Explorative metabarcoding of Abies balsamea L. Mill. endomycobiota

Thèse

Julien Ponchart

Doctorat en sciences forestières
Philosophiæ doctor (Ph. D.)

Québec, Canada

© Julien Ponchart, 2019

Explorative metabarcoding of *Abies balsamea* L. Mill. endomycobiota

Thèse

Julien Ponchart

Sous la direction de :

Louis Bernier, directeur de recherche
Jean Bérubé, codirecteur de recherche

Résumé

L'étude des champignons endophytes, ainsi que celle des autres composants du microbiote des plantes, a fortement bénéficié du développement des techniques de séquençage à haut débit à la fin des années 2000. Ces progrès technologiques ont notamment permis la popularisation du métabarcoding, une approche servant à identifier les composants de la biodiversité et d'étudier leur distribution au sein d'échantillons environnementaux grâce à leur contenu en ADN. L'abondance des données produites, ainsi que la standardisation de la préparation des échantillons permises par ces techniques de séquençage ont modifié en profondeur la perception de l'ampleur de la diversité fongique. Cependant, les préceptes de l'endophytologie fongique restent majoritairement dictés par les études basées sur les mises en culture bien qu'elles ne parviennent, comme l'ont démontré les études moléculaires, qu'à récolter une partie de la diversité fongique. Les techniques de séquençage à haut débit ne sont pas sans biais elles aussi puisqu'elles tendent à majorer les estimations de la biodiversité même avec les analyses les plus poussées. Les objectifs principaux de cette thèse étaient tout d'abord de développer une approche analytique rigoureuse afin d'estimer de façon conservatrice la biodiversité associée aux données issues du pyroséquençage 454; puis, de développer une meilleure compréhension de la structure de l'endomycobiote des arbres en milieu forestier tout en remettant en question les conclusions des études basées sur les mises en culture.

La surestimation de la biodiversité est essentiellement liée à la conservation de séquences erronées qui participent à la formation du nombre important de singletons et doubletons généralement observés avec les techniques de séquençage à haut débit. Trois sources d'erreurs prédominent: la formation de chimères, la substitution de nucléotides lors de l'amplification et les erreurs de séquençage. Nous avons posé l'hypothèse que la sélection d'un sous fragment du code-barres moléculaire fongique, basée sur des propriétés particulières, pourrait si ce n'est identifier formellement ces séquences comme erronées, du moins limiter leur effet sur l'estimation de la biodiversité. Le fragment que nous avons considéré se compose du résidu de la petite sous-unité ribosomique (pSSU) situé à la suite de l'amorce ITS1F, et de l'espaceur transcrit interne 1 (ITS1). Nous avons montré qu'utiliser ce fragment pour analyser les données permet d'améliorer la sensibilité de la détection des chimères. La substitution de nucléotides ainsi que les erreurs de séquençage sont des phénomènes rares, et les séquences erronées sont donc faiblement représentées et relativement similaires à des séquences réelles et abondantes. Nous avons donc posé l'hypothèse qu'inclure le pSSU, dont la variabilité est plus faible que celle de l'ITS1, puisse étouffer l'impact de ces erreurs. Les séquences potentiellement erronées ont été regroupées avec les séquences réelles et abondantes dont elles déviaient, permettant ainsi de réduire la formation des singletons et des doubletons. Suite à cela, nous avons donc développé une méthode afin d'extraire directement le fragment pSSU-ITS1 des amplicons du code-barres fongique.

À partir de l'endomycobiote d'un unique sapin baumier que nous avons analysé afin d'évaluer notre traitement de données dans notre premier chapitre, nous avons observé qu'utiliser le fragment pSSU-ITS1 en lieu et place de seulement l'ITS1 n'affecte pas les conclusions sur la structure de la communauté des champignons endophytes. Bien qu'il faille le considérer dans le cadre d'un échantillonnage limité, nous avons évalué, semble-t-il pour la première fois, l'ampleur de la diversité des champignons endophytes recueillis dans un arbre à un moment donné et extrapolé cette richesse à 2 536 ± 73 mOTUs. Nous avons confirmé dans notre second chapitre que les champignons endophytes présentent une certaine spécificité de tissu puisque

l’endomycobiote des branches de sapins baumiers se divise selon le type de tissu considéré plutôt que de former une entité ubiquitaire uniformément répartie dans l’ensemble des branches. Enfin, dans notre dernier chapitre, nous avons montré que les mécanismes impliqués dans la colonisation de la plante hôte par les champignons endophytes se révèlent d’une complexité et d’une dynamique plus importantes que le processus d’accumulation passive suggéré par les études basées sur les mises en culture: les quatre dernières cohortes d’aiguilles de sapins baumiers que nous avons étudiées présentaient une diversité relativement conservée, mais des communautés différentes.

Abstract

As for the studies of other members of the plant microbiota, fungal endophytology has vastly benefited from the development of High Throughput Sequencing techniques in the late 2000s. This technological progress has notably allowed for the popularization of metabarcoding, i.e. a DNA-based approach to identify biodiversity components from environmental samples and study the community composition and distribution. The massive production of data, and the standardization in the sample preparations associated with such methods, have deeply modified the perception of the extent of the fungal biodiversity. Yet fungal endophytology precepts remain largely inherited from culture-dependent methods which have been shown to yield a more fractioned portion of the biodiversity than the molecular-based approach, as many fungi are not amenable to standard culturing. HTS techniques are not without drawbacks either as they tend to inflate the biodiversity estimates even with state of the art analysis. The main goals of this thesis were first to develop a more rigorous approach to analyse data obtained from 454 pyrosequencing, one of the original HTS techniques, in order to estimate conservatively the biodiversity; and then to develop a better understanding of the structure of forest trees endomycobiota and challenge earlier conclusions based on culture-dependent methods.

Inflation of the biodiversity is mostly due to remaining undetected erroneous sequences partially forming the large number of singletons and doubletons generally observed with HTS based studies. Three sources of error are significant: PCR chimeras, PCR single base substitutions, and sequencing error. Here we hypothesized that the selection of a sub-region of the fungal barcode displaying particular characteristics might, if not formally assess erroneous sequences as such, at least limit their impact on the estimation of the diversity. We thus considered a fragment composed of the partial ribosomal small sub-unit immediately following the ITS1F primer in addition of the ITS1 sub-locus (pSSU-ITS1). We showed that basing the analysis on the pSSU-ITS1 fragment enhances the sensitivity of chimera detection. As PCR single base substitutions and sequencing errors remain rare events, spurious sequences are rare too and somewhat similar to true abundant sequences. We hypothesized that the presence of the pSSU, whose variability is lower than that of the ITS1 sub-locus, might buffer these errors. Putative rare spurious sequences were grouped with the true abundant sequences they deviated from, thus reducing the proportion of singletons and doubletons. We then developed an approach to readily extract this pSSU-ITS1 fragment from fungal ITS amplicons.

We observed from the endomycobiota of a single balsam fir that we produced to test our data treatment in the first chapter that considering the pSSU-ITS1 fragment did not alter the conclusions on the structure of the fungal endophytic community from ITS1 analysis. While it has to be considered with appropriate reservations due to the limited sampling, we also estimated, for the first time to the best of our knowledge, the extent of the fungal endophyte biodiversity harboured by a single tree at a precise time with an extrapolation of 2 536 ± 73 mOTUs. In the second chapter on the endomycobiota present in the different tissue types of balsam fir branches, we confirm that some tissue specificity is exhibited by fungal endophytes as our results suggest that the aerial endomycobiota of balsam fir trees might be fractioned in distinct communities depending on the tissue types. Finally, in the third chapter, we reveal that the mechanisms of colonization of the host plant by fungal endophytes might be more complex and dynamic that the suggested passive accumulation hinted by culture-dependent methods. The last four cohorts of needles from balsam fir sampled displayed relatively similar diversities, but harboured distinct communities.

Table of contents

Résumé ii
Abstract iv
Table of contents v
List of figures viii
Figures viii
Supplementary Figure ix
List of tables x
Tables x
Supplementary Table xi
List of abbreviations, acronyms xii
Acknowledgements xvi
Preface xviii
Introduction 1
1. Endophytism: definition and distribution 1
2. Fungal endophyte classes 2
3. Colonization through space and time 4
4. Assessing fungal diversity 6
5. Balsam fir in the context of global change 8
6. Scope of the thesis. 9
Chapter 1 SSU to the rescue: conserving the original partial SSU fragment ahead of the ITS1 sub-locus enhances the detection of putative chimeras and limits the formation of possibly spurious taxa in fungal metabarcoding 10
1.1 Résumé 11
1.2 Abstract 12
1.3 Introduction 13
1.4 Materials and Methods 15
1.4.1 Sampling and sample preparation 15
1.4.2 DNA extraction and library construction 16
1.4.3 Bioinformatic analyses 17
1.4.4 Ecological analyses 19
1.5 Results 20
1.5.1 SSU effect 20

1.5.2 Community analysis 22
1.6 Discussion 24
1.7 Conclusion 28
1.8 Acknowledgements 29
1.9 References 29
Chapter 2 Diversity, community composition and structure of the endomycobiota of balsam fir trees (*Abies balsamea* [L.] Mill) as a function of tissue type 45
2.1 Résumé 46
2.2 Abstract 47
2.3 Introduction 48
2.4 Materials and Methods 50
2.4.1 Sample preparation, DNA extraction, PCR amplification, and sequencing 50
2.4.2 Data processing and statistical analyses 52
2.5 Results 54
2.6 Discussion 56
2.7 Conclusion 60
2.8 Acknowledgements 60
2.9 References 61
Chapter 3 Endomycobiotas of the last four cohorts of *Abies balsamea* needles display relatively similar diversities but different compositions 80
3.1 Résumé 81
3.2 Abstract 82
3.3 Introduction 83
3.4 Materials and Methods 85
3.4.1 Sample preparation, DNA extraction, PCR amplification, and sequencing 85
3.4.2 Data processing and statistical analyses 86
3.5 Results 89
3.6 Discussion 91
3.7 Conclusion 94
3.8 Acknowledgements 95
3.9 References 95
Conclusion 112
1. Stringent data analysis. 113
2. Foliar vs Aerial 114

3. Time of exposure 115
4. Limits 115
5. Perspectives 116
Bibliography 119

List of figures

Figures

Figure 1.1: Organization of ribosomal genes and target regions of 5' extremity oligonucleotide for 5.8S. Colours in the close-up of the beginning of the 5.8S indicate the frequency of the most common nucleotide among the aligned fungal sequences......34

Figure 1.2: Species accumulation curves (solid lines) and bootstrap estimates (long-dashed lines) of species richness for ITS1 sub-locus based standard analysis (red) and pSSU-ITS1 fragment based analysis with the ITS3a-ITSx combination (blue). Dotted lines represent the 95% confidence intervals......35

Figure 1.3: Relative abundance plots of fungal endophyte communities associated with the whole tree (A) by plant system based on the ITS1 sub-locus standard analysis (B) and the pSSU-ITS1 fragment (C)......36

Figure 1.4: Results of diversity (A, B) and evenness (C, D) analyses between the root and the aerial systems based on the ITS1 sub-locus standard analysis (A, C) and the pSSU-ITS1 fragment (B, D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis test or after ANOVA test with $p \leq 0.05$......37

Figure 1.5: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition between the root system (brown circles) and the aerial system (needle samples: open green triangles, remaining aerial samples: closed green triangles) for the standard analysis based on the ITS1 sub-locus (A) and the pSSU-ITS1 fragment (B). One-way analysis of similarity (ANOSIM) indicates significant differences ($P < 0.05$) in community composition between the root and aerial systems. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM......38

Figure 1.6: mOTUs co-occurrence networks displaying a root module and an aerial module for the standard analysis based on the ITS1 sub-locus (A) and the pSSU-ITS1 fragment (B). Brown nodes represent root specialists, light blue aerial specialists, and light green needles specialists. Dark blue nodes represent aerial-needle affinity, and dark green root-needle affinity. Red nodes represent generalist taxa, and grey non habitat preference assigned. Grey edges were used when at least one of the nodes was assigned only to the kingdom level, blue edge only to Ascomycota, plum only to Basidiomycota, orange when both nodes were assigned to the same class, and red when both nodes were assigned to different classes. Pairwise Spearman's rank correlations were considered if $\rho > 0.6$, and P-value < 0.01......39

Figure 2.1: Results of diversity (A, B) and evenness (C, D) analyses results among the stands (A, C) and tissue types (B, D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis and DSCF tests or after ANOVA and Tukey tests with $p \leq 0.05$......68

Figure 2.2: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition among the stands (open circles for Laverdière 01, closed circles for Laverdière 02, open triangles for La Chute 01, closed triangles for La Chute 02) and the tissue types (green for needles, purple for buds, brown for bark, yellow for wood). One-way analysis of similarity (ANOSIM) indicates significant differences (P < 0.05) in community composition both among the stands and the tissue types. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM..69

Figure 2.3: Relative abundance plots of fungal endophyte communities by stand (A) and tissue type (B)...70

Figure 2.4: mOTUs co-occurrence network displaying two main modules. Nodes colour represents taxonomy assignment (as defined for Figure 2.3). Pairwise Spearman's rank correlations were considered if ρ > 0.6, and P-value < 0.01...71

Figure 3.1: Species accumulation curves (solid lines) and bootstrap estimate (long-dashed lines) of species richness for the full dataset. Dotted lines represent the 95% confidence intervals..101

Figure 3.2: Results of diversity (A, B) and evenness (C, D) analyses results among stands (A, C) and needle cohorts (B, D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis and DSCF tests or after ANOVA and Tukey tests with p ≤ 0.05..102

Figure 3.3: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition among the stands (A) (circles: Laverdiere 01, squares: Laverdiere 02, triangles: La Chute 01, diamonds: La Chute 02) and needle cohorts (B) (circles: current year, squares: 1-year-old, triangles: 2-year-old, diamonds: 3-year-old). One-way analysis of similarity (ANOSIM) indicates significant differences (P < 0.05) in community composition both between the stands and the tissue types. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM...103

Figure 3.4: Relative abundance plots of fungal endophyte communities by stand (A) and needle cohort (B)...104

Figure 3.5: mOTUs co-occurrence network displaying two main modules. Pairwise Spearman's rank correlations were considered if ρ > 0.6, and P-value < 0.01...105

Supplementary Figure

Figure S2.1: Species accumulation curves (solid lines) and bootstrap estimate (long-dashed lines) of species richness for the full dataset (A) and for each tissue type (B) (Needles: green, Buds: purple, Bark: brown, Wood: yellow). Dotted lines represent the 95% confidence intervals..79

List of tables

Tables

Table 1.1: Influence of ITS extraction methods, fragment used for chimera detection, and for clustering..40

Table 1.2: Kruskal-Wallis tests of the average relative abundance of fungal phylum and classes between the root and aerial systems for the standard analysis based on the ITS1 sub-locus. Significant differences (p ≤ 0.05) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA test instead of Kruskal-Wallis test..41

Table 1.3: Kruskal-Wallis tests of the average relative abundance of fungal phylum and classes between the root and aerial systems for the pSSU-ITS1 fragment based analysis. Significant differences (p ≤ 0.05) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA test instead of Kruskal-Wallis test..42

Table 2.1: Characteristics of Forêt Montmorency sites sampled in August 2012...72

Table 2.2: Analyses of diversity (Fisher's alpha), evenness (Pielou's evenness), and similarity (ANOSIM using Jaccard index). Richness and evenness values are based on 100 sub-samplings. Values for the community are based on the sub-sampling displaying less dissimilarity among groups (smaller R value). Significant effects are highlighted in green p-value ≤ 0.05...73

Table 2.3: Results of PERMANOVA with 999 permutations of the Jaccard dissimilarities for fungal mOTU community structure (DF, degrees of freedom; SS, sum of squares; MS, mean sum of square; F, pseudo-F by permutation)..74

Table 2.4: Kruskal-Wallis and Dwass Steele Critchlow Fligner tests of the average relative abundance of fungal phylum and classes among the stands (A) and tissue types (B). Significant differences (p ≤ 0.05) are highlighted in green, or yellow if the DSCF could not distinguish the effect due to low abundance. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA and Tukey tests instead of Kruskal-Wallis and DSCF tests..75

Table 3.1: Characteristics of Forêt Montmorency sites sampled in August 2012...106

Table 3.2: Results of PERMANOVA with 999 permutations of the Jaccard dissimilarities for fungal mOTU community structure (DF, degrees of freedom; SS, sum of squares; MS, mean sum of square; F, pseudo-F by permutation)..107

Table 3.3: Kruskal-Wallis and Dwass Steele Critchlow Fligner tests of the average relative abundance of taxonomic assignments among stands (A) and needle cohorts (B). Significant differences ($p \leq 0.05$) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA and Tukey tests instead of Kruskal-Wallis and DSCF tests..........108

Supplementary Table

Table S1.1: Combination of CLAM tests results to attribute categories to mOTUs present in needle samples..........43

List of abbreviations, acronyms

°C	degree Celsius
%	percentage
$	US dollar
A. b	*Abies balsamea*
ANOSIM	analysis of similarities
ANOVA	analysis of variance
Bp	basepair
BSA	bovine serum albumine
DF	degrees of freedom
DSCF	Dwass Steele Critchlow Fligner
DNA	deoxyribonucleic acid
dNTP	desoxyribonucleotide triphosphate
h	hour
ha	hectare
HMMs	hidden Markov models
HTS	high throughput sequencing
Hz	hertz
ITS	internal transcribed spacer
km, m, cm, mm	kilometre, metre, centimetre, millimetre
km^2, mm^2	square kilometre, square millimetre
LCA	lowest common ancestor
m.a.s.l.	metres above sea level
mg, µg	milligram, microgram
MID	multiplex identifier
Min	minute
mM	millimolar
mOTU	molecular operational taxonomic unit
MS	mean sum of square
Myr	million year-old
NCBI	national center for biotechnology information
nMDS	non-metric multidimensional scaling
PCR	polymerase chain reaction
PERMANOVA	permutational multivariate analysis of variance
pSSU	partial small sub-unit
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
S	Svedberg unit
sd	standard deviation
sec	second
SLP	single-linkage preclustering
SS	sum of squares

SSU	small sub-unit
UV	ultraviolet

To Herbert Copeland & Robert Whittaker for booting Fungi out of the Plant Kingdom

The truth is, most of us discover where
we are heading when we arrive
Bill Watterson - 1990

Acknowledgements

First of all, I like to thank my director, Dr Louis Bernier. I realise it might not have been evident to inherit a half-way through Ph.D student at a Christmas party. I am grateful for your help and advices transitioning from lab work to actually writing this dissertation.

I am also thankful to my co-director, Dr Jean A. Bérubé. You made me think of going for a doctorate in the first place and provide me with the opportunity to do so. I appreciate the confidence you show in me to complete this study.

I am immensely grateful to Dr Elizabeth A. Arnold. My visit to your laboratory remains for me the highlight of my Ph.D. I am thankful for your teachings, advices and kindness. You made me realize I may not be completely out of place in a laboratory.

To my folks, there is no word to match your unbounded support which led me to here and now. You have gotten me back on my feet since as long as I can stand, and likely more often in the past years than when I was a toddler. I can't but hope to prove myself worthy of such a love.

To my sister, Dr Pauline "Geeeorges" Ponchart, the best sibling I could have wished for (I do apologize for that biting accident in the late 1980s by the way). Let's face it, this all started with you explaining me about nucleosides, nucleotides, nucleic acids and amino acids (which also explain my reservation to ask you to teach me anything nowadays).

I sincerely thank Dr Jana U'Ren. I appreciate the time you took both to discuss about data treatment and teach me ecological analysis without which, the following chapters would be limping. Your mentorship allowed me to step up my game.

In all fairness, I could not not mention Patrick N. Gagné (well actually I could, it has been done before). You allowed me to almost immediately test couple of ideas I had thanks to the scripts you developed for me, bypassing the maddening and long process to learn computer talk (which I reserve for the upcoming months).

I am extremely grateful to the Centre for Forest Research and its bursary service which allowed me to present my work in different congresses, but also to go to the University of Arizona for an internship.

Due to the time frame of this project, a myriad of people have contributed either professionally or socially to my well-being. From the Laurentian Forestry Centre, I like to thanks the 2.24 alumni: Agathe Vialle, Franck Stephani, Guillaume Nicolas, Sylvie Carles and Karl Lundén (as well as Julie Godbout if the wifi signal is strong enough). From the same building I also thank Marie-Josée Bergeron, Amélie Potvin, Josyanne Lamarche, Nicole Lecours and Julie Dubé. The move to the Institute of Integrative Biology and Systems was made easy thanks to Martha Nigg, Hélène Vignaud, Éléonore Durand, Marine Vautier but also Sébastien Gérardi, Julien Prunier and Jérôme Laroche. Special thanks to the Jouven-Bleuvre, for either making me run a semi marathon (Clara), or teaching me forestry on site (Jean Marc). I apologize for stealing Watson when you were on holidays.

I also would like to thank Sophie-Anne Landry and Stéphanie Boivin from the “administrative” office of the faculty of forestry for their kindness and patience, and also the program director André Desrochers for allowing me non-unique last session.

The last though goes for old Ph.D students who might have lost their footings along the way. I can’t advice you enough to read “*South*” from Sir Ernest Shackleton, it will do the rest...

Preface

The following thesis is organized in three scientific papers written in English and edited correspondingly with the recommendations of the Faculté des études supérieures et postdoctorales (FESP) of Laval University. It is enclosed by a general introduction and a conclusion both written in English.

The first paper, *SSU to the rescue: conserving the original partial SSU fragment ahead of the ITS1 sub-locus enhances the detection of putative chimeras and limits the formation of possibly spurious taxa in fungal metabarcoding*, is to be submitted to Methods in Ecology and Evolution. The co-authors are Louis Bernier and Jean A. Bérubé.

The second paper, *Diversity, community composition and structure of the endomycobiota of balsam fir trees (Abies balsamea [L.] Mill.) as a function of tissue type*, is to be submitted to Frontiers in Microbiology. The co-authors are Louis Bernier and Jean A. Bérubé.

The last paper, *Last four cohorts of Abies balsamea needles display relatively similar diversities but different endomycobiotas*, is to be submitted to Fungal Ecology. The co-authors are Louis Bernier and Jean A. Bérubé.

I am the first author for each paper. I was responsible for the experimental designs, sampling, laboratory work, bioinformatics processing of genetic sequences, statistical analysis of data, and editing.

Introduction

1. Endophytism: definition and distribution

For a long time, the definition of the term “endophyte” attributed to De Bary (1866) (Arnold 2008) did not exceed its etymology, with “endo-” meaning within, and “-phyte” meaning plant; an endophyte was thus an organism living within a plant. Hence, this definition was mostly restricted to a location. Nowadays endophytes are microorganisms that live all or part of their life cycle within plants without causing any apparent damage or symptom to their host (Petrini 1991; Wilson 1995; Schultz & Boyle 2005). In addition to the location, this definition thus provides the effect of the interaction/association on the plant health status as not eliciting symptoms of disease. These two aspects seem to be the only properties endophytes have in common. Interestingly, the definition of endophytes differs from that of epiphytes which only mentions the position on the surface of plant. Endophytes do not belong to a restricted number of taxonomic ranks, nor do they exhibit similarities in their process of colonization. Furthermore, they display a wide range of functions. From this lack of homogeneity, it seems apparent that rather than describing a type of organisms, the term endophyte actually characterizes the transient or extended part of the life cycle of microorganisms that live within plants when they are not causing any apparent damage or symptom to their host. According to this definition, both latent pathogens and dormant saprotrophs are then considered as endophytes as they are present within the plant at a time of their life cycle when they are not triggering apparent symptoms on their hosts (Osono 2006; Saikkonen 2007; Newton et al., 2010; Boberg et al., 2011). Endophytes are most likely to be involved in neutral, commensalism, and mutualism interactions for instance. While the term applies mostly to both bacteria and fungi, the focus hereby will only be placed on the latter kingdom.

Endophytism is a long established association between Plants and Fungi (>400Myr) and is thought to have contributed to land conquest by plants, alongside with mycorrhizae (another Plant/Fungi association) (Krings et al., 2007; Rodriguez et al., 2009). Among fungi, endophytism is exhibited by an hyperdiverse group of species whose richness was estimated to 0.5 - 1 million species (Sieber 2007; Bills 1996). It represents a non negligible portion of the 2.2 - 3.8 million estimated species composing the fungal kingdom (Hawksworth & Lücking, 2017). Furthermore, endophytism is also widespread among plants as every plant sampled was found to harbour fungal endophytes, and it is generally considered that they are virtually found in every plant (Petrini 1986; Saikkonen et al., 1998; Rodriguez et al., 2009). It thus implies that not only this association is recovered from every major lineage of land plants, from grasses to non vascular plants, fern and allies,

conifers, and angiosperms, but also that it is retrieved from all terrestrial ecosystems (both natural and agro-ecosystems), occurring in different biomes, from the tropics to the tundra (Arnold & Lutzoni, 2007; Rodriguez et al., 2009). Endophytism is not limited to terrestrial ecosystems as it is also observed among aquatic plants (Sandberg et al., 2014). Despite the abundance of plant species colonized and the diversity of fungal endophytes, not all fungi are found within each plant. Mechanisms of recruitment of the community of fungal endophytes (endomycobiota) and even of colonization are not yet fully understood. But far from constituting random associations, fungal endophytes appear to show a specificity to their host species and sometimes even to a particular genotype of such host species (Ahlholm et al., 2002; Balint et al., 2013; Rajala et al., 2013). It also has to be noted that a certain seasonal variation of the composition of the endomycobiota is observed (Osono & Mori 2005; Guo et al,. 2008; Osono 2008) suggesting a dynamic process in the recruitment and survival of this community, possibly subject both to the environment but also to the already settled endophyte species.

2. Fungal endophyte classes

Traditionally, fungal endophytes are separated in two main groups based on phylogeny and life history traits. While the clavicipitaceous endophytes are likely the most studied group, they only represent a small number of phylogenetically related ascomycetous species confined to the clavicipitaceae family in the hypocreales order (Rodriguez et al., 2009). These clavicipitaceous endophytes are also restricted to some cool- and warm-season grasses present in both northern and southern hemispheres (Bischoff & White, 2005; Rodriguez et al., 2009). The particular attention they receive relies on their significant agricultural impact. Following the discovery by Bacon (Bacon et al., 1977; Porras-Alfaro & Bayman, 2011) that clavicipitaceous endophytes of pasture grasses were toxic to cattle, an estimated cost of $600 million a year to the livestock industry was attributed to the associated toxicity syndromes (Hoveland, 1993; Porras-Alfaro & Bayman, 2011). This is to be linked with one functional trait of clavicipitaceous endophytes: conferring enhanced resistance of host plants to herbivory either from insect feeding or mammalian herbivores (Rodriguez et al., 2009) usually resulting from the production of chemicals, notably alkaloids. Similarly these endophytes may confer disease resistance either by production of antifungal compounds, stimulation of plant defences, or simply trophic competition with pathogens (Rodriguez et al., 2009). Other potential functional roles include increase of plant biomass, as well as enhanced tolerance to abiotic stresses such as drought and metal contamination. The latter rely notably on the development of an extensive root system, increasing the area that can be prospected for both soil moisture and nutrients (Rodriguez et al., 2009). Clavicipitaceous endophytes represent a relatively more accessible model to study endophytology as the colonized plants usually harbour one dominant fungal isolate/genotype

(Wille et al., 1999; Rodriguez et al., 2009). The infection is also systemic, typically occurs in the intercellular spaces within the plant shoots, and transmission is primarily vertical (inherited) (Rodriguez et al., 2009). The infested plants also display a lower diversity of tissue, are less expanded in size, and more readily cultivable for experiments than woody plants. Yet, examining native plants under natural conditions is of uttermost importance as the endophytes-conferred benefits appear to be influenced notably by host genotype and environmental conditions, and might be missed in controlled experiments.

Nonclavicipitaceous endophytes represent a highly diverse polyphyletic group of species and the vast majority of fungal endophytes (Rodriguez et al., 2009), and are therefore the focus of this study. They mostly belong to the Dikarya sub-kingdom, and more particularly to the Ascomycota phylum. They are also the most widely distributed among plant lineages and across the globe. As for clavicipitaceous endophytes, they have been shown to confer to their host fitness benefits such as biotic and abiotic stress tolerance, improved nutrient uptake and increased growth and yields (Rodriguez et al., 2009). Rodriguez et al. (2009) separated them provisionally into three functional groups, mostly based on host colonization and transmission. Non-clavicipitaceous endophytes belonging to the first group colonize roots, stems, and leaves, and are able to form extensive infections within the intercellular spaces. Transmission is often vertical (inherited), but horizontal transmission is also observed. They are of most importance as they may be required for the normal development of some plants (Garbary & Macdonald, 1995; Rodriguez et al., 2009). Members of the second group are restricted to the root system and more often referred to as dark septate endophytes due to their display of darkly melanized septa. They may colonize both inter- and intracellular spaces (Rodriguez et al., 2009). They are found worldwide and are prevalent in high-stress environments (Rodriguez et al., 2009). They are especially associated with fine roots of conifers in boreal and temperate forests but not restricted to these plants nor habitats. They overall show a large host range and/or a lack of host specificity (Rodriguez et al., 2009; Jumpponen & Trappe 1998; Mandyam & Jumpponen 2005). Their transmission is considered to be horizontal and possibly originating from the soil fungal community. Finally, the last group of non-clavicipitaceous endophytes is confined to the above-ground plant tissues. Members of this group are hyperdiverse and display an extremely high diversity within individual host tissues, plants and populations (Rodriguez et al, 2009). They are likely to be the group containing most of the fungal endophytes biodiversity. They form highly localized infections of an estimated area of 2 mm^2 within inter-cellular spaces as observed in leaf tissue (Lodge et al., 1996; Arnold et al., 2000; Rodriguez et al., 2009). They are horizontally transmitted. Due to sampling strategies employed in the work presented thereafter, this last category of non-clavicipitaceous endophytes is likely to contain most of the species retrieved.

3. Colonization through space and time

Little is known about the mechanisms involved in the horizontal transmission of the endomycobiota. The environment seems to play a key role, notably as a source of inoculum. Environmental conditions, such as high relative humidity, might also influence favourably the colonization (Arnold & Herre, 2003; Rodriguez et al., 2009). Colonization is considered to be passive and to happen through the natural openings of the plant such as the stomata and the lenticels, but also via lesions and wounds (Johnston et al., 2006; Saikkonen 2007), and remain confined to the intercellular spaces. Microscopy studies do not reveal the establishment of any particular structures (Deckert et al., 2001; Johnston et al., 2006). However, colonization is not random, and differences both in diversity and composition exist between the endophytic- and the epiphytic communities (Santamaria & Bayman, 2005; Osono 2008). This suggests mechanisms of selection or recruitment of the endomycobiota from the plant. This is reinforced by the apparent host specificity of certain endophytes as mentioned earlier, specificity that might extend to a particular host genotype (Ahlholm et al., 2002; Balint et al., 2013; Rajala et al., 2013). Yet, the colonization is far from being homogeneous, notably due to the fact that infections are highly localized (Lodge et al., 1996; Arnold et al., 2000; Rodriguez et al., 2009). Variations in the environmental conditions, even at a micro-scale, also appear to influence the composition of the community recruited. Within the same trees, Unterseher et al. (2007) showed that some species of fungal endophytes prefer the shaded lower forest canopy to the sunny upper canopy. The latter is considered less stable with shorter periods of moisture on the leaf surface, broader fluctuation in temperatures, and possible higher UV radiation intensity. Schematic representation of leaf colonization by fungal endophytes shows a densely packed mosaic of diverse endophyte species (Lodge et al., 1996; Rodriguez et al., 2009). Differences also exist between the different parts of the leaf with distinct communities observed between the petiole and the blade (Hata & Futai 1996; Hata & Sone 2008).

Disparities in fungal endophyte distribution are not only observed across space but also along time. Within the same trees, seasonal variations in both diversity and composition of the endophytic community are usually observed (Osono & Mori 2005; Guo et al., 2008; Osono 2008). It is conceivable that plants are continuously exposed to an airborne inoculum whose composition evolves during the year, resulting in successive waves of colonization allowing additional endophytic species to settle and/or to replace part of the community established previously. As for the mechanisms of recruitment of the endomycobiota in itself, the dynamics and the evolution of its composition along time are far from being understood. Yet there is a delay in the

assemblage of the endomycobiota since endophytic infection is virtually null following the flushing of the foliage (confirming the horizontal transmission), whereas epiphytic colonization is almost immediate (Hata et al., 1998; Osono & Mori 2005; Guo et al., 2008; Osono 2008). As mentioned earlier, environmental factors such as precipitations and temperature (Cordier et al., 2012) may influence the efficiency of fungal endophytes to colonize plant tissues. These factors are also subject to variations along the year with different patterns depending on geographical position. The effect of time on the endomycobiota is difficult to evaluate. It is usually considered that the density of infection tends to increase with tissue age (notably for the foliage). Yet it has been shown that rather than being linked directly with the absolute age of the tissue, the density of infection in plants is more likely to be influenced by the duration of exposure to the environment (Arnold & Herre, 2003). Variation along time of tissue properties such as water or phenolics content (Hatcher 1990, Unterseher et al., 2007), and physiological changes might also alter the plant properties as an ecological niche and affect in turn the endomycobiota. The settled community of fungal endophytes might also modify in a similar way the habitat conditions to develop a less hostile environment (Hata et al., 1998). Again, the paucity of studies on this particular subject makes it difficult to precisely understand the interactions between the plant and its endomycobiota. But it also has to be noted that the fact that endophytes do not trigger symptoms on their host does not imply the absence of defense reactions from the plant, and endophytism is sometimes referred to as a balanced antagonism (Schulz et al., 1999). The switch between endophytic and pathogenic lifestyles may also result in a more aggressive colonization of the host, suggesting that the plant contains in some way the development of fungal endophytes (Eaton et al., 2011).

Part of the difficulty of studying interactions between the endomycobiota and its host lays on the fact that infections are highly localized, overall heterogeneous, and that they involve a community of hundreds of fungal endophyte species whose composition changes over time. This task is made even more difficult as fungal endophytes are not the only kind of organisms associated with plants. Fungal endophytes are indeed part of a larger entity, the plant microbiota which includes all the microorganisms, such as bacteria, fungi, protists and viruses, which are associated both superficially and internally with the plant. Overall, the plant microbiota plays key roles in different aspects of host metabolism and physiology, improving fitness, enhancing nutrient acquisition, but also providing resistance to pathogens and/or stress tolerance (Friesen et al., 2011; Bulgarelli et al., 2013, Hardoim et al., 2015). Its role also extends to a broader scope than its host as it may also influence plant community structure and ecosystem functioning (Porras-Alfaro & Bayman 2011; Rout 2014; Laforest-Lapointe et al., 2017). Increasingly, the approach underlying plant microbiota studies (and microbiota studies in general) is to include all components over all the taxonomic kingdoms involved, in order to decipher the interactions network. Such an approach will unarguably lead to a better understanding of the role of the

different members of the plant microbiota. For instance, one of the most cited examples of endophytic fungus benefiting its host is *Curvularia protuberata* which was originally thought to confer thermotolerance to its host *Dichanthelium lanuginosum*, allowing both of them to grow at high soil temperature in Yellowstone National Park (Redman et al., 2002). It has since been shown that it is actually an endohyphal virus, found within *Curvularia protuberata*, that conferred heat tolerance (Marquez et al., 2007). As the plant microbiota is composed of a myriad of microorganisms interacting both with the plant and among themselves (not to mention with the environment), deciphering such a complex network of interactions is a daunting task for which not every laboratory is equipped for. Yet, a contribution to that aspect might be to evaluate the composition and distribution of a particular subset of the plant microbiota, i.e. the endomycobiota.

4. Assessing fungal diversity

As mentioned earlier, estimates of the fungal diversity ranges from 1.5 million species (Hawksworth 1991) to 5.1 – 6 million (O'Brien et al., 2005; Blackwell 2011; Taylor et al., 2014), with more recent work refining this estimate to 2.2 – 3.8 million species (Hawksworth & Lücking, 2017). With only ca. 120 000 species described (Hawksworth & Lücking, 2017), the dimension of the fungal kingdom remains widely unknown with > 95% the species being undescribed. Turning to plants, notably the endomycobiota which is thought to be composed of 0.5 to 1 million species (Sieber 2007; Bills 1996), should thus allow a better grasp of that hidden diversity. Early days of endophytology relied on microscopy (De Bary 1866) but most knowledge on the subject comes from culture-dependent studies, notably from the works of G. C. Carroll, F. E. Carroll, O. Petrini, P. J. Fisher, J. K. Stone, and D. Wilson in the late 1970s and 1980s. Identifications were then based on morphotypes with the limitation that not every specimen sporulated in culture, the lack of sexual structures depriving scientists of useful informations to assess the nomenclature. The use of a molecular DNA-based approach to discriminate species quickly arose to complement morphotyping (Gardes & Bruns, 1991) and has somewhat superseded the culture-dependent approach during the last decade, notably due to the development of High Throughput Sequencing (HTS) techniques (Margulies et al., 2005). This concept of using DNA to identify species is far from being specific to fungal endophytes and is known as barcoding (Hebert et al., 2003). It facilitates the identification of microscopic organisms whose features may be difficult to observe, but also of macroscopic organisms at an early lifestage for instance, when not every character may be fully developed to allow species diagnosis. Different loci or groups of loci are targeted depending on the type of organism. Briefly, the concept relies on the fact that different species will contain different sequences of the barcode regions, and that ideally, every member of the same species will display the same sequence (or at least that interspecific variation

always exceeds intraspecific variation). Concerning the fungal kingdom, the official barcode is the nuclear ribosomal internal transcribed spacer (ITS) (Schoch et al., 2012, but in use for two decades beforehand).

The nuclear ribosomal transcribed spacer is a multi-copy (60 to 220 repeats (Cassidy et al., 1984; Russel et al., 1984)), tripartite segment, roughly 550-base pairs long, which is easily amplified by the Polymerase Chain Reaction (Nilsson et al., 2008). Nilsson et al. (2008) found the weighted average of intraspecific ITS variability across the kingdom Fungi to be 2.51% with a standard deviation of 4.57 (based on 4 185 fully identified fungal species). This variability differs among the different phyla ranging from 1.96% (standard deviation of 3.73) for the Ascomycota, to 7.46% (standard deviation of 4.14) for the Glomeromycota. It has to be noted that while this is of tremendous importance and extremely informative, it remains that this work is based on a relatively low number of fungal species (4 185 species) and that it will deserve a revision as data for such evaluation become available. Due to the existence of this intraspecific ITS variability, and in order to allow an effective use of barcoding for species identification, it is necessary to record every version of the ITS for each species. Yet such a level of completion is far from being achieved as of the 120 000 formally recognised species, only 34 878 are represented in Genbank (Hawksworth & Lücking, 2017). If initiatives such as UNITE (Koljalg et al., 2004) intend to develop an ITS reference database for fungal species, availability of ITS data remains at best fragmentary. Inclusion of the ITS barcode sequences in the description of species may prove to be useful in order to improve the level of completion of ITS reference database, but with around 1 800 fungal species newly described per year, reaching such goal may prove to be fastidiously long. Benefits of such an approach may also exceed species identification by barcoding to simplification of the nomenclature, allowing the removal of unnecessary synonymy. While Hawksworth (1991) mentioned a 2.5:1 level of synonymy, we found for fungal species associated with *Abies balsamea* (L.) Mill., the host on which this study was based, a 10.9:1 synonymy ratio (or 549 fungal species and a total of 5 994 synonyms). A “one species = one name” approach combined with listing the different versions of the barcode should thus prove useful for a better determination of the organism considered.

Barcoding is not limited to species identification (i.e. determining a name). For instance, it is also used to evaluate patterns of distribution even for species that may neither have their ITS sequences referenced nor be part of the minority of species described. The concept of molecular Operational Taxonomic Units (mOTUs), which act as a proxy for species, relies on forming groups of sequences based on a similarity threshold. It bypasses the shortcoming due to the low level of completion of the ITS reference database. While inherent to molecular based approach, it may be applied to both culture- and culture-independent methods. For the former

it may confirm that each morphotype represents a single species, while for the latter it forms the basis of observations. In the case of culture-independent methods, the term metabarcoding is usually preferred. Typically, metabarcoding involves the collection of environmental samples (mostly tree branches in the studies reported in this thesis, but may be as diverse as water sampling from lakes or even dust from a cabin), global DNA extraction, specific amplification of the barcode sequences, and sequencing. Development of the HTS techniques in the late 2000s has contributed to massively popularize metabarcoding studies as they permit sequencing millions of amplicons simultaneously. Before the development of HTS, a cloning step was necessary to segregate the amplicons and Sanger sequencing was performed one sequence at a time, thereby limiting the sequencing depth allowed per sample. While the massive amount of data is certainly a benefit, development of HTS is also accompanied by the production of shorter sequences that do not cover the full length of the fungal barcode. As mentioned earlier, the nuclear ribosomal Internal Transcribed Spacer is a tripartite segment composed of the rapidly evolving ITS1, the highly conserved 5.8S, and the moderately to rapidly evolving ITS2 (Nilsson et al., 2008). Studies based on HTS techniques have thus targeted either the ITS1, or the ITS2 sub-locus, with a preference for ITS1 as it displays on average a variability exceeding that of ITS2 (Nilsson et al., 2008) and may be more represented in nucleotidic databases. As with the full length fragment, intra-specific variability exists, and since mOTUs are formed based on a similarity threshold, it has to be stressed that no single value can perfectly delimit every species. A compromise of either 95 or 97% of similarity is usually employed and broadly accepted for studying global patterns of distribution and organisation of the biodiversity. The metabarcoding approach is even credited with being the major source for discovering novel fungal taxa (Hawksworth & Lücking, 2017).

5. Balsam fir in the context of global change

Balsam fir (*Abies balsamea* [L.] Mill.) is an important constituent of the boreal forest of eastern North America and the second most common species in Quebec after black spruce (*Picea mariana* Mill. B.S.P.). Consultation of the fungus-host database (Farr & Rossman, 2016) in the early days of this project showed a higher number of fungal species associated with balsam fir than with black spruce. The species displays sensitivity to climate change, and predictions show that it could disappear from about 20% of its distribution range (established on the 1961 – 1990 period), and remain present in less favourable conditions in about 40% of this area (Périé et al., 2014). This may favour the establishment of deciduous pioneer tree species such as white birch and aspen (Price et al., 2013). Those modifications in the distributions of tree species will likely be accompanied by long-term changes in the fungal community (Price et al., 2013). Such environment, with an expected increase of both abiotic and biotic stresses notably due to the progressive influence of climate change on boreal forests,

may represent an extremely alluring opportunity to study fungal endophytes and the possible evolution in the composition of the community recruited to face these environmental challenges. The increasing pressure of climate change on boreal forests may also lead to a better understanding of the functional roles of some members of the endomycobiota. Recording the biodiversity of the endomycobiota may also be considered in the context of conservation biology, in order to evaluate the dimension of the fungal kingdom before the possible disappearance of some of these species.

6. Scope of the thesis.

The main purpose of this Ph.D project is to develop a better understanding of the structure of forest trees endomycobiota using High Throughput Sequencing techniques and to challenge earlier conclusions based on culture-dependent methods. The work described in the thesis is structured in three chapters focusing on (1) the selection of a sub-fragment of the nuclear ribosomal Internal Transcribed Spacer to perform metabarcoding studies; (2) the effect of tissue type on the diversity and composition of the endomycobiota within branches; (3) the influence of the time of exposure to the environment on the diversity and composition of the endomycobiota of different needle cohorts.

Chapter 1 SSU to the rescue: conserving the original partial SSU fragment ahead of the ITS1 sub-locus enhances the detection of putative chimeras and limits the formation of possibly spurious taxa in fungal metabarcoding

Julien P. Ponchart[1,§], Louis Bernier[1,2], Jean A. Bérubé[3].

[1] Centre for Forest Research, Université Laval, Québec, Qc, Canada
[2] Institute for System and Integrative Biology (IBIS), Université Laval, Québec, Qc, Canada
[3] Laurentian Forest Centre, Canadian Forest Service, Natural Resources Canada, Québec, Qc, Canada
[§] Corresponding author: Julien P. Ponchart, email: julponchart@gmail.com

1.1 Résumé

Dans cette étude, nous avons évalué les avantages de conserver le fragment de pSSU précédant le sous-locus ITS1 chez les champignons. Nous avons montré que sa présence améliore la sensibilité de la détection des chimères (par un facteur 30) et limite la formation de singletons (-38%) et de doubletons (-22,7%), réduisant ainsi la diversité extrapolée (-21,9%). Nous avons développé une alternative au logiciel ITSx pour extraire plus facilement notre fragment d'intérêt (pSSU-ITS1) à partir d'amplicons fongiques. Notre approche, basée sur la détection d'un motif unique (ITS3a), présente une efficacité similaire à celle d'ITSx dans la détection de séquences fongiques non seulement à partir de notre jeu de données mais également pour la base de données de référence d'ITS fongiques (UNITE v7.1). Les analyses écologiques effectuées sur le sous locus ITS1 ou le fragment pSSU-ITS1 révèlent un patron similaire dans la structure de la communauté de l'endomycobiote d'un individu de sapin baumier.

1.2 Abstract

High throughput sequencing (HTS) techniques have brought new insights into the scale of diversity and distribution of the plant microbiota, including endophytic fungi. However, such data sets often contain a large number of singletons and doubletons for which it is hard to differentiate sequencing artefacts from biological reality. In the absence of a complete, high quality reference database, no firm solution exists to determine whether these singletons and doubletons represent members of a rare biosphere or spurious taxa. Here, we hypothesized that the selection of a particular sub-region of the fungal barcode might reduce their formation. The aim of this study was to evaluate if (1) conserving the original 5' extremity of the fungal amplicons, usually the partial fragment of the ribosomal small sub-unit (pSSU) immediately adjacent to the ITS1F primer ahead of the ITS1 sub-locus, impacts the detection of chimeras; (2) as the pSSU displays a lower variability, its inclusion in addition to the ITS1 sub-locus might also buffer the variation due to putative sequencing errors, and reduce the formation of those singletons and doubletons; (3) performing ecological analyses based on pSSU-ITS1 leads to similar conclusions on the community structure as analyses based on ITS1. For this study we analysed the fungal endophytic community (endomycobiota) of a single individual of *Abies balsamea* (L.) Mill. based on 108 samples distributed among both the aerial and root systems. We first show that conserving the pSSU ahead of the ITS1 sub-locus enhances the sensitivity of chimera detection (30 fold rate increase) while limiting the formation of singletons (-38%) and doubletons (-22.7%) and reducing thus the overall observed and extrapolated diversity (-20.8, and -21.9%, respectively). We then develop an alternative approach to the commonly used ITSx software to readily extract our fragment of interest (pSSU-ITS1) from fungal amplicons, based on the detection of a single pattern (ITS3a). We show that our ITS3a-based approach presents a similar efficiency to ITSx in detecting fungal sequences not only from our dataset but also for the fungal ITS reference database (UNITE v7.1). Finally, we demonstrate that ecological analyses performed on either ITS1 or pSSU-ITS1 fragment reveal a similar pattern in the community structure of the endomycobiota of a single balsam fir tree.

1.3 Introduction

With the development of High Throughput Sequencing (HTS) techniques, DNA metabarcoding studies have improved our understanding of the extent and distribution of biological diversity. These methods have been applied successfully in domains ranging from the human (Hamady & Knight, 2009) and plant microbiotas (Jumpponen & Jones, 2009), to marine ecosystems (Leray & Knowlton, 2015), soil ecosystems (Buée et al., 2009; Yoccoz et al., 2012), air (Fröhlich-Nowoisky et al., 2009; Yamamoto et al., 2012), and indoor environments (Amend et al., 2010), with recent extensions to conservation purposes (Ji et al., 2013; Thomsen & Willerslev, 2015). However, such approaches are limited to determining associations rather than interactions as they can only detect the presence of organisms and cannot differentiate living from dead cells, although complementation with RNA metabarcoding might discriminate the “active” biodiversity (Stecher et al., 2016).

Metabarcoding approaches are not without drawbacks and methodological artefacts also exist. They originate from three main sources of errors: PCR chimeras, PCR single base substitutions, and sequencing errors (Quince et al., 2011). If PCR linked errors are independent of the HTS (no effect on the incidence frequency), their impact (i.e. the number of sequences concerned) is exacerbated by the large amount of data produced. Sequencing errors, on the other hand, are inherent to the HTS employed. While most of these issues have been addressed (Quince et al., 2009; Huse et al., 2010; Quince et al., 2011; Schloss et al., 2011), and even with state of the art analysis, estimates of the biodiversity based on metabarcoding approaches are globally considered to be inflated. This is notably due to remaining erroneous sequences putatively forming undetected chimeras (Deiner et al., 2016) and a large proportion of singletons and doubletons (Tedersoo et al., 2010; Unterseher et al., 2011; Lindahl et al., 2013; U’Ren et al., 2014). Whereas some of these less represented molecular Operational Taxonomic Units (mOTUs) compose the “rare biosphere”, others were shown to be spurious variants of more frequent mOTUs (U’Ren et al., 2014).

Assigning singletons and doubletons to one or the other categories mentioned above is impossible in the absence of a complete, high quality reference database. The lack of such database is notably worth mentioning for the fungal barcode, the nuclear ribosomal internal transcribed spacer (ITS) (Schoch et al., 2012). With ca. 120 000 species described from the 2.2 – 3.8 million species (Hawksworth & Lücking, 2017), > 95% of the fungal diversity remains unknown, and only about 30% of the described species are referenced molecularly by their barcode region in the GenBank database (Hawksworth & Lücking, 2017), and ca. 14% in the fungal reference database Unite (Koljalg et al., 2005). If remaining erroneous sequences may not be

unquestionably identified as such, their impact might be limited by exploiting their proprieties: sequences with errors are likely to be rare, and they should be similar to a true abundant sequence (Quince et al., 2011). For the chimeras, acknowledgement of the mechanisms involved in their formation might improve their detection. In both cases, the attentive consideration of the ITS sub-region to be used for fungal metabarcoding studies might partially lift the uncertainty of the spurious nature of some rare mOTUs.

Selection of a ITS sub-region has to be considered as HTS techniques used for metabarcoding studies can not reach the ca. 450-800 bp of the fungal barcode. It is notably the case for 454 pyrosequencing whose particularity is to generate intermediate length reads (Margulies et al., 2005), i.e. the sequencing is randomly interrupted before reaching the 3' extremity of the amplicons. As the chimera detection exploits the sequence abundances (Quince et al., 2009), the ITS1 sub-locus (which tends to be favoured over the ITS2 as it generally displays a higher variability (Nilsson et al., 2008)) is extracted in order to increase the accuracy of these abundances. This lead to the removal of the original 5' extremity of the amplicons, usually the partial fragment of the ribosomal small sub-unit (pSSU) immediately adjacent to the widely used ITS1F primer (Gardes & Bruns, 1993). The formation of chimeras happens when a partially amplified sequence fragment during one cycle of the PCR serves as primer for a different sequence in the following cycle (Quince et al., 2011). Chimera detection is performed on sequences which do not display the original extremities of the amplicons. The absence of these short conserved extremities during chimera detection might thus impair the process.

The extraction of the ITS1 sub-locus is also performed as the less variant flanking regions may distort the sequence clustering and/or may skew the similarity searches for taxonomic affiliation (Nilsson et al., 2010) depending on the length of the residual portions left. But despite being more conserved than the ITS1, the SSU remains variable enough to prevent the use of pattern matching to remove it (Nilsson et al., 2010). Considering a pSSU-ITS1 fragment as a potential sub-barcode allows also the selection of a segment with properties similar to those of the full length ITS. The ITS contains in its middle the highly conserved 5.8S which constitutes about one third of the global length of the barcode region (Yang et al., 2018). The pSSU immediately following the ITS1F is ca. 50bp and represents about one quarter of the pSSU-ITS1 length. It should thus decrease the overall variability and allow to use the same similarity threshold as with the full-length ITS barcode. As the accuracy of nucleotide assignment decreases with the progression of the elongation (Margulies et al., 2005), keeping the original 5' extremity of the sequences might also allow to buffer the sequencing errors, more likely to occur towards the 3' extremity.

The aim of this study was to examine the potential advantages of using a pSSU-ITS1 fragment over the ITS1 sub-locus as a sub-region of the ITS for metabarcoding studies employing HTS techniques. This approach was tested on the fungal endophytic community (endomycobiota) of a single individual of *Abies balsamea* (L.) Mill. from which 108 samples distributed among both the aerial and root systems were analyzed. We first tested the effect of conserving the pSSU as original 5’ extremities of the fungal amplicons on the detection of chimeras, hypothesizing that its removal might impede the process. Secondly, we determined the effect of conserving the less variable pSSU in front of the ITS1 sub-locus on the clustering into mOTUs, notably on the proportion of singletons and doubletons formed. Thirdly, we developed an alternative to the ITSx software (Bengtsson-Palme et al., 2013) to readily extract the pSSU-ITS1 fragment, and evaluated its efficiency compared to ITSx to detect fungal sequences from the ITS reference database UNITE. Last, we performed a range of ecological analyses including diversity, community composition and correlation-based network analysis to infer if metabarcoding studies based on pSSU-ITS1 resulted in observations different from those obtained by standard ITS1 sub-locus analysis.

1.4 Materials and Methods

1.4.1 Sampling and sample preparation

Sampling was performed in September 2013 at the Montmorency forest, ca. 70km north of Quebec-City. *Abies balsamea* is the dominant tree species in the area and covered more than 75% of the 4.7 ha study plot. The tree selected had no visible damage or symptoms of disease, was situated at 762 m.a.s.l. (47°19.489’ N - 71°05.627’ W), about 11 meters tall, and its age was evaluated to 25-30 years. Branches were collected at six different heights (2 m, 3 m, 4 m, 5.5 m, 6.5 m, and 7.5 m) from two opposite sectors. Two tree-ring cores (0.5 cm in diameter) were also collected from the trunk at five different heights (0.2 m, 1.2 m, 2.2 m, 3 m, 4 m) orthogonally of each other. Roots were sampled from 30 cm under the soil surface and priority was given to the thinner portion of the root system including rootlets (i.e., only roots $\leq$ 5cm in diameter were kept). Due to the complexity of interconnected root networks, only two fragments (each one about 1.2 m in length) could be assigned to the designated tree.

Each branch was sub-sampled as follows: 40 symptomless needles, 7 buds, 4 fragments of bark and wood (deprived of needles; 5 cm each). The last four annual growth increments of the main axis of the branch were equally sampled (10 needles for each year, and one 5 cm wood/bark fragment) in order to retrieve maximum biodiversity in sampled plant parts (exposed to up to 4 years of colonization). Following surface sterilisation (see Stefani & Bérubé 2006), needles were cut into 1mm sections while buds, bark, and wood were ground into a fine powder in a sterile mortar with a pestle and liquid nitrogen. Wood from tree ring cores was not surface sterilized (as no epiphytes were expected), but bark extremities were removed before grinding. Root fragments were rinsed in two distilled water baths before sub-sampling. Because the distribution of rootlets was too scattered to define distinct sample units, they were pooled per fragment and ground into fine powder (using liquid nitrogen) following superficial sterilization. Wood and bark from the root were collected as five 5cm long pieces of both 1cm and 0.5cm in diameter per fragment. After superficial sterilization, pieces were separated into bark and wood, and then ground. In total 108 sample units from the balsam fir tree were obtained.

1.4.2 DNA extraction and library construction

Assays were carried out to determine the optimal mass for each of the tree tissue types for DNA extraction. It was initially done using 60mg and the quantity of material was adjusted according to results (data not shown). Final samples consisted of 100mg of needles and wood, 30mg for buds and rootlets, and 60mg for bark. DNA was extracted using the DNeasy® Plant kit (Qiagen GmbH, Hildren, Germany). A sterilised 3mm tungsten carbide bead (Qiagen) was added to each sample, which was then subjected to a two step grinding in a Mixer Mill Retsch MM300 (Qiagen) for 2 min at 30 Hz. Then, 400 µL of extraction buffer AP1 (500 µL for the wood samples due to absorbance properties of the tissue), 2 µL of RNase A and a tip of Dx antifoaming reagent (Qiagen) were added. Samples were subjected to a third grinding of 2 min at 15 Hz in the Mixer Mill. DNA was extracted following the manufacturer's instructions except that extracted DNA was eluted with 75 µL of buffer in two steps (50µL then 25µL).

The ITS region was amplified using the specific fungal primer ITS1F (Gardes & Bruns, 1993) and the universal primer ITS4 (White et al., 1990) adapted for the 454 Roche pyrosequencing Lib-L technology. ITS1F featured the adaptor A and a 10bp Roche MID barcode and ITS4 carried the adaptor B. Each sample was amplified in three independent reactions (using the same MID). The amplification reaction contained 25 µg of Bovine Serum Albumine (BSA, Sigma-Aldrich, St Louis, MO, USA), 1.25 X PCR buffer (Invitrogen, Carlsbad, CA,

USA), 2.5 mM of $MgCl_2$, 250 µM of each dNTP, 400nM of each of the forward and the reverse primers, 1 unit of Platinum Taq DNA polymerase (Invitrogen), and 1 µL of template DNA. The PCR cycle parameters consisted of an initial denaturation at 94 °C for 3 min, then 32 cycles of denaturation at 94 °C for 30 sec, annealing at 52 °C for 30 sec and extension at 72 °C for 1 min, followed by a final extension step at 72 °C for 10 min.

Triplicate PCR products were pooled and purified using the Agencourt Ampure XP magnetic PCR clean-up system (Beckman Coulter, Brea, CA, USA). Primer-dimers were eliminated using a 1.8:1 volume ratio of magnetic beads for amplicons. A 0.6:1 volume ratio was used for discarding fragments smaller than 150-200 bp. DNA was eluted in 26 µL of EB buffer (Qiagen) and concentrations were measured using the Quant-iT Picogreen dsDNA assay kit (Invitrogen, Eugene, OR, USA). The 108 samples were divided into four amplicons libraries in order to reach a significant sequencing depth and were in equimolar proportions within each library. Following pre-processing of the data, a fifth library composed of the 27 samples with the lowest number of reads was constructed. For each library, amplicons length was verified using an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Unidirectional sequencing was performed with Roche GS FLX+ system (Roche-454 Life Sciences, Branford, CT, USA) at McGill University and Genome Quebec Innovation Centre (Montreal, QC, Canada).

1.4.3 Bioinformatic analyses

Data were treated as described in U'Ren et al., (2014) with slight modifications. Denoising was performed using the mothur (v1.33.1, Schloss et al., 2009) implementation of the PyroNoise algorithm (Quince et al., 2009). Despite computational cost, the mindelta parameter set on 10^{-6} was preferred to the fixed iteration number of 1000 for the shhh.flows command. Following denoising, all sequences containing any mismatch to the primer and the barcode, any ambiguous base call, or any homopolymer repeat longer than 9bp were removed. ITS1 sub-loci were extracted with ITSx (Bengtsson-Palme et al., 2013). Putative chimeric sequences were then detected with chimera.perseus in mothur and removed. Sequences were clustered into molecular Operational Taxonomic Units (mOTUs) as described in U'Ren et al. (2014). Unique sequences were processed with ESPRIT (Sun et al., 2009) to perform pairwise alignments. Using the single-linkage preclustering (SLP) algorithm developped by Huse et al. (2010), the distance matrix was processed to a precluster step with a width of 0.03. Clustering was done at a 95% sequence similarity with mothur using the average neighbour method. Representative sequences of each mOTU were retrieved from the preclustering

step and all unique sequences were conserved to observe intra-mOTUs variation. These sequences were then blasted (Altschul et al., 1990) against the nt database (downloaded on September 24th, 2016). Kingdom assignment was assessed using the Lowest Common Ancestor algorithms featured in Megan (Huson et al., 2007). The following parameters were used: min support=1, minscore=200, top%=10, min comp=0.3, %ID filter (U'Ren et al., 2014). All mOTUs which were not directly assigned to the fungal kingdom were checked manually to ensure their taxonomic affiliation and only fungal mOTUs were retained for further analyses. This approach was preferred over the rdp method (see below), as the latter detected false positives (data not shown).

To observe the potential effect of conserving the original 5' extremity of the amplicons on the detection of chimeras, sequences for which the ITS1 sub-locus were retrieved with ITSx were re-extracted to feature the pSSU-ITS1 fragment. To this end all different versions of the first 50bp of the 5.8S (obtained with ITSx) served as exact patterns in fqgrep (ver 0.4.4, Indraniel et al., 2016) and the -c parameter was used to highlight their positions. Shell sed commands based on the colour tag in front of the pattern were used to remove every nucleotide from that point resulting in sequences only displaying the pSSU-ITS1 fragment. Detection and removal of putative chimeras were performed as described above. In order to single out the effect of the pSSU-ITS1 in the detection of chimera and not on the clustering a version of this dataset displaying only the ITS1 was generated and the remaining parts of the analysis were conducted as described above. A third version of the analysis was also performed on the original pSSU-ITS1 dataset following the removal of the chimeras to evaluate the potential effect of the pSSU-ITS1 on the clustering.

To readily extract the pSSU-ITS1 fragment, an alternative to the ITSx software was tested. Using ITS sequences of 308 cultures of fungal endophytes isolated from balsam fir needles (personal data), a putative frame of the 5.8S was drawn following alignment with ClustalW (default parameters, Larkin et al., 2007) and visualized with Bioedit (Hall, 1999). The most frequent version of the first 60 base pairs was selected to serve as a query. A dataset of 569 420 nucleotidic sequences was downloaded from NCBI on September 26th 2016, using the keywords "internal transcribed spacer*" and "5.8S" and selecting the fungal kingdom as organisms. Sequences shorter than 60bp, containing ambiguities or homopolymers longer than 9bp were removed, redundancy was also eliminated with mothur. A total of 413 263 sequences were conserved to constitute a custom database. The blast of the query against this database was run with a gap open penalty of 25 and a gap extension penalty of 2. It resulted in 321 735 sequences aligned across the 60bp. The oligonucleotide ITS3a (CTTTCAACAACGGATCTCTT) was defined to be at the start of this alignment and three mismatches

were allowed as it corresponded to the number of bases with frequencies lower than 99.5% (Figure 1.1). This approach for detecting fungal sequences was validated against the Unite v7.1 fungal ITS reference database (accessed in November 2016) using fqgrep to assess the presence of this pattern and results compared to those obtained with ITSx.

Following the denoising and trimming of the sequences of our dataset, the extraction of the pSSU-ITS1 fragments was performed using our ITS3a based approach, before proceeding with the rest of the analysis (chimera detection and clustering based on pSSU-ITS1). Finally, a combination of our ITS3a based approach with ITSx was also evaluated: all sequences for which ITS3a could not be detected were passed through ITSx. Sequences for which the ITS1 sub-locus was detected with ITSx were re-extracted to feature the pSSU-ITS1 fragment as described above. These sequences were then pooled with those extracted with fqgrep ITS3a and the analysis was then resumed. Representative sequences of each mOTU for this approach were deposited in GenBank (Submission number: SUB5671914)

1.4.4 Ecological analyses

Ecological analyses were performed to evaluate if considering the pSSU-ITS1 fragment instead of the ITS1 sub-locus led to the observation of different organisation structures of the fungal endophytic community. While the original analysis based on the ITS1 sub-locus served as a standard, only the combination of our ITS3a based approach with ITSx was used for the pSSU-ITS1 fragment. The individual-based species-accumulation curves using the number of observed mOTUs (Sobs), its 95% confidence intervals and bootstrap estimates of species richness were computed in mothur, sampling every 1000 sequences. The curves were plotted with the R program version 3.2.3 (R Core Development Team, 2015). Total mOTU richness was extrapolated using the bootstrap values from the specpool function of the vegan package (Oksanen et al., 2007; Smith and van Belle, 1984).

To reduce the stochasticity of the following observations, one hundred sub-samplings (selecting randomly 3 900 sequences per sample) were performed using mothur on the original sample-by-mOTU feature table deprived of singletons and doubletons. Three root samples were removed as they contained fewer sequences. Possible effect of the system (root or aerial) on the diversity was evaluated using the Fisher’s alpha index (Fisher et al., 1943), a robust measure relatively insensitive to rare species (Colwell 2009; Magurran 2004),

while evenness was measured with the Pielou's evenness index (Pielou 1966; Jost 2010). Analyses were based on the mean average values of the indexes computed from the 100 sub-samplings (with vegan R package). Analysis of community similarity (One-way ANOSIM; Clarke 1993) between the two systems was conducted using Jaccard's index in PAST (Hammer et al., 2001). Results were visualized with non-metric multidimensional scaling (nMDS) in PAST (Taguchi & Oono, 2005) and run on the 100 sub-samplings. All statistics were performed with R unless specified otherwise.

The structure of the endomycobiota was also evaluated using correlation-based network analysis. All mOTUs composed of less than 3 sequences, or distributed in fewer than 3 samples were discarded from the original sample-by-mOTU feature table to reduce the network complexity through decreasing sample-specificity. All possible pairwise Spearman's rank correlations between mOTUs were calculated using the Hmisc R package (Harrell 2008). Only correlations with a Spearman's correlation coefficient (ρ) > 0.6 and a P-value < 0.01 were considered. The network was visualized and its modules detected using Cytoscape (ver 3.3.0, Shannon et al., 2003). Taxonomic assignments to the class level served as attributes to the nodes. These assignments were defined with the mothur classify.seqs command using rdp implementation (Wang et al., 2007) and UNITE v7.1 as template, with 1000 iterations, a cutoff of 60 (but support of 80), and kmer size of 8. The multinomial species classification method (CLAM test from vegan R package, "supermajority" rule; Chazdon et al., 2011) was used to assess if a fungal mOTU was preferentially retrieved from either the root or the aerial system. Results from the CLAM test also served as attributes to the nodes of the network. The position of the needle community (the most studied tissue of plant microbiota) relative to the endomycobiota of the root and aerial systems, was also evaluated (see Table S1.1 for details) to delve deeper into the community structure.

1.5 Results

1.5.1 SSU effect

Sequencing of the libraries produced 1 069 756 reads which were reduced to 1 039 365 trimmed sequences following the denoising step (Table 1.1). Detection of chimeras was influenced by the fragment considered with a higher number of sequences assessed as putative chimeras when performing the analysis on pSSU-ITS1 than on ITS1 alone. Extraction with ITSx retrieved 941 825 sequences of which 1 269 were considered as putative chimera when the analysis was performed on the ITS1 sub-locus (Table 1.1). When using the same

subset of 941 825 sequences but displaying the pSSU-ITS1 fragment this time, 37 278 sequences were assessed as putative chimeras (Table 1.1). While 1 051 sequences were commonly detected as chimeras independently of the presence of the pSSU, 218 were assessed as putative chimeras only when considering the ITS1 sub-locus, and 36 227 specifically with pSSU-ITS1 (data not shown).

Clustering was affected by the fragment considered: keeping the pSSU ahead of ITS1 led to a reduction of the number of mOTUs formed. The effect was evaluated using the subset of 941 825 sequences extracted with ITSx deprived of the 37 278 putative chimeras assessed on pSSU-ITS1 fragment (Table 1.1). Following the removal of 11 922 non-fungal sequences, clustering based on the ITS1 sub-locus of the remaining 892 625 sequences (mean length of 185.5bp) resulted in 2 542 mOTUs (including 207 doubletons and 1 110 singletons) (Table 1.1). Clustering on the same subset of 892 625 sequences based on the pSSU-ITS1 fragment (mean length of 234.7bp) induced a 20.8% reduction in observed richness (2 013 mOTUs from 2 542). This decrease concerned mostly the formation of singletons (688 from 1 110, 38% reduction) and doubletons (160 from 207, 22.7% reduction). The formation of mOTUs composed of at least three sequences was reduced by less than 5% (1 165 from 1 225 mOTUs) (Table 1.1). The estimation of the extrapolated richness was also reduced by 21.9% (2 502 ± 72 mOTUs from 3 204 ± 83) (Table 1.1).

To facilitate the direct extraction of the pSSU-ITS1 fragment, we developed an approach based on the detection of a single pattern (ITS3a) using fqgrep. The validity of this method was evaluated using the Unite v7.1 fungal ITS reference database (accessed on November 20th 2016) which contained 54 749 fungal sequences for a total of 54 873 sequences (the database also displayed few sequences previously misassigned as fungal). About half of these fungal sequences (24 288) were assigned up to the species level and distributed among 16 531 full binomial names. The ITS3a pattern was identified in 53 287 sequences (16 374 binomial names). Extraction with ITSx yielded a similar proportion with 54 032 sequences (16 447 binomial names). The ITS3a and ITSx approaches detected a common set of 52 972 sequences (16 331 binomial names). This set lacked 315 sequences (43 binomial names) retrieved specifically with our ITS3a approach, and 1 060 sequences (116 binomial names) retrieved with ITSx. The combination of both methods permitted the retrieval of 54 347 sequences (16 490 binomial names). The remaining 526 sequences (85 binomial names) of the Unite v7.1 database were never detected. Similar results were observed on our balsam fir dataset, with the ITSx and ITS3a approaches yielding a similar proportion of sequences (941 825 and 943 539, respectively) (Table 1.1). While most of these sequences (940 233 sequences) were recovered by both ITSx and ITS3a, a few others were specifically detected by one method only (1 592 by ITSx, 3 306 by ITS3a).

The combination of both methods allowed to recover more sequences (945 131 sequences) than each individual method (Table 1.1), potentially representing a more complete richness.

The ITS3a-based approach combined with ITSx was thus used to compare the results between the standard analysis based on the ITS1 sub-locus and the one conducted on the pSSU-ITS1 fragment. For the ITS1 analysis, 928 348 sequences were conserved for clustering (mean length of 185.1 bp) following the removal of putative chimeras (1 269 sequences), and non fungal sequences (12 208) (Table 1.1). 2 546 mOTUs were formed, including 207 doubletons and 1 115 singletons. Alternatively, for the pSSU-ITS1 analysis once the 37 406 putative chimeras and 12 110 non fungal sequences had been removed, the clustering was performed on 895 615 sequences (mean length of 238.1bp) and led to an observed richness of 2 041 mOTUs (among which 164 doubletons and 700 singletons) (Table 1.1). The extrapolation of the total richness associated to a single balsam fir tree was estimated to 3 208 ± 83 mOTUs when based on the ITS1 sub-locus and 2 536 ± 73 mOTUs with pSSU-ITS1 (Table 1.1). In both cases, about 80% of the expected richness was thus collected. In spite of the number of samples and effort made on the sequencing depth, the resulting accumulation curves were not saturated (Figure 1.2).

1.5.2 Community analysis

Overall numbers of mOTUs featured in the 100 sub-samplings were similar with 1 215 mOTUs retrieved with the ITS1 approach and 1 170 with pSSU-ITS1. Yet the proportion of mOTUs which could be assigned at least to the Class level was higher when considering the ITS1 sub-locus (ca. 60% for about 50% when analysis based on pSSU-ITS1) (Figure 1.3). These mOTUs belonged mostly to the Dikarya sub-kingdom (ITS1: 73.5%; pSSU-ITS1: 67.01%), but an important fraction could only be assigned to the Ascomycota phylum (ITS1: 13.99%; pSSU-ITS1: 18.89%) or to the fungal kingdom (ITS1: 23.79%; pSSU-ITS1: 30.68%). Zygomycota and Chytridiomycota were also detected but in lesser proportion (1.98% and 0.165%, respectively for ITS1; 2.14 and 0.17%, respectively for pSSU-ITS1) (Figure 1.3). Rozellomycota (0.4%) and Glomeromycota (0.165%) were also represented in the ITS1 analysis but not in the pSSU-ITS1 analysis. However, the sequences composing the mOTUs from these phyla were in the pSSU-ITS1 dataset (with the exception of 3 sequences considered as chimeras), and in most cases formed identical mOTUs but with different representative sequences leading to an assignment to the kingdom rather than the phylum level.

Whereas there were differences in the proportion of assignment to the different taxonomic levels, observations of the taxonomic structures between the root and the aerial system were mostly similar independently of the fragment considered for the analysis. The relative abundance of Ascomycota was higher within the aerial system than within the root system (Figure 1.3, Table 1.2 and Table 1.3), but the latter was enriched in Basidiomycota and Zygomycota. This was essentially due to Agaricomycetes and Mortierellales, respectively (Figure 1.3, Table 1.2 and Table 1.3). Despite this general trend, relative abundances of Leotiomycetes, Pezizomycetes, and Saccharomycetes were higher in the root system (Table 1.2 and Table 1.3). The same was observed for the Archaeorhizomycetes in the analysis based on the ITS1 sub-locus, but the abundances were similar between the two systems for the pSSU-ITS1 analysis. Proportion of Sordariomycetes was not influenced by the system (Table 1.2 and Table 1.3). Cystobasidiomycetes, Exobasidiomycetes, Agaricostilbomycetes, and unclassified Basidiomycota were found in higher proportion in the aerial system for the ITS1 analysis, but it was restricted to the Exobasidiomycetes class for the pSSU-ITS1 analysis.

Diversity analysis led to similar results for both ITS1 and pSSU-ITS1. Diversity of the root system was lower than that of the aerial system when measured with the Fisher's alpha index (ITS1: Kruskal–Wallis $\chi^2 = 47.546$; $P = 5.372 \times 10^{-17}$; pSSU-ITS1: $\chi^2 = 45.351$; $P = 1.647 \times 10^{-11}$, Figure 1.4.a and Figure 1.4.b) and evenness evaluated with Pielou's index was also lower in the roots (ITS1: ANOVA $F = 24.982$; $P = 2.38 \times 10^{-6}$; pSSU-ITS1: $F = 28.543$; $P = 5.516 \times 10^{-7}$, Figure 1.4.c and Figure 1.4.d). Both systems were colonized by distinct communities (ITS1 ANOSIM $R = 0.7002$; $P = 0.001$; pSSU-ITS1: $R = 0.7002$; $P = 0.001$, Figure 1.5.a and Figure 1.5.b). The influence of the systems on the fungal endomycobiota structure of this single tree was also observed with the co-occurrence analysis. In both cases the overall network (ITS1: 338 mOTUs establishing 838 pairwise correlations; pSSU-ITS1: 292 nodes and 706 edges) could be divided into 2 main modules, each representative of one system (Figure 1.6). Isolated correlations were grouped into an accessory module (ITS1: 67 nodes – 48edges; pSSU-ITS1: 56 nodes – 37 edges) (data not shown). Interestingly, most of the mOTUs defined as specialists of needles were found in the root module (ITS1: 14 of the 18; pSSU-ITS1: 12 of the 15) and arranged in a sub-module connected to the main root module (Figure 1.6). A micro aerial sub-module connected to the main root module was also observed. Most of the edges linked the same class nodes as those defined by the CLAM analysis (ITS1: 562; pSSU-ITS1: 483) (Figure 1.6). Conclusions on the taxonomic associations were hindered as a minority of edges had both their extremities assigned to the class level (ITS1: 229, 77 paired the same class, 152 linked different classes; pSSU-ITS1: 116, 54 same class, 62 different classes).

1.6 Discussion

The main goal of this study was to assess the potential benefits to consider a pSSU-ITS1 fragment as a fungal barcode sub-region over the ITS1 sub-locus. The necessity to define a standardized targeted ITS sub-region comes from the limitations of the HTS employed for metabarcoding studies to produce the full length fungal barcode. It is the case for 454 pyrosequencing whose particularity is to generate intermediate length reads with sequences usually randomly interrupted within the 5.8S or ITS2 sub-locus. Another issue with metabarcoding studies is the inflation of the biodiversity estimates due to the formation of a large proportion of singletons and doubletons. Most of them are considered as possible spurious taxa and find their origins in three sources of errors: PCR chimeras, PCR single base substitutions, and sequencing errors (Quince et al., 2011). We hypothesized that the selection of the pSSU-ITS1 sub-fragment might improve the evaluation of these types of errors.

A ca. 30-fold increase in the detection of putative chimeras was observed when the analysis was based on the pSSU-ITS1 fragment rather than on the ITS1 sub-locus. Interestingly, when performing the clustering into mOTUs based on the ITS1 sub-locus, the same richness was observed independently of the fragment used for the detection of chimeras. Sequences that were specifically considered as chimeric based on the pSSU-ITS1 analysis were distributed among 474 mOTUs in the analysis on the ITS1 sub-locus. They were found in low occurrence within each mOTU (account for an average of 6.72%, sd = 16.27%, but median at 1.73%). This result suggests that the breaking point between the two parental sequences from which the chimera originated is likely to be situated somewhere before the transition between the SSU and the ITS1 sub-locus. While the choice of fragment for chimera detection does not seem to impact the richness observed in this particular study, it implies the conservation of ITS1 sub-locus issued from spurious sequences composed of the SSU of one species and the ITS1 of another.

The need to select a sub-region of the fungal barcode for chimera detection is a particularity of the 454 pyrosequencing data due to the intermediate length reads produced. More recent HTS techniques (notably Illumina paired-end sequencing) circumvent this problem by targeting directly from the amplification either the ITS1 or ITS2 sub-locus, and the sequences obtained display the full length of the amplicons. The targeted sub-locus is usually embedded with short conserved extremities (partial fragments of SSU and 5.8S for ITS1; partial fragments of 5.8S and LSU for ITS2). Yet, because high throughput data treatments are inherited from 454 pyrosequencing, chimera detection still often follows the extraction of either ITS1 or ITS2. Therefore, we

recommend implementing the chimera detection step prior to the extraction of the targeted sub-locus, as removal of the original extremities impairs the process.

In this study, we also hypothesized than conserving the pSSU fragment might buffer the variation due to sequencing errors and that some singletons might group with the mOTUs from which they possibly deviate, thereby reducing the over-estimation of biodiversity. A possible drawback of this approach might be the fusion of closely related taxa into a single mOTU. This problem is similar to the choice of a similarity threshold for the clustering into mOTUs, independently of the selected fragment, as no single value might account for the difference of intra-species variations. Since the conservation of the pSSU affected mostly the formation of singletons (-38%) and doubletons (-22.7%), rather than the formation of mOTUs composed of at least 3 sequences (-4.9%), we considered the latter effect negligible. Inspection of intra-mOTU variation did not reveal any clear fusion of different taxa together but, given the low level of completion of the reference database and the fact that about 50% of mOTUs could not be assigned lower than the phylum level, it is likely that such an event could not be detected.

The main difference in the ecological analyses between the ITS1 sub-locus and pSSU-ITS1 fragment based approaches was on the proportion of mOTUs assigned to the class level. Taxonomic affiliation was based on rdp approach against the UNITE database which is composed of sequences deprived of any SSU fragment, thus disabling any possibility to skew the similarity search. The impact of the presence of the pSSU was not on the taxonomic assignment provided per se, but on the confidence estimates that accompany each assignment. The bootstrap confidence estimation considered the overall number of eight-character sub-sequences (kmers) composing the query, and then randomly chose one-eighth of the kmers to calculate the joint probability with sequences from the reference database (Wang et al., 2007). The presence of the pSSU in the query increases the total number of kmer compared to the ITS1 sub-locus only, and as it is absent from sequences in Unite, it might lower the confidence estimates under the selected threshold of 80. In turn this might lead to an assignment to a higher taxonomic level (from Class to Phylum for instance). In the absence of a complete, high quality reference database, taxonomic assignment remains somewhat trivial. It is likely that in the presence of better matches for each query, the presence of the pSSU would not lower the bootstrap confidence estimate under the threshold value.

The method we developed to detect fungal ITS yielded a similar proportion of sequences as the ITSx software, both for the Unite fungal ITS reference database (>97% recovery) and the balsam fir endomycobiota dataset (>88%) produced for this study. Yet, each approach specifically detected a few supplementary sequences (these sequences were rarely engaged into a mOTU specific to one approach). This result was due to the distinctive principles applied to recover fungal sequences. Our approach relied on screening the sequences for the presence of a single oligonucleotide (ITS3a) located at the start of the 5.8S gene, allowing up to three mismatches in its composition. Because screening was limited to the 3' extremity, our approach was less restraining than the approach implemented in ITSx which needs to determine both the SSU and 5.8S extremities to extract the full-length ITS1 sub-locus. ITSx relies on the recognition of pre-defined patterns in the form of hidden Markov models (HMMs) (Nilsson et al., 2010; Bengtsson-Palme et al., 2013) which confers a higher plasticity (Eddy 2004), notably for the detection of more deviant rRNA genes. This was observed in the Unite database, as 35.7% of the sequences specifically detected by ITSx belonged to the Cantharellales order (most of them assigned to the Tulasnellaceae family).

Such plasticity to improve the detection of the more deviant rRNA genes was not without drawbacks. While both approaches detected false positives (i.e. non fungal ITS sequences), and despite the fact that some of them were commonly retrieved by the two methods, ITSx had a slightly higher rate of false positives than ITS3a. For instance, 18 of the 52 972 sequences detected in Unite by both methods were not of fungal origin. For the 1 060 sequences specifically detected by ITSx, 43 were non fungi, whereas just 1 of the 315 sequences only detected with fqgrep on ITS3a was not assigned to that kingdom. Therefore, while our approach might be more fungal specific, it appeared to be more restrictive as the detection of more deviant rRNA genes is impaired, whereas ITSx seemed to be more adapted for these particular sequences but to the detriment of fungal specificity as more non-fungal sequences are also extracted. The time of analysis was also impacted by the approach used. Detection with fqgrep based on ITS3a only took a couple of minutes for either the Unite fungal ITS database composed of 54 873 sequences or our balsam fir endomycobiota dataset composed of over 220 000 unique sequences, whereas analysis with ITSx required about 2h for UNITE and >14h for the balsam fir endomycobiota dataset.

It is worth emphasizing that both approaches shared the same limitation, in that the screened motifs (either ITS3a or the HMMs) were defined from fungal ITS sequences deposited in public International Nucleotidic Sequence Databases which are far from representing the complete diversity of fungi. It might imply an important selective bias as the majority of the fungi are neither described nor referenced molecularly (Yahr et

al., 2016). As of November 2016, Hawksworth & Lücking (2017) estimated to 34 878 the number of fungal species with a full binomial name (of ca. 120 000 species currently described) to be referenced in the Genbank database. With such a low level of completion and considering the recent estimate of 2.2 to 3.8 million fungal species (Hawksworth & Lücking, 2017), it is more than likely that current metabarcoding approaches can not yet recover the full extent of the fungal diversity. It has to be noted that even the combination of the two approaches (first ITS3a, then ITSx on the sequences not displaying this oligonucleotide) could not extract all sequences from UNITE. Based on binomial names, 85 species were present among the 526 sequences not detected. It is worth noting that 41 of these species were extracted (represented by other sequences). For the 44 species undetected, 41 were only represented by one sequence.

On an ecological point of view, for this particular dataset, a difference in richness, evenness and species distribution of the fungal community was observed between the root and the aerial systems. Both systems interact with different environments and might be subjected to different factors shaping the fungal composition of their associated community (Arnold 2007; Coince et al., 2014). The soil is the most important reservoir of fungal diversity (Peay et al., 2016) and a relatively stable environment where microorganisms can survive in a dormant state (Vorholt 2012), whereas the aerial parts of the plant are exposed to a more dynamic environment subjected to more important changes in temperature, radiation, and moisture (Arnold 2007; Turner et al., 2013). Biodiversity was found to be higher in the aerial parts than in the root system. We acknowledge a possible bias in our sampling design (despite a similar number a sample units for both systems, root samples came from two 1.2m fragments due to difficulties on site to ascertain the root network of the selected tree, whereas aerial samples were distributed among 12 branches and the trunk). Endosphere studies of both Agaves and Cacti species displayed similar richness between the two systems despite significantly lower raw read counts for leaf endosphere (Coleman-Derr et al., 2016) or stem endosphere (Fonseca-Garcia et al., 2016).

The community inhabiting needles appeared to be more similar to the community inhabiting the root system than to that recovered from the aerial system. This might be linked with a functional purpose. It has been shown that some foliar endophytes have the ability to degrade litter and that the colonization of the foliage as endophytes gives them a position of precursor in the successional cohort of microorganisms involved in litter decomposition (Osono 2006). Root associated fungal endophytes are usually detected in soils (Fonseca-Garcia et al., 2016; Peay et al, 2016) and many are known to be successful opportunistic early decomposers of litter (Peay et al., 2016), suggesting a possible overlap between foliar endophytes and root endophytes. Our

study, however, was not designed with this goal in mind, and further work on this topic should include separation of needle samples according to age, and inclusion of litter samples, soil samples from immediate surface horizon to the rhizosphere, and root samples.

1.7 Conclusion

Generalisation of culture-independent methods have revolutionised the study of plant fungal endomycobiota with metabarcoding and have revealed a magnitude of biodiversity superior to that observed with culturing approach. Yet, HTS techniques are known to display methodological biases which usually lead to an overestimation of biodiversity notably due to the conservation of erroneous sequences despite state of the art data analysis. This usually results in the formation of a cohort of singletons and doubletons during the clustering into mOTUs. In the absence of a complete reference ITS database, it is impossible to detect these spurious sequences, which thus remain in the dataset during the analysis. In this study, we attempted to address indirectly this problem in tempering rare events by targeting both chimera detection and mOTU clustering based on the selection of a wider barcode sub-fragment (pSSU-ITS1). Retention of the original 5' extremity of the fungal amplicons led to a 30-fold increase in the detection of putative chimeras. The lower variability of the pSSU also appeared to buffer sequencing errors within the ITS1 sub-locus and reduced the formation of singletons by 38% and of doubletons by 22.7%. This approach also permitted to evaluate conservatively the endomycobiota of a single tree, a non negligible variable for adapting sequencing depth and sampling strategy. The development and globalisation of HTS techniques for metabarcoding have demonstrated the urge to develop a more complete reference database. It paradoxically makes an argument for more culture- and specimen-dependent sequencing to infer the biological reality of the sequences generated and allow discriminating rare taxa from spurious ones. The completion of such a database will not only require referencing more species, but also more individuals per species in order to determine intra-species variation. It seems likely that until a complete reference database is developed, metabarcoding studies will not be able to take full advantage of the mass of data produced.

1.8 Acknowledgements

We kindly thank Julie Boudreau from McGill University and Genome Quebec Innovation Centre (Montreal, Qc, Canada) for technical assistance on 454 pyrosequencing. We are indebted to Pr Elizabeth A. Arnold (University of Arizona, Tucson, Az, USA) who provided helpful comments and invaluable suggestions to improve the overall quality of this manuscript. We are grateful to Dr Jana U'Ren (University of Arizona, Tucson, Az, USA) for useful discussions on pyrosequencing data processing and her great help on ecological analysis. We also thank Patrick Gagné (Université Laval, Qc, Canada) for feedbacks on bioinformatic analysis.

1.9 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Amend, A. S., Seifert, K. A., & Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count?. *Molecular ecology*, *19*(24), 5555-5565.

Arnold, A. E. (2007). Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal biology reviews*, *21*(2-3), 51-66.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... & Amend, A. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution*, *4*(10), 914-919.

Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R. H., Uroz, S., & Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, *184*(2), 449-456.

Chazdon, R. L., Chao, A., Colwell, R. K., Lin, S. Y., Norden, N., Letcher, S. G., ... & Arroyo, J. P. (2011). A novel statistical method for classifying habitat generalists and specialists. *Ecology*, *92*(6), 1332-1343.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral ecology*, *18*(1), 117-143.

Coince, A., Cordier, T., Lengellé, J., Defossez, E., Vacher, C., Robin, C., ... & Marçais, B. (2014). Leaf and root-associated fungal assemblages do not follow similar elevational diversity patterns. *PloS one*, *9*(6), e100668.

Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., ... & Tringe, S. G. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytologist*, *209*(2), 798-811.

Colwell, R. K. (2009). Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology*, 257-263.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... & Pfrender, M. E. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular ecology*, *26*(21), 5872-5895.

Eddy, S. R. (2004). What is a hidden Markov model?. *Nature biotechnology*, *22*(10), 1315.

Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 42-58.

Fonseca-García, C., Coleman-Derr, D., Garrido, E., Visel, A., Tringe, S. G., & Partida-Martínez, L. P. (2016). The cacti microbiome: interplay between habitat-filtering and host-specificity. *Frontiers in microbiology*, *7*, 150.

Fröhlich-Nowoisky, J., Pickersgill, D. A., Després, V. R., & Pöschl, U. (2009). High diversity of fungi in air particulate matter. *Proceedings of the National Academy of Sciences*, *106*(31), 12814-12819.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*, *2*(2), 113-118.

Hall, T. A. (1999, January). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Vol. 41, No. 41, pp. 95-98). [London]: Information Retrieval Ltd., c1979-c2000..

Hamady, M., & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome research*, *19*(7), 1141-1152.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis Palaeontol. Electronica 4: 1–9.

Hawksworth, D. L., & Lücking, R. (2017). Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiology spectrum*, *5*(4).

Harrell, Frank E. (2008). "Hmisc: Harrell Miscellaneous." R package version 3.4-4.

Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, *12*(7), 1889-1898.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, *17*(3), 377-386.

Indraniel et al.. (2016). fqgrep: v0.4.4. Zenodo. 10.5281/zenodo.45105

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... & Larsen, T. H. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters*, *16*(10), 1245-1257.

Jost, L. (2010). The relation between evenness and diversity. *Diversity*, *2*(2), 207-232.

Jumpponen, A., & Jones, K. L. (2009). Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate Quercus macrocarpa phyllosphere. *New Phytologist*, *184*(2), 438-448.

Kõljalg, U., Larsson, K. H., Abarenkov, K., Nilsson, R. H., Alexander, I. J., Eberhardt, U., ... & Pennanen, T. (2005). UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, *166*(3), 1063-1068.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. *bioinformatics*, *23*(21), 2947-2948.

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076-2081.

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., ... & Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers–a user's guide. *New Phytologist*, *199*(1), 288-299.

Magurran, A.E. (2004). Measuring biological diversity. Blackwell, Oxford, United Kingdom.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... & Dewell, S. B. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376.

Nilsson, R. H., Kristiansson, E., Ryberg, M., Hallenberg, N., & Larsson, K. H. (2008). Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary bioinformatics*, *4*, EBO-S653.

Nilsson, R. H., Veldre, V., Hartmann, M., Unterseher, M., Amend, A., Bergsten, J., ... & Abarenkov, K. (2010). An open source software package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology. *Fungal Ecology*, *3*(4), 284-287.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, and M. H. H. Stevens. (2007). Vegan: Community ecology package. R package, version 1.8 – 8

Osono, T. (2006). Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. *Canadian Journal of Microbiology*, *52*(8), 701-716.

Peay, K. G., Kennedy, P. G., & Talbot, J. M. (2016). Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology*, *14*(7), 434.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, *13*, 131-144.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., ... & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, *6*(9), 639.

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, *12*(1), 38.

R Core Team (2015). R: A language and environment for statistical computing.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, *75*(23), 7537-7541.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS one*, *6*(12), e27310.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., ... & Miller, A. N. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241-6246.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-2504.

Smith, E. P., & van Belle, G. (1984). Nonparametric estimation of species richness. *Biometrics*, 119-129.

Stecher, A., Neuhaus, S., Lange, B., Frickenhaus, S., Beszteri, B., Kroth, P. G., & Valentin, K. (2016). rRNA and rDNA based assessment of sea ice protist biodiversity from the central Arctic Ocean. *European journal of phycology*, *51*(1), 31-46.

Stefani, F. O. P., & Bérubé, J. A. (2006). Biodiversity of foliar fungal endophytes in white spruce (Picea glauca) from southern Québec. *Botany*, *84*(5), 777-790.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M. L., McKendree, W., & Farmerie, W. (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic acids research*, *37*(10), e76-e76.

Taguchi, Y. H., & Oono, Y. (2004). Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics*, *21*(6), 730-740.

Tedersoo, L., Nilsson, R. H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., ... & Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New phytologist*, *188*(1), 291-301.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4-18.

Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome. *Genome biology*, *14*(6), 209.

U'Ren, J. M., Riddle, J. M., Monacell, J. T., Carbone, I., Miadlikowska, J., & Arnold, A. E. (2014). Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Molecular Ecology Resources*, *14*(5), 1032-1048.

Unterseher, M., Jumpponen, A. R. I., Oepik, M., Tedersoo, L., Moora, M., Dormann, C. F., & Schnittler, M. (2011). Species abundance distributions and richness estimations in fungal metagenomics–lessons learned from community ecology. *Molecular Ecology*, *20*(2), 275-285.

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, *10*(12), 828.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, *73*(16), 5261-5267.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. L. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*(1), 315-322.

Yahr, R., Schoch, C. L., & Dentinger, B. T. (2016). Scaling up discovery of hidden diversity in fungi: impacts of barcoding approaches. *Phil. Trans. R. Soc. B*, *371*(1702), 20150336.

Yamamoto, N., Bibby, K., Qian, J., Hospodsky, D., Rismani-Yazdi, H., Nazaroff, W. W., & Peccia, J. (2012). Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *The ISME journal*, *6*(10), 1801.

Yang, R. H., Su, J. H., Shang, J. J., Wu, Y. Y., Li, Y., Bao, D. P., & Yao, Y. J. (2018). Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PloS one*, *13*(10), e0206428.

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., ... & Sønstebø, J. H. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, *21*(15), 3647-3655.

Figure 1.1: Organization of ribosomal genes and target regions of 5' extremity oligonucleotide for 5.8S. Colours in the close-up of the beginning of the 5.8S indicate the frequency of the most common nucleotide among the aligned fungal sequences.

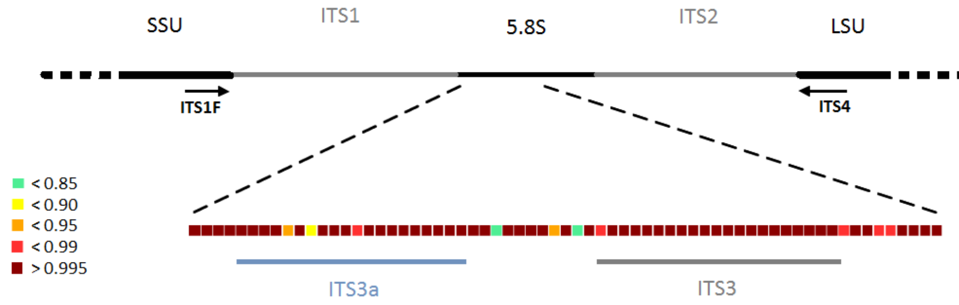


Figure 1.2: Species accumulation curves (solid lines) and bootstrap estimates (long-dashed lines) of species richness for ITS1 sub-locus based standard analysis (red) and pSSU-ITS1 fragment based analysis with the ITS3a-ITSx combination (blue). Dotted lines represent the 95% confidence intervals.

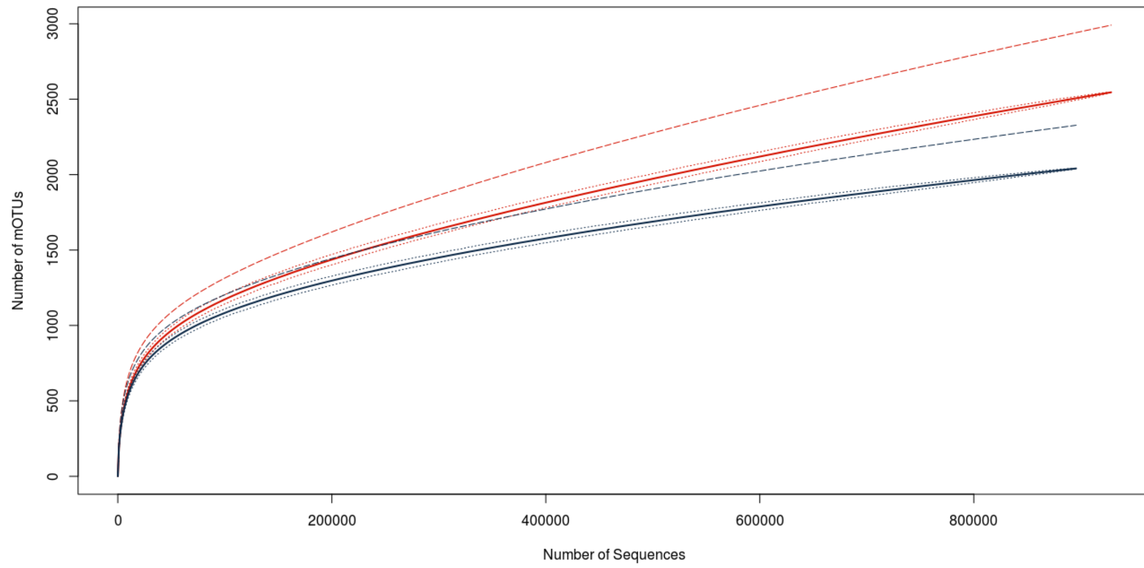


Figure 1.3: Relative abundance plots of fungal endophyte communities associated with the whole tree (A) by plant system based on the ITS1 sub-locus standard analysis (B) and the pSSU-ITS1 fragment (C)

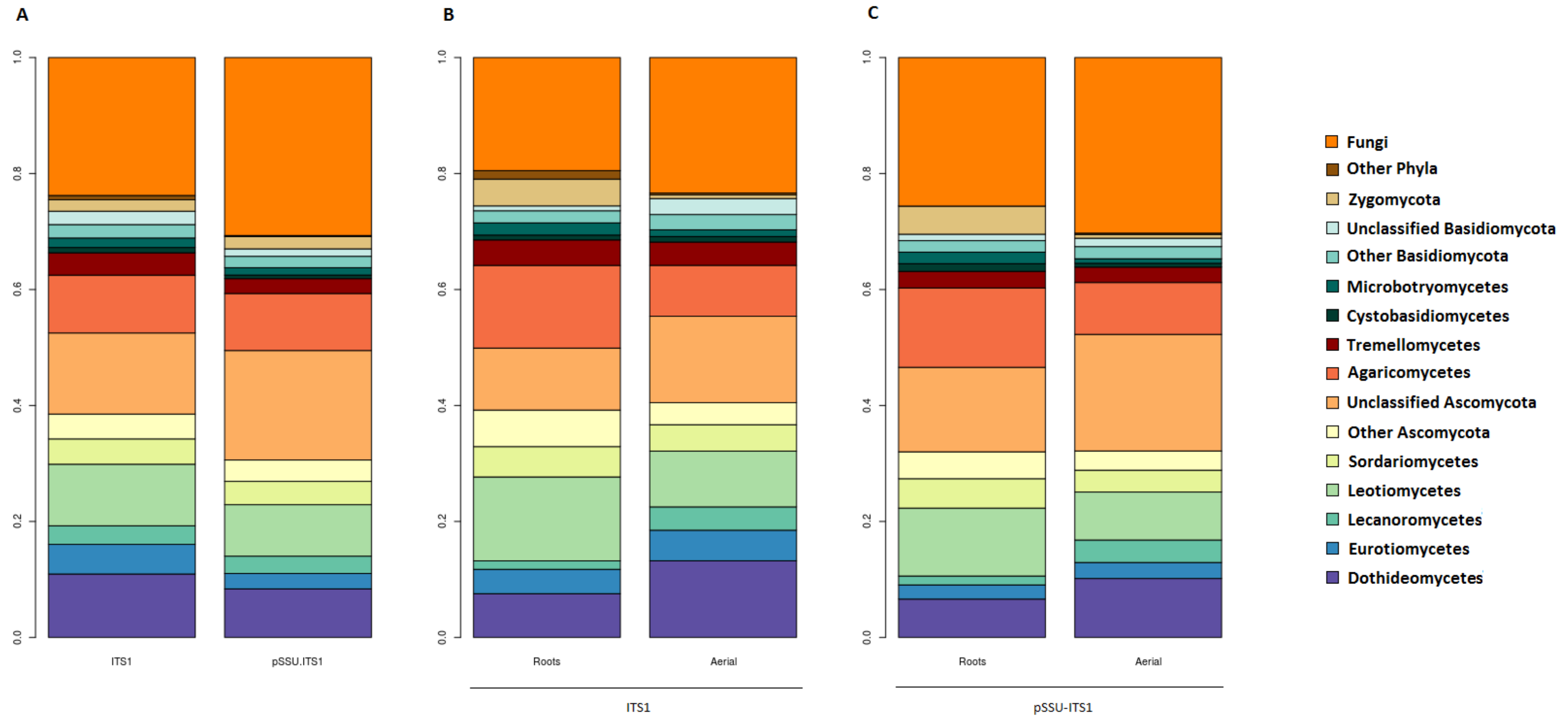


Figure 1.4: Results of diversity (A, B) and evenness (C, D) analyses results between the root and the aerial systems based on the ITS1 sub-locus standard analysis (A,C) and the pSSU-ITS1 fragment (B,D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis tests or after ANOVA test with $p \leq 0.05$.

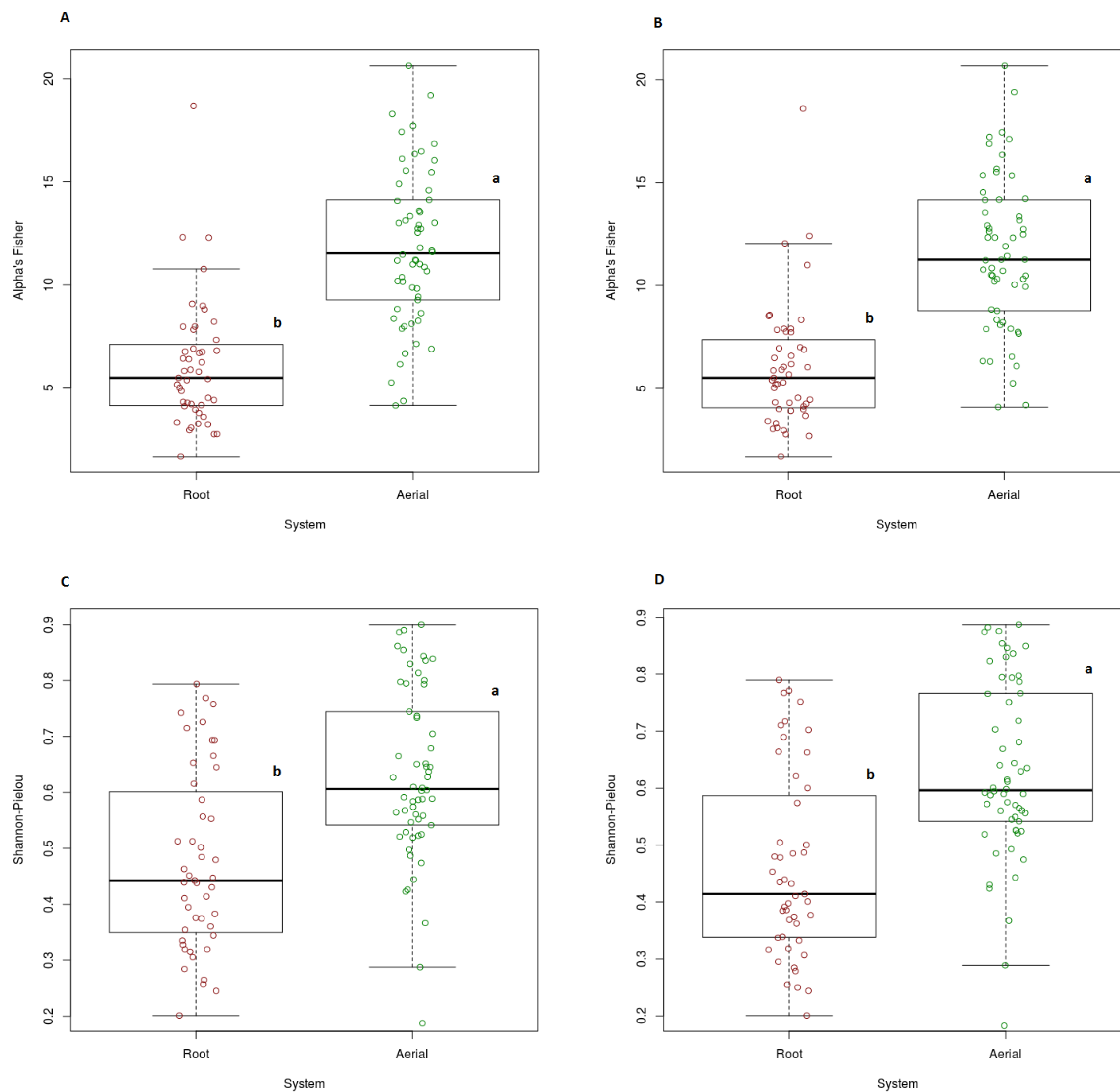


Figure 1.5: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition between the root system (brown circles) and the aerial system (needle samples: open green triangles, remaining aerial samples: closed green triangles) for the standard analysis based on the ITS1 sub-locus (A) and the pSSU-ITS1 fragment (B). One-way analysis of similarity (ANOSIM) indicates significant differences ($P < 0.05$) in community composition between the root and aerial systems. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM.

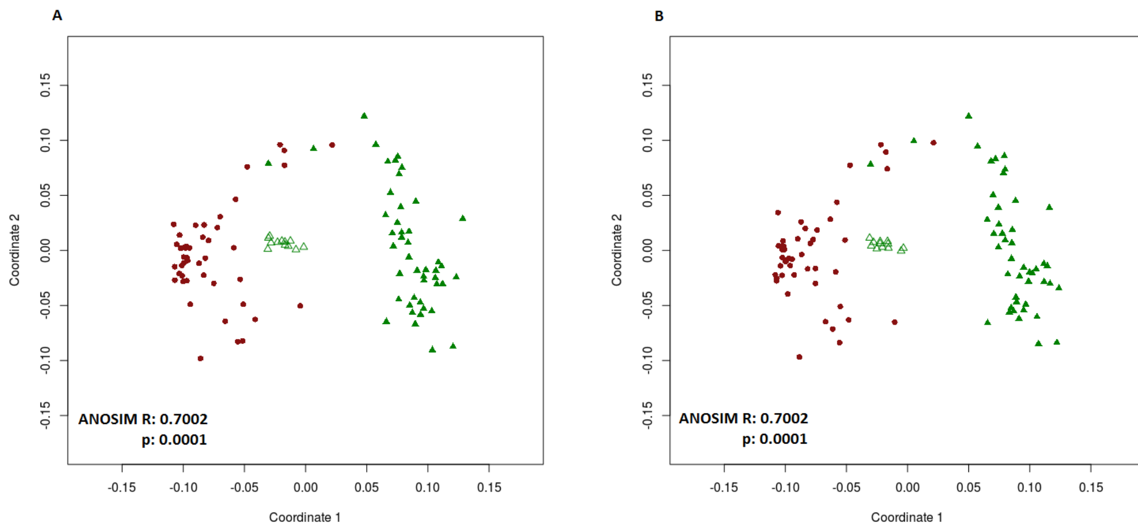


Figure 1.6: mOTUs co-occurrence networks displaying a root module and an aerial module for the standard analysis based on the ITS1 sub-locus (A) and the pSSU-ITS1 fragment (B). Brown nodes represent root specialists, light blue aerial specialists, and light green needles specialists. Dark blue nodes represent aerial-needle affinity, and dark green root-needle affinity. Red nodes represent generalist taxa, and grey non habitat preference assigned. Grey edges were used when at least one of the nodes was assigned only to the kingdom level, blue edge only to Ascomycota, plum only to Basidiomycota, orange when both nodes were assigned to the same class, and red when both nodes were assigned to different classes. Pairwise Spearman's rank correlations were considered if $\rho > 0.6$, and P-value < 0.01.

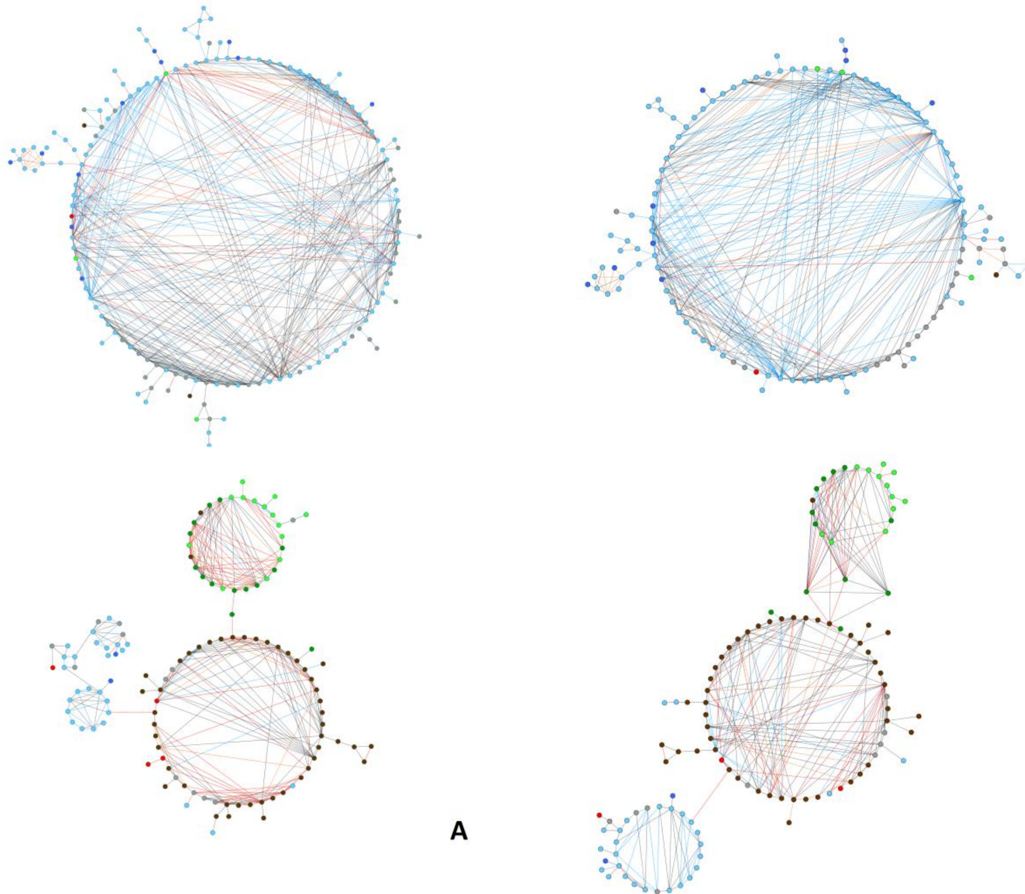


Table 1.1: Influence of ITS extraction methods, fragment used for chimera detection, and for clustering

Methods	ITSx ITS1 ITS1	ITSx pSSU-ITS1 ITS1	ITSx pSSU-ITS1 pSSU-ITS1	ITS3a fqgrep	ITS3a-ITSx
Raw	1 069 756				
Denoised	1 061 579				
Trimmed	1 039 365				
Program for extraction	ITSx			fqgrep	
Extracted	941 825			943 539	945 131
Fragment for chimeras	ITS1	pSSU-ITS1		pSSU-ITS1	pSSU-ITS1
Number of chimeras	1 269	37 278		37 392	37 406
Fragment for clustering	ITS1	ITS1	pSSU-ITS1	pSSU-ITS1	pSSU-ITS1
Conserved sequences	940 556	904 547		906 147	907 725
Fungal sequences	928 348	892 625		894 366	895 615
mOTUs	2 546	2 542	2 013	1 961	2 041
>2 sequences	1 224	1 225	1 165	1 144	1 177
Doubletons	207	207	160	147	164
Singletons	1 115	1 110	688	670	700
Extrapolation	3 208	3 204	2 502	2 433	2 536
SDE Extrapolaton	83	83	72	68	73
Mean Length (bp)	185.109	185.456	234.683	238.081	238.137

Table 1.2: Kruskal-Wallis tests of the average relative abundance of fungal phylum and classes between the root and aerial systems for the standard analysis based on the ITS1 sub-locus. Significant differences (p ≤ 0.05) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA test instead of Kruskal-Wallis test

Taxonomic assignment	Degrees freedom	Number of samples	X2	P	Root system	Aerial system
Ascomycota *	1	105	86.23	2.861x10-15	b	a
Basdiomycota *	1	105	36.755	2.23x10-8	a	b
Zygomycota	1	105	26.021	3.378x10-7	a	b
Chytridiomycota	1	105	2.4781	0.1154	a	a
Glomeromycota	1	105	0.62834	0.428	a	a
Rozellomycota	1	105	7.7695	5.314x10-3	a	b
Fungi	1	105	0.52099	0.4704	a	a
Dothideomycetes	1	105	42.785	6.11x10-11	b	a
Eurotiomycetes	1	105	42.979	5.533x10-11	b	a
Lecanoromycetes	1	105	52.704	3.877x10-13	b	a
Leotiomycetes	1	105	25.855	3.681x10-7	a	b
Sordariomycetes	1	105	71.954x10-3	0.7885	a	a
Pezizomycetes	1	105	46.747	8.076x10-12	a	b
Saccharomycetes	1	105	13.857	1.972x10-4	a	b
Taphrinomycetes	1	105	6.0384	0.014	b	a
Archaeorhizomycetes	1	105	36.585	1.461x10-9	a	b
Orbiliomycetes	1	105	5.9238	0.01494	b	a
Pseudeurotiaceae	1	105	4.6089	0.03181	b	a
Other Ascomycota	1	105	10.672	1.088x10-3	b	a
Unclassified Ascomycota	1	105	39.567	3.17x10-10	b	a
Agaricomycetes	1	105	46.938	7.326x10-12	a	b
Tremellomycetes	1	105	0.7545	0.3851	a	a
Cystobasidiomycetes	1	105	14.45	1.439x10-4	b	a
Microbotryomycetes	1	105	3.7031	0.05431	a	a
Pucciniomycetes	1	105	0.60409	0.437	a	a
Exobasidiomycetes	1	105	52.642	4.003x10-13	b	a
Malasseziales	1	105	5.9237x10-3	0.9387	a	a
Agaricostilbomycetes	1	105	5.1019	0.0239	b	a
Ustilaginomycetes	1	105	1.234	0.2666	a	a
Unclassified Basidiomycota	1	105	16.532	4.784x10-5	b	a
Mortierellales	1	105	25.098	5.45x10-7	a	b
Mucorales	1	105	1.3303	0.2487	a	a

Table 1.3: Kruskal-Wallis tests of the average relative abundance of fungal phylum and classes between the root and aerial systems for the pSSU-ITS1 fragment based analysis. Significant differences ($p \leq 0.05$) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA test instead of Kruskal-Wallis test

Taxonomic assignment	Degrees freedom	Number of samples	X2	P	Root system	Aerial system
Ascomycota	1	105	118.13	<2.2x10-6	b	a
Basdiomycota	1	105	38.799	4.698x10-10	a	b
Zygomycota	1	105	25.94	3.522x10-7	a	b
Chytridiomycota	1	105	2.4781	0.1154	a	a
Fungi	1	105	3.2446	0.0717	a	a
Dothideomycetes	1	105	22.748	1.847x10-6	b	a
Eurotiomycetes	1	105	18.41	1.782x10-5	b	a
Lecanoromycetes	1	105	52.75	3.788x10-13	b	a
Leotiomycetes *	1	105	13.312	4.158x10-4	a	b
Sordariomycetes	1	105	0.3859	0.5345	a	a
Pezizomycetes	1	105	47.961	4.348x10-12	a	b
Saccharomycetes	1	105	9.3767	2.198x10-3	a	b
Taphrinomycetes	1	105	5.9592	0.0146	b	a
Archaeorhizomycetes	1	105	2.4918	0.1144	a	a
Orbiliomycetes	1	105	7.0691	7.842x10-3	b	a
Pseudeurotiaceae	1	105	5.4402	0.0197	b	a
Other Ascomycota	1	105	10.673	1.087x10-3	b	a
Unclassified Ascomycota	1	105	71.827	1.773x10-13	b	a
Agaricomycetes	1	105	45.003	1.967x10-11	a	b
Tremellomycetes	1	105	0.3665	0.5449	a	a
Cystobasidiomycetes	1	105	1.9045	0.1676	a	a
Microbotryomycetes	1	105	3.4866	0.0619	a	a
Pucciniomycetes	1	105	0.0267	0.8703	a	a
Exobasidiomycetes	1	105	52.693	3.901x10-13	b	a
Malasseziales	1	105	0.0298	0.863	a	a
Agaricostilbomycetes	1	105	1.6363	0.2008	a	a
Ustilaginomycetes	1	105	1.234	0.2666	a	a
Unclassified Basidiomycota	1	105	3.4983	0.0614	a	a
Mortierellales	1	105	25.015	5.689x10-7	a	b
Mucorales	1	105	1.3303	0.2487	a	a

Table S1.1: Combination of CLAM tests results to attribute categories to mOTUs present in needle samples

Categories	Root system vs aerial system	Root samples vs non foliar aerial samples	Root system vs needle samples	Non foliar aerial samples vs needle samples
Aerial-Needle Affinity	Specialist Aerial	Generalist	Specialist Needle	Generalist
Aerial-Needle Affinity	Specialist Aerial	Specialist Aerial	Generalist	Generalist
Aerial-Needle Affinity	Specialist Aerial	Specialist Aerial	Specialist Needle	Generalist
Aerial-Needle Affinity	Specialist Aerial	Specialist Aerial	Too Rare	Generalist
Aerial-Needle Affinity	Specialist Aerial	Too Rare	Specialist Needle	Generalist
Generalist	Generalist	Generalist	Generalist	Generalist
Generalist	Generalist	Generalist	Generalist	Specialist Aerial
Generalist	Generalist	Generalist	Generalist	Too Rare
Generalist	Generalist	Generalist	Specialist Needle	Generalist
Generalist	Generalist	Generalist	Specialist Roots	Specialist Aerial
Generalist	Generalist	Generalist	Too Rare	Generalist
Generalist	Generalist	Specialist Aerial	Specialist Roots	Specialist Aerial
Generalist	Generalist	Specialist Roots	Generalist	Generalist
Generalist	Generalist	Too Rare	Too Rare	Too Rare
Generalist	Generalist	Generalist	Generalist	Specialist Needle
Generalist	Generalist	Generalist	Specialist Needle	Specialist Needle
Generalist	Generalist	Generalist	Specialist Roots	Generalist
Generalist	Generalist	Generalist	Specialist Roots	Too Rare
Generalist	Generalist	Generalist	Too Rare	Specialist Aerial
Generalist	Generalist	Generalist	Too Rare	Too Rare
Generalist	Generalist	Specialist Aerial	Too Rare	Specialist Aerial
Root-Needle Affinity	Generalist	Specialist Roots	Generalist	Specialist Needle
Root-Needle Affinity	Generalist	Specialist Roots	Generalist	Too Rare
Root-Needle Affinity	Specialist Roots	Specialist Roots	Generalist	Generalist
Root-Needle Affinity	Specialist Roots	Specialist Roots	Generalist	Specialist Needle
Root-Needle Affinity	Specialist Roots	Specialist Roots	Generalist	Too Rare
Specialist Aerial	Specialist Aerial	Specialist Aerial	Generalist	Specialist Aerial
Specialist Aerial	Specialist Aerial	Specialist Aerial	Specialist Needle	Specialist Aerial
Specialist Aerial	Specialist Aerial	Specialist Aerial	Specialist Roots	Specialist Aerial
Specialist Aerial	Specialist Aerial	Specialist Aerial	Too Rare	Specialist Aerial
Specialist Aerial	Specialist Aerial	Specialist Aerial	Too Rare	Too Rare
Specialist Aerial	Specialist Aerial	Specialist Aerial		Specialist Aerial
Specialist Aerial	Specialist Aerial	Specialist Aerial		Too Rare

Specialist Needle	Generalist	Specialist Roots	Specialist Needle	Specialist Needle
Specialist Needle	Generalist	Too Rare	Specialist Needle	Specialist Needle
Specialist Needle	Specialist Aerial	Generalist	Specialist Needle	Specialist Needle
Specialist Needle	Specialist Aerial	Specialist Aerial	Specialist Needle	Specialist Needle
Specialist Needle	Specialist Aerial	Too Rare	Specialist Needle	Specialist Needle
Specialist Needle	Specialist Aerial		Specialist Needle	Specialist Needle
Specialist Roots	Specialist Roots	Generalist	Specialist Roots	Generalist
Specialist Roots	Specialist Roots	Generalist	Specialist Roots	Specialist Aerial
Specialist Roots	Specialist Roots	Specialist Roots	Specialist Roots	Generalist
Specialist Roots	Specialist Roots	Specialist Roots	Specialist Roots	Specialist Needle
Specialist Roots	Specialist Roots	Specialist Roots	Specialist Roots	Too Rare
Specialist Roots	Specialist Roots	Generalist	Specialist Roots	Too Rare
Specialist Roots	Specialist Roots	Specialist Roots	Specialist Roots	Specialist Aerial
Specialist Roots	Specialist Roots	Specialist Roots	Specialist Roots	
Specialist Roots	Specialist Roots	Specialist Roots	Too Rare	Too Rare
Specialist Roots	Specialist Roots	Specialist Roots	Too Rare	
Specialist Roots	Specialist Roots	Too Rare	Too Rare	
Too Rare	Too Rare	Too Rare	Generalist	Too Rare
Too Rare	Too Rare	Too Rare	Too Rare	Too Rare
Too Rare	Too Rare	Specialist Aerial		Too Rare

Chapter 2 Diversity, community composition and structure of the endomycobiota of balsam fir trees (*Abies balsamea* [L.] Mill) as a function of tissue type

Julien P. Ponchart[1,§], Louis Bernier[1,2], Jean A. Bérubé[3].

[1] Centre for Forest Research, Université Laval, Québec, Qc, Canada
[2] Institute for System and Integrative Biology (IBIS), Université Laval, Québec, Qc, Canada
[3] Laurentian Forest Centre, Canadian Forest Service, Natural Resources Canada, Québec, Qc, Canada
[§] Corresponding author: Julien P. Ponchart, email: julponchart@gmail.com

2.1 Résumé

Le but de cette étude était d'identifier l'endomycobiote associé aux branches de sapin baumier et de comparer sa diversité et sa composition en fonction du type de tissu. À partir de 256 échantillons provenant de 16 individus en bonne santé récoltés à la Forêt Montmorency, nous avons montré que la diversité de l'endomycobiote diffère en fonction du type de tissu. Les diversités observées dans les échantillons d'aiguille et du bois étaient supérieure à celle des bourgeons et de l'écorce. Les différents types de tissus présentent des communautés distinctes, ce qui empêche la définition d'une communauté de base partagée par tous les types de tissus. Nous préconisons donc d'inclure l'ensemble des différents types de tissu afin de mieux comprendre les interactions entre plante et microorganismes associés. En conclusion, nos résultats suggèrent que l'endomycobiote aérien associé aux branches de sapins baumier pourrait être fractionné en communautés distinctes selon les types de tissus.

2.2 Abstract

Understanding tissue-specificity of the plant microbiota is critical for understanding fungal diversity and functional traits. Yet, when considering the aerial system, most of our knowledge comes from the foliar community and the conclusions tend to serve for the whole system. The aim of this study was to identify the aerial endomycobiota associated with *Abies balsamea* branches and to compare its diversity and composition as a function of tissue type at a regional scale. Based on 256 samples from 16 healthy looking individuals of balsam fir trees collected at 2 locations within the Forêt Montmorency research forest north of Quebec City, we first show that the diversity of the endomycobiota differs as a function of tissue type. Communities associated with needle and wood were more diverse than bud and bark communities. Then, we demonstrate that the different tissue types displayed distinct communities which impaired the definition of a precise core community shared by all tissue types. We thus advocate refining the scope of resolution to the tissue type level for a better understanding of the interactions between the plant and its associated microorganisms. Together, our results suggest that the aerial endomycobiota of balsam fir trees might be fractioned in distinct communities depending on the tissue types.

2.3 Introduction

The plant microbiota, i.e. the microorganisms comprising bacteria, fungi, protists, and viruses associated with the interior and surface of plants, plays key roles in regulating host metabolism and physiology, enhancing nutrient acquisition, and providing resistance to antagonist and abiotic stressors (Friesen et al., 2011; Bulgarelli et al., 2013; Hardoim et al., 2015). As such, plant-associated microbes can shape plant fitness and thus influence plant community structure and ecosystem functioning (Porras-Alfaro & Bayman 2011; Rout 2014; Laforest-Lapointe et al., 2017). The importance of this microbiota is such that plants are now viewed as multiorganismal networks rather than autonomous entities (Hacquard & Schadt 2015; Bordenstein & Theis, 2015; Vandenkoonhuyse et al., 2015). This holobiont concept, consisting of the host plant and the myriad of microorganisms which confer additional functional traits was defined and extended to the hologenome (Rosenberg 2007; Bordenstein & Theis, 2015). This concept considers the combination of the host genome to those of these microorganisms as the unit of natural selection in evolution. Briefly, it relies on the fact that the potential for adaptability of the hologenome exceeds that of the plant genome on its own, mostly due to the rapidity with which microorganisms adapt to environmental changes (Rosenberg et al., 2007; Rout 2014; Bordenstein & Theis, 2015).

Several aspects of plant microbiota studies have benefited from technological advancements: biodiversity studies, both with metabarcoding (either DNA- or RNA-based) and metagenomics approaches (Jumpponen & Jones 2009; U'Ren et al., 2014; Bulgarelli et al., 2015); but also interactions studies between the plant and its microorganisms or between different members of this microbial community with metatranscriptomics, metaproteomics, and metabolomics (Delmotte et al., 2009; Larsen et al., 2011; Knief et al., 2012; Schenk et al., 2012). The development of microfluidic technology to be coupled with these techniques also promises further progress in the understanding of the plant microbiome (Stanley & van der Heijden, 2017). While diversity studies might appear more trivial, they remain of primordial importance to identify which species compose the microbiota, which factors influence their distribution patterns, and which species might actually interact with each other.

Fungal endophytes are microorganisms living within a plant without causing any apparent damage or symptom to their host (Petrini 1991; Wilson 1995; Schultz & Boyle 2005) and constitute thus part of the plant microbiota. They are a highly diverse group of fungi composed of as many as 0.5 to 1 million species (Sieber 2007; Bills 1996) therefore constituting a possibly important portion of the hidden fungal diversity. Furthermore,

endophytism is also widespread among plants as every plant sampled was found to harbour fungal endophytes, and it is generally considered that they are virtually found in every plant (Petrini 1986; Saikkonen et al., 1998; Rodriguez et al., 2009). Yet questions regarding the abundance, distribution and ecology of fungal endophytes are still unresolved and factors shaping the endomycobiota composition are largely unknown (Persoh 2013). Decades of endophytology studies have shown that endophytic fungi are far from being distributed randomly within space and time (Unterseher et al., 2007). Whereas pioneering works on endophytology surveyed the different tissue types of tree branches (Petrini & Fisher 1988; Fisher & Petrini 1990; Fisher et al., 1993), recent studies have focused on the leaf-associated community (Zimmerman & Vitousek 2012; U'Ren & Arnold 2016; Oono et al., 2017).

Foliar based endomycobiota studies have notably demonstrated that recruitment mechanisms of the fungal endophyte community by the plant display a higher complexity that the passive colonization generally described. Fungal endophytes show specificity to their host species (Hata & Futai, 1996; Joshee et al., 2009) and sometimes even to a particular genotype of these host species (Ahlholm et al., 2002; Balint et al., 2013; Rajala et al., 2013). Variations of the environmental conditions, even at a micro-scale such as light regime within the canopy, influence the composition of the endomycobiota within the same tree (Unterseher et al, 2007). Endophytic- and epiphytic communities also differ both in diversity and composition despite their immediate proximity (Santamaria & Bayman, 2005; Osono 2008). With such entanglements within the same tissue, limiting fungal endophyte community studies to a single tissue might appear as an opportunity to develop gradually a better understanding of endophytism. Yet there is a lack of representativity of the different aerial tissues as most endophytology studies focus almost exclusively on the leaf-associated community. While foliar endophyte based studies are justified by the fact that leaves represent a particular biodiversity hotspot, notably for tropical trees (Arnold & Lutzoni 2007; Redford et al., 2010; Rastogi et al., 2013), it has been suggested that sapwood might harbour a similar, if not greater, diversity than leaves (Gazis & Chaverri 2010). Gazis et al. (2012) even discovered a new fungal taxonomic class, the Xylonomycetes, in the sapwood of *Hevea brasiliensis*. Moreover, the biochemical properties of the aerial tissues differ markedly, suggesting that they might represent different ecological niches harbouring different communities. Previous studies have also established that some endophytic fungi show a certain degree of tissue specificity (Carroll et al., 1977; Fisher & Petrini 1988). Including all the aerial tissues in endomycobiota studies might allow a better comprehension of the mechanisms involved in the recruitment of the fungal endophyte community but also to recover a more extensive part of their diversity.

The aim of this study was thus to investigate the aerial endomycobiota of balsam fir (*Abies balsamea* (L.) Mill.), and to compare the biodiversity and community composition of each of the different branch tissue types. To this end, needle, buds, bark and wood were sampled and their fungal endophytes identified by using High Throughput Sequencing (HTS) techniques. We hypothesized that the overall diversity harboured by branches might exceed the one recovered from needles only; and that the aerial endomycobiota, rather than representing a single entity, might at least partially be fractioned into distinct communities according to tissue type. We also investigated whether a core community of ubiquitous fungal endophytes was present among the different tissue types.

2.4 Materials and Methods

2.4.1 Sample preparation, DNA extraction, PCR amplification, and sequencing

Sampling was conducted at the end of August 2012 at the Montmorency research forest (about 70km north of Quebec City) which is located at the southern portion of the North American boreal zone (Brandt 2009). Balsam fir is the main tree species in this forest. Samples were collected from four different stands, two near the Laverdière site and distant of 150m from each other (Lav01 & Lav02), and two near La Chute and separated by 400m (LaC01 & LaC02). The Laverdière and La Chute sites are distant of about 5km from each other (Table 2.1). At each stand, four trees which did not display any visible damage or symptom of disease were selected and for each tree, four branches were collected (at 2 and 3m from the ground and from two opposite directions), resulting overall in 16 trees and 64 branches sampled. From each branch we collected 30 symptomless 2-year-old needles and 8 buds. Branch sampling was completed by collecting a 5 cm segment of bark and wood for each of the four most recent growth period.

Samples were surface sterilized through a succession of six baths following Stefani & Bérubé (2006). The first three baths (1.2% sodium hypochlorite for 2 minutes, 76% ethanol for 1 minute, and 19% ethanol solution for 1 minute) were followed by three distilled water baths to rinse the samples. Needles were cut into 1 mm segments under sterile conditions. Buds were ground into a fine powder using a sterile mortar with a pestle and liquid nitrogen. Branch segments were separated into bark and wood and then ground into a fine powder in a sterile mortar with a pestle and liquid nitrogen. From each branch, a total of 100 mg of needles, 30 mg of buds, 60 mg of bark, and 100 mg of wood were used as input for DNA extraction performed with DNeasy®

Plant kit (Qiagen GmbH, Hildren, Germany). Samples were ground twice in a Mixer Mill Retsch MM300 (Qiagen) for 2 min at 30Hz using a sterilized 3mm tungsten carbide bead (Qiagen). After addition of 400µL of AP1 extraction buffer (500µL for the wood samples due to higher absorbance), 2µL of RNase A and a tip of Dx antifoaming, samples were ground a third time for 2 min at 15Hz. DNA was then extracted following the manufacturer's instructions but eluted with 75µL of buffer in two steps (50µL then 25µL).

The ITS region was amplified with modified primer set ITS1-F (Gardes & Bruns, 1993) and ITS4 (White et al., 1990). ITS1-F featured the adaptor A and a 10bp Roche Multiplex IDentifier barcode (MID, allowing post sequence analysis), and ITS4 carried the adaptor B. As needle samples yielded low concentrations post PCR, possibly due to the presence of inhibitors (data not shown), the genomic DNA of these samples was purified using the Agencourt Ampure XP magnetic clean-up system (Beckman Coulter, Brea, CA, USA) with a volume ratio of 1.8:1 of magnetic beads for genomic DNA. All samples were amplified in triplicate (using the same MID). The amplification reaction contained 25 µg of Bovine Serum Albumine (BSA, Sigma-Aldrich, St Louis, MO, USA), 1.25 X PCR buffer (Invitrogen, Carlsbad, CA, USA), 2.5 mM of $MgCl_2$, 250 µM of each dNTP, 400 nM of each of the forward and reverse primers, 1 unit of Platinum Taq DNA polymerase (Invitrogen), and 1 µL of template DNA. The PCR parameters consisted of an initial denaturation at 94 ˚C for 3 min, then 32 cycles of denaturation at 94 ˚C for 30 sec, annealing at 52 ˚C for 30 sec and extension at 72 ˚C for 1 min, followed by a final extension step at 72 ˚C for 10 min. Amplifications were conducted both in MJ Research PTC200 and BioRad T100 thermocyclers.

Triplicates were pooled and purified using the Agencourt Ampure XP magnetic PCR clean-up system. Primer-dimers were dismissed using a 1.8:1 volume ratio of magnetic beads for amplicons. A second purification with a 0.6:1 volume ratio was used to discard fragments smaller than 150-200 bp. Elution in 26 µL of EB buffer (Qiagen), from the original 55 µL of amplicons increased the concentrations which were measured using the Quant-iT Picogreen dsDNA assay kit (Invitrogen, Eugene, OR, USA). Four amplicon libraries (one for each tissue type) composed of samples in equimolar proportions were built. After preprocessing of the data, a fifth library composed of the 22 samples with the lowest number of reads was constructed (11 needle samples and 11 wood samples). Due to technical problem with one MID, one needle sample had to be incorporated into an independent library. For each library, amplicon length was verified using an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA).

Unidirectional sequencing was performed with Roche GS FLX+ system (Roche-454 Life Sciences, Brandford, CT, USA) at McGill University and Genome Quebec Innovation Centre (Montreal, QC, Canada). Each library was sequenced on a 1/4th plate region. The fifth library was sequenced on two independent 1/4th regions.

2.4.2 Data processing and statistical analyses

Denoising was performed using the mothur (v1.33.1 Schloss et al., 2009) implementation of the PyroNoise algorithm (Quince et al., 2009), with the mindelta parameter set on 10^{-6} rather than for 1000 iterations. Analyses were performed on the ITS sub-fragment delimited by the primer ITS1-F and the oligonucleotide ITS3a (CTTTCAACAACGGATCTCTT (Ponchart et al., 2019)). Conserving a portion of the ribosomal small subunit (pSSU) in front of the full length ITS1 sub-locus enhances the detection of putative chimeras for 454 pyrosequencing data and reduces the formation of rare possible spurious mOTUs (Ponchart et al., 2019). The presence of the ITS3a pattern with 3 mismatches allowed was assessed with fqgrep (ver 0.4.4, Indraniel et al., 2016) and the -c parameter was used to highlight its position. Shell sed commands based on the colour tag in front of the pattern were used to remove every nucleotide from that point resulting in sequences only displaying the pSSU-ITS1 fragment. All sequences which did not display this pattern were run through ITSx (Bengtsson-Palme et al., 2013) targeting fungal sequences. Only sequences in which both ITS1 and 5.8S were detected were conserved. Partial SSUs were retrieved to match the same sub-portion of the amplicons (post ITS1-F – ante ITS3a).

All sequences containing any mismatch to the primer and the barcode, any ambiguous base call, or any homopolymer repeat longer than 9bp were removed. To take better advantage of the sequencing depth later on, samples of the same tissue from the same height and tree were merged together yielding 128 sample-units. Putative chimeric sequences were then assessed with chimera.perseus on mothur and discarded. All unique sequences were processed with ESPRIT (Sun et al., 2009) to perform pairwise alignments. Using the single-linkage preclustering (SLP) algorithm developed by Huse et al., (2010), the distance matrix was processed to a precluster step with a width of 0.03. The clustering into molecular Operational Taxonomic Units (mOTUs) was done at a 95% sequence similarity threshold with mothur using the average neighbour method. Representative sequences of each mOTU were retrieved from the preclustering step and all unique sequences were conserved and deposited in GenBank (Submission number: SUB5671970).

Taxonomy of these representative sequences was assessed with the classify.seqs command from mothur with the implementation of the rdp method (Wang et al., 2007). The update of Unite (Koljalg et al., 2005) v7 from November 20th, 2016 served as template (UNITEv6_sh_99_s.fasta and UNITEv6_sh_99_s.tax). The command was run with a cutoff of 60, and 1000 iterations. The default kmer size of 8 was kept. The compilation of the assignment was based on bootstrap value support of 75 for the Class level. If the class was mentioned as *incertae sedis*, the assignment was made to the next lowest level supported by a bootstrap value of 90 at least if possible or to a higher level otherwise. Refinements of the assignment were attempted by blasting the sequences against the NCBI nt database (downloaded on September 24th, 2016).

Species-accumulation and rarefaction curves were calculated with mothur, sampling every 1000 sequences, and represented with the R program version 3.2.3 (R Core Development Team, 2015). Extrapolation of the richness was evaluated using the specpool command of the vegan package (Oksanen 2007; Smith and van Belle, 1984) with the bootstrap estimate. To ensure the validity of sample comparisons, we sub-sampled to select randomly 1 600 sequences per sample. One hundred sub-samplings were performed to reduce the stochasticity of the observations. Diversity was measured using Fisher's alpha (Fisher et al., 1943) and evenness was measured with the Pielou's evenness index (Pielou 1966). In each case singletons and doubletons were excluded prior to calculations. Both metrics were computed with the vegan R package. Normality was assessed using the Shapiro-Wilk test. A parametric one-way ANOVA coupled with a Tukey's range test was used to evaluate the relation of stand, and tissue type to evenness. A non-parametric Kruskal-Wallis test followed by a Dwass Steele Critchlow Fligner test (DSCF, run with the npar1way procedure in SAS 9.4 (SAS Institute Inc., 2012)) was used for the same purposes for diversity. Unless otherwise noted, all statistical analyses were performed with R.

We used one-way ANOSIM (analysis of similarity (Clarke 1993)) to compare community structure. Analyses were based on Jaccard's index and were performed in PAST (Hammer et al., 2001). Singletons and doubletons were excluded prior to calculations. Results were visualized with non-metric multidimensional scaling (nMDS) in PAST (Taguchi & Oono, 2004) and replotted with R. Permutational multivariate analysis of variance (PERMANOVA (Anderson 2001)), implemented with the adonis function from the vegan R package, was used to quantify sources of variation in the community structure (using Jaccard dissimilarity measure, with 999 permutations). To visualize homogeneity in the endomycobiota structure, correlation-based network analysis was used to infer co-occurrence patterns of sample-units based on their mOTU composition. Singletons and doubletons were removed from the original sample-by-mOTU feature table and the sequence

counts were replaced with relative abundance per sample. This substitution allowed accounting for the difference in sequencing depth among samples (Ju et al., 2014). mOTUs occurring in fewer than 3 samples were discarded to reduce the network complexity through decreasing sample-specificity. Pairwise Spearman's rank correlations were calculated using Hmisc package (Harrell 2008) in R. Only correlations with a Spearman's correlation coefficient (ρ) > 0.6 and a P-value < 0.01 were considered. Visualization of the network and detection of its modules were conducted using Cytoscape (ver 3.3.0, Shannon et al., 2003). The same analysis was performed to infer mOTUs co-occurrence patterns based on their distribution among samples.

2.5 Results

The sequencing of the libraries produced 934 467 reads; the pSSU-ITS1 fragment was extracted for 830 104 fungal trimmed sequences. After removal of putative chimeras, 713 105 sequences, with a mean length of 227.9 bp, were considered for the clustering which resulted in 2 446 mOTUs (including 204 doubletons and 858 singletons). Saturation was not observed for the dataset (Figure S2.1.a) or for any individual tissue (Figure S2.1.b). Extrapolation of the richness was estimated to 3 069 ± 64 mOTUs, indicating that ca. 80% of the expected species richness was collected. The number of mOTUs resulting from the 100 sub samplings ranged from to 1 272 to 1 317 with a mean value of 1 299.33 mOTUs (± 9.19). All 1 384 mOTUs composed of at least 3 sequences were represented among the sub samplings.

Variation in diversity could not be attributed to stand (Figure 2.1.a and Table 2.2), but diversity did differ by tissue type (Figure 2.1.b and Table 2.2). Needles displayed the highest diversity, followed by wood and buds, and then bark (Figure 2.1.b). Distribution of the mOTUs was also more even within wood and needle samples, than in bark and bud samples (Figure 2.1.d). Evenness did not differ by stand (Figure 2.1.c and Table 2.2). Fungal communities differed in composition among stands (with the exception of La Chute 01 and La Chute 02 which harboured similar communities for 96 of the 100 sub-samplings) and among tissues types (Table 2.2 and Figure 2.2). Tissue types explained a more important part of the variation (22.4%) than stands (4%) (Table 2.3). The interaction between tissue type and stands was also significant and explained 8.1% of the variance (Table 2.3). Within each tissue, variation in the composition of the associated communities was observed between stands (Table 2.3). Analysis of sample co-occurrence (based on mOTUs composition) revealed an

intra-group heterogeneity (either for samples from the same tissue, or from the same stand). Only 9 bark samples were engaged in 5 edges (3 pairs, 1 trio) (data not shown).

The 1 384 mOTUs composed of at least 3 sequences were taxonomically assigned among four phyla. Most belonged to the Dikarya (751 mOTUs affiliated to the Ascomycota and 233 to the Basidiomycota), whereas "Zygomycota" and Chytridiomycota were represented by 6 and 2 mOTUs, respectively. About 28% of the mOTUs could be identified only at the level of Fungi, and 20% at the level of phylum. Ascomycota were distributed among 12 classes and 1 family, Basidiomycota among 9 classes and 1 order, Zygomycota among 2 orders, and Chytridiomycota among 1 class. The stands had different fungal communities based on species composition, but the relative abundances of the different taxonomic ranks were similar among the four stands, with the exception of the Ascomycota assigned to Dothideomycetes, Orbiliomycetes and *Knufia*, and the Basidiomycota assigned to Pucciniomycetes (Table 2.4.a and Figure 2.3.a). With the exception of Dothideomycetes, which represented 174 mOTUs, the three remaining categories were relatively uncommon with 5, 3 and 7 mOTUs. Differences in relative abundance were observed among tissue types both at the phylum and the class levels (Table 2.4.b and Figure 2.3.b).

The community inhabiting the needle samples (Figure 2.3.b and Table 2.4.b) was the one for which taxonomic assignment was the most precise, with the lowest proportion of mOTUs only assigned to the kingdom level and to unclassified Ascomycota. This community also displayed the highest proportion of Basidiomycota, mainly distributed among Agaricomycetes, Tremellomycetes, and Exobasidiomycetes. Ascomycota of the needle samples were mainly attributed to Dothideomycetes, unclassified Ascomycota, and Leotiomycetes. Sordariomycetes, Lecanoromycetes, and Saccharomycetes were most abundant within needle samples and wood samples, whereas Pezizomycetes were most abundant within needle samples only. The endophyte fungal community within the bud samples (Figure 2.3.b and Table 2.4.b) displayed the highest proportion of mOTUs only assigned to the kingdom. Agaricomycetes were at their lowest, and Basidiomycota were mostly distributed among Exobasidiomycetes, Tremellomycetes, and Unclassified Basidiomycota. Apart from Ascomycota only assigned to the phylum, Dothideomycetes, Leotiomycetes, and Sordariomycetes were the most represented Ascomycota.

The bark sample community (Figure 2.3.b and Table 2.4.b) displayed both the highest proportion of Ascomycota and the lowest proportion of Basidiomycota. Among the Ascomycota, the proportions of mOTUs only assigned to the phylum, and to Leotiomycetes were at their highest, whereas Dothideomycetes and Sordariomycetes were at their lowest. For the Basidiomycota, all assignments were less common within bark samples. The fungal community within the wood samples (Figure 2.3.b and Table 2.4.b) was mainly composed of Ascomycota, mostly distributed among unclassified Ascomycota, Dothideomycetes, and Leotiomycetes. Along with the community from needle samples, it had the highest proportion of Sordariomycetes, Lecanoromycetes, and Saccharomycetes. For the Basidiomycota, with the exception of the Agaricomycetes, wood samples displayed the highest proportions for all assignments.

Of the 681 mOTUs composed of at least 3 sequences and present in at least 3 samples which were retained for the co-occurrence analysis, 154 mOTUs established 373 pairwise correlations ($\rho > 0.6$ and a P-value < 0.01). About half of them represented isolated correlations forming pairs (46 mOTUs in 23 edges), trios (15 mOTUs in 11 edges) and three low complexity structures (19 mOTUs – 23 edges) (data not shown). The remaining 72 mOTUs established 315 edges and were organized in two modules (Figure 2.4). The first module was composed of 35 mOTUs and 245 edges. While it appears extremely dense and composed of more cosmopolitan mOTUs (every mOTU but one was found at every stand), no clear structure could be observed although 23 of the 35 mOTUs were not featured in the associated bark community. Blasts of their representative sequences against the Unite database were inconclusive as none of the mOTU displayed any hits (based on coverage of at least 90% of the sequence and similarity of at least 97%) to fully identified sequences. The second module, composed of 37 mOTUs and 76 edges, was more disparate and no clear structure was observed either. The same limitation on the taxonomic assignment was encountered.

2.6 Discussion

The main goal of this study was to survey the aerial endomycobiota of balsam fir trees and to define whether it constituted a singular entity or if it might be fractioned into distinct communities. Knowledge of the plant microbiota in general and of fungal endophytes in particular, is derived mainly from studies on foliar tissue (Vorholt 2012; Penuelas & Terradas 2014; Arnold 2007; Arnold & Lutzoni 2007; Bálint et al., 2013). And while the conclusions are usually assumed to be true for the whole aerial system, no evidence exists to confirm this. By standardizing sample preparation, high throughput sequencing techniques have allowed a more accessible

and extensive approach for studying plant microbiotas using metabarcoding (Ottesen et al., 2013; Colemann-Derr et al., 2015, Fonseca-Garcia et al., 2016). Studies of plant microbiotas recovered from specific tissue might lead to a better understanding of their structure and of the interactions with the host and among the different members of these communities.

Differences in diversity were observed among communities from different tissues, with needle endomycobiota displaying the highest diversity. This predominance could be explained by the presence of stomata serving as multiple points of entrance in addition to a possible access to the photo-assimilates. While needles provide a relative protection against environmental stresses (notably against UV radiation, changes in temperature and relative humidity) in comparison with their surface (Leveau 2009; Redford et al., 2010; Rostagi et al., 2013), they are also a privileged site for exchange with this environment. The community from wood samples had a higher diversity than the community from bark samples. Wood represents a more stable and more insulated substrate than bark which is more exposed to the elements. However, previous studies based on culture dependent methods tend to indicate that communities from bark usually display a higher diversity than those from wood (Petrini & Fisher 1990; Fisher & Petrini 1990; Wang & Guo 2007). The difference observed could be due to the difference in the methods employed. In previous studies, xylem fragments were used for plating, whereas wood was crushed into a fine powder in this study, thus increasing dramatically the surface exposed for the metabarcoding. In culture-dependent methods, differences in incubation conditions and drying regimes increase water loss and oxygen concentration. This influences which species develop overtly on the growth medium and, consequently which portion of the community present in the original substrate is retrieved (Oses et al., 2008; Parfitt et al., 2010). Nonetheless, culture dependent methods, while being limited to fungi able to grow on medium, allow the recovery of the active diversity, whereas the metabarcoding approach detects DNA of all present species, active or not. The community from bud samples displayed an intermediate diversity between that of communities from wood and bark samples. At first glance, this relatively high diversity is surprising given the shorter time of exposure to the environment but buds present the particularity of being extremely sticky, and the efficiency of the superficial sterilization for this tissue might therefore be questionable.

Needles, buds, bark, and wood harboured distinct fungal communities. This suggests tissue specificity and partitioning of the aerial endomycobiota into the different ecological niches that the tissue types constitute. It is likely that such segregation is also accompanied by the display of distinct functional roles. Yet this aspect remains the main limitation of fungal endophyte metabarcoding studies as endophyte species for which the

ecological role is known remain the exception and since the vast majority of endophytes has yet to be adequately characterized. Due to those drawbacks, interpretations of results from metabarcoding studies are often restricted to distribution patterns. The community associated with bark displayed the least heterogeneity in its composition as observed from both the nMDS and the sample co-occurrence analysis. It also displayed the lowest diversity. As a dying tissue, bark might represent a less appealing substrate for endophytes and competition with the epiphytic community might be more pronounced. Most of the mOTUs composing the main module in the co-occurrence analysis were not featured in bark either. Samples from wood, which is the only aerial tissue not directly exposed to the environment, harboured the endophytic community displaying the most apparent heterogeneity. The nMDS clearly showed this disparity. This result might indicate a lower selectivity in the recruitment of the fungal cohort associated with this tissue, or that wood serves as a privileged shelter for opportunistic fungi. In support of the latter hypothesis, we observed that wood tissue had the second most diverse community and presented the highest evenness which suggests a more balanced competition between members of the community.

Relative abundance of the unidentified mOTUs (only assigned to the kingdom, or to a phylum – mostly Ascomycota) generally accounted for the main proportion of taxonomical assignment for each tissue (56.1%, 58.1%, and 45.32% for bud, bark, and wood samples, respectively) with the exception of the needle samples (36.71%), possibly reflecting the focus on foliar endophytes in previous studies. This suggests that switching the focus of endophytology studies from the leaf associated community to the different aerial tissue types will allow a better recovery of the overall diversity displayed by fungal endophytes. With respect to this suggestion, wood might be a candidate of major interest since it is less transient than buds and bark, and presented the second highest diversity in this study. Whereas different studies have focused on wood-associated fungi, they were mainly concerned with species implicated in wood decay (Oses et al., 2008; Giordano et al., 2008; Parfitt et al., 2010), rather than considering the tissue as a niche occupied by fungal endophytes (but see Hutchinson 1999; Gazis & Chaverri, 2010 – the latter suggesting that the sapwood of *Hevea brasiliensis* might even harbour a greater diversity of fungal endophytes than its leaves).

Interestingly, while needle- and wood samples both harboured distinct fungal communities based on species composition, the taxonomic structures of the communities from the two tissues shared similar proprieties based on relative abundance of the assignment. For instance, communities from these tissues had a more elevated abundance of Basidiomycota and Sordariomycetes which are known plant decomposers (notably xylariacious ascomycetes) (Osono 2006; Oses et al., 2008; Parfitt et al., 2010). It has been suggested that

some fungi with the ability to degrade litter or decay wood might occupy plant tissue as endophytes to gain a place as precursors for the earlier stage of plant material decomposition (Osono 2006; Oses et al., 2008; Parfitt et al., 2010). Screening for the presence of the 11 wood decay fungi reported by Parfitt et al. (2010) revealed mOTUs with potential assignment to *Hypoxylon fuscum*, *Nemania serpens*, *Hypoxylon fragiforme.* Other mOTUs also displayed blast hits to known wood decay fungi such as *Daldinia loculata*, *Hypoxylon macrocarpum*, and *Hypoxylon perforatum* but surprisingly all of these wood decay species were only detected in needle samples and not in wood samples (data not shown). It would be interesting to investigate further if the presence of wood decay fungi within the needles (1) represents only an “accidental” colonization of the tissue, facilitated by the entrance via the stomata; (2) suggests that needles (and leaves in general) are a favoured point of entrance for the aerial endomycobiota migrating onwards to the wood; (3) is a preliminary step of infection, with needles providing a massive surface of incubation where fungi can develop latently as endophytes to increase their putative inoculum charge before switching to pathogenicity and reach their targeted zone. In any case, it has to be stressed that none of our results can argue for a movement of the observed species within the plant due to the method used.

Another similarity between needle- and wood samples was that they both harboured the highest abundance of Lecanoromycetes (lichen-forming fungi). Such abundance was surprising within wood samples since these organisms are more expected in bark samples. Gazis et al. (2012) however have shown in their study of endophytic fungi from the sapwood of *Hevea* spp. that assignment to Lecanoromycetes might be misleading and that some could actually belong to the new discovered class of Xylonomycetes for which no representative can be found in the Unite database. Thus, the 27 mOTUs assigned to Lecanoromycetes found in the wood samples received a particular attention. While half of these mOTUs displayed relative good quality hits (≥ 95% similarity on ≥ 90% of the sequence length) to identified Lecanoromycetes with a binomial name, none presented hits with the same properties to the four referenced Xylonomycetes species in nt (based on accession number: *Xylona heveae*: JX838222-JX838234, NR_121539; *Symbiotaphrina buchneri*: KY105569, DQ248313; *Symbiotaphrina kochii*: DQ248314, KC215110, KC215113; *Trinosporium guianense*: JX069869) (data not shown). The endosubstratal thallus of these Lecanoromycetes mOTUs probably penetrates more extensively the branches, reaching as far as the living wood. Slight brushing and superficial sterilization might explain their absence from the bark community.

2.7 Conclusion

This study focused on the endomycobiota of balsam fir trees at a small geographical scale and will have to be followed by larger studies including more tree species on a more global scale. Nevertheless, our results hint towards a possible fragmentation of the aerial endomycobiota depending on the type of tissue considered. Fungal endophytic communities retrieved from needle, bud, bark, and wood samples were all different in terms of species composition, and displayed distinct diversities. Whereas there is no doubt that a more extensive approach to the study of the plant microbiota will benefit the understanding of its structure and its interactions both with the host and among its different members, a partitioned approach per tissue type might achieve as much in refining the scope of the observations. It will be interesting to generalize this kind of observations, in studying the impact of different factors, such as tree species, host genotypes, seasonal sampling date, and other environmental factors for each of the tissue types and to compare their effects to those documented on foliar communities. The search for the hidden fungal diversity might also gain from such a partitioned approach, as foliar endophytes have received much of the attention in endophytology, to the detriment of the communities inhabiting other plant parts which are usually overlooked. Community in wood samples might be of primordial importance: we found it to be of the same amplitude as the one from the foliar samples, and it plays a key role in decay and thus in nutrient cycling which is crucial to the functioning of woodland ecosystems. Communities from buds and bark should not be put aside; however, as more than half of their diversity could only be assigned to the kingdom level or the Ascomycota phylum.

2.8 Acknowledgements

We kindly thank Julie Boudreau from McGill University and Genome Quebec Innovation Centre (Montreal, Qc, Canada) for technical assistance on 454 pyrosequencing. We are indebted to Pr Elizabeth A. Arnold (University of Arizona, Tucson, Az, USA) who provided helpful comments and invaluable suggestions to improve the overall quality of this manuscript. We are grateful to Dr Jana U'Ren (University of Arizona, Tucson, Az, USA) for useful discussions on pyrosequencing data processing and her great help on ecological analysis. We also thank Patrick Gagné (Université Laval, Qc, Canada) for feedbacks on bioinformatic analysis.

2.9 References

Ahlholm, J. U., Helander, M., Henriksson, J., Metzler, M., & Saikkonen, K. (2002). Environmental conditions and host genotype direct genetic diversity of Venturia ditricha, a fungal endophyte of birch trees. *Evolution*, *56*(8), 1566-1573.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, *26*(1), 32-46.

Arnold, A. E. (2007). Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal biology reviews*, *21*(2), 51-66.

Arnold, A. E., & Lutzoni, F. (2007). Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots?. *Ecology*, *88*(3), 541-549.

Bálint, M., Tiffin, P., Hallström, B., O'Hara, R. B., Olson, M. S., Fankhauser, J. D., ... & Schmitt, I. (2013). Host genotype shapes the foliar fungal microbiome of balsam poplar (Populus balsamifera). *PLoS One*, *8*(1), e53987.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... & Amend, A. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution*, *4*(10), 914-919.

Bills, G. F. (1996) in Endophytic Fungi in Grasses and Woody Plants, eds. Redlin, S. C. & Carris, L. M. (APS Press, St. Paul, MN), pp. 31–65.

Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biology*, *13*(8), e1002226.

Brandt, J. P. (2009). The extent of the North American boreal zone. *Environmental Reviews*, *17*(NA), 101-161.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual review of plant biology*, *64*, 807-838.

Bulgarelli, D., Garrido-Oter, R., Münch, P. C., Weiman, A., Dröge, J., Pan, Y., ... & Schulze-Lefert, P. (2015). Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell host & microbe*, *17*(3), 392-403.

Carroll, F. E. (1977). Preliminary studies on the incidence of needle endophytes in some European conifers. *Sydowia*, *29*, 87-103.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, *18*(1), 117-143.

Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T., ... & Tringe, S. G. (2016). Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytologist*, *209*(2), 798-811.

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., ... & Vorholt, J. A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proceedings of the National Academy of Sciences*, *106*(38), 16428-16433.

Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 42-58.

Fisher, P. J., & Petrini, O. (1988). Tissue specificity by fungi endophytic in Ulex europaeus. *Sydowia*, *40*, 46-50.

Fisher, P. J., & Petrini, O. (1990). A comparative study of fungal endophytes in xylem and bark of Alnus species in England and Switzerland. *Mycological Research*, *94*(3), 313-319.

Fisher, P. J., Petrini, O. S. B. C., & Sutton, B. C. (1993). A comparative study of fungal endophytes in leaves, xylem and bark of Eucalyptus in Australia and England. *Sydowia*, *45*(2), 338-345.

Fonseca-García, C., Coleman-Derr, D., Garrido, E., Visel, A., Tringe, S. G., & Partida-Martínez, L. P. (2016). The cacti microbiome: interplay between habitat-filtering and host-specificity. *Frontiers in microbiology*, *7*.

Friesen, M. L., Porter, S. S., Stark, S. C., von Wettberg, E. J., Sachs, J. L., & Martinez-Romero, E. (2011). Microbially mediated plant functional traits. *Annual review of ecology, evolution, and systematics*, *42*, 23-46.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*, *2*(2), 113-118.

Gazis, R., & Chaverri, P. (2010). Diversity of fungal endophytes in leaves and stems of wild rubber trees (Hevea brasiliensis) in Peru. *Fungal ecology*, *3*(3), 240-254.

Gazis, R., Miadlikowska, J., Lutzoni, F., Arnold, A. E., & Chaverri, P. (2012). Culture-based study of endophytes associated with rubber trees in Peru reveals a new class of Pezizomycotina: Xylonomycetes. *Molecular phylogenetics and evolution*, *65*(1), 294-304.

Giordano, L., Gonthier, P., Varese, G. C., Miserere, L., & Nicolotti, G. (2009). Mycobiota inhabiting sapwood of healthy and declining Scots pine (Pinus sylvestris L.) trees in the Alps. *Fungal Diversity*, *38*(69), e83.

Hacquard, S., & Schadt, C. W. (2015). Towards a holistic understanding of the beneficial interactions across the Populus microbiome. *New Phytologist*, *205*(4), 1424-1430.

Hammer, Ř., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis–Palaeontol. Electron. 4: 9pp.

Hardoim, P. R., Van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., ... & Sessitsch, A. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293-320.

Harrell, Frank E. 2008. “Hmisc: Harrell Miscellaneous.” R package version 3.4-4.

Hata, K., & Futai, K. (1996). Variation in fungal endophyte populations in needles of the genus Pinus. *Canadian Journal of Botany*, *74*(1), 103-114.

Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, *12*(7), 1889-1898.

Hutchison, L. J. (1999). Wood-inhabiting microfungi isolated from Populus tremuloides from Alberta and northeastern British Columbia. *Canadian journal of botany*, *77*(6), 898-905.

Indraniel et al.. (2016). fqgrep: v0.4.4. Zenodo. 10.5281/zenodo.45105

Joshee, S., Paulus, B. C., Park, D., & Johnston, P. R. (2009). Diversity and distribution of fungal foliar endophytes in New Zealand Podocarpaceae. *Mycological research*, *113*(9), 1003-1015.

Ju, F., Xia, Y., Guo, F., Wang, Z., & Zhang, T. (2014). Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environmental microbiology*, *16*(8), 2421-2432.

Jumpponen, A., & Jones, K. L. (2009). Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate Quercus macrocarpa phyllosphere. *New Phytologist*, *184*(2), 438-448.

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., ... & Vorholt, J. A. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *The ISME journal*, *6*(7), 1378.

Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E et al. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 166 : 1063–1068.

Laforest-Lapointe, I., Paquette, A., Messier, C., & Kembel, S. W. (2017). Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature*, *546*(7656), 145-147.

Larsen, P. E., Sreedasyam, A., Trivedi, G., Podila, G. K., Cseke, L. J., & Collart, F. R. (2011). Using next generation transcriptome sequencing to predict an ectomycorrhizal metabolome. *BMC systems biology*, *5*(1), 70.

Leveau, J. (2009). Microbiology: Life on leaves. *Nature*, *461*(7265), 741-742.

Oksanen, J., R. Kindt, P. Legendre, B. O ' Hara, and M. H. H. Stevens. 2007. Vegan: Community ecology package. R package, version 1.8 – 8

Oono, R., Rasmussen, A., & Lefèvre, E. (2017). Distance decay relationships in foliar fungal endophytes are driven by rare taxa. *Environmental microbiology*, *19*(7), 2794-2805.

Oses, R., Valenzuela, S., Freer, J., Sanfuentes, E., & Rodriguez, J. (2008). Fungal endophytes in xylem of healthy Chilean trees and their possible role in early wood decay. *Fungal Divers*, *33*, 77-86.

Osono, T. (2006). Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. *Canadian Journal of Microbiology*, *52*(8), 701-716.

Osono, T. (2008). Endophytic and epiphytic phyllosphere fungi of Camellia japonica: seasonal and leaf age-dependent variations. *Mycologia*, *100*(3), 387-391.

Ottesen, A. R., Peña, A. G., White, J. R., Pettengill, J. B., Li, C., Allard, S., ... & Strain, E. (2013). Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato). *BMC microbiology*, *13*(1), 114.

Parfitt, D., Hunt, J., Dockrell, D., Rogers, H. J., & Boddy, L. (2010). Do all trees carry the seeds of their own destruction? PCR reveals numerous wood decay fungi latently present in sapwood of a wide range of angiosperm trees. *Fungal Ecology*, *3*(4), 338-346.

Peñuelas, J., & Terradas, J. (2014). The foliar microbiome. *Trends in plant science*, *19*(5), 278-280.

Peršoh, D. (2013). Factors shaping community structure of endophytic fungi–evidence from the Pinus-Viscum-system. *Fungal Diversity*, *60*(1), 55-69.

Petrini, O. (1986). Taxonomy of endophytic fungi of aerial plant tissues. *Microbiology of the phyllosphere/edited by NJ Fokkema and J. van den Heuvel*.

Petrini, O., & Fisher, P. J. (1988). A comparative study of fungal endophytes in xylem and whole stem of Pinus sylvestris and Fagus sylvatica. *Transactions of the British Mycological Society*, *91*(2), 233-238.

Petrini, O., & Fisher, P. J. (1990). Occurrence of fungal endophytes in twigs of Salix fragilis and Quercus robur. *Mycological Research*, *94*(8), 1077-1080.

Petrini, O. (1991). Fungal endophytes of tree leaves. In *Microbial ecology of leaves* (pp. 179-197). Springer, New York, NY.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, *13*, 131-144.

Ponchart, J. P., Bernier, L., Bérubé, J. A. (2019) SSU to the rescue: conserving the original partial SSU fragment ahead of the ITS1 sub-locus enhances the detection of putative chimeras and limits the formation of possibly spurious taxa in fungal metabarcoding. *To be submitted.*

Porras-Alfaro, A., & Bayman, P. (2011). Hidden fungi, emergent properties: endophytes and microbiomes. *Annual review of phytopathology*, *49*.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., ... & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, *6*(9), 639-641.

R Core Team (2015). R: A language and environment for statistical computing.

Rajala, T., Velmala, S. M., Tuomivirta, T., Haapanen, M., Müller, M., & Pennanen, T. (2013). Endophyte communities vary in the needles of Norway spruce clones. *Fungal biology*, *117*(3), 182-190.

Rastogi, G., Coaker, G. L., & Leveau, J. H. (2013). New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. *FEMS microbiology letters*, *348*(1), 1-10.

Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y., & Fierer, N. (2010). The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environmental microbiology*, *12*(11), 2885-2893.

Rodriguez, R. J., White Jr, J. F., Arnold, A. E., & Redman, R. S. (2009). Fungal endophytes: diversity and functional roles. *New phytologist*, *182*(2), 314-330.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., & Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nature reviews. Microbiology*, *5*(5), 355.

Rout, M. E. (2014). The plant microbiome. *Genomes of Herbaceous Land Plants*, *69*, 279-309.

Santamaría, J., & Bayman, P. (2005). Fungal epiphytes and endophytes of coffee leaves (Coffea arabica). *Microbial Ecology*, *50*(1), 1-8.

Saikkonen, K. (2007). Forest structure and fungal endophytes. *Fungal Biology Reviews*, *21*(2-3), 67-74.

SAS Institute Inc. 2012

Schenk, P. M., Carvalhais, L. C., & Kazan, K. (2012). Unraveling plant–microbe interactions: can multi-species transcriptomics help?. *Trends in biotechnology*, *30*(3), 177-184.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, *75*(23), 7537-7541.

Schulz, B., & Boyle, C. (2005). The endophytic continuum. *Mycological research*, *109*(6), 661-686.

Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

Sieber, T. N. (2007). Endophytic fungi in forest trees: are they mutualists?. *Fungal Biology Reviews*, *21*(2), 75-89.

Smith, E.P & van Belle, G. (1984). Nonparametric estimation of species richness. Biometrics 40, 119–129.

Stefani, F. O. P., & Bérubé, J. A. (2006). Biodiversity of foliar fungal endophytes in white spruce (Picea glauca) from southern Québec. *Botany*, *84*(5), 777-790.

Stanley, C. E., & van der Heijden, M. G. (2017). Microbiome-on-a-Chip: New Frontiers in Plant–Microbiota Research. *Trends in Microbiology*.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M. L., McKendree, W., & Farmerie, W. (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic acids research*, *37*(10), e76-e76.

Taguchi, Y. H., & Oono, Y. (2004). Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics*, *21*(6), 730-740.

Unterseher, M., Reiher, A., Finstermeier, K., Otto, P., & Morawetz, W. (2007). Species richness and distribution patterns of leaf-inhabiting endophytic fungi in a temperate forest canopy. *Mycological Progress*, *6*(3), 201-212.

U'Ren, J. M., Riddle, J. M., Monacell, J. T., Carbone, I., Miadlikowska, J., & Arnold, A. E. (2014). Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Molecular Ecology Resources*, *14*(5), 1032-1048.

U’Ren, J. M., & Arnold, A. E. (2016). Diversity, taxonomic composition, and functional aspects of fungal communities in living, senesced, and fallen leaves at five sites across North America. *PeerJ*, *4*, e2768.

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., & Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytologist*, *206*(4), 1196-1206.

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature reviews. Microbiology*, *10*(12), 828.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, *73*(16), 5261-5267.

Wang, Y., & Guo, L. D. (2007). A comparative study of endophytic fungi in needles, bark, and xylem of Pinus tabulaeformis. *Botany*, *85*(10), 911-917.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*(1), 315-322.

Wilson, D. (1995). Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 274-276.

Zimmerman, N. B., & Vitousek, P. M. (2012). Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences*, *109*(32), 13022-13027.

Figure 2.1: Results of diversity (A, B) and evenness (C, D) analyses results among the stands (A, C) and tissue types (B, D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis and DSCF tests or after ANOVA and Tukey tests with $p \leq 0.05$.

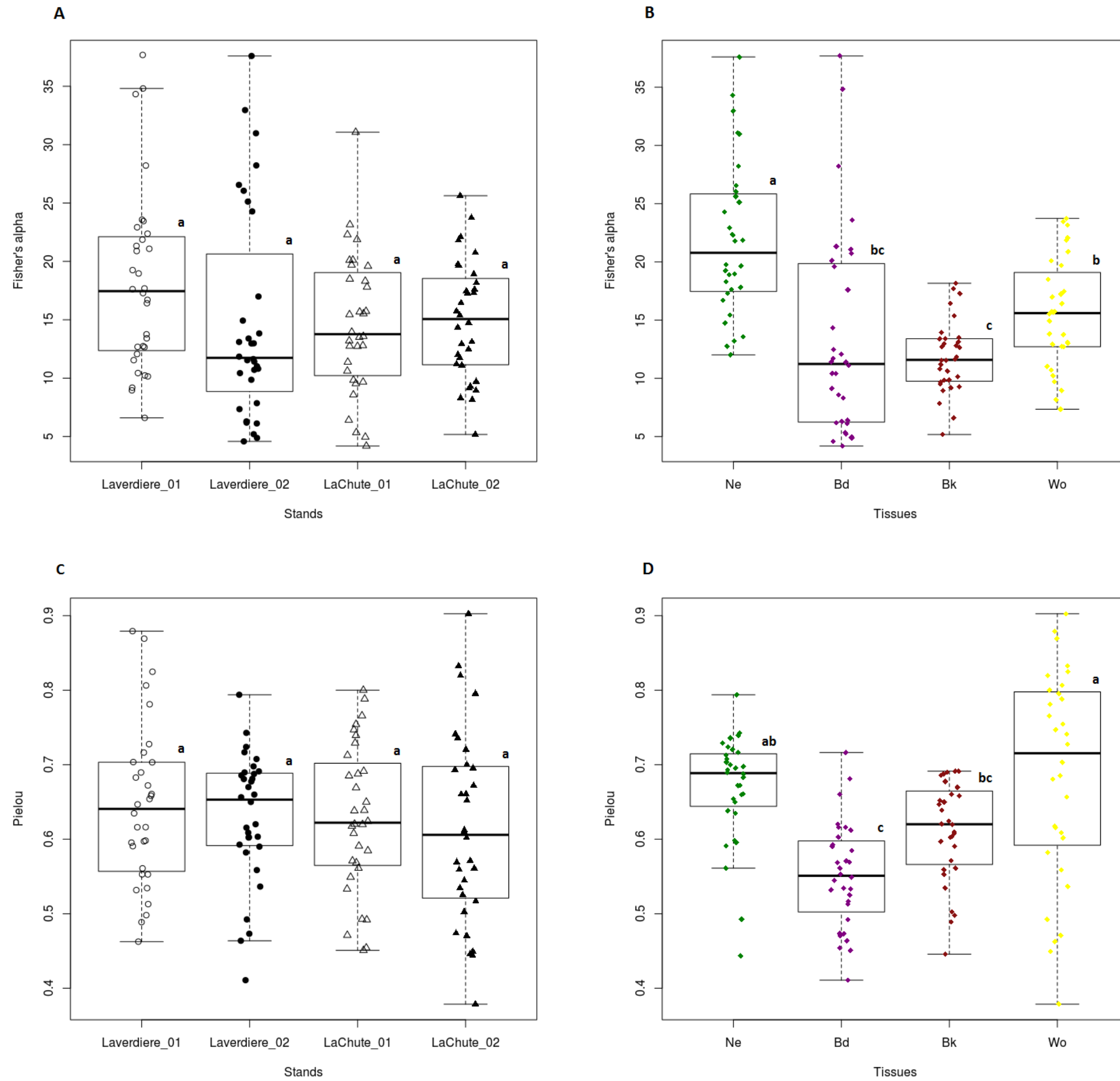


Figure 2.2: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition among the stands (open circles for Laverdière 01, closed circles for Laverdière 02, open triangles for La Chute 01, closed triangles for La Chute 02) and the tissue types (green for needles, purple for buds, brown for bark, yellow for wood). One-way analysis of similarity (ANOSIM) indicates significant differences ($P < 0.05$) in community composition both among the stands and the tissue types. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM.

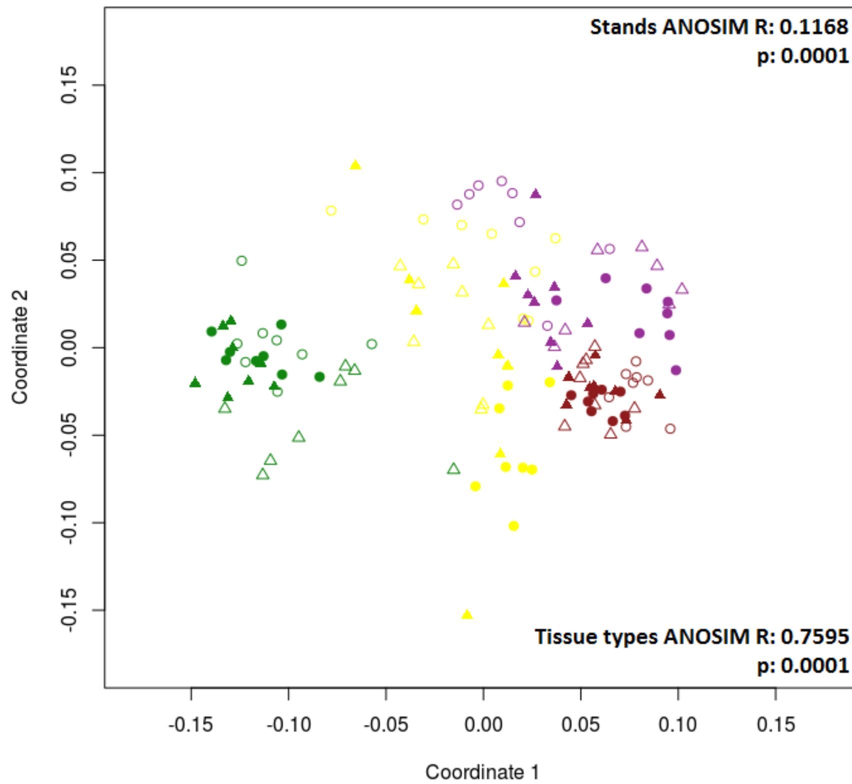


Figure 2.3: Relative abundance plots of fungal endophyte communities by stand (A) and tissue type (B)

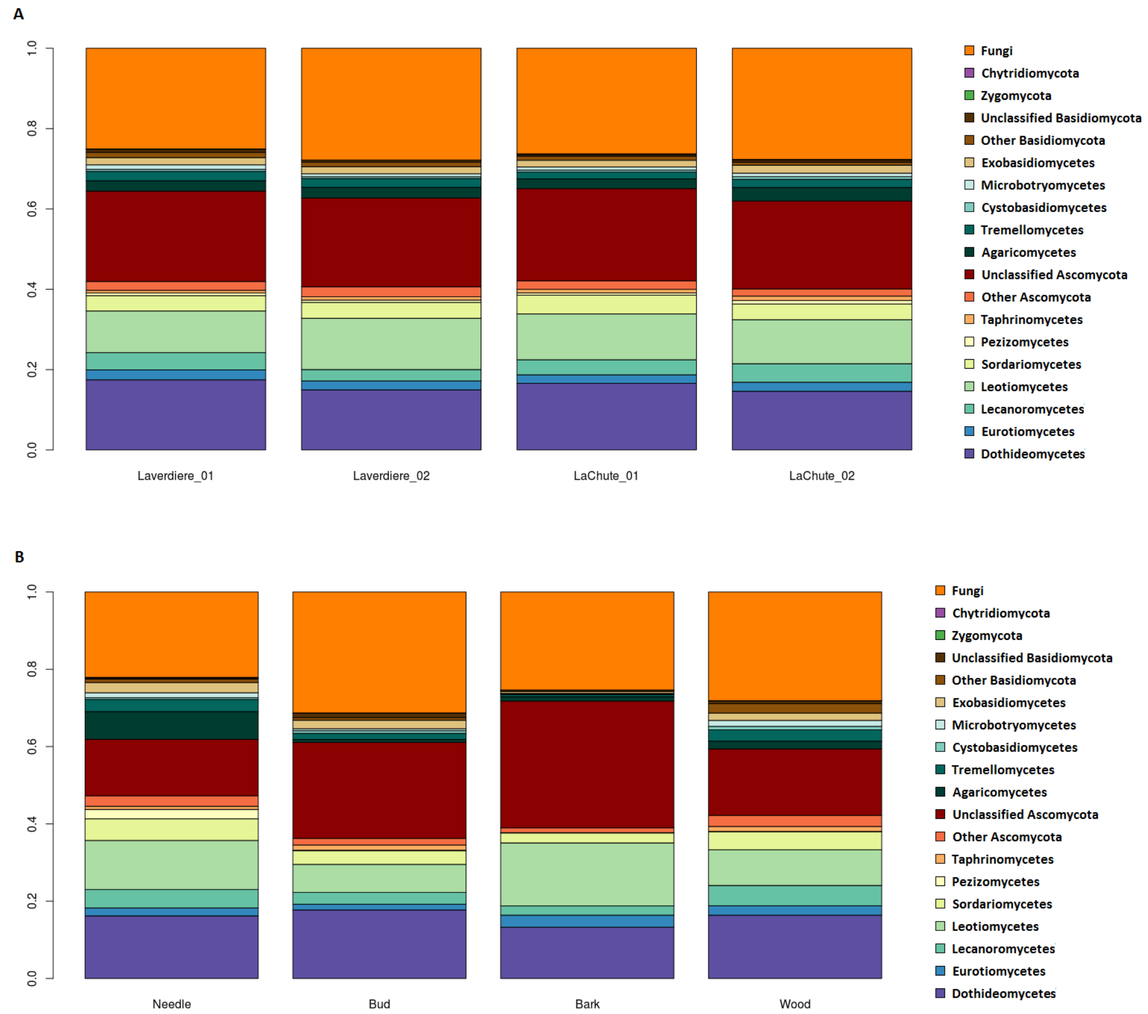


Figure 2.4: mOTUs co-occurrence network displaying two main modules. Nodes colour represents taxonomy assignment (as defined for Figure 2.3). Pairwise Spearman's rank correlations were considered if $\rho > 0.6$, and P-value < 0.01.

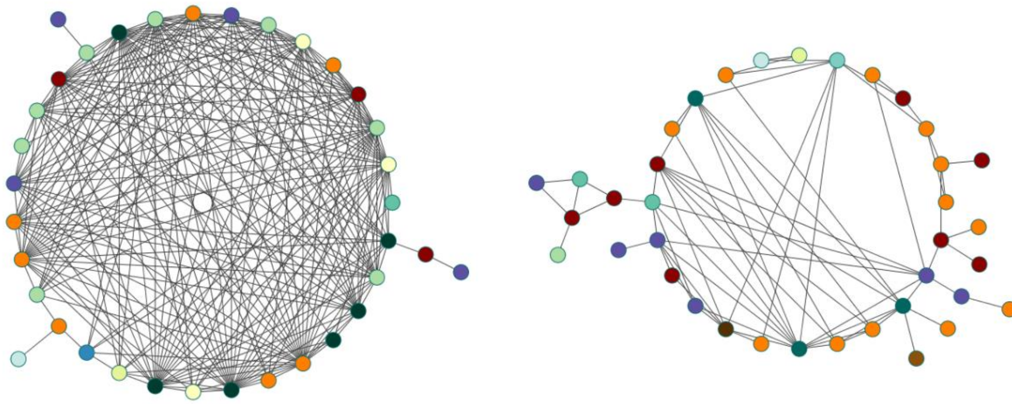


Table 2.1: Characteristics of Forêt Montmorency sites sampled in August 2012

	Laverdière		**La Chute**	
	Lav1	Lav2	LaC1	LaC2
Latitude	47°18.983' N	47°18.906' N	47°20.071' N	47°19.861' N
Longitude	071°09.666' W	071°09.686' W	071°05.888' W	071°05.916'W
Elevation (m.a.s.l.)	709m	706m	723m	723m
*A.b*coverage	N.A.	30.00%	90.00%	70.00%
Accompagnying tree species	*Picea mariana*	*Betula papyrifera*	*Picea mariana*	*Picea mariana* *Picea glauca*

Table 2.2: Analyses of diversity (Fisher's alpha), evenness (Pielou's evenness), and similarity (ANOSIM using Jaccard index). Richness and evenness values are based on 100 sub-samplings. Values for the community are based on the sub-sampling displaying less dissimilarity among groups (smaller R value). Significant effects are highlighted in green p-value ≤ 0.05

	Diversity		**Evenness**		**Community**	
	Kruskal-Wallis		ANOVA		ANOSIM	
	X2	p-value	F	p-value	R	p-values
Stands	5.1541	0.1609	0.4328	0.7299	0.1168	0.0001
Tissues	40.791	7.242x10-9	13.73	8.649x10-8	0.7595	0.0001

Table 2.3: Results of PERMANOVA with 999 permutations of the Jaccard dissimilarities for fungal mOTU community structure (DF, degrees of freedom; SS, sum of squares; MS, mean sum of square; F, pseudo-F by permutation)

	DF	SS	MS	F	R2	P
Aerial system						
Tissue	3	11.639	3.8795	12.822	0.22444	0.001
Stand	3	2.102	0.7006	2.3155	0.04053	0.001
Tissue:Stand	9	4.228	0.4698	1.5527	0.08154	0.001
Residuals	112	33.887	0.3026		0.65349	
Total	127	51.856			1	
Needle						
Stand	3	1.5301	0.51002	1.7801	0.16018	0.001
Residuals	28	8.0222	0.28651		0.83982	
Total	31	9.5523			1	
Bud						
Stand	3	1.854	0.61801	2.1116	0.1845	0.001
Residuals	28	8.1947	0.29267		0.8155	
Total	31	10.0487			1	
Bark						
Stand	3	1.3131	0.43771	1.722	0.15576	0.001
Residuals	28	7.1173	0.25419		0.84424	
Total	31	8.4304			1	
Wood						
Stand	3	1.6328	0.54426	1.4441	0.13399	0.007
Residuals	28	10.553	0.37689		0.86601	
Total	31	12.1858			1	

Table 2.4: Kruskal-Wallis and Dwass Steele Critchlow Fligner tests of the average relative abundance of fungal phylum and classes among the stands (A) and tissue types (B). Significant differences ($p \leq 0.05$) are highlighted in green, or yellow if the DSCF could not distinguish the effect due to low abundance. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA and Tukey tests instead of Kruskal-Wallis and DSCF tests. DF: Degrees freedom.

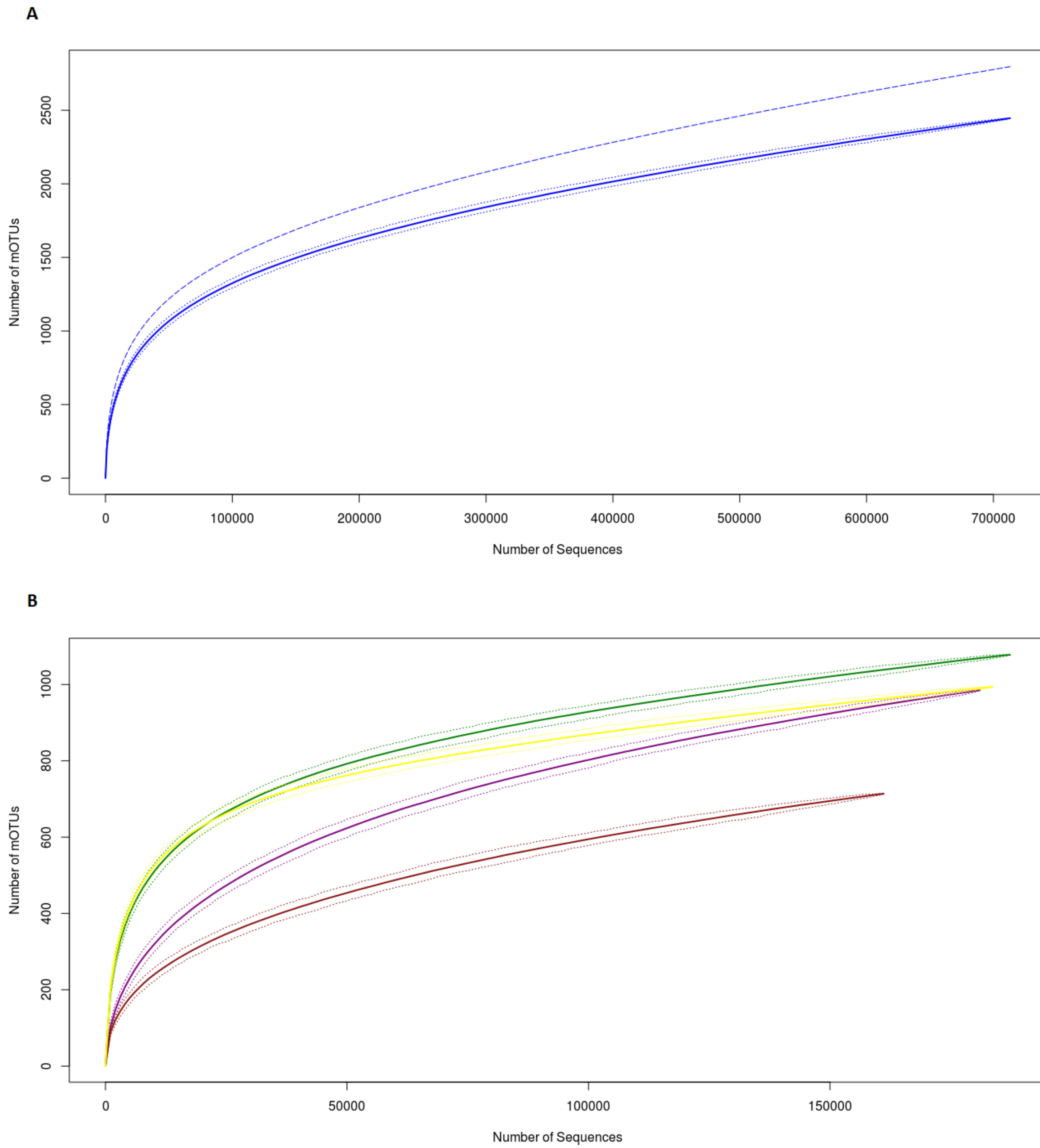
Taxonomic assignment	DF	Nb samples	X2	P	Lav01	Lav02	LaC01	LaC02
Ascomycota *	3	128	1.1944	0.3148	a	a	a	a
Basdiomycota	3	128	1.3786	0.7106	a	a	a	a
Zygomycota	3	128	2.0596	0.5601	a	a	a	a
Chytridiomycota	3	128	3.8705	0.2758	a	a	a	a
Fungi	3	128	2.9148	0.405	a	a	a	a
Dothideomycetes	3	128	10.1296	0.0175	a	ab	ab	b
Eurotiomycetes	3	128	1.1925	0.7548	a	a	a	a
Lecanoromycetes	3	128	6.9581	0.0732	a	a	a	a
Leotiomycetes *	3	128	1.3620	0.2576	a	a	a	a
Sordariomycetes	3	128	1.4904	0.6845	a	a	a	a
Pezizomycetes	3	128	0.8973	0.8261	a	a	a	a
Saccharomycetes	3	128	1.4909	0.6844	a	a	a	a
Taphrinomycetes	3	128	2.5613	0.4643	a	a	a	a
Archaeorhizomycetes	3	128	6.0472	0.1093	a	a	a	a
Orbiliomycetes	3	128	8.0389	0.0452	ab	a	ab	b
Pseudeurotiaceae	3	128	0.5184	0.9148	a	a	a	a
Arthoniomycetes	3	128	3.0000	0.3916	a	a	a	a
Digitodochium	3	128	3.7696	0.2874	a	a	a	a
Knufia	3	128	21.9007	6.841x10-5	a	ab	b	b
Lichinomycetes	3	128	3.7677	0.2877	a	a	a	a
Microcyclospora	3	128	1.0622	0.7862	a	a	a	a
Pleurophoma	3	128	6.0472	0.1093	a	a	a	a
Pseudoveronaea	3	128	3.3652	0.3387	a	a	a	a
Robillarda	3	128	3.0000	0.3916	a	a	a	a
Unclassified Ascomycota	3	128	0.3924	0.9418	a	a	a	a
Agaricomycetes	3	128	0.9301	0.8182	a	a	a	a
Tremellomycetes	3	128	1.7614	0.6234	a	a	a	a
Cystobasidiomycetes	3	128	0.7882	0.8523	a	a	a	a
Microbotryomycetes	3	128	4.5208	0.2104	a	a	a	a
Pucciniomycetes	3	128	17.8364	4.754x10-4	a	b	b	b
Exobasidiomycetes	3	128	0.2809	0.9636	a	a	a	a
Malasseziales	3	128	0.8073	0.8477	a	a	a	a

Agaricostilbomycetes	3	128	2.7586	0.4304	a	a	a	a
Ustilaginomycetes	3	128	3.0000	0.3916	a	a	a	a
Dacrymycetes	3	128	3.0000	0.3916	a	a	a	a
Unclassified Basidiomycota	3	128	0.8936	0.827	a	a	a	a
Mortierellales	3	128	0.5696	0.9034	a	a	a	a
Mucorales	3	128	3.0000	0.3916	a	a	a	a
Chytridiomycetes	3	128	3.8705	0.2758	a	a	a	a

Taxonomic assignment	DF	Nb samples	X2	P	Needle	Bud	Bark	Wood
Ascomycota *	3	128	29.5070	1.8150x10-14	b	b	a	b
Basdiomycota	3	128	79.4834	<2.2000x10-16	a	c	d	b
Zygomycota	3	128	11.5744	0.0090	a	a	a	a
Chytridiomycota	3	128	2.3099	0.5106	a	a	a	a
Fungi	3	128	39.0829	1.6770x10-8	c	a	b	b
Dothideomycetes	3	128	17.5510	5.4430x10-4	a	a	b	a
Eurotiomycetes	3	128	9.1311	0.0276	ab	b	a	ab
Lecanoromycetes	3	128	25.1538	1.4340x10-5	a	b	b	a
Leotiomycetes *	3	128	38.5970	<2.2000x10-16	b	c	a	c
Sordariomycetes	3	128	27.2544	5.2070x10-6	a	bc	c	ab
Pezizomycetes	3	128	99.9444	<2.200x10-16	a	b	b	b
Saccharomycetes	3	128	23.9686	2.5360x10-5	a	b	b	a
Taphrinomycetes	3	128	29.5098	1.7500x10-16	a	a	b	a
Archaeorhizomycetes	3	128	6.0472	0.1093	a	a	a	a
Orbiliomycetes	3	128	2.7021	0.4399	a	a	a	a
Pseudeurotiaceae	3	128	32.7502	3.6360x10-7	a	b	b	a
Arthoniomycetes	3	128	3.0000	0.3916	a	a	a	a
Digitodochium	3	128	8.7982	0.0321	a	a	a	a
Knufia	3	128	5.1051	0.1643	a	a	a	a
Lichinomycetes	3	128	36830	0.2978	a	a	a	a
Microcyclospora	3	128	5.2516	0.1543	a	a	a	a
Pleurophoma	3	128	2.0160	0.5691	a	a	a	a
Pseudoveronaea	3	128	18.7237	3.1180x10-4	a	a	a	a
Robillarda	3	128	3.0000	0.3916	a	a	a	a
Unclassified Ascomycota	3	128	91.4744	<2.2000x10-16	c	b	a	c
Agaricomycetes	3	128	73.5875	7.2740x10-16	a	c	bc	b
Tremellomycetes	3	128	39.2416	1.5430x10-8	a	b	c	ab
Cystobasidiomycetes	3	128	13.3818	0.0039	ab	a	b	a
Microbotryomycetes	3	128	40.5668	8.0800x10-9	a	b	b	a
Pucciniomycetes	3	128	6.4833	0.0903	a	a	a	a
Exobasidiomycetes	3	128	32.5715	3.9660x10-7	a	a	b	a
Malasseziales	3	128	66.7476	2.1210x10-14	b	c	c	a
Agaricostilbomycetes	3	128	5.7060	0.1268	a	a	a	a
Ustilaginomycetes	3	128	3.0000	0.3916	a	a	a	a
Dacrymycetes	3	128	3.0000	0.3916	a	a	a	a
Unclassified Basidiomycota	3	128	22.7639	4.5230x10-5	a	a	b	a

Mortierellales	3	128	15.4799	1.4490x10-3	a	a	a	a
Mucorales	3	128	3.0000	0.3916	a	a	a	a
Chytridiomycetes	3	128	2.3099	0.5106	a	a	a	a

Figure S2.1: Species accumulation curves (solid lines) and bootstrap estimate (long-dashed lines) of species richness for the full dataset (A) and for each tissue type (B) (Needles: green, Buds: purple, Bark: brown, Wood: yellow). Dotted lines represent the 95% confidence intervals.



Chapter 3 Endomycobiotas of the last four cohorts of *Abies balsamea* needles display relatively similar diversities but different compositions

Julien P. Ponchart[1,§], Louis Bernier[1,2], Jean A. Bérubé[3].

[1] Centre for Forest Research, Université Laval, Québec, Qc, Canada
[2] Institute for System and Integrative Biology (IBIS), Université Laval, Québec, Qc, Canada
[3] Laurentian Forest Centre, Canadian Forest Service, Natural Resources Canada, Québec, Qc, Canada
[§] Corresponding author: Julien P. Ponchart, email: julponchart@gmail.com

3.1 Résumé

Le but de cette étude était d'identifier l'endomycobiote foliaire associé au sapin baumier et de mesurer l'effet du temps d'exposition à l'environnement sur sa biodiversité et sa composition. À partir de 256 échantillons provenant de 16 individus en bonne santé récoltés à la Forêt Montmorency, nous avons montré que la diversité parmi les quatre dernières cohortes d'aiguilles est similaire avec seulement une augmentation temporaire pour les aiguilles de 2 ans. Chaque cohorte héberge une communauté distincte pour laquelle une hétérogénéité relative a été observée empêchant la définition d'une communauté partagée entre les différentes cohortes. Ceci suggère un processus plus dynamique de colonisation par les endophytes fongiques que celui d'accumulation passive suggeré par les méthodes dépendant de mise en culture. Ces résultats suggèrent que l'endomycobiote foliaire du sapin baumier pourrait être plus sensible à l'environnement que préalablement anticipé et modifier rapidement sa composition pour s'adapter à l'évolution des conditions affectant son hôte.

3.2 Abstract

Over the last 50 years, studies of fungal endophytes have increased, notably in the early 2010s as High-Throughput Sequencing methods emerged. Yet knowledge on endophytology depends mainly on culture-dependent methods and remains to be challenged using culture-independent methods. The aim of this study was to identify the foliar endomycobiota associated with *Abies balsamea* and measure the effect of time of exposure to the environment on both its biodiversity and composition. Based on 256 samples from 16 healthy looking individuals of balsam fir trees collected at 2 locations within the Forêt Montmorency research forest north of Quebec City, we first show that the diversity among the last four cohorts of needles was mainly conserved with only a temporary increase for 2-year old needles, while evenness remained stationary. Then, we demonstrate that each cohort harboured a distinct community, for which a relative heterogeneity was observed, impairing the definition of a core microbiota spread over the different needle ages. This hints towards a more dynamic process of colonization of foliar tissue by fungal endophytes than the accumulative one observed with culture-dependent methods. Together, our results suggest that the foliar endomycobiota of balsam fir trees might be more sensitive to the environment than anticipated and quickly shifts its composition to adapt to changing conditions affecting its host.

3.3 Introduction

Fungal endophytes live within a plant without causing any apparent damage or symptom (Petrini 1991; Wilson 1995; Schultz & Boyle 2005). They are part of the myriad of microorganisms, including fungi (alongside epiphytes, pathogens, and mycorrhizae), bacteria, protists, and viruses, which form the plant microbiota, i.e the microorganisms associated with the interior and surfaces of plants. As such, fungal endophytes play a key role providing their host with means of resistance to both abiotic and biotic stresses. They have been linked with host tolerance to drought, nutritional stress, herbivory, disease, and also to confer host fitness benefits such as plant defence, plant growth enhancement and increased reproductive success (see Rodriguez et al., 2009; Porras-Alfaro & Bayman 2011; Rai & Agarkar 2016; Hardoim et al., 2015 for reviews). Their roles could also extend to the ecosystem as they might also influence plant community structure and ecosystem functioning (Porras-Alfaro & Bayman 2011; Rout 2014). Their importance is such that Rai & Agarkar (2014) qualified them as the soldiers of plants, protecting and maintaining the integrity of the hosts and eventually the ecosystem.

However, fungal endophytes for which the ecological role is known remain the exception, and the term does not describe a particular type of interaction with the plant. For instance, in addition to mutualists, latent pathogens and dormant saprotrophs are also considered as endophytes (Osono 2006; Saikkonen 2007; Newton et al., 2010; Boberg et al., 2011). While it is important to categorize these ecological functions, in order to potentially take advantage of their benefits, a better understanding on how fungal endophytes interact with the plant, among themselves, with other components of the plant microbiota, and with the environment is also needed. With so many parties involved, deciphering such a complex interactions network is challenging. This is accentuated by difficulties in identifying the participants, as the composition of the endomycobiota remains widely unknown for any plant host. A better comprehension of colonization mechanisms and persistence within the plant and through time is also needed in order to potentially use fungal endophytes notably as biocontrol agents. While the endomycobiota is transmitted horizontally, the process is considered to be passive and colonization happens via stomata, lenticels, or wounds before extending to extracellular spaces (Johnston et al., 2006; Saikkonen 2007).

Yet, fungal endophytes show specificity to their host species, and even to the genotype of their host (Ahlholm et al., 2002; Balint et al., 2013; Rajala et al., 2013). Differences in diversity and composition are also observed for foliage endophytes between endophytic- and epiphytic communities despite their immediate proximity

(Santamaria & Bayman, 2005; Osono 2008). Even the delay of colonization differs, with epiphytic colonization being almost immediate after flushing of the foliage, whereas endophytic colonization is virtually non-existent at this same time (Hata et al., 1998; Osono & Mori 2005; Guo et al, 2008; Osono 2008). Disparities even exist between the parts of the leaf with distinct communities between the petiole and the blade (or between the base and the edge of needle) being displayed (Hata & Futai 1996; Hata & Sone 2008). Seasonal variations of diversity and composition of fungal endophyte assemblages have also been observed, as well as an effect of age for evergreen trees (Osono & Mori 2005; Guo et al., 2008; Osono 2008). Diversity tends to increase from the flushing to the senescence of the leaves, and while it is not always simple to distinguish between isolation frequency and richness from culture-depending studies (Espinosa-Garcia & Langenheim 1990; Kowalski 1993; Hata et al., 1998, Osono 2008), diversity is usually considered to increase with age. All these hint towards possibly more complex mechanisms than just a passive colonization and endomycobiota are far from being distributed randomly within space and time (Unterseher et al., 2007).

Conifers are an interesting target for studying the persistence and dynamics of the foliar endomycobiota as they retain their needles for several years (8 to 13 years for *Abies balsamea* (Bakuzis & Hansen 1965; Anderson 1897)). Time of exposure is likely to be accompanied by variation in foliage properties such as water, nitrogen, fibre, and phenolics contents (Hatcher 1990), but also with physiological changes. For instance, the photosynthetic capacities of needles decrease from the second year on (Bazukis & Hansen 1965; Clark 1956), and host defences weaken too with age (Coley 1988; Coley & Barone 1996; Nascimento et al., 2015), notably as the composition of the wax layer on the needle surface changes as needles grow older (Jalkanen et al., 1981). These modifications might affect the needle as an ecological niche and along impact the composition of its microbiota. Moreover, it is likely that colonisation of needles by endophytes alters habitat conditions to develop a less hostile environment within the needles which then turn into a more favourable environment for additional endophytic colonization for specific species or a possible wider range of species (Hata et al., 1998). As needles from different cohorts are subjected to different successive colonization events over time, investigating their endomycobiota separately might allow to obtain a more accurate representation of their diversity than selecting a particular cohort would.

The aim of this study was thus to investigate the foliar endomycobiota of balsam fir (*Abies balsamea* (L.) Mill.), the effect of time of exposure to the environment on its biodiversity and on community composition. To this end, needles from the last four growth periods were sampled and their fungal endophytes identified by using High Throughput Sequencing (HTS) techniques. We anticipated from the literature that the number of species

retrieved would increase with time of exposure to the environment before possibly reaching saturation due to limited colonizable space available. Evenness was monitored in order to quantify the possible establishment of a more specific component of the endomycobiota along the time of exposure. We hypothesized a decrease in evenness as more selective species might colonize the needles to the detriment of more “accidental” colonizers. We also hypothesized that prospecting endomycobiota from different cohorts would allow us to determine a core community grouping fungal endophyte species more specific to this host and which could benefit the plant at the time of the sampling.

3.4 Materials and Methods

3.4.1 Sample preparation, DNA extraction, PCR amplification, and sequencing

Sampling was conducted at the end of August 2012 at the Montmorency research forest (about 70km north of Quebec City) which is located at the southern portion of the North American boreal zone (Brandt 2009). Balsam fir is the main species in this forest. Samples were collected from four different stands, two near the Laverdière site and distant of 150m from each other (Lav01 & Lav02), and two near La Chute and separated by 400m (LaC01 & LaC02). The Laverdière and La Chute sites are distant of about 5km from each other (Table 3.1). At each stand, four trees which did not display any visible damage or symptom of disease were selected and, for each tree, four branches were collected (at 2 and 3m from from the ground and from two opposite directions), resulting overall in 16 trees and 64 branches sampled.

From each branch, we collected 30 symptomless needles for each of the four most recent growth periods, and each cohort was treated independently (a total of 7 680 needles were thus sampled in this study). Samples were surface sterilized through a succession of six baths following Stefani & Bérubé (2006). The first three baths (1.2% sodium hypochlorite for 2 minutes, 76% ethanol for 1 minute, and 19% ethanol solution for 1 minute) were followed by three distilled water baths to rinse the samples. Needles were then cut into 1mm parts under sterile conditions and 100mg served as input for DNA extraction, which was performed with DNeasy® Plant kit (Qiagen GmbH, Hildren, Germany). Needles were ground twice in a Mixer Mill Retsch MM300 (Qiagen) for 2 min at 30Hz using a sterilized 3mm tungsten carbide bead (Qiagen). After addition of 400µL of AP1 extraction buffer, 2µL of RNase A and a tip of Dx antifoaming, samples were ground a third time

for 2 min at 15Hz. DNA was then extracted following the manufacturer's instructions but eluted with 75µL of buffer in two steps (50µL then 25µL).

Prior to amplification of the ITS region, genomic DNA was purified using the Agencourt Ampure XP magnetic clean-up system (Beckman Coulter, Brea, CA, USA) with a volume ratio of 1.8:1 of magnetic beads for genomic DNA. This step was carried out as original amplifications yielded low concentrations, possibly due to the presence of inhibitors (data not shown). Modified primer set ITS1-F (Gardes & Bruns, 1993) and ITS4 (White et al., 1990) was used to amplify the ITS region (ITS1-F featuring the adaptor A and a 10bp Roche Multiplex IDentifier barcode [MID, allowing post sequence analysis], and ITS4 carrying the adaptor B). All samples were amplified six times (using the same MID). The amplification reaction contained 25 µg of Bovine Serum Albumine (BSA, Sigma-Aldrich, St Louis, MO, USA), 1.25 X PCR buffer (Invitrogen, Carlsbad, CA, USA), 2.5 mM of $MgCl_2$, 250 µM of each dNTP, 400 nM of each of the forward and reverse primers, 1 unit of Platinum Taq DNA polymerase (Invitrogen), and 1 µL of template DNA. The PCR program consisted of an initial denaturation at 94 ˚C for 3 min, then 32 cycles of denaturation at 94 ˚C for 30 sec, annealing at 52 ˚C for 30 sec and extension at 72 ˚C for 1 min, followed by a final extension step at 72 ˚C for 10 min.

Reactions of the same samples were pooled and purified using the Agencourt Ampure XP magnetic PCR clean-up system. Primer-dimers were eliminated using a 1.8:1 volume ratio of magnetic beads for amplicons. In order to discard fragments smaller than 150-200 bp, a second purification with a 0.6:1 volume ratio was used. Elution in 26 µL of EB buffer (Qiagen), from the original 110 µL of amplicons, increased the concentrations which were measured using the Quant-iT Picogreen dsDNA assay kit (Invitrogen, Eugene, OR, USA). Four amplicon libraries (one for each cohort) composed of samples in equimolar proportions were built. For each library, amplicon length was verified using an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Unidirectional sequencing was performed with Roche GS FLX+ system (Roche-454 Life Sciences, Brandford, CT, USA) at McGill University and Genome Quebec Innovation Centre (Montreal, QC, Canada).

3.4.2 Data processing and statistical analyses

Analyses were performed on the ITS sub-fragment delimited by the primer ITS1-F and the oligonucleotide ITS3a (CTTTCAACAACGGATCTCTT (Ponchart et al., 2019)). Conserving a portion of the ribosomal small

subunit (pSSU) in front of the full length ITS1 sub-locus enhances detection of putative chimeras for 454 pyrosequencing data and reduces the formation of rare possible spurious mOTUs (Ponchart et al., 2019). Denoising was performed using the mothur (v1.33.1, Schloss et al., 2009) implementation of the PyroNoise algorithm (Quince et al., 2009) with the mindelta parameter set on 10^{-6} rather than for 1000 iterations. Following the denoising, the presence of the ITS3a pattern with 3 mismatches allowed was assessed with fqgrep (ver 0.4.4, Indraniel et al., 2016) and the -c parameter was used to highlight its position. Shell sed commands based on the colour tag in front of the pattern were used to remove every nucleotide from that point resulting in sequences only displaying the pSSU-ITS1 fragment. All sequences which did not display this pattern were run through ITSx (Bengtsson-Palme et al., 2013) targeting fungal sequences. Only sequences with both ITS1 and 5.8S detected were conserved. Partial SSU were retrieved to match the same sub-fragment of the amplicons (post ITS1-F – ante ITS3a). All sequences containing any mismatch to the primer and the barcode, any ambiguous base call, or any homopolymer repeat longer than 9bp were removed.

To take better advantage of the sequencing depth later on, needles of the same cohort from the same height of a tree were combined yielding 128 sample-units. Putative chimeric sequences were then assessed with chimera.perseus on mothur and discarded. All unique sequences were processed with ESPRIT (Sun et al., 2009) to perform pairwise alignments. Using the single-linkage preclustering (SLP) algorithm developed by Huse et al., (2010), the distance matrix was processed to a precluster step with a width of 0.03. The clustering into molecular Operational Taxonomic Units (mOTUs) was done at a 95% sequence similarity threshold with mothur using the average neighbour method. Representative sequences of each fungal mOTU were retrieved from the preclustering step and all unique sequences were conserved and deposited in GenBank (Submission number: SUB5671995).

The above sequences were then blasted against the NCBI nt database (downloaded on September 24th, 2016). Kingdom assignment was assessed using the Lowest Common Ancestor algorithms featured in Megan (Huson et al., 2007). The following parameters were used: min support=1, minscore=200, top%=10, min comp=0.3, %ID filter (U'Ren et al., 2014). All mOTUs which were not directly assigned to the fungal kingdom were manually checked to ensure their taxonomic affiliation and all non fungal mOTUs were removed. The implementation in mothur of the rdp method (Wang et al., 2007) allowed refining the taxonomy with the classify.seqs command. The v7 of Unite from November 20th, 2016 served as template (UNITEv6_sh_99_s.fasta and UNITEv6_sh_99_s.tax). A cutoff of 60, 1000 iterations and the default kmer size of 8 were used as parameters. For the class level, the compilation of the assignment was based on a

bootstrap value support of 80. When the class was mentioned as *incertae sedis*, the assignment was made to the next lowest level supported by a bootstrap value of 90 at least or to a higher level otherwise.

Species-accumulation and rarefaction curves were calculated with mothur, sampling every 1000 sequences, and represented with the R program version 3.2.3 (R Core Development Team, 2015). Extrapolation of the richness was evaluated using the specpool command of the vegan package (Oksanen 2007; Smith and van Belle, 1984) with the bootstrap estimate. Singletons and doubletons were then excluded prior to the following analyses. To ensure the validity of sample comparisons, we sub-sampled to select randomly 1000 sequences per samples. Four samples were removed as they contained fewer sequences. One hundred sub-samplings were performed to reduce the stochasticity of the observations. Diversity was measured using Fisher's alpha (Fisher et al., 1943), whereas evenness was measured with the Pielou's evenness index (Pielou 1966). Both metrics were computed with the vegan R package. Normality was assessed using Shapiro-Wilk test. A parametric one-way ANOVA coupled with a Tukey's range test was used to evaluate the relation of stands, and needle-age to the diversity. A non-parametric Kruskal-Wallis test followed by a Dwass Steele Critchlow Fligner test (DSCF ran with the npar1way procedure in SAS 9.4 (SAS Institute Inc., 2012)) was used for the same purposes for evenness. Unless otherwise noted, all statistical analyses were performed in R.

We used one-way ANOSIM (Analyses of similarity) to compare community structure. Analyses were based on Jaccard's index (Clarke 1993) and performed in PAST (Hammer et al., 2001). Results were visualized with non-metric multidimensional scaling (nMDS) in PAST (Taguchi & Oono, 2004) and replotted in R. Permutational multivariate analysis of variance (PERMANOVA (Anderson 2001)), implemented with the adonis function from the vegan R package, was used to quantify sources of variation in the community structure (using Jaccard dissimilarity measure, with 999 permutations). To visualize homogeneity in the endomycobiota structure, correlation-based network analysis was used to infer the co-occurrence patterns of sample-units based on their mOTU composition. Singletons and doubletons were removed from the original sample-by-mOTU feature table and the sequence counts were replaced by relative abundance per sample. This substitution allowed to account for the difference in sequencing depth among samples (Ju et al., 2014). mOTUs occurring in fewer than 3 samples were discarded to reduce the network complexity through decreasing sample-specificity. Pairwise Spearman's rank correlations were calculated using Hmisc package (Harrell 2008) in R. Only correlations with a Spearman's correlation coefficient (ρ) > 0.6 and a P-value < 0.01 were considered. Visualization of the network and detection of its modules were conducted using Cytoscape

(ver 3.3.0, Shannon et al., 2003). The same analysis was performed to infer mOTU co-occurrence patterns based on their distribution among samples.

3.5 Results

Sequencing of the libraries produced 658 346 reads; pSSU-ITS1 fragment was extracted for 564 260 trimmed sequences. After removal of putative chimeras (88 233 sequences) and non fungi (93 sequences), 475 934 sequences, with a mean length of 245 bp, were considered for the clustering which resulted in 2 318 mOTUs (including 180 doubletons and 983 singletons). Saturation was not observed for the dataset (Figure S3.1). Extrapolation of the richness was estimated to 2 900 ± 66 mOTUs, indicating that ca. 80% of the expected richness was collected. The number of mOTUs resulting from the 100 sub-samplings ranged from 1 006 to 1 070 with a mean value of 1 042.32 mOTUs (±9.16). All but 4 of 1 155 mOTUs composed of at least 3 sequences were represented among the sub-samplings (each specific to the samples discarded).

Diversity slightly differed by stand (F=5.299, p-value: 1.825×10^{-3}). Only the two stands with extreme values (Laverdiere-01 for highest diversity, and La Chute-01 for lowest diversity) had distinct diversities, but both shared similar diversity with the two remaining stands (Laverdiere-02, and La Chute-02) (Figure 3.2.a). Variation in diversity could also be attributed to needle age (F=6.5182, p-value: 4.019×10^{-4}), but only because two-year old needles displayed higher diversity (Figure 3.2.b). No difference in evenness was observed neither between stands (χ^2=3.3924, p-value: 0.335) nor between needle ages (χ^2=0.9791, p-value: 0.8063) (Figure 3.2.c and 3.2.d). Fungal communities differed in composition among stands (Figure 3.3.a), with each stand having distinct communities for 95 of the sub-samplings, but Laverdiere-02 and La Chute-01 displaying a similar community for 5 of the 100 sub-samplings. Differences in fungal composition were also observed among needle ages (Figure 3.3.b), with each cohort presenting distinct communities for 60 of the sub-samplings, but with current year and 1-year-old needles sharing the same communities for 40 of the sub-samplings. Both stands and needle ages explained a similar portion of variation in fungal communities (5.6% for stands, and 6.5% for needle-ages) (Table 3.2).

The 1 155 mOTUs composed of at least 3 sequences were taxonomically assigned among four phyla. Most belonged to the Dikarya (663 mOTUs affiliated to the Ascomycota, and 208 to the Basidiomycota), whereas

“Zygomycota” and Chytridiomycota were represented by 7 and 1 mOTUs respectively. About 24% of the mOTUs could be identified only at the level of Fungi, and 20.4% could not be attributed to a level lower than phylum. Ascomycota were distributed among 11 classes and 1 family (10 mOTUs were attributed to a level lower than class), Basidiomycota among 9 classes and 1 order, “Zygomycota” within 1 order, and Chytridiomycota within 1 class. Usual foliar endophyte classes dominated the Ascomycota, notably Dothideomycetes (152 mOTUs), Leotiomycetes, Sordariomycetes, and Eurotiomycetes (92, 82 and 34 respectively). Lecanoromycetes were also relatively well represented with 25 mOTUs. Agaricomycetes and Tremellomycetes dominated the Basidiomycota (99 and 35 mOTUs respectively).

Interestingly, while mOTU composition differed among sites and needle ages, the taxonomic structure and relative abundance of the different phyla and classes of these communities were globally conserved (Table 3.3, Figure 3.4). Most of the differences observed (global proportion of Ascomycota, Leotiomycetes, Eurotiomycetes, Lecanoromycetes, Pezizomycetes, and unclassified Ascomycota) were due to variation among stands rather than between the two sites (Laverdière vs La Chute) with the exception of the Agaricomycetes which occurred in lower proportion at La Chute. For the needle ages, the differences observed (Orbiliomycetes, Exobasidiomycetes, Pseudoveronaea) did not appear to be related with the ageing process of the tissue, except for Leotiomycetes whose relative abundance tended to increase with needle age, and Pezizomycetes which displayed the opposite tendency.

Analysis of sample co-occurrence (based on mOTUs composition) revealed a relative intra-group heterogeneity (either for samples from the same stand, or from the same cohort). The resulting network was composed of 14 nodes (samples) engaged in 23 edges and organized around a main module (12 nodes – 22 edges) and one isolated pair (data not shown). No clear preference in association (either same site, stand, or needle age) was detected.

Of the 595 mOTUs composed of at least 3 sequences and present in at least 3 samples which were retained for the co-occurrence analysis, 68 mOTUs established 98 pairwise correlations ($\rho > 0.6$ and a P-value < 0.01). About half of them represented isolated correlations forming pairs (34 mOTUs in 17 edges) and one trio (3 mOTUs in 3 edges) (data not shown). The remaining 31 mOTUs established 88 edges and were organized in two modules (Figure 3.5). The first module was composed of 15 mOTUs and 45 edges. While it appears more

dense and composed of more cosmopolite mOTUs (retrieved from 65 to 128 samples), no clear structure could be observed (either relative to stand or needle age). Blasts of their representative sequences against the NCBI nt database were inconclusive as 10 of the 15 mOTUs did not display any hits (based on a coverage of at least 90% of the sequence and similarity of at least 97%). The second module, composed of 16 mOTUs and 33 edges was more disparate (mOTUs were retrieved from 3 to 26 samples) but no clear structure was observed either. About half of these mOTUs did not display any hits when blasted against the nt database for the same coverage and similarity mentioned previously (7 of the 16 mOTUs).

3.6 Discussion

The main goal of this study was to survey the foliar endomycobiota of balsam fir trees and evaluate the possible difference in diversity and composition among the last four cohorts of needles. Existing data, as it is often the case in endophytology, come mostly from culture-dependent methods which have established that needle age, or at least time of exposure to the environment, might have an influence on diversity, with richness tending to increase with leaf age (Espinosa-Garcia and Langenheim 1990; Hata et al., 1998; Arnold and Herre. 2003; Osono 2008, Nascimento et al., 2015). However, as the majority of endophytic fungi are not amenable to culture, the conclusions of these studies rely on a fraction of the communities. Use of HTS, which usually reveals a more extensive portion of these communities, might help either nuance or confirm these observations. Yet, it has to be noted that not all species detected by culturing are recovered by molecularly-based methods (Arnold et al., 2007), and that HTS themselves also display some biases. They are notably known for generating a large proportion of rare taxa considered as possibly spurious for the most part. Moreover HTS detect DNA of species present in the sample but not necessarily “active” or interacting with the plant species.

With the exception of 2-year-old needles which displayed a higher diversity, no effect of age on diversity was detected. Endophytic colonization of woody plants is considered to occur by horizontal transmission (Arnold & Herre 2003; Saikkonen et al., 2004; Guo et al., 2008; Nascimento 2015), and needles virtually contain no endophytes at an early stage of their growth (Hata et al., 1998; Guo et al., 2008). Yet, and despite the fact needle elongation of balsam fir continues into the second season (Bakuzis & Hansen 1965), current year needles which started flushing around mid-June at the Forêt Montmorency site (thus ca. 2 months prior to sampling) had a similar diversity as 1- and 3-year-old needles which had been exposed to different seasonal

colonizations. This result is not completely in opposition with results from culture-dependent studies as Arnold and Herre (2003) demonstrated for endophyte-free seedlings of *Theobroma cacao* that it is not so much the absolute leaf age as the time of exposure to the environment that influences the density of endophyte infection in spite of differences in leaf chemistry and leaf toughness. However, the time scale in this study is more extended, and possible differences in terms of physiology, chemistry, and host defence between the different needle cohorts might be more marked and should not be completely dismissed.

Notwithstanding 2-year-old needles, it seems likely that fungal endophytes could readily colonize the extent of accessible space within the foliage. Colonization of endophytes is thought to be passive and through the stomata, before expanding to the intercellular spaces. Given the limited volume, it is plausible that a two-month-long exposure could be sufficient to reach a quasi complete colonization of the available space. The 2-year-old cohort could constitute a turning point for the fungal community as the photosynthetic capacity of balsam fir needles starts to decline at 2 years (Clarke 1956; Bakuzis & Hansen 1965). This physiological change could be accompanied by others (notably in plant defence), and favour colonization by dormant saprophytic fungi or latent pathogens over mutualists. This turn-over could explain a temporary increase of the diversity but due to the limitations of reference databases, taxonomic identity is usually difficult to attribute and with it, ecological roles of the species detected remain unknown.

Similar evenness among the needle cohorts and relative conservation in the diversity suggest that either the community is fixed upon initial colonization or that the evolution of its dynamics might be better observed at a shorter time scale. The latter is more plausible as both community composition and isolation frequencies of endophytic fungi have been shown to change with seasonal variations (Hata et al., 1998; Unterseher et al., 2007; Guo et al., 2008). Prolonged time of exposure to the environment, change in physiology and/or physical alterations (microscopic wounds) of the needles are usually cited to explain these variations. Fungus-fungus interactions might also alter the habitat conditions for the endophytes (Hata et al., 1998; Terhonen 2011).

Fungal endophytes are known to provide their host with tolerance to both abiotic and biotic stresses (drought, humidity, insolation intensity, and insect attacks for example). As the environmental conditions vary along the year, so are the intensities of these different stresses. The recruitment of the endophytic community might be tuned to better answer these temporary pressures leading to an evolution in the composition of the

endomycobiota to reflect the plant needs. This could join up with the hologenome theory which hypothesizes that a given host would rely on its micro-organisms to adapt quicker to a selective pressure (Rosenberg et al., 2007; Rout 2014; Bordenstein & Theis 2015). Rather than to be accumulated (the diversity we estimated on balsam fir remained globally stable between the different cohorts of needles), it is likely that fungal species with different ecological roles partially supersede each other, re-balancing the community as a result of the competition between settled fungi and new colonizers. Such dynamics have been observed for substrate competition in wood decaying fungi, notably with *Resinium bicolor* being able to establish in, and exclude other species from inoculated wood blocks (Holmer and Stenlid 1996). It also has to be noted that as evenness relies on the abundance of the mOTUs, conclusions should be considered carefully as Amend et al., (2010) demonstrated that abundance of reads from pyrosequencing is semi-quantitative.

Each cohort displayed a distinctive community, but current and one-year-old needles displayed a mixed result, with their communities being different for 60 of the 100 sub-samplings, but identical for the remaining 40 sub-samplings. Comparative data on the composition of communities among different leaf-age categories are difficult to compile. While this topic has been studied, notably with culture-dependent methods, focus is usually confined to the isolation rate of the few most frequent species, rather than on comparison of overall communities (Espinosa-Garcia & Langenheim 1990; Kowalski 1993; Hata et al., 1998, Osono 2008). Because retrieval of species by culturing methods yields a relatively low number of species, community compositions are usually listed but, if differences do appear, they usually resulted from few sporadic species scattered among different leaf ages which impair the assessment of a possible difference of the overall community among different ages.

A certain variation within group (either from the same cohort or same stand) was observed notably from the nMDS, and the co-occurrence analysis did not allow isolating a module that could be easily linked to either a specific cohort or a stand. This could be due to a relatively low number of samples which impaired the collection of a sufficiently representative community for each category. It has to be noted that while endophytic communities show specificity to their plant host species, studies have also demonstrated that different genotypes from a same plant species displayed distinctive communities (Ahlholm et al., 2002; Balint et al., 2013; Rajala et al., 2013). The possible presence of different *Abies balsamea* genotypes (a parameter that we did not evaluate) among the stands might then explain a relative disparity. Also, homogeneity of the endomycobiota within a single host species or even within a particular tissue of that host is yet to be assessed.

Yet some variability has been observed notably depending on the position in the canopy and light availability (Unterseher et al., 2007; Saikkonen 2007).

Co-occurrence analysis did not allow determining a core-community spread over the different cohorts. This result, together with the intra-group variations, relatively conserved diversity and stable evenness across the different cohorts, and distinct community composition, appears to contradict results from culture-dependent methods in which older leaves display higher diversity than younger ones (Espinosa-Garcia & Langenheim 1990; Osono 2008). Conclusions from these studies tend to suggest an accumulative process of infection by fungal endophytes due to a longer time of exposure to the environment and successional colonization events. On the contrary, our results suggest a more dynamic process with at least a partial turn-over in the composition of the foliar endomycobiota over time. This possibility is reinforced by the seasonal variations reported in other studies of conifer needles (Hata et al., 1998; Guo et al., 2008) and deciduous foliage (Unterseher et al., 2007) which hint towards a possible delicately balanced adaptation of the endomycobiota to environmental conditions where advantageous species for the host at a given time might be recruited. However, our study design precludes definitive conclusion, and a more adapted sampling strategy monitoring regularly at a smaller time scale (semimonthly, for instance) targeted cohorts of needles over several years might not only allow to capture a more extensive diversity of the endomycobiota, but also to develop a better comprehension of its dynamics.

3.7 Conclusion

While endophytology studies concentrate mainly on leaf-associated fungal communities as they represent a biodiversity hotspot, understanding of colonization mechanisms and persistence of the fungal endophytes within leaves and through time remains fragmentary. The literature, notably reports based on culture-dependent methods, suggests a relatively straightforward accumulative process by which diversity tends to increase from the flushing to the senescence of the leaves. Our results based on metabarcoding of the endomycobiota associated with the last four cohorts of balsam fir needles suggest a more dynamic process. The diversity was relatively well conserved among the four cohorts implying that the two months of exposure of current year needles to the environment was sufficient to reach similar diversity as in older needles. As evenness was stable among the four cohorts, this excludes the accumulation of a specific portion of the endomycobiota over time. Distinct communities were also harboured by the four cohorts. Together with the

seasonal variation in the composition of fungal endophyte community observed in other studies, these results hint toward a constant evolution of the endomycobiota composition through time, with at least a partial replacement of the settled community. The design of our study resulted in the comparison of needle cohorts which differed in one-year increments in time of exposure to the environment. In order to develop a better understanding of the dynamics of the colonization and recruitment mechanisms of the endomycobiota, this period will have to be shortened considerably. A regular follow-up of the endomycobiota of the same cohort over several years might also prove to be an interesting monitoring strategy. The focus of such studies will also have to be widened to include more coniferous species on a more global scale. The search for the hidden fungal diversity might also benefit from considering different cohorts rather than a single one as we found that they harboured distinct communities.

3.8 Acknowledgements

We kindly thank Julie Boudreau from McGill University and Genome Quebec Innovation Centre (Montreal, Qc, Canada) for technical assistance on 454 pyrosequencing. We are indebted to Pr Elizabeth A. Arnold and Dr Jana U'Ren (University of Arizona, Tucson, Az, USA) for useful discussions on pyrosequencing data processing and their great help on ecological analysis. We also thank Patrick Gagné (Université Laval, Qc, Canada) for feedbacks on bioinformatic analysis.

3.9 References

Ahlholm, J. U., Helander, M., Henriksson, J., Metzler, M., & Saikkonen, K. (2002). Environmental conditions and host genotype direct genetic diversity of Venturia ditricha, a fungal endophyte of birch trees. *Evolution*, *56*(8), 1566-1573.

Amend, A. S., Seifert, K. A., & Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count?. *Molecular ecology*, *19*(24), 5555-5565.

Anderson, A. P. (1897). Comparative anatomy of the normal and diseased organs of Abies balsamea affected with *Aecidium elatinum*. *Botanical Gazette*, *24*(5), 309-344.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, *26*(1), 32-46.

Arnold, A. E., & Herre, E. A. (2003). Canopy cover and leaf age affect colonization by tropical fungal endophytes: ecological pattern and process in *Theobroma cacao* (Malvaceae). *Mycologia*, *95*(3), 388-398.

Arnold, A. E., Henk, D. A., Eells, R. L., Lutzoni, F., & Vilgalys, R. (2007). Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia*, *99*(2), 185-206.

Bálint, M., Tiffin, P., Hallström, B., O'Hara, R. B., Olson, M. S., Fankhauser, J. D., ... & Schmitt, I. (2013). Host genotype shapes the foliar fungal microbiome of balsam poplar (*Populus balsamifera*). *PLoS One*, *8*(1), e53987.

Bazukis, E., Hansen, H., (1965). *Balsam Fir: A Monographic Review*. University of Minnesota Press.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... & Amend, A. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution*, *4*(10), 914-919.

Boberg, J. B., Ihrmark, K., & Lindahl, B. D. (2011). Decomposing capacity of fungi commonly detected in *Pinus sylvestris* needle litter. *Fungal Ecology*, *4*(1), 110-114.

Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biology*, *13*(8), e1002226.

Brandt, J. P. (2009). The extent of the North American boreal zone. *Environmental Reviews*, *17*(NA), 101-161.

Clark, J. (1956). Photosynthesis of white spruce and balsam fir. Bi-m. Progr. Rep. Div. For. Biol. Dep. Agric. Can. 12(5):1-2.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, *18*(1), 117-143.

Coley, P. D. (1988). Effects of plant growth rate and leaf lifetime on the amount and type of anti-herbivore defense. *Oecologia*, *74*(4), 531-536.

Coley, P. D., & Barone, J. A. (1996). Herbivory and plant defenses in tropical forests. *Annual review of ecology and systematics*, *27*(1), 305-335.

Espinosa-Garcia, F. J., & Langenheim, J. H. (1990). The endophytic fungal community in leaves of a coastal redwood population diversity and spatial patterns. *New Phytologist*, *116*(1), 89-97.

Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 42-58.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*, *2*(2), 113-118.

Guo, L. D., Huang, G. R., & Wang, Y. (2008). Seasonal and tissue age influences on endophytic fungi of *Pinus tabulaeformis* (Pinaceae) in the Dongling Mountains, Beijing. *Journal of integrative plant biology*, *50*(8), 997-1003.

Hammer, Ř., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis–Palaeontol. Electron. 4: 9pp.

Hardoim, P. R., Van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., ... & Sessitsch, A. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293-320.

Harrell, Frank E. 2008. "Hmisc: Harrell Miscellaneous." R package version 3.4-4.

Hata, K., & Futai, K. (1996). Variation in fungal endophyte populations in needles of the genus Pinus. *Canadian Journal of Botany*, *74*(1), 103-114.

Hata, K., Futai, K., & Tsuda, M. (1998). Seasonal and needle age-dependent changes of the endophytic mycobiota in *Pinus thunbergii* and *Pinus densiflora* needles. *Canadian Journal of Botany*, *76*(2), 245-250.

Hata, K., & Sone, K. (2008). Isolation of endophytes from leaves of *Neolitsea sericea* in broadleaf and conifer stands. *Mycoscience*, *49*(4), 229-232.

Hatcher, P. E. (1990). Seasonal and age-related variation in the needle quality of five conifer species. *Oecologia*, *85*(2), 200-212.

Holmer, L., & Stenlid, J. (1996). Diffuse competition for heterogeneous substrate in soil among six species of wood-decomposing basidiomycetes. *Oecologia*, *106*(4), 531-538.

Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, *12*(7), 1889-1898.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, *17*(3), 377-386.

Indraniel et al.. (2016). fqgrep: v0.4.4. Zenodo. 10.5281/zenodo.45105

Jalkanen, R., Huttunen, S., & Väisänen, T. (1981). The wax structure of the developing needles of *Pinus sylvestris* progenies infected by *Lophodermella sulcigena*.

Johnston, P. R., Sutherland, P. W., & Joshee, S. (2006). Visualising endophytic fungi within leaves by detection of (1→ 3)-ß-d-glucans in fungal cell walls. *Mycologist*, *20*(4), 159-162.

Ju, F., Xia, Y., Guo, F., Wang, Z., & Zhang, T. (2014). Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed waste water treatment plants. *Environmental microbiology*, *16*(8), 2421-2432.

Kowalski, T. (1993). Fungi in living symptomless needles of *Pinus sylvestris* with respect to some observed disease processes. *Journal of phytopathology*, *139*(2), 129-145.

Nascimento, T. L., Oki, Y., Lima, D. M. M., Almeida-Cortez, J. S., Fernandes, G. W., & Souza-Motta, C. M. (2015). Biodiversity of endophytic fungi in different leaf ages of *Calotropis procera* and their antimicrobial activity. *Fungal ecology*, *14*, 79-86.

Newton, A. C., Fitt, B. D., Atkins, S. D., Walters, D. R., & Daniell, T. J. (2010). Pathogenesis, parasitism and mutualism in the trophic space of microbe–plant interactions. *Trends in microbiology*, *18*(8), 365-373.

Oksanen, J., R. Kindt, P. Legendre, B. O ' Hara, and M. H. H. Stevens. (2007). Vegan: Community ecology package. R package, version 1.8 – 8

Osono, T., & Mori, A. (2005). Seasonal and leaf age-dependent changes in occurrence of phyllosphere fungi of giant dogwood. *Mycoscience*, *46*(5), 273-279.

Osono, T. (2006). Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. *Canadian Journal of Microbiology*, *52*(8), 701-716.

Osono, T. (2008). Endophytic and epiphytic phyllosphere fungi of Camellia japonica: seasonal and leaf age-dependent variations. *Mycologia*, *100*(3), 387-391.

Petrini, O. (1991). Fungal endophytes of tree leaves. In *Microbial ecology of leaves* (pp. 179-197). Springer, New York, NY.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, *13*, 131-144.

Ponchart, J. P., Bernier, L., Bérubé, J. A. (2019). SSU to the rescue: conserving the original partial SSU fragment ahead of the ITS1 sub-locus enhances the detection of putative chimeras and limit the formation of possible spurious rare taxa in fungal metabarcoding.

Porras-Alfaro, A., & Bayman, P. (2011). Hidden fungi, emergent properties: endophytes and microbiomes. *Annual review of phytopathology*, *49*.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., ... & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, *6*(9), 639-641.

R Core Team (2015). R: A language and environment for statistical computing.

Rai, M., & Agarkar, G. (2016). Plant–fungal interactions: What triggers the fungi to switch among lifestyles?. *Critical reviews in microbiology*, *42*(3), 428-438.

Rajala, T., Velmala, S. M., Tuomivirta, T., Haapanen, M., Müller, M., & Pennanen, T. (2013). Endophyte communities vary in the needles of Norway spruce clones. *Fungal biology*, *117*(3), 182-190.

Rodriguez, R. J., White Jr, J. F., Arnold, A. E., & Redman, R. S. (2009). Fungal endophytes: diversity and functional roles. *New phytologist*, *182*(2), 314-330.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., & Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nature reviews. Microbiology*, *5*(5), 355.

Rout, M. E. (2014). The plant microbiome. *Genomes of Herbaceous Land Plants*, *69*, 279-309.

Saikkonen, K., Wäli, P., Helander, M., & Faeth, S. H. (2004). Evolution of endophyte–plant symbioses. *Trends in plant science*, *9*(6), 275-280.

Saikkonen, K. (2007). Forest structure and fungal endophytes. *Fungal Biology Reviews*, *21*(2-3), 67-74.

Santamaría, J., & Bayman, P. (2005). Fungal epiphytes and endophytes of coffee leaves (Coffea arabica). *Microbial Ecology*, *50*(1), 1-8.

SAS Institute Inc., 2012

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, *75*(23), 7537-7541.

Schulz, B., & Boyle, C. (2005). The endophytic continuum. *Mycological research*, *109*(6), 661-686.

Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

Smith, E.P & van Belle, G. (1984). Nonparametric estimation of species richness. Biometrics 40, 119–129.

Stefani, F. O. P., & Bérubé, J. A. (2006). Biodiversity of foliar fungal endophytes in white spruce (*Picea glauca*) from southern Québec. *Botany*, *84*(5), 777-790.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M. L., McKendree, W., & Farmerie, W. (2009). ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic acids research*, *37*(10), e76-e76.

Taguchi, Y. H., & Oono, Y. (2004). Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics*, *21*(6), 730-740.

Terhonen, E., Marco, T., Sun, H., Jalkanen, R., Kasanen, R., Vuorinen, M., & Asiegbu, F. (2011). The effect of latitude, season and needle age on the mycota of Scots pine (*Pinus sylvestris*) in Finland.

Unterseher, M., Reiher, A., Finstermeier, K., Otto, P., & Morawetz, W. (2007). Species richness and distribution patterns of leaf-inhabiting endophytic fungi in a temperate forest canopy. *Mycological Progress*, *6*(3), 201-212.

U'Ren, J. M., Riddle, J. M., Monacell, J. T., Carbone, I., Miadlikowska, J., & Arnold, A. E. (2014). Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Molecular Ecology Resources*, *14*(5), 1032-1048.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, *73*(16), 5261-5267.

White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications*, *18*(1), 315-322.

Wilson, D. (1995). Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 274-276.

Figure 3.1: Species accumulation curves (solid lines) and bootstrap estimate (long-dashed lines) of species richness for the full dataset. Dotted lines represent the 95% confidence intervals.

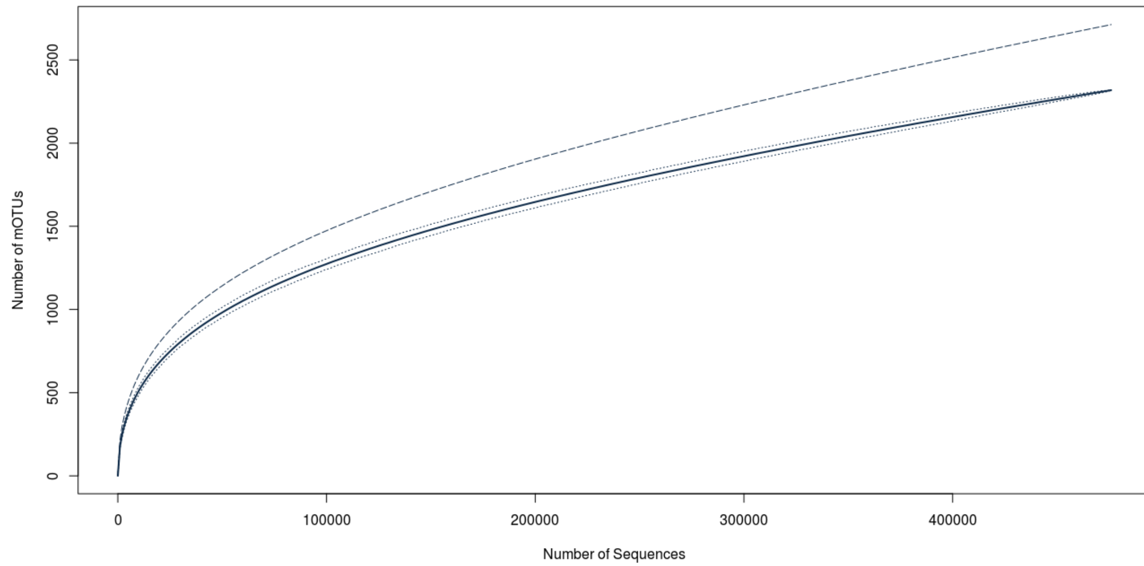


Figure 3.2: Results of diversity (A, B) and evenness (C, D) analyses results among the stands (A, C) and needle cohorts (B, D). Different letters on the top of the box plots represent significant differences after Kruskal-Wallis and DSCF tests or after ANOVA and Tukey tests with $p \leq 0.05$.

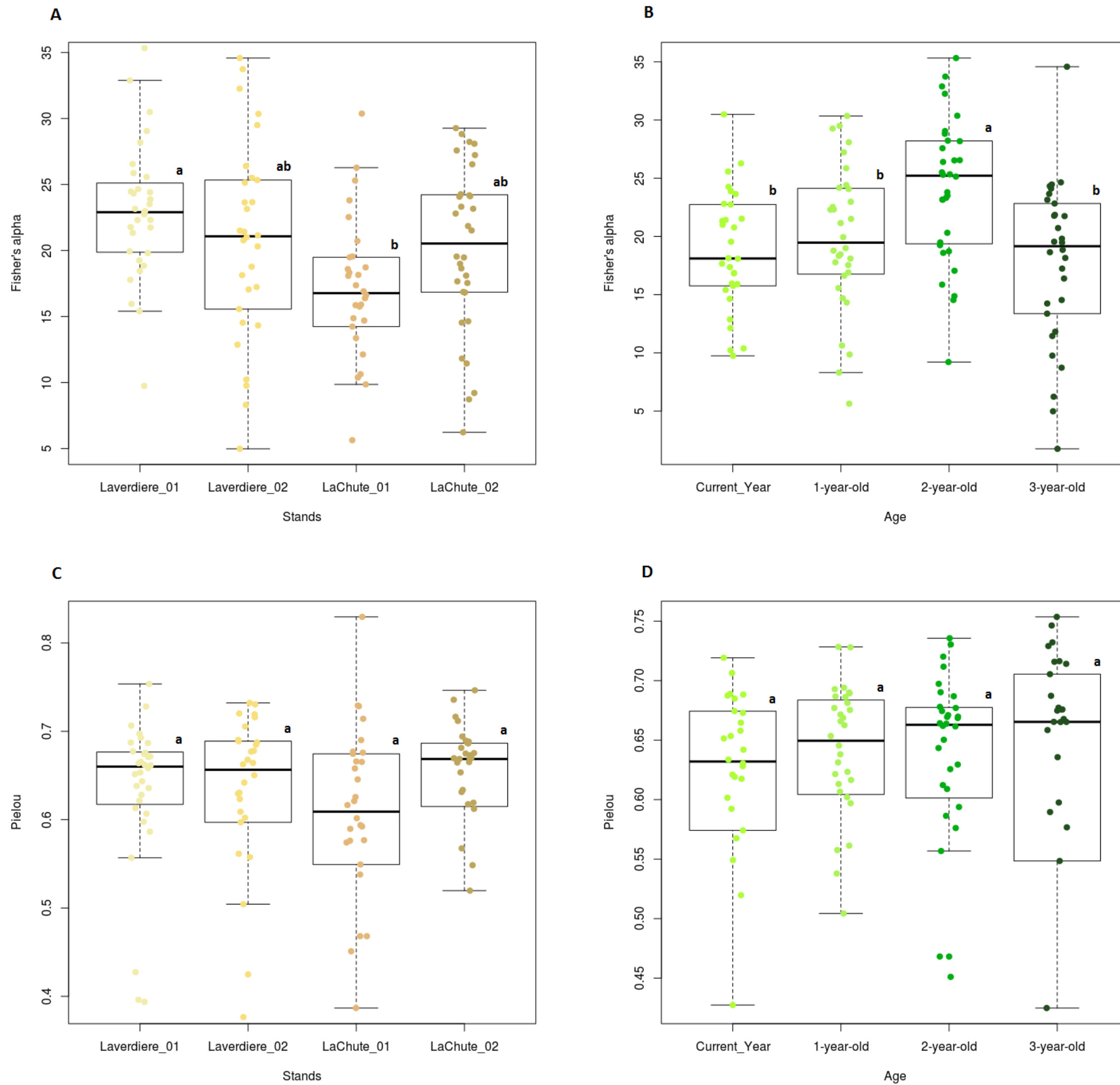


Figure 3.3: Non-metric multidimensional scaling (nMDS) plots of endophytic fungal communities, showing the differences in community composition among the stands (A) (circles: Laverdiere 01, squares: Laverdiere 02, triangles: La Chute 01, diamonds: La Chute 02) and the needle cohorts (B) (circles: current year, squares: 1-year-old, triangles: 2-year-old, diamonds: 3-year-old). One-way analysis of similarity (ANOSIM) indicates significant differences ($P < 0.05$) in community composition both among the stands and the cohorts. The Jaccard index was used to quantify community similarity for nMDS and ANOSIM.

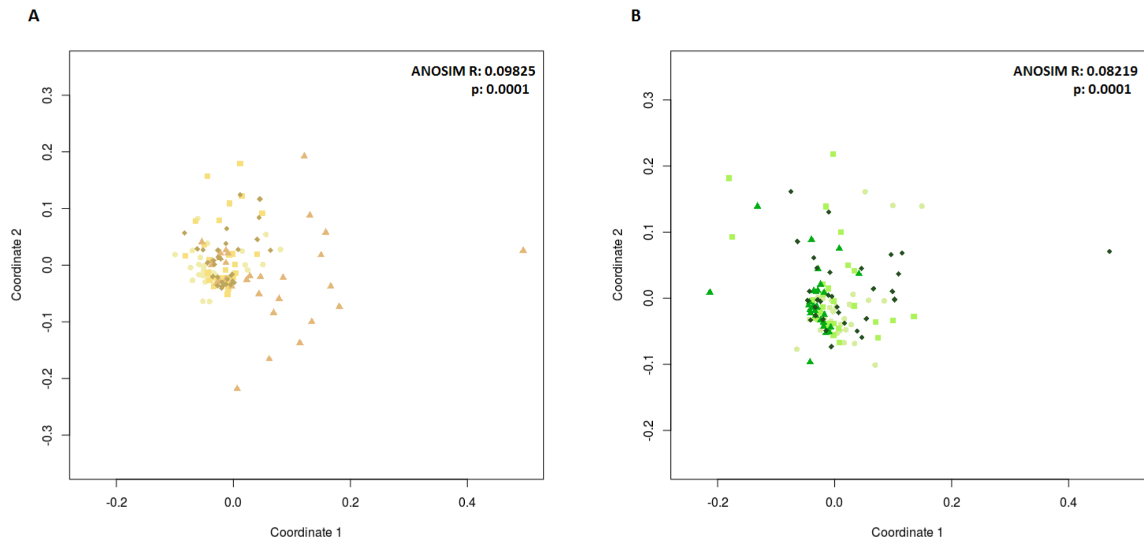


Figure 3.4: Relative abundance plots of fungal endophyte communities by stand (A) and needle cohort (B)

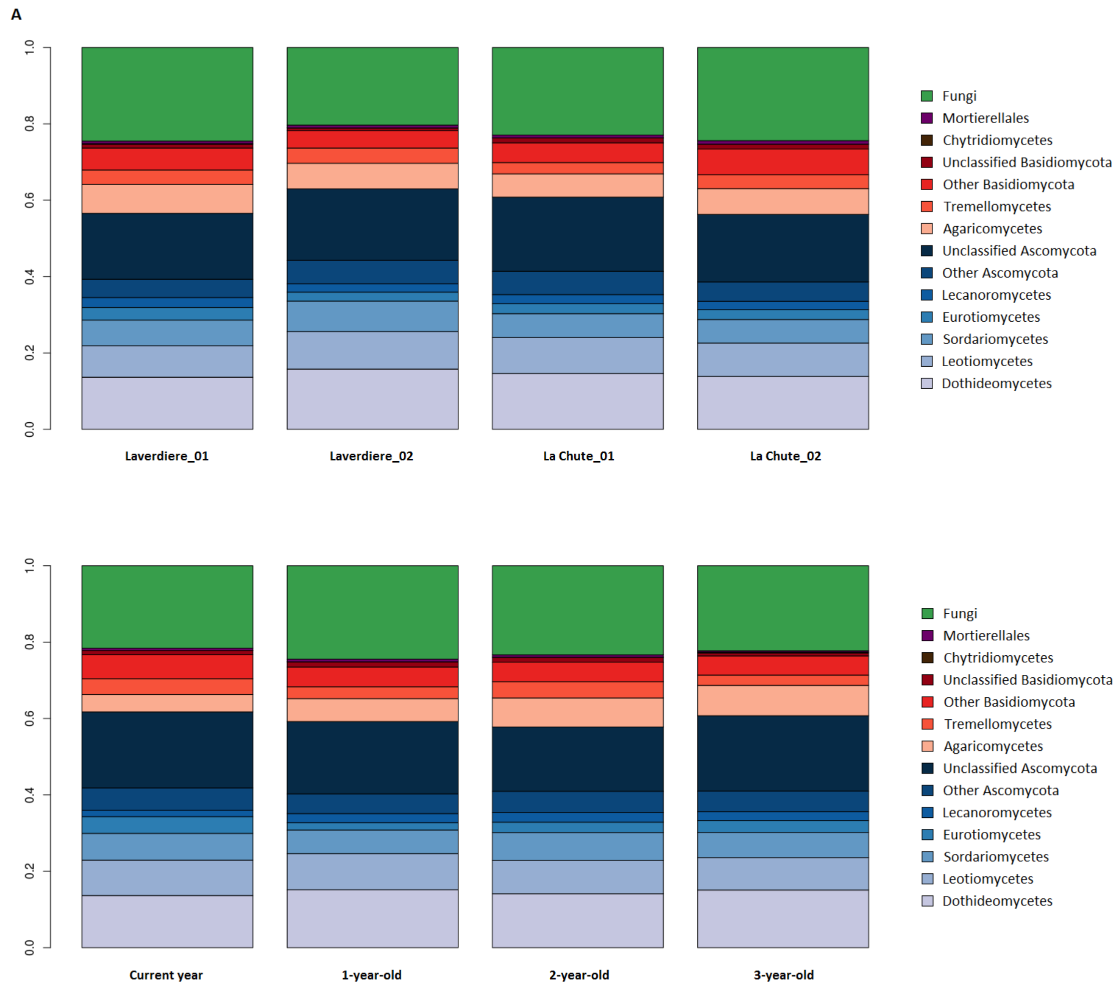


Figure 3.5: mOTUs co-occurrence network displaying two main modules. Pairwise Spearman's rank correlations were considered if $\rho > 0.6$, and P-value < 0.01.

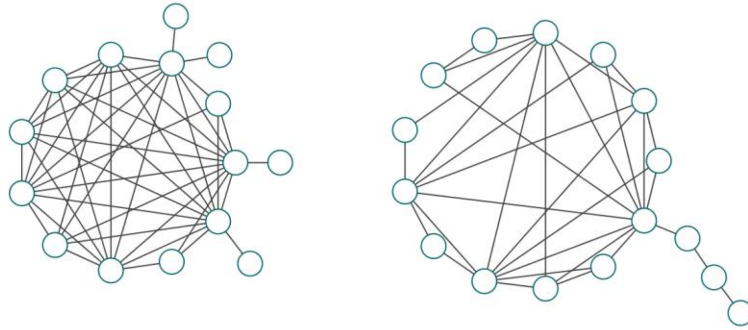


Table 3.1: Characteristics of Forêt Montmorency sites sampled in August 2012

	Laverdière		La Chute	
	Lav1	Lav2	LaC1	LaC2
Latitude	47°18.983' N	47°18.906' N	47°20.071' N	47°19.861' N
Longitude	071°09.666' W	071°09.686' W	071°05.888' W	071°05.916'W
Elevation (m.a.s.l.)	709m	706m	723m	723m
*A.b*coverage	N.A.	30.00%	90.00%	70.00%
Accompagnying tree species	*Picea mariana*	*Betula papyrifera*	*Picea mariana*	*Picea mariana* *Picea glauca*

Table 3.2: Results of PERMANOVA with 999 permutations of the Jaccard dissimilarities for fungal mOTU community structure (DF, degrees of freedom; SS, sum of squares; MS, mean sum of square; F, pseudo-F by permutation)

	DF	SS	MS	F	R2	P
Stands	3	2.138	0.7127	2.4447	0.0547	0.001
Ages	3	2.525	0.84163	2.887	0.06459	0.001
Stands:Ages	9	2.942	0.32691	1.1214	0.07527	0.121
Residuals	108	3.485	0.29153		0.80545	
Total	123	39.09			1	

Table 3.3: Kruskal-Wallis and Dwass Steele Critchlow Fligner tests of the average relative abundance of taxonomic assignments among stands (A) and needle cohorts (B). Significant differences ($p \leq 0.05$) are highlighted in green. Alphabetical order follows decreasing values. * indicates data were subjected to ANOVA and Tukey tests instead of Kruskal-Wallis and DSCF tests

Taxonomic assignment	DF	Nb samples	X2	P	Lav01	Lav02	LaC01	LaC02
Ascomycota *	3	124	6.4441	4.403x10-3	b	b	a	b
Basdiomycota *	3	124	2.6209	0.05389	a	a	a	a
Chytridiomycota	3	124	2.875	0.4113	a	a	a	a
Zygomycota	3	124	1.8753	0.5987	a	a	a	a
Fungi *	3	124	2.4315	0.06845	a	a	a	a
Dothideomycetes *	3	124	2.3818	0.07288	a	a	a	a
Leotiomycetes	3	124	14.239	2.597x10-3	b	a	a	ab
Sordariomycetes	3	124	1.7525	0.6253	a	a	a	a
Eurotiomycetes	3	124	10.432	0.01523	a	b	ab	b
Lecanoromycetes *	3	124	6.2884	5.335x10-4	a	ab	b	a
Taphrinomycetes	3	124	4.7082	0.1945	a	a	a	a
Pezizomycetes	3	124	13.8870	3.063x10-3	ab	a	b	a
Saccharomycetes	3	124	3.2535	0.3542	a	a	a	a
Orbiliomycetes	3	124	6.8822	0.07575	a	a	a	a
Pseudeurotiaceae	3	124	7.0115	0.0715	a	a	a	a
Knufia	3	124	11.09	0.01125	a	a	a	a
Microcyclospora	3	124	4.1120	0.2496	a	a	a	a
Alatosessilispora	3	124	2.2613	0.5200	a	a	a	a
Archaeorhizomycetes	3	124	8.4986	0.03676	a	a	a	a
Chalara	3	124	3.1333	0.3715	a	a	a	a
Digitodochium	3	124	2.3863	0.4962	a	a	a	a
Lichinomycetes	3	124	2.0186	0.5686	a	a	a	a
Pseudoveronaea	3	124	1.9195	0.5893	a	a	a	a
Tumularia	3	124	2.875	0.4113	a	a	a	a
Vestigium	3	124	1.0679	0.7848	a	a	a	a
Unclassified Ascomycota	3	124	10.305	0.01614	ab	b	a	ab
Agaricomycetes	3	124	16.2560	1.005x10-3	a	ab	c	bc
Tremellomycetes	3	124	1.7296	0.6304	a	a	a	a
Microbotryomycetes	3	124	1.4858	0.6856	a	a	a	a
Exobasidiomycetes	3	124	1.7867	0.6178	a	a	a	a
Cystobasidiomycetes	3	124	1.6417	0.65	a	a	a	a
Malasseziales	3	124	5.8296	0.1202	a	a	a	a

Pucciniomycetes	3	124	6.7363	0.08079	a	a	a	a
Agaricostilbomycetes	3	124	1.9278	0.5875	a	a	a	a
Dacrymycetes	3	124	2.8750	0.4113	a	a	a	a
Ustilaginomycetes	3	124	2.8750	0.4113	a	a	a	a
Unclassified Basidiomycota	3	124	4.5246	0.2101	a	a	a	a
Chytridiomycetes	3	124	2.875	0.4113	a	a	a	a
Mortierellales	3	124	1.8753	0.5987	a	a	a	a

Taxonomic assignment	DF	Nb samples	X2	P	Current year	1-year-old	2-year-old	3-year-old
Ascomycota *	3	124	2.1203	0.1012	a	a	a	a
Basdiomycota *	3	124	1.3485	0.262	a	a	a	a
Chytridiomycota	3	124	3.1333	0.3715	a	a	a	a
Zygomycota	3	124	4.946	0.1758	a	a	a	a
Fungi *	3	124	0.286	0.8354	a	a	a	a
Dothideomycetes *	3	124	0.6088	0.6105	a	a	a	a
Leotiomycetes	3	124	12.009	7.353x10-3	b	ab	ab	a
Sordariomycetes	3	124	2.8635	0.4132	a	a	a	a
Eurotiomycetes	3	124	5.2899	0.1518	a	a	a	a
Lecanoromycetes *	3	124	0.8853	0.4509	a	a	a	a
Taphrinomycetes	3	124	2.5964	0.4581	a	a	a	a
Pezizomycetes	3	124	10.2150	0.01682	a	a	b	ab
Saccharomycetes	3	124	0.12405	0.9888	a	a	a	a
Orbiliomycetes	3	124	9.5932	0.02236	ab	a	a	b
Pseudeurotiaceae	3	124	0.5224	0.9139	a	a	a	a
Knufia	3	124	5.3365	0.1487	a	a	a	a
Microcyclospora	3	124	1.9880	0.5749	a	a	a	a
Alatosessilispora	3	124	7.7383x10-3	0.9998	a	a	a	a
Archaeorhizomycetes	3	124	3.7470	0.2901	a	a	a	a
Chalara	3	124	2.875	0.4113	a	a	a	a
Digitodochium	3	124	1.1418	0.767	a	a	a	a
Lichinomycetes	3	124	2.1509	0.5417	a	a	a	a
Pseudoveronaea	3	124	29.1	2.133x10-6	b	b	a	b
Tumularia	3	124	2.875	0.4113	a	a	a	a
Vestigium	3	124	1.0651	0.7855	a	a	a	a
Unclassified Ascomycota	3	124	7.2251	0.06506	a	a	a	a
Agaricomycetes	3	124	7.8207	0.04987	a	a	a	a
Tremellomycetes	3	124	1.0258	0.7950	a	a	a	a
Microbotryomycetes	3	124	5.4581	0.1412	a	a	a	a
Exobasidiomycetes	3	124	12.309	6.397x10-3	ab	ab	a	b
Cystobasidiomycetes	3	124	2.615	0.4549	a	a	a	a
Malasseziales	3	124	9.8709	0.0197	a	a	a	a
Pucciniomycetes	3	124	6.504	0.0895	a	a	a	a
Agaricostilbomycetes	3	124	5.8631	0.1185	a	a	a	a
Dacrymycetes	3	124	3.1333	0.3715	a	a	a	a
Ustilaginomycetes	3	124	3.1333	0.3715	a	a	a	a

Unclassified Basidiomycota	3	124	1.6824	0.6408	a	a	a	a
Chytridiomycetes	3	124	3.1333	0.3715	a	a	a	a
Mortierellales	3	124	4.946	0.1758	a	a	a	a

Conclusion

While the term “endophyte” was coined nearly 150 years ago (De Bary 1866), knowledge about fungal forest endophytology derives mostly from culture-dependent studies from the late 1970s to the early 1990s. The molecular-based approach supplemented and eventually partially replaced culture-based studies later on, but it is with the development of High-Throughput Sequencing techniques in the late 2000s and the hype surrounding microbiota research that it was demonstrated that the culture-based method only recovered a small portion of the endophyte community as most fungi are not amenable to standard culturing. Yet, despite this change of paradigm in term of diversity, most of the observations from culture-dependent studies remained to be disputed either to infer or nuance the conclusions drawn from them. The work presented here proposed thus an update on some critical points shaping the tree endomycobiota using a combination of High-Throughput Sequencing technique, stringent bioinformatics analyses, and biodiversity analyses with *Abies balsamea* endomycobiota as support.

The main purpose of the work presented here was to develop a better understanding of the structure of tree endomycobiota, and to infer the variation in composition of this community at a small geographical scale. To reach this goal, the first objective was to develop a more stringent analysis of 454-pyrosequencing data in order to provide a more conservative view of the fungal endophytic community diversity and assembly. The first chapter thus presented a reinforcement of putative chimeric sequence detection, and a better account of potential undetected erroneous sequences. Then, the second objective was to compare the diversity, community composition, and structure of the endomycobiota from needles and from other surrounding aerial tissues. In the second chapter, we employed a simple design over different neighbouring forest stands to test whether describing only the foliar endomycobiota is sufficient to characterize the overall aerial fungal endophytic community. Finally, the third chapter aimed to compare the diversity, community composition, and structure of the endomycobiota among different needle cohorts and observe the changes as time of exposure to the environment increased. This chapter presented results that are important for understanding the dynamics and persistence of tree foliar endomycobiota over time, as well as holding potential impacts for the use of fungal endophytes notably as biological control agents against foliar tree pathogens.

1. Stringent data analysis.

While limits inherent to High-Throughput Sequencing techniques have previously been addressed and methods have been developed to prevent either artefacts from amplification (Schloss et al., 2011) or from data analysis (Quince et al., 2009; Huse et al., 2010; Quince et al., 2011), it is accepted that even with state of the art data analysis, undetectable erroneous sequences remain and inflate the recovered biodiversity. Our first chapter, which aimed at assessing a conservative estimation of the endomycobiota biodiversity and its structure, provided an original approach for dealing with these undetectable putative spurious sequences from fungal metabarcoding data. Our study relied on using a portion of the more conserved ribosomal small subunit ahead of the Internal Transcribed Spacer 1 to buffer potential sequencing errors. The main results of the first chapter include: (1) an increased sensitivity in the detection of putative chimeric sequences (30 fold rate increase); and (2) the significant diminution of rare molecular operational taxonomic units (-38% of singletons and -22.7% of doubletons) thus limiting the inflation in biodiversity usually observed without interfering with the more represented mOTUs (-4.9% of mOTUs composed of at least three sequences).

Assessing the biological reality of sequences is a problem which emerged with the development of culture-independent methods. This is exacerbated by the relative low level of completion of the ITS database: of the 2.2 - 3.8 million fungal species estimated, 120 000 are currently described, and only 34 878 of the latter are referenced with sequences in Genbank (Hawksworth & Lücking 2017). Removal of any sequences during data treatment is therefore a delicate issue, and exercising caution in doing so is usually recommended. With the exception of the detection of putative chimeras, our approach offers a compromise, as the sequences identified as potentially undetectable erroneous are not deleted from the dataset but appear to join the sequences they deviated from in more abundant mOTUs. The similarity threshold for the clustering based on the pSSU-ITS1 fragment was not adjusted from the full length ITS fragment as one of its limitations as a barcode is that no single cut-off value can be determined to demarcate intra- from interspecific variability across the fungal kingdom (Nilsson et al., 2008). Yet the 95% value which is commonly used was retained to conservatively assess the biodiversity as it impacted mostly rare occurring mOTUs, a characteristic of spurious taxa. We acknowledge the limitation of our method, but due to the nature of the problem, a more definitive grasp of the situation could only be reached by achieving a more comprehensive completion of fungal ITS reference databases, a drawback shared by all culture-independent approaches.

2. Foliar vs Aerial

Leaves are a critically important surface of exchange between plants and the environment. With an estimated global area of ca. 5 x 10^8 km^2 (Vorholt 2012), they dominate the phyllosphere but also fungal endophytology studies, to a point where observations from this tissue prevail for the whole aerial system. While fungal communities associated with other aerial parts received some attention, these studies focused mostly on the epiphytic component (in the case of bark, and buds), or relied on a more restricted functional-based approach (wood decaying fungi and their enzymatic activities, for instance). Our second chapter, in which we compare the biodiversity and composition of the endophytic community associated with every tissue of balsam fir branches, provided a more comprehensive view of tree endomycobiota. Our study design took advantage of the possible standardization of samples offered by HTS methods for comparing immediately adjacent tissues collected from the branches (a similar attempt with culture-dependent method would require different incubation conditions depending on the tissue, hence influencing the diversity yielded and impairing comparison between tissues). The central results of this second chapter include: (1) the presence of distinctive communities associated with the different aerial parts, both in terms of diversity and composition, leading to a more composite view of the aerial endomycobiota of trees; (2) the possibility of different strategies in the recruitment of the fungal community according to the tissue considered.

Communities from needles displayed the highest diversity but distinct communities were detected among the different tissues despite their immediate proximity. This suggests a precise selection of the communities, and possibly different mechanisms to regulate their composition depending on the tissue. Different dynamics seem at play too, possibly influenced by differences of exposition to the environment. Our results provide a more fragmented view of the tree endomycobiota and argue in favour of studying each tissue both separately and altogether in order to characterize their biodiversity but also to determine the factors shaping these communities. Such an approach would also improve the recovery of species composing the endomycobiota, sometimes referred to as the hidden diversity. Considering the proximity, it is likely that communities from different tissues interact too. A more complete comprehension of the different mechanisms involved in the recruitment and regulation of the endomycobiota must comprise an understanding of such mechanisms within the different tissues.

3. Time of exposure

Fungal endophytes display interesting potential functions for their host. Yet, contrary to grass endophytes which are transmitted vertically (inherited) and colonize the plant systemically, the transmission of fungal tree endophytes is horizontal and the colonization more limited. In order to take advantage of such functions, it is necessary to develop beforehand a better understanding of the colonization process, and the persistence and dynamics of the community within the host tissue over time. Our third chapter, in which we characterized the differences in biodiversity and community composition among the last four cohorts of needles, provided a more dynamic view than the apparent species accumulation seemingly hinted by culture-dependent studies. Our study design allowed us to assess the impact of additional years of exposure to the environment on the dynamics of the endomycobiota. The central results of this third chapter include: (1) the observation of a relatively stable diversity over time contradicting an expected enrichment with increased exposure to the environment; (2) the presence of distinct communities associated with the different cohorts.

As change in community composition is also observed with seasonal variations (Hata et al., 1998; Unterseher et al., 2007; Guo et al., 2008), this study suggests that the endomycobiota is neither fixed nor only increases from simple accumulation upon initial colonization. Rather, its composition evolved with time, likely to respond to the plant needs toward the environment. With a relative stability of diversity over time, it is plausible that fungal species displaying different ecological roles partially supersede each other, re-balancing the community as a result of the competition between already settled fungi and new colonizers. A certain lack of homogeneity in the community composition was also observed within the same cohort, possibly explained by the highly localised infections by tree fungal endophytes in leaves that are usually restricted to a ca. 2 mm^2 area. Switching the focus to the individual needle as sampling unit might allow a better understanding of the cohesion or disparity of the endomycobiota within the cohort.

4. Limits

While HTS techniques have allowed for a more extensive detection of the fungal endophyte community than culture-dependent studies, they are not without drawbacks. If culture-based studies were limited by the amenability of fungi to grow on the medium and under the conditions of incubation selected, determination of the biodiversity with HTS is dependent on the efficiency of the primers selected. With barely 1% of the

estimated fungal species both described and referenced in databases, the degree of species coverage across the fungal kingdom by existent primers is debatable and it is expected that HTS techniques do not permit a fully comprehensive retrieval of the endomycobiota. Moreover, DNA metabarcoding is restricted to determining associations over interactions as it does not allow distinguishing the active microbiota from species for which the DNA is present and amplifiable. However, complementation with RNA metabarcoding might discriminate this active biodiversity (Stecher et al., 2016).

The primary aim of metabarcoding is to identify the species present in the environmental samples collected, and afterwards to allow one to perform further ecological analyses such as studying patterns of distribution for instance. This key part is thus highly dependent on the quality and level of completion of a reference database. As mentioned earlier, taxonomic assignment is particularly an issue in the fungal kingdom as of the 2.2 - 3.8 million expected species, only 120 000 are currently described, and 34 878 of the latter are referenced molecularly. Although initiatives such as Unite (Koljalg et al., 2004) allow in some cases a more accurate annotation of the mOTUs by providing a curated reference database, more often, information about the taxa recovered depends on the details filled in the metadata of Genbank "uncultured fungus" entries (which, in some cases, is the only possibility to infer the potential biological reality of the mOTUs recovered).

Focus on a single component of the plant microbiota (endomycobiota for instance) has also its limitations and might hinder a more global understanding of the contribution of the different taxa to the host. The most cited example of benefits offered by endophytic fungi is perhaps the symbiosis observed between *Curvularia protuberata* and the tropical panic grass *Dichanthelium lanuginosum* which allows both organisms to grow at high soil temperatures in Yellowstone National Park (Redman et al., 2002). While initially attributed to the fungus, the heat tolerance is actually conferred by an endohyphal virus found within *Curvularia protuberata* (Marquez et al., 2007).

5. Perspectives

There is no doubt that the development of HTS techniques has contributed to unveil an important part of microbial diversity previously inaccessible with culture-based studies. Paradoxically, and rather than leading to the demise of specimen-dependent approaches, it has also provided a strong argument for reinforcing the use

of such methods. While barcoding of herbarium specimens is a necessity for advancing the level of completion of reference databases and potentially providing more information for taxonomic assignment of data issued from metabarcoding studies, the adoption of a High-Throughput Culturing approach might also prove to be beneficial if only to address the biological reality of certain sequences issued from culture-independent studies (and infer primer efficiency across the fungal kingdom). A specimen-based approach also removes the restriction to a single-locus identification and eventually allows for genome sequencing.

While the prior proposition might be considered a technological step back, a step forward would be a multi "-omics" approach. For instance, the combination of metagenomics with metametabolomics and metaproteomics might allow a better grasp of the functional potential and activities of the retrieved community in addition to attempts to characterize its composition with metabarcoding. Metatranscriptomic studies might also be considered for investigating if the difference in gene expression profiles between different tissues of the trees explains or relates to some extent to the difference in the community composition observed. The same approach could be used on trees of the same species but from different genotypes to determine which genes might be involved in the recruitment and regulation of the endophytic fungal community associated with these trees.

A better apprehension of the recruitment mechanisms also requires a better understanding of the available pool of fungal species surrounding the host in the local environment. Thus monitoring fungal communities retrieved from the soil, air, precipitations but also the epiphytic component associated with the targeted host might provide supplementary informations, notably on the life cycle of the members of the endomycobiota. Monitoring the community composition shifts alongside to the variations of different environmental variables, expected to reflect both biotic and abiotic stresses on the host, might hint towards possible biological roles of yielded species. It is likely that a sampling strategy less sparse over time (semimonthly rather than annually) and the establishment of time series protocols would permit not only a better view of the dynamics of the endomycobiota but also allow a better recovery of the fungal endophyte community and diminish the overall hidden diversity of fungi. Future studies will need to widen their focus to all the components of the plant microbiota across different kingdoms, and disentangle the interaction networks among all these microorganisms, with the host, but also with the environment. They will then have to identify which part of the microbiota constitutes the holobiont (i.e. the host plant and the myriad of microorganisms which confer additional functional traits), and how the hologenome (combination of the host genome with those of the

previously mentioned microorganisms) manages to modulate both its expression and composition to enhance the potential of adaptability of this multiorganismal network to environmental changes.

Bibliography

Ahlholm, J. U., Helander, M., Henriksson, J., Metzler, M., & Saikkonen, K. (2002). Environmental conditions and host genotype direct genetic diversity of Venturia ditricha, a fungal endophyte of birch trees. *Evolution*, *56*(8), 1566-1573.

Arnold, A. E., Maynard, Z., Gilbert, G. S., Coley, P. D., & Kursar, T. A. (2000). Are tropical fungal endophytes hyperdiverse?. *Ecology letters*, *3*(4), 267-274.

Arnold, A. E., & Herre, E. A. (2003). Canopy cover and leaf age affect colonization by tropical fungal endophytes: ecological pattern and process in Theobroma cacao (Malvaceae). *Mycologia*, *95*(3), 388-398.

Arnold, A. E., & Lutzoni, F. (2007). Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots?. *Ecology*, *88*(3), 541-549.

Arnold, A. (2008). Hidden within our botanical richness, a treasure trove of fungal endophytes. *The Plant Press*, *32*, 13-15.

Bacon, C. W., Porter, J. K., Robbins, J. D., & Luttrell, E. S. (1977). Epichloe typhina from toxic tall fescue grasses. *Appl. Environ. Microbiol.*, *34*(5), 576-581.

Bálint, M., Tiffin, P., Hallström, B., O'Hara, R. B., Olson, M. S., Fankhauser, J. D., ... & Schmitt, I. (2013). Host genotype shapes the foliar fungal microbiome of balsam poplar (*Populus balsamifera*). *PLoS One*, *8*(1), e53987.

Bills, G. F. (1996). Isolation and analysis of endophytic fungal communities from woody plants. *Endophytic fungi in grasses and woody plants*, 31-65.

Bischoff, J. F., & White, J. F. (2005). Evolutionary development of the Clavicipitaceae. *Mycology Series*, *23*, 505.

Blackwell, M. (2011). The Fungi: 1, 2, 3… 5.1 million species?. *American journal of botany*, *98*(3), 426-438.

Boberg, J. B., Ihrmark, K., & Lindahl, B. D. (2011). Decomposing capacity of fungi commonly detected in *Pinus sylvestris* needle litter. *Fungal Ecology*, *4*(1), 110-114.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual review of plant biology*, *64*, 807-838.

Cassidy, J. R., Moore, D., Lu, B. C., & Pukkila, P. J. (1984). Unusual organization and lack of recombination in the ribosomal RNA genes of Coprinus cinereus. *Current genetics*, *8*(8), 607-613.

Cordier, T., Robin, C., Capdevielle, X., Fabreguettes, O., Desprez-Loustau, M. L., & Vacher, C. (2012). The composition of phyllosphere fungal assemblages of E uropean beech (F agus sylvatica) varies significantly along an elevation gradient. *New Phytologist*, *196*(2), 510-519.

De Bary, A. (1866). Morphologie und physiologie der pilze, flechten und myxomyceten. W. Engelmann.

Deckert, R. J., Melville, L. H., & Peterson, R. L. (2001). Epistomatal chambers in the needles of Pinus strobus L.(eastern white pine) function as microhabitat for specialized fungi. *International Journal of Plant Sciences*, *162*(1), 181-189.

Eaton, C. J., Cox, M. P., & Scott, B. (2011). What triggers grass endophytes to switch from mutualism to pathogenism?. *Plant Science*, *180*(2), 190-195.

Farr, D.F., & Rossman, A.Y. Fungal Databases, Systematic Mycology and Microbiology Laboratory, ARS, USDA.

Friesen, M. L., Porter, S. S., Stark, S. C., von Wettberg, E. J., Sachs, J. L., & Martinez-Romero, E. (2011). Microbially mediated plant functional traits. *Annual review of ecology, evolution, and systematics*, *42*, 23-46.

Garbary DJ, Macdonald KA. 1995. The Ascophyllum polysiphonia
Mycosphaerlla symbiosis. 4. Mutualism in the Ascophyllum mycosphaerella
interaction. Botanica Marina 38: 221–225.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology*, *2*(2), 113-118.

Guo, L. D., Huang, G. R., & Wang, Y. (2008). Seasonal and tissue age influences on endophytic fungi of *Pinus tabulaeformis* (Pinaceae) in the Dongling Mountains, Beijing. *Journal of integrative plant biology*, *50*(8), 997-1003.

Hardoim, P. R., Van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., ... & Sessitsch, A. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews*, *79*(3), 293-320.

Hata, K., & Futai, K. (1996). Variation in fungal endophyte populations in needles of the genus Pinus. *Canadian Journal of Botany*, *74*(1), 103-114.

Hata, K., Futai, K., & Tsuda, M. (1998). Seasonal and needle age-dependent changes of the endophytic mycobiota in *Pinus thunbergii* and *Pinus densiflora* needles. *Canadian Journal of Botany*, *76*(2), 245-250.

Hata, K., & Sone, K. (2008). Isolation of endophytes from leaves of *Neolitsea sericea* in broadleaf and conifer stands. *Mycoscience*, *49*(4), 229-232.

Hatcher, P. E. (1990). Seasonal and age-related variation in the needle quality of five conifer species. *Oecologia*, *85*(2), 200-212.

Hawksworth, D. L. (1991). The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological research*, *95*(6), 641-655.

Hawksworth, D. L., & Lücking, R. (2017). Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiology spectrum*, *5*(4).

Hebert, P. D., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313-321.

Hoveland, C. S. (1993). Importance and economic significance of the Acremonium endophytes to performance of animals and grass plant. *Agriculture, ecosystems & environment*, *44*(1-4), 3-12.

Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, *12*(7), 1889-1898.

Johnston, P. R., Sutherland, P. W., & Joshee, S. (2006). Visualising endophytic fungi within leaves by detection of (1→ 3)-ß-d-glucans in fungal cell walls. *Mycologist*, *20*(4), 159-162.

Jumpponen, A. R. I., & Trappe, J. M. (1998). Dark septate endophytes: a review of facultative biotrophic root-colonizing fungi. *The New Phytologist*, *140*(2), 295-310.

Kõljalg, U., Larsson, K. H., Abarenkov, K., Nilsson, R. H., Alexander, I. J., Eberhardt, U., ... & Pennanen, T. (2005). UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, *166*(3), 1063-1068.

Krings, M., Taylor, T. N., Hass, H., Kerp, H., Dotzler, N., & Hermsen, E. J. (2007). Fungal endophytes in a 400-million-yr-old land plant: infection pathways, spatial distribution, and host responses. *New Phytologist*, *174*(3), 648-657.

Laforest-Lapointe, I., Paquette, A., Messier, C., & Kembel, S. W. (2017). Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature*, *546*(7656), 145-147.

Lodge, D. J., Fisher, P. J., & Sutton, B. C. (1996). Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. *Mycologia*, 733-738.

Mandyam, K., & Jumpponen, A. (2005). Seeking the elusive function of the root-colonising dark septate endophytic fungi. *Studies in mycology*, *53*, 173-189.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... & Dewell, S. B. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376.

Márquez, L. M., Redman, R. S., Rodriguez, R. J., & Roossinck, M. J. (2007). A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *science*, *315*(5811), 513-515.

Newton, A. C., Fitt, B. D., Atkins, S. D., Walters, D. R., & Daniell, T. J. (2010). Pathogenesis, parasitism and mutualism in the trophic space of microbe–plant interactions. *Trends in microbiology*, *18*(8), 365-373.

Nilsson, R. H., Kristiansson, E., Ryberg, M., Hallenberg, N., & Larsson, K. H. (2008). Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary bioinformatics*, *4*, EBO-S653.

O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J. M., & Vilgalys, R. (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Applied and environmental microbiology*, *71*(9), 5544-5550.

Osono, T., & Mori, A. (2005). Seasonal and leaf age-dependent changes in occurrence of phyllosphere fungi of giant dogwood. *Mycoscience*, *46*(5), 273-279.

Osono, T. (2006). Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. *Canadian Journal of Microbiology*, *52*(8), 701-716.

Osono, T. (2008). Endophytic and epiphytic phyllosphere fungi of Camellia japonica: seasonal and leaf age-dependent variations. *Mycologia*, *100*(3), 387-391.

Périé, C., de Blois, S., Lambert, M. C., & Casajus, N. (2014). *Effets anticipés des changements climatiques sur l'habitat des espèces arborescentes au Québec*. Ministère des Ressources naturelles du Québec.

Petrini, O. (1991). Fungal endophytes of tree leaves. In *Microbial ecology of leaves* (pp. 179-197). Springer, New York, NY.

Porras-Alfaro, A., & Bayman, P. (2011). Hidden fungi, emergent properties: endophytes and microbiomes. *Annual review of phytopathology*, *49*.

Price, D. T., Alfaro, R. I., Brown, K. J., Flannigan, M. D., Fleming, R. A., Hogg, E. H., ... & Pedlar, J. H. (2013). Anticipating the consequences of climate change for Canada's boreal forest ecosystems. *Environmental Reviews*, *21*(4), 322-365.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., ... & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature methods*, *6*(9), 639.

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, *12*(1), 38.

Rajala, T., Velmala, S. M., Tuomivirta, T., Haapanen, M., Müller, M., & Pennanen, T. (2013). Endophyte communities vary in the needles of Norway spruce clones. *Fungal biology*, *117*(3), 182-190.

Redman, R. S., Sheehan, K. B., Stout, R. G., Rodriguez, R. J., & Henson, J. M. (2002). Thermotolerance generated by plant/fungal symbiosis. *Science*, *298*(5598), 1581-1581.

Rodriguez, R. J., White Jr, J. F., Arnold, A. E., & Redman, A. R. A. (2009). Fungal endophytes: diversity and functional roles. *New phytologist*, *182*(2), 314-330.

Rout, M. E. (2014). The plant microbiome. *Genomes of Herbaceous Land Plants*, *69*, 279-309.

Russell, P. J., Wagner, S., Rodland, K. D., Feinbaum, R. L., Russell, J. P., Bret-Harte, M. S., ... & Metzenberg, R. L. (1984). Organization of the ribosomal ribonucleic acid genes in various wild-type strains and wild-collected strains of Neurospora. *Molecular and General Genetics MGG*, *196*(2), 275-282.

Saikkonen, K. (2007). Forest structure and fungal endophytes. *Fungal Biology Reviews*, *21*(2-3), 67-74.

Sandberg, D. C., Battista, L. J., & Arnold, A. E. (2014). Fungal endophytes of aquatic macrophytes: diverse host-generalists characterized by tissue preferences and geographic structure. *Microbial ecology*, *67*(4), 735-747.

Santamaría, J., & Bayman, P. (2005). Fungal epiphytes and endophytes of coffee leaves (Coffea arabica). *Microbial Ecology*, *50*(1), 1-8.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS one*, *6*(12), e27310.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., ... & Miller, A. N. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241-6246.

Schulz, B., Römmert, A. K., Dammann, U., Aust, H. J., & Strack, D. (1999). The endophyte-host interaction: a balanced antagonism?. *Mycological Research*, *103*(10), 1275-1283.

Schulz, B., & Boyle, C. (2005). The endophytic continuum. *Mycological research*, *109*(6), 661-686.

Sieber, T. N. (2007). Endophytic fungi in forest trees: are they mutualists?. *Fungal Biology Reviews*, *21*(2-3), 75-89.

Stecher, A., Neuhaus, S., Lange, B., Frickenhaus, S., Beszteri, B., Kroth, P. G., & Valentin, K. (2016). rRNA and rDNA based assessment of sea ice protist biodiversity from the central Arctic Ocean. *European journal of phycology*, *51*(1), 31-46.

Taylor, D. L., Hollingsworth, T. N., McFarland, J. W., Lennon, N. J., Nusbaum, C., & Ruess, R. W. (2014). A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs*, *84*(1), 3-20.

Unterseher, M., Reiher, A., Finstermeier, K., Otto, P., & Morawetz, W. (2007). Species richness and distribution patterns of leaf-inhabiting endophytic fungi in a temperate forest canopy. *Mycological Progress*, *6*(3), 201-212.

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, *10*(12), 828.

Wille, P. A., Aeschbacher, R. A., & Boller, T. (1999). Distribution of fungal endophyte genotypes in doubly infected host grasses. *The Plant Journal*, *18*(4), 349-358.

Wilson, D. (1995). Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos*, 274-276.