on Land. *Microbiology and Molecular Biology Reviews* **79**: 243–262.

Tunlid A, Rineau F, Smits M, Shah F, Nicolas C, Johansson T, Persson P, Martin F. **2013**. *Genomics and Spectroscopy Provide Novel Insights into the Mechanisms of Litter Decomposition and Nitrogen Assimilation by Ectomycorrhizal Fungi*. Springer, Berlin, Heidelberg.

Ullrich R, Nuske J, Scheibner K, Spantzel J, Hofrichter M. **2004**. Novel Haloperoxidase from the Agaric Basidiomycete Agrocybe aegerita Oxidizes Aryl Alcohols and Aldehydes. *Applied and Environmental Microbiology* **70**: 4575–4581.

Valette N, Perrot T, Sormani R, Gelhaye E, Morel-Rouhier M. **2017**. Antifungal activities of wood extractives. *Fungal Biology Reviews* **31**: 113–123.

van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. **1998**. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**: 69–72.

van Der Heijden MGA, Bardgett RD, Van Straalen NM. **2008**. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* **11**: 296–310.

Van Der Wal A, Ottosson E, De Boer W. 2015. Neglected role of fungal community composition in explaining variation in wood decay rates. *Ecology* **96**: 124–133.

Vanderwolf KJ, Malloch D, Mcalpine DF, Forbes GJ. 2013. A world review of fungi, yeasts, and slime molds in caves. *International Journal of Speleology* **42**: 77–96.

Watt M, Hugenholtz P, White R, Vinall K. 2006. Numbers and locations of native bacteria on field-grown wheat roots quantified by fluorescence in situ hybridization (FISH). *Environmental Microbiology* **8**: 871–884.

Weng J-K, Chapple C. 2010. The origin and evolution of lignin biosynthesis. *New Phytologist* **187**: 273–285.

Willats WGT, Mccartney L, Mackie W, Knox JP. 2001. Pectin: Cell biology and prospects for functional analysis. *Plant Molecular Biology* **47**: 9–27.

Worrall JJ, Anagnost SE, Zabel RA. 1997. Comparison of wood decay among diverse lignicolous fungi. *Mycologia* **89**: 199–219.

Wu YT, Wubet T, Trogisch S, Both S, Scholten T, Bruelheide H, Buscot F. 2013. Forest Age and Plant Species Composition Determine the Soil Fungal Community Composition in a Chinese Subtropical Forest. *PLoS ONE* **8**.

Xie N, Chapeland-Leclerc F, Silar P, Ruprich-Robert G. 2014. Systematic gene deletions evidences that laccases are involved in several stages of wood degradation in the filamentous fungus Podospora anserina. *Environmental Microbiology* **16**: 141–161.

Yamashita S, Masuya H, Abe S, Masaki T, Okabe K. 2015. Relationship between the decomposition process of coarse woody debris and fungal community structure as detected by high-throughput sequencing in a deciduous broad-leaved forest in Japan. *PLoS ONE* **10**: 1–16.

Yoshida T, Sugano Y. 2015. A structural and functional perspective of DyP-type peroxidase family. *Archives of Biochemistry and Biophysics* **574**: 49–55.

Zak DR, Holmes WE, White DC, Peacock AD, Tilman D. 2003. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology* **84**: 2042–2050.

Zámocký M, Gasselhuber B, Furtmüller PG, Obinger C. 2014. Turning points in the evolution of peroxidase-catalase superfamily: Molecular phylogeny of hybrid heme peroxidases. *Cellular and Molecular Life Sciences* **71**: 4681–4696.

Zámocký M, Hofbauer S, Schaffner I, Gasselhuber B, Nicolussi A, Soudi M, Pirker KF, Furtmüller PG, Obinger C. 2015. Independent evolution of four heme peroxidase superfamilies. *Archives of Biochemistry and Biophysics* **574**: 108–119.

Chapter 2

Evolution and distribution of the Dye decolorizing peroxidase (DyP) gene family in the Fungi suggest participation of these enzymes in both intracellular and extracellular biological processes.

2.0 Premise

Phylogenetic and phylogenomic represent two approaches that can contribute to a better understanding of protein functions in a specific group of organisms. As part of the thesis' objective to decipher the role of fungal DyP peroxidases in fungal biology and their potential contribution to organic matter and more specifically lignin degradation, we first carried out an in-depth analysis of the molecular evolution and distribution of the corresponding DyP coding genes in the Fungi. We also specifically addressed the potential co-evolution of this gene family with specific trophic modes in the Fungi.

A technical output of this study is the establishment of an exhaustive dataset of all fungal DyP coding gene sequences present in public databases. Such a dataset will be used in the fifth chapter of this thesis to design DyP-specific probes for sequence capture and the study of the diversity of expressed DyP genes in different natural environments.

2.1 Introduction

Lignin is a complex plant cell-wall polymer made of different phenylpropanoid monomers that represent about 30% of the organic carbon available in the biosphere (Boerjan et al., 2003). One of its functions in living plants is to protect cell wall polysaccharides (cellulose, hemicellulose) from degradation by hydrolytic enzymes secreted by microbial pathogens. At the ecosystem scale, degradation of dead plant biomass, including lignin, by saprotrophic microorganisms is however required to recycle the carbon and other nutrients (such as nitrogen and phosphorus) trapped in this biomass.

A specific functional group of saprotrophic fungi, known as the white rot fungi (abbreviated WR), is considered as the main group of microorganisms capable of performing enzymatic lignolysis. Phylogenomic studies support the assumption that the WR phenotype is an ancestral character that arose once early in the evolution of the Basidiomycota, possibly at the end of the carboniferous period (Floudas et al., 2012). Subsequently, the WR phenotype was lost several times independently during the course of evolution of the Basidiomycetes to give rise to different lineages of either brown rot (BR) saprotrophic fungi or mutualistic ectomycorrhizal (EM) fungi that associate to plant roots (Koheler et al., 2015; Martin et al., 2016; Nagy et al., 2017). Both BR and EM fungi have lost the capacity of performing enzymatic lignolysis although being capable of modifying lignin structure using oxidizing mechanisms (Rineau et al., 2012; Arantes et al., 2012).

From a mechanistic point of view, biochemical studies have demonstrated that several enzymes belonging to the so-called secreted,

heme-containing, fungal class II peroxidases (E.C. 1.11.1.-) participate directly or indirectly to lignin hydrolysis. Among these class II peroxidases, lignin peroxidases (LiP, E.C. 1.11.1.14) found in the Polyporales and versatile peroxidases (VP, E.C. 1.11.1.16) interact directly with lignin to cleave non-phenolic bonds that link phenylpropanoid monomers. As for the class II Manganese peroxidases (Mn-P, E.C. 1.11.1.13), but also VPs (Salame et al., 2014), they oxidize Mn^{2+} to Mn^{3+} which chelates organic acids and cleaves the phenolic structures of lignin. A recent report also indicates that Mn-Ps, in combination with specific organic acids, can cleave non phenolic bonds (Fernández-Fueyo et al., 2015). The essential role of class II peroxidases in lignolysis is further supported by the observation that diversification of this protein family correlates with diversification of WR fungi and that the transition from a WR to a BR or EM trophic mode almost always goes along with a loss or a sharp contraction of this gene family in the fungal genomes.

Despite the acknowledged contribution of class II peroxidases in lignin degradation, their exclusive role in this process is debated. On the one hand substantial lignin degradation in specific ecosystems seems to occur in the absence of organisms producing these enzymes (Li et al., 2017) and, on the other hand, other microbial enzymes have been shown to be able to cleave phenolic and/or non-phenolic bonds of artificial lignin substrates or to oxidize Mn^{2+} to Mn^{3+} and therefore to behave similarly to either class II LiPs or Mn-Ps. One such enzyme family is the "Dye decolorizing peroxidase" (DyP) one (pfam No. PF04261, E.C. 1.11.1.19) (Celis & Dubois 2015). These heme-containing peroxidases are unrelated to class II peroxidases and were first identified for their unique capacity at oxidizing high-redox

anthraquinone dyes (Kim et al., 1995), a property which make them good candidates for the remediation of dye-contaminated waste-water, including carotenoides used in the dairy industry (Scheibner et al., 2008). DyPs present a ferredoxin-like fold (Singh & Eltis, 2015), which is unique among peroxidases, and are present in all three domains of life (Bacteria, Archae and Eukarya) as opposed to class II peroxidases that are restricted to the Fungi. However, global phylogenies of the DyP family all show that, with a very few exceptions, fungal sequences are monophyletic and form a well-supported clade that does not include sequences from other taxonomic groups (Sugano et al., 2009; Zamocky et al., 2015). Different level of biochemical information exist for at least 9 fungal DyPs (all from Basidiomycota), including experimental 3D structures for 2 of them (Yoshida et al., 2011; Strittmatter et al., 2013) and identification of amino acids participating to catalysis through site directed mutagenesis and biophysical measurements (Linde et al., 2015a).

The roles of DyPs in fungal biology has never been addressed by forwards genetic approaches and they are sometime presented as participating to detoxification of naturally-occurring soil borne or plant molecules (Valette et al., 2017), possibly because of the nature and range of substrates they can oxidize. Additionally, and for the same reason, in both the Bacteria and the Fungi, they have been repeatedly suspected to participate to ligninolysis (Liers et al., 2010; Celis & Dubois, 2015; Cragg et al., 2015; de Gonzal et al., 2016). In the Fungi, this statement is further supported by the identification of DyPs in the secretome of fungi growing on lignin-rich substrates (Salvachúa et al., 2016) and by the observation that as for class II peroxidases, DyPs family diversification parallels the evolution of WR fungi and regresses in BR and

EM lineages (Nagy et al., 2017).

The objective of the present study was to further evaluate, using phylogenetic and correlative approaches, the potential biological functions of DyPs in the Fungi, including their putative contribution to ligninolysis. The strategy we followed was to establish a comprehensive view of the evolution of the DyP family in the Fungi, using all currently available genomic data, and to confront this data to our current knowledge on DyP catalytic mechanisms and evolution of trophic modes in the Fungi.

2.2 Materials & Methods

2.2.1 Species and sequence datasets

Publically available DyP sequences (**Attachment A**) were extracted from the GenBank, (www.ncbi.nlm.nih.gov/genbank/) and the JGI Mycocosm (https://genome.jgi.doe.gov/programs/fungi/index.jsf; Grigoriev et al., 2014) databases as well as from our own unpublished whole genome sequences from 155 basidiomycete and ascomycete species. DyP protein sequences were identified in public databases using the Blastp algorithm (Altschul et al., 1990) using as queries the protein sequences of two biochemically characterized DyPs from *Auricularia auricula-judae* (accession No. JQ650250) and *Bjerkandera adusta* (accession No. CDN40127.1) and an E-value threshold of 10^{-5}. DyP sequences extraction from the unpublished genomes was done using HMMER (http://hmmer.org/). A HMMER profile was first created starting from a Clustal Omerga (Sievers et al., 2015) alignment of 150 DyP sequences

from the public databases. DyP sequences extracted by HMMER were further confronted to the GenBank-nr database using Blastp. Four bacterial outgroup sequences were selected to be representative of the bacterial gene diversity in the DyP clade C (Zamocky et al., 2015).

When available, Class II Peroxidase (member of the AA2 family in the CAZy database (www.cazy.org; Lombard et al., 2014)) gene copy numbers per fungal genome were obtained from the Mycocosm database. When not already available, Class II Peroxidases were searched for among fungal proteomes by using blastp with proteins AAO47909.1 (Versatile Peroxidase from *Bjerkandera adusta*) and P49012.1 (Lignin Peroxidase from *Phanerochaete chrysosporium*) as query sequences. The AA2 sequences were verified as described above for the DyP sequences extracted via HMMER from the unpublished genomes.

2.2.2 Phylogenetic and network analyses

Only putatively full-length sequences, starting with a methionine and ending with a stop codon, longer than 1000 bp, and without gaps in the regions where amino-acids essential for catalysis have been identified (as described in Linde at al., 2015b), were included in the gene phylogeny. In the case of the *Aspergillus* and the *Penicillium* genera for which the genomes of numerous species were available, only sequences from *Aspergillus oryzae* and *Penicillium italicum* were retained to be included in the phylogenetic analysis. Sequences were aligned using Muscle (Edgar, 2004). ModelFinder (Kalyaanamoorthy et al., 2017), as implemented on the IQ-Tree web Server

(Trifinopoulos et al., 2015), was used to find the best substitution model (LG+R). IQ-Tree web Server was also used to perform a Maximum Likelihood phylogenetic analysis using IQ-Tree (Nguyen et al., 2015). One thousand bootstrap replicates were conducted to assess tree topology.

The species phylogeny was obtained by modifying the publically available tree (Tr106951 - Zhao et al., 2017) downloaded from TreeBase (www.treebase.org/). Species taxonomy was assigned using the MycoBank database (www.mycobank.org). Both gene and species trees were edited using the on-line software iTOL (https://itol.embl.de; Letunic and Bork 2016).

Sequence Similarity Networks (SSN) were computed using the EFI - Enzyme Similarity Tool (Gerlt et al., 2015) starting from the complete set of DyP protein sequences. The lowest alignment score limit for the output file was 52. The full-network output file was visualized with Cytoscape 3.5.1 (Shannon et al., 2013). The network is represented with an "organic" layout and only edges with an alignment score > 100 are visualized.

2.2.3 Prediction of sequence features

Prediction of N-terminal signal peptides for eukaryotic secreted proteins was performed with SignalP (www.cbs.dtu.dk/services/SignalP/; Nielsen 2017) with default settings except for the D-cutoff value which was set at 0.34 for both noTM and TM networks. Other protein features were predicted using tools of the Sequence Manipulation Suite (www.bioinformatics.org/sms2/index.html). This was the case for the prediction of the No. of specific amino acids (Protein Stats) or of the global

protein hydropathy using "GRAVY" for "Grand AVerage of HYdropathy".

The PCA analysis was performed using the prcomp funcion of R and the ggbiplot package version 0.55 (Vu 2011). Kmeans analysis was performed with the kmeans function of R.

The statistical analyses (gene numbers between trophic guilds comparisons) were performed using the non parametric test Mann-Whitney and Dunn's Post-hoc test. The analyses were performed in R with dunn.test package v1.3.4.

2.3 Results

2.3.1 Global distribution of DyP genes in the Fungi

As in November 2017, among the 472 fungal species genomes inspected, DyP coding genes were exclusively identified in the Dikarya (Ascomycota + Basidiomycota) and not in any of the other fungal taxonomic groups (220 genomes inspected). In the Ascomycota, 154 DyP genes were identified in 82 species, all from the Pezizomycotina (40% of the species inspected) and none from the Saccharomycotina and Taphrinomycotina yeast species clades (50 inspected).

In the Basidiomycota DyPs were absent from the early-branching clades of the Ustilagomycotina, and present in the Puccimycotina and Agaromycmycotina. In this latter group 297 sequences were identified in 110 species (23% of inspected species genomes). Overall, Basidiomycota differed from Ascomycota by a higher incidence of species with DyP genes (61,2%

versus 34,6%) and, when present, a higher mean number of sequences per genome (3,7 versus 1,8)(**Figure 01**).

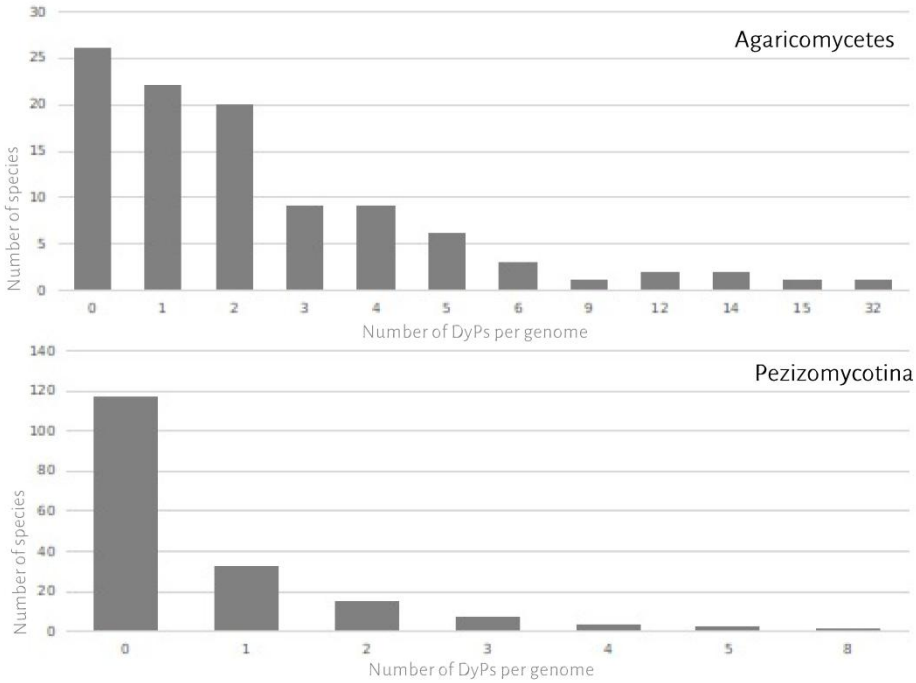


Figure 01 - Distribution of DyP gene copy numbers in the genomes of 177 Pezizomycotina and of 102 Agaromycetes species

2.3.2 Evolution of fungal DyP proteins

All fungal DyP amino acid sequences were used to compute a sequence similarity network, which (give value of "pairwise alignment score") individualized 10 well separated clusters of sequences (**Figure 02**). Following a phylogenetic analysis, each sequence cluster was used to define as many monophyletic clades (**Figure 03**). Only one of these clades/clusters (clade VI.1) encompassed sequences from both Basidiomycota and Ascomycota. On the opposite, an other clade include sequences from "discrete" taxonomic groups.

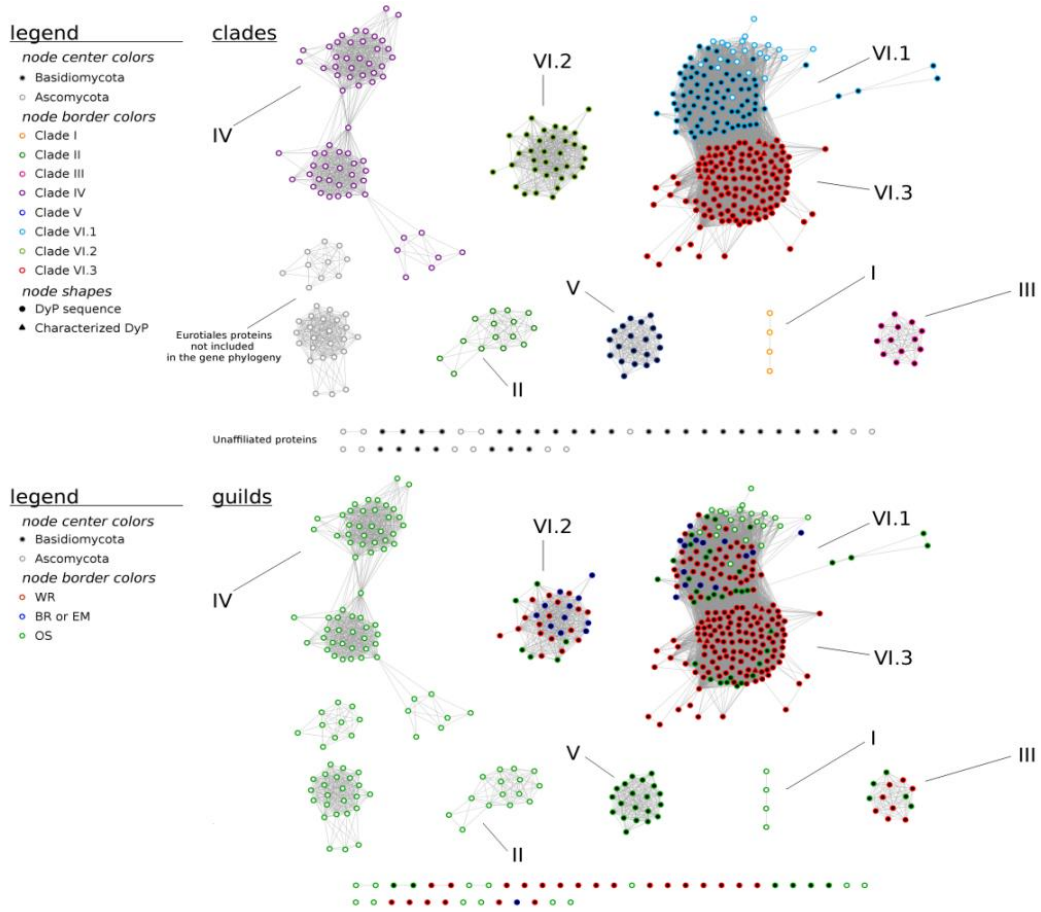


Figure 2 - Fungal DyP protein Sequence Similarity Networks (SSN) used to delineate different clades within this protein family. 452 publically-available identified in 60 Pezizomycotina and 297 Basidiomycota genomes as well as 76 Agaricomycetes DyP proteins characterized from a biochemical point of view were used to construct the SSN composed of two elements. The nodes represent a single DyP sequence and the edges, the lines that connect two or more nodes with a similarity greater than 150 (evaluated as pairwise alignement scores). (A) Visualization of the different individual "clades", each with a different colour code. (B) Distribution of the Agaricomycetes proteins according to the trophic mode of the species they originate from, either white rot (red) or brown rot + ectomycorrhizal (blue).

This is the case of clade V with sequences from the plant pathogens belonging to the Puccimycetes (with the exception of a sequence originating from the Agaricales *Termitomyces* sp.). It could means that the number of Y and W residues is proportional with the sequence length. Y, W and length are independent from the other two variables. pI and GRAVY are inversely correlated between them. Sequences with a signal peptide cluster along the GRAVY vector and with a kmeans analysis we confirmed that GRAVY value and signal peptide presence are correlated. On this base we can affirm that the PCA separated most of the proteins according to the different clades they belong to (**Figure 04**). For example, proteins in clade VI.3 had on average both higher hydropathy indices and lower isoelectric point values compared to proteins from other clades, while proteins from clade III were characterized by both low hydropathy indices and high numbers of tyrosines.

For each protein sequences, we estimated or computed a number of features that either reflect the global protein structure (GRAVY hydropathy index, Isolelectric point, sequence length) or that could be related to catalytic activity as in the case of the number of tyrosine and tryptophan residues that can participate to electron transfer. When incorporated in a PCA analysis tryptophan abundance (W), tyrosine abundance (Y) and sequence length are strictly related.

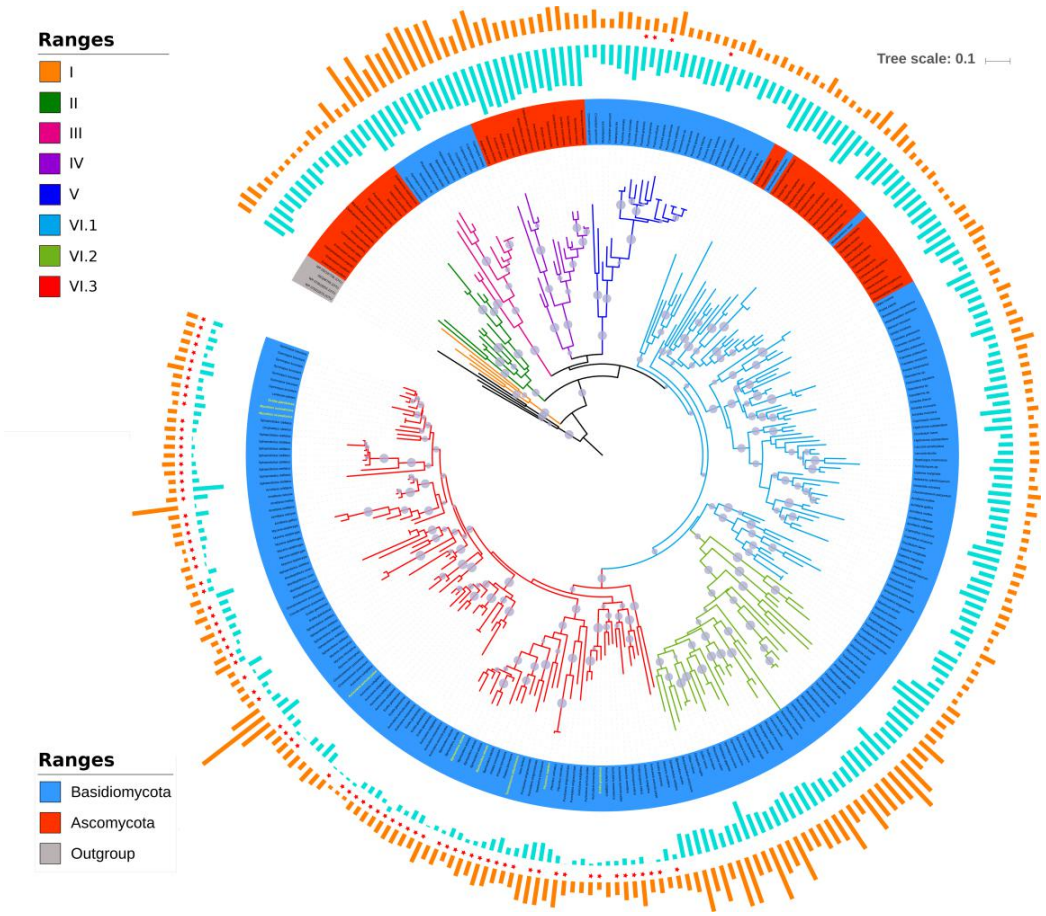


Figure 3 - Phylogenetic relationships (PhyML tree) between 298 publically-available DyP protein sequences identified in 58 Pezizomycotina and 235 Basidiomycota genomes. Sequences from the Agaricomycetes DyP proteins characterized from a biochemical point of view were also included and highlighted in yellow. 4 homologous bacterial sequences were used to root the tree. Grey dots on the branches indicate those that are supported by >0.75 bootstrap values (1000 replicates). Branch colors identify the 8 different fungal DyP clades as delineated in a protein similarity network analysis (Fig. 2). On the margin of the tree are indicated, from the inside to the outside, (i) the name and taxonomic origin of the sequence (either Pezizomycotina or Basidiomycota), (ii) the value of the protein hydropathy (GRAVY) index (light blue bars), (iii) the presence of a signal peptide for secretion on the N-terminal part of the protein (asterisks) and, (iv) the absolute number of tyrosine residues (orange bars).

The presence of a signal peptide for secretion out of the cell was a unique feature of almost all (80,85%) proteins belonging to clade VI.3 as well as of

several of the proteins from the Pucciniales (clade V). In line with this observation, all 9 fungal DyPs studied thus far from a biochemical point of view belonged to clade VI.3. All of these proteins were indeed originally characterized from fungal culture filtrates, which exhibited a so-called "dye-decolorizing" activity.

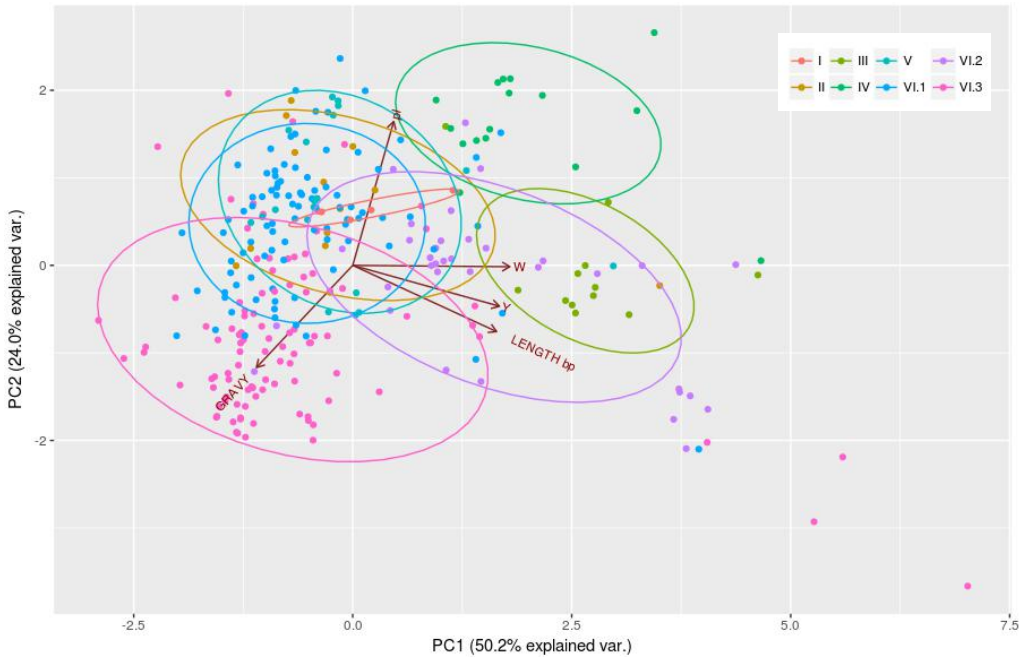


Figure 4 - Principal Component Analysis (PCA): PCA graphical representation reports the PC1 and PC2 axis which together explain the 75% of the variance. Dark red arrows are the eigenvectors. In "SP" graph each dot represent a DyP, when it is blue a signal-peptide were predicted, while pink dots corresond to putative intracellular DyPs. The two DyP distributions are approxymatively indicated with an ellipsoids of the correspondent color. In "clades" graph nodes and ellipsoids colours correspond to the phylogenic clade membership; correspondence is reported in figure.

Inspection of amino-acid residues present at positions essential for catalysis in the AauDyPI enzyme from *Auricularia auricula-judae* (Strittmatter et al., 2015; Linde et al., 2015), highlighted differences between the different clades (**Figure 05**).

Figure 5 - Relative occurrence of different amino acids in DyP proteins belonging to clades I to VI.3 (as defined in Fig. 02) at positions identified as essential for catalysis for the AauDtP from *Auricularia auricula-judae* (GenBank accession No. JQ650250) which belong to clade VI.3. Size of the amino acid one letter code is proportional to its frequency in the corresponding clade. Amino acid are colored according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino acids are black.

Positions in the distal side of the heme pocket are well conserved across all the clades, except in the case of Clade III where the aspartate residue is

replaced by a glycine (position 168 in AauDyPI) one and phenylalanine frequently replaced by valine (pos. 359). In the case of the proximal side of the heme pocket, while the histidine and aspartate residues implicated in Fe2+ ligation are well conserved across clades, the central valine (pos. 253) residue present in the AauDyPI protein is frequently replaced by isoleucine, methionine or phenylalanine residues, the latter two being predominant in clades I and IV, and III, respectively (Fig. xy). Finally, the radical-forming residues tyrosine and tryptophan, present at positions 337 and 377 of the AauDyPI protein are conserved in all clades, except in the clades VI.1 and VI.2 where they are frequently replaced by phenylalanine, another aromatic amino acid.

2.3.3 DyPs and the evolution of trophic modes in the Basidiomycota

Mapping the occurrence of DyP genes on a consensus Basidiomycota species tree (Zhao et al., 2017) suggests that clades VI appeared and diverged in the ancestor of the Agaromycetes, after the separation between the Agaromycetes and the Dacrymycetales/Tremellales (**Figure 06**). Clade III, whose phylogenetic link to clades VI.1, VI.2 and VI.3 cannot be resolved may have appeared latter in the evolution of the Agaromycetes. Reconciliation between species and gene trees suggests that the distribution of DyP genes in extant Agaromycetes species is the consequence of multiple and independent events of gene losses and amplifications that have affected all four major DyP clades present in this taxonomic group during evolution. Consequently, extant Agaromycetes species possess between zero and 32 (*Sphaerobolus*

stellatus) DyP genes which belong to either a single or up to three different clades. Today's most widespread clade is clade VI.1 that has been affected by the lowest number of gene loss events (the most noticeable one being a putative loss in the ancestor of the Boletales) and few events of gene expansion. On the contrary, the evolutionary history of clade VI.3, that encompasses extracellular DyPs, was characterized by more frequent episodes of gene loss but also of rapid expansions, as observed in several of today's species which can possess more than three copies of clade VI.3 DyPs (*Auricularia auricula-judae, Exidia glandulosa, Sphaerolobus stellatus, Bjerkandera adusta, Chodrostereum purpureum, Moniliophtora roreri, Gymnolopus luxurians, Mycena epipterygia*).

For example, the fungus *S. stellatus* processes three paralogous groups of clade VI.3 DyP sequences that encompassed 9, 6, 1 sequences. While several of these paralogous sequences diverged before the separation from the Cantarellales, diversification within each of these paralogous group is recent and seemingly occurred after the split from the Auriculariales.

To evaluate a potential contribution of Agaromycetes DyP genes to organic matter and more specifically to lignin degradation, we compared the prevalence of the different DyP clades in groups of species that differed with respect to their trophic modes ("trophic guilds"). Using literature data and experts' opinions, we defined four trophic modes. "White rot" (WR) species *sensu stricto* were the saprotrophic or pathogenic lignicolous (living on wood) species that degrade lignin and produce a "white rot"; according to Floudas et al., 2012, all of these species possess, often numerous, class II peroxidase genes in their genomes.

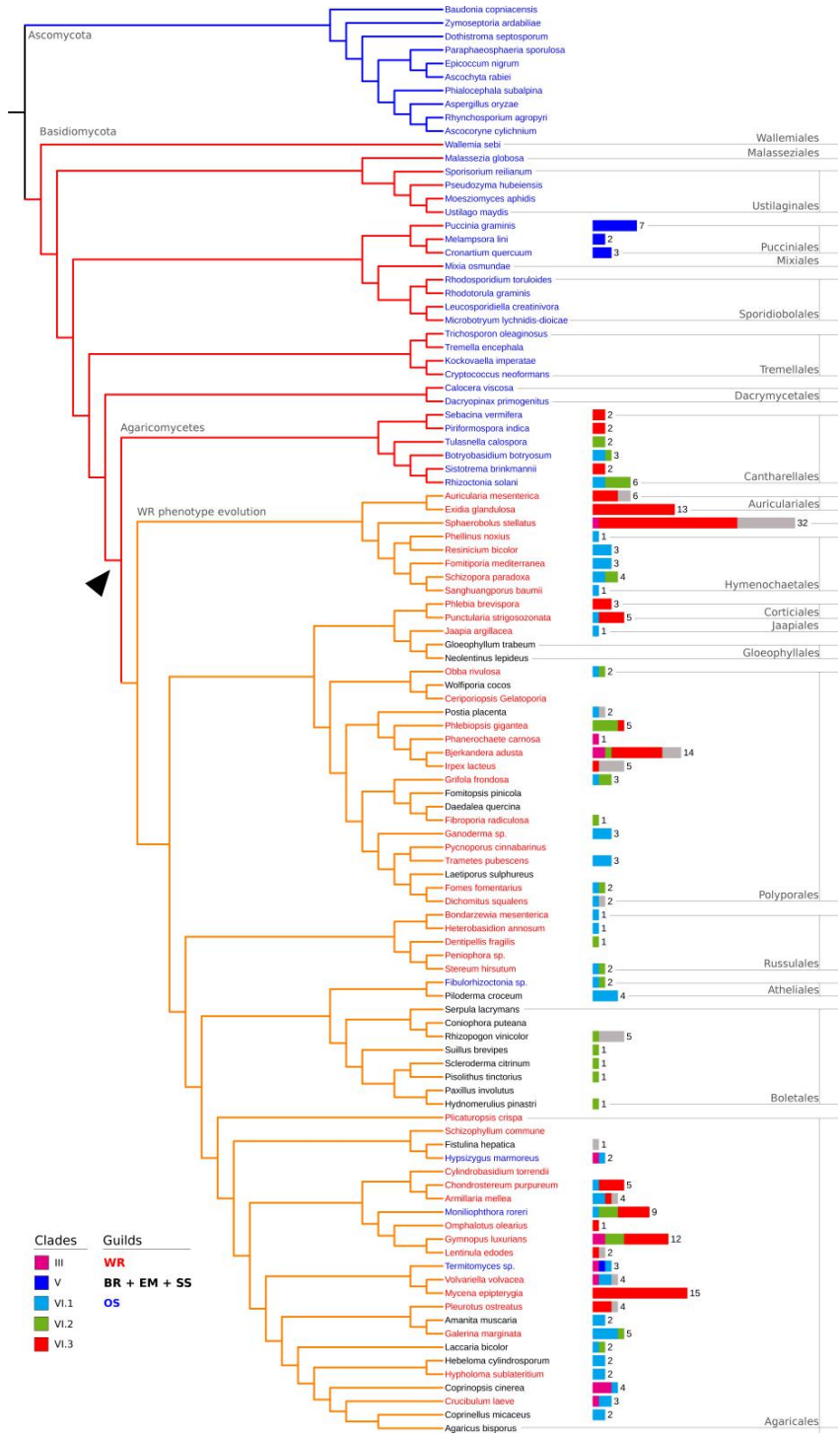


Figure 06 - DyP distribution in different Basidiomycota species with different trophic guild of affinity. Red labels of the tree are the WR species, black labels are species which lost the ability to enzymatically degrade lignin (BR and EM). Blue labels are species of other trophic guilds.

"Brown rot" (BR) species are the saprotrophic lignicolous species producing a brown rot phenotype that results from a selective degradation of the plant cell wall polysaccharides along with a non-enzymatic modification of the lignin polymer. "Ectomycorrhizal" (EM species) are mutualist symbionts of mainly woody plants. Both BR and EM have independently evolved from WR lineages, a transition accompanied by a loss or strong reduction in the number of class II peroxidases in their genomes. Finally, a fourth group named "other saprotrophs" (OS) encompassed all other species that are either litter saprotrophs (e.g. *Agaricus bisporus*), coprophilic species (e.g. *Coprinopsis cinerea*) or endophytic ones (e.g. *Sebacina vermifera*). Because these species do not grow on wood, their modes of degrading the different plant organic polymers do not exactly fit the white rot/brown rot definition although several authors classify several of these species as primarily white rot ones (Eastwood et al., 2011; Nagy et al., 2016).

A first observation from the cluster analysis is that clade VI.3 (putatively extracellular DyPs) encompasses sequences exclusively from WR and OS species and none from BR or EM ones. This distribution is specific to clade VI.3, indeed all other three clades with Agaromycetes sequences (III, VI.1 and VI.2) contained sequences from EM, OS and BR+EM species. Furthermore, although clade VI.3 sequences are absent from nearly two thirds of the species classified as WR (e.g. many Polyporales or the Agaricales *Galerina marginata, Hypholoma sublatericium* or *Crucibulum leave*), Chi2 tests strongly supported ($P<0.01$) the hypothesis that the occurrence of clade VI.3 sequences is more frequent in WR (or WR+OS) species than in BR+EM species (**Table 01**).

		Agaricomycetes			
		WR	BR+MYC	WR+SS	BR + MYC
CLADE VI.1	present	25	7	28	7
	absent	38	11	39	11
	Chi score	*0.0037*		*0.0493*	
	p-value	*0.951563*		*0.824222*	
CLADE VI.2	present	10	11	10	11
	absent	53	7	57	7
	Chi score	*149.189*		*162.698*	
	p-value	*0.000112*		*0.000055*	
CLADE VI.3	present	23	0	23	0
	absent	40	18	44	18
	Chi score	*75.072*		*68.089*	
	p-value	*0.006145*		*0.009071*	

Table 01 - Chi squared test on the occurrence of clades VI.1, VI.2 and VI.3 sequences between WR and others trophic guilds. Presence/absence comparison was done in two different cases. WR vs BR+MYC compare number of present/absent species between WRF and species which lost enzymatic ability to degrade lignin (MYC+BR). WR+SS vs MYC+BR comparison includes in the first category WRF and soil saprotrophs.

On the contrary, while the occurrence of clade VI.1 was similar (P>0.8) among WR (or WR+SS) and BR+EM species, we noted a higher incidence (P<0.01) of clade VI.2 genes in BR+EM species compared to in WR (or WR+SS) ones.

2.4 Discussion

Our analyses confirm previous observations (Linde et al., 2015b) on fungal DyP distribution made using more restricted number of fungal species genomes. In the kingdom Fungi, DyP coding genes of the clade D, as defined by Zamocky et al., 2015, appear to be only present in Dikarya species belonging to the Pezizomycetes in the Ascomycota and the Pucciniales and Agaromycetes in the Basidiomycota. Prevalence, average number of gene copy numbers and DyP phylogenetic diversity are all higher among Agaricomycetes species compared to Pezizomycetes and Pucciniales ones. These data indicate that fungal DyP genes cannot be regarded as so-called "housekeeping genes" but participate either to non-essential functions in fungal biology or to functions that can also be performed by other, non-homologous, genes in species without DyP genes.

The access to a large (more than 450) number of fungal DyP sequences allowed identification of distinct clades that diverged from each other at different periods during the evolution of the Dikarya. In the Agaricomycetes, and to a far lesser extent in the Pezizomycetes, diversification did not parallel taxa diversification since DyP genes belonging to different (up to three) clades can co-exist in the genome of a single extant fungal species. DyP clades, which form distinct clusters in network analyses, can also be distinguished from each other based on "global" protein characteristics (hydropathy indices, isoelectric points, absolute numbers of aromatic residues, as well as the nature of the residues present at key positions in the polypeptide). These different protein features are not necessarily controlled by the protein primary structure, since

non-homologous or distantly related proteins can for example share similar hydropathy indices or isoelectric points. These observations suggest that the divergence between DyPs may have led to significant changes in catalytic properties, substrate range and cellular functions. At present, it is difficult to speculate on the nature of these changes since current functional information on fungal DyPs regards exclusively phylogenetically-related DyP enzymes all belonging to a single clade (VI.2) which seems to correspond to the only – apart from a few DyPs from pathogenic rust fungi – secreted fungal DyPs. Both the position of this clade in the global fungal DyP phylogeny and its exclusive distribution in extant Agaromycetes species suggest that ancestral fungal DyP sequences, putatively inherited from bacteria, were intracellular as are still a majority of extant fungal DyPs. Acquisition of a signal peptide by fungal DyPs may thus represent a recently acquired derived character.

The four main DyP clades (III, VI.1, VI.2, VI.3) present in extant Agaromycetes have thus followed different and apparently independent evolutionary trajectories, being independently eliminated, recruited or amplified in specific lineages. It seems for example that clade VI.3 (extracellular DyPs) experienced more frequent episodes of both gene diversification and loss than clade VI.1. Consequently clade VI.1 is more prevalent in present Agaromycetes species compared to clade VI.3, but with on average lower number of copies per genome (1.5 versus 3.9). Another consequence of these distinct trajectories is that clade VI-I is present in most of the major orders that form the Agaromycetes while current genome sequencing data, suggest that clade VI.3 could have been lost early during the diversification of the Boletales and the Russulales. This conclusion needs

however to be substantiated by additional genome data from these two orders.

Besides the presence/absence and number of genes in specific taxa, DyP genes along with other genes known to encode enzymes active on the main components of plant biomass (e.g. cellulose, hemicellulose, lignin) have been extensively studied with respect to their distribution and abundance in fungal trophic groups that are distributed in several of the major fungal lineages. Several of these studies have more specifically focused on the evolution of the so-called "white rot" (WR) phenotype, a unique evolutionary event that has tentatively been mapped in the Agaromycetes, after the separation of the Cantharellales from the rest of the Agaromycetes. This evolutionary step had important consequences on the terrestrial carbon cycle as it led to the diversification of a unique microbial trophic group capable of hydrolyzing using complex enzyme cocktails all components of the plant biomass, including lignified woody structures. Within the WR fungi, several lineages independently lost the ability to delignify woody plant tissues to become either saprotrophic "Brown rot" (BR) species or ectomycorrhizal (EM) symbionts of plants. Comparative genomic studies confronting the genome contents of WR versus BR or of WR versus EM species highlighted in both cases a drastic reduction, if not a loss, of many extracellular hydrolytic enzyme coding genes in BR and EM lineages. In this general scheme of trophic mode evolution in the Agaromycetes, Nagy et al. (2017) as one of the fungal gene families whose pattern of diversification coincides with the diversification of WR species and which furthermore experienced a contraction in gene copy numbers in BR and EM lineages. While our results

globally support these observations, the separate analysis of the different clades tell a slightly different story. Indeed, of the different fungal DyP clades, only one of them, the one encompassing the secreted ones (clade VI.3) is strictly associated to WR species, although being absent from many WR taxa. Members of the other clades are indifferently distributed in WR, BR, EM and other saprotrophic species. From this observation and the fact that clade VI.3 DyPs are extracellular, it is tempting to hypothesize that these enzymes play a role in the degradation/modification of extracellular plant polymers such as lignin. Their absence in many WR taxa suggests however that they are unessential in this process and could therefore represent only one of the several biochemical "options" that fungi have evolved to degrade lignin. This hypothesis is supported by a low prevalence of clade VI.3 DyPs among WR polyporales, a group of fungi in which has evolved lignin class II peroxidases, a unique class of enzyme active on the non-phenolic moieties of lignin. It cannot however be excluded that both intracellular and extracellular DyPs contribute to the oxidative detoxification of small molecules, such as the constituents of wood extractives as suggested by Valette et al., 2017 or also potentially the products of lignin hydrolysis.

Through the identification of distinct fungal DyP clades, with different evolutionary histories, the present study should stimulate further biochemical studies targeting more precisely putatively intracellular enzymes that have never been characterized thus far. Besides elucidating their potential natural substrate ranges and therefore cellular functions, these intracellular DyPs could also be tested on "non biological" substrates of biotechnological relevance such as recalcitrant industrial dyes. As several extracellular DyPs

have been patented for different applications such as Dye removal from foodstuff (Zorn et al., 2011 - patent US7981636B2) or increased digestibility of biomass (Zorn et al., 2009 - patent WO2009EP58871), it is expected that intracellular DyPs that may be active under different environmental conditions may represent novel industrial targets for novel applications.

2.5 Acknowledges

We would like to acknowledge Dr. Harald Kellner and Enrico Buettner (Technische Universität Dresden) for kindly share of unpublished genomes. Francesco Venice (University of Turin) for the bioinformatic support.

2.6 Bibliography

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Arantes V, Jellison J, Goodell B. 2012. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied Microbiology and Biotechnology* **94**: 323–338.

Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* **54**: 519–546.**Celis AI, Dubois JL. 2015**. Substrate, product, and cofactor: The extraordinarily flexible relationship between the CDE superfamily and heme. *Archives of Biochemistry and Biophysics* **574**: 3–17.

Cragg SM, Beckham GT, Bruce NC, Bugg TDH, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison J, McGeehan JE, *et al.* 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology* **29**: 108–119.

de Gonzalo G, Colpa DI, Habib MHM, Fraaije MW. 2016. Bacterial enzymes involved in lignin degradation. *Journal of Biotechnology* **236**: 110–119.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, *et al.* 2011. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science* **333**: 762–765.**Edgar RC. 2004**. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Fernández-Fueyo E, Linde D, Almendral D, López-Lucendo MF, Ruiz-Dueñas FJ, Martínez AT. 2015. Description of the first fungal dye-decolorizing peroxidase oxidizing manganese(II). *Applied Microbiology and Biotechnology* **99**: 8927–8942.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science* **336**: 1715–1719.

Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1854**: 1019–1037.

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–D704.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Kim SJ, Ishikawa K, Hirai M, Shoda M. 1995. Characteristics of a newly isolated fungus, Geotrichum candidum Dec 1, which decolorizes various dyes. *Journal of Fermentation and Bioengineering* **79**: 601–607.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **44**: W242–W245.

Li H, Yelle DJ, Li C, Yang M, Ke J, Zhang R, Liu Y, Zhu N, Liang S, Mo X, *et al.* 2017. Lignocellulose pretreatment in a fungus-cultivating termite. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 4709–4714.

Liers C, Bobeth C, Pecyna M, Ullrich R, Hofrichter M. 2010. DyP-like peroxidases of the jelly fungus Auricularia auricula-judae oxidize nonphenolic lignin model compounds and high-redox potential dyes. *Applied Microbiology and Biotechnology* **85**: 1869–1879.

Linde D, Pogni R, Cañellas M, Lucas F, Guallar V, Baratto MC, Sinicropi A, Sáez-Jiménez V, Coscolín C, Romero A, *et al.* 2015a. Catalytic surface radical in dye-decolorizing peroxidase: a

computational, spectroscopic and site-directed mutagenesis study. *Biochemical Journal* **466**: 253–262.

Linde D, Ruiz-Dueñas FJ, Fernández-Fueyo E, Guallar V, Hammel KE, Pogni R, Martínez AT. 2015b. Basidiomycete DyPs: Genomic diversity, structural-functional aspects, reaction mechanism and environmental significance. *Archives of Biochemistry and Biophysics* **574**: 66–74.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-5.

Martin F, Kohler A, Murat C, Veneault-Fourrey C, Hibbett DS. 2016. Unearthing the roots of ectomycorrhizal symbioses. *Nature Reviews Microbiology* **14**: 760–773.

Nagy LG, Riley R, Bergmann PJ, Krizsan K, Martin FM, Grigoriev I V., Cullen D, Hibbett DS. 2017. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Molecular Biology and Evolution* **34**: 35–44.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Nielsen H. 2017. Predicting Secretory Proteins with SignalP. In: Humana Press, New York, NY, 59–73.

Rineau F, Roth D, Shah F, Smits M, Johansson T, Canbäck B, Olsen PB, Persson P, Grell MN, Lindquist E, *et al.* 2012. The ectomycorrhizal fungus Paxillus involutus converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. *Environmental Microbiology* **14**: 1477–1487.

Salame TM, Knop D, Levinson D, Mabjeesh SJ, Yarden O, Hadar Y. 2014. Inactivation of a Pleurotus ostreatus versatile peroxidase-encoding gene (mnp2) results in reduced lignin degradation. *Environmental Microbiology* **16**: 265–277.

Salvachúa D, Katahira R, Cleveland NS, Khanna P, Resch MG, Black BA, Purvine SO, Zink EM, Prieto A, Martínez MJ, *et al.* 2016. Lignin depolymerization by fungal secretomes and a microbial sink. *Green Chem.* **18**: 6046–6062.

Scheibner M, Hülsdau B, Zelena K, Nimtz M, De Boer L, Berger RG, Zorn H. 2008. Novel peroxidases of Marasmius scorodonius degrade β-carotene. *Applied Microbiology and Biotechnology* **77**: 1241–1250.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, *et al.* 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539–539.**Singh R, Eltis LD. 2015**. The multihued palette of dye-decolorizing peroxidases. *Archives of Biochemistry and Biophysics* **574**: 56–65.

Strittmatter E, Liers C, Ullrich R, Wachter S, Hofrichter M, Plattner DA, Piontek K. 2013. First crystal structure of a fungal high-redox potential dye-decolorizing peroxidase substrate interaction sites and long-range electron transfer. *Journal of Biological Chemistry* **288**: 4095–4102.

Sugano Y, Sasaki K, Shoda M. 1999. cDNA cloning and genetic analysis of a novel decolorizing enzyme, peroxidase gene dyp from Geotrichum candidum Dec 1. *Journal of Bioscience and Bioengineering* **87**: 411–417.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research* **44**: 232–235.

Valette N, Perrot T, Sormani R, Gelhaye E, Morel-Rouhier M. 2017. Antifungal activities of wood extractives. *Fungal Biology Reviews* **31**: 113–123.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yoshida T, Tsuge H, Konno H, Hisabori T, Sugano Y. 2011. The catalytic mechanism of dye-decolorizing peroxidase DyP may require the swinging movement of an aspartic acid residue. *FEBS Journal* **278**: 2387–2394.

Zámocký M, Hofbauer S, Schaffner I, Gasselhuber B, Nicolussi A, Soudi M, Pirker KF, Furtmüller PG, Obinger C. 2015. Independent evolution of four heme peroxidase superfamilies. *Archives of Biochemistry and Biophysics* **574**: 108–119.

Zhao RL, Li GJ, Sánchez-Ramírez S, Stata M, Yang ZL, Wu G, Dai YC, He SH, Cui BK, Zhou JL, *et al.* 2017. A six-gene phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective. *Fungal Diversity* **84**: 43–74.

Zorn H, Szweda R, Kumar M, Wilms J. 2009. Method for modifying non-starch carbohydrate material using peroxidase enzyme. Patent WO2009EP58871.

Chapter 3

RNA extraction from decaying wood for (meta)transcriptomic analyses

3.0 Premise

The study of microbial diversity and microbial gene expression in environmental samples requires the development of specific protocols to extract nucleic acids (DNA and RNA) from these complex matrices. As part of this thesis to evaluate the role of DyP peroxidase in plant organic matter degradation, we developed a protocol to extract RNA from decaying wood, one of the material with the highest concentration in lignin. This protocol was implemented on a large variety of wood samples and validated by the PCR amplification of different fungal and bacterial protein-coding genes from reverse-transcribed wood RNA.

This protocol was further used in the following two chapters to study the fungal diversity and DyP gene expression by gene capture in wood samples collected in different geographic sites in Italy and France.

Results reported in this chapter have been published in the Canadian Journal of Microbiology: **Adamo M, Voyron S, Girlanda M, Marmeisse R. 2017.** RNA extraction from decaying wood for (meta)transcriptomic analyses. Canadian journal of microbiology **10**: cjm-2017-0230. doi: 10.1139/cjm-2017-0230.

3.1 Introduction

Wood is a structural component of perennial treeplants whose durability, strength, as well as flexibility allows them to reach large dimensions despite being subjected to important mechanical constraints. Because of these characteristics, wood is extensively used as a construction material and for the manufacturing of numerous goods such as furniture.Wood durability is synonymous of delayed and slow decomposition by microorganisms. These characteristics can, in part, be explained by the specific chemical com-position of this material. Wood is essentially composed of the secondary cell walls of dead plant cells and is rich in cellulose, hemicellulose, and impregnated by lignin, a polyphenolic polymer recalcitrant to (bio)chemical at-tack and that restricts the diffusion of microbial hydrolytic enzymes (Pereira et al.,, 2004). Many woods are also impregnated by large quantities of phenolics or other plant secondary metabolites that act as antimicrobial agents (Valette et al.,, 2017). Finally, most woods have an unbalanced chemical composition, being extremely poor in macronutrients such as nitrogen (Cornwell et al.,,2009; Rinne et al.,, 2017), thus representing a harsh environment to microbial life.Despite these characteristics, no wood is truly recalcitrant to microbial attack largely performed by guilds of specifically adapted saprotrophic and (or) pathogenic fungi, whose actions contribute, from an ecological point of view, to the recycling of the organic carbon (Lonsdale et al.,, 2008; Arnstadt et al.,, 2016) trapped in this dead plant biomass (estimated in the range of 50–200 m^3 ha^{-1} in temperate unmanaged forests (Albrecht 1991).

From a mechanistic point of view, wood degradation is largely the result of the action of extracellular hydrolytic enzymes and also of low molecular weight agents (Mn^{3+}, reactive oxygen species) of microbial origin (Blanchette 1995; Arantes et al.,, 2012). Noticeably, several fungal taxa (essentially in the Basidiomycota) secrete complex enzyme cocktails capable of hydrolyzing all structural components of wood, cellulose (by cellulases), hemicellulose (by hemicellulases), and lignin (by so-called class II per-oxidases) (Eastwood et al.,, 2011; Talbot et al.,, 2015). Lignivorous fungal taxa, however, differ from each other by their exact mode of deconstructing wood components (Cragg et al.,, 2015). Schematically, so-called white-rot fungi secrete large amounts of peroxidases that remove lignin, leaving behind a white fibrous material enriched in cellulose (Otjen & Blanchette 1986). As for brown-rot fungi, they presumably produce large amounts of diffusible low molecular weight oxidants that degradepolysaccharides, thus leaving behind a brownish brittle material enriched in chemically modified lignin (Aranteset al.,,, 2012). This functional classification scheme has received experimental support from biochemical studies carried out on simplified lignocellulosic substrates(Ruiz-Dueñas et al.,, 2009; Fernández-Fueyo et al.,, 2012) and from comparative genomic studies that revealed a differential distribution of genes encoding enzymes active on polysaccharides or lignin between brown-rot and white-rot species (Eastwood et al.,, 2011; Floudas et al.,, 2012; Kohler et al.,, 2015). However, extensive sampling of ba-sidiomycete genomes has revealed a high diversity of fungal wood decay mechanisms (Riley et al.,, 2014).Despite this apparent wealth of knowledge on wood degradation by individual species of fungi, the exact mode and pace of wood degradation in the field need to be

clarified. Indeed, under natural conditions, wood degradation is accomplished by not just one but several fun-gal species that act either simultaneously or successively on the same substrate (Rajala et al.,, 2011; Yamashita et al.,, 2015; Purahong et al.,, 2016). Additional points that need to be addressed in wood degradation regard the microbial response to potentially toxic resins and phenolics that accumulate in wood (Valette et al.,, 2017) and adaptive responses to macronutrient limitation (Rajala et al.,, 2011). Finally, a controversial point regards the role of bacteria in this process (Bugg et al.,, 2011; Purahong et al.,, 2014;Johnston et al.,, 2016). Molecular surveys of bacterial communities present in decomposing wood samples demonstrate that they are diversified and differ between plant species and decomposition stages (Purahong et al.,, 2014)although their actual contribution to wood lignocellulose degradation remains uncertain (Johnston et al.,,2016). Metatranscriptomics, i.e., the qualitative and quantitative analysis of the pool of genes expressed by a microbial community, constitutes an experimental approach to access the functions expressed by this community.The functional annotation of metatranscriptomic mRNA (mRNA) sequences gives information on the relativec ontribution of different pathways to a specific biogeochemical process (Damon et al.,, 2012; Baldrian & López-Mondéjar 2014), while the taxonomic annotation of the same mRNAs indicates which microbial taxa con-tribute to this process and suggests potential synergies and (or) functional complementarities between taxa. Thus far, metatranscriptomics has not been implemented on decaying wood. One challenge resides in the development of a reliable protocol to extract pure RNAfrom such a low living biomass substrate rich in polysaccharides and

phenolics, known to be potent inhibitors of many enzymes used in molecular biology, such as reverse transcriptase or DNA polymerases. While several protocols have been published for the extraction of RNAs from woody tissues of living plants (Chang et al.,, 1993; Moser et al.,, 2004; le Provost et al.,, 2007; Gambino et al.,, 2008; Lorenz et al.,, 2010), none of them seems to have been implemented on decaying dead wood. In the present study, we evaluate a new protocol to extract pure, undegraded RNA from wood samples representative of several of the major botanical families that dominate in Northern hemisphere forests.

3.2 Materials & Methods

3.2.1 Plant material

Decomposing wood samples from 22 tree species were collected in different geographic sites in France and Italy (**Figure 01; Table 01**). For each plant species, a variable number of wood pieces of different size classes (from twigs to trunk fragments) and at different stages of decomposition were randomly sampled. In addition to tree-specificwood samples, we also collected 5 forest-specific samples in 5 geographic sites in France and Italy (**Table 01**). Forests in these sites, and therefore the corresponding composite wood samples, were either monospecific (e.g., pure Larix decidua stands at Colle della Lombarda) or mixed-species forests (up to 4 angiosperm dominant tree species at the Mandria Natural Park (Manpa) site). In eachforest, we collected about 100 fragments of decomposingwood along two 20-m-long transects.

Wood fragments were brought to the laboratory and after removing bark fragments, wood was reduced to sawdust using a stainless steel grater. For each plant species or forest, all individual samples were mixed together in equal quantities and 0.1 g (fresh mass) aliquots of the resulting composite samples were placed in 2ml micro-centrifuge tubes, frozen in liquid nitrogen, and stored at −70 °C until RNA extraction. Wood water content was estimated after drying the samples for 24 h at 70 °C (**Table 01**).

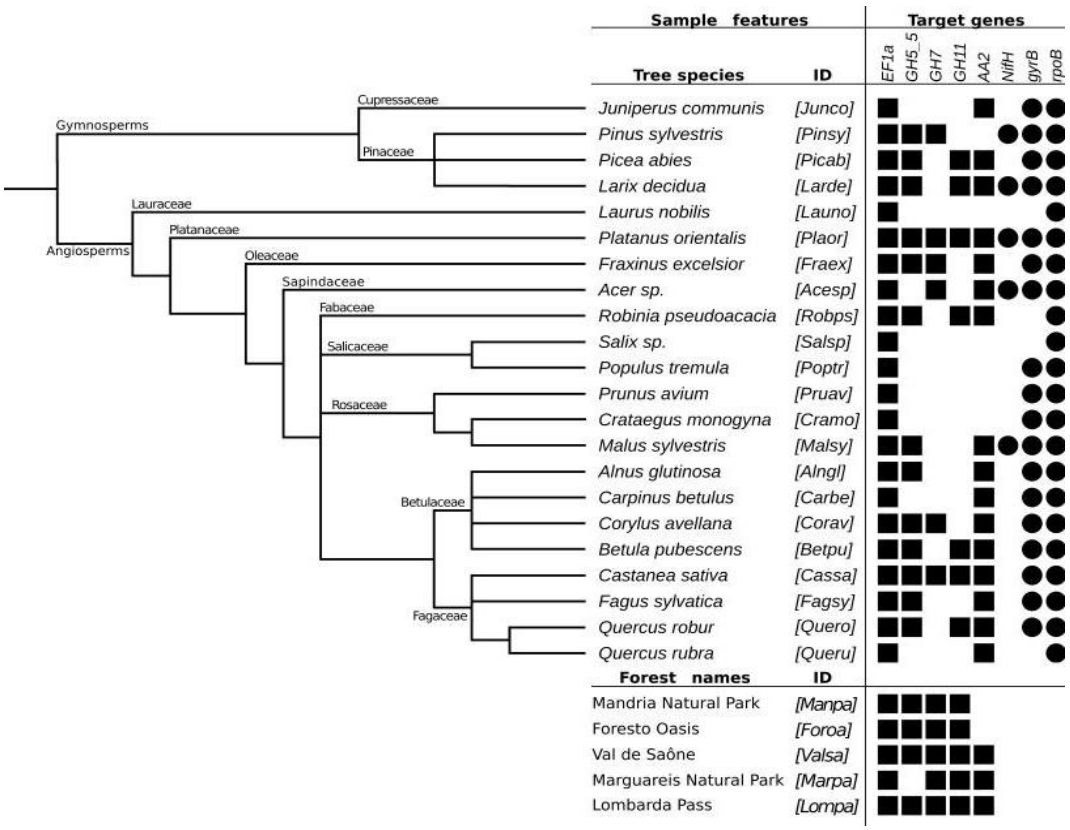


Figure 01: Taxonomic position of the tree species from which decaying wood RNA was extracted and outcome of the PCRs performed on single-stranded (tree-specific RNA extracts) or double-stranded cDNA (forest-specific extracts) to amplify 5 eukaryote-specific (EF1α , GH5_5, GH7, GH11, AA2) and 3 bacteria-specific (nifH, gyrB, rpoB) gene fragments. Black squares and circles indicate successful PCR amplifications for the eukaryote- and bacteria-specific genes, respectively. Bacterial genes were not amplified from forest-specific cDNAs that were synthesized from eukaryotic poly-A mRNA.

3.2.2 RNA extraction

Pestles, mortars, all glass- and plastic-ware, and aqueous solutions were autoclaved before use. Molecular biology grade chemicals were dissolved in water treated overnight with 0.1% (v/v) diethyl-pyrocarbonate and then autoclaved.

One hundred milligrams of each frozen wood sample were ground to powder under liquid nitrogen in a mortar. 4ml of water-equilibrated (acidic) phenol and 4 mL of autoclaved extraction buffer (0.1 mol Tris–HCl, pH 8.0; 0.1 mol NaCl; 20 mmol Na 2 EDTA; 1% (m/v) polyvinylpyrrolidone-40; with or without 2% (m/v) cetyltrimethylammonium bromide (CTAB)) were then added to the mortar immediately after nitrogen evaporation.Once the suspension was completely thawed, it was transferred as ca. 1.5 mL aliquots in 6–8 centrifuge tubes(2 mL). After a centrifugation at 7000g for 10 min at 4 °C, the upper aqueous phase was transferred to new 2 mL tubes and mixed thoroughly by inversion for ca. 1 min to an equal volume of phenol:chloroform:isoamyl alcohol(24:25:1 by volume, pH 8.0). After a second identical centrifugation step, the upper aqueous phase was re-extracted 1 or 3 times with an equal volume of chloroform and recentrifuged as before. Following transfer of the aqueous phase to a new 1.5 mL tube, the nucleic acids were precipitated by adding an equal volume of 2-propanol. After mixing by inversion and an incubation at −20 °C for at least for 2 h, the nucleic acids were pelleted by centrifugation at 18,000g for 30 min at 4 °C. Pellets were dissolved in 500 µL of water to which was added an equal volume of 6 mol of LiCl for the selective precipitation of RNA. After an overnight incubation at 4 °C, RNA was recovered by centrifugation at 18,000g for 30

min at 4 °C. Residual salts were removed by wash-ing the pellet with 200 µL of 70% ethanol. After a last centrifugation at 18 000g for 5 min at 4 °C, the pellets were dried 10 min on ice and dissolved in 50 µl of water.

3.2.3 RNA post-extraction treatments

For each RNA sample in 50 µl of water, residual DNA was removed using 2 U of RNase-free DNase I according to the manufacturer's instructions (DNA-free TM kit, Ambion, Waltham, Massachusetts, USA). In case RNA preparations did not satisfy purity criteria, as judged by the color of the extract and low OD_{260}/OD_{280} and OD_{260}/OD_{230} values, RNA was further purified using an RNA Clean & Concentrator TM-5 kit from Zymo Research,Irvine, California, USA, to remove potential reverse transcriptase (RT) and PCR inhibitors.

3.2.4 RNA quality

RNA integrity was evaluated following either non denaturing agarose gel electrophoresis (1.2% agarose in 0.5×TBE buffer) and ethidium bromide staining or by capillary electrophoresis using a Bioanalyzer 2100 (Agilent,Santa Clara, California, USA) and a RNA Pico Chip (Agi-lent). RNA purity was assessed by spectrophotometry (spectrophotometer Nanodrop ND-1000; Thermo FisherScientific, Waltham, Massachusetts, USA) at wavelengths230, 260, 270, and 280 nm. RNA concentration was evaluated by fluorimetry using the Qubit RNA HS Assay kit and Qubit Fluorimeter 2.0.

Sample ID		Plant species	Sampling date	Location	Coordinates	Wood moisture [%]	RNA yield [µg/100 mg]	260:280	260:230	260:270
Acesp	a	*Acer sp.*	10/25/2015	Coubon (FRA)	44°99'N, 3°90'E	68.96 ± 9.43	1.03 ± 0.33*	1.38 ± 0.10	0.65 ± 0.08	1.03 ± 0.11
Alngl	a	*Alnus glutinosa*	10/25/2015	Coubon (FRA)	44°99'N, 3°90'E	62.23 ± 0.57	1.41 ± 0.55*	1.62 ± 0.27	0.72 ± 0.08	1.13 ± 0.1
Betpu	a	*Betula pubescens*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	44.1 ± 0.83	2.36 ± 1.5	1.61 ± 0.33	0.70 ± 0.11	1.12 ± 0.1
Carbe	a	*Carpinus betulus*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	47.22 ± 1.98	2.11 ± 1.45	1.60 ± 0.27	0.81 ± 0.35	1.15 ± 0.12
Cassa	a	*Castanea sativa*	11/8/2015	Entracque (ITA)	44°23'N, 7°42'E	37.06 ± 13.39	0.59 ± 0.36*	1.50 ± 0.23	0.65 ± 0.36	1.12 ± 0.14
Corav	a	*Corylus avellana*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	54.34 ± 10.17	1.09 ± 0.91*	1.48 ± 0.18	0.80 ± 0.24	1.11 ± 0.13
Cramo	a	*Crataegus monogyna*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	26.13 ± 0.56	0.8 ± 0.2*	1.38 ± 0.15	0.72 ± 0.22	1.09 ± 0.1
Fagsy	a	*Fagus sylvatica*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	46.33 ± 0.68	1.42 ± 2.46	1.53 ± 0.21	1.13 ± 0.30	0.7 ± 0.1
Fraex	a	*Fraxinus excelsior*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	30.68 ± 5.8	1.21 ± 0.59*	1.64 ± 0.26	1.01 ± 0.29	1.12 ± 0.12
Junco	a	*Juniperus communis*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	40.5 ± 0.68	7.633 ± 0.77*	1.26 ± 0.02	0.77 ± 0.02	0.94 ± 0.11
Larde	a	*Larix decidua*	6/18/2015	Vinadio (ITA)	44°20'N, 7°14'E	38.72 ± 4.38	0.97 ± 0.24*	0.96 ± 0.01	0.30 ± 0.00	1.26 ± 0.12
Launo	a	*Laurus nobilis*	12/8/2015	S. Margherita Ligure (ITA)	44°33'N, 9°22'E	47.66 ± 0.41	2.13 ± 0.54	1.37 ± 0.08	0.90 ± 0.30	1.07 ± 0.1
Malsy	a	*Malus sylvestris*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	50.99 ± 0.91	1.64 ± 1.46	1.46 ± 0.12	1.03 ± 0.19	1.13 ± 0.15
Picab	a	*Picea abies*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	67.74 ± 0.79	2.24 ± 2.32*	1.50 ± 0.23	0.68 ± 0.08	0.98 ± 0.14
Pinsy	a	*Pinus sylvestris*	10/25/2015	Solignac sur Loire (FRA)	44°58'N, 3°53'E	51.33 ± 1.07	2.54 ± 2.73	1.4 ± 0.07	0.63 ± 0.03	1.11 ± 0.11
Plaor	a	*Platanus orientalis*	11/5/2015	Torino (ITA)	45°05'N, 7°68'E	71.74 ± 0.56	0.252 ± 0.24*	1.54 ± 0.05	1.02 ± 0.06	1.03 ± 0.17
Poptr	a	*Populus tremula*	10/25/2015	Solignac sur Loire (FRA)	44°58'N, 3°53'E	43.3 ± 1.38	2.59 ± 0.66	1.34 ± 0.02	0.75 ± 0.06	1.12 ± 0.12
Pruav	a	*Prunus avium*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	62.59 ± 1.0	0.95 ± 0.39*	1.44 ± 0.04	0.87 ± 0.13	1.08 ± 0.1
Quero	a	*Quercus robur*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	29.44 ± 3.87	0.796 ± 0.83	1.47 ± 0.02	0.76 ± 0.13	1.52 ± 0.11
Queru	a	*Quercus ruber*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	69.14 ± 0.25	0.41 ± 0.17*	1.37 ± 0.20	0.55 ± 0.16	1 ± 0.1
Robps	a	*Robinia pseudoacacia*	11/3/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	65.94 ± 1.49	0.99 ± 0.95*	1.36 ± 0.17	0.76 ± 0.26	0.72 ± 0.18
Salsp	a	*Salix sp.*	10/24/2015	Entracque (ITA)	44°23'N, 7°42'E	37.31 ± 3.35	1.77 ± 1.21	1.39 ± 0.06	0.78 ± 0.07	1.10 ± 0.1
Manpa	b	*Fagus sylvatica; Corylus avellana; Quercus robur; Carpinus betulus*	2/28/2015	Venaria Reale (ITA)	45°18'N, 7°55'E	54.81 ± 3.87	1.30 ± 0.29	1.35 ± 0.01	0.76 ± 0.06	1.4 ± 0.09
Foroa	b	*Quercus pubescens, Prunus avium; Cotynus coggygria*	4/1/2015	Bussoleno (ITA)	45°14'N, 7°10'E	57.98 ± 1.03	4.42 ± 1.07	1.18 ± 0.02	0.66 ± 0.02	1.25 ± 0.14
Valsa	b	*Acer campestre; Fraxinus excelsior*	6/10/2015	Laives (FRA)	46°67'N, 4°82'E	59.36 ± 5.88	0.49 ± 0.044*	1.77 ± 0.14	1.08 ± 0.07	1.04 ± 0.14
Marpa	b	*Abies alba; Fagus sylvatica; Laburnum alpinum*	6/15/2015	Chiusa di Pesio (ITA)	44°20'N, 7°68'E	50.16 ± 11.365	4.57 ± 2.10	1.44 ± 0.05	0.79 ± 0.05	1.14 ± 0.12
Lompa	b	*Larix decidua*	6/18/2015	Vinadio (ITA)	44°20'N, 7°14'E	38.72 ± 4.38	0.97 ± 0.24*	0.96 ± 0.01	0.30 ± 0.00	1.15 ± 0.12

Table 01: Origin and characteristics of the decaying wood samples and of the corresponding RNA extracts with the standard method described in the paper.

3.2.5 cDNA synthesis

Five hundred nanograms of total tree-specific RNA were used for cDNA synthesis in the presence of 4 µmol of random hexamers (10 µl final volume). The mixture was heated 5 min at 70 °C and kept on ice for at least2 min before adding 10 µl of reaction mix (4 µl of 5×buffer (Thermo Fisher Scientific); 2 µl of dNTPs, 10 µmol each; 1.5 µl of RNAsin at 40 $U \cdot \mu l^{-1}$; 2 µlL of 5% bovine serum albumin (BSA); 1 µl of M-MLV Reverse Transcriptase at 200 $U \cdot \mu l^{-1}$ (Thermo Fisher Scientific); 0.5 µl of RNA-grade water). After 1 h at 42 °C, the enzyme was inactivated by incubating 10 min at 70 °C.

For the synthesis of cDNAs from polyadenylated eukaryotic mRNAs, 500ng of total forest-specific RNA were used with the Mint-2 cDNA synthesis and amplification kit (Evrogen, Moscow, Russian Federation) following the manufacturer's protocol. Briefly, first-strand cDNA synthesis was primed using a CDS adaptor, which anneals to the 3' poly-A tail of eukaryotic mRNA. Upon reaching the5= end of the mRNA the Mint RT enzyme adds a few deoxycytidine (C) nucleotides at the end of all of theneo-synthesized single-stranded cDNAs (ss-cDNAs). These extra Cs are used to anchor a so-called PlugOligo adaptor whose complementary sequence is added by the RT enzyme at the 3= end of all ss-cDNAs. As a result, all ss-cDNAs are bordered at both their 3= and 5= ends with ac ommon 23-nt-long M1 sequence present in both the CDS and PlugOligo adapters that is used to PCR amplify all double-stranded cDNAs (ds-cDNAs).

3.2.6 PCR amplifications

Internal fragments of 8 eukaryotic or bacterial protein-coding genes were amplified using published primer pairs.The EF1α primers (Rehner and Buckley 2005) target the housekeeping and constitutively expressed eukaryotic elongation factor 1α. The GH7 (Edwards et al., 2008) and GH5_5 (Barbi et al., 2014) primers target the 2 correspond-ing fungal families of cellulases. The GH11 primers (Barbi et al., 2014) target fungal hemicellulases (endoxylanases).The AA2 primers (Barbi et al., 2014) target fungal (basidiomycetes) class II peroxidases involved in lignin hydrolysis. nifH primers (Poly et al., 2001) target bacterial nitrogenase encoding genes, and the rpoB (Mollet et al., 1997) and gyrB (Barret et al., 2015) primers target the bacterial housekeeping and constitutively expressed RNA polymerase and DNA gyrase β-subunits encoding genes, respectively.

PCRs were performed in 25 µl final volumes contain-ing either 1 µl of tree-specific ss-cDNA or 50 ng of forest-specific ds-cDNA, 2.5 µl of 10× Taq buffer (Thermo FisherScientific), 0.1 mmol dNTPs, 2 µmol of forward and re-verse primers, 0.3% bovine serum albumin, 1 U of Taq polymerase (Thermo Fisher Scientific), 13.4 µl of water. Amplifications were performed using a T100 TM Thermal Cycler (Bio-Rad Laboratories, Hercules, California, USA)and the cycling parameters given in the publications describing each of the primer pairs. Fungal rDNA internal transcribed spacer (ITS) region was amplified from 2 tree-specific samples using primers ITS1f (Gardes & Bruns 1993) and ITS4 (White et al., 1990) using the following program: initial denaturation for 4 min at 94 °C, 30 cycles of denaturation (30 s at 94 °C), annealing (45 s at 50 °C), and elongation (1 min at 72 °C), followed by a final

elongation step for 5 min at 72 °C.

3.2.7 Sequencing and sequence analysis

PCR products were purified using a QIAquick PCR Purification kit from Qiagen (Hilden, Germany) and cloned into the pCR4-TOPO vector using the TOPO TA Cloning Kit for Sequencing according to the manufacturer's instructions (Invitrogen Life Technology, Karlsruhe, Ger-many). Where possible, 5 positive clones for each genefamily were sequenced (Sanger sequencing; Biofidal, Vaulx en Velin, France). Nucleotide sequences were manually edited and compared to the GenBank database us-ing the BLASTX (protein coding sequences) or BLASTN(ITS sequences) programs. Sequences were deposited in the EMBL, GenBank, and DDJB databases under accession Nos. MF580440–MF580468.

3.3 Results and discussion

Different protocols and commercial RNA extraction kits were first tested on 4 decaying wood samples: 2 from angiosperms, *Platanus orientalis* (Plaor) and *Quercus robur* (Quero), and 2 from gymnosperms, *Juniperus communis* (Junco) and *Picea abies* (Picab). Two commercial kits designed to extract RNA from soil (RNA Power Soil kit from MO BIO laboratories, Carlsbad, California, USA, and Zymo BIOMICS TM RNA/DNA Mini kit (Zymo Research)) failed to extract RNA from the samples or resulted in the extraction of extremely low quantities of RNA that could not be visualized on agarose gels. The Qiagen

RNeasy Plant Mini kit (Qiagen), designed to extract RNA from live plant tissues, produced inconsistent results.

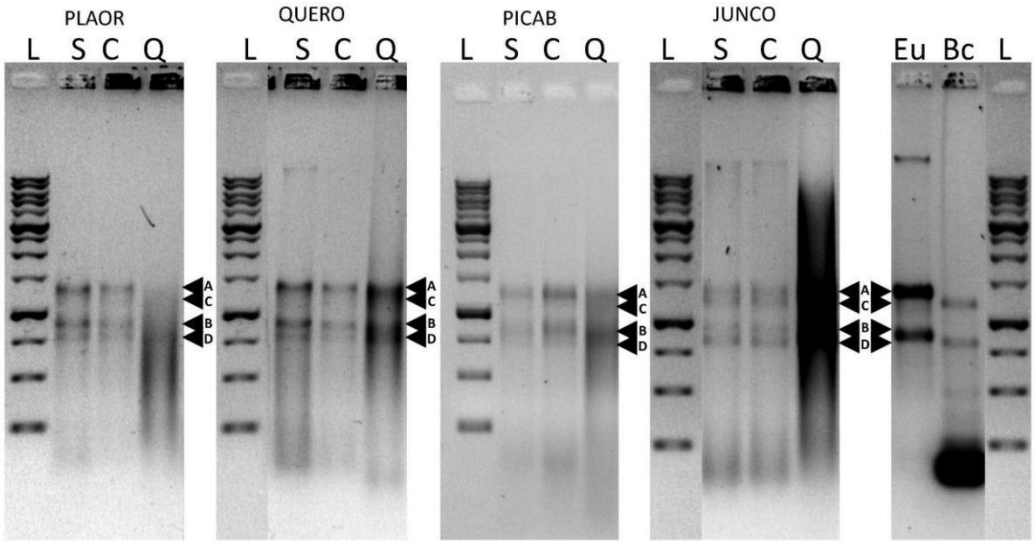


Figure 03: *Specific electrophoretic profile of 4 selected decaying wood RNA extracts. RNA extracts were size fractionated by electrophoresis on non-denaturing agarose gels stained with ethidium bromide. Each extract is characterized by prominent ribosomal RNA fragments that co-migrate with bacterial (lane Bc; RNA extracted from the bacterium Escherichia coli) and eukaryotic (lane Eu; RNA extracted from the fungus Tulasnella calospora) ribosomal RNA. Eukaryotic 28S (A) and 18S (B) and bacterial 23S (C) and 16S (D) ribosomal RNA. For each sample a comparison is showed between RNA extracted using the following protocols: standard protocol described in this article (lane S), standard protocol described in this article with the addition of 2% CTAB in the extraction buffer (lane C), and RNeasy Plant Mini Kit from Qiagen (lane Q). A 1 kb DNA ladder from EuroClone (Pero, Italy) was used as molecular weight marker (lane L).*

Although all 4 Qiagen RNA extracts presented satisfactory OD_{260}/OD_{280} ratios above 1.36, following agarose gel electrophoresis RNA extracted from Plaor and Picab appeared significantly degraded (**Figure 02**). We then implemented an original extraction procedure that combined steps from different protocols (Moser et al., 2004; Gambino et al., 2008), including the

one by Chang et al., (1993) for lignified live tissues of *Pinus taeda*. This protocol was implemented with or without a detergent (2% CTAB)in the extraction buffer. Both versions resulted in the extraction of seemingly undegraded RNA from all 4 decaying wood samples, as judged from the sharp ribosomal RNA bands resolved following both agarose (Fig. 2)and capillary (Agilent bioanalyzer; **Figure 03**) gel electrophoresis. Both versions of the protocol differed, however, with respect to the final extraction yield. Unexpectedly, inclusion of CTAB in the extraction buffer, which was expected to improve cell lysis, resulted in consistently lower RNA quantities (down to ca. 10-foldin the case of the Junco sample) compared with the protocol from which CTAB was omitted (**Table 02**).

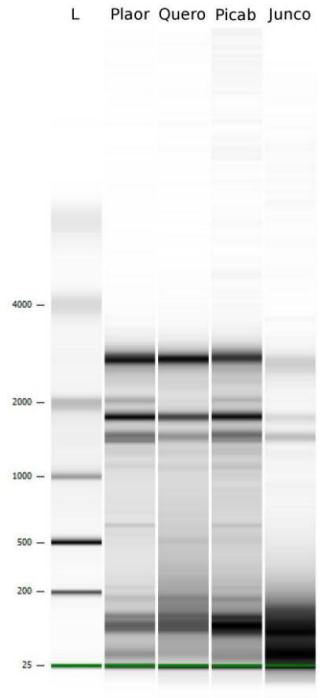


Figure 4: *Separation of four selected decaying wood RNA extracts (see abbreviations in Figure 1), obtained using the standard protocol described in the manuscript by capillary electrophoresis (Bionalyzer 2100 and a RNA Pico Chip (Agilent)) Occurrence of sharp bands of ribosomal RNA suggested an absence of RNA degradation during the extraction procedure.*

In terms of purity, the RNA extracts had OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios below 2.0 expected for pure RNA preparations (**Table 1**).

	Buffer type			
Sample name	**Standard**	**CTAB**	**p value**	
Plaor	0.252 ± 0.238	0.064 ± 0.015	0.835	
Quero	0.796 ± 0.826	0.125 ± 0.015	0.492	
Picab	2.45 ± 2.316	0.093 ± 0.016	0.040	*
Junco	7.633 ± 7.283	0.869 ± 0.775	0.00001	**

Table 2: *Addition of CTAB in the RNA extraction buffer reduces the extraction yield (expressed in µg of RNA per 100 mg of wood ± standard deviation). RNA were extracted using either the standard protocol (Standard) without CTAB in the extraction buffer or with a modified version (CTAB) in which 2% CTAB were added to the buffer. Difference were compared using a simple χ2 test for each of wood samples. *, p< 0.05; **, p < 0.01.*

These low ratios suggest the presence of contaminating molecules, such as proteins, phenolics, and polysaccharides. To exclude contamination by phenol used for protein denaturation and removal, we measured the OD_{260}/OD_{270} ratio after 1 or 3 successive chloroform extractions.

Three successive extractions did not significantly change this ratio close to one, thus ruling out phenol contamination during extraction. The protocol, without CTAB in the extraction buffer and with a single chloroform extraction to remove phenol, was then tested on decaying wood samples from

an additional 18 gymnosperm and angiosperm tree species. Besides being representative of several of the major Northern hemisphere temperate plant families with tree species (e.g., Pinaceae, Salicaceae, Fagaceae, Betulaceae), these plant species differ from each other with respect to their wood structure and chemistry (Schwarze 2007). Wood samples were collected at different stages of decomposition and several of those sampled were clearly affected by either white or brown rot fungal species.

Total RNA was successfully extracted from all of the 22 (18+4) decaying wood samples, as well as from composite wood samples from five forests, with yields ranging from 2.52 ± 2.4 µg.g-1 (*Platanus orientalis*) to 76.33 ± 7.7 µg.g-1 of wood (*Populus tremula*) for individual tree species. Variation in yield could not be related to the plant taxonomic origin nor to the type of wood (**Table 01**) and may simply reflect the –uncontrollable- degree of microbial colonization of the samples and the global activity of the microbial communities. Following agarose gel electrophoresis, all RNA extracts presented clear bands of ribosomal RNA (rRNA) indicative of an absence of degradation during extraction. On agarose gels, most extracts presented an unusual and distinctive electrophoretic profile characterized by up to four rRNA bands of, sometime, similar intensities (**Figure 02 and Figure 04 and Figure 05**). When run side by side with RNA extracted from a bacterium and a fungus, these four bands co-migrated with the prokaryotic (16S and 23S) and eukaryotic (18S and 28S) cytoplasmic rRNA molecules (**Figure 02**) suggesting that most wood-extracted RNA are composed of equal parts of bacterial and eukaryotic RNA. Such a situation contrasts with most preparations of environmental RNA usually dominated by bacterial RNA (Bailly et al., 2007;

Yadav et al., 2016).

In terms of purity, most wood RNA extracts had OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios below the 2.0 ratio expected for pure RNA preparations (Table 1). In cases where the RNA extract was colored and/or was characterized by an OD_{260}/OD_{280} ratio below 1.0. RNA was further purified using an RNA binding resin ("RNA Clean & Concentrator™-5 kit"). Use of this kit usually did not affect the OD_{260}/OD_{230} ratio but increased the OD_{260}/OD_{280} one.

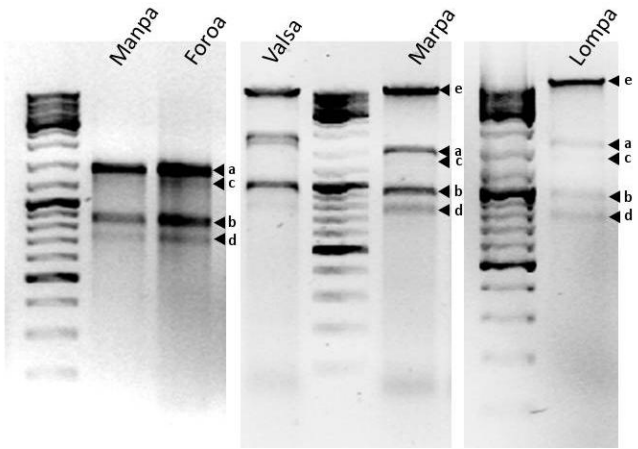


Figure 05: *Specific electrophoretic profile of the five forest-specific wood RNA extracts. RNA extracts were size fractionated by electrophoresis on non-denaturing agarose gels stained with ethidium bromide. Most extracts are characterized by four prominent ribosomal RNA fragments that co-migrate with eukaryotic 28S (a) and 18S (b) and bacterial 23S (c) and 16S (d) ribosomal RNA. (e), metagenomic DNA that could be removed using a DNase I treatment. 1 Kbp Gene Ruler by Thermo Fisher Scientific was used as ladder.*

Despite the putative presence of these contaminants, RNA samples extracted from all 22 tree species were successfully converted into single stranded cDNAs from which were amplified different gene fragments of either fungal or bacterial origin. Sequences from genes constitutively

expressed by all eukaryotic (EF1α) or bacterial (rpoB and gyrB) species were amplified from all or almost all (the gyrB sequence could not be amplified from 4 samples) cDNA samples. Sequences from genes absent from many fungal (GH5_5, GH7, GH11 and AA2 genes) or bacterial (nifH) species were amplified from between 27% (GH11 and nifH) to 59% (AA2 peroxidases) of the samples (**Figure 01**).

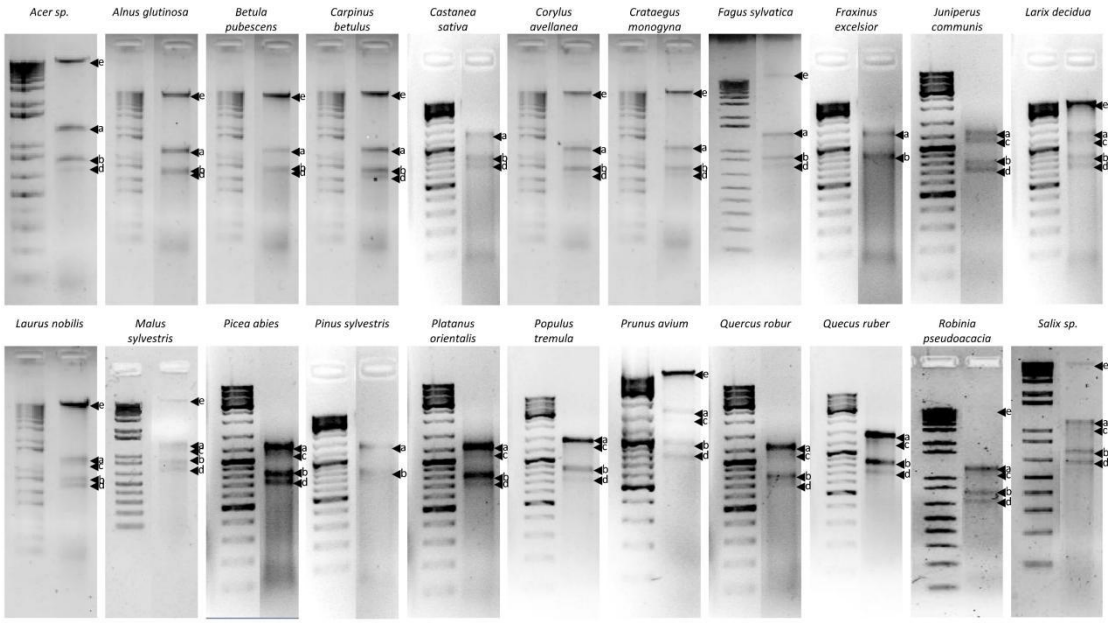


Figure 06: *Specific electrophoretic profile of the 22 tree-specific wood RNA extracts. RNA extracts were size fractionated by electrophoresis on non-denaturing agarose gels stained with ethidium bromide. Most extracts are characterized by four prominent ribosomal RNA fragments that co-migrate with eukaryotic 28S (a) and 18S (b) and bacterial 23S (c) and 16S (d) ribosomal RNA. (e), metagenomic DNA that could be removed using a DNase I treatment. 1 Kbp Gene Ruler by Thermo Fisher Scientific was used as ladder for C. sativa, F. excelsior, J. communis, L. decidua, P. abies, P. sylevstris, P. orientalis, P. tremula, P. avium, Q. robur, Q. ruber samples. In all the others samples 1 kbp Plus Ladder by Thermo Fisher Scientific was used.*

Only in the cases of *Platanus orientalis* and *Castanea sativa* it was possible to amplify all five fungal genes (Fig. 1). In the case of *Picea abies* and *Platanus orientalis* cDNA from which AA2, GH11, rpoB gene fragments were amplified (**Figure 01**), sequencing of PCR products confirmed that the amplified sequences indeed belonged to these different gene families. None of the different GH11 and AA2 sequences were identical to homologous sequences deposited in databases (percentages of identity to known sequences being below 92%). Furthermore, for these two wood samples, amplification and sequencing from cDNA of the ITS fungal barcode sequence (Schoch et al., 2012) identified several taxa typically associated with decaying wood such as the Agaricomycetes *Mycena haematopoda* and *Brevicellicium olivascens*.

Failure to amplify fungal lignocellulolytic gene transcripts from single-stranded cDNAs directly synthesized from environmental (soil) RNA has been reported several times in the literature (see Weber and Kuske, 2011) Absence of amplification can result from the genuine absence of the corresponding transcripts in metatranscriptomes or to their extreme rarity. To circumvent this problem, in the literature (Luis et al., 2005; Bailly et al., 2007; Edwards et al., 2008; Damon et al., 2012), the corresponding sequences are usually amplified from “pre-amplified” double-stranded (ds) eukaryotic cDNAs obtained using the “template switching protocol” (Matz et al., 1999) as implemented by commercial kits, such as the “Mint” or “SMART” ones (from Evrogen and Clontech Laboratories, Mountain View, CA USA respectively). ds-cDNA synthesis and amplification was achieved successfully for all five RNA preparations from "forest wood samples" (**Figure 06**). PCR amplification of GH7, GH11, EF1α eukaryotic gene fragments was obtained for all five cDNA

preparations. By contrast, amplification of these gene fragments was successful in 27%, 32% and 100% respectively of the 22 single stranded cDNA preparation obtained from the RNA extracted from single tree species. However, in the case of the GH*5*_*5* and AA2 gene fragments amplification was only successful for 80% (4/*5*) and 60% (3/*5*) of the five ds-cDNA samples, respectively.

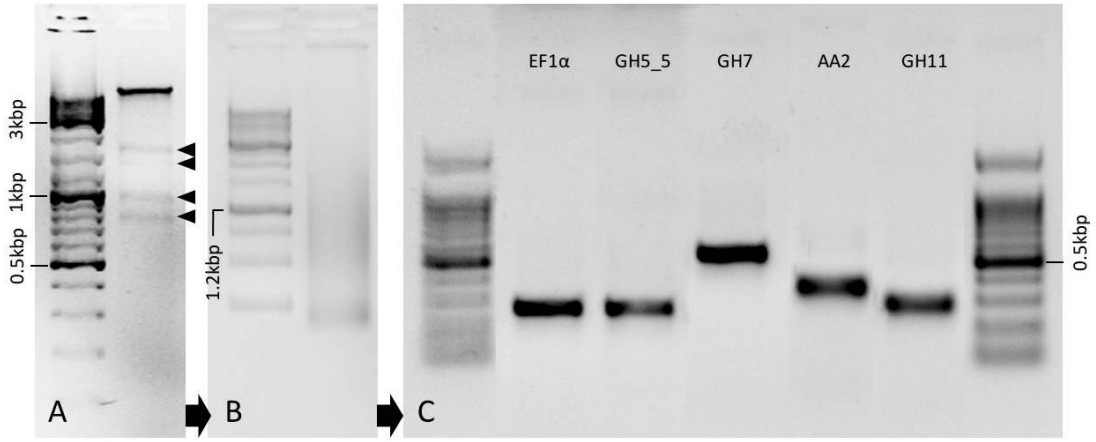


Figure 7: *Conversion of eukaryotic poly-A mRNA extracted from a forest-specific decaying wood sample (A) into large quantities (µg amounts) of double-stranded cDNA (B) that is used as matrix for the PCR amplification of different eukaryote-specific gene fragments (C). (A) The Valsa decaying wood RNA extract presents the four characteristic ribosomal RNA bands (arrowheads) observed in other wood extracts (Fig. 2). 1 Kbp Gene Ruler by Thermo Fisher Scientific was used as ladder in the first gel lane. (B) Double-stranded cDNAs synthesized using the template switching protocol (as implemented in the Mint kit) and further amplified by PCR range in size from ca 100 bp to more than 3 kbp. 1 Kbp DNA Ladder by EuroClone was used in (B) and Gel Pilot 100 bp Plus Ladder by Qiagen was used in (C).*

3.4 Conclusions

We developed a versatile RNA extraction and purification protocol from decaying wood samples that functions on a wide variety of woods, irrespective

of their original physical structure, chemical composition and state of decomposition. We therefore anticipate that this protocol could be implemented successfully on wood samples from other tree species.

All extracted RNA could be converted into cDNA using either a conventional reverse transcription approach or the template switching one. From these cDNAs a number of eukaryotic and bacterial genes were amplified, including genes important for wood degradation such as fungal genes encoding enzymes participating to lignocellulose degradation or bacterial nifH genes that could drive organic nitrogen entry in this nitrogen-limited ecosystem. Furthermore, the successful conversion of wood-extracted RNA into cDNA represents an essential step towards a metatranscriptomic analysis of wood degradation through either the systematic sequencing of all mRNAs after removal of rRNA (McGrath et al., 2008) or the targeted sequencing of specific enzyme-coding gene families involved in lignocellulose degradation (Kellner et al., 2014; Barbi et al., 2014; Baldrian and Lòpez-Mondéjar 2014). Finally, decaying wood cDNAs could also represent a rich source of enzymes relevant to various industrial processes such as biomass treatment, second generation biofuel production and biorefinery. These cDNAs could be screened for such enzymes following cloning and functional expression in a suitable microbial host cell (Kellner et al., 2011; Bragalini et al., 2014; Marmeisse et al., 2017).

In the course of this study we also observed that decaying wood RNA is seemingly composed of eukaryotic and bacterial RNA in similar proportions. This unusual ratio, as bacteria "over dominate" most natural microbial ecosystems such as soil (Plassart et al., 2012), could promote decaying wood as

a model ecosystem to dissect the functional interactions between bacteria and fungi and their relative and complementary contributions to key ecosystem processes such as biomass degradation.

3.5 Acknowledgments

Experiments were performed in the framework of the PeroxiDiv project co-funded by the ANR (France; ANR-RF-2015-01) and the DFG (Germany). Martino Adamo was supported by a grant from the Italian Ministry of Education, University and Research and the CMIRA program of region Rhône-Alpes. We would like to acknowledge Dr. Anna Fusconi (University of Turin) for critically reading the manuscript.

3.6 Bibliography

Albrecht, L. 1991. Die Bedeutung des toten Holzes im Wald. Forstwiss. Centralbl. **110:** 106–113.

Arantes V, Jellison J, Goodell B. 2012. Peculiarities of brown-rot fungi and biochemical Fenton reaction with regard to their potential as a model for bioprocessing biomass. *Applied Microbiology and Biotechnology* **94**: 323–338.

Arnstadt T, Hoppe B, Kahl T, Kellner H, Krüger D, Bässler C, Bauhus J, Hofrichter M. 2016. Patterns of laccase and peroxidases in coarse woody debris of Fagus sylvatica, Picea abies and Pinus sylvestris and their relation to different wood parameters. *European Journal of Forest Research* **135**: 109–124.

Baldrian P, López-Mondéjar R. 2014. Microbial genomics, transcriptomics and proteomics: New discoveries in decomposition research using complementary methods. *Applied Microbiology and Biotechnology* **98**: 1531–1537.

Barbi F, Bragalini C, Vallon L, Prudent E, Dubost A, Fraissinet-Tachet L, Marmeisse R, Luis P. 2014. PCR primers to study the diversity of expressed fungal genes encoding lignocellulolytic enzymes in soils using high-throughput sequencing. *PLoS ONE* **9**: 1–22.

Barbi F, Prudent E, Vallon L, Buée M, Dubost A, Legout A, Marmeisse R, Fraissinet-Tachet L, Luis P. 2016. Tree species select diverse soil fungal communities expressing different sets of lignocellulolytic enzyme-encoding genes. *Soil Biology and Biochemistry* **100**: 149–159.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O. 2015. Emergence shapes the structure of the seed microbiota. *Appl. Environ. Microbiol.* **81**: 1257–1266.

Blanchette, R.A. 1995. Degradation of the lignocellulose complex in wood. *Can. J. Bot.* **73**: 999–1010.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* **28**: 1883.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**: 113–116.

Cornwell WK, Cornelissen JHC, Allison SD, Bauhus J, Eggleton P, Preston CM, Scarff F, Weedon JT, Wirth C, Zanne AE. 2009. Plant traits and wood fates across the globe: Rotted, burned, or consumed? *Global Change Biology* **15**: 2431–2449.

Cragg SM, Beckham GT, Bruce NC, Bugg TDH, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison J, McGeehan JE, *et al.* 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology* **29**: 108–119.

Damon C, Lehembre F, Oger-Desfeux C, Luis P, Ranger J, Fraissinet-Tachet L, Marmeisse R. 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLoS ONE* **7**.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker SE, Barry K, Bendiksby M, *et al.* 2011. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science* **333**: 762–765.

Edwards IP, Upchurch RA, Zak DR. 2008. Isolation of fungal cellobiohydrolase I genes from sporocarps and forest soils by PCR. *Applied and Environmental Microbiology* **74**: 3481–3489.

Fernández-Fueyo E, Linde D, Almendral D, López-Lucendo MF, Ruiz-Dueñas FJ, Martínez AT. 2015. Description of the first fungal dye-decolorizing peroxidase oxidizing manganese(II). *Applied Microbiology and Biotechnology* **99**: 8927–8942.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science* **336**: 1715–1719.

Gambino G, Perrone I, Gribaudo I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* **19**: 520–525.

Gardes M, Bruns T. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of …. *Molecular Ecology Notes*: 9–10.

Johnston SR, Boddy L, Weightman AJ. 2016. Bacteria in decomposing wood and their interactions with wood-decay fungi. *FEMS microbiology ecology* **92**: 1–12.

Kellner H, Luis P, Pecyna MJ, Barbi F, Kapturska D, Kruger D, Zak DR, Marmeisse R, Vandenbol M, Hofrichter M. 2014. Widespread occurrence of expressed fungal secretory peroxidases in forest soils. *PLoS ONE* **9**.

Kellner H, Luis P, Portetelle D, Vandenbol M. 2011. Screening of a soil metatranscriptomic library by functional complementation of Saccharomyces cerevisiae mutants. *Microbiological Research* **166**: 360–368.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Le Provost G, Herrera R, Paiva JA, Chaumeil P, Salin F, Plomion C. 2007. A micromethod for high throughput RNA extraction in forest trees. *Biological Research* **40**: 291–297.

Lonsdale D, Pautasso M, Holdenrieder O. 2008. Wood-decaying fungi in the forest: Conservation needs and management options. *European Journal of Forest Research* **127**: 1–22.

Lorenz WW, Yu Y-S, Dean JFD. 2010. An Improved Method of RNA Isolation from Loblolly Pine (*P. taeda* L.) and Other Conifer Species. *Journal of Visualized Experiments*: 1–6.

Luis P, Kellner H, Martin F, Buscot F. 2005. A molecular method to evaluate basidiomycete laccase gene expression in forest soils. *Geoderma* **128**: 18–27.

Marmeisse R, Bailly J, Damon C, Lehembre F, Lemaire M, Wésolowski-Louvel M, Fraissinet-Tachet L. 2011. Soil Eukaryotic Diversity: A Metatranscriptomic Approach. *Handbook of Molecular Microbial Ecology I*: 597–602.

Marmeisse R, Kellner H, Fraissinet-Tachet L, Luis P. 2017. Discovering Protein-Coding Genes from the Environment: Time for the Eukaryotes? *Trends in Biotechnology* **xx**: 1–12.

Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A. 1999. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Research* **27**: 1558–1560.

McGrath KC, Thomas-Hall SR, Cheng CT, Leo L, Alexa A, Schmidt S, Schenk PM. 2008. Isolation and analysis of mRNA from environmental microbial communities. *Journal of Microbiological Methods* **75**: 172–176.

Mollet C, Drancourt M, Raoult D. 1997. rpoB sequence analysis as a novel basis for bacterial identification. *Molecular Microbiology* **26**: 1005–1011.

Moser C, Gatoo P, Moser M, Pindo M, Velasco R. 2004. Isolation of Functional RNA From Small Amounts. *Molecular Biotechnogy* **26**: 95–99.**Pereira H, Graca J, Rodrigues JC. 2004**. Wood Chemistry in Relation to Quality. *ChemInform* **35**: 80–81.

Otjen, L., and Blanchette, R.A. 1986. A discussion of microstructural changes in wood during decomposition by white rot basidiomycetes. *Can. J. Bot.* **64**: 905–911.

Plassart, P, Terrat, S., Thomson, B., Griffiths, R., Dequiedt, S., Lelievre, M. 2012. Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. *PLoS One* **7**: 15-18.

Poly F, Monrozier LJ, Bally R. 2001. Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil. *Research in Microbiology* **152**: 95–103.

Purahong W, Arnstadt T, Kahl T, Bauhus J, Kellner H, Hofrichter M, Krüger D, Buscot F, Hoppe B. 2016. Are correlations between deadwood fungal community structure, wood physico-chemical properties and lignin-modifying enzymes stable across different geographical regions? *Fungal Ecology* **22**: 98–105.

Purahong W, Hoppe B, Kahl T, Schloter M, Schulze ED, Bauhus J, Buscot F, Krüger D. 2014. Changes within a single land-use category alter microbial diversity and community structure: Molecular evidence from wood-inhabiting fungi in forest ecosystems. *Journal of Environmental Management* **139**: 109–119.

Rehner SA, Buckley E. 2005. A Beauveria phylogeny inferred from nuclear ITS and EF1- sequences: evidence for cryptic diversification and links to Cordyceps teleomorphs. *Mycologia* **97**: 84–98.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, *et al.* 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proceedings of the National Academy of Sciences* **111**: 9923–9928.

Rinne KT, Rajala T, Peltoniemi K, Chen J, Smolander A, Mäkipää R. 2017. Accumulation rates and sources of external nitrogen in decaying wood in a Norway spruce dominated forest. *Functional Ecology* **31**: 530–541.

Ruiz-Dueñas FJ, Martínez AT. 2010. Structural and functional features of peroxidases with a potential as industrial biocatalysts. In: Biocatalysis Based on Heme Peroxidases: Peroxidases as Potential Industrial Biocatalysts. 37–59.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., *et al.* 2012. Nuclear ribosomal internal tran- scribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 6241–6246.

Schwarze FWMR. 2007. Wood decay under the microscope. *Fungal Biology Reviews* **21**: 133–170.

Talbot JM, Martin F, Kohler A, Henrissat B, Peay KG. 2015. Functional guild classification predicts the enzymatic role of fungi in litter and soil biogeochemistry. *Soil Biology and Biochemistry* **88**: 441–456.

Valette N, Perrot T, Sormani R, Gelhaye E, Morel-Rouhier M. 2017. Antifungal activities of wood extractives. *Fungal Biology Reviews* **31**: 113–123.

Weber CF, Kuske CR. 2011. Reverse transcription-PCR methods significantly impact richness and composition measures of expressed fungal cellobiohydrolase I genes in soil and litter. *Journal of Microbiological Methods* **86**: 344–350.

White, T., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In PCR protocols: a guide to methods and applications. Edited by M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White. *Academic Press, London*, UK. pp. 315–322.

Yadav, R., Bragalini, C., Fraissinet-Tachet, L., Marmeisse, R., and Luis, P. 2016. Metatranscriptomics of soil eukaryotic communities. In Microbial environmental genomics (MEG). Edited by F. Martin and S. Uroz. *Humana Press*, New York, USA. pp. 273–287.

Yamashita S, Masuya H, Abe S, Masaki T, Okabe K. 2015. Relationship between the decomposition process of coarse woody debris and fungal community structure as detected by high-throughput sequencing in a deciduous broad-leaved forest in Japan. *PLoS ONE* **10**: 1–16.

Chapter 4

Metabarcoding on both environmental DNA and RNA better resolves differences between fungal communities.

4.0 Premise

In the present thesis, we formulated the hypothesis that a potential role of fungal DyP peroxidases in organic matter could be reflected by differences in DyP gene diversity and expression levels in environments that differ from each other with respect to lignin input and turnover: wood, forest soils and grassland soils.

As a prerequisite to this study, we first compared the fungal diversity in samples representative of each of these environments collected in different geographic sites in order to take into account the largest number of fungal taxa. Fungal diversity was assessed by metabarcoding using both DNA and RNA directly extracted from these environmental samples. As an output of this study, we evaluated different bioinformatics methods to address fungal diversity and suggest that the combined study of environmental DNA and RNA is particularly appropriated to highlight the impact of different environmental parameters on fungal community composition.

4.1 Introduction

Metabarcoding is the combined use of universal DNA barcodes and high-throughput sequencing technologies to characterize microbial communities from genetic material directly extracted from environmental samples (soil, plants, sediment, water, etc.). This strategy is increasingly used to assess biodiversity and the impacts of anthropogenic and ecological factors on the environment (Creer et al., 2016). Metabarcoding is an efficient method that simultaneously targets multiple species. In the case of fungi, metabarcoding complements or can substitute traditional ecosystem biomonitoring protocols often based on the collection and expert identification of individual species (Blaser et al., 2013; Purahong et al., 2017). Metabarcoding also allows identification of the numerous fungal species which are not cultivable (Creer et al., 2016).

In fungal ecology, most (almost all) metabarcoding studies make use of "environmental DNA" to characterize fungal communities. Few studies used "environmental RNA" or both environmental DNA and RNA to assess fungal diversity (Anderson & Parkin 2007; Bastias et al., 2007; Anderson et al., 2008; Damon et al., 2010; Baldrian et al., 2012; Barnard et al., 2013; Žifčáková et al., 2016). As RNA has a higher turnover rate compared to DNA and supposedly degrades faster than DNA upon cell death, its use as a matrix for metabarcoding, should better describe active microbial communities, leaving aside non-active or dead microbial cells that could contribute to the environmental DNA pool. In that respect, the relative abundance of a specific RNA sequence should better reflect the physiological status and prevalence of

the corresponding microbial taxa within the microbial communities (Blazewicz et al., 2013). While many metabarcoding studies carried out in different ecosystem, on different microbial taxa (either bacteria of eukarya) and considering both DNA and RNA highlighted a strong correlation between “active” (RNA based) and the “total" (DNA based) microbial communities, (Visco et al., 2015; Dowle et al., 2015; Pochon et al., 2015; Laroche et al., 2016), other studies identified differences between the different datasets (Laroche et al., 2016; Guardiola et al., 2016). For example, a greater taxonomic diversity often deduced from the DNA dataset (Pawlowski et al., 2014; Guardiola et al., 2016). This could be a consequence of the presence of DNA from dead or resting or dead organisms (legacy DNA) that do not or no longer participate to ecosystem processes (Lennon & Jones 2011). At the same time, several studies (Griffits et al., 2000; Laroche et al., 2016; Guardiola et al., 2016; Zifcakova et al., 2016) also revealed the presence of RNA-specific taxa; these OTUs could be the result of amplification/sequencing artefacts or chimaeras or again, the detection of rare but very active taxa. From a technical point of view, the use of RNA instead of DNA for metabarcoding is however often more challenging and time consuming as RNA extraction from environmental samples is often less straightforward than DNA extraction and RNA needs to be converted to cDNA before PCR-amplification.

In metabarcoding studies, it is therefore essential to identify and eliminate artefactual sequences and spurious taxa whose presence may interfere in data analysis and mask or the contrary exacerbate differences between datasets. The most frequently employed strategy is the "trimming strategy" which considers that the sequences with a low number of reads in the

whole dataset have a higher probability to represent unauthentic taxa or technical artefacts than abundant sequences. As a consequence, a minimal sequence read abundance threshold (typically between 2 and 10 reads) is often implemented to eliminate rare sequences (Blaalid et al., 2013; Brown et al., 2015; Denef et al., 2016; Laroche et al., 2017; Rinne et al., 2017). In the case of studies that generate both RNA and DNA-based datasets on the same environmental samples, it has also been proposed to consider only sequences present in both datasets (Pochon et al., 2015; Laroche et al., 2016).

In fungal taxonomy and community ecology, the most study DNA barcode is the nuclear ribosomal RNA (rRNA) intergenic spacer (ITS) sequence that contains the short 5.8S rRNA and separates the 18S and 28S rRNA subunits (Schoch et al., 2012; Blaalid et al., 2013). This fast-evolving, intron-like sequence is first transcribed as part of a large rRNA precursor and then sequentially eliminated during maturation of this precursor inside and then outside of the nucleus (Lalev & Nazar 2001; Ferreira-Cerca et al., 2005). The ITS can therefore be amplified using primers located within the rRNA genes from either DNA or cDNA matrice. When used as barcode in fungal community studies, the ITS is almost exclusively amplified from environmental DNA (but see Baldrian et al., 2012 and Barnard et al., 2013 for its amplification from RNA). The bioinformatic treatments of the data most commonly consist in the successive elimination of low quality sequences, the detection and removal of potential chimeric sequences, the clustering of the remaining sequence reads based on a sequence similarity threshold (usually equal or above 97%) and finally the exclusion of clusters containing a single sequence (so-called "singletons").

The aim of the present study was to compare DNA barcodes generated from DNA and RNA co-extracted from different grassland soils, forest soils and decomposing wood samples collected in geographic sites differing from each other with respect to soil characteristics, climate and vegetation. In mycology, it is widely accepted, especially as regards "macrofungi" that the three studied environments (grassland soils, forest soils and decomposing wood) host distinctive fungal assemblages and that geography, climate and vegetation exert a strong impact on fungal species distribution. Consequently, it is expected that comparison of corresponding metabarcoding data should lead to a clear grouping of communities sampled in similar environments while clearly separating the different contrasted geographic sites. We therefore evaluated different bioinformatic treatments of the DNA and RNA metabarcoding datasets with respect to their capacity of emphasizing similarities between environments and differences between sites. Additionally we emphasized and tried to interpret the differences that emerged from the parallel analysis of the DNA and RNA datasets.

4.2 Materials & Methods

4.2.1 Site selection

The four selected study sites are located in North-West Italy. These sites represent different of the distinct climatic and biogeographic zones of these area, the continental one found at low elevations (Mandria site), the sub-Mediterranean xeric zone (Foresto site) and the medium/high altitude alpine one (Creus and Maddalena sites). All sites were in protected areas and

site selection was also based on the co-occurrence of adjacent forested and natural grassland plots of high plant biodiversity and naturalistic importance (**Table 01**) (Sindaco et al., 2009, Biondi et al., 2010). Besides geography and climate, sites differ considerably with respect to geology and soil features (**Table 02**). By collecting samples in these different protected undisturbed areas we therefore expected to access different highly diversified fungal communities (Zak et al., 2003; Wu et al., 2013; Tedersoo et al., 2014; Barbi et al., 2016).

4.2.2 Sampling method

A common sampling protocol was followed in all sites. In both grassland and forest plots, 20 soil cores (8 cm in diameter, 15 cm in depth) were regularly collected along two distinct 20 m-long linear transects. After litter and plant removal, each sample was sieved (2 mm mesh size) and all samples from the same grassland/forest plot were mixed together in equal amounts to constitute a single composite sample that was frozen in liquid nitrogen and stored at -75°C before DNA/RNA extraction. About 100 wood pieces were regularly collected in the vicinity of the two transects used for forest soil collection. Wood samples represented different size classes (from twigs to trunk fragments), stages of decomposition and the different tree species present on the sampling site. After removing bark fragments, wood was reduced to sawdust using a stainless steel grater. For each forest, all samples were mixed together in equal amounts, frozen in liquid nitrogen and stored at -70 °C until RNA/DNA extraction.

4.2.4 cDNA synthesis

500 ng of RNA were used for cDNA synthesis in the presence of 4 µmol of random hexamers (10 µL final volume). The mixture was heated 5 min at 70°C and kept on ice for at least two minutes before adding 10 µL of reaction mix (4 µL of 5x buffer (Thermo Fisher Scientific); 2 µL of 10 µmol dNTPs; 1.5 µL RNAsine at 40 U.µL-1; 2 µl of 5 % Bovine Serum Albumin (BSA); 1 µL of M-MLV Reverse Transcriptase at 200 U.µL-1 (Thermo Fisher Scientific); 0.5 µL RNA grade water). After 1 h at 42 °C, the enzyme was inactivated by incubating 10 min at 70°C.

4.2.3 RNA/DNA extraction

RNA was extracted from 2 g of soil using the RNA Power Soil extraction Kit from MOBIO laboratories (Carlsbad, CA USA) according to the manufacturer's instructions. DNA was co-extracted using the PowerSoil DNA Elution Kit by MOBIO laboratories. Purity of the extracts was evaluated by spectrophotometry (Nanodrop ND-1000 spectrophotometer, Thermo Fisher Scientific, Waltham, MA USA) and quantified by fluorimetry using the Qubit dsDNA BR Assay kit and Qubit Fluorometer 2.0 (Thermo Fisher Scientific).

Wood RNA and DNA were co-extracted from 100 mg of wood samples following the protocol described in Chapter 03 (Adamo et al., 2017).

SITE ID	SCI Site Code	Collection date dd/mm/yyyy	Location	Coordinates	Altitude (m asl)	Mean annual temperture [°C]	Mean annual precipitati	Parental rock	Sample ID	Sample Type	Ecological features
Mandria	IT1110079	28/02/2015	Venaria Reale, TO	45°18'N, 7°55'E	300	12.3	860	Quaternary sediments	MB	Forest soil*	Sub-Atlantic and medio-European oak or oakhornbeam forests of the *Carpinion betuli* (9160)
									MP	Grassland soil*	Molinia meadows on calcareous, peaty or clayey-siltladen soils (*Molinion caeruleae*)(6410)
									MW	Decaying wood **	*Quercus robur; Carpinus betulus; Acer campestre; Corylus avellana*
Foresto	IT1110030	1/4/2015	Bussoleno, TO	45°14'N, 7°10'E	500	11,4	799	Limestone	FB	Forest soil*	Pannonian woods with *Quercus pubescens* (91H0)
									FP	Grassland soil*	Semi-natural dry grasslands and scrubland facies on calcareous substrates (Festuco-Brometalia)(*important orchid sites)(6210*)
									FW	Decaying wood **	*Quercus pubescens, Prunus avium; Cotynus coggygria*
Creus	IT1160057	15/06/2015	Chiusa di Pesio, CN	44°20'N, 7°68'E	1200	8,21	1289	Quartzites	CB	Forest soil*	Acidophilous Picea forests of the montane to alpine levels (*Vaccinio-Piceetea*) with an *Abies alba* prevalence (9410 - 42.25)
									CP	Grassland soil*	Mountain hay meadows (6520)
									CW	Decaying wood **	*Abies alba; Fagus sylvatica; Laburnum alpinum*
Lombarda	IT1160023	18/06/2015	Vinadio, CN	44°20'N, 7°14'E	2000	2,9	695	Gneiss	LB	Forest soil*	Alpine *Larix decidua* forest (9420)
									LP	Grassland soil*	Species-rich *Nardus* grasslands, on siliceous substrates in mountain areas (6230)
									LW	Decaying wood **	*Larix decidua*

* Habitat descripton based on "Italian Interpretation Manual of the 92/43/ECC Directive Habitat - http://vnr.unipg.it/habitat/index.jsp
** dominant tree species

Table 01 – Main characteristics of the sampling sites: Coordinates are in WGS84 datum; Ecological features refer to Biondi et al., 2010.

Sample ID	Clay (< 2 m)	Faint silt (2/20 m)	Coarse silt (20/50 m)	Faint sand (50/200 m)	Coarse sand (200/2000 m)	Nitrogen (N)	Phosphorous (P2O5)	Oragnic Carbon (C)	Organic matter	Volatiles	C/N	Cation exchange capacity	Calcaire (CaCO3)	pH	Extractives	Klason lignin content	acid-soluble lignin
	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/100g	-	cmol⁻	g/kg	-	%	%	%
MB	197	295	267	167	74	2,22	0,017	33,4	57,7	9,12	15	5,55	< 1	4,53	N/A	N/A	N/A
MP	194	312	290	163	41	2,07	0,026	22,7	39,3	7,49	11	6,1	< 1	5,7	N/A	N/A	N/A
MW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,15	33,75	2,93
FB	124	140	126	347	263	3,95	0,03	71,6	124	11,7	18,1	22,8	218	8,08	N/A	N/A	N/A
FP	117	139	143	413	188	3,95	0,02	71,9	124	10	18,2	18,7	198	8,14	N/A	N/A	N/A
FW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	UN	UN	UN
CB	211	305	88	98	298	5,65	0,053	103	178	22,2	18,3	13,1	< 1	3,93	N/A	N/A	N/A
CP	200	267	109	114	310	5,44	0,073	56,5	97,8	13,3	10,4	8,57	< 1	4,72	N/A	N/A	N/A
CW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,06	44,31	1,56
LB	269	247	82	118	284	5,72	0,053	87	150	20,4	15,2	11,6	< 1	4,81	N/A	N/A	N/A
LP	340	270	76	107	207	7,96	0,081	99,3	172	23,1	12,5	10,9	< 1	4,99	N/A	N/A	N/A
LW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	4,56	48,02	0,9

Table 02 – Physicochemical characteristics of the studied soil and wood samples: Soil analyses were performed by the "laboratoire INRA d'analyse des sols d'Arras" (https://www6.hautsdefrance.inra.fr/las) using standard protocols including ISO protocols. "Volatiles" represent mass loss after combustion at 550°C. Wood lignin and extractive contents were assayed by Dr. Harald Kellner, Technical University of Dresden (D). N/A, not applicable; UN, not available.

4.2.5 PCR amplification and sequencing

Both cDNA and DNA amplifications were performed following a nested PCR approach with the same set of primers, but starting from 20 ng of extracted DNA and 1 µl of cDNA. In the first reaction the entire ITS region (ITS1-5.8S-ITS2) was amplified using the fungal-specific primers ITS1F - ITS4 (White et al., 1990; Gardes & Bruns 1993). In the second reaction the fITS9 - ITS4 primers were used (Ihrmark et al., 2012) to amplify the shorter (ca 200 bp) ITS2 region suitable for Illumina sequencing and phylogenic analyses (Ihrmark et al., 2012; Schoch et al., 2012; Blaalid et al., 2013). The nested PCR approach was adopted to avoid the massive amplification of plant sequences because of the nature of the samples (soil and decaying wood).

The first PCRs were performed with 2.5 µL of 10 x Taq Buffer (Thermo Fisher Scientific), 0.1 mmol dNTPs; 2 µmol of forward and reverse primers; 0.3% BSA; 1U Taq DNA polymerase (Thermo Fisher Scientific); up to 25 µL (final reaction volume) of water. For the ITS1F - ITS4 primer pair, the PCR conditions used were: 5min at 95°C, 35 cycles of 30 s at 94°C, 45 s at 54°C and 1min at 72°C, followed by 10 min at 72°C. 1 µL of each PCR product was used as template in the nested PCR reaction where 2.5 µL of 10 x Taq Buffer (Thermo Fisher Scientific), 0.1 mmol dNTPs; 2 µmol of forward and reverse primers; 0.3% BSA; 1U Taq DNA polymerase (Thermo Fisher Scientific); 13.4 µL water were mixed in a 25 µL final volume. PCR conditions were 30 s at 98°C, 30 cycles of 10 s at 98°C, 30 s at 64°C and 20 s at 72°C, followed by 10 min at 72°C. All PCRs were performed in using a T3000 thermal cycler (Biometra GmbH, Gottingen, Germany).

PCR products were controlled by electrophoresis on 1% agarose gels, and the four PCR replicates of each sample were pooled in equal amounts and purified using the Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's instructions. After purification the PCR products were quantified with Qubit dsDNA BR Assay kit and Qubit Fluorometer 2.0 (Thermo Fisher Scientific) in order to prepare the libraries for a paired-end sequencing with the Illumina MiSeq technology (2x250 bp) by Fasteris (Plan-les-Ouates, Switzerland).

4.2.6 Bioinformatic analyses

Bioinformatic analyses were performed as described in the study Voyron et al.,(2016).

Paired-end reads from libraries were merged using PEAR v.0.9.8 (Zang et al., 2014), with quality score threshold settled at 28 and minimum reads length settled at 200 bp. Reads were then processed using the Quantitative Insights Into Microbial Ecology (QIIME) v.1.8 software package (Caporaso et al., 2010). Sequences were re-orientated when necessary, and de-multiplexed based on the primers tags. Chimeric sequences were identified and removed using USEARCH61 (Edgar, 2010), as implemented in the QIIME pipeline. Operational taxonomic units (OTUs) were determined using an open reference-based clustering strategy, with the USEARCH61 method, at 98% similarity.

Taxonomic assignations were performed with Mothur v 1.35.1 and the UNITE database version 7.2 for Mothur (Schloss et al., 2009; Koljalg et al.,

2013; http://unite.ut.ee, last access: 12 May 2017). The OTU representative sequences of this study were submitted to GeeBank with the following accession numbers: AAxxxxxx - ZZxxxxxx.

4.2.7 Statistical analyses

Dataset definition

Starting from the bioinformatic pipeline output three separated datasets were created: the *allreads* dataset is composed of the all the reads generated by the described pipeline and its samples were rarefied to a common number of reads per sample with *rrarefy* of the R package VEGAN (version 2,4-3) (Oksanen et al., 2017), rarefaction threshold: 8470 reads. The *10reads* dataset contains only OTUs which globally have more than 10 reads; rarefaction threshold, 7833 reads. The *shared* dataset contains only OTUs which are represented in both RNA and DNA samples; rarefaction threshold, 4239 reads.

Dataset comparison

Multivariate homogeneity of group dispersions was first assessed using the *betadisper* and *permutest* (with 1000 permutations) functions in the R package VEGAN. Using the same R package one-way and two-way PERMANOVAS were performed using the function *anosim*. Differences in fungal communities composition among samples were visualized with a "nonmetric multidimensional scaling ordination" (NMDS) carried out with *NMDS* function of the VEGAN R package. Relative contribution of the different studied variables to the total sample variance was assessed by

variance partitioning (Peres-Neto et al., 2006) using the function *varpart* (R package VEGAN). Indicator species analysis (Dufrene & Legendre 1997) was carried out using the *multipatt* function in the INDICSPECIES R package v1.7.3, with 999 permutations (De Caceres & Legendre 2009), in order to assess which taxa (OTUs) were significantly associated with a specific habitat.

Statistical comparisons between numeric distributions were performed with a kruskall-wallis non parametric test (R-Base package) and a Dunn's Post-hoc test (dunn.test package v1.3.5) (Dinno 2015). Frequencies were compared by means of the Chi-Square test, with the Pearson's correction, when zeros were present in the frequency distributions (Tallarida & Murray 1987). Calculations were performed using WPS Office - Spreadsheets (Kingsoft Office Software).

Species richness indices calculations (Shannon, Simpson and inverted Simpson analysis) (Simpson 1949; Spellerberg & Fedor 2003) were performed using functions *H*, *D* and *iD* of the R package PHYLOSEQ (version 1.22.3) (McMurdie & Holmes 2013). Charts and basic calculations were performed using WPS Office - Spreadsheets (Kingsoft Office Software) or the GGPLOT2 (v2.2.1) package for R (Wickham 2009). Ternary plots were created by means of the GGTERN (v2.2.1) package for R (www.ggtern.com).

Network analysis

Non-random co-occurrence hypothesis between taxa was tested using the checkerboard score (C-score) under a null model preserving samples frequencies (Stone and Roberts, 1990). Network co-occurrence was evaluated calculating all the possible Spearman's rank correlations between OTUs with

more than 10 reads in a rarefied OTU Table with the R script published by Williams et al., 2014.

Co-occurrence relationships were considered as valid only with Spearman's correlation coefficient (ρ) greater than 0.6 and a statistically significant P-value equal or lower to 0.01 (Junker and Schreiber, 2008). Description of the networks topology was performed calculating several parameters: connectedness, transitivity, centrality and modularity (Barberán et al., 2012). All statistical analyses were carried out in the R environment (www.r-project.org) using the packages VEGAN and IGRAPH (Csardi and Nepusz, 2006). Networks were explored and visualized with the interactive platform Cytoscape 3.5.1 (Shannon et al., 2003).

4.3 Results

4.3.1 High-throughput sequencing output

DNA and RNA were extracted from each of the 12 wood or soil samples; after conversion of RNA to cDNA, fungal ITS2 was amplified by PCR from all 24 DNA and cDNA extracts. All PCR controls with water instead of a DNA/cDNA matrix were negative. Libraries were sequenced on an IlluminaTM MiSeq instrument (2x250 bp reads), yielding a total of 2,390,074 (DNA) and 1,499,965 (cDNA) sequence reads. After removal of unmatched and low-quality reads, sequence clustering at a 98% sequence identity threshold produced a total of 5841 (DNA) and 2995 (cDNA) OTUs after removal of singletons (*all reads* datasets).

4.3.2 Effect of combined DNA/RNA analyses and trimming methodologies on read and OTU distribution

OTU table data were analyzed separately on three different datasets (called "*all reads", "10 reads"* and "*shared"*)(**Table 03**). *All reads* corresponds to the DNA and cDNA OTUs (and corresponding reads) after removal of singletons, *10 reads* excludes all OTUs with less than 10 reads in the global (all samples together) DNA or cDNA datasets, leaving the singleton concept for a more restrictive concept (**Figure 01**). Finally, the shared dataset contains only OTUs present in both the *10 reads DNA* and *10 reads cDNA* datasets.

Dataset	Reads		OTUs		Rarefaction threshold (reads)
	DNA	RNA	DNA	RNA	
allreads	239974	150154	5841	2295	8470
10reads	103125	104483	2568	2073	7833
shared	78435	87702	2024	1820	4239

*pre-rarefaction dataset data

Table 03 – Total numbers of reads and OTUs for each of the three datasets compared in the present study: "reads" numbers represent Illumina MiSeq sequences that were retained for clustering after having removed low quality reads and potential chimeras. The "allread" dataset was obtained after removing unique sequences across the combined DNA+RNA dataset; the "10reads" dataset was obtained after removing sequences found less than 10 times across the combined DNA+RNA dataset; the "shared" dataset was obtained after removing from "10 reads" dataset all sequences that were specific to either the DNA or RNA dataset.

Removal of OTUs with less than 10 reads eliminated preferentially OTUs that were specific to either the DNA or cDNA datasets, thus increasing the proportion (from 40 to 68%) of shared OTUs. Elimination of OTUs with less than 10 reads had however little impact on the proportion of reads that were attributed to the shared OTUs (**Figure 01**).

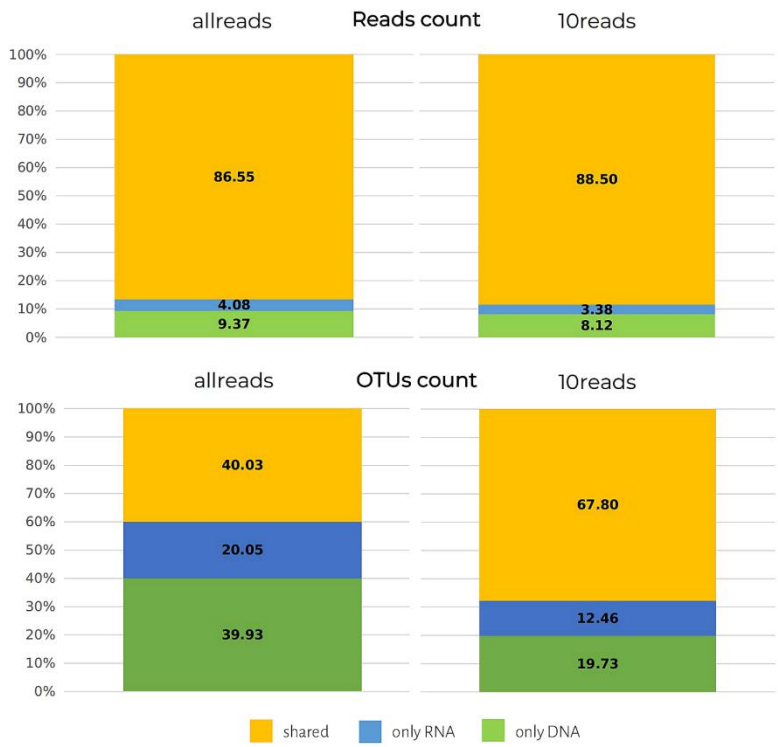


Figure 01 - Effect of removing rare OTUs on the proportion of DNA-specific or RNA-specific reads and OTUs: Removing OTUs with less than 10 reads led to a strong increase in the proportion of OTUs identified in both the DNA and cDNA datasets ("shared") and removed many of the OTUs specific to either the DNA or cDNA datasets.

Regarding the *shared* OTUs that were defined using the global sequence dataset (data from all samples mixed together)(**Table 04**), we observed that at the level of each individual sample at least more than 76% of the OTUs were present as both DNA and cDNA reads in the same sample (**Figure 02**). Shared OTUs represented only by either DNA or cDNA sequence reads within a specific sample always represented a minor fraction of the total number of reads (between 0.05 and 9.31%; **Figure 02**) and were therefore retained in the following analyses.

		DNA		RNA		Total (DNA + RNA)	
site	habitat	OTUs	reads	OTUs	reads	OTUs	reads
MANDRIA	Fsoil	119	7607	155	7830	183	15437
	Gsoil	161	5781	147	6861	185	12642
	Wood	162	6118	95	6393	204	12511
	TOT	442	19506	397	21084	228	40590
FORESTO	Fsoil	207	7492	244	7205	279	14697
	Gsoil	181	6909	228	7079	297	13988
	Wood	50	7483	125	7721	147	15204
	TOT	438	21884	597	22005	192	43889
CREUS	Fsoil	171	6893	147	7757	296	14650
	Gsoil	143	7243	139	7717	282	14960
	Wood	167	7553	123	7753	193	15306
	TOT	481	21689	409	23227	248	44916
LOMBARDA	Fsoil	254	5906	192	6975	298	12881
	Gsoil	167	5211	63	6975	184	12186
	Wood	242	4239	162	7436	271	11675
	TOT	663	15356	417	21386	234	36742

* non-rarefied *shared* dataset data

Table 04 – Overview of the total numbers of reads and fungal OTUs obtained using the "shared" sequence dataset for each of the 12 studied soil and wood samples. Fsoil, forest soil samples; Gsoil, grassland soil samples.

When looking at the global dataset (all samples together), at a coarse taxonomic level, dataset reduction did not significantly affect the relative proportions of reads assigned to specific taxonomic groups (**Figure 03**). The most obvious differences concerned comparisons between the DNA and RNA (cDNA) datasets where a higher proportion of reads were assigned to Ascomycota and less to Basidiomycota in the RNA datasets. The symbiotic Glomeromycota were far more abundant in the RNA dataset.

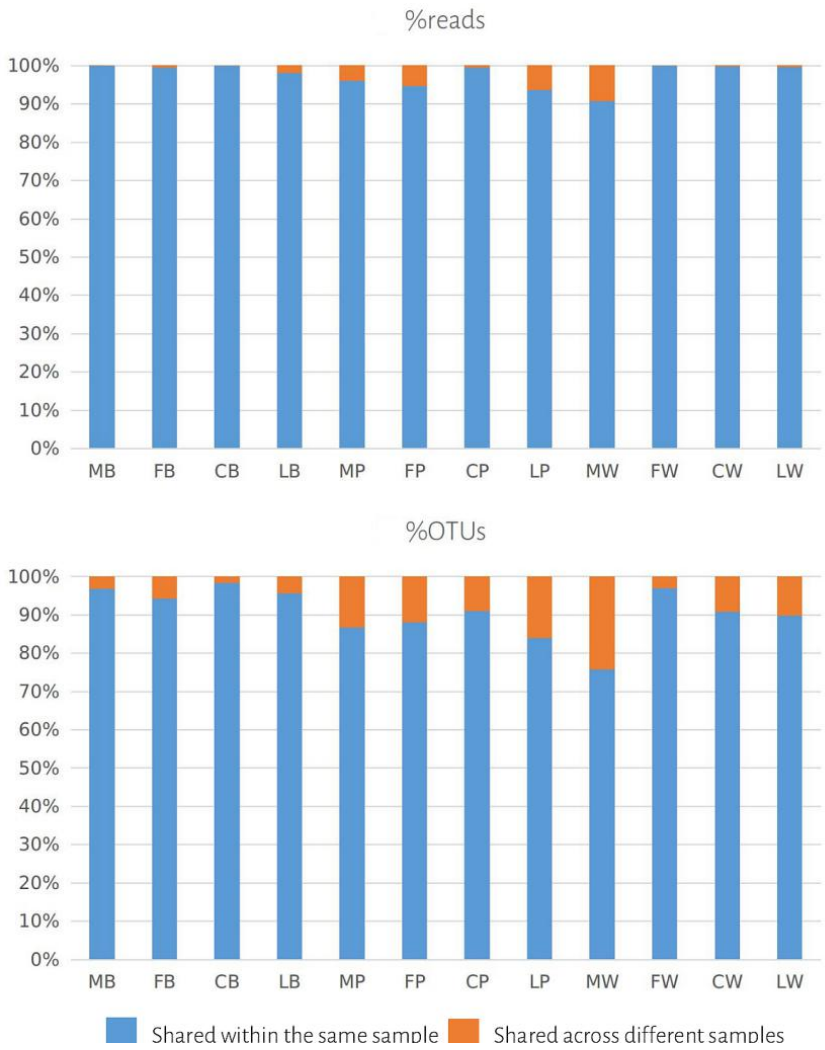


Figure 02 - A majority of the OTUs identified in each individual sample is represented by both DNA and cDNA reads within the same sample. For each of the 12 studied environmental samples (see abbreviations in Table 01) we calculated (blue bars) the proportion of OTUs identified by both DNA and cDNA reads within the same sample sequence dataset (i.e. "shared" OTUs) as well as the proportion of sequence reads affiliated to these shared OTUs. On the contrary, orange bars correspond to the proportion of OTUs only represented by either DNA sequence reads or cDNA sequence reads in the sample's sequence dataset.

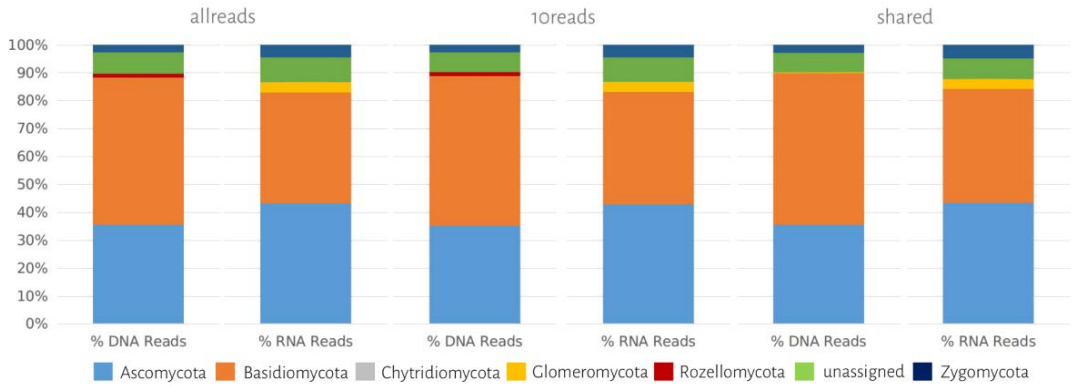


Figure 03 – All samples together, at the phylum level the taxonomic composition of the different DNA and RNA datasets are similar.

Similarly, dataset reduction did not also significantly affect diversity indices (Shannon, Simpson, iSimpson) (**Table 05**).

SHANNON	allreads	10reads	shared
allreads	x		
10reads	0,514920602	x	
shared	0,19247417	0,5148621	x

SIMPSON	allreads	10reads	shared
allreads	x		
10reads	0,866544761	x	
shared	0,743930363	0,87174446	x

iSIMPSON	allreads	10reads	shared
allreads	x		
10reads	0,864483931		
shared	0,734351681	0,862104957	x

Table 05: Pairwise comparisons (P values, two way T tests) of the different diversity indices calculated for the three different datasets.

Concentrating on the "*shared*" dataset, within each individual site, pairwise comparisons of taxon distribution between DNA and cDNA datasets

and between habitats gave variable results and no general rule could be deduced from these confrontations (**Figure 04**). For example, in the Lombarda site, all three habitats had a specific taxon distribution with the DNA dataset but could not be separated with the cDNA one. On the contrary, in the Foresto and Creus sites, none of the habitats could be separated based on the DNA datasets while several, but not all, of the habitats had a statistically significant taxon distribution with the cDNA data.

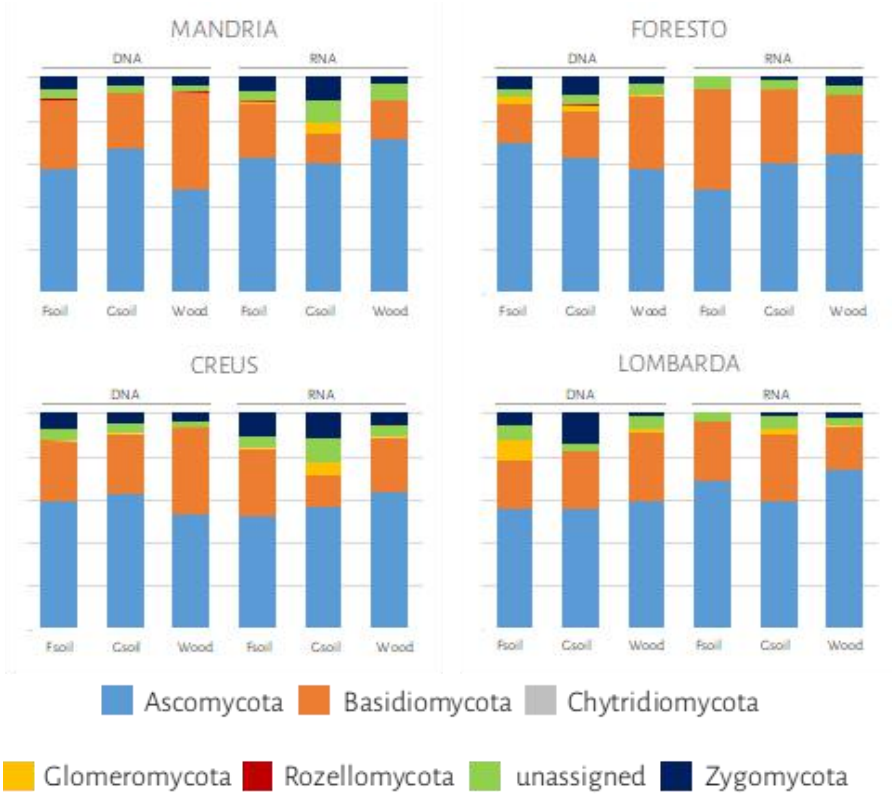


Figure 04: Taxon distribution (phylum level) in each of the different 12 studied samples. OTUs were identified using either the sharedDNA or sharedRNA read datasets.

At an intermediate taxonomic level (class level), considering the five classes with the highest number of identified OTUs (more than 15), three

different patterns were observed. For one class (Sordariomycetes) there was an equal number of OTUs identified by either a higher number of RNA or DNA reads. For three classes (Agaricomycetes, Tremellomycetes and Leotiomycetes), a majority of the OTUs were over-represented in the DNA dataset while for the Eurotiomycetes OTUs over-represented in the RNA dataset dominated (**Figure 05**).

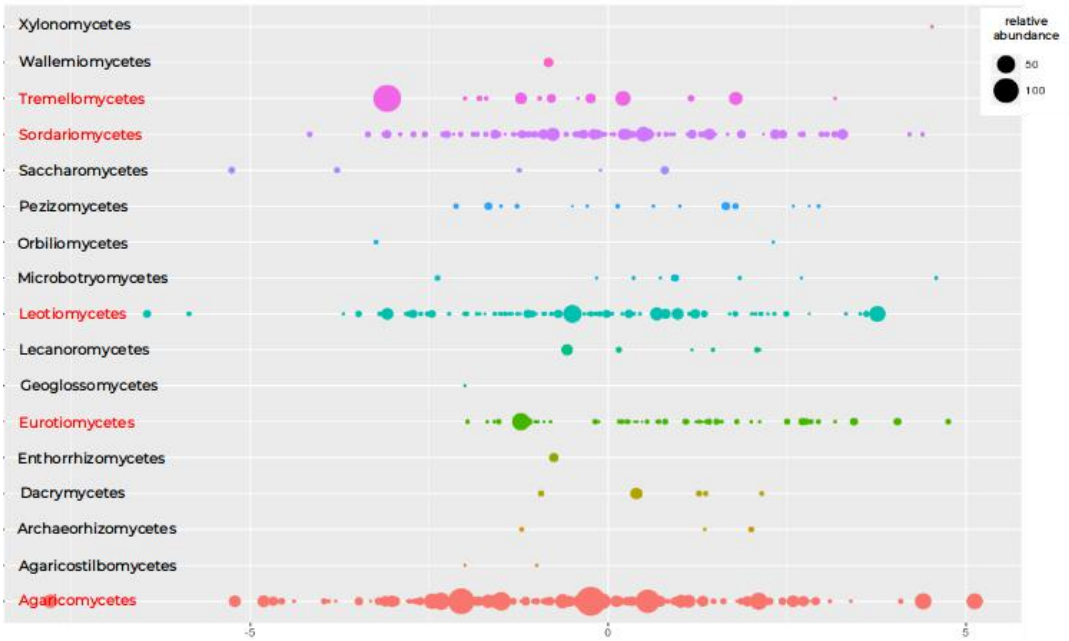


Figure 05: OTU distribution according to their relative abundance (read number) in either the sharedDNA or sharedRNA global datasets (all samples together). For each OTU (filled circles) the log2 transformed value of the ratio [No. of reads the RNA dataset] : [No. of reads the DNA dataset] (log2(RNA:DNA)) was computed and plotted on a horizontal axis for each fungal class. The size of the symbols reflect the relative abundance total no. of reads) of the taxa in the dataset. Classes in red contain more than 15 OTUs.

At the finest taxonomic level (OTU or species level), concentrating on the OTUs identified by the highest numbers of reads in the entire dataset (all sites

together, both DNA and cDNA reads), we observed important differences between them in term of relative abundance in the DNA versus cDNA datasets (**Figure 06**), with species either over-represented in the DNA or cDNA datasets or equally represented in both. These global mean values masked sometime important between samples variations. For example, while the five species with the highest mean overrepresentation in the global RNA dataset were always characterized by a higher overrepresentation as RNA sequences in all of the samples in which they were found, for most of the species with a globally equal representation in the global RNA/DNA datasets, these mean values masked highly variable ratios in the different individual datasets (**Figure 06**). We did not observe any obvious link between this distribution and the taxonomy or trophic mode of the species.

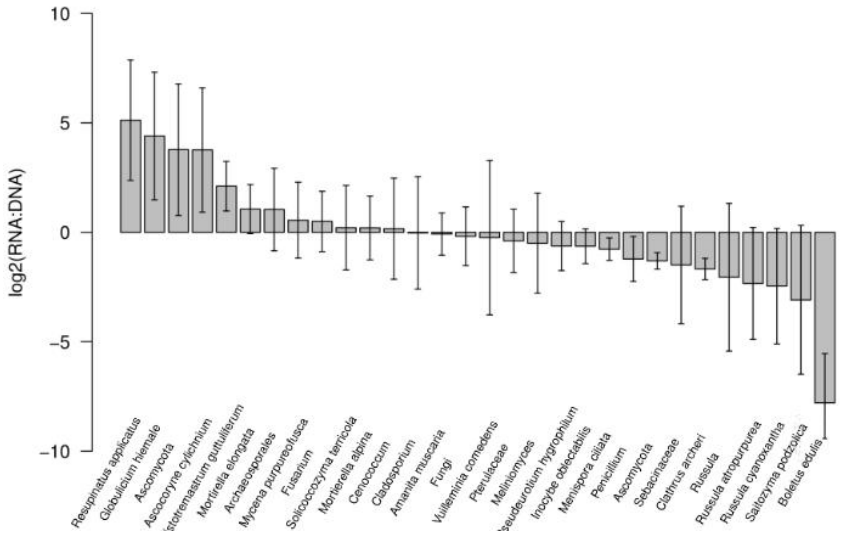


Figure 06: The 30 most abundant OTUs (abundance defined by the absolute number of reads in the sharedDNA+cDNA global dataset) differ from each other with respect to their relative abundance in the sharedDNA and sharedcDNA datasets. Log2 of the (RNA:DNA) ratio was calculated for each of the individual sample in which the taxon was present. Bars indicate the standard deviation of the mean, thus illustrating that for several of the taxa their relative abundance as DNA or cDNA reads varied considerably between samples.

4.3.3 Habitat and geographic site effects

One-way PERMANOVAs, coupled with two-way PERMANOVA analyses, were implemented on the three different datasets, each composed of 24 samples, using abundance-base Bray Curtis indices, to evaluate the impact of OTU selection on the separation of fungal communities according to geographic sites and "habitat" (grassland soils, forest soils and dead wood). Regarding the habitat effect, while the "wood" *versus* "soils" comparisons were highly significant (P<0.01) for all three datasets, significant (P<0.05) differences between grassland and forest soils were only observed using the "*shared*" dataset (**Table 06**). Regarding the site effects, of the six pairwise comparisons, five were found different (P<0.05) using *shared* dataset against only three or four for the "*10 reads*" and "*all reads*" datasets, respectively (**Table 06**). The seemingly highly different Mandria (continental climate, low altitude plain) and Creus (Alpine climate, medium altitude mountain) sites could nevertheless not be differentiated in all three analyses.

Dataset	habitat vs habitat			site vs site					
	Fsoil vs Gsoil	Fsoil vs Wood	Gsoil vs Wood	Mandria vs Foresto	Mandria vs Creus	Mandria vs Lombarda	Foresto vs Creus	Foresto vs Lombarda	Creus vs Lombarda
allreads		**	**	*		*	*		*
10reads		**	**	*		*			*
shared	*	**	**	*		*	*	*	*

* = p-value < 0.05; ** = p-value < 0.01

Table 06: Two-way PERMANOVA analyses of the differences existing between habitats and geographic sites highlight the impact of removing rare or spurious OTUs from datasets in ecological analyses. PERMANOVA analyses were performed using abundance-base Bray Curtis indices calculated separately for each 24 datasets (3 habitats x 4 sites x 2 DNA and cDNA datasets).

Regarding the "*shared*" dataset, NMDS ordination of the different samples was performed separately using the "*shared DNA+cDNA*", the "*shared DNA*"

and the "*shared cDNA*" datasets. These two latter datasets take into account the frequency of each shared OTU exclusively in the DNA and cDNA datasets, respectively. While a clear separation of the samples according to the substrate was observed when using the "*shared DNA+cDNA*" dataset, in the case of the "*shared DNA*" we observed an overlap between the forest soil and wood sampled and an overlap between the forest and grassland samples in the case of the "*shared cDNA*" dataset (**Figure 07**). This result suggests that the forest soil samples may globally share more taxa with decaying wood but that many of the taxa active in grassland soils are also active in forest ones.

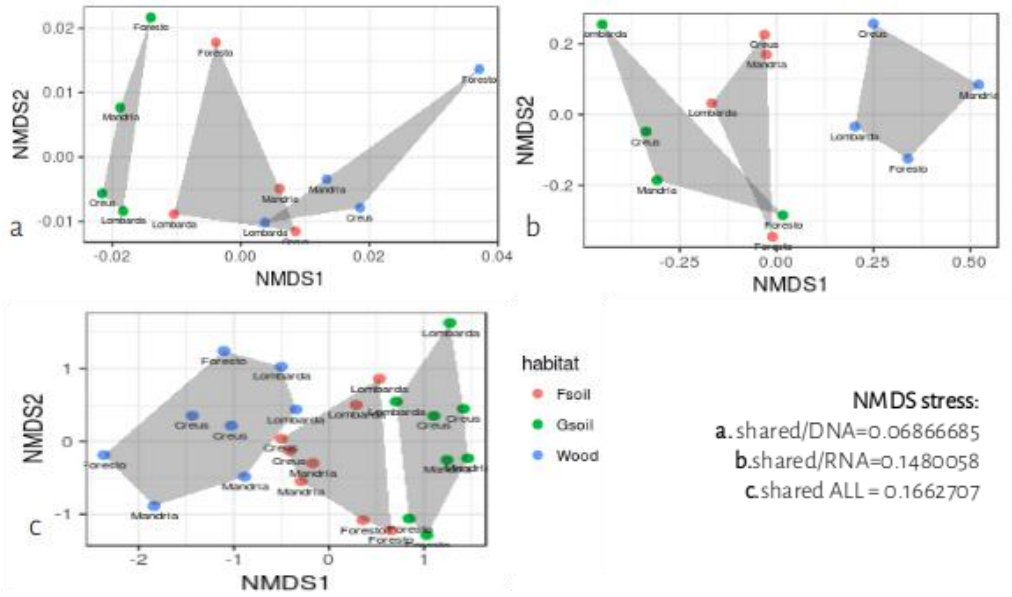


Figure 07: NMDS multivariate ordinations of the shared datasets performed separately for the sharedDNA (a), sharedRNA (b) and sharedDNA + sharedRNA ("sharedAll") (c) datasets show that only the inclusion of both the sharedDNA and sharedRNA datasets in the analysis unambiguously separated the different samples according to the habitat they originate from.

Ternary plots of OTU distribution by habitat abundance illustrated these differences between "*shared cDNA*" and "*shared DNA*" datasets. The "*shared cDNA*" dataset appeared characterized by a lower number of OTUs shared

between the three habitats compared to the "*shared DNA*" one (**Figure 08**).

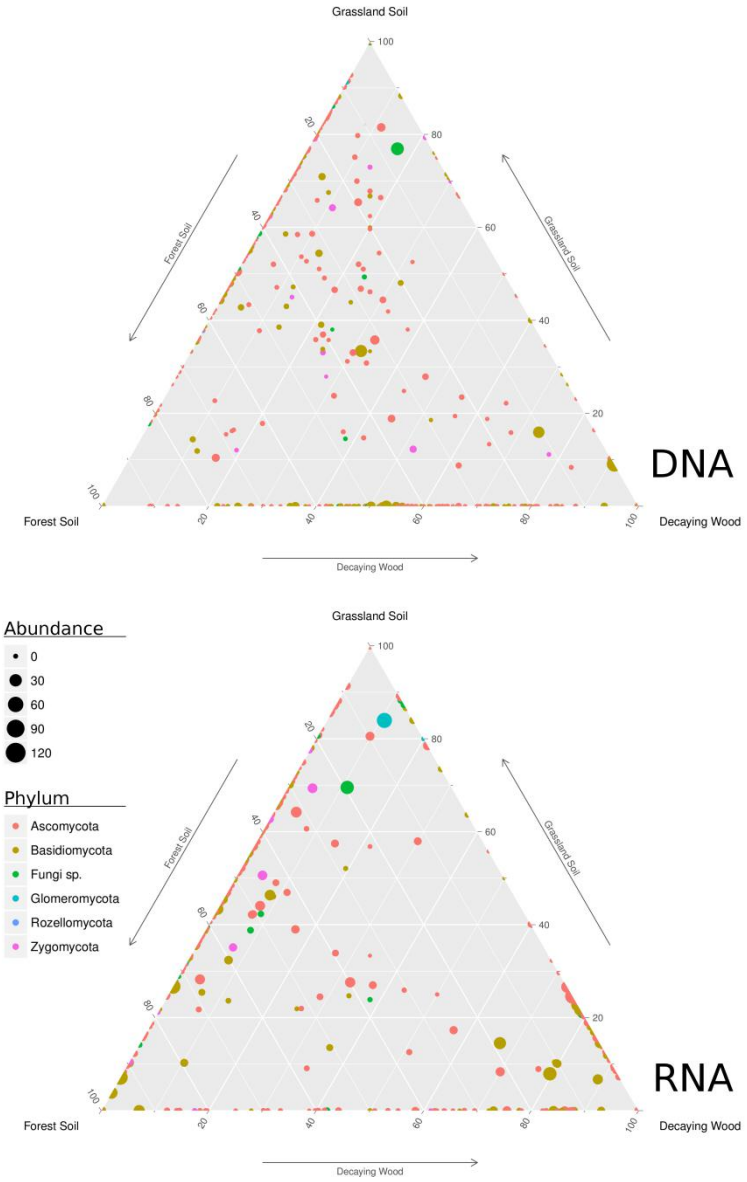


Figure 8 : Ternary plots by habitat abundance illustrating the distribution of each individual taxon (closed circles whose sizes reflect their abundance in terms or read numbers) in each of the three studied habitats. Plots were drawn separately for the sharedDNA and sharedRNA global datasets to illustrate that the relative abundance of taxa in the 3 habitats varies depending on the nucleic acid used for metabarcoding.

In variance partitioning analyses we observed a 14.3% increase in the

proportion of total variance explained by the habitat and site effects and their interaction when comparing the "*all reads*" (24% of the variance explained) and "*shared*" (28% of the variance) datasets (**Figure 09**).

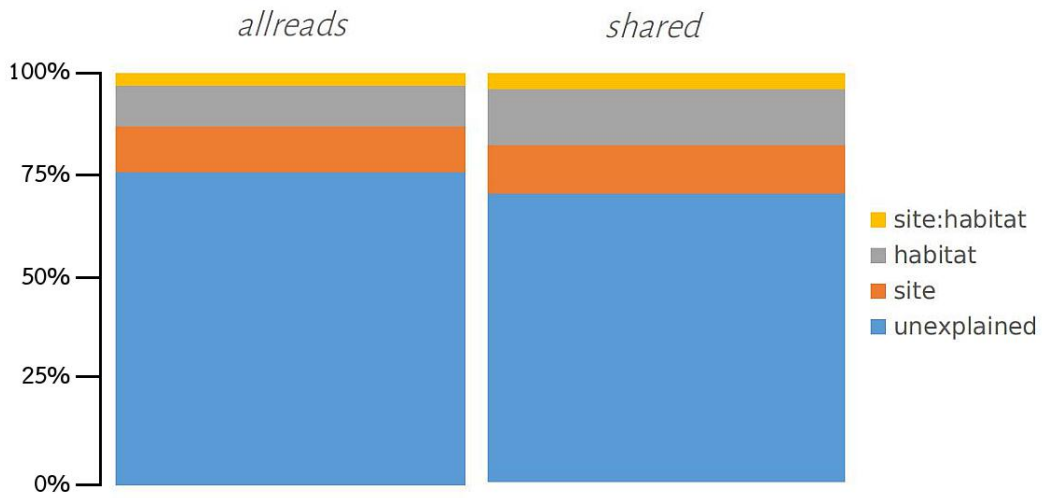


Figure 9 : In a variance partitioning analysis, the percentage of variance explained by the two studied variables (sites and habitat) and their interaction (site:habitat) is significantly higher when using the "sharedDNA+RNA" than when using the allreads dataset from which only singletons were removed.

By using the three "*all reads*", "*10 reads*" and "*shared*" datasets we performed indicator species analysis, to identify indicator species for each of the three habitats. For all three habitats we observed that not only the "*shared*" dataset gave the lowest number of indicator species, but also that all, or almost all, of the indicator species identified with this dataset were also identified with the two others. The opposite was not true as a large fraction of the indicator species identified with the "*all reads*", "*10 reads*" datasets were either specific to them or shared by only these two sets of data (**Figure 10**).

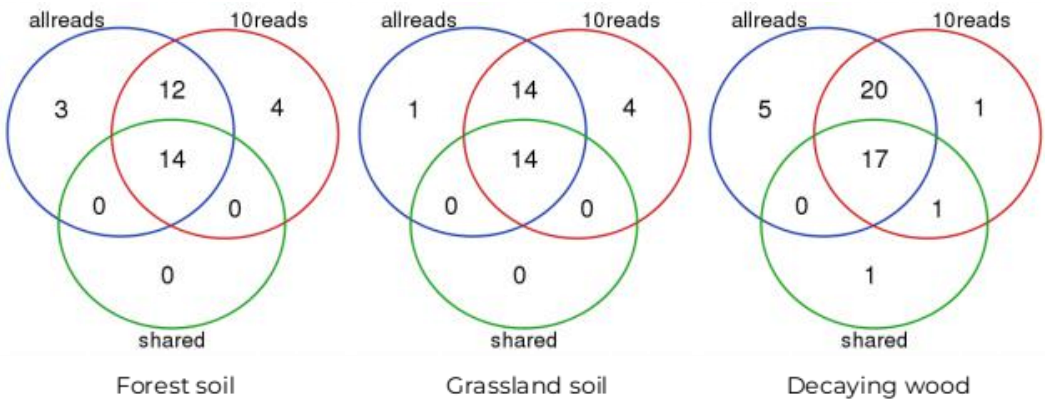


Figure 10 : Indicator species analysis. Venn diagrams comparing for each habitat the numbers and distributions of the different indicator species identified using the "allreads", "10reads" and "shared" datasets.

None of the shared grassland soil indicator species belonged to the Basidiomycota, they dominated by saprotrophic/plant pathogenic Ascomycota (9 species) and also contained saprotrophic Zygomycota *Mortierella* OTUs and one Glomeromycota plant symbiotic species. On the opposite, shared forest soil indicator species were dominated by basidiomycete taxa (9 OTUs versus 4 ascomycete OTUs), and at least seven of them were symbiotic ectomycorrhizal taxa. The taxonomic distribution of the shared decomposing wood OTUs was intermediate (7 Basidiomycota, 9 Ascomycota), and many of the Basidiomycota OTUs are "typical" wood decomposing taxa. (**Table 07**).

OTU ID	habitat	Indicative sp.	phylum	ecology	p-value	
OTU2372	Fsoil	Sebacinaceae sp.	Basidiomycota		0.001	***
OTU3896	Fsoil	Cortinarius alboviolaceus	Basidiomycota	ECM	0.003	**
OTU4167	Fsoil	Herpotrichiellaceae sp.	Ascomycota		0.002	**
OTU11949	Fsoil	Ascomycota sp.	Ascomycota		0.004	**
OTU2042	Fsoil	Saccharomycetales sp.	Ascomycota		0.007	**
OTU1883	Fsoil	Hydnum_repandum	Basidiomycota	ECM	0.011	*
OTU2854	Fsoil	Inocybe_napipes	Basidiomycota	ECM	0.031	*
OTU4215	Fsoil	Inocybe_assimilata	Basidiomycota	ECM	0.027	*
OTU135	Fsoil	Russula_ochroleuca	Basidiomycota	ECM	0.024	*
OTU1950	Fsoil	Laccaria_laccata	Basidiomycota	ECM	0.021	*
OTU3919	Fsoil	Cortinarius_diasemospermus	Basidiomycota	ECM	0.026	*
OTU12086	Fsoil	Ascomycota sp.	Ascomycota		0.026	*
OTU129	Fsoil	Clavulina sp.	Basidiomycota	ECM	0.047	*
OTU4882	Fsoil	Trichoderma_virens	Ascomycota	PLANT PATHOGEN	0.040	*
OTU32	Gsoil	Mortirella_elongata	Zygomycota	SAPROTROPH	0.001	***
OTU11200	Gsoil	Helotiales sp.	Ascomycota		0.005	**
OTU4165	Gsoil	Hypocreales sp.	Ascomycota		0.033	*
OTU4504	Gsoil	Trichoderma_asperellum	Ascomycota	PLANT PATHOGEN	0.045	*
OTU11683	Gsoil	Paraphoma sp.	Ascomycota	PLANT PATHOGEN	0.049	*
OTU210	Gsoil	Mortierella sp.	Zygomycota	SAPROTROPH	0.018	*
OTU462	Gsoil	Paraglomus sp.	Glomeromycota	AM	0.019	*
OTU4965	Gsoil	Fungi sp.	Fungi sp.		0.019	*
OTU4554	Gsoil	Fusarium_oxysporum	Ascomycota	PLANT PATHOGEN	0.023	*
OTU106	Gsoil	Mortierella sp.	Zygomycota	SAPROTROPH	0.019	*
OTU4605	Gsoil	Ascomycota sp.	Ascomycota		0.019	*
OTU4693	Gsoil	Fusarium_solani	Ascomycota	PLANT PATHOGEN	0.026	*
OTU4375	Gsoil	Acremonium_persicinum	Ascomycota	LICHENICOLOUS	0.038	*
OTU10750	Gsoil	Ascomycota sp.	Ascomycota		0.037	*
OTU2440	Wood	Exophiala_moniliae	Ascomycota		0.002	**
OTU1443	Wood	Pterulaceae sp.	Basidiomycota		0.003	**
OTU12085	Wood	Fungi sp.	Fungi sp.		0.005	**
OTU2197	Wood	Xylodon_crustosus	Basidiomycota	CORTICICOLOUS	0.011	*
OTU4282	Wood	Basidiodendron_caesiocinereum	Basidiomycota	CORTICICOLOUS	0.025	*
OTU10426	Wood	Ascomycota sp.	Ascomycota		0.006	**
OTU10766	Wood	Meliniomyces sp.	Ascomycota	ENDOPHYTE	0.011	*
OTU3759	Wood	Ascomycota sp.	Ascomycota		0.010	**
OTU12052	Wood	Leotiomycetes sp.	Ascomycota		0.041	*
OTU11586	Wood	Ascomycota sp.	Ascomycota		0.017	*
OTU12228	Wood	Ascomycota sp.	Ascomycota		0.025	*
OTU1288	Wood	Hydnaceae sp.	Basidiomycota	ECM	0.023	*
OTU1622	Wood	Peniophorella_praetermissa	Basidiomycota	CORTICICOLOUS	0.029	*
OTU2108	Wood	Athelia sp.	Basidiomycota	CORTICICOLOUS	0.019	*
OTU3105	Wood	Hyphodontia_pallidula	Basidiomycota	ECM	0.023	*
OTU4270	Wood	Exophiala sp.	Ascomycota		0.023	*
OTU485	Wood	Mycena sp.	Basidiomycota	WRF	0.021	*
OTU4603	Wood	Herpotrichiellaceae sp.	Ascomycota		0.043	*
OTU4377	Wood	Herpotrichiellaceae sp.	Ascomycota		0.049	*

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001

Table 07 : Indicator species identified for each of the 3 habitats using the shared DNA+RNA dataset. "ecology" refers to the known trophic modes of the taxa. ECM, symbiotic ectomycorrhizal taxa, AM, symbiotic arbusculat mycorrhizal taxa, WRF, saprotrophic white rot taxa, corticicolous, saprotrophic wood associated taxa. p-values refer to significance of indicator values from 999 permutations (Monte Carlo Test); asterisks in the last column refer to the Monte Carlo Test p-value .

4.3.4 Network analysis

Network analysis is a powerful way to identify groups of taxa that are

likely to interact with each other within species-rich communities. In a preliminary analysis, we used the shared dataset to investigate the impact of metabarcoding using either DNA or RNA matrices on community networks. We further simplified the analysis by focusing not on all individual OTUs but on Basidiomycete families.

	DNA	cDNA	DNA + cDNA
edges	171	152	188
nodes	58	54	57
transitivity	0.5603774	0.5337	0.590425532
modularity	0.55261448	0.5921	0.574284178
communities	12	7	7

Table 08 : Main features of the networks of interactions calculated for Basidiomycete families using the the sharedDNA, sharedRNA and sharedDNA+RNA datasets.

We thus computed and draw three different networks for the "*shared DNA*", "*shared cDNA*" and "*shared DNA+cDNA*" datasets. These three networks did not significantly differ from each other as regard several of their global features (numbers of nodes and edges, transitivity, modularity)(**Table 08**). They however differ from each other with regard of the number of "communities" that can be deduced from these analysis, seven in the case of the "*shared cDNA*" and "*shared DNA+cDNA*" datasets, but 12 in the case of the "*shared DNA*" one. Furthermore, visual inspection of the networks highlighted obvious "qualitative" differences between them (**Figure 11**). For example, several families (e.g. the Sporidiobolales, Suilaceae and Mycenaceae) are poorly connected in the DNA network but become highly connected to many other families in either the *"shared cDNA*" or "*shared DNA+cDNA*" networks (**Figure 11**).

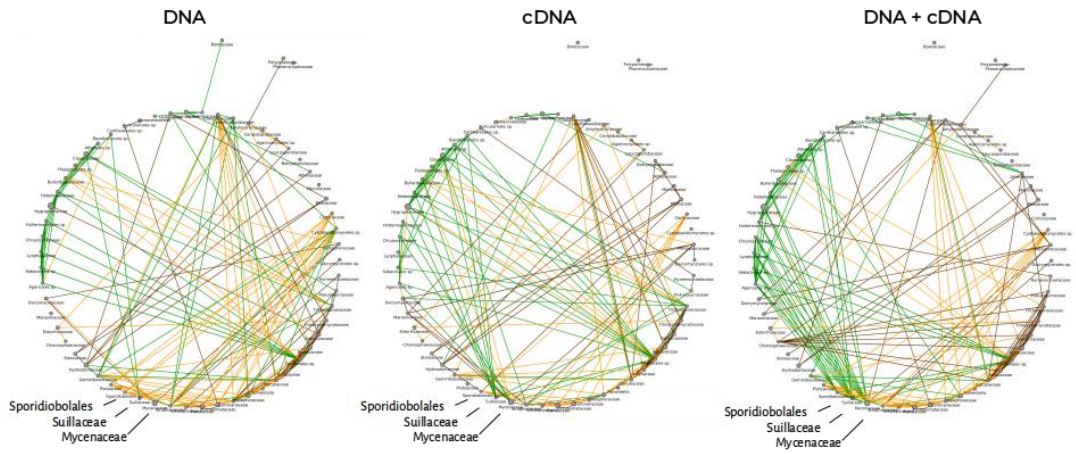


Figure 11 : Basidiomycota family network analysis performed separately using the sharedDNA, sharedRNA and sharedDNA+RNA datasets. Each point represent a Basidiomycete family, edges link families that were found to co-occur with each other in either the forest soils (color green lines, 71 edges in DNA, 58 edges in RNA, 84 edges in DNA+RNA), the grassland soils color orange lines, (77 edges in DNA, 52 edges in RNA, 71 edges in DNA+RNA) or the decomposing wood samples (color brown lines, 50 edges in DNA, 50 edges in RNA, 77 edges in DNA+RNA). Networks were calculated and drawn separately for the sharedDNA, sharedRNA and sharedDNA+RNA samples. The Sporidiobolales, Suillaceae and Mycenaceae represent three of the families for which their patterns of interactions with other families obviously differ between the different networks.

4.4 Discussion

4.4.1 Consequences of combined DNA/RNA analyses and trimming methodologies on read and OTU distribution

In our global dataset, we observed that the number of DNA-based OTUs is higher than the number of RNA-based ones. An excess of DNA-based OTUs can be explained by the presence in soil and wood of "legacy DNA", *i.e.* DNA from dead organisms that represent taxa that were once present but disappeared. This concept is based on the observation that extracellular DNA

persists for a longer time in the environment than RNA (Carvalhais et al., 2012; Prüfer et al., 2014). In the literature, however, the proportion of OTUs found only in DNA or RNA datasets varies substantially between studies. While some studies effectively found more DNA-based than RNA-based OTUs in their datasets (Pawlowski et al. (2014) and Pochon et al. (2015), others (Hu et al., 2016; Laroche et al., 2016) reported the opposite. In the specific case of fungal communities, Barnard et al., 2013 and Žifcáková et al., 2016 also reported an excess of DNA-based OTUs while diversity indices calculated on either DNA or RNA data were similar.

Although DNA-based OTUs outnumbers RNA-based once, we also observed a number of RNA-specific OTUs not observed in the DNA dataset. Several of these RNA-specific OTUs could represent RT-PCR artefacts or sequencing errors (Egge et al., 2013; Ficetola et al., 2015). The RNA conversion to cDNA requires a reverse transcriptase (RT) which can introduce punctual mutations in some of the cDNA sequences (Svarovskaia et al., 2003; Houseley & Tollervey, 2010). RT enzymes can also switch from one RNA molecule to another one, thus creating chimeric cDNA sequences (Cocquet et al., 2006). Random hexamer used to prime cDNA synthesis are known to introduce some nucleotide bias in the sequences (Hansen et al., 2010).

Finally, an insufficient sequencing depth can also explain differences between DNA and RNA datasets, including the presence of several of the RNA-specific taxa. Indeed, rare but physiologically active taxa have a higher probability to be absent in only one dataset simply because of their rarity.

To remediate to these problems, in molecular ecology, rare (very rare) OTUs are usually removed from the datasets in order to (i) minimize the

impact of sequencing errors, (ii) minimize the contribution of free DNA from dead organisms, (iii) minimize the random effect of sequencing depth and finally (iv) remove rare but genuine OTUs that are unlikely to play a significant role to the biological processes under study.

In most cases, in fungal ecology, OTUs identified by a single read in the whole dataset (called singletons) are eliminated (e.g. Baldrian et al., 2012; Voyron et al., 2016; Rinne et al., 2017). In other fields higher thresholds are sometime preferred (Laroche et al., 2017). In our case we implemented a threshold of 10 reads and observed that this resulted in a very significant increase in the proportion of shared RNA/DNA OTUs (**Figure 01**) compared to the datasets where only true singletons were removed. Under the hypothesis that shared RNA/DNA OTUs are unlikely to be technical artefacts because observed in two independent datasets with sequences generated from different matrices with two distinct protocols, we can conclude that the removal of OTUs with less than ten reads certainly improves the "quality" of the dataset by removing both spurious sequences and OTUs.

The second "trimming approach" was to eliminate OTUs that were specific to either the DNA or RNA datasets from which the 10 reads threshold had already been implemented. While this approached removed 62% of the initial OTUs, it eliminated only 11% of the reads (**Figure 01**). This demonstrates that DNA and RNA-specific OTUs predominate among rare OTUs, thus accentuating the doubtful nature of rare OTUs (artefacts, inactive or dead taxa). Within the "shared" dataset, defined by pooling all environmental samples, a minority (24% in the worst case, but less than 10% in two thirds of the samples) of the selected OTUs is not shared within individual samples, but

with RNA/DNA datasets from other sample. Although these OTUs could have also been eliminated we retained them because they represented a minority of the OTUs and several of them, although not shared in a given sample were shared in others.

4.4.2 Combined DNA/RNA analysis and fungal OTU abundance

In molecular microbial ecology, the relative abundance (in term of number of reads) of individual OTUs DNA and RNA metabarcoding datasets is usually used as a proxy to infer their relative abundance and level of activity in the community, respectively. The relationship between RNA-based read numbers and species "activity" can be demonstrated when cell number is known (Blazewicz et al., 2013), and is generally accepted for unicellular organisms (Denef et al., 2016; Laroche et al., 2017). For filamentous fungi, this direct correlation is difficult to establish, but several studies suggest this may be globally true (Baldrian et al., 2012; Blazewicz et al., 2013; Barnard et al., 2013; Žifcáková et al., 2016). A relative indication of activity and/or dormancy of a specie could then be indirectly evaluated by the rRNA:rDNA read ratio. At the level of individual OTUs we recorded important variations between OTUs, but for many of them the RNA/DNA read ratio was highly variable from one sample to the other. This could be taken as an indication that a given RNA/DNA ratio is not an intrinsic feature of a taxon but indeed reflects the level of activity of the corresponding taxon in the different sites at the time of sampling. At a higher taxonomic level, in the case of bacteria, a taxonomic dependence between rRNA and rDNA reads abundance and taxonomy (Denef et al., 2016). In our case, an evaluation at the class level revealed important

variations between taxa belonging to a specific family. However, specific fungal classes show a clear enrichment in either putatively "active" (high rRNA/rDNA ratios) taxa (e.g. the Tremelomycetes) or "inactive" (low rRNA/rDNA ratios) taxa (e.g. the Eurotiomycetes). These trends deserve further studies as rRNA/rDNA ratios also integrate species-specific features such as the relative number of rDNA copy number per haploid genome that can vary considerably and the density of nuclei per volume of cytoplasm, which can also vary considerably between fungal species.

4.4.3 Habitat and geographic site effects

It is expected that an accurate treatment of metabarcoding datasets that remove spurious taxa and better reflects their relative abundance within the studied communities must ultimately lead to better statistical support for genuine differences existing between the studied samples. This was tested by comparing fungal communities in highly contrasted habitats and sites for which "common mycological knowledge" indicates that significant differences in fungal community composition must exist. We observed that only the "*shared*" dataset allowed to discriminate all habitats and almost all sites, while both the "*all reads*" and "*10 reads*" datasets failed to give significant statistical supports to differences between two of the habitats and to up to half of the pairwise comparisons between sites. It would be of interest to repeat these analyses using (i) datasets from which had been eliminated OTUs represented with more than 10 reads, but also (ii) datasets specific for different taxonomic groups such as the Ascomycota and the Basidiomycota as what we call

common mycological knowledge on differences between habitats and sites is essentially based on data from macromycetes affiliated essentially to the Basidiomycota.

The "shared" dataset also seemingly improved the delineation of habitat-specific indicator species, as almost all species identified within this dataset were already detected by analyzing the "*all reads*" and "*10 reads*" datasets. By comparison, the separate analysis of the "*all reads*" and "*10 reads*" datasets gave a number of dataset-specific indicator species as well as a number of species exclusive of these two datasets. Functional annotation of the indicator species from decaying wood and forest soil habitats highlighted the robustness of this analysis as they were dominated by wood decomposers and ectomycorrhizal taxa, respectively. Although Ascomycete taxa dominate the grassland soil mycoflora, the total absence of Basidiomycete taxa among grassland indicator species was unexpected and indicate that many of the grassland basidiomycete OTUs have a restricted geographic distribution or are sensitive to site specific features (soil composition, local flora).

4.5 Conclusions

Accuracy of metabarcoding techniques used in ecological studies to probe microbial communities can be substantially affected by the nature of the nucleic acid matrix used (DNA versus RNA). Different datasets can indeed affect the outcome of downstream statistical analyses and modify our perception of environmental factore on microbial communities. Dataset trimming strategies relatively stronger than usually carried out (in fungal

ecology at least) are recommended. Trimmed *10reads* and *shared* datasets show higher ability to correctly discriminate environmental variables.

Removing OTUs not shared between DNA and RNA datasets improved the sensitivity and accuracy of the results. This is true for the sites and habitats analyzed in the present study. Further investigations are needed to evaluate the extent of these findings, trying at the same time to better understand if the RNA:DNA ratio could be used as a true indication of fungal OTUs activity.

4.5 Bibliography

Adamo M, Voyron S, Girlanda M, Marmeisse R. 2017. RNA extraction from decaying wood for (meta)transcriptomic analyses. *Canadian journal of microbiology* **10**: cjm-2017-0230.

Anderson IC, Parkin PI. 2007. Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *Journal of Microbiological Methods* **68**: 248–253.

Anderson IC, Parkin PI, Campbell CD. 2008. DNA- and RNA-derived assessments of fungal community composition in soil amended with sewage sludge rich in cadmium, copper and zinc. *Soil Biology and Biochemistry* **40**: 2358–2365.

Baldrian P, Kolařík M, Štursová M, Kopecký J, Valášková V, Větrovský T, Žifčáková L, Šnajdr J, Rídl J, Vlček Č, *et al.* 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME Journal* **6**: 248–258.

Barberán A, Bates ST, Casamayor EO, Fierer N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal* **6**: 343–351.

Barbi F, Prudent E, Vallon L, Buée M, Dubost A, Legout A, Marmeisse R, Fraissinet-Tachet L, Luis P. 2016. Tree species select diverse soil fungal communities expressing different sets of lignocellulolytic enzyme-encoding genes. *Soil Biology and Biochemistry* **100**: 149–159.

Barnard RL, Osborne CA, Firestone MK. 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *ISME J* **7**: 2229–2241.

Bastias BA, Anderson IC, Xu Z, Cairney JWG. 2007. RNA- and DNA-based profiling of soil fungal communities in a native Australian eucalypt forest and adjacent Pinus elliotti plantation. *Soil Biology and Biochemistry* **39**: 3108–3114.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* **13**: 218–224.

Blaser S, Prati D, Senn-Irlet B, Fischer M. 2013. Effects of forest management on the diversity of deadwood-inhabiting fungi in Central European forests. *Forest Ecology and Management* **304**: 42–48.

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal* **7**: 2061–2068.

Brown EA, Chain FJJ, Crease TJ, Macisaac HJ, Cristescu ME. 2015. Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities? *Ecology and Evolution* **5**: 2234–2251.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.

Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. 2012. Application of metatranscriptomics to soil environments. *Journal of Microbiological Methods* **91**: 246–251.

Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**: 127–131.

Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM. 2016. The ecologist's field guide to sequence-based identification of biodiversity (R Freckleton, Ed.). *Methods in Ecology and Evolution* **7**: 1008–1018.

Csardi GNT. 2006. The igraph software package for complex network research. *InterJournal Complex Systems* **1695**: 1695.

Damon C, Barroso G, Férandon C, Ranger J, Fraissinet-Tachet L, Marmeisse R. 2010. Performance of the COX1 gene as a marker for the study of metabolically active Pezizomycotina and Agaricomycetes fungal communities from the analysis of soil RNA. *FEMS Microbiology Ecology* **74**: 693–705.

De Cáceres M, Legendre P. 2009. Associations between species and groups of sites: Indices and statistical inference. *Ecology* **90**: 3566–3574.

Denef VJ, Fujimoto M, Berry MA, Schmidt ML. 2016. Seasonal succession leads to habitat-dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. *Frontiers in Microbiology* **7**: 606.

Dinno A. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *Stata Journal* **15**: 292–300.

Dowle E, Pochon X, Keeley N, Wood SA. 2015. Assessing the effects of salmon farming seabed enrichment using bacterial community diversity and high-throughput sequencing (P Sobecky, Ed.). *FEMS Microbiology Ecology* **91**: fiv089.

Dufrene M, Legendre P, Monographs SE, Aug N. 2013. Species Assemblages and Indicator Species : The Need for a Flexible Asymmetrical Approach. *Ecological Monographs* **67**: 345–366.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B. 2013. 454 Pyrosequencing to Describe Microbial Eukaryotic Community Composition, Diversity and Relative Abundance: A Test for Marine Haptophytes (S Lin, Ed.). *PLoS ONE* **8**: e74371.

Ferreira-Cerca S, Pöll G, Gleizes PE, Tschochner H, Milkereit P. 2005. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Molecular Cell* **20**: 263–275.

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, *et al.* 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* **15**: 543–556.

Gardes M, Bruns T. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of …. *Molecular Ecology Notes*: 9–10.

Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Applied and Environmental Microbiology* **66**: 5488–5491.

Guardiola M, Wangensteen OS, Taberlet P, Coissac E, Uriz MJ, Turon X. 2016. Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ* **4**: e2807.

Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* **38**: e131–e131.

Houseley J, Tollervey D. 2010. Apparent Non-Canonical Trans-Splicing Is Generated by Reverse Transcriptase In Vitro (T Preiss, Ed.). *PLoS ONE* **5**: e12271.

Hu SK, Campbell V, Connell P, Gellene AG, Liu Z, Terrado R, Caron DA. 2016. Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS Microbiology Ecology* **92**: fiw050.

Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE, *et al.* 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* **82**: 666–677.

Junker BH, Schreiber F. 2008. *Analysis of biological networks*. Wiley-Interscience.

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, *et al.* 2013. Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology* **22**: 5271–5277.

Lalev AI, Nazar RN. 2001. A Chaperone for Ribosome Maturation. *Journal of Biological Chemistry* **276**: 16655–16659.

Laroche O, Wood SA, Tremblay LA, Ellis JI, Lejzerowicz F, Pawlowski J, Lear G, Atalah J, Pochon X. 2016. First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. *Marine Environmental Research* **120**: 225–235.

Laroche O, Wood SA, Tremblay LA, Lear G, Ellis JI, Pochon X. 2017. Metabarcoding monitoring analysis: the pros and cons of using co-extracted environmental DNA and RNA data to assess offshore oil production impacts on benthic communities. *PeerJ* **5**: e3347.

Lennon JT, Jones SE. 2011. Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology* **9**: 119–130.

Oksanen J. 2013. Vegan: ecological diversity. *Available from: hokudai. ac. jp [Accessed 15 August 2014]*.

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources* **14**: 1129–1140.

Peres-Neto PR, Legendre P, Dray S, Borcard D. 2006. Variation partitioning of species data metrices: estimation and comparison of fractions. *Ecology* **87**: 2614–2625.

Pochon X, Wood SA, Keeley NB, Lejzerowicz F, Esling P, Drew J, Pawlowski J. 2015. Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. *Marine Pollution Bulletin* **100**: 370–382.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C, *et al.* 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43–49.

Purahong W, Pietsch KA, Lentendu G, Schöps R, Bruelheide H, Wirth C, Buscot F, Wubet T. 2017. Characterization of unexplored deadwood mycobiome in highly diverse subtropical forests using culture-independent molecular technique. *Frontiers in Microbiology* **8**.

Rinne KT, Rajala T, Peltoniemi K, Chen J, Smolander A, Mäkipää R. 2017. Accumulation rates and sources of external nitrogen in decaying wood in a Norway spruce dominated forest. *Functional Ecology* **31**: 530–541.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, *et al.* 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**: 7537–7541.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, *et al.* 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **109**: 6241–6246.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.

Simpson EH. 1949. Measurement of Diversity. *Nature* **163**: 688–688.

Sindaco R, Savoldelli P, Selvaggi A. 2009. *La Rete Natura 2000 in Piemonte - I Siti di Importanza Comunitaria*. Torino: Regione Piemonte.

Spellerberg IF, Fedor PJ. 2003. A tribute to Claude-Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' Index. *Global Ecology and Biogeography* **12**: 177–179.

Stone L, Roberts A. 1990. The checkerboard score and species distributions. *Oecologia* **85**: 74–79.

Svarovskaia ES, Cheslock SR, Zhang W-H, Hu W-S, Pathak VK. 2003. Retroviral mutation rates and reverse transcriptase fidelity. *Front Biosci* **8**.

Tallarida RJ, Murray RB. 1987. Chi-Square Test. In: Manual of Pharmacologic Calculations. New York, NY: Springer New York, 140–142.

Tedersoo L, Wardle DA, Lindahl BD. 2014. Disentangling global soil fungal diversity. *Science* **346**: 1052–1053.

Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J. 2015. Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environmental Science and Technology* **49**: 7597–7605.

Voyron S, Ercole E, Ghignone S, Perotto S, Girlanda M. 2017. Fine-scale spatial distribution of orchid mycorrhizal fungi in the soil of host-rich grasslands. *New Phytologist* **213**: 1428–1439.

White T, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Shinsky J, White T, eds. PCR Protocols: A Guide to Methods and Applications. San Diego: Academic Press, 315–322.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Williams RJ, Howe A, Hofmockel KS. 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* **5**: 1–10.

Wu YT, Wubet T, Trogisch S, Both S, Scholten T, Bruelheide H, Buscot F. 2013. Forest Age and Plant Species Composition Determine the Soil Fungal Community Composition in a Chinese Subtropical Forest. *PLoS ONE* **8**.

Zak DR, Holmes WE, White DC, Peacock AD, Tilman D. 2003. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology* **84**: 2042–2050.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614–620.

Žifčáková L, Větrovský T, Howe A, Baldrian P. 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environmental Microbiology* **18**: 288–301.

Chapter 5

Diversity of fungal DyPs expressed in different habitats

5.0 Premise

In this chapter we tested our initial hypothesis that a potential role of fungal DyP peroxidases in organic matter could be reflected by differences in DyP gene diversity and expression levels in environments that differ from each other with respect to lignin input and turnover: wood, forest soils and grassland soils.

To that aim, we developed and implemented a solution hybrid selection capture (SHS) of DyP sequences expressed in these different environments. Captured environmental DyP cDNAs were subjected to high-throughput gene sequencing to reconstitute full-length or partial gene sequeces that were positioned in the global DyP phylogenetic tree.

This chapter makes use of the results obtained in all of the previous chapters as it necessitated (i) the design of capture probes based on the known DyP diversity (Chapter 02), (ii) the use of RNA extracted from wood (Chapter 03) and soils (Chapter 04)), and (iii) the confrontation of the results to the known DyP phylogeny (Chapter 02) and fungal diversity in the studied environments (Chapter 04).

5.1 Introduction

Plant biomass degradation represents an essential step for the terrestrial biogeochemical cycles of carbon, nitrogen and phosphorus, three elements that are abundant in complex organic molecules (Bardgett et al., 2005). While several groups of macro and microorganisms are involved in this complex process (German et al., 2011), fungi are considered as the main players in plant organic matter (POM) decay in terrestrial environments (Bardgett et al., 2005; Damon et al., 2012; Schneider et al., 2012). POM degradation is a largely enzymatic process. During the course of evolution, a vast array of different enzyme families active on plant cell wall (PCW) polymers have been selected and their combined action on POM contribute to its degradation (Floudas et al., 2012; Nagy et al., 2017).

Many of these enzymes have been extensively studied not only because of their central role in POM degradation, but also because of their participation to other biological processes such as virulence of plant pathogens and of their contribution to numerous biotechnological processes (Dashtban et al., 2009; Alvarez et al., 2016). Several of these enzymes have been classified in specialized databases (Lombard et al., 2014; Fawal et al., 2013) according to phylogenetic relatedness and mode of action on plant polymers (essentially polysaccharides and lignin). Despite the wealth of knowledge that has accumulated on both POM degradation and the contribution and mode of action of several enzyme families involved, it is suspected that additional enzyme families, yet to be described, may contribute to this process (Kellner et al., 2014). This is the case for lignin degradation. Lignin represents the

second most abundant plant polymer after cellulose (Welker et al., 2015). It is an irregular polymer resulting from the polymerization of between one to three, depending on the plant species, different phenylpropanoid monomers (Pereira et al., 2004). It is described as a "recalcitrant molecule" whose enzymatic degradation is difficult to achieve (Hibbet et al., 2016). Because of this characteristic, it is believed that the presence and abundance of lignin in plant cell walls represent an obstacle to microbial attack in living plants and contribute to the durability of wood, a highly lignified plant material (Bugg et al., 2011).

Thus far, only one group of enzymes has been formally described as participating both directly and indirectly to lignin hydrolysis. These enzymes belong to the class II peroxidases (E.C. 1.11.1.x) family found in the Fungi. Among class II peroxidases, so-called lignin peroxidases (LiP, E.C. 1.11.1.14) are active on the non-phenolic moieties of lignin, thus contributing to its depolymerisation (Hofrichter et al., 2010; Lundell et al., 2010). Other class II peroxidases, called manganese peroxidases (MnP, E.C. 1.11.1.13) act indirectly by oxidizing Mn^{3+} to Mn^{2+} that contributes to the hydrolysis of lignin phenolic moieties (Hofrichter et al., 2010; Lundell et al., 2010). Finally, a third group of class II peroxidases, versatile peroxidases (VP, E.C. 1.11.1.16) display both lignin and Mn peroxidase activities (Hofrichter et al., 2010; Lundell et al., 2010). These three classes of related enzymes are specific to the Agaromycetes and recent phylogenomic studies (Floudas et al., 2012; Koheler et al., 2015; Nagy et al., 2016) suggest that the diversification of the corresponding gene families correlates with the diversification of so-called white rot fungi, the only functional group of microorganisms formally demonstrated to be

capable of degrading lignin enzymatically (Bugg et al., 2011).

Besides class II peroxidases, it is suspected that other enzymes could participate to lignin degradation. On the one hand, significant lignin loss has been monitored in ecosystems from which fungi with class II peroxidases are most likely absent (Li et al., 2017) and on the other hand, several biochemical studies have identified other secreted microbial enzymes capable of either oxidizing Mn^{3+} to Mn^{2+} or of being active on so-called low molecular weight lignin model compounds (veratryl alcohol; adlerol, a nonphenolic β-O-4 lignin model dimer) (Liers et al., 2010; Fernández-Fueyo et al., 2015). Despite these lines of evidence, their actual role in PCW lignin degradation is still debated because of the simplistic nature of artificial lignin substrates and because of the catalytic properties (e.g. substrate affinity, optimum pH) of the corresponding enzymes (Eichlerová et al., 2000). "Dye decolorizing heme peroxidases" (DyPs) are among these enzymes suspected to participate to lignin hydrolysis in both bacteria and fungi (Zamocky et al., 2015). This particular peroxidase family, phylogenetically distinct from class II peroxidases, was first described as a fungal family of extracellular enzymes capable of decolorizing industrial anthraquinone dyes (Yoshida & Sugano, 2015; Linde et al., 2015). Specific fungal members of this family are active on artificial lignin models (Liers et al., 2013; de Gonzalo et al., 2016) and/or Mn^{3+} (Fernández-Fueyo et al., 2015) and addition of this enzyme to enzyme cocktails used to degrade plant biomass improves the digestibility of this material (Zorn et al., 2009 - patent WO2009EP58871.). In phylogenomic studies the DyP gene family is also one of the gene families whose diversification correlates with the diversification of white rot fungal species

(Nagy et al., 2016; see chapter DYP). A detailed phylogenetic analysis of this gene family in the Fungi (Linde et al., 2015; chapter 02) showed that it can be divided in several distinct clades and that only one of these clades seems to correspond to secreted enzymes. This later clade encompasses all biochemically characterized fungal DyPs and is only present in white rot species, while members of the other clades encoding putative intracellular enzymes are present in the genomes of fungal species, irrespective of their trophic mode. All these different observations suggest a role for extracellular DyPs in lignin degradation although they could also be involved in the detoxification of soluble fungistatic plant secondary metabolites such as wood extractives as suggested by (Valette et al., 2017).

To further investigate the potential role of fungal DyP enzymes in lignin degradation we carried out a genetic profiling of fungal DyP genes expressed in different natural environments, which differ with respect to lignin content and degradation rates. We made the hypothesis that genes encoding extracellular DyPs should be more diverse and present higher transcription levels compared to those encoding intracellular enzymes in environments with higher lignin degradation rates. The three selected environments/substrates are (i) decaying wood, a substrate highly enriched in lignin and hosting fungal communities also enriched in lignicolous species, (ii) forest soil, a substrate with high inputs of lignin-enriched POM and numerous lignicolous fungal species and (iii) grassland soil, a substrate with high inputs of herbaceous POM poor in lignin and which paradoxically often accumulates lignin or poorly-degraded lignin possibly because of the presence of saprotrophic fungal communities encompassing few lignicolous species

(Thevenot et al., 2010). From a technical point of view, DyP genes expressed by the different fungal species present in the environmental samples were isolated by "solution hybrid selection capture" (SHS) of DyP cDNA synthesized from total environmental RNA (Denonfoux et al., 2013; Bragalini et al., 2014; Gasc et al., 2016). This procedure that does not rely on PCR, combined with high-throughput sequencing, allows the isolation and reconstruction of full-length or nearly full-length genes for precise phylogenetic affiliation and eventual heterologous production of the enzyme for biochemical characterization. This procedure that uses multiple degenerate oligonucleotidic probes has also been shown to lead to the selection of novel sequences highly divergent from known ones (Parisot et al., 2012).

5.2 Materials & Methods

5.2.1 Environmental samples

Soil and wood samples were collected in five geographic sites located in Italy and France (**Table 01**). In order to take into account different fungal communities and the largest number of fungal taxa, sites were selected in different bio-geographic regions of Europe (Atlantic/continental, Alpine and Mediterranean), encompass different plant species and associations and highly contrasted climatic and soil conditions (**Table 02**). One grassland soil, one forest soil and one decomposing wood composite sample was collected in each of the five sites (*i.e.* 15 environmental samples). Soil and wood sampling procedures were similar at each site to allow comparisons across sites

(**Chapter 04**). Soil and wood samples were processed within 24 hours following collection and kept frozen at -75°C before DNA and RNA extraction (**Chapter 04**).

5.2.2 RNA extraction and cDNA synthesis

Environmental RNA was extracted from 2 g of each soil sample using the MoBio "Power Kit RNA soil extraction" (Hilden, Germany) as described in the kit manual. Total RNA was extracted from 200 mg of decomposing wood as described in Adamo et al. (2017). When necessary ($OD_{260/280}$ ratio of the RNA preparation below 1.0), the "RNA Clean & Concentrator Kit" by Zymo Research (Irvine, CA USA) was used to remove contaminating molecules.

The Mint-2 cDNA synthesis and amplification kit (Evrogen, Moscow, Russian Federation) was used for the synthesis of cDNAs from polyadenylated eukaryotic mRNAs, using 500 ng of total environmental RNA. Briefly, first-strand cDNA synthesis was primed using a CDS adaptor, which anneals to the 3' poly-A tail of eukaryotic mRNA. Upon reaching the 5' end of the mRNA the Mint reverse transcriptase (RT) enzyme adds a few deoxycytidine (C) nucleotides at the end of all of the neo-synthesized single-stranded cDNAs (ss-cDNAs). These extra Cs are used to anchor a so-called PlugOligo adaptor whose complementary sequence is added by the RT enzyme at the 3' end of all ss-cDNAs. As a result, all ss-cDNAs are bordered at both their 3' and 5' ends with a common 23-nt-long M1 sequence present in both the CDS and PlugOligo adapters. All neosynthesized ss-cDNAs are then converted and amplified as double stranded (ds) cDNAs in a single PCR reaction using the M1

primer and the Evrogen “Encyclo Plus PCR kit”.

5.2.3 Solution hybrid selection capture

Capture probe design and synthesis

All publicly available (April 2015) DyP sequences of fungal origin were identified by BLAST search and collected from GenBank (http://www.ncbi.nlm.nih.gov/genbank/), the Joint Genome Institute database (Grigoriev et al., 2014, http://jgi.doe.gov/) and PeroxiBase (http://peroxibase.toulouse.inra.fr/). From these 751 DNA coding sequences (CDSs), from 72 basidiomycete and 36 ascomycete species, 80 nucleotide-long degenerated probes were designed using the KASpOD software (Parisot et al., 2012). These 69 probes were separated in two groups, a group of 18 probes specific for ascomycetous DyP sequences; a group of 51 probes designed to be specific with DyPs belonging to the Basidiomycota.

Each synthesized ssDNA oligonucleotids included the 80 nt sequence of the probe (X) bordered at both the 5' and 3' end by a short adapter sequence used to convert the probe as dsDNA by PCR and to initiate transcription using the T7-RNA polymerase (Grinke et al., 2009): 5'-TCGCACCAGCGTGT-(X)-CACTGCGGCTCCT-3'. After PCR amplification and gel purification, the dsDNA oligonucleotides were used as templates to synthesize botinylated RNA probes by *in vitro* transcription using the “MEGAscript T7 Transcription kit” (Ambion, USA) and biotin-dUTP (TEBU Bio, France). RNA probes were then mixed together in equimolar amounts by taking into account the degree of degeneration of each probe.

SITE ID	SCI Site Code	Collection date	Location	Coordinates	Altitude (m asl)	Mean annual temperture [°C]	Mean annual precipitation [mm]	Parental rock	Sample ID	Sample Type	Ecological features
Mandria	IT1110079	2/28/2015	Venaria Reale, TO	45°18'N, 7°55'E	300	12.3	860	Quaternary sediments	MB	Forest soil*	Sub-Atlantic and medio-European oak or oakhornbeam forests of the Carpinion betuli (9160)
									MP	Grassland soil*	Molinia meadows on calcareous, peaty or clayey-siltladen soils (*Molinion caeruleae*) (6410)
									MW	Decaying wood **	*Quercus robur; Carpinus betulus; Acer campestre; Corylus avellana*
Foresto	IT1110030	4/1/2015	Bussoleno, TO	45°14'N, 7°10'E	500	11,4	799	Limestone	FB	Forest soil*	Pannonian woods with *Quercus pubescens* (91H0)
									FP	Grassland soil*	Semi-natural dry grasslands and scrubland facies on calcareous substrates (Festuco-Brometalia) (*important orchid sites) (6210*)
									FW	Decaying wood **	*Quercus pubescens, Prunus avium; Cotynus coggygria*
Saone	FR2600976	6/8/2015	Ouroux-sur-Saône, 01	46°42'N, 4°54'E	185	11	781	Quaternary sediments	SB	Forest soil*	Alluvional forests at *A. glutinosa* and *F. excelsior* (91E0)
									SP	Grassland soil*	Managed low-nutrient meadows (6510)
									SW	Decaying wood **	*Salix alba, Fraxinus excelsior, Alnus glutinosa*
Creus	IT1160057	6/15/2015	Chiusa di Pesio, CN	44°20'N, 7°68'E	1200	8,21	1289	Quartzites	CB	Forest soil*	Acidophilous Picea forests of the montane to alpine levels (Vaccinio-Piceetea) with an *Abies alba* prevalence (9410 - 42.25)
									CP	Grassland soil*	Formazioni erbose secche seminaturali e facies coperte da cespugli su substrato calcareo (Festuco-Brometalia) (6210)
									CW	Decaying wood **	*Abies alba; Fagus sylvatica; Laburnum alpinum*
Lombarda	IT1160023	6/18/2015	Vinadio, CN	44°20'N, 7°14'E	2000	2,9	695	Gneiss	LB	Forest soil*	Alpine *Larix decidua* forest (9420)
									LP	Grassland soil*	Species-rich *Nardus* grasslands, on siliceous substrates in mountain areas (6230)
									LW	Decaying wood **	*Larix decidua*

Table 01 - Main characteristics of the sampling sites: Compared to Chapter 04, an additional geographic site from France ("Saone") was included in the present study.

Sample ID	Clay (< 2 m)	Faint silt (2/20 m)	Coarse silt (20/50 m)	Faint sand (50/200 m)	Coarse sand (200/2000 m)	Nitrogen (N)	Phosphorous (P2O5)	Oragnic Carbon (C)	Organic matter	Volatiles	C/N	Cation exchange capacity	Calcaire (CaCO3)	pH	Extractives	Klason lignin content	acid-soluble lignin
	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/100g	-	cmol⁻	g/kg	-	%	%	%
MB	197	295	267	167	74	2,22	0,017	33,4	57,7	9,12	15	5,55	<1	4,53	N/A	N/A	N/A
MP	194	312	290	163	41	2,07	0,026	22,7	39,3	7,49	11	6,1	<1	5,7	N/A	N/A	N/A
MW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,15	33,75	2,93
FB	124	140	126	347	263	3,95	0,03	71,6	124	11,7	18,1	22,8	218	8,08	N/A	N/A	N/A
FP	117	139	143	413	188	3,95	0,02	71,9	124	10	18,2	18,7	198	8,14	N/A	N/A	N/A
FW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	UN	UN	UN
SB	561	358	50	18	13	4,32	0,065	48,4	83,8	16,1	11,2	37,5	<1	6,56	N/A	N/A	N/A
SP	716	224	43	12	5	8,91	0,038	92	159	27	10,3	60	<1	6,59	N/A	N/A	N/A
SW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1,61	33,65	2,75
CB	211	305	88	98	298	5,65	0,053	103	178	22,2	18,3	13,1	<1	3,93	N/A	N/A	N/A
CP	200	267	109	114	310	5,44	0,073	56,5	97,8	13,3	10,4	8,57	<1	4,72	N/A	N/A	N/A
CW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,06	44,31	1,56
LB	269	247	82	118	284	5,72	0,053	87	150	20,4	15,2	11,6	<1	4,81	N/A	N/A	N/A
LP	340	270	76	107	207	7,96	0,081	99,3	172	23,1	12,5	10,9	<1	4,99	N/A	N/A	N/A
LW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	4,56	48,02	0,9

Table 02 : Physicochemical characteristics of the studied soil and wood samples. Soil analyses were performed by the "laboratoire INRA d'analyse des sols d'Arras" (https://www6.hautsdefrance.inra.fr/las) using standard protocols including ISO protocols. "Volatiles" represent mass loss after combustion at 550°C. Wood lignin and extractive contents were assayed by Dr. Harald Kellner, Technical University of Dresden (D). N/A, not applicable; UN, not available. Site abbreviations as presented in Table 1.

Sequence Capture

cDNA sequence capture was performed as described by Denonfoux et al., (2013) and Bragalini et al., (2014). briefly: 2µg of heat denaturated cDNA were hybridized to 500ng of a mix of biotinylated RNA probes for 24h at 65°C in microcentrifuge tubes. The probes/cDNA hybrids were then specifically captured on streptavidin-coated paramagnetic beads (Dynabeads M-280 Streptavidin, Invitrogen, Waltham, MA USA). After different washing steps to eliminate contaminating cDNAs, the captured cDNAs were detached from the beads using 0.1M NaOH and purified using the "Qiaquick PCR purification kit" (Qiagen, Hilden, Germany). Captured cDNAs were further amplified using "Encyclo Plus PCR kit" and the M1 primer (Schmidt & Mueller 2011); were purified using the "Qiaquick PCR purification kit" and subjected to a second cycle of cDNA capture identical to the first one. At the end of the second capture cycle, purified and re-amplified cDNAs were quantified by spectrophotometry (NanoDrop TM 2000, Thermo Scientific, Waltham, MA USA). Size distribution of the captured amplified cDNAs was evaluated by running 250ng on an ethidium bromide-stained 1% agarose gel. The enrichment in target sequences along the sequence capture protocol was evaluated via a semi-quantitative PCR using DyP-specific primers (Kellner et al., 2014; Bragalini et al., 2014).

High-throughput sequencing sample preparation

Since known fungal DyP CDSs range in size from 1500 bp to 2000 bp, captured cDNA were first size fractionated to sequence only putatively full-length cDNAs and to eliminate truncated sequences and shorter and

longer contaminating DNA molecules. A first size fractionation of 1 µg of cDNA (between 1400 and 2000 nt) was performed using a "BluePippin" apparatus (Sage Science, Beverly, MA, USA). Fractionated cDNA was quantified by fluorimetry using the "Qubit dsDNA HS Assay Kit" and "Qubit Fluorimeter 2.0" (Thermo Fisher Scientific) and re-amplified by PCR using the M1 primer and "KAPA Hi-Fi *Taq* polymerase" (KAPA Biosystems, Wilmington, MA USA). The 50 µl PCR mix contained 10 ng of cDNA, 10 µl of 5X KAPA HiFi Fidelity Buffer, 1.5 µl 10mM dNTPs, 10 µl of the 10 µM concentrated M1 primer and 1U of KAPA HiFi Taq polymerase. After an initial denaturation at 94°C for 3 min, cDNA fragments were amplified for 25 cycles comprising 20 s at 98°C, 15 s at 66°C and 1min at 72°C. After a final elongation at 72°C for 3 min, 5 µl of PCR products were run on an ethidium bromide-stained 1% agarose gel. The Beckman Coulter "Agencourt® AMPure® XP kit" (Brea, CA USA) was used to perform a final purification and partial size fractionation of the captured cDNA. A 0.5x concentration of beads was used in order to remove small unspecific fragments from the eluted material. Purified cDNA quality was assessed by means of the "Qubit dsDNA HS Assay Kit" and "Qubit Fluorimeter 2.0", while the solution purity was assessed by measuring the OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios (NanoDrop TM 2000).

5.2.4 High-throughput sequencing

Two sequencing high-throughput technologies were implemented. In a preliminary experiment, one captured cDNA sample was sequenced in

parallel using the single molecule PacBio RSII platform and a SMRT cell M1 ("Plateforme Génomique INRA", Castanet-Tolosan, France) and the Illumina MiSeq 2x250bp ("I.G.A. Technologies", Udine, Italy). All other captured cDNA samples were sequenced using the Illumina HiSeq 2000 2x250 bp technology (I.G.A. Technologies, Udine, Italy) which uses the same protocol as the Illumina MiSeq 2x250bp for the sample preparation.

5.2.5 Bioinformatic analyses

PacBio data treatment

PacBio raw data (bax.h5 files) were converted in fastq format and demultiplexed using sequence tags to attribute read sequences to each sample. Adapter sequences from Mint2 cDNA kit were trimmed using custom Perl script. For each sample, trimmed sequences were clustered with CD-HIT-EST (version 4.6, Limin et al, 2012) using a 96 % identity threshold. Then sequences inside each cluster were aligned with CAP3 (Huang et al, 1999). Finally, frequent deletion/insertion errors were corrected manually using the alignment from CAP3 in order to determine a composite sequence for each cluster.

Illumina MiSeq and HiSeq data treatment

Adapter sequences were eliminated using custom Perl script enabling the detection of partial adapter sequences. Sequence quality was evaluated with PRINSEQ (Schmieder and Edwards, 2011) ; bases with quality lower than 20 were eliminated and only sequences longer than 60 bases were kept. Then

trimmed sequences were assembled using IDBA-UD (Peng et al, 2012) (default parameters) and resulting contigs were further assembled with CAP3 (Huang et al, 1999) (default parameters) to obtain longer contigs.

Search for peroxidase sequences

Similarity search between the custom peroxidase database and obtained sequences (composite sequences from PacBio data, trimmed Illumina reads, contigs from Illumina reads) was performed using DIAMOND (Buchfink et al, 2015), with the BLASTx command in « sensitive » mode and a maximal evalue of $1e^{-05}$. Sequences with high similarity were counted.

Abundance of complete DyP sequences (from PacBio treatment) in environmental samples

Illumina HiSeq reads were mapped to the 6 complete sequences derived from PacBio data using Bowtie2 (Langmead and Salzberg, 2012) and the sensitive mode. Number of reads aligned to each complete sequences was reported.

5.2.6 Phylogenetic analysis

Captured DyP sequences were added to the DyP "global phylogeny dataset" (**Chapter 02**), which includes almost all *ca.* 800 publically available full-length fungal DyP sequences; starting with a methionine and ending with a stop codon, longer than 1000 bp, and without gaps in the regions where amino-acids essential for catalysis have been identified (as described in Linde at al., 2015b). In the case of the *Aspergillus* and the *Penicillium* genera for

which the genomes of numerous species were available, only sequences from *A. oryzae* and *P. italicum* were included in the phylogenetic analysis. Sequences were aligned using Muscle (Edgar, 2004). ModelFinder (Kalyaanamoorthy et al., 2017), as implemented on the IQ-Tree web Server (Trifinopoulos et al., 2015), was used to find the best substitution model (LG+R). IQ-Tree web Server was also used to perform a Maximum Likelihood phylogenetic analysis using IQ-Tree (Nguyen et al., 2015). One thousand bootstrap replicates were conducted to assess tree topology.

Sequence Similarity Networks (SSN) were computed using the EFI - Enzyme Similarity Tool (Gerlt et al., 2015) starting from the complete set of DyP protein sequences. The lowest alignment score limit for the output file was 52. The full-network output file was visualized with Cytoscape 3.5.1 (Shannon et al., 2013). The network is represented with an “organic” layout and only edges with an alignment score > 100 are visualized.

5.2.7 Heterologous expression of DyP genes in *Podospora anserina*

Sequence optimization and synthesis

Nucleotide sequences of environmental DyP coding sequences were modified to meet *Podospora anserina* codon usage (http://www.kazusa.or.jp/codon) while maintaining the original amino acid sequences of the environmental proteins. For that purpose, the online codon optimization software COOL (http://cool.syncti.org) was used, keeping the default settings. "Optimized" nucleotide coding sequences without stop codons were synthetized *in vitro* (Sigma-Aldrich) with on their 5'-end four

randomly selected nucleotides followed by an *XbaI* cloning site upstream of the ATG and on their 3'-end a *HindIII* site followed by four randomly selected nucleotide. Synthesized sequences were first cloned in the *XbaI-HindIII* cloning sites of the pYes2.1V5-His yeast expression plasmid. The *XbaI-HindIII* fragments containing the DyP sequences were then subcloned in the *XbaI-HindIII* cloning sites of the *P. anserina* pBC-Hygro expression vector (Silar, 1995). Cloning in these sites placed the DyP ORFs under the control of the strong pBC-Hygro expression vector gene promoter (gene promoter name inside vector map is T3) and created a C-terminal in-frame fusion of the DyP polypeptide with a 9-amino acid long Influenza virus hemagglutinin (HA) antigenic sequence. The pBC-Hygro plasmid confers chloramphenicol resistance to *E. coli* and hygromycin B resistance to *P. anserina*.

The "S" wild-type strain of *P. anserina* whose genome was entirely sequenced (Espagne et al., 2008) was used as recipient strain for the different pBC-Hygro + DyP plasmids. *P. anserina* protoplast transformation was performed as described by Brygoo and Debuchy (1985) and transformants were selected on a medium supplemented with 75 $mg.ml^{-1}$ of hygromycin. Culture conditions, media and genetic methods are described in Esser (1974).

Western Blot detection of the HA-tagged DyP proteins produced in P. anserina

Mycelia of *P. anserina* transformants grown for 60 hours on a hygromycin-supplemented selective medium were placed in a 2 ml FastPrep (MP Biomedicals) tube containing 50µl of lysis-solution (Tris-HCl pH 7.5 20 mM; NaCl 100 mM; Na_2EDTA 1 mM; Triton X-100 0.05%) and 250 µl of glass

beads (0.25 - 0.5 mm diameter). Cell disruption was performed by shaking at maximal speed in a "FastPrep-24™ 5G" Instrument (MP Biomedicals) for 40 sec. After addition of 120µl of loading buffer 5x (Tris pH 6.8 0.4M; 1% β-mercaptohetanol; 1% SDS; 30% glycerol and as bromophenol blue as need to turn the solution blue), the samples were incubated for 5 min at 100°C, centrifuged for 5 min at 23,000 rpm and the protein-containing supernatant stored at -70°C.

Denatured proteins were separated by electrophoresis (100V, 100mA) on a 10% polyacrylamide denaturing gel (Biorad Criterion TGX stain free 12%) in a TrisHCl, 25 mM; glycine 192 mM; SDS 0.1% buffer. After separation, proteins were transferred by western blot on a nitrocellulose membrane which was incubated in 5% Biorad Blotting-grade Blocker in 1% TBS buffer for 1.5 h. The membrane was then incubated for 24 h in the presence of a 15,000 times diluted mouse monoclonal anti-HA primary antibody (Sigma H9658). Afterwards the membrane was incubated 1 h in the presence of a rabbit antimouse IgG antibody conjugated to a horseradish peroxidase (Sigma A9044). After subsequent washing steps, peroxidase activity was visualized using the Biorad Clarity Western ECL Substrate detection kit.

5.3 Results

5.3.1 Explorative probe design

751 putatively full-length or partial fungal DyP coding sequences were identified in public databases. Many of them were identified in full genome sequences of 35 Ascomycota and 72 Basidiomycota species. From these sequences, two sets of phylum-specific degenerated explorative 80 bp-long probes were generated. The degree of degeneration of Ascomycota probes (18 probes, 12 degenerated; DyP.A probes set) ranged from 8 to 1024. The degree of degeneration of Basidiomycota probes (51 probes, 40 degenerated; DyP.B probes set) ranged from 8 to 1024 (see **Attachment B**). Because of the high level of sequence diversity found among fungal DyP sequences, these different probes did not cover the entire diversity of this gene family. Percentage of coverage of Ascomycota sequences, for a single probes, ranged from 2.3% to 12.3%; for Basidiomycota, from 0.2% to 4.5%. Globally the Ascomycota and Basidiomycota probes were susceptible to hybridize to 40.41% and 36.38% of the known Ascomycota and Basidiomycota known sequences, respectively. These figures must be considered as minimal figures as probes designed on full-length sequences may eventually match the unknown moieties of partial DyP sequences.

5.3.2 Validation of the DyP sequence capture protocol:

Ascomycetes and Basidiomycete probes were separately labelled and used to capture DyP sequences from a single cDNA sample originating from RNA

extracted from the SB forest soil sample. After two successive rounds of sequence capture, PCR-amplified cDNAs ranged between *ca.* 50 and 2000 bp (results not shown). Semi quantitative PCRs performed using DyP-specific primers and different quantities of cDNAs before and after one and two cycles of capture clearly showed an increase in abundance of DyP sequences before and after capture (**Figure 01**). Since known full-length fungal DyP open reading frames (ORFs) range in size between 1400 and 2000 bp, amplified cDNA fragments of between 1400 and 2000 bp in length were size-selected before being submitted to high-throughput sequencing.

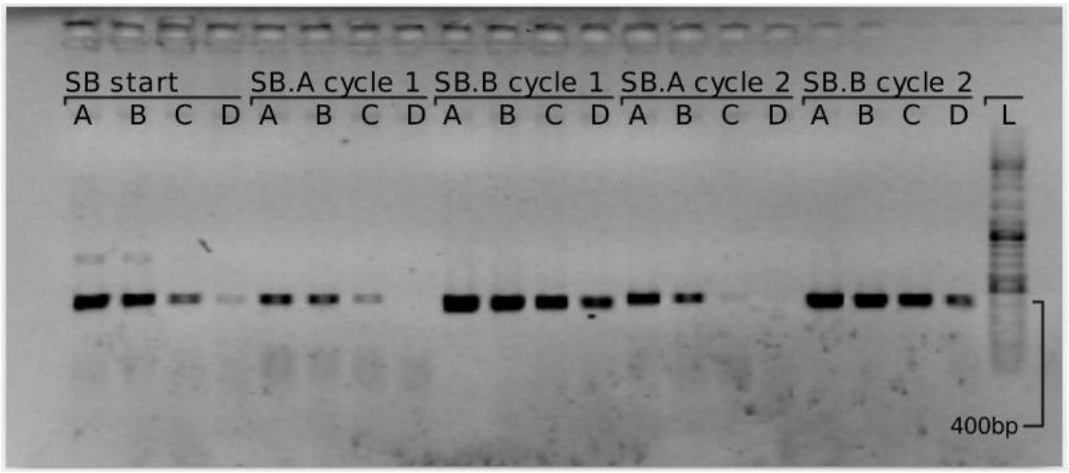


Figure 01 : Visualization by semi-quantitative PCR of DyP sequence enrichment along the two successive cycles of sequence capture from the SB cDNA soil sample. PCRs using DyP-specific primers were performed on different quantities (A = 10 ng; B = 1 ng; C = 0.1 ng and D = 0.01 ng) of cDNA before (SB start) or after one (cycle1) or two (cycle2) cycles of capture using either ascomycete (SB.A) or basidiomycete (SB.B) probes. At least in the case of basidiomycete probes we observe a clear amplification signal at the lowest cDNA quantities, after both one and 2 cycles of capture, thus demonstrating target sequence enrichment.

Both cDNA samples captured using either ascomycete-specific probes (SB.A) or basidiomycete-specific ones (SB.B) were separately sequenced using the PacBio single molecule sequencing platform (**Attachment B**) and the Illumina MiSeq paired end (2x250 bp) one (**Table 04**) which required a mechanical fragmentation of the captured cDNAs. All sequences identified following treatment of the PacBio data were also found among sequences reconstructed using MiSeq short reads; thus validating the correctness of the MiSeq assembly. We found however additional partial length DyP sequences in the MiSeq assembly, probably because this latter technology has a far higher throughput compared to the PacBio one. For the sequencing of the subsequent cDNA samples we therefore favored the Illumina technology.

Regarding the PacBio sequence output, the 260,265 subreads represented 4,521 Circular Consensus Sequences (CCSs, i.e. sequenced cDNA molecules), of which only 4,341 CCSs passed quality filters (average length of 1319 bp). Regarding sequences captured using the DyP.A probes, 249 (i.e. 34,8%) of the CCS clusters (96% identity threshold) were DyPs; in the case of the SB.B, 124 (i.e. 55.8%) of the CCS clusters where DyPs. Further clustering and manual correction of the different CCSs led to the identification of 6 different DyP sequences in the SB.A and of 6 in the SB.B datasets, respectively. This observation suggested that the explorative probes did not inadvertently captured other gene families and that the numerous singletons most likely represent contaminating cDNA sequences that were not eliminated during the washing steps of the streptavidin-coated beads after the capture.

Sequencing Platform	Illumina MiSeq	
Captured Genes / Sample	**DyPA SB**	**DyPB SB**
Paired end raw reads (n)	**11305818**	**10140462**
Mean lenght (nt)	300	300
Total size (nt)	3394890714	3045604579
Paired end reads after adapters/quality trimming (n)	**9089390**	**7646732**
Total size (nt)	1977009028	1653786455
Average length (nt)	217	216
DyPA (n)	**82647**	**154537**
DyPB (n)	**778218**	**1004011**
Contigs using IDBA_UD+CAP3 (n)	**11184**	**3132**
Average length (nt)	615	1966
Range length (nt)	200-3677	200-242399
DyPA (n)	**6**	**7**
DYPB (n)	**36**	**20**

Table 04 - Illumina MISeq sequencing data assembly data for the SB cDNA captured separately by ascomycete (DyPA-SB) and basidiomycete (DyPB-SB) DyP probes: All contigs corresponding to putative DyP sequences were blasted against all available DyP sequences. Best hits against an Ascomycete sequence were counted as DyPA (penultimate line), best hits against an Ascomycete sequence was counted as DyPB (ultimate line)

Unexpectedly, probes designed for Ascomycete and Basidiomycete genes captured exactly the same 6 full-length cDNA sequences,. Regarding the partial sequences obtained after assembly of the MiSeq data, alignment of the deduced protein sequences to known ones showed that they were truncated in either their 3' or 5' ends (**Figure 03**).

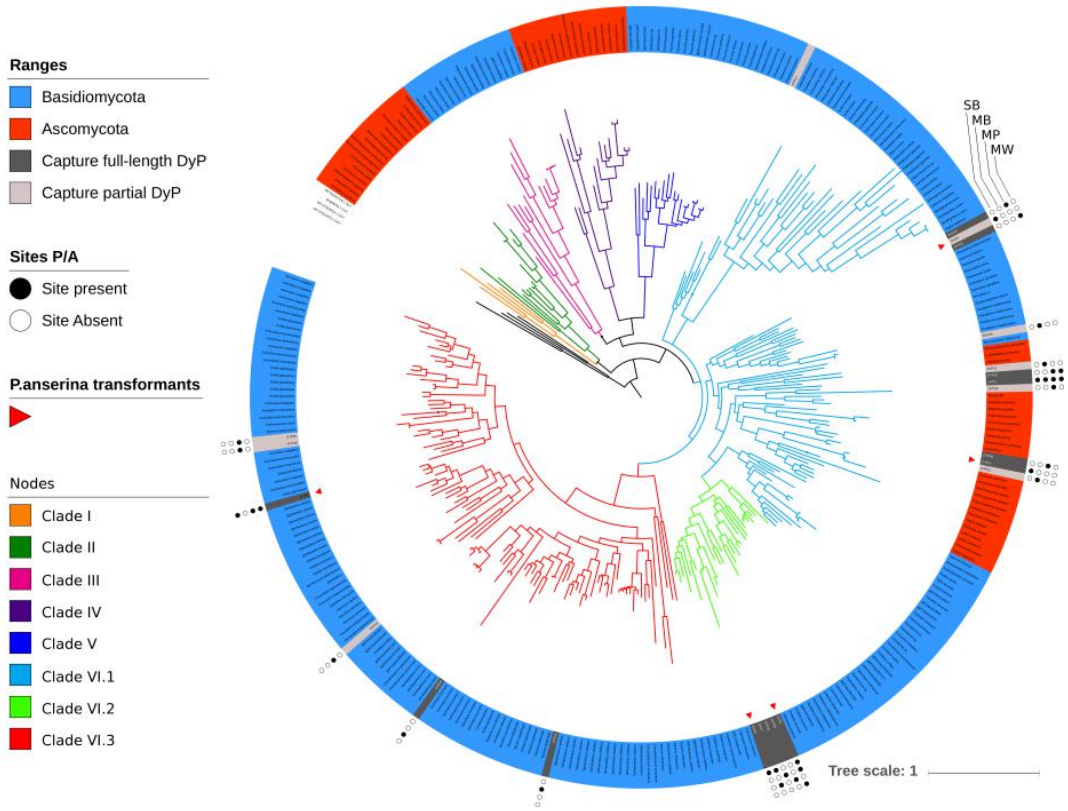


Figure 02 – Distribution of partial and full-length captured genes in the global phylogeny of fungal DyP sequences as presented in Chapter 02: different color codes in the outside ring indicate if the captured sequences are full-length or partial, if the reference sequences are from basidiomycetes or ascomycetes. Red arrowheads point to the 5 genes that were introduced by transformation in P. anserina. Occurrence of the different sequences in the different 4 studied samples is indicated on the outside of the tree.

This phylogenetic analysis also showed that the environmental sequences were distributed in 3 of the clades that formed this gene family. These results suggested that the sequence capture protocol did not target a restricted group of phylogenetically-related DyP sequences and that ascomycete and basidiomycete sets of probes could reasonably be mixed in future sequence capture experiments.

5.3.3 Sequence capture from all other environmental samples

For the other 14 cDNA samples from decomposing wood, grassland and forest soils, DyP sequence capture was performed using a single set of 69 probes encompassing probes originally designed to target Basidiomycota or Ascomycota DyP sequences. In all cases, semi-quantitative PCRs detected a significant enrichment in DyP sequences after either one and/or two cycles of sequence capture (results not shown).

	Illumina Hi-Seq sequencing Platform		
	MB	**MP**	**MW**
Paired end raw reads (n)	**44650320**	**47842380**	**56208136**
Mean lenght (nt)	250	250	250
Total size (nt)	1116258000	7514693500	1196059500
Paired end reads after adapters/quality trimming (n)	**28169782**	**30027108**	**27086472**
Total size (nt)	5758391530	600434999	5411424102
Average length (nt)	204	199	199
DyPA (n)	**161134**	**1855892**	**275440**
DyPB (n)	**5935867**	**9727636**	**5800785**
Contigs using IDBA_UD+CAP3 (n)	**3154**	**3333**	**4766**
Average length (nt)	588	533	516
Range length (nt)	200-4655	200-3578	200-3154
DyPA (n)	**6**	**5**	**5**
DYPB (n)	**95**	**73**	**87**

Table 05 - Illumina HiSeq sequencing data assembly data for the three MB, MP and MW cDNA captured ueing ascomycete and basidiomycete DyP probes mixed together: All contigs corresponding to putative DyP sequences were Blasted against all available DyP sequences. Best hits against an Ascomycete sequence were counted as DyPA (penultimate line), .best hits against an Ascomycete sequence was counted as DyPB (ultimate line)

Sized captured cDNAs were sequenced using the Illumina HiSeq (2x250bp) technology in two successive runs. In the first run 3 indexed samples were pooled. In the second run, the remaining 11 indexed samples

were sequenced. As the bioinformatics analysis of the results of the second run has not yet been completed, in the present chapter I only present the results of the first run, which concerns samples MB, MP and MW from the Mandria site (**Table 05**).

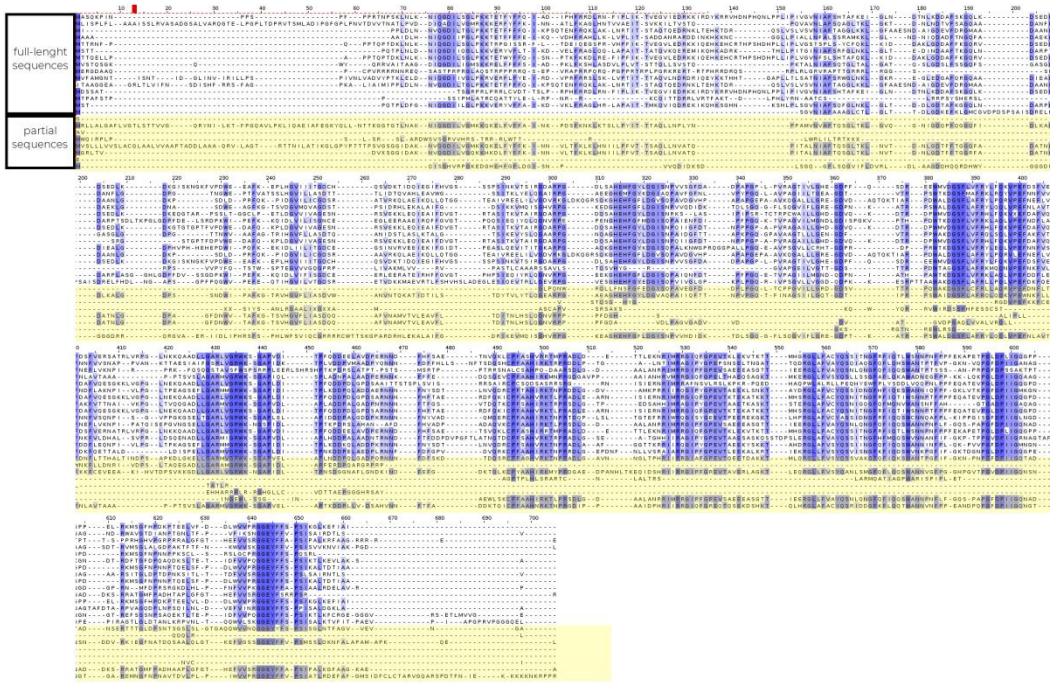


Figure 03: Alignment of captured full-length and partial DyP amino acid sequences showing that partial sequences are truncated at either their N-terminal or C-terminal ends.

Sequence assembly yielded a total of 23 DyP sequences (contigs), of which 15 were putatively full length and 8 partial (**Figure 02**). Partial sequences were truncated at either their 5' or 3' ends (**Figure 03**). Mapping reads onto contigs can be used as a proxy to estimate the relative abundance of the sequences in each of the sequences of the samples (**Figure 04**).

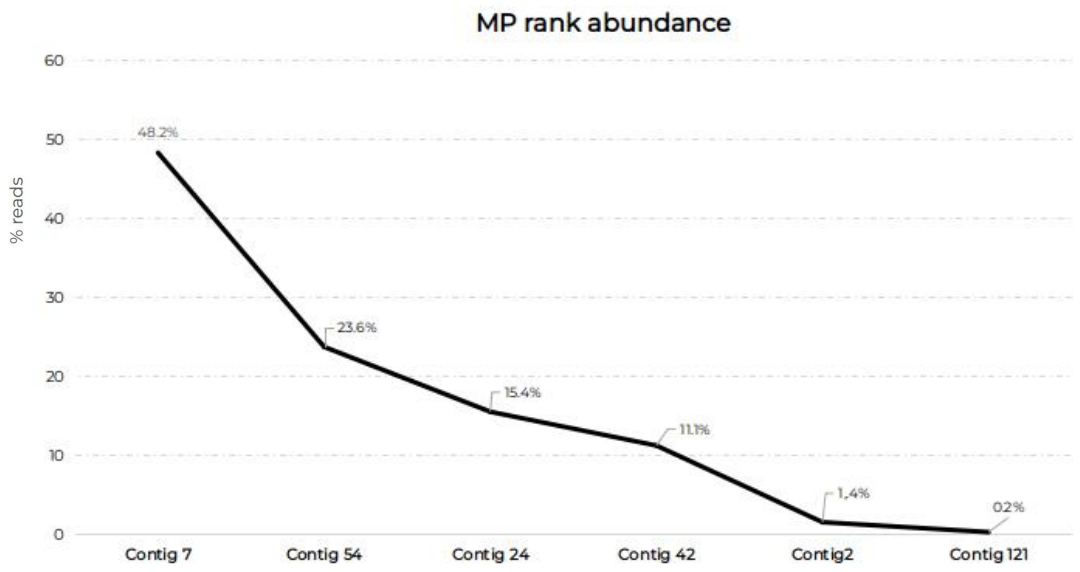


Figure 04: Abundance of each individual full-length DyP sequences in the MP sample as estimated by mapping individual DyP HiSeq reads to each contig sequence.

Although 18 of the 23 different DyP sequences (97% identity) were specific to a single sample, 5 of them were detected in 2, 3 or 4 different samples from one or more site (**Figure 02**); only one of the sequences was found in all samples.

5.3.4 Phylogenetic distribution of the captured DyPs

Full length and partial sequences obtained by sequence capture, were added to the global "DyP-type Peroxidases D" phylogeny (**Chapter 02**). At the moment sequences have been generated from 4 samples from 2 different sites.

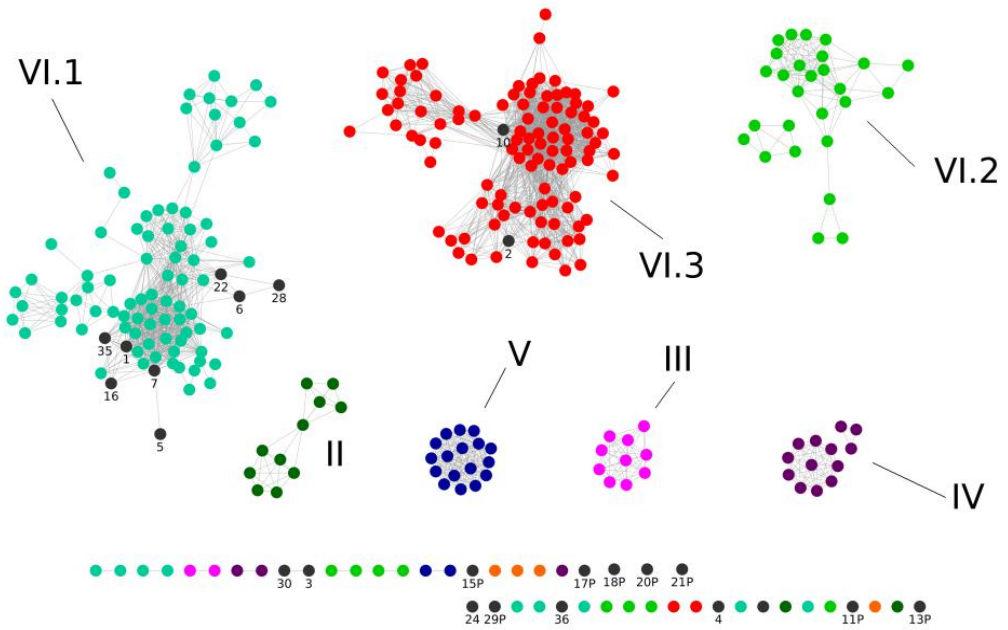


Figure 05 – Positioning of captured DyPs (black dots) in a protein Sequence Similarity Networks (SSN) of fungal DyP sequences phylogeny database and captured sequences: Fungal DyP protein Sequence Similarity Networks (SSN) were used to delineate the different clades (see chapter 02) was complemented with captured DyP sequences. The nodes represent a single DyP sequence and the edges, the lines that connect two or more nodes with a similarity greater than 150 (evaluated as pairwise alignment scores). In this analysis (as opposed to the phylogenetic analysis), most partial and several full-length captured DyPs (e.g. the ones from cluster VI.2) could not be affiliated to any clade.

Both Sequence Similarity Network (SSN, **Figure 05**) and/or phylogenic analyses (**Figure 02**) placed the different captured DyPs, in 3 of the clades defined in a previous global phylogenetic analysis of this gene family (**Chapter 02**). Sequences from clade VI-1 seemingly originated from either basidiomycete or ascomycete species, while captured sequences that fall in clades VI-2 or VI-3 were probably all from basidiomycetes (**Figure 02**). Six captured DyPs (3 full-length, 3 partial) fall in clade VI.3, which corresponds to

secreted DyPs exclusively found in WR saprotrophic species although, thus far, only one of these captured sequences originates from a decaying wood sample.

5.3.5 Heterologous expression of captured DyPs in *Podospora anserina*

To test the functionality of captured DyPs, in a pilot experiment we selected 5 of them from the SB soil sample. These 5 DyPs were distributed for two of them in clade VI.1 (one putative ascomycete sequence and one putative basidiomycete one), two of them in clade VI.2 and one of them in clade VI.3, which corresponds to secreted DyPs (**Figure 02**). As a host microorganism for production of the proteins we selected the filamentous Ascomycete *P. anserina* for four main reasons: (i) its genome sequence (Espagne et al., 2008) indicated that it has no endogenous DyP coding gene, (ii) it is a transformable fungus for which expression plasmids are available (Silar 1995), (iii) several mutants deleted of one or several "phenol oxidase" encoding genes including laccases (Xie et al., 2014) and class II or unspecific peroxygenases (P. Silar, personal communication) are available and could facilitate downstream biochemical characterization of DyP enzymes, and (iv) production of functional fungal DyPs has been reported in filamentous fungi (Sugano et al., 2000; Lauber et al., 2017), while bacterial hosts tend to produce non functional proteins in inclusion bodies (Linde et al., 2015).

To potentially favor gene transcription and protein characterization, the DyP genes were synthesized using the *P. anserina* codon usage code, cloned downstream of a strong constitutive *P. anserina* promoter and their 3'-end

fused in-frame to an HA antigen coding sequence. For each of the 5 genes, between 4 and 15 hygromycin-resistant putative *P. anserina* transgenic mycelia were selected and at least 2 of them were directly screened by western blot for the presence of the recombinant protein, using a commercial anti-HA antibody, for the presence of the protein in cell lysates.

For all 15 transformants analyzed, western blots failed to reveal a positive signal in the 50-55 kDa range of protein sizes which corresponds to the theoretical molecular masses of the different DyPs (**Figure 05**). In coomassie-stained protein gels, this zone was furthermore not characterize by the presence of an additional visible protein band that would differentiate proteins extracts from transformants from the one of the untransformed recipient strain. In several cases however, faint putative hybridization signals were observed in zones corresponding to proteins of 75-125 kDa in mass (**Figure 06**).

From this negative result, we analyzed 3 DYP1, 5 DYP2 and 3 DYP3 putative transgenic mycelia by RT-PCR using internal DYP1, 2 and 3 specific primers. Of these 12 putative transformants, a positive amplification signal of the expected size was only recorded for one DYP1 transformant. The other mycelia may therefore be false hygromycin-resistant transformants or may be true transformants that do not express the transgene. For the transformant that gave a positive amplification signal using DYP internal primers, no RT-PCR amplification signal was however observed when the reverse primer hybridized to the 3' HA tag coding sequence. This observation suggested that the transgene mRNA could be truncated and may not cover the 3'-end of the gene.

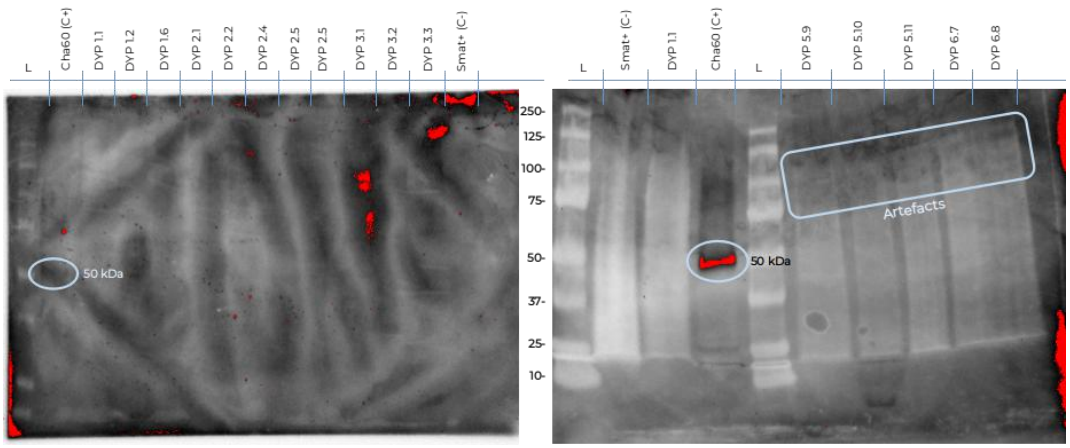


Figure 06 - P. anserina DyP transformants western-blots: P. anserina DyP transformants were tested on a western-blot using the HA-tag, in order to evaluate the eventual expression of the DyPs by the fungus. No positive responses were shown. L = ladder; Cha60 =positive control; Smat+ = negative control; DyPx.y = P. anserina transformant name. Right size to recover DyPs proteins (50kDa) is highlighted with a ring, where it is expressed by Cha60.

To go further, we performed shotgun protein mass spectrometry sequencing on different trypsin-digested protein samples from 3 transformants. For the 3 transformants we sequenced the proteins in the range of 75-250 kDa where a putative hybridization signal was observed when using an anti-HA antibody. In addition, for the DYP1.1 transformant giving a positive signal in RT-PCR, we also sequenced the proteins in the range of 50 kDa, the size of the expected recombinant protein. A recombinant DYP protein was only unambiguously identified in this latter sample. However, only four peptides covering *ca* 13% (68/514 aa) of the entire DYP1.1 sequence were detected by mass spectrometry (**Scheme 01**), thus suggesting that the corresponding polypeptide may not be particularly abundant in the corresponding extract.

DYP1 (57,70kDa)

MASQKPINPPSPFPPRTNPSKLNLKNIQGDILSGLPKKTETFYFFQIADIPHFRRDLRN
FIPLIKTVEGVIEDRKKIRDYKRRVHDNPHQNLPPLIPIVGVNIAFSHTAFKEIGLNDT
NLKDDAFSKGQLKDSEDLKDKGISKNGKFVPDWEEAFKEPLHGVIIITGDCHQSVDK
TIDQIEGIFHVGSSSPSINKVTSIRGDARPGDLSAHEHFGYLDGISNPVVSGFDADPAP
GPLPVRAGTIVLGHEGDPFQSDREEWMVDGSFLVFRYLFQKVPEFDSFVERSATRLV
RPGLNKKQAADLLGARLVGRWKSGAPVDITPFQDDEELAVDPERNNDFHFSAETN
VQKLCPFASHVRRTMPRADLDETTLEKNRIMRRGIQFGPEVTKLEKVTKTTMHGR
GLLFACYQSSITNGFRFIQTLWSNNPNFPPFEKAPETPGLDPLIGQGPPELRKMSGF
HPDKPTEELVFDDLWVVPRGGEYFFSPSIKGLKEFIAIKLIDYPYDVPDYASL*

Scheme 01: Positioning on the captured DyP1 amino acid sequence that was expressed in P. anserina of the different peptides that were identified by mass spectrometry on protein extracts from a Podospora transformant. Peptides are highlighted in blue. The C-terminal part of the protein highlighted in yellow corresponds to the HA-tag that was not detected by mass spectrometry.

5.4 Discussion

The main objective of the present study was to compare the phylogenetic diversity of fungal DyP genes expressed in environments that differ from each other by lignin input and degradation potential in order to indirectly evaluate the possible contribution of this fungal gene family to lignin degradation. The results obtained thus far are not yet sufficient to conclude on this hypothesis. Indeed, we obtained DyP sequence data for only four samples representing two of the five studied geographic sites. However, the present study demonstrates the validity of the experimental approach and the raw sequence data for the three remaining sites have already been obtained but need to be analyzed to reconstruct DyP sequences.

The present study nevertheless validates the sequence capture approach

coupled to high-throughput sequencing to study the diversity of genes expressed in the environment by species-rich fungal communities. In a previous study (Bragalini et al., 2014), captured cDNA encoding GH11 hemicellulases that were cloned and analyzed by Sanger sequencing. In the present study, we show that high-throughput sequencing can confidently reconstruct numerous full-length or nearly full-length captured environmental cDNA sequences whose relative abundance in the original sample can be evaluated by counting the number or reads that map to each of these sequences. This conclusion is supported by the fact that two different sequencing technologies (PacBio single molecule sequencing and Illumina short read sequencing) implemented on a common sample produced *in fine* the same sequences, thus excluding that these sequences represent technical artefacts. Furthermore, in phylogenetic analyses, most of these sequences branched in sequence clades defined by sequences identified in fungal proteomes. Noticeably, captured sequences that branched outside of well-defined clades were mostly partial sequences, thus suggesting that this phylogenetic placement may be erroneous, resulting from the inclusion in the analysis of a too low number of phylogenetic-informative characters.

A conclusion that needs to be confirmed by the analysis of the additional environmental samples is that each sample seems to be dominated by a low number (in the range of 5-14) of abundantly expressed DyP genes, which correspond to those that can be reconstructed entirely and to which map the highest number of short Illumina reads. Phylogenetic placement of these full-length sequences suggests that they originate for most of them from basidiomycetes taxa and for few of them from ascomycetes. This observation

apparently challenges the taxonomic surveys of the corresponding fungal communities which, at least for grassland soils and to a lesser extent for forest soils and decaying wood are dominated by ascomycete taxa (**Chapter 04**). One of the most likely explanation for this apparent incongruence is that DyP genes are far less represented in Ascomycete taxa compared to Basidiomycete ones both in term of percentage of taxa possessing DyP genes and in term of number of DyP genes per genome (**Chapter 02**). This situation could be also exacerbated by a potentially lower expression level of DyP genes in Ascomycetes compared to Basidiomycetes.

Another observation from the phylogenetic analysis if that captured sequences seem not to be randomly distributed across the global phylogenetic tree of the fungal DyP family but rather form discrete clusters that encompass sequences from one or several samples. While we cannot entirely rule out cross contaminations between the different cDNA samples during the sequence capture, we can also propose that a few related taxa abundant in several of the samples are the major producers of DyP enzymes. In support for this hypothesis, the phylogenetic analysis of known DyP genes demonstrated that specific fungal taxa or related ones can possess different but highly similar DyP genes originating from recent duplication events (**Chapter 02**). Up to know this hypothesis cannot be directly confirmed as none of the captured genes are thus far identical or highly similar to known ones from known taxa.

The phylogenetic analysis placed the captured sequences in different "locations" in the three most prominent fungal DyP clades (clades VI.1, VI.2 and VI.3 as defined in Chapter 02). From a technical point of view this

observation suggests indirectly that the capture probes did not inadvertently target a limited group of related sequences. Another observation that goes in the same direction is that the probes designed on ascomycete sequences captured unrelated basidiomycete sequences, thus suggesting that at least several of these probes are highly generalists. We cannot however exclude that an absence of sequences from other clades (e.g. clades I, II, III, IV and V) among the captured sequences does not result from a selectivity of the probes against sequences from these clusters although several of the probes in theory target known sequences from these clusters. An alternative explanation is that because these latter clusters encompass fewer sequences from fewer taxa compared to clades VI.1, VI.2 and VI.3, the probability of capturing sequences from them may inevitably be lower. In the same line, capturing sequences that belong to thus far "uncharacterized" fungal DyP clusters, if any, may require analyzing a higher number of environmental samples than we currently did.

Besides a putative association between ecosystem types and the relative abundance of specific DyP clades, it is essential to get information on the catalytic properties and substrate range of phylogenetically distinct DyPs to definitively conclude on the roles of these enzymes in lignin degradation or in the detoxification of toxic "environmental molecules". To that aim we evaluated the filamentous ascomycete *P. anserina* as a host for a routine production of functional recombinant environmental DyPs. Thus far, this preliminary experiment failed to give convincing results as none of the transformants we analyzed seem to produce, or in negligible quantities, the studied DyP polypeptides. Failure to produce recombinant proteins in a

microbial host are numerous and may be host-specific (absence of transcription of the transgene, protein misfolding leading to its degradation or aggregation in insoluble inclusion bodies…). Regarding fungal DyPs, all of the studied ones belonged to clade VI.3 that encompass secreted enzymes that were mostly purified as native proteins from culture filtrates of wild type non recombinant fungal strains. The production of recombinant extracellular fungal DyP has only been reported for 3 of them, one in the filamentous ascomycete *Trichoderma reesei*. This scarcity of information and the total absence of information for intracellular fungal DyPs make it difficult to orientate experiments designed to produce these proteins in a recombinant host and to precisely choose *a priori* the most appropriate microbial host. One possible strategy could simply be to first randomly screen different potential hosts with a limited number of transgenes.

5.6 Acknowledgements

Results presented in this chapter have largely benefited from a collaboration with colleagues (Pierre Peyret, Eric Peyretaillade, Cyrielle Gasc, Rocio Prego and Sophie Marre) from the laboratory MEDIS (INRA-Université d'Auvergne, Clermont-Ferrand) who designed the capture probes and performed the bioinformatics analyses of the high-throughput sequencing results. Laurent Vallon and Patricia Luis (Unité d'Ecologie Microbienne, CNRS-INRA-Université Lyon 1) participated to the capture. The work on *Podospora* benefited from a collaboration with the group of Philippe Silar (P. Silar, R. Ferrari) from the LIED laboratory (CNRS-Université Paris Diderot)

who provided the plasmid vector for expression in *Podospora* and performed the transformations. Finally, proteomic analyses of *Podospora* transformants were largely performed by Jeanne Doré (Unité d'Ecologie Microbienne, CNRS-INRA-Université Lyon 1) and Cindy Dieryckx and Vincent Girard (Plateforme de protéomique, CNRS-Université Lyon 1- Bayer CropScience Joint laboratory).

5.7 Bibliography

Adamo M, Voyron S, Girlanda M, Marmeisse R. 2017. RNA extraction from decaying wood for (meta)transcriptomic analyses. *Canadian journal of microbiology* **10**: 17-30,.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Álvarez C, Reyes-Sosa FM, Díez B. 2016. Enzymatic hydrolysis of biomass from wood. *Microbial Biotechnology* **9**: 149–156.

Bardgett R. 2005. *The Biology of Soil*. Oxford University Press.

Bragalini C, Ribière C, Parisot N, Vallon L, Prudent E, Peyretaillade E, Girlanda M, Peyret P, Marmeisse R, Luis P. 2014. Solution hybrid selection capture for the recovery of functional full-length eukaryotic cDNAs from complex environmental samples. *DNA Research* **21**: 685–694.

Brygoo Y, Debuchy R. 1985. Transformation by integration in Podospora anserina. *MGG Molecular & General Genetics* **200**: 128–131.

Buchfink B, Chao X, Huson DH., 2015. Fast and Sensitive Protein Alignment using DIAMOND, *Nature Methods* **12**: 59–60.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. 2011. Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* **28**: 1883.

Damon C, Lehembre F, Oger-Desfeux C, Luis P, Ranger J, Fraissinet-Tachet L, Marmeisse R. 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. *PLoS ONE* **7**.

Dashtban M, Schraft H, Qin W. 2009. Fungal bioconversion of lignocellulosic residues; Opportunities & perspectives. *International Journal of Biological Sciences* **5**: 578–595.

de Gonzalo G, Colpa DI, Habib MHM, Fraaije MW. 2016. Bacterial enzymes involved in lignin degradation. *Journal of Biotechnology* **236**: 110–119.

Denonfoux J, Parisot N, Dugat-Bony E, Biderre-Petit C, Boucher D, Morgavi DP, Paslier D Le, Peyretaillade E, Peyret P. 2013. Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Research* **20**: 185–196.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Eichlerová I, Homolka L, Nerud F, Zadrazil F, Baldrian P, Gabriel J. 2000. Screening of Pleurotus ostreatus isolates for their ligninolytic properties during cultivation on natural substrates. *Biodegradation* **11**: 279–287.

Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury J-M, Ségurens B, Poulain J, *et al.* 2008. The genome sequence of the model ascomycete fungus Podospora anserina. *Genome Biology* **9**: 7-77.

Esser K. 1974. Podospora anserina. In: Bacteria, Bacteriophages, and Fungi. Boston, MA: Springer US, 531–551.

Fawal N, Li Q, Savelli B, Brette M, Passaia G, Fabre M, Mathé C, Dunand C. 2012. PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Research* **41**: 441–444.

Fernández-Fueyo E, Linde D, Almendral D, López-Lucendo MF, Ruiz-Dueñas FJ, Martínez AT. 2015. Description of the first fungal dye-decolorizing peroxidase oxidizing manganese(II). *Applied Microbiology and Biotechnology* **99**: 8927–8942.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martinez AT, Otillar R, Spatafora JW, Yadav JS, *et al.* 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science* **336**: 1715–1719.

Gasc C, Peyretaillade E, Peyret P. 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. *Nucleic Acids Research* **44**: 4504–4518.

Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1854**: 1019–1037.

German DP, Weintraub MN, Grandy AS, Lauber CL, Rinkes ZL, Allison SD. 2011. Optimization of hydrolytic and oxidative enzyme methods for ecosystem studies. *Soil Biology and Biochemistry* **43**: 1387–1397.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, *et al.* 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**: 182–189.

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, *et al.* 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–D704.

Hibbett D, Blanchette R, Kenrick P, Mills B. 2016. Climate, decay, and the death of the coal forests. *Current Biology* **26**: R563–R567.

Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T. 2010. New and classic families of secreted fungal heme peroxidases. *Applied Microbiology and Biotechnology* **87**: 871–897.

Huang X, Madan A., 1999. CAP3: A DNA sequence assembly program. *Genome Research* **9**: 868-877.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Kellner H, Luis P, Pecyna MJ, Barbi F, Kapturska D, Kruger D, Zak DR, Marmeisse R, Vandenbol M, Hofrichter M. 2014. Widespread occurrence of expressed fungal secretory peroxidases in forest soils. *PLoS ONE* **9**.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, *et al.* 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410–415.

Langmead B, Salzberg S,. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Lauber C, Schwarz T, Nguyen QK, Lorenz P, Lochnit G, Zorn H. 2017. Identification, heterologous expression and characterization of a dye-decolorizing peroxidase of Pleurotus sapidus. *AMB Express* **7**: 164.

Li H, Yelle DJ, Li C, Yang M, Ke J, Zhang R, Liu Y, Zhu N, Liang S, Mo X, *et al.* 2017. Lignocellulose pretreatment in a fungus-cultivating termite. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 4709–4714.

Liers C, Pecyna MJ, Kellner H, Worrich A, Zorn H, Steffen KT, Hofrichter M, Ullrich R. 2013. Substrate oxidation by dye-decolorizing peroxidases (DyPs) from wood- and litter-degrading agaricomycetes compared to other fungal and plant heme-peroxidases. *Applied Microbiology and Biotechnology* **97**: 5839–5849.

Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu and Weizhong Li. 2012. CD-HIT: accelerated for clustering the next generation sequencing data. *Bioinformatics* **28** (23): 3150-3152.

Linde D, Ruiz-Dueñas FJ, Fernández-Fueyo E, Guallar V, Hammel KE, Pogni R, Martínez AT. 2015. Basidiomycete DyPs: Genomic diversity, structural-functional aspects, reaction mechanism and environmental significance. *Archives of Biochemistry and Biophysics* **574**: 66–74.

Lundell TK, Mäkelä MR, Hildén K. 2010. Lignin-modifying enzymes in filamentous basidiomycetes - Ecological, functional and phylogenetic review. *Journal of Basic Microbiology* **50**: 5–20.

Nagy LG, Riley R, Bergmann PJ, Krizsan K, Martin FM, Grigoriev I V., Cullen D, Hibbett DS. 2017. Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Molecular Biology and Evolution* **34**: 35–44.

Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson KH, *et al.* 2016. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Molecular Biology and Evolution* **33**: 959–970.

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Parisot N, Denonfoux J, Dugat-Bony E, Peyret P, Peyretaillade E. 2012. KASpOD-a web service for highly specific and explorative oligonucleotide design. *Bioinformatics* **28**: 3161–3162

Peng Y, et al. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth, *Bioinformatics* **28**: 1420-1428.

Pereira H, Graca J, Rodrigues JC. 2004. Wood Chemistry in Relation to Quality. *ChemInform* **35**: 80–81.

Schmieder R, Edwards R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863-864.

Schneider T, Keiblinger KM, Schmid E, Sterflinger-Gleixner K, Ellersdorfer G, Roschitzki B, Richter A, Eberl L, Zechmeister-Boltenstern S, Riedel K. 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *The ISME Journal* **6**: 1749–1762.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.

Silar P. 1995. Two new easy to use vectors for transformations. *Fungal Genet. Newsl.* **42**: 73.

Sugano Y, Nakano R, Sasaki K, Shoda M. 2000. Efficient heterologous expression in Aspergillus oryzae of a unique dye-decolorizing peroxidase, DyP, of Geotrichum candidum Dec 1. *Applied and environmental microbiology* **66**: 1754–8.

Thevenot M, Dignac MF, Rumpel C. 2010. Fate of lignins in soils: A review. *Soil Biology and Biochemistry* **42**: 1200–1211.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research* **44**: 232–235.

Valette N, Perrot T, Sormani R, Gelhaye E, Morel-Rouhier M. 2017. Antifungal activities of wood extractives. *Fungal Biology Reviews* **31**: 113–123.

Welker CM, Balasubramanian VK, Petti C, Rai KM, De Bolt S, Mendu V. 2015. Engineering plant biomass lignin content and composition for biofuels and bioproducts. *Energies* **8**: 7654–7676.

Xie N, Chapeland-Leclerc F, Silar P, Ruprich-Robert G. 2014. Systematic gene deletions evidences that laccases are involved in several stages of wood degradation in the filamentous fungus *Podospora anserina*. *Environmental Microbiology* **16**: 141–161.

Yoshida T, Sugano Y. 2015. A structural and functional perspective of DyP-type peroxidase family. *Archives of Biochemistry and Biophysics* **574**: 49–55.

Zámocký M, Hofbauer S, Schaffner I, Gasselhuber B, Nicolussi A, Soudi M, Pirker KF, Furtmüller PG, Obinger C. 2015. Independent evolution of four heme peroxidase superfamilies. *Archives of Biochemistry and Biophysics* **574**: 108–119.

Zorn H, Szweda R, Kumar M, Wilms J. 2009. Method for modifying non-starch carbohydrate material using peroxidase enzyme. Patent WO2009EP58871.

Chapter 06

General conclusions

My PhD work was positioned in the field of microbial ecology both in terms of technics that were developed and implemented and in terms of concepts that were followed. These concepts and tools were complemented by concepts and tools from other fields such as general ecology, molecular evolution, genomics and fungal molecular genetics. The opportunity to develop a research project at the conjunction of these different fields represents, I think, one of the main achievements of my PhD.

Four keywords can be extracted from my PhD title “Diversity of fungal DyP-type peroxidases and their potential contribution to organic matter degradation”: “diversity”, “fungi”, “DyP” and “organic matter degradation”.

Regarding the fungal DyP peroxidase gene family, I would like to highlight two observations. The first one is that, contrarily to what was "commonly thought", **a minority of fungal DyPs seems to correspond to secreted enzymes**, while a majority of them are most likely intracellular. This observation suggests that different DyPs may play different functions and that if DyPs are directly involved in lignin degradation, this should be a specific feature of the secreted ones. The second observation is that **secreted DyPs seem to be specifically associated to the white rot phenotype** that corresponds to the ability at degrading lignin enzymatically. This association was not observed for other putatively intracellular DyPs, thus adding an additional argument in favor of the **direct participation of fungal DyPs to plant organic matter (lignin?) degradation**. Additionally, our observation that all fungal DyPs characterized thus far correspond to secreted DyPs and our suggestion that intracellular DyPs should participate to distinct biological processes should encourage specific functional studies on these latter intracellular

enzymes. These studies could be motivated by purely fundamental arguments but also by applied ones. It could indeed be hypothesized that intracellular DyPs could present catalytic properties and substrate ranges (including "artificial molecules") distinct from the ones of extracellular DyPs that have already been patented for industrial applications. A preliminary attempt to produce functional recombinant DyPs from two of the intracellular clades in the filamentous Ascomycete *P. anserina* has failed or at least led to the production in too small quantities to allow a functional characterization of the enzymes. Reasons for this failure can be multiple and have not been explored further. Production of recombinant DyPs will certainly necessitates the implementation of systematic screening protocols to evaluate different expression hosts (bacteria, yeasts, filamentous fungi), different plasmid constructs (constitutive *versus* inducible expression), different gene constructs (with or without HA or his tags), and different culture conditions (e.g. media supplemented or not with heme).

In term of **fungal diversity**, the major objective of my PhD thesis was to appreciate the diversity of DyP genes expressed in different environments that differ from each other with respect to lignin output and degradation potential. In a preliminary study (Chapter 04), I evaluated the fungal communities present in the different study sites and environments by means of metabarcoding on both soil and wood extracted DNA and RNA. The combined use of environmental RNA and DNA had rarely been reported for the analysis of fungal community structure. We found that it improved resolving differences in fungal community composition between both habitats and geographic sites that differ from each other with respect to

geology and soil, climate and vegetation. We formulated the hypothesis that the combined analysis of DNA and RNA "improved" the quality of the dataset by allowing a better detection and elimination of the spurious taxa that can result from sequencing artefacts or that represent dead or dormant fungal cells.

To evaluate **DyP diversity** in the different habitats, we choose to isolate, by sequence hybridization capture, full-length genes from environmental RNA. This technical approach is certainly much more demanding both in terms of "wet lab" experiments and bioinformatics treatment of the data, compared to PCR amplification of the genes using gene-specific primers. Sequence hybridization capture if however know to present many advantages over PCR to reveal the true diversity of a gene family. (i) Sequence capture that makes use of numerous probes designed over the entire length of the genes is less biased compared to PCR that uses a unique pair of primers, (ii) through the use of "exploratory probes", sequence capture can potentially lead to the isolation of sequences highly divergent from the one already present in databases and (iii) sequence capture allows for the recovery of full-length genes, thus permitting more accurate phylogenetic placements and their expression to isolate and characterize the corresponding recombinant proteins.

The results that were obtained first demonstrate that sequence capture can be implemented on a "significant" (15) number of environmental samples and that the direct sequencing of captured cDNA, without cloning, can indeed lead to the reconstruction of novel full-length DyP sequences. These sequences fall in different clades that were identified during the global

phylogenetic survey of the fungal DyP family and belong to either the Basidiomycota or the Ascomycota.

By the end of my PhD thesis, sequences of DyP genes were not available for all environmental samples and therefore our original hypothesis regarding the differential DyP distribution across habitats could not be tested. Once available, we will then test if the relative proportion of DyPs in the different clades differs significantly across habitats and more specifically if "secreted" DyPs are more abundant in habitats with high lignin input and degradation capacity (wood and forest soils) compared to habitats with lower values both both these parameters (grassland soils). If this turned true this would represent an "ecological argument" in favor of the contribution of fungal DyPs to lignin degradation. Besides the case of DyPs, this approach could be used to infer the ecological consequences of phylogenetic diversification in other gene families.

This page was intentionally left blank

Alla fine di questi tre anni di “lavoro” (Di Vietro et al., 2013) vorrei ringraziare molte persone, ognuna con i suoi meriti e la sua storia, quindi grazie a tutti i dottorandi, borsisti, postdoc, tesisti, tecnici, ricercatori, professori, studenti, amici e vagabondi con cui ho avuto a che fare in questi tre anni tra Torino, Villeurbanne, Zittau & Clermont-Ferrand.

Ci sono persone senza le quali questa tesi non sarebbe stata scritta o sarebbe molto diversa.

Grazie a Silvia Perotto, Mariangela Girlanda, Roland Marmeisse e Patricia Luis per aver creduto in me.

Grazie Sam e Stefi, per il vostro contributo, mai scontato e mai mancato. Un contributo scientifico e umano che dovreste essere orgogliosi di apportare ogni giorno e dispensare a tutti senza chiedere nulla in cambio.

Grazie a Stefano Ghignone per aver perso un’intera giornata di lavoro ad ubuntizzarmi il PC.

Many many tanks alla “tulcasnella crew”: Jaco, Matte & Silvia Jr.

Grazie a Simo, Salvo, Vale e di nuovo Matte, con cui ho passato i momenti migliori tra ‘ste 4 mura.

Grazie a Jaco … 1000 di queste avventure!

Grazie a T’ai, fedele compagna di lavoro, sicuro che al prossimo giro saremo di nuovo insieme.

Grazie a tutta l’equipe 2 di Lyon (Patricia, Laurent, Mylene, Jeanne, Antoine, Florian, Laurence, Delphine) e a chi ho avuto il piacere di conoscere dell’equipe 7. Abbiamo passato un anno insieme; forse non ci siamo mai conosciuti veramente, ma per me resta un’esperenza, che ricorderò sempre con piacere.

Grazie ancora a Roland per l’ospitalità.

Attachment A

Complete list of all the fungal species considered in Chapter 02 redaction. Each species is complete of: main taxonomic information, trophic guild, number of DyPs found in the genome, source of origin of genomic data, reference for the genome publication (when applicable). Phylum abbreviations : Bas. = Basidiomycota; Asc. = Ascomycota. **Source:** JGI = JGI MycoCosm Genomes Portale; NCBI = GeneBank; ZITTAU = Department of Bio-and Environmental Sciences, International Institute Zittau, Technische Universitat Dresden, Zittau, Germany. Trophic Guilds legend: BR = brown-rot; EM = ectomycorrhizic; SA = saprotrophic microfungus; SS = soil saprotroph; SY = symbiotroph; WA = wood associated; WR = white-rot. SA, SS, SY and WA trophic guilds could be reported as OS (other saprotroph).

Species	Phylum	Class	Order	Family	Guild	Source	Copies	Sequences type	Published in
Agaricus bisporus	Bas	Agaricomycetes	Agaricales	Agaricaceae	SS	JGI	0	Genome	Morin et al., 2012
Amanita muscaria	Bas	Agaricomycetes	Agaricales	Amanitaceae	EM	JGI	2	Genome	Kohler et al., 2015
Amanita thiersii	Bas	Agaricomycetes	Agaricales	Amanitaceae	SS	JGI	1	Genome	Hess J et al., 2014
Armillaria gallica	Bas	Agaricomycetes	Agaricales	Physalacriaceae	WR	NCBI	4	Genome	Sipos et al., 2017
Armillaria mellea	Bas	Agaricomycetes	Agaricales	Physalacriaceae	WR	JGI	4	Genome	Collins et al., 2013
Armillaria ostoyae	Bas	Agaricomycetes	Agaricales	Physalacriaceae	WR	NCBI	4	Genome	Sipos et al., 2017
Armillaria solidipes	Bas	Agaricomycetes	Agaricales	Physalacriaceae	WR	NCBI	4	Genome	Sipos et al., 2017
Ascochyta rabiei	Asc	Dothideomycetes	Pleosporales	Helotiaceae	SA	JGI	1	Genome	Verma S et al., 2016
Ascocoryne cylichnium	Asc	Leotiomycetes	Helotiales	Helotiaceae	WA	ZITTAU	1	Genome	Unpublished
Ascocoryne sarcoides	Asc	Leotiomycetes	Helotiales	Helotiaceae	WA	JGI	1	Genome	Gianoulis et al., 2012
Aspergillus aculeatus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	de Vries et al., 2017
Aspergillus bombycis	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Moore et al., 2016
Aspergillus carbonarius	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	de Vries et al., 2017
Aspergillus clavatus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	Arnaud et al., 2012
Aspergillus cristatus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Ge Y et al., 2016

Aspergillus flavus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Arnaud et al., 2012
Aspergillus fumigatus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	Fedorova et 2008
Aspergillus glaucus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	de Vries et al., 2017
Aspergillus lentulus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	2	Genome	Kusuya et al., 2015
Aspergillus oryzae	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	2	Genome	Arnaud et al., 2012
Aspergillus parasiticus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	1	Genome	Faustinelli 2016
Aspergillus sydowii	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	de Vries et al., 2017
Aspergillus terreus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	Arnaud et al., 2012
Aspergillus thermomutatus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	1	Genome	Dufresne et al., 2017
Aspergillus turcosus	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	2	Genome	Dufresne et al., 2017
Aspergillus udagawae	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	Kusuya Y et al., 2015
Aspergillus versicolor	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	de Vries et al., 2017
Aspergillus wentii	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	de Vries et al., 2017
Auricularia auricula-judae	Bas	Agaricomycetes	Auriculariales	Auriculariaceae	WR	ZITTAU	6	Genome	Unpublished
Auricularia mesenterica	Bas	Agaricomycetes	Auriculariales	Auriculariaceae	WR	ZITTAU	6	Genome	Unpublished
Auricularia subglabra	Bas	Agaricomycetes	Auriculariales	Auriculariaceae	WR	JGI	9	Genome	Floudas et al., 2012
Baudoinia compniacensis	Asc	Dothideomycetes	Capnodiales	Teratosphaeriaceae	SA	JGI	1	Genome	Ohm RA et al., 2012
Bjerkandera adusta	Bas	Agaricomycetes	Polyporales	Phanerochaetaceae	WR	JGI	14	Genome	Binder M et al., 2013
Bondarzewia mesenterica	Bas	Agaricomycetes	Russulales	Bondarzewiaceae	WR	ZITTAU	1	Genome	Unpublished
Botryobasidium botryosum	Bas	Agaricomycetes	Cantharellales	Botryobasidiaceae	BR	JGI	3	Genome	Riley R et al., 2014
Botrytis cinerea	Asc	Leotiomycetes	Helotiales	Sclerotiniaceae	SA	NCBI	1	Genome	Staats M et al., 2012
Byssochlamys spectabilis	Asc	Eurotiomycetes	Eurotiales	Thermoascaceae	SA	JGI	1	Genome	Oka T et al., 2014
Calocera cornea	Bas	Dacrymycetes	Dacrymycetales	Dacrymycetaceae	BR	JGI	0	Genome	Nagy et al., 2016
Calocera viscosa	Bas	Dacrymycetes	Dacrymycetales	Dacrymycetaceae	BR	JGI	0	Genome	Nagy et al., 2016
Cenococcum geophilum	Asc	Dothideomycetes	Incertae sedis	Gloniaceae	SY	JGI	1	Genome	Peter M et al., 2016
Ceriporiopsis (Gelatoporia)	Bas	Agaricomycetes	Polyporales	Meruliaceae	WR	JGI	0	Genome	Fernandez-Fueyo E et
Chondrostereum purpureum	Bas	Agaricomycetes	Agaricales	Cyphellaceae	WR	ZITTAU	5	Genome	Unpublished genome
Cladophialophora yegresii	Asc	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	SA	JGI	1	Genome	Teixeira MM et al., 2017
Cladosporium fulvum	Asc	Dothideomycetes	Capnodiales	Davidiellaceae	SA	JGI	1	Genome	de Wit PJ et al., 2012
Cladosporium sphaerospermum	Asc	Dothideomycetes	Capnodiales	Davidiellaceae	SA	JGI	2	Genome	Ng KP et al., 2012
Coniophora puteana	Bas	Agaricomycetes	Boletales	Coniophoraceae	BR	JGI	0	Genome	Floudas et al., 2012

Coniosporium apollinis	Asc	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	SA	JGI	1	Genome	Teixeira MM et al., 2017
Coprinellus micaceus	Bas	Agaricomycetes	Agaricales	Psathyrellaceae	SS	ZITTAU	2	Genome	Unpublished
Coprinopsis cinerea	Bas	Agaricomycetes	Agaricales	Psathyrellaceae	SS	JGI	5	Genome	Stajich JE et al., 2010
Cronartium quercuum	Bas	Pucciniomycetes	Pucciniales	Cronartiaceae	WA	JGI	3	Genome	Pendleton AL et al., 2014
Crucibulum laeve	Bas	Agaricomycetes	Agaricales	Nidulariaceae	WR	ZITTAU	3	Genome	Unpublished
Cryptococcus neoformans	Bas	Tremellomycetes	Tremellales	Incertae sedis	SA	JGI	0	Genome	Loftus BJ et al., 2005
Cylindrobasidium torrendii	Bas	Agaricomycetes	Agaricales	Physalacriaceae	WR	JGI	0	Genome	Floudas et al., 2015
Cyphellophora europaea	Asc	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	SA	JGI	1	Genome	Teixeira MM et al., 2017
Dacryopinax primogenitus	Bas	Dacrymycetes	Dacrymycetales	Dacrymycetaceae	BR	JGI	0	Genome	Floudas et al., 2012
Daedalea quercina	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	BR	JGI	0	Genome	Nagy et al., 2016
Dentipellis fragilis	Bas	Agaricomycetes	Russulales	Hericiaceae	WR	ZITTAU	1	Genome	Unpublished genome
Diaporthe ampelina	Asc	Sordariomycetes	Diaporthales	Diaporthaceae	WA	JGI	1	Genome	Morales-Cruz 2015
Diaporthe helianthi	Asc	Sordariomycetes	Diaporthales	Diaporthaceae	SA	NCBI	1	Genome	Baroncelli 2016
Dichomitus squalens	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	NCBI	2	Genome	Floudas et al., 2012
Dothistroma septosporum	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	JGI	1	Genome	de Wit et al., 2012
Endocarpon pusillum	Asc	Eurotiomycetes	Verrucariales	Verrucariaceae	SY	JGI	4	Genome	Wang et al., 2014
Epicoccum nigrum	Asc	Dothideomycetes	Incertae sedis	Incertae sedis	SA	NCBI	1	Genome	Fokin et al., 2017
Eurotium rubrum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Kis-Papo et al., 2014
Eutypa lata	Asc	Sordariomycetes	Xylariales	Diatrypaceae	WA	JGI	1	Genome	Blanco-Ulate ., 2013
Exidia glandulosa	Bas	Agaricomycetes	Auriculariales	Exidiaceae	WR	JGI	14	Genome	Nagy et al., 2016
Fibroporia radiculosa	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	WR	JGI	1	Genome	Tang JD et al., 2012
Fibulorhizoctonia sp.	Bas	Agaricomycetes	Atheliales	Atheliaceae	SA	JGI	2	Genome	Nagy et al., 2016
Fistulina hepatica	Bas	Agaricomycetes	Agaricales	Fistulinaceae	BR	JGI	1	Genome	Floudas et al., 2015
Fomes fomentarius	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	ZITTAU	2	Genome	Unpublished
Fomitiporia mediterranea	Bas	Agaricomycetes	Hymenochaetales	Hymenochaetaceae	WR	JGI	4	Genome	Floudas et al., 2012
Fomitopsis pinicola	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	BR	JGI	0	Genome	Floudas et al., 2012
Fomitopsis rosea	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	BR	ZITTAU	0	Genome	Unpublished
Galerina marginata	Bas	Agaricomycetes	Agaricales	Strophariaceae	WR	JGI	5	Genome	Riley R et al., 2014
Ganoderma applanatum	Bas	Agaricomycetes	Polyporales	Ganodermataceae	WR	ZITTAU	0	Genome	Unpublished genome
Ganoderma sp.	Bas	Agaricomycetes	Polyporales	Ganodermataceae	WR	JGI	3	Genome	Binder M et al., 2013
Glarea lozoyensis	Asc	Leotiomycetes	Helotiales	Helotiaceae	SA	JGI	2	Genome	Chen L et al., 2013

Gloeophyllum trabeum	Bas	Agaricomycetes	Gloeophyllales	Gloeophyllaceae	BR	JGI	0	Genome	Floudas et al., 2012
Grifola frondosa	Bas	Agaricomycetes	Agaricales	Schizophyllaceae	WR	NCBI	3	Genome	Min et al., 2017
Gymnopus luxurians	Bas	Agaricomycetes	Agaricales	Omphalotaceae	WR	JGI	12	Genome	Kohler et al., 2015
Hebeloma cylindrosporum	Bas	Agaricomycetes	Agaricales	Strophariaceae	EM	JGI	2	Genome	Kohler et al., 2015
Heterobasidion annosum	Bas	Agaricomycetes	Russulales	Bondarzewiaceae	WR	JGI	1	Genome	Olson A et al., 2012
Hortaea werneckii	Asc	Dothideomycetes	Dothideales	Dothideaceae	SA	JGI	2	Genome	Lenassi M et al., 2013
Hydnomerulius pinastri	Bas	Agaricomycetes	Boletales	Paxillaceae	BR	JGI	1	Genome	Kohler et al., 2015
Hypholoma sublateritium	Bas	Agaricomycetes	Agaricales	Strophariaceae	WR	JGI	2	Genome	Kohler et al., 2015
Hypsizygus marmoreus	Bas	Agaricomycetes	Agaricales	Lyophyllaceae	WA	NCBI	2	Genome	Wu et al., 2017
Irpex lacteus	Bas	Agaricomycetes	Polyporales	Phanerochaetaceae	WR	NCBI	1	Genome	Jia and Yao 2017
Jaapia argillacea	Bas	Agaricomycetes	Jaapiales	Jaapiaceae	WR	JGI	1	Genome	Riley et al., 2014
Kockovaella imperatae	Bas	Tremellomycetes	Tremellales	Incertae sedis	SY	JGI	0	Genome	Mondo et al., 2017
Kretzschmaria deusta	Asc	Sordariomycetes	Xylariales	Xylariaceae	WA	ZITTAU	0	Genome	Unpublished
Laccaria amethystina	Bas	Agaricomycetes	Agaricales	Hydnangiaceae	EM	JGI	2	Genome	Kohler et al., 2015
Laccaria bicolor	Bas	Agaricomycetes	Agaricales	Hydnangiaceae	EM	JGI	2	Genome	Martin F et al., 2008
Laetiporus sulphureus	Bas	Agaricomycetes	Polyporales	Laetiporaceae	BR	JGI	0	Genome	Nagy et al., 2016
Lentinula edodes	Bas	Agaricomycetes	Agaricales	Marasmiaceae	WR	NCBI	2	Genome	Shim et al., 2016
Leucoagaricus gongylophorus	Bas	Agaricomycetes	Agaricales	Agaricaceae	SS	JGI	0	Genome	Aylward et al., 2013
Leucosporidiella creatinivora	Bas	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	SA	JGI	0	Genome	Mondo et al., 2017
Malassezia globosa	Bas	Exobasidiomycetes	Malasseziales	Malasseziaceae	SA	JGI	0	Genome	Xu J et al., 2007
Malassezia sympodialis	Bas	Exobasidiomycetes	Malasseziales	Malasseziaceae	SA	JGI	0	Genome	Gioti A et al., 2013
Melampsora lini	Bas	Pucciniomycetes	Pucciniales	Melampsoraceae	SA	JGI	2	Genome	Nemri A et al., 2014
Metarhizium majus	Asc	Sordariomycetes	Hypocreales	Clavicipitaceae	SA	NCBI	1	Genome	Hu et al., 2014
Microbotryum lychnidis-dioicae	Bas	Microbotryomycetes	Microbotryales	Microbotryaceae	SA	JGI	0	Genome	Perlin MH et al., 2015
Mixia osmundae	Bas	Mixiomycetes	Mixiales	Mixiaceae	SA	JGI	0	Genome	Toome M et al., 2014
Moesziomyces aphidis	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Lorenz S et al., 2014
Moniliophthora perniciosa	Bas	Agaricomycetes	Agaricales	Marasmiaceae	WR	JGI	6	Genome	Mondego JM et al., 2008
Moniliophthora roreri	Bas	Agaricomycetes	Agaricales	Marasmiaceae	SA	NCBI	12	Genome	Meinhardt 2014
Mycena chlorophos	Bas	Agaricomycetes	Agaricales	Mycenaceae	WR	NCBI	3	Genome	Tanaka et al., 2014
Mycena epipterygia	Bas	Agaricomycetes	Agaricales	Mycenaceae	WR	ZITTAU	15	Genome	Unpublished
Mycetinis scorodonius	Bas	Agaricomycetes	Agaricales	Omphalotaceae	WR	NCBI	2	Characterized	Scheibner., 2008

Mycosphaerella graminicola	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	JGI	1	Genome	Goodwin SB et al., 2011
Neolentinus lepideus	Bas	Agaricomycetes	Gloeophyllales	Gloeophyllaceae	BR	JGI	0	Genome	Nagy et al., 2016
Neosartorya fischeri	Asc	Eurotiomycetes	Eurotiales	Trichocomaceae	SA	JGI	2	Genome	Arnaud et al., 2012
Neurospora crassa	Asc	Sordariomycetes	Sordariales	Sordariaceae	SA	JGI	1	Genome	Baker et al., 2015
Neurospora tetrasperma	Asc	Sordariomycetes	Sordariales	Sordariaceae	SA	JGI	1	Genome	Ellison et al., 2011
Obba rivulosa	Bas	Agaricomycetes	Polyporales	Incertae sedis	WR	JGI	2	Genome	Miettinen et al., 2016
Omphalotus olearius	Bas	Agaricomycetes	Agaricales	Omphalotaceae	WR	JGI	1	Genome	Wawrzyn et al., 2012
Paraconiothyrium sporulosum	Asc	Dothideomycetes	Pleosporales	Montagnulaceae	SA	JGI	1	Genome	Zeiner et al., 2016
Paxillus adelphus	Bas	Agaricomycetes	Boletales	Paxillaceae	EM	JGI	0	Genome	Kohler et al., 2015
Paxillus involutus	Bas	Agaricomycetes	Boletales	Paxillaceae	EM	JGI	0	Genome	Kohler et al., 2015
Penicillium antarcticum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	Nielsen et al., 2017
Penicillium arizonense	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	5	Genome	Grijseels et al., 2014
Penicillium chrysogenum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	4	Genome	van den Berg et al., 2008
Penicillium coprophilum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	Nielsen et al., 2017
Penicillium decumbens	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	Nielsen et al., 2017
Penicillium expansum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	3	Genome	Yang et al., 2014
Penicillium flavigenum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	5	Genome	Nielsen et al., 2017
Penicillium freii	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	2	Genome	Wingfield et al., 2016
Penicillium griseofulvum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	1	Genome	Banani et al., 2016
Penicillium italicum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	1	Genome	Ballester et al., 2015
Penicillium nalgiovense	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	8	Genome	Nielsen et al., 2017
Penicillium nordicum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	5	Genome	Crovadore et 2016
Penicillium polonicum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	3	Genome	Nielsen et al., 2017
Penicillium solitum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	4	Genome	Nielsen et al., 2017
Penicillium steckii	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Nielsen et al., 2017
Penicillium subrubescens	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Peng M et al., 2017
Penicillium thymicola	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	5	Genome	Nguyen et al., 2016
Penicillium verruculosum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	NCBI	2	Genome	Hu et al., 2016
Penicillium vulpinum	Asc	Eurotiomycetes	Eurotiales	Aspergillaceae	SA	JGI	2	Genome	Nielsen et al., 2017
Peniophora sp.	Bas	Agaricomycetes	Russulales	Peniophoraceae	WR	NCBI	0	Genome	Nagy et al., 2016
Pestalotiopsis clavispora	Asc	Sordariomycetes	Amphisphaeriales	Pestalotiopsidaceae	WA	ZITTAU	0	Genome	Unpublished genome

Phanerochaete carnosa	Bas	Agaricomycetes	Polyporales	Phanerochaetaceae	WR	JGI	1	Genome	Suzuki H et al., 2012
Phanerochaete chrysosporium	Bas	Agaricomycetes	Polyporales	Phanerochaetaceae	WR	JGI	0	Genome	Ohm RA et al., 2014
Phellinus noxius	Bas	Agaricomycetes	Hymenochaetales	Hymenochaetaceae	WR	NCBI	1	Genome	Chung et al., 2017
Phialocephala scopiformis	Asc	Leotiomycetes	Helotiales	Vibrisseaceae	SY	JGI	1	Genome	Walker AK et al., 2016
Phialocephala subalpina	Asc	Leotiomycetes	Helotiales	Vibrisseaceae	SY	NCBI	1	Genome	Schlegel et al., 2017
Phialophora attae	Asc	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	SA	NCBI	1	Genome	Moreno et al., 2015
Phlebia brevispora	Bas	Agaricomycetes	Corticiales	Corticiaceae	WR	JGI	3	Genome	Binder M et al., 2013
Phlebiopsis gigantea	Bas	Agaricomycetes	Polyporales	Phanerochaetaceae	WR	JGI	5	Genome	Hori C et al., 2014
Piloderma croceum	Bas	Agaricomycetes	Atheliales	Atheliaceae	EM	JGI	4	Genome	Kohler et al., 2015
Piriformospora indica	Bas	Agaricomycetes	Sebacinales	Sebacinaceae	SY	JGI	2	Genome	Zuccaro A et al., 2011
Pisolithus microcarpus	Bas	Agaricomycetes	Boletales	Pisolithaceae	EM	JGI	0	Genome	Kohler et al., 2015
Pisolithus tinctorius	Bas	Agaricomycetes	Boletales	Pisolithaceae	EM	JGI	1	Genome	Kohler et al., 2015
Pleurotus ostreatus	Bas	Agaricomycetes	Agaricales	Pleurotaceae	WR	JGI	4	Genome	Riley R et al., 2014
Pleurotus sapidus	Bas	Agaricomycetes	Agaricales	Pleurotaceae	WR	NCBI	1	Characterized	Lauber et al., 2017
Plicaturopsis crispa	Bas	Agaricomycetes	Agaricales	Incertae sedis	WR	JGI	0	Genome	Kohler et al., 2015
Postia caesia	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	BR	ZITTAU	2	Genome	Unpublished
Postia placenta	Bas	Agaricomycetes	Polyporales	Fomitopsidaceae	BR	JGI	1	Genome	Martinez et al., 2009
Pseudocercospora musae	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	NCBI	1	Genome	Chang et al., 2016
Pseudozyma antarctica	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Morita T et al., 2013
Pseudozyma hubeiensis	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Konishi M et al., 2013
Puccinia coronata	Bas	Pucciniomycetes	Pucciniales	Pucciniaceae	SA	JGI	3	Genome	Nazareno et al., 2017
Puccinia graminis	Bas	Pucciniomycetes	Pucciniales	Pucciniaceae	SA	JGI	7	Genome	Duplessis et al., 2011
Puccinia striiformis	Bas	Pucciniomycetes	Pucciniales	Pucciniaceae	SA	NCBI	2	Genome	Cantu et al., 2011
Puccinia triticina	Bas	Pucciniomycetes	Pucciniales	Pucciniaceae	SA	JGI	3	Genome	Cuomo et al., 2017
Punctularia strigosozonata	Bas	Agaricomycetes	Corticiales	Punctulariaceae	WR	JGI	5	Genome	Floudas et al., 2012
Pycnoporus cinnabarinus	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	JGI	0	Genome	Levasseur A et al., 2014
Rachicladosporium antarcticum	Asc	Dothideomycetes	Capnodiales	Cladosporiaceae	SA	NCBI	1	Genome	Coleine et al., 2017
Rachicladosporium sp.	Asc	Dothideomycetes	Capnodiales	Cladosporiaceae	SA	NCBI	1	Genome	Coleine et al., 2018
Resinicum bicolor	Bas	Agaricomycetes	Hymenochaetales	Rickenellaceae	WA	ZITTAU	3	Genome	Unpublished
Rhizoctonia solani	Bas	Agaricomycetes	Cantharellales	Ceratobasidiaceae	SA	JGI	5	Genome	Wibberg et al., 2013
Rhizopogon vesiculosus	Bas	Agaricomycetes	Boletales	Rhizopogonaceae	EM	JGI	1	Genome	Mujic AB et al., 2017

Rhizopogon vinicolor	Bas	Agaricomycetes	Boletales	Rhizopogonaceae	EM	JGI	1	Genome	Mujic AB et al., 2017
Rhodosporidium toruloides	Bas	Incertae sedis	Sporidiobolales	Incertae sedis	SA	NCBI	0	Genome	Zhu et al., 2012
Rhodotorula graminis	Bas	Microbotryomycetes	Sporidiobolales	Sporidiobolaceae	SA	JGI	0	Genome	Firrincieli et al., 2015
Rhodotorula sp.	Bas	Microbotryomycetes	Sporidiobolales	Sporidiobolaceae	SA	NCBI	0	Genome	Goordial et al., 2016
Rhynchosporium agropyri	Asc	Leotiomycetes	Helotiales	Incertae sedis	SA	NCBI	2	Genome	Torriani et al., 2014
Rhynchosporium secalis	Asc	Leotiomycetes	Helotiales	Incertae sedis	SA	NCBI	1	Genome	Torriani et al., 2015
Rosellinia necatrix	Asc	Sordariomycetes	Xylariales	Xylariaceae	WA	NCBI	1	Genome	Kanematsu al., 2017
Sanghuangporus baumii	Bas	Agaricomycetes	Hymenochaetales	Hymenochaetaceae	WR	NCBI	1	Genome	Zhu et al., 2016
Schizophyllum commune	Bas	Agaricomycetes	Agaricales	Schizophyllaceae	WR	JGI	0	Genome	Ohm RA et al., 2010
Schizopora paradoxa	Bas	Agaricomycetes	Hymenochaetales	Schizoporaceae	WR	JGI	4	Genome	Min B et al., 2015
Scleroderma citrinum	Bas	Agaricomycetes	Boletales	Sclerodermataceae	EM	JGI	1	Genome	Kohler et al., 2015
Sclerotinia borealis	Asc	Leotiomycetes	Helotiales	Sclerotiniaceae	SA	NCBI	1	Genome	Mardanov al., 2016
Sclerotinia sclerotiorum	Asc	Leotiomycetes	Helotiales	Sclerotiniaceae	SA	JGI	1	Genome	Amselem et al., 2011
Sebacina vermifera	Bas	Agaricomycetes	Sebacinales	Sebacinaceae	SY	JGI	2	Genome	Kohler et al., 2015
Septoria musiva	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	JGI	1	Genome	Ohm RA et al., 2012
Septoria populicola	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	JGI	1	Genome	Ohm RA et al., 2012
Serpula himantioides	Bas	Agaricomycetes	Boletales	Serpulaceae	BR	ZITTAU	0	Genome	Unpublished
Serpula lacrymans	Bas	Agaricomycetes	Boletales	Serpulaceae	BR	JGI	0	Genome	Eastwood et al., 2011
Sistotrema brinkmannii	Bas	Agaricomycetes	Cantharellales	Hydnaceae	WR	ZITTAU	2	Genome	Unpublished
Sistotremastrum niveocremeum	Bas	Agaricomycetes	Trechisporales	Trechisporaceae	BR	JGI	0	Genome	Nagy et al., 2016
Sistotremastrum suecicum	Bas	Agaricomycetes	Trechisporales	Trechisporaceae	BR	JGI	0	Genome	Nagy et al., 2016
Sordaria macrospora	Asc	Sordariomycetes	Sordariales	Sordariaceae	SA	NCBI	2	Genome	Nowrousian al., 2010
Sphaerobolus stellatus	Bas	Agaricomycetes	Geastrales	Sphaerobolaceae	WR	JGI	32	Genome	Kohler et al., 2015
Sporisorium reilianum	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Schirawski al., 2010
Stereum hirsutum	Bas	Agaricomycetes	Russulales	Stereaceae	WR	JGI	2	Genome	Floudas et al., 2012
Suillus brevipes	Bas	Agaricomycetes	Boletales	Suillaceae	EM	JGI	1	Genome	Branco S et al., 2015
Suillus luteus	Bas	Agaricomycetes	Boletales	Suillaceae	EM	JGI	1	Genome	Kohler et al., 2015
Termitomyces albuminosus	Bas	Agaricomycetes	Agaricales	Lyophyllaceae	SY	NCBI	1	Characterized	Johjima et al., 2002
Termitomyces sp.	Bas	Agaricomycetes	Agaricales	Lyophyllaceae	SY	NCBI	3	Genome	Poulsen et al., 2014
Tilletiaria anomala	Bas	Exobasidiomycetes	Georgefischeriales	Tilletiariaceae	WA	JGI	0	Genome	Toome M et al., 2014
Trametes cinnabarina	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	NCBI	0	Genome	Busk et al., 2014

Trametes pubescens	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	NCBI	3	Genome	Granchi Z et al., 2017
Trametes versicolor	Bas	Agaricomycetes	Polyporales	Polyporaceae	WR	JGI	2	Genome	Floudas et al., 2012
Tremella encephala	Bas	Tremellomycetes	Tremellales	Tremellaceae	WA	JGI	0	Genome	Mondo SJ et al., 2017
Tremella mesenterica	Bas	Tremellomycetes	Tremellales	Tremellaceae	WA	JGI	0	Genome	Floudas et al., 2012
Trichoderma gamsii	Asc	Sordariomycetes	Hypocreales	Hypocreaceae	SY	NCBI	2	Genome	Baroncelli al., 2016
Trichosporon asahii	Bas	Tremellomycetes	Tremellales	Trichosporonaceae	SA	JGI	0	Genome	Yang R et al., 2003
Trichosporon oleaginosus	Bas	Tremellomycetes	Tremellales	Trichosporonaceae	SA	JGI	0	Genome	Kourist R et al., 2015
Tulasnella calospora	Bas	Agaricomycetes	Cantharellales	Tulasnellaceae	SY	JGI	2	Genome	Kohler et al., 2015
Ustilago hordei	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Laurie JD et al., 2012
Ustilago maydis	Bas	Ustilaginomycetes	Ustilaginales	Ustilaginaceae	SA	JGI	0	Genome	Kamper et al., 2006
Volvariella volvacea	Bas	Agaricomycetes	Agaricales	Pluteaceae	WR	JGI	4	Genome	Bao D et al., 2013
Wallemia ichthyophaga	Bas	Wallemiomycetes	Wallemiales	Wallemiaceae	SA	JGI	0	Genome	Zajc J et al., 2013
Wallemia sebi	Bas	Wallemiomycetes	Wallemiales	Wallemiaceae	SA	JGI	0	Genome	Padamsee 2012
Wolfiporia cocos	Bas	Agaricomycetes	Polyporales	Incertae sedis	BR	JGI	0	Genome	Floudas et al., 2012
Xylaria hypoxylon	Asc	Sordariomycetes	Xylariales	Xylariaceae	WA	ZITTAU	1	Genome	Unpublished
Xylaria longipes	Asc	Sordariomycetes	Xylariales	Xylariaceae	WA	ZITTAU	2	Genome	Unpublished
Xylaria polymorfa	Asc	Sordariomycetes	Xylariales	Xylariaceae	WA	ZITTAU	1	Genome	Unpublished
Zymoseptoria ardabiliae	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	JGI	1	Genome	Stukenbrock , 2012
Zymoseptoria tritici	Asc	Dothideomycetes	Capnodiales	Mycosphaerellaceae	SA	NCBI	2	Genome	Torriani et al., 2008

Attachment B

Exploratory probes used to capture DyP cDNAs from environmental samples: Probes were specific to either Ascomycota or Basidiomycota DyP sequences. Coverage represent the percentage of known DyP sequences (either from basidiomycetes or ascomycetes) to which each individual probe can hybridize.

Ascomycota specific probes

Sequence (from 5' to 3')	ID	Degenerated	No. of	Coverage
CAGACAACATTCAGGGCAGCATCTGGCCCCGTCTRCCYAARWMYTATGAATCCTAYCTSTTYT	DYP A1	RYRWMYYSY	512	11,1111
ACACCTYAGAGATATYCTYGACAATGGRGARRTCACCACYGGCACTCAATGCGAGRACCACCTG	DYP A2	YYYRRRYRR	512	8,1871
TCGGTCGACACGGACAACATCCAAGGGAGCATTTGGCCCCGTCTCCCTAAATACTATGAAACC	DYP A3		0	6,4327
GGGCCACAAAGGTGAYCCSAATMTGAAAAGAGGCCCTGAGTGGGCTAAGGAGGGCAGTTTCC	DYP A4	YSM	8	2,3392
AGTGGTYGACACYGACAACATCCAGGGMAGCATCTGGCCYCGTCTCCCCAAGTWYKMKGAAT	DYP A5	YYMYWYKMK	1024	9,3567
CGATATTAATAAGGACAACATYCAAGGCAACATCTGGCCGGGTKTGCCGAAGCTTYATGAATT	DYP A6	YKY	8	4,0936
TGTCCCGGCCAAGATTCWGSCCTTGCCAGGCGTYAACATTGCCTTTGCCTCGACTGGTCTAG	DYP A7	WSY	8	8,1871
ACATCCGAAACGGATTCAACTTCCTGACTACCCGCTGGGCAAGCAACCACCACTTCCCCGACC	DYP A8		0	12,2807
ATGGTGTTCCGCACATATGAACAAAGAACACCGGAGTTCGTTGCCTGGTGTGCCGCGAACACC	DYP A9		0	2,3392
GTCAAGACACMATCGAYCAAGGCRYMATCCTCTGCGGCCGMCCCGGCGACACYCAAGCCTCC	DYP A10	MYRYMMYY	256	4,0936
AAGGCGACATCCTCCTCAACGGTCTCCCCAAAGAAGTCGAAACGTTCTGGTTCTTCGACATCG	DYP A11		0	2,9240
ATCACTGATCTCAGCAATGTCCAAGGSGACATTCTCCTCAARGGCTTGACGAAAGAGGTCGAG	DYP A12	SRY	8	2,3392
CAAGACCCTTCGCCTCAGCCGGATTGGGCGACGGAAGGTAGCTTCCTGGTATTCCGCAAGCT	DYP A13		0	2,3392
ATGAATAACACGGATGCCTTCAACTAYMRCCCTRYTGACCAGAYCAAGTGYCCTTAYGCMTCT	DYP A14	YMRRYYYYMY	1024	5,2632
TGCCSTCTGAAGTCRACCCCGCCAACATCCAGGGCAGTATCTGGCCCCGTCTCCCMCGGTACT	DYP A15	SRMY	16	4,6784
TCAAGGCRMTGAAGGRRTATCTGACTGGCCCCCCTGATTAYCCGCCAGAGCTTGCCGGCCCT	DYP A16	RMRRY	32	3,5088
AATCTAATAACCATGAGCAATTCATCAAGAGCGGTGAACGTATCCATCCAGACTGAAAGTCCC	DYP A17		0	2,3392
AGRCTWGAYCACCACCCMCCTCGRTCWATCGGTCTSCCTGATGGCAAATATCCMACCGAGGC	DYP A18	RWYMRWSM	512	2,3392

Basidiomycota specific probes

Sequence (from 5' to 3')	ID	Degenerated	No. of	Coverage
ACCGAGCTCTACTACTTCTTCCAAATCACCAACGCGTCAGATTTCCGCAAACATCTCATCAAAT	DYP B13		0	0,9940
ATGCTTGCAGCGGATCCTAAGCGCAACAACGATTTCAAGATTGAAGGGGAGATCAACTCACAA	DYP B15		0	0,7952
ACGCATGGCTCAAAGATGGCTCGTTTATGGTTTTCCGACAGCTCCAACAACTCGTACCCGAGT	DYP B21		0	2,9821
ACCTAACCGGGTCGATTTAAAAAACGTCCAAGGCGATATTATCATTGGTTTGCAAAAACGTTA	DYP B22		0	1,1928
AGTCTCGACTAGCCTTCCGTCGAACATTGAAAGACAAGCTTCTGCCTCTGATCACCACCACTC	DYP B23		0	1,5905
ATCCCACCCTTCGATCCTGCCAACGTCCAGGGCGACATCCTTGCCGGCCTCCCGAAGAAGGTC	DYP B32		0	1,5905
TTCCATATTAGGTGACTTTATACTTGTTCCCATTCCTGGAAGCAAATCGACAGGCCAGATATA	DYP B40		0	0,1988
ACCGAAAGGAGGAGAATATTTCTTCGTGCCCTCAATATCAGGCCTTCGGGAGGTAATTGTCGG	DYP B44		0	0,3976
GACGACGTCCATCCGATGAGATCGAGCCGGAGTGCATCAGGGAGGCACCGCTGCTACCTGAT	DYP B46		0	0,3976
TTTCGGTGCTCGATTGATTGGACGGTGGAAGAGTGGTGCTCCTATCGATTTGGCACCTCTTC	DYP B47		0	1,1928
TGATGTCGAAGAAGAAGTAAGTTTGGGTCTTCTTTGGAAGGCCGCTCAAGATGTCGCCCTGA	DYP B51		0	1,5905
GCACGCCGCCCACGAAAGAGGGCGTGTACCGCATCAAGCTGTATGGGCAGAGTCCGGAGCAR	DYP B17	R	2	0,7952

AATACAAGGGAAAGGAACACTTCGGATATCAGGACGGTGTATCYCAGCCATCGTTGAGAGGCT	DYP B18	Y	2	0,7952
CAGCGCGATTAGCAGGCGCACTTACAATCAGCTTGTTACMTGTGCTTACRAGGTCGAAGGAG	DYP B14	MR	4	0,9940
TACAAGGCGACATTTTGGTTGGMATGAAGAAACAGAAGGAACGCTTYGTCTTTTTCCAMGTC	DYP B4	MYM	8	2,5845
GCTCGAACAGAAYGTGCTTTTTCTCGAAGAYTATGTCAACAAGAACTGGCTGTCTATTCCAGC	DYP B12	YYR	8	1,1928
ATGCGCTGGTGGACTACCTGCGTCGCTCTCACTTCGTTARTACCTTACRCCCTCYTGCCCAGT	DYP B24	RRY	8	1,1928
AGGAACAACAAKTTCGACTTYGGTGATTCCTTAGTTCGAGGCGATCAGACGAAGTGCCCTTTY	DYP B26	KYY	8	0,7952
AGGAGCTYTGGGGAGCTAGGWTGATYGGTCGTTGGAAATCYGGGGCTCCTATAGCATTGTGT	DYP B1	YWYY	16	1,1928
GARAGYGTCGGACTCAATCTCGATGATATCCARGGTGACATCCTCGTRGGCATGAAGAAGGAC	DYP B45	RYRR	16	1,9881
GAAAAGRAWGCCATCAGTTATGACCAGGTGGAAGCTCTTGTGACGAAGTGCTTGMGRCGGAR	DYP B41	RWMRR	32	0,3976
TCCAAGGACGGYTCWTTCCTCGCGTTCCGCCAGCTTCAGCARCTCGTTCCMGARTTCCACAA	DYP B37	YWRMRS	64	1,1928
CCAGGCYTTCAGGCAGAGGCTCAAGCTGCTCATYCCTCTGATCACCACCACCMCYCAGGTCM	DYP B48	YYMYMW	64	0,9940
TGGYCTTTGYGCTRCARGGCGTGTATKCTGTMCCAAATCTTYCATCCAATGACCTGAAACCTA	DYP B2	YYRRKMY	128	1,1928
ACTKGAYGGCAGCTTCCTYGCCYTCCGGTACCTCTTCCAGCTMGTTCCCGAGTTCRACACSTT	DYP B33	KYYYMRS	128	1,7893
GGSCGMTGGAARTCGGGAGCGCCCATYGATCTGACACCWACCGCGGAYGACCCCGCKCTCGG	DYP B36	SMRYWYKY	256	1,3917
GGGCCATCGACGGGASYTTCATGGCSTTCCGRCACYTGCARCAGAARGTGCCCGAGTTYMACG	DYP B5	SYSRYRRYM	512	2,7833
TTTGGAYCTTRAMAACATYCARGGYGAYATCTTGAGCGGCCTTCCAAAGAAGACCCAAACCTA	DYP B39	YRMYRYYY	256	1,7893
GCCGGCCTCCCGAAGAAGGTCCAGCACTACCTGYTCTTCCAGATCGAYGACRAYGTCMMSGC	DYP B31	YYRYMMSHM	768	2,5845
CGGATCCYAAWCAAGCWTGGCTCAAAGAYGGRTCRTTTATGGYTTTCAGAGAGCTCCAACAA	DYP B19	YWWYRRYYY	512	4,5726
GCGGCRCGCATGTTCGGYCGYTGGCCAAGYGGCGCACCAAYCGATCTCACSCCCTTYGCSGAC	DYP B27	RYYYYSYSY	512	1,3917
CTTGGAYGGSATYTCCCAACCSGCRGTCAAGGAWTTCGACACGAAGCCSAACCCMGGMCAGG	DYP B29	YSYSRWSMM	512	2,9821
GAAGCCTYSAGTCWCGCGCCGAKCTKYTGGGWGCGCGCATGGTCGGSCGCTGGAARTCGGGA	DYP B34	YSWKKYWSR	512	1,9881
CARAYGATATTCAAGGYGACATTCTCGTSGGWATGCAYAAGCARAAGCAGCTRTTCTACTTCT	DYP B38	RYYSWYRRY	512	2,1869
ACGGGTARYTGGGTYCAGGGCTTYGCAGGAACAAGTRTTCAYGGCGTATTTYTGATTGCGTCY	DYP B42	RYYYRYYYS	512	1,3917
GTACAACGTCTCTRTTRATYAGTCCTCCTGGAATTCCCCCKCTCCCTCAGGCTCRGCCRYCTG	DYP B49	RRYKRRYSR	512	1,1928
GACGAAGATGGGTTCCCTRRWGGATGTCCACTTCAGAAYGACCCTRTRCGTCTTGTYAAYAGG	DYP B50	RRWYRRYYW	512	0,3976
CTCGWRRTAATCTTCCTGGAAGCAACGATCATKTATTCAGACGRGCTAGTATTMCYTATGGC	DYP B3	WRRKRMYRRS	1024	1,1928
CAYCYTSCCKCTGGATAACATCCARGGCGAYATCYTGRTYGGAATGAAGAAACAGAARGAACGC	DYP B6	YYSKRYYRYR	1024	1,9881
TACCAGTCGAAGAYRTACACACTKCCRSAGTTYGTRATCCCRAAGGGCGGMGARTACTTCTTC	DYP B7	YRKRSYRRMR	1024	1,1928
AACAYCYTSCCKCTGGATAACATCCARGGCGAYATCYTGRTYGGAATGAAGAAACAGAARGAAC	DYP B8	YYSKRYYRYR	1024	1,5905
TCGGCCTCAGCAACRTYCAAGGCGAYGTCATWCCYGGTCTSCCYAAGGCYCTCGAATWYTTCT	DYP B9	RYYWYSYYWY	1024	1,3917
GTGCGAGGCAGTCGCGCACAACGCCTCTCCTCACSARCTTYCCKGGRCARGCCCCGCTRCCTT	DYP B10	SRYKRRRYMR	1024	0,9940
CCTCTSAARSYRTCGGACTCAATCTCGATGATATYCAGGGYGACATYCTYGTKGGCATGAAGAA	DYP B11	SRSYRYYYYK	1024	3,3797
TTYARGCGCGACCTYGCGAACTACACMCCSACRACRTCCGAWGAYGTTCTTGACAACCTKCGC	DYP B16	YRYMSRRWYK	1024	1,5905
AAAAGGTCAAAAASTTGAYGCCRTSRAMAACTTGGGAGAYCCGGTRGATCAAGCARCGAAAAAR	DYP B20	SYRSRMYRRR	1024	3,1809
TCGTKCCCGAGTACAASAAGTKSCTSCTCGACAACGCCGTYCAGAGCCCSGCCGGMAAYCTCAC	DYP B25	KSKSSYSMYS	1024	1,1928
TAGCGGCAACAACCATGGCRTATTCATCATCTGCTCTRRATYCCARSAGAGYTGCCTAKMKGA	DYP B28	RRRYRSYKMK	1024	1,7893
CGCMGCSTACCAGAGCAACATCGTYAACGGCTTCCARTTCATCCAGCACAGYTGGGCSAACAM	DYP B30	MSYRYSMYRR	1024	2,3857
TCARKCRCGCGCSGATCTSCTGGGAGCGCGCATGGTYGGSCGMTGGAARTCGGGAGCGCCCA	DYP B35	RKRSSYSMRY	1024	3,7773
GCAAGACWCGYCCSCGTGCCGATCTCGTYGCRCCKGCSAACAGCATCATYCGCAGYGGMATCC	DYP B43	WYSYRKSYYM	1024	1,1928

Degeneration code:

R = G or A; K= G or T; S = G or C; W = A or T; M = A or C; Y = T or C; D = G or A or T; V = G or A or C; B = G or T or C; H = A or T or C; N = G or A